HELLENIC REPUBLIC
National and Kapodistrian University of Athens

"ALEXANDER FLEMING"
Biomedical Sciences Research Center

The role of lncRNAs in cellular homeostasis and carcinogenesis

by

Vathrakokoili-Pournara Anna, BSc

Submitted in the fulfilment of requirements of the degree of Master of Science in Molecular Biomedicine at the Medical School, National and Kapodistrian University of Athens and BSRC Fleming

Pantelis Hatzis Group

Institute for Fundamental Biomedical Research (IFBR)

Biomedical Sciences Research Center ‘Alexander Fleming’

July 2020

ABSTRACT

The canonical WNT pathway plays a pivotal role in stem cell maintenance and affects the expression profile of target genes that are implicated in development, cell proliferation and self-renewal of the intestinal epithelium. Constitutive and deviating activation of the pathway, caused by prominent mutations in its main components, such as APC, AXIN and β-catenin, is among the primary factors leading to the onset of Colorectal Cancer. Thus far, the mechanisms involved in WNT-dependent regulation of the target genes, due to their apparent complexity, remain elusive. Nonetheless, solid evidence indicates that a plethora of targets and regulatory elements of the effector β-catenin/TCF4 transcriptional complex reside in non-coding regions, reporting the existence of long non-coding RNAs that mediate transcriptional responses.

Our research, surprisingly, has identified lnc-IGSF9, a lncRNA whose expression is down-regulated in colorectal cancer patients and also negatively regulated by the β-catenin/TCF4 canonical pathway. Overexpression of lnc-IGSF9 in CRC cell lines increases the expression of genes that control cell adhesion, ultimately promoting cell differentiation. Amongst the up-regulated genes, we distinguished IGSF9, a neighbouring gene that resides in the vicinity of lnc-IGSF9, as a putative mediator of lnc-IGSF9 function. IGSF9 is a cell adhesion protein, whose reduced expression strongly correlates with Colorectal Cancer manifestation. Further functional experiments demonstrate that lnc-IGSF9 and IGSF9 follow a similar expression pattern, with lnc-IGSF9 affecting directly IGSF9 expression but not vice versa. Analysis of ChIP-seq experiments against β-catenin in CRC cell lines revealed two strong binding sites in the locus of interest, particularly in the promoter and in a putative enhancer of lncIGSF9. Excision of these two loci and subsequent expression analysis of the clones showed that the absence of the putative enhancer does not impact lnc-IGSF9 expression, which indicates that its importance in the mechanism may be minor. On the other hand, excision of the promoter resulted in hardly no expression of lnc-IGSF9 as expected, however little impact on IGSF9 expression was observed attributed to its already low expression in CRC cell lines.

In parallel, we investigated the role of CDX2 transcription factor, a protein that was identified as interactor of lnc-IGSF9 in RNA pull-down experiments. Re-analysis of publicly available ChIP-seq data in LS174T cells demonstrated the global binding pattern of CDX2. Specifically, CDX2 strongly binds to both the promoter and the putative intronic enhancer region of lnc-IGSF9 implying its role on the regulation of lnc-IGSF9 expression. Follow-up experiments clearly show that *in-trans* overexpression of CDX2, yields the upregulation of lnc-IGSF9 and IGSF9 as well, proposing a regulatory mechanism between those three components which may be subverted during carcinogenesis. In conclusion, further investigation should be done in order to delineate the exact mechanism of action of lnc-IGSF9 by thorough examination of their putative mediators CDX2 and IGSF9 as well.

ACKNOWLEDGEMENTS

During my master's in Molecular Biomedicine, hosted by BSRC Fleming, I had the opportunity to meet and interact with creative people, highly motivated by their research who helped me expand my knowledge and develop my personal and scientific skills. This thesis is the outcome of one-year research at the BSRC Fleming at Pantelis Hatzis Group.

Before I get into the thick of the things, I would like to thank some people who played a pivotal role in the completion of this thesis. First of all, I would like to thank deeply my supervisor Dr. Pantelis Hatzis for his valuable guidance and suggestions throughout this far from ordinary year. His strong knowledge and expertise on the field helped me gain new perspective and engage novel ideas and concepts in Molecular Biology. I would like also to thank him for his great support and friendly advice for the next step of my studies.

I would also like to extend my thanks to Dr. Panayiota Kafasla and Dr. Despina Sanoudou, whom I both met during their lectures at my master's course, for happily accepting to be part of the committee as well as for introducing their research and expose us to up-to-date advances.

Special thanks should be given to all my lab partners. Vasiliki Zarkou for her unique guidance and always bright input at every step of this project, as well as, for being patient and always willing to share her knowledge and expertise. Alexandros Galaras for the great collaboration and support throughout the year as well as for motivating and encouraging me to get involved in Bioinformatics. His knowledge sharing and discussions on the topic were really valuable to me. Many thanks also to Dr. Eszter Szantai for her always kind help at the bench and friendly attitude. Finally, I want to thank Marili Vasilogianni, the what-you-wish-for lab mate, for being supportive and helpful during a year that was critical for both of us.

Also, many thanks to Dr. Panagiotis Moulos for the guidance and advice on his field of expertise as well as for learning to act independently, a skill you acquire always the hard way, but totally essential for everyone in research. Moreover, I would like to thank all the members from the Fousteri and Verykokakis Lab for the daily interaction and the positive atmosphere that they bring.

Finally, I would like to thank family and friends for their support and encouragement at every step.

Thank you very much.

Table of contents

List of Figures

Introduction

Results

Introduction

1.1 Mammalian Intestinal Epithelium

1.1.1 Background

The intestinal tract is a widely studied organ with a plethora of functions. The intestinal lumen is lined with a specialized simple epithelium, which is responsible for the primary functions of digestion and water and nutrient absorption. In parallel, it acts as a barrier against luminal pathogens preventing primary infections in the gut. It is also a major player in the regulation of metabolic and immune processes in the human body. All these different functions reflect the complexity of this organ and pinpoint the interplay that exists between the extensive cellular and non-cellular components that make up the intestinal tract including: epithelial cells, immune cells, stromal cells, hormones and neurotransmitters, nutrients, the microbiome and many more.[1,2]

The gut is anatomically divided into the small intestine and the colon. The small intestine can be subdivided into the duodenum, the jejunum, and the ileum, which are covered internally by an epithelial monolayer that is lined with mesenchymal cells. The intestinal epithelium is the most actively self-renewing tissue of the adult human body, and adult mammals in general. It is arranged in a series of finger-like projections into the lumen called villi, and invaginations into the mesenchyme called crypts. The bottom of the small intestinal crypt contains actively cycling adult intestinal stem cells (ISCs), also known as crypt base columnar cells, which sit between the terminally differentiated Paneth cells.[3] The remainder of the crypt is largely occupied by transit-amplifying cells, which are estimated to divide twice a day and after 4-5 divisions they move up the crypt. Cells then can exit the crypt compartment and simultaneously undergo differentiation as they enter the villus, where they become either secretory (goblet, Paneth and enteroendocrine) cells or enterocytes (absorptive cells). Although the colon lacks villi, the organization is roughly the same, except that Paneth cells are not present and differentiated cells occupy a large part of the crypt. [Figure 1-1]

1.1.2 Intestinal stem cells (ISCs)

All intestinal epithelium cells are replaced every 3–4 days in mice and this renewal rate is speculated to be approximately every week in the human colon[4]. This rapid renewal is likely to be important for limiting the amount of damaged epithelial cells due to the many bacteria and toxic chemicals that pass by inside the lumen and which are constantly in direct contact with these cells. The incredible epithelial turnover is sustained by the ISCs. Lineage tracing technology expanded our knowledge of ISCs. In a remarkable study from the laboratory of Hans Clevers, leucine-rich-repeat-containing G-protein-coupled receptor 5 expressing (Lgr5+) cells were demonstrated to function as bona fide stem cells[5]. These Lgr5+ cells, are slender cells

squeezed in between the Paneth cells and were already previously described as crypt base columnar cells (CBCs) by Cheng and Leblond[6].

The intestine demonstrates impressive regeneration potential in case of intestinal injury. Interestingly, depletion of all Lgr5+ ISCs does not result in crypt loss and complete regeneration of the affected crypts occurs[7]. In response to loss of Lgr5+ cells, two cell types are believed to be responsible for replacing the ISC pool and sustaining epithelial homeostasis: 1) slow-cycling, quiescent cells at the + 4 position (also called '+ 4' cells) within crypts and 2) absorptive and secretory progenitors[8,9]. Although, CBC cells display functional marker expression differences based on their location within the crypt bottom, they seem uniformly capable of multipotent behavior, albeit in different circumstances. Two factors seem important for this bidirectional conversion: 1) the intrinsic ability to switch cell fate, e.g. by chromatin remodeling[10], and 2) receiving niche signals for reversibly gaining ISC phenotype and functionality [11]. Recent studies have indicated that even terminally differentiated Paneth cells and late-stage entero-endocrine cells, still have the capacity to switch back to an ISC state, indicating that conceivably any intestinal epithelial cell is equipped with a 'switch-back' potential.

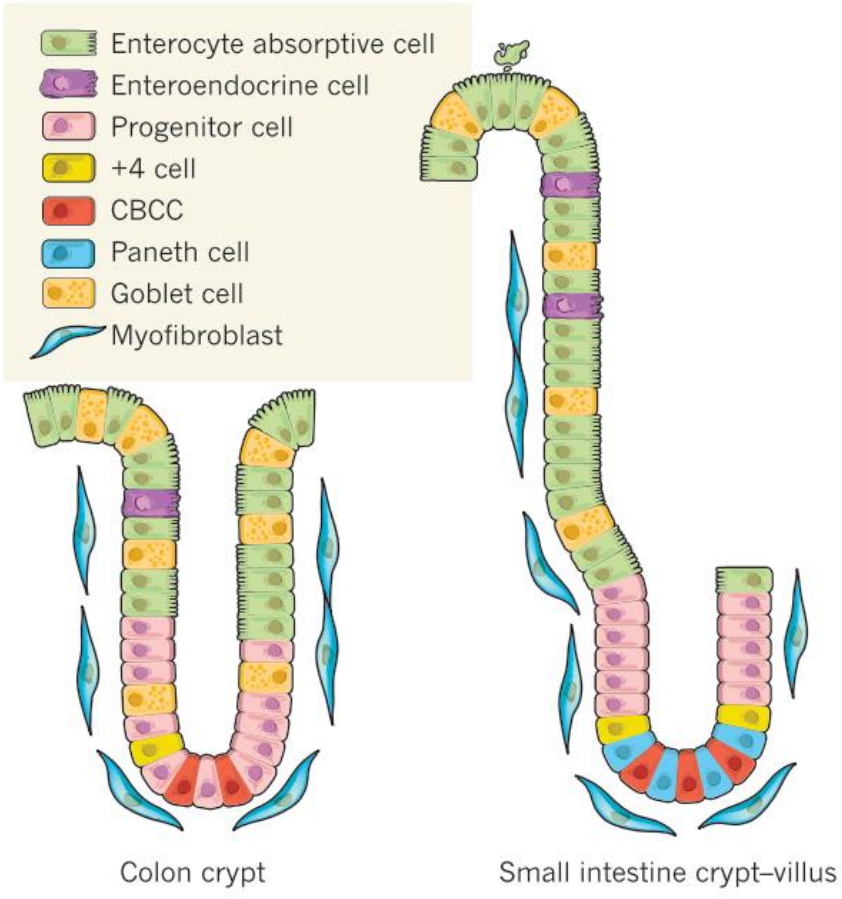


Figure 1-1 The organization of the colon crypt and the small intestinal crypt–villus.
Cellular organisation of the colon(left), Crypts and villi of small intestine (right).As one can observe, Paneth cells are not detected in the colon, yet a Paneth-like cell has been suggested to be present at the crypt bottom. All four lineages (three in the colon) — appear in different but set ratios.

1.1.3 Tumour Initiation in the Intestine

1.1.3.1 Intestinal Cancer Stem Cells

It is generally believed that the main source for mutations in the intestine is due to DNA replication errors and carcinogenic exposure. Given the high proliferation of the Transit-Amplifying (TA) cells and the more restricted zone of proliferation of the ISCs, the TA cells are more susceptible to acquire a mutation. However, these cells would differentiate, migrate to the villi and be shed from the intestine within 3–4 days. As a result, the short lifespan of these proliferating cells reduces the risk of tumour initiation. An additional mechanism to prevent accumulation of mutated cells in the crypt is the neutral drift of the ISCs. Until recently it was believed that ISCs divide by asymmetric cell division, which means an ISC gives rise to one TA cell and one ISC. Instead, the ISC division follows a principle of random replacement of ISCs, a process called 'neutral drift' [12,13]. This means that a single ISC in a crypt can be replaced by any of the other ISCs in the crypt. In a scenario with 5 functional ISCs, a marked wild-type stem cell has a 1/5 (20%) chance to populate the whole crypt and replace all the other ISCs.

Furthermore, Vermeulen et al studied the consequences of oncogenic mutations on ISC fitness. Firstly, they validate that a neutral mutation in an ISC has a high risk of being replaced by a normal stem cell within the crypt due to the stochastic replacement of ISC inside the crypt. Following that, they determined the clonal advantage of lineages harboring either heterozygous or homozygous inactivating mutations of APC gene – a key gene, whose inactivation is closely linked with Colorectal Cancer manifestation. The probability for an APC mutation, to populate the whole crypt is 42%, which means in the majority of cases the mutated cell will be replaced by one of its wild-type stem cell neighbors and will be consequently extinct. Then, they studied the advantage of a Kras mutation (KrasG12D) to populate the whole crypt- the so-called fixation-, which is even higher (about 72%). Although these studies have not considered the different positions of the tracked stem cell at start of the observation, it demonstrates the mutation itself has a major impact on the stem cell fitness. It shows that even if a stem cell acquires a mutation, there is a high chance that the cell will be lost, even if it has an advantage on the stem cell fitness. This might explain why CRC takes years to develop, even in patients with a genetic predisposition (germline $APC^{mut/+}$) to familiar adenomatous polyposis (FAP) disease [14].

1.1.3.2 Colorectal Cancer

Colorectal cancer, or alternatively cancer of the colon and the rectum is the second leading cause of cancer-related death in developed countries, and almost half of the population will develop at least one benign intestinal tumour during their lifetime. Treatment regimens for advanced CRC involve a combination of chemotherapies that are toxic and largely ineffective yet have remained the backbone of therapy over the last decade[15]. CRC may be hereditary or sporadic, accounting for 80% of all patients affected by the disease. The hereditary form is related to two familial syndromes, familial adenomatous polyposis, in which the appearance of multiple intestinal polyps, is observed and hereditary non-polyposis colorectal cancer.

Individuals who develop FAP have a mutation in the APC (adenomatous polyposis of the colon) tumour suppressor gene, whereas those who develop hereditary nonpolyposis colorectal cancer have mutations in genes involved in DNA repair and mismatch repair (MMR) genes[16,17]. Furthermore, the sporadic form, is related to inflammatory bowel conditions such as Crohn's disease and ulcerative colitis, as well as to eating habits such as red meat consumption and low fiber intake[18]. *APC* mutant CRC accounts for more than 600,000 deaths annually worldwide, a number greater than *KRAS* mutant lung or pancreas cancer. Hence, strategies to exploit APC alterations in CRC have broad clinical potential[17]. Moreover, activating mutations in the Wnt pathway initiate the overwhelming majority of CRC cases[19]

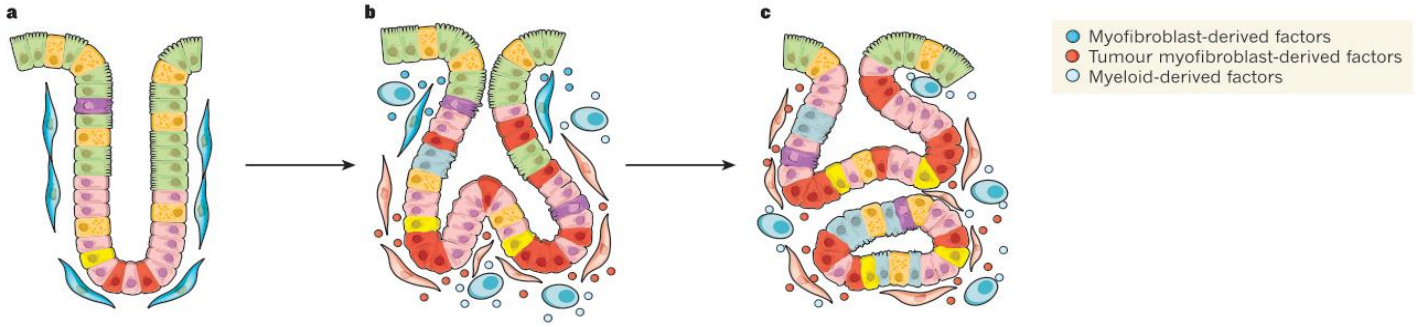


Figure 1-2 Colorectal Cancer Manifestation
a) Normal organization of the intestinal crypt. b) transformation towards an adenoma accompanied with crypt fission and alteration towards inflammation. c) accumulation of further lesions, CRC establishment.

1.1.5 Organoid models

Organoids are three-dimensional (3D) in vitro grown structures derived from adult and embryonic stem cells (ASCs and PSCs respectively) and can self-organize into a near-native microanatomy with organ-specific differentiated cell types and tissue compartmentalization. The first adult stem cell-derived organoid cultures were established from Lgr5-expressing mouse intestinal stem cells that were placed in conditions mimicking the intestinal stem cell niche [Figure 1-3]. By providing R-spondin-1, epidermal growth factor (EGF) and Noggin, and embedment of the cells in an extracellular matrix-providing basement membranes extract, the Lgr5-expressing stem cells received the signals necessary to self-renew, proliferate and form differentiated offspring, resembling the intestinal epithelium[20]. Since then, organoid cultures have been established for a variety of human tissues. Patient-derived tumor organoids, PDTO models show improved resemblance to the original tumour compared to 2D cultured cancer cell lines. Thereby, organoid cultures bridge the gap between in vitro 2D cancer cell line cultures and in vivo PDTXs. Organoid cultures additionally allow for genetic engineering to study the effects of oncogenic mutations in detail. Using the CRISPR/Cas9 technique to manipulate patient-derived organoids can shed ample of light on the impact of specific mutations in cancer onset and progression.

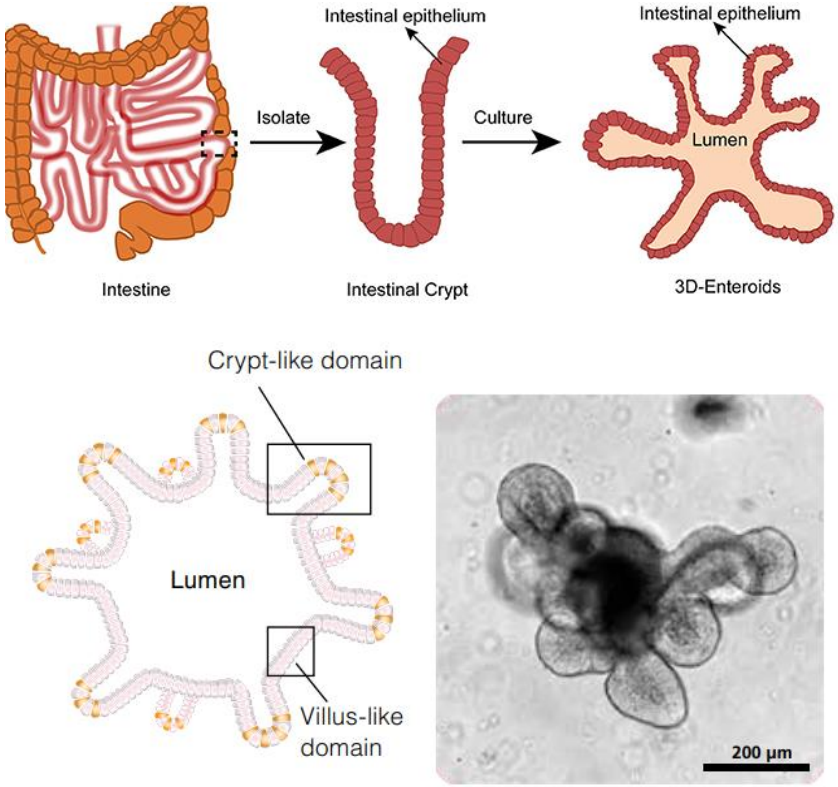


Figure 1-3 Intestinal organoids isolation and morphology
Intestinal crypts can be isolated from surgery sections or biopsies, followed by culturing into 3-D enteroids, the so-called mini-guts (up). Outline of a mature intestinal organoid (down left) and brightfield image of mature (day 5) mouse intestinal organoid (down right). The central lumen is surrounded by an epithelial monolayer with budding crypt-like domains.

1.2 WNT Pathway

1.2.1 Background

In 1973, the wingless(wg) gene was discovered during a mutagenesis screening for temperature-sensitive mutants in *Drosophila melanogaster*[21]. The foundation research for Wnt signal transduction was carried out in the 1980s and 1990s, and it established that the gene products of the Drosophila wingless (wg) and murine proto-oncogene Int1 (now called Wnt1) are orthologous[22]. The name Wnt is an amalgation of wingless and Int1 and stands for "Wingless-related integration site".

Wnts are secreted factors that regulate cell growth, motility, and differentiation during embryonic development, as well as, stem cell proliferation, and tissue homeostasis in adults. They act in a paracrine fashion by activating diverse signaling cascades inside target cells. In particular, Wnts are hydrophobic, and Cys-rich glycolipoproteins whose functions are evolutionary conserved throughout metazoans[23–25]. Abnormalities in the Wnt signaling pathway causes pathological conditions such as birth defects, cancers, and other diseases[26]. In humans, there are 19 genes encoding WNTs that connect to various receptors and stimulate different intracellular signal transduction pathways[27]. These pathways have been classified into

either canonical (β-catenin dependent) or non-canonical (β-catenin independent) signaling pathways. This categorization can only serve as a rough guide, as within canonical and non-canonical WNT signaling various sub-branches are used in different cellular contexts[28,29].Frizzled proteins act as common receptor s for both the β-catenin-dependent and β-catenin-independent pathways, and the use of LRP5 and LRP6 or ROR1 and ROR2 determines β-catenin-dependent and β-catenin-independent pathways, respectively. The complex interplay between receptors and co-receptors determines the downstream effects of signal transduction.

1.2.2 Wnt ligands

Wnt proteins, which are 40 kDa in size and rich in cysteines, are modified by the attachment of a lipid, palmitoleic acid[30–33]. This modification is shared between all Wnts and is brought about by a special palmitoyl transferase: Porcupine[31]. This lipid functions primarily as a binding motif for the Wnt receptor, FZD[34]. The lipid may contribute to restricting Wnt spreading and its range of action turns Wnt proteins hydrophobic. During maturation of Wnts, the transmembrane protein Wntless/Evi (Wls) binds to the lipidated forms and this is required for transferring Wnts to the plasma membrane to become secreted [35–38]. How extracellular Wnt signals are transferred to target cells remains unknown, but available evidence suggests that the proteins are not present in a free form. More likely, Wnt proteins are incorporated into secretory vesicles or exosomes[39,40]. These vesicles contain Wls as well as the mature Wnt signals[40], in such a form that the Wnt protein is present on the outside of the vesicle, available for binding to FZDs. In another model, Wnt transfer involves direct contact between cells mediated by receptors FZD and the transmembrane E3 ligases Rnf43/Znrf3 [41].

1.2.3 The Canonical WNT pathway

The canonical WNT pathway is the best-characterized pathway and its key switch is the cytoplasmic protein β-catenin, whose stability is controlled by a destruction complex. In the absence of Wnt ligands, cytoplasmic β-catenin is constantly degraded by the action of the destruction complex(DC), which is composed of a scaffold protein Axin, the tumour suppressor adenomatous polyposis coli gene product (APC), casein kinase 1 (CK1), and glycogen synthase kinase 3 (GSK3). CK1 and GSK3 sequentially phosphorylate the amino terminal region of β-catenin, resulting in β-catenin recognition by b-TRCP, an E3 ubiquitin ligase subunit, and it subsequent ubiquitination and proteasomal degradation[42]. This continual elimination of β-catenin prevents it from reaching the nucleus. Therefore, WNT target genes, although bound by the T cell factor/lymphoid enhancer factor (TCF/LEF) family of proteins, are repressed because, in the Wnt ‘‘off’’ state, Groucho proteins interact with TCFs, repressing and rendering them inactive.[43]

The WNT/β-catenin pathway is activated when a Wnt ligand binds to the seven-pass transmembrane Frizzled (Fz or Fzd) receptor and its coreceptor, low-density lipoprotein receptor-related protein 6 (LRP6), or its close relative LRP5. The formation of a likely Wnt-Fz-LRP6 complex, together with the recruitment of the scaffold protein Dishevelled (Dvl), results in LRP6 phosphorylation and activation and the recruitment of the Axin complex to the receptors. These events lead to inhibition of Axin-mediated β-catenin phosphorylation

and thereby to the stabilization of β-catenin, which accumulates and travels to the nucleus to form complexes with TCF/LEF ultimately activating the WNT transcriptional repertoire [44] .

1.2.4 Wnt antagonists and agonists

There are several secreted protein families, which antagonize or enhance WNT/ β-catenin signaling. Secreted Frizzled-related proteins (sFRPs) and Wnt inhibitory protein (WIF) both bind to Wnt ligands, and in the case of sFRPs, also to Fz, and therefore function as Wnt antagonists for both canonical and noncanonical signaling[51]. This Wnt-binding property suggests that sFRPs and WIF may also regulate Wnt stability and diffusion/distribution extracellularly beyond just acting as Wnt inhibitors. Some sFRPs have also functions independent of the WNT pathway.

Two distinct classes of Wnt inhibitors are the Dickkopf (DKK) family and the WISE/SOST family. DKK proteins, exemplified by DKK1, are LRP5/6 ligands/antagonists and are considered as specific inhibitors of Wnt/β-catenin signaling. Although two different models for DKK1 action have been proposed[53,54], recent biochemical and genetic studies have argued against a model in which DKK1 inhibits WNT signaling via inducing LRP6 internalization/degradation through transmembrane Kremen (Krm) proteins[53,55,56]. DKK1 disruption of Wnt-induced Fz-LRP6 complex remains a more likely mechanism[57], with Krm playing a minor regulatory role only in specific tissues. WISE and SOST constitute another family of LRP5/6 ligands/antagonists. Like DKK1, SOST can disrupt Wnt-induced Fz-LRP6 complex in vitro[57]. Both DKK1 and SOST are strongly implicated in human diseases. Finally, Shisa proteins represent a distinct family of Wnt antagonists that bind to Fz proteins in the ER and prevent them from reaching the cell surface, thereby inhibiting Wnt signaling in a cell-autonomous way[58].

Norrin and R-spondin proteins are two families of agonists for WNT/β-catenin signaling. Norrin is a specific ligand for FZD4 and acts through FZD4 and LRP5/6 during retinal vascularization. Rspo proteins exhibit synergy with Wnt, Fz, and LRP6 [59], and show genetic interaction with LRP6 during embryogenesis. Rspo genes are often co-expressed with, and depend on, Wnt for expression and may represent a means of positive feedback that reinforces Wnt signaling[58]. Mutations in Norrin and Rspo genes cause distinct hereditary diseases.

1.2.5 B-catenin and TCF/LEF interaction and outputs

Because WNT/β-catenin signaling regulates proliferation, fate specification, and differentiation in numerous developmental stages and adult tissue homeostasis, Wnt target genes are diverse (Vlad et al., 2008) and cell and context specific[45]. An emerging feature is that WNT signaling components are often regulated positively or negatively by TCF/β-catenin [45,59]. Wnt induction of Axin2, DKK1, and Naked and suppression of Fz and LRP6 constitute negative feedback loops that dampen Wnt signaling. On the contrary, Wnt induction of Rspo and TCF/LEF genes constitutes positive feed-forward circuits that reinforce WNT signaling, a feature that has been exploited during colon carcinogenesis[61] . These various WNT pathway self-regulatory loops are mostly utilized in a cell-specific manner.

1.2.6 WNT signaling in CRC

Mutations that promote constitutive activation of the WNT signaling pathway lead to cancer. The best-known example of a disease involving a WNT pathway mutation that produces tumours is familial adenomatous polyposis (FAP), an autosomal, dominantly inherited disease in which patients display hundreds or thousands of polyps in the colon and rectum. This disease is caused most frequently by truncations in ***APC*** [67,68], which promote aberrant activation of the WNT pathway leading to adenomatous lesions due to increased cell proliferation. Mutations in β-catenin and APC have also been found in sporadic colon cancers and a large variety of other tumour types. It is arguable that every colorectal tumour develops, at least in part, as a result of WNT pathway activation.

The best characterized alternatives to APC mutations are activating mutations of ***β-catenin***[69,70].These either delete the whole of exon 3 or target individual serine or threonine residues encoded by this exon. These serines/threonines (codons 45, 41, 33 and 37) are phosphorylated by the degradation complex that contains APC, and hence their mutation causes β-catenin to escape from proteasomal degradation. Mutations at these sites are not found together with APC mutations, showing that alterations in the two genes may be mutually exclusive, although their functional effects are unlikely to be identical given that the C-terminal functions of APC that are removed by almost all pathogenic mutations do not appear to be replicated in the β-catenin protein.

Axin (AXIN1) is an important part of the complex that phosphorylates beta-catenin and hence tags it for degradation. Axin probably acts as a scaffold protein in this context and its loss would be predicted to cause effectively increased WNT signaling. At least two, S215L and L396M, presumably somatic AXIN1 mutations have been found in colorectal cancer cell lines[71]. Further somatic AXIN mutations have been found in sporadic colorectal cancers[72]. In addition, at least some of the colorectal cancers with AXIN1 mutations also have APC mutations, showing that AXIN1 changes cannot be the sole cause of Wnt activation in these tumours.

In some tissues, ***AXIN2*** may substitute for AXIN1 in the β-catenin degradation complex. Mutations in the AXIN2 gene have been reported in colorectal cancers[73], primarily involving insertions or deletions within short, coding oligonucleotide repeats. Although most of these cancers showed nuclear β-catenin expression, no mutations in β-catenin or APC were detected. Almost all mutations in AXIN2 in these cancers were heterozygous, with WT AXIN2 allele present. Intriguingly, germline mutations in AXIN2 have been suggested as predisposing to colorectal cancer[74]. However, the principal phenotype in these families is tooth agenesis.

In addition, ***TCF4*** contains an oligonucleotide repeat tract in the 3' region of the gene that frequently undergoes slippage in microsatellite-unstable colorectal cancers. It has been proposed that this change creates a truncated, more active form of the transcription factor, although TCF4 has multiple splice variants that make functional assessment problematic[72,75] .

Finally, epigenetic changes, such as ***promoter methylation*** in the APC gene or in other pathway members such as WNT inhibitors (DDK, WIF) [76],- have been found as early events in tumourigenesis. Additionally, given the crosstalk and overlap between molecular pathways, especially in cell signaling, it is entirely expected that Wnt activation in colorectal tumours will be influenced by alterations in genes that act

primarily in other pathways. One example is the effect of ***Kras mutations*** – present in 30–40% of colorectal cancers – that can induce WNT signaling in vitro by increasing β-catenin stability, perhaps through inhibition of the kinase activity of GSK-3beta[77].

1.3 Long-non-coding RNAs

1.3.1 Genomics

Much of the non-protein-coding portion of the human genome has historically been regarded as "junk" DNA. However, the continuing development of next-generation sequencing technologies—has allowed an thorough examination of the non-coding genome with unprecedented resolution and scale. It is known that only ~2% of the human genome encodes proteins, but the majority of the gemomw appears to be transcribed under some conditions[78]. Among the various types of non-protein-coding transcripts, a class referred to as long noncoding RNAs (lncRNAs) has attracted increasing attention.

LncRNAs are defined as transcripts of more than 200 nucleotides in length typically ranging from 1000–10,000 nucleotides, with little to no protein-coding potential. That being said, it should also be mentioned that some transcripts annotated as lncRNAs in fact can, encode small peptides as they contain small ORFs[79,80]. LncRNAs resemble mRNAs as they are generally transcribed by RNA polymerase II, 5' capped, 3' polyadenylated, and often undergo splicing via canonical genomic splice motifs[81–83]

There is a broad range of estimates for the number of lncRNA genes in mammals, ranging from less than 20,000 to over 100,000 in humans[84,85]. Nevertheless, the function and biological relevance of the vast majority of lncRNAs remains enigmatic. The existence or production of a transcript does not automatically imply its functionality. Indeed, we must assume until proven otherwise that of the tens of thousands of annotated lncRNAs, those that function independently of the DNA sequence from which they are transcribed may represent a small minority. Importantly, even if a small percentage of lncRNAs are functional, they would still constitute a major gene class with hundreds or possibly thousands of members. Even a scenario in which only 10% are functional implies the existence of more than 1000 human loci generating non-coding RNAs with biological roles. From a variety of screens and expression analyses, it is increasingly evident that changes in the expression levels of many lncRNAs are correlated with developmental processes and disease states such as carcinogenesis, an indication that many of them may carry some functional role. Indeed, since the early studies that demonstrated the central role of Xist in the process of X-chromosome inactivation[86,87], a growing body of evidence has described a myriad of functions for lncRNAs in many cellular processes, such as gene imprinting[88], differentiation and development[89], antiviral response [90], and vernalization in plants [91].

1.3.2 LncRNA Identification

Because of their poly(A) tails and other mRNA-like features, lncRNAs are represented in typical cDNA cloning, tiling array, and RNA-seq data sets. The first large-scale catalog of putatively non-coding transcripts came from the FANTOM project[92,93], which used cDNA cloning followed by Sanger sequencing and reported >34,000 long noncoding RNAs expressed in different mouse tissues, of which 3,652 had confident support[94]. Subsequent studies refined EST- and cDNA-based lincRNA catalogs in mouse and human, which comprise the current RefSeq and Ensembl lncRNA annotations [95,96]. Over the same period, tiling microarrays were used to identify and annotate transcribed regions[97,98] , which was potentially more sensitive than cloning but suffered from reduced dynamic range and difficulties in defining splice junctions and connecting transcribed regions into transcript models. More recently, high-throughput sequencing of millions of short RNA fragments (RNA-seq) is enabling transcript models to be reconstructed, either with the aid of a reference genome[99,100] or without it [101]. RNA-seq can yield billions of strand-specific and possibly paired-end reads, and those can be sufficient for reconstruction of even very lowly abundant transcripts [100,102]. Despite the advantages of RNA-seq in terms of sensitivity and accessibility, assembly of transcript models from short reads still has limitations, because of the relatively small portion of the full transcript accounted for each read and from sequence redundancies in the genome that can yield biases. It remains difficult to distinguish between lncRNAs and fragments of alternative mRNA isoforms or pseudogenes. Focusing only on spliced transcripts helps improve specificity[100] but misses some bona fide single exon lncRNAs, such as Malat1 and Neat1[103]. Therefore, lncRNA databases (e.g., RefSeq and Ensembl) still rely on cDNA sequences obtained using Sanger sequencing[95], but this has started undergoing revisions, as read lengths for high-throughput sequencing methods continue to improve and as multiple data sets are produced and integrated.

1.3.3 Conservation

LncRNAs are poorly conserved across species in terms of their nucleotide sequence. In particular, compared to protein-coding sequences, most of which are highly conserved throughout vertebrates, lncRNA sequences evolve very rapidly. Less than 6% of zebrafish lncRNAs have detectable sequence conservation with human or mouse lncRNAs[104] , and only 12% of human and mouse lncRNAs appear to be conserved in the other species[100,105]. However, many lncRNAs have evolutionary conserved function, secondary structure and regions of short sequence homology[106–108].

1.3.4 Expression Levels

Compared to mRNA expression, lncRNA expression is typically more variable between tissues, with many lncRNAs preferentially expressed in brain and testis[94,95,100]. The tissue-specific expression of a lncRNA is

likely to reflect a particular function in that tissue. LncRNA expression is often unique to specific cell types, developmental time frames, and disease. Also, lncRNAs can be found in different cellular compartments including both the cytoplasm and nucleus to which they seem to predominantly localize [95]. Expression similarity between a lncRNA gene and its closest protein-coding neighbor is generally not greater than that between two adjacent protein-coding genes. The median lncRNA expression level is only about 1/10 of the median mRNA level [94,95,97,109,110].The extent to which the lower level of expression is caused either by less efficient transcription or more efficient degradation of lncRNAs, remains unknown. Two studies, one using a transcription inhibitor and the other using pulse-chase analysis, both concluded that mRNAs and lnc-RNAs have similar half-life distributions[111,112]. Thus, at least the lincRNAs that accumulate to sufficient levels for quantification in such studies are not preferentially destabilized by pathways that degrade aberrant mRNA molecules.

1.3.5 Classification of lncRNAs

There are different ways of categorizing lncRNAs: firstly, lncRNAs can be classified based on their cellular localisation, dividing them into *cytoplasmic* or *nuclear* lncRNA. A second type of lncRNA classification is based on the location at which the lncRNA functions relative to its gene locus. *Trans- acting* lncRNAs are transcribed, processed and translocate to exert their function elsewhere, akin to mRNAs. Their destination does not depend on site of transcription. Accordingly, as long as their levels are properly maintained, transcribing these lncRNAs from a different genomic location or supplanting them into the system should not interfere with their function (as a result their loss of function can be rescued by exogenous expression). A few examples of such lncRNAs have now been extensively characterized[113–115], and many additional lncRNAs have been ascribed trans activities[116–119]. In contrast, *cis- acting* lncRNAs are those whose activities are based at and dependent on the loci from which they are transcribed. Transcripts with the potential of acting *in cis* likely make up a remarkable portion of known lncRNAs: the majority of lncRNAs are enriched in the chromatin fraction and are specifically tethered to chromatin — presumably at their sites of transcription[120]. This may indicate that the actions of these lncRNAs are focused close to their loci. In addition, the fairly low levels at which lncRNAs are generally expressed, sometimes just a few molecules per cell naturally favor an *in cis* mechanism of action [95,102,113]. Finally, lncRNAs can be grouped according to their genomic location. They can reside between in intergenic regions(*intergenic lncRNA*), transcribed from a promoter of a protein-coding gene, yet in the opposite direction (*bidirectional lncRNA*), originate from the antisense RNA strand of another coding gene (*antisense lncRNA)*, or overlap with one or more introns/exons of different protein-coding genes in the sense RNA strand (*sense-overlapping lncRNAs*) [Figure 1-4] [121].

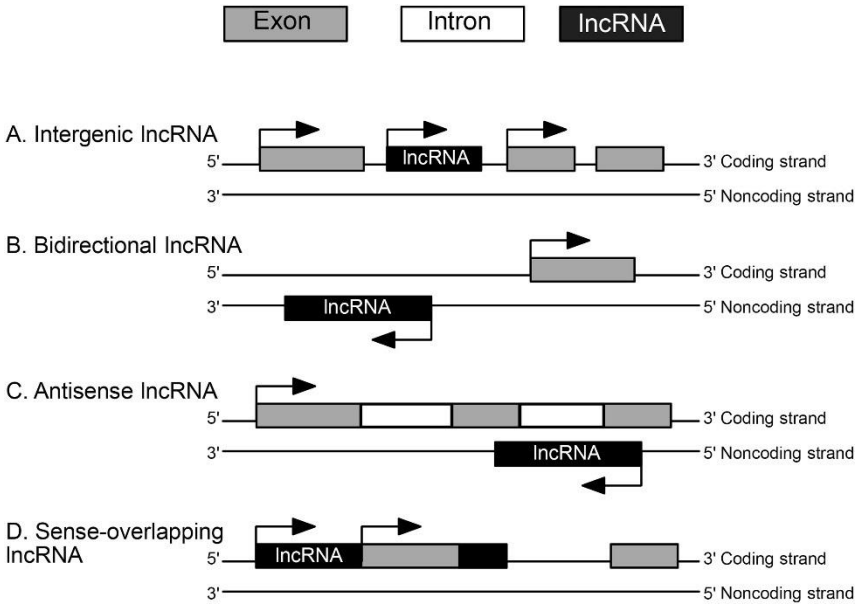


Figure 1-4 | LncRNA classification based on genomic location.
LncRNA classification based on genomic location. **A)** Intergenic lncRNAs are located in intergenic regions. **B)** Bidirectional lncRNAs are transcribed from the same promoter as a protein-coding gene, but in the opposite direction. **C)** Antisense lncRNAs originate from the antisense RNA strand of a gene. **D)** Sense-overlapping lncRNAs overlap with one or more introns and/or exons of a gene in the sense RNA strand direction.

1.3.6 Molecular mechanisms of lncRNAs

Categorising nuclear lncRNAs based on their biological function can be quite challenging due to their large numbers and the multitude of biological processes that they may be involved in. Notwithstanding, lncRNAs can be classified as: guides, dynamic scaffolds and molecular decoys. With these classifications not being mutually exclusive, oftentimes well-studied lncRNAs have been found to act with more than one mode, illustrating the complexity of lncRNA molecular mechanisms[Figure 1-5].

Guides

LncRNA guides are required for the proper localization of factors at specific genomic loci for regulation of the genome. These transcripts bind to regulatory or enzymatically active proteins, such as transcription factors and chromatin modifiers, to direct them to precise locations in the genome. One well-studied guide lncRNA is HOTAIR which functions in *trans* to direct the chromatin modifier Polycomb Repressive Complex 2 (PRC2) to the HOXD genomic locus and, when aberrantly overexpressed, to cancer-related genes, leading to gene repression . The lncRNA MEG3 can recruit PRC2 to target genes via triple-helix formation with the DNA. An example of a lncRNA involved in chromosomal targeting via three-

dimensional organization is Firre. Firre is transcribed from a genomic locus that escapes X chromosome inactivation (XCI). Firre can act *in trans* to form nuclear domains via interactions with hnRNP U to mediate co-localization of multiple chromosomal loci from chromosomes 2, 9, 15, and 17. Additionally, Firre acts *in cis* to help maintain XCI by positioning the inactive X chromosome near the nucleolus while also preserving H3K27me3. Thus, Firre is suggested to convey specificity in the organization of proper chromosomal domains within the nucleus through sequence specific interactions which may serve as a localization signal to initiate or maintain specific nuclear sub-compartments. Specific targeting by guide lncRNAs is stimulated by RNA-DNA, RNA-RNA and RNA-Protein interactions.

Dynamic scaffolds

LncRNAs acting as dynamic molecular scaffolds play structural roles by providing a platform for the assembly of enzymatic complexes and other regulatory co-factors. These often short-lived ribonucleoprotein (RNP) complexes can target specific genomic locations for regulation of gene expression. For instance, telomerase RNA TERC is a paradigm of an RNA scaffold that assembles the telomerase complex, which maintains the ends of telomeres, combining reverse transcriptase activity with telomere targeting proteins in one RNP. TERC served as a useful initial model to test whether newly identified lncRNAs can form stable, homogeneous RNPs. However, little evidence exists for any recently identified lncRNA to act as a stable molecular scaffold like TERC. Instead, lncRNAs may interact with proteins in more dynamic, low-affinity interactions, such as mRNAs during maturation. Dynamic interactions due to lncRNAs are also made with other non-canonical RNA-binding proteins such as chromatin modifying complexes. The lncRNAs TUG1, MALAT1 and ANRIL function as dynamic scaffolds linking chromatin modifying complexes PRC2 and PRC1. In particular, the imprinting-associated lncRNA Kcnq1ot1 scaffolds PRC2 and G9a to promote H3K27me3 and H3K9me3 for targeted genomic repression.

Decoys

The main function of decoy lncRNAs is to limit the availability of specific regulatory factors by acting as molecular sinks or sponges. This class of RNA regulates gene expression by sequestering RNA-binding proteins, transcription factors, microRNAs, catalytic proteins and subunits of larger modifying complexes[122–125]. By keeping these factors away from interacting with their target, decoys act by negatively regulating effector factors. A classic example is the lncRNA PANDA. Upon DNA damage, PANDA associates with the transcription factor NF-YA to prevent p53-mediated apoptosis. NF-YA activates several key genes for apoptosis and cell senescence; however, PANDA binding to NF-YA titrates the latter away from target gene, thereby decreasing expression of apoptotic and senescence-related genes [126,127].

Additionally, several lncRNAs such as MEG3 and TUG1, have been shown to sequester various microRNAs from protein and mRNA targets, resulting in altered protein translation and degradation [128–131]. MicroRNA sponge function in lncRNAs, referred to as the competitive endogenous RNA (ceRNA) hypothesis is highly controversial. It proposes that specific transcripts can impair microRNA activity through sequestration, effectively de-repressing targets of that miRNA. This suggestion has opened a debate, mainly due to the argument that physiological expression levels of individual lncRNAs would not be sufficient to suppress microRNA activity. However, subtle regulation by lowly expressed lncRNAs could be magnified through downstream processes, mainly through the upregulation of transcription factors that signal to multiple effector targets thus amplifying the outcome. Despite the controversy surrounding the ceRNA hypothesis, it is widely acknowledged as a possible generic mechanism for modulating gene expression [130,132,133].

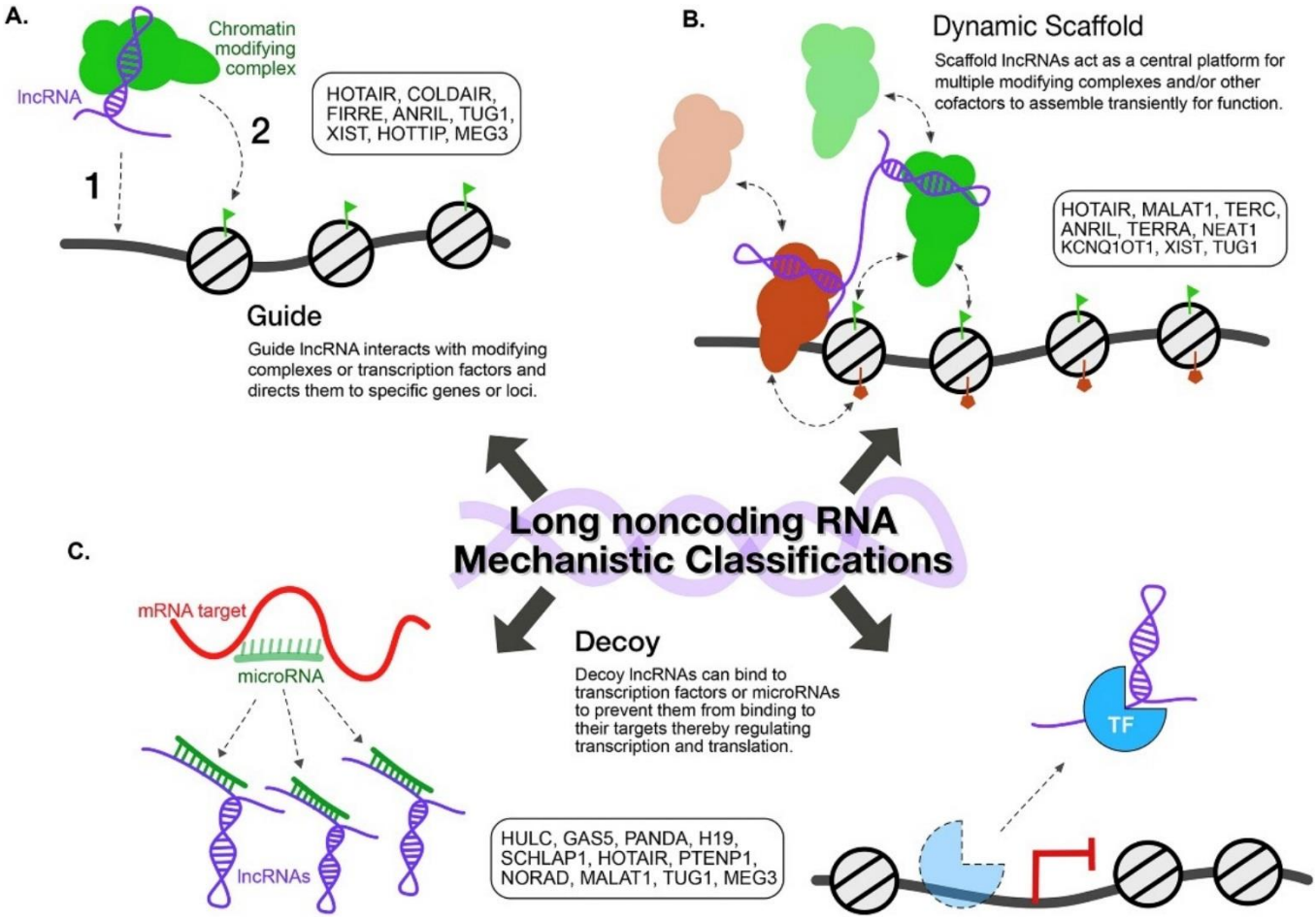


Figure 1-5 | Functional classification for lncRNAs

A) lncRNAs can act as guides to target chromatin-modifying complexes to specific genomic locations for the regulation of gene expression. B) lncRNAs can act as dynamic scaffolds for cofactors to transiently assemble together. C) lncRNAs can bind to microRNAs or transcription factors as decoys to sequester them away from their targets, affecting transcription and translation.

1.3.7 LncRNAs in carcinogenesis

Cancer is primarily caused by genetic alterations that result in aberrant gene expression. Genome-wide association studies in cancer have revealed that more than 80% of cancer-associated SNPs occur in non-coding regions of the genome where, in some cases, cancer-associated loci are transcribed into lncRNAs which play roles in tumorigenesis[128,134–136]. Numerous lncRNAs have been identified to be aberrantly expressed in various cancers in different the tissues and organs[129]. LncRNAs are a very heterogeneous group of transcripts in terms of their mechanism of action. Their expression has been correlated with distinct sets of genes that influence cell cycle regulation, survival, mobility, immune response and pluripotency, among other functions, which contribute to the transformed phenotype of cancer cells [128,134–136].

Many lncRNAs whose cellular roles have been characterized, function as oncogenes that promote tumor growth and are often overexpressed in cancer. HOTAIR is one of the most well-studied oncogenic lncRNAs and was initially characterized as a regulator of the HOX family of genes, which help control cellular identity [137]. However, a more global role for HOTAIR in controlling gene repression through targeting of PRC2 and LSD1/CoREST/REST was uncovered[138,139]. HOTAIR overexpression has been associated with poor outcomes in breast and several other cancers, possibly by increasing metastasis and tumour invasiveness [129,140].

Some lncRNAs act as safeguards against cancer development by preventing proliferation, activating apoptosis, maintaining genomic stability, or promoting the expression of tumor suppressor. For instance, MEG3 is one of the most well-characterized tumor suppressive lncRNAs. In addition to regulating the TGF-b pathway noted above, MEG3 downregulates MDM2 expression and increases p53 protein levels regulation of these and other pathways by MEG3 leads to decreased cell proliferation[141–143].

Many lncRNAs display both tumor suppressive and oncogenic functions. For example, in breast cancer, the lncRNA NKILA negatively regulates nuclear factor kB (NF-kB) signalling and downstream inflammation. In mouse xenograft models human breast cancer cell lines overexpressing NKILA, showed reduced metastasis, pointing to a tumor suppressive role for the lncRNA[144]. However, it was also recently shown that increased NKILA expression can promote tumor immune evasion, an oncogenic property, through induction of cell death of cytotoxic T lymphocytes (CTLs) and TH1 cells[145].

1.3.8 Methodologies of functional characterization of lncRNAs

1.3.8.1 CRISPR/Cas9 KO

The discovery of CRISPR/Cas9, has provided a novel set of tools to solve problems in studying the lncRNA function[146]. CRISPR-based lncRNA knockout is a very effective method to investigate lncRNAs' functionality. Straightforward insertion/deletions created by single double strand breaks are unlikely to cause functional ablations of noncoding genes. Therefore, one should consider a more comprehensive approach with respect to lncRNA function. One strategy which ensures complete ablation would be to delete the entire genomic region associated with a lncRNA by introducing two double strand breaks at the same time. Other approaches could include excision of the promoter, transcription start site(TSS) or one or more exons[147]. This approach is highly effective and has been expanded to genome-wide screens of lncRNA depletion[148]. It should be noted that deletion of the genomic DNA cannot elucidate whether the phenotypic effect is due to loss of the noncoding transcript or potential regulatory sequences in the genomic region itself[149,150].

1.3.8.2 CRISPRa and CRISPRi

Apart from CRISPR knock out (CRISPR KO), the CRISPR-Cas9 system has other diverse applications that can contribute to the elucidation of lncRNA functions. The catalytically inactive/dead Cas9 mutant (dCas9) has been engineered to act as a transcriptional activator or suppressor of the target gene by eliminating both the enzymatic activities of the RuvC and HNH domains of the Cas9 nuclease[151–153]. Although, this mutant Cas9 lacks endonuclease activity, it is still capable of binding to its guide RNA and the DNA strand that is being targeted. Many studies have demonstrated that Cas9 can be fused to different domains to affect activation or repression transcription, and these strategies are now termed as CRISPR activation (CRISPRa) and CRISPR interference (CRISPRi), respectively[152–155].

CRISPRa now is generally used to activate the expression of endogenous human genes and non-coding regions as well. The most effective CRISPRa system appears to be the synergistic activation mediator or SAM system. This system utilizes dCas9 with a sgRNA encoding MS2 RNA aptamers at the tetraloop and the second stem–loop to recruit the MS2 coat protein (MCP) that is fused to two activators, p65 and heat shock factor 1 (HSF1). Additionally, the strong activator Viral Protein 64 (VP64) is fused to dCas9.

Transcription repression by nuclease-deficient Cas 9 (dCas9) can be achieved by fusing dCas9 with different repressor domains, including MAX-interacting protein 1 (MXI1), Krüppel-associated box (KRAB) domain or four concatenated mSin3 domains (SID4X), to either amino or carboxyl termini of dCas9. By using this approach non-coding RNAs can be silenced and thus, the output of their transcriptional repression can be observed.

1.4 Linking lncRNAs to the Wnt pathway

1.4.1 Paradigms of WNT-associated lncRNAs

The WNT signaling cascade is highly conserved among species and controls a plethora of biological processes during embryonic and adult development in animals. It is apparent that abnormal activity of the pathway can result in serious developmental disorders and diseases and, most prominently, in cancer[161]. Although the main molecular players of the pathway have been well characterized, additional mechanistic layers of the regulation of its target genes remains to be delineated. The discovery of long non-coding RNAs (lncRNAs) that are regulated by WNT and/or participate in WNT pathway modulation is particularly interesting and has shed light on the hypothesis that some regulatory components of the pathway are likely to exist in non-coding regions of the genome[162,163] .In fact, recent studies have highlighted the involvement of non-coding transcripts in a variety of key signaling networks [164,165], such as Notch [166], TGFb [167,168], p53 [169] and WNT. From these studies it is obvious that lncRNAs can impact many cellular pathways and biological processes in multiple levels, and most importantly they seem to play significant role in tissue-specific tumorigenesis.

LncRNAs implicated in the WNT pathway act through various mechanisms from the transcriptional to the post-translational level. The human 8q24 gene desert, located upstream of the MYC gene, contains multiple regulatory elements; one of these encompasses the SNP rs6983267, which maps to a functional TCF4 binding site in CRC cells and affects the binding of the WNT-regulated transcription factor TCF4 and the subsequent recruitment of β-catenin[170]. The same region has been shown to express distinct WNT- regulated lncRNAs in different human tumors including CCAT1-L, CCAT1-S, CCAT2 and CASC11[164]. CCAT2 (colon cancer associated transcript 2) is a lncRNA of ~400nt that is transcribed in the sense orientation from the highly conserved 8q24 chromosomal region next to the MYC gene and encompasses the rs6983267 SNP[171]. CCAT2 has been shown to physically interact with the TCF4 transcription factor, increasing its transcriptional activity in CRC[172]. The resulting activation of the WNT/β-catenin- mediated transcriptional machinery results in increased expression of genes involved in genomic instability and excessive cell proliferation, promoting cancer growth[171].

LncRNAs have increasingly been recognized as essential organizers of chromosomal architecture[173]. WiNTRLINC1 (WNT- regulated lincRNA1) is a direct WNT/β-catenin target gene that is located in the vicinity of the *ASCL2* gene locus and participates in the formation of an intra-chromosomal loop[174]. Loss of WiNTRLINC1 results in increased apoptosis and G2-cell cycle arrest in colon cancer cells by decreasing the expression of its neighbor ASCL2 gene, a transcription factor that controls intestinal stem cell maintenance[174]. Chromosome conformation capture experiments (3C) revealed the formation of a chromatin loop bridging the TSS region of *WiNTRLINC1* with an enhancer located immediately downstream of the *ASCL2* locus. Moreover, the WiNTRLINC1 transcript itself was shown to be required for loop formation and *ASCL2* expression by affecting the recruitment of Pol II, TCF4/β-catenin and the Mediator complex to the *ASCL2* regulatory regions. *ASCL2*, in turn also binds to the *WiNTRLINC1* promoter region, activating the expression of WiNTRLINC1 and completing a positive feedback loop. This *WiNTRLINC1/ASCL2* regulatory loop is frequently found amplified in patients with colorectal cancer and is positively correlated with worse

disease state, increased metastatic rate and decreased patient survival[174]. Thus, aberrant WNT pathway activation increases the expression of WiNTRLINC1 and results in ASCL2 activation, potentially enhancing stemness and carcinogenesis in the intestine.

1.4.2 Lnc-IGSF9

1.4.2.1 Lnc-IGSF9 and cancer

LncIGSF9/LINC01133 is a newly identified cancer-associated lncRNA. Recent studies have shown that it can act both as oncogene and tumor-suppressor depending on the tissue and type of cancer. Its oncogenic role has been demonstrated in non-small cell lung cancer, in cervical squamous cell carcinoma, in osteosarcoma, in pancreatic ductal adenocarcinoma (PDAC) and in hepatocellular carcinoma[175–178]. As an oncogenic lncRNA, the high expression of LINC01133 is driven by the transcription factor C/EBPβ and gene amplification in tissues where LINC01133 promotes tumorigenesis through targeting tumor-suppressive miRNAs (i.e., miR-442a) or proteins (i.e., Klf2, E-cadherin, p21, Dkk1) or upregulation of oncogenic mediators (Cyclin G1 and phosphorylated PI3K/AKT). On the other hand, its role as a tumor-suppressor has been observed in oral squamous cell carcinoma, in gastric cancer, in colorectal cancer, in breast cancer and ovarian tumor[177,178]. As a tumor-suppressor, there is a reciprocal inhibition between LINC01133 and pro-EMT TGF-β signaling. In gastric cancer, LINC01133 also inhibited oncogenic WNT/β-catenin signaling via targeting miR-106a-3p[177].

In recent studies on colorectal cancer, LINC01133 was found to physically interact with SRSF6, whose overexpression promoted EMT and metastasis, indicating that LINC01133-mediated regulation of EMT was effected, at least in part, via inhibiting SRSF6. Consistent with its anti-EMT function, clinical sample analysis revealed that LINC01133 expression was negatively associated with vimentin (a mesenchymal marker) and positively associated with E-cadherin (an epithelial marker). Low tumor expression of LINC01133 was correlated with poor survival of CRC patients[179]. These studies indicated that LINC01133 downregulation contributes functionally to CRC progression, where this lncRNA could serve as a prognostic factor. Although these studies have elucidated some of the potential roles of LINC01133, its nuclear roles and exact mechanism of function in normal intestine and colon as well as in colorectal cancer have not been delineated and require further investigation.

Despite its double role in tumorigenesis, deregulation of LINC01133 has been shown to be associated with clinicopathological parameters and patient survival in multiple cancer types, pointing to the prognostic value of this lncRNA. Moreover, the tissue-specific functions of LINC01133 might preclude its systemic inhibition or reactivation for cancer therapy. Development of methods for tissue-specific targeting of LINC01133 is therefore warranted. Taken together, LINC01133 is a tumor-associated lncRNA with clinical potential.

1.4.2.2 Establishing the premises

Our group has identified WNT-regulated lncRNAs by utilizing the CRC cell line Ls174T engineered to overexpress in an inducible manner a small hairpin RNA (shRNA) against β-catenin or a dominant negative mutant form of TCF4, in order to shut down the WNT pathway[172]. ChIP-seq experiments with antibodies against TCF4, β-catenin, H3K4me3 and the largest subunit of RNA polymerase II, along with RNA-seq experiments before and after WNT pathway abrogation, revealed sites of WNT-depended transcription[172]. Among others, lnc-IGSF9 was identified as a gene that is negatively regulated by the WNT pathway. The lnc-IGSF9 gene is located on the long arm of chromosome 1 (hg19 coordinates: chr1:159931014-159948876), consists of three exons and produces a transcript of ~1154 nt. The transcript is 5'-capped and polyadenylated. Based on PhyloCSF software [180] and in vitro transcription/translation experiments, lncIGSF9 does not have protein coding potential. In addition, the lncRNA is localized mainly in the nucleus, as revealed by cell fractionation and FISH experiments. Following that, reanalysis of the RNA-seq data from The Cancer Genome Atlas (TCGA) revealed that lnc-IGSF9 is significantly down-regulated in patient-derived adenomas and carcinomas compared to normal tissue, indicating that it could be used as a predictive biomarker for colorectal cancer. Taking advantage of colorectal cancer cell lines engineered to express the dCas9 protein fused to transcriptional activators we managed to overexpress lnc-IGSF9 in cis (unpublished data). Follow-up RNA-seq experiments demonstrated a long list of differentially expressed genes, some of which potentially directly affected by lnc-IGSF9 induction. Among them is the IGSF9 gene, which the lncRNA was named after, as it is located in the vicinity of the lncIGSF9 locus; IGSF9 was one of the most upregulated genes and was identified as putative target of lnc-IGSF9 (unpublished data).

Moreover, one of the first approaches to further delineate the function of lnc-IGSF9 was to identify lnc-IGSF9-interacting proteins through RNA-pull down experiments coupled with mass-spectrometry. The proteins that were found to interact with this specific transcript are mainly associated with the organization of genome architecture. Interestingly, homeobox protein CDX2, a master regulator of intestinal epithelium differentiation and development, which is also essential for the binding of TCF4 on specific WNT targets, was found to interact directly with the lnc-IGSF9 transcript[181,182]. Finally, after reanalysis of publicly available CDX2 ChIP-seq experiments in LS174 cells we observed strong binding of CDX2 in the promoters of the lncIGSF9 and IGSF9 genes, an indication that CDX2 may be strongly implicated in the regulation of these genes.

Materials and Methods

Cell Culture

LS174T and HT29 colorectal cancer cells were used for the functional experiments of this study. Both cell lines carry APC mutations (active WNT pathway). Cells were cultured in DMEM (Gibco) medium supplemented with 10 % FBS (fetal bovine serum) and a cocktail of antibiotics (Amphotericin, Penicilin-Streptomycin, Gentamicin). Cells were maintained in flasks or culture plates at 37°C degrees and with 5% supplied CO_2 Absence of mycoplasma infection was confirmed regularly (PCR mycoplasma detection kit, ABM).

For WNT activation we utilized HT29 cells. CHIR99021 (a GSK-3 inhibitor) was used to chemically inhibit the GSK-3 kinase an activate the Wnt pathway. At 80% confluency, 24h after plating, HT29 cells were treated with 5 mM CHIR99021. Control cells were treated with DMSO. 24h hours later cells were harvested, and RNA was extracted.

The HEK-LentiX 293T cell line was used for production of lentiviruses. The Lenti-X 293T cell line is a subclone of the transformed human embryonic kidney cell line, HEK 293, which is highly transfectable and supports high levels of viral protein expression. When transfected with the Lenti-X Packaging System and a lentiviral vector, these cells are capable of producing lentiviral titers as high as >108 ifu/ml

Lentivirus Production

For lentiviral production, $40x10^6$ HEK LentiX-293T cells were plated in $10cm^2$ plates, that were previously coated with 1mg/ml collagen (Collagen Type I, Rat tail, Corning) and cultured for 24h. On the second day cells were transfected with a DNA mix consisting of 3.75 μg plasmid vector (shRNA, sgRNA or the sequence that codes for the lncRNA/protein of interest) , 1.32 μg of pMD2.G (Addgene, #12259) plasmid vector that expresses lentiviral envelop proteins and 2.43 μg of psPAX vector that expresses proteins that are necessary for the packaging of the lentiviruses. LentiX-293T cells were transfected with JetPrime (PolyPlus) in a 1:2 ratio (1μg DNA: 2μl JetPrime Reagent). The culture medium was replaced 8h after transfection with fresh medium and cells were incubated for 48h. After checking transfection efficacy by fluorescent microscope, the virus-containing medium was collected and centrifuged at 500g, 4°C for 5 minutes to precipitate dead cells. The virus supernatant was collected and filtered through a 0.45μm sterile filter. Fresh viruses were used directly to transduce cells or stored at -80°C for future use. In order to enhance transduction efficiency, we used the cationic polymer Polybrene (Hexadimethrine bromide, 8μg per ml of culture medium).

In cis overexpression of lnc-IGSF9

In order to overexpress lnc-IGSF9 transcript in cis, we used LS174T colorectal cancer cell line that has been engineered to constitutively express a mutant form of Cas9 (dCas9), fused to VP64 transactivating domains domain as well as the transcriptional activators HSF1 and p65. dCas9 can precisely target sites through hybridization with a guide RNA but has no nuclease activity and as a result it cannot cleavage DNA. Instead it functions as a scaffold that recruits the transcriptional machinery to activate locus-precise gene expression. To achieve lnc-IGSF9 overexpression, two sgRNAs were designed that target the promoter of the gene as well as a non-targeting sgRNA as control (NTG). The sgRNAs were cloned into the pLV-U6-gRNA-diMS2-EF1Alpha-Puro-T2A-BFP vector.

For the transduction experiments, $3X10^5$ cells were plated and simultaneously transduced with 500µl of lentiviruses expressing sgRNA1 or sgRNA2 and NTG as a control. Transduction was assisted with the addition of the cationic polymer Polybrene, which significantly enhances infection efficacy. Transduction efficiency was checked in a fluorescent microscope as the sgRNA vector contains sequences for the expression of Blue Fluorescent Protein (BFP).

Single guide RNAs

Primer name	Target site	Sense oligo	Antisense oligo
sgRNA1	GGAAAATGAGACAAGGAA	GGAAAATGAGACAAGGAA TCTCATTCCTCCCCAACAA	TTAGCTCTTAAACTTCCTTG TCTCATTTTCCCCAACAAG
sgRNA2	GGAAAATGAGACAAGGAA	TTGGGGAGGAAAGCCAAA GAAAGTTTAAGAGC	TTAGCTCTTAAACTTTCTTT GGCTTTCCTCCCCAACAAG
NTG	GGAAAATGAGACAAGGAA	TTGGGACCAGGATGGGCAC CACCCGTTTAAGAGC	TTAGCTCTTAAACGGGTGG TGCCCATCCTGGTCCCAAC AAG

Knock down of β-catenin, lnc-IGSF9 and *IGSF9*

In order to perform knock down experiments, we utilized a lentiviral system for overexpressing shRNAs against target transcripts. shRNAs oligos were designed one for each target sequence (e.g., b-catenin, lnc-IGSF9, IGSF9 and scrambled sequence used as a control in the experiments). The shRNA oligos were cloned into the pLB vector. Stbl2 competent bacterial cells were transformed with recombinant constructs and grew in agar plates with ampicillin. Single colonies were selected, cultured in Luria Broth medium at 38°C for 18 h and DNA was extracted. Successful recombinant constructs were identified after double digestion with the SlaI and XbaI restriction enzymes, generating a fragment of 374 bp compared to a 334 bp fragment from the empty vector.

For the knock-down experiments, $3x10^5$ LS174T cells were plated in 6-well plates and transduced with 300 µl of lentiviruses overexpressing shRNA against β-catenin, lnc-IGSF9, IGSF9 according to the set-up of each experiment. 24 h later cells were washed carefully twice with 10X PBS to remove viruses from their culture medium. 72 h after transduction cells were harvested and whole RNA was isolated.

The shRNA constructs were transfected into HEK-Lenti-X cells to produce lentiviruses.

The shRNA sequences designed for each gene are reported in the following table:

Primer name	Sense oligo	Antisense oligo
Lnc-IGSF9 sh1	TGGGAGGAGGTAAAGAGTAGTTCAA GAGACTACTCTTTACCTCCTCCCTTTT TTC	TCGAGAAAAAAGGGAGGAGGTAAAG AGTAGTCTCTTGAACTACTCTTTACCT CCTCCCA
IGSF9 sh1	TTGGAATTGCTGGAGACTTTTTTTCA AGAGAAAAAAGTCTCCAGCAATTCCA TTTTTTC	TCGAGAAAAAATGGAATTGCTGGAG ACTTTTTTCTCTTGAAAAAAAGTCTC CAGCAATTCCAA
β-catenin sh2	TCCATGGAACCAGACAGAAATTCAAGAGATTT CTGTCTGGTTCCATGGTTTTTTC	TCGAGAAAAAACCATGGAACCAGAC AGAAATCTCTTGAATTTCTGTCTGGTT CCATGGA
Scrambled (scr)	TGTACAGCCGCCTCAATTCTTTCAAG AGAAGAATTGAGGCGGCTGTACTTTT TTC	TCGAGAAAAAAGTACAGCCGCCTCA ATTCTTCTCTTGAAAGAATTGAGGCG GCTGTACA

In-trans overexpression of CDX2

In order to overexpress CDX2, we cloned the flag-CDX2 ORF into the LeGO-iG2 lentiviral vector. The construct N-flag-CDX2-pcDNA3 containing flag-CDX2 ORF was digested with BamHI and SlaI restriction enzymes to purify flag-CDX2 DNA. LeGO-iG2 was digested with the NotI restriction enzyme. The flag-CDX2 fragment was ligated into digested LeGO-iG2 by using T4 ligase (NEB) overnight at 16^{o}C. DH5a competent cells were transformed with the recombinant constructs and grew in agar plates with ampicillin. Single colonies were selected, cultured in Luria Broth medium at 38^{o}C for 18 h and DNA was extracted. Successful recombinant constructs were distinguished by NheI digestion.

3x10^{5} LS174T cells were plated per well and 1ml LeGO-iG2-CDX2 was used for the transduction. LeGO-iG2 expressing virus was used as a control. The experimental procedure and protocol follow the same steps as described in previous sections.

RNA extraction and quantitative PCR

Total RNA was isolated with TRI reagent (Molecular Research Center). The extracted RNA (1 µg) was then used for cDNA synthesis with reverse transcription, utilizing the M-MLV Reverse Transcriptase (Thermo Fisher Scientific) in 20 µl reactions. The produced cDNA was diluted with ddH2O to 400 µl and 4 µl used as a template for each PCR reaction. We performed qPCR with the SYBR-Green PCR master mix (Applied Biosystems) and 1 µM of the forward and reverse primer. We generated duplicates for each reaction for greater reproducibility and we quantified the expression of the house-keeping genes RPLP1 and GAPDH for normalization.

Primer sequences

Primer name	Sequence
RPLP1 F1	AAGCAGCCGGTGTAAATGTTGAGC
RPLP1 R1	CATTGCAGATGAGGCTCCCAATGT
GAPDH F1	ACATCAAGAAGGTGGTGAAGCAGG
GAPDH R1	TGTCGCTGTTGAAGTCAGAGGAGA

lnc-IGSF9 F4	GAATGGTTGGGAGGAGGTAAAG
lnc-IGSF9 R4	CTGGGCTCAAGGAATCTGAATAG
IGSF9 F1	CCGAGAGATGAATGTGGATGG
IGSF9 R1	GGAGAACGAAGGAAAGATGAGG
IGSF9 F3	CTGGCTTTGATGGTGGTTATCT
IGSF9 R3	ACCCAGTCATGGTGCATTC
CTNNB1 F1	TGCAGTTCGCCTTCACTATGGACT
CTNNB1 R1	GATTTGCGGGACAAAGGGCAAGAT
CDX2 F2	TCGGCAGCCAAGTGAAA
CDX2 R2	GATGGTGATGTAGCGACTGTAG
ASCL2 F2	CGCGAGCTACTCGACTTCTCC
ASCL2 R2	GAGCGCGGGCCGGTCCA
WiNTRLINC1 F4	GGAATATTCTCAGAGCTCCAGAGC
WiNTRLINC1 R4	GGCGTGAGGTGTGGACAGCTGCC
AXIN2 F2	CAGCAGAGGGACAGGAATC
AXIN2 R2	CAGTTTCTTTGGCTCTTTGTG
HOXC5 F1	CCCTTTGCTGTCCCATAGTC
HOXC5 R1	AGGAAGGACCCAGAGTCAATA
FOXQ1 F1	CCCAGGCTTCGTCTTATTTCT
FOXQ1 R1	GTGGAAAGGTTCCCTGATGT

CRISPR/Cas9 Knock-out of lnc-IGSF9 promoter and enhancer

2.8.1 Construct Preparation

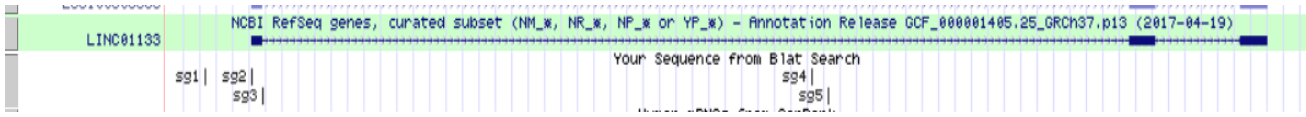
In order to knock-out the promoter and the enhancer of lnc-IGSF9 we used the pX333 vector **(Addgene #64073).** The vector is designed to be used for tandem expression of two sgRNAs from two independent U6

promoters. Cas9 is expressed from the same vector by the Cbh promoter. Four different single guides targeting the lincIGSF9 locus were designed to facilitate the excision of the promoter and the enhancer of lnc-IGSF9. sgRNA1 and sgRNA2 target two sites surrounding the promoter, while sgRNA4 and sgRNA5 target the intronic enhancer of lnc-IGSF9.

The oligo sequences of sgRNA used are listed in the following table and a snapshot from genome browser is provided to illustrate the exact target sites.

	Target sequence	Sense oligo	Antisense oligo
lncIGSF9 sg1 KO	GCCCCATGAAGTTTG CACCT	caccgGCCCCATGAAGTTT GCACCT	aaacAGGTGCAAACTTCAT GGGGCc
lncIGSF9 sg2 KO	AGGCTGTTACATCAG AATTG	caccgAGGCTGTTACATCA GAATTG	aaacCAATTCTGATGTAACA GCCTc
lncIGSF9 sg4 KO	TAAATCACACGCTGC CCACG	caccgTAAATCACACGCTG CCCACG	aaacCGTGGGCAGCGTGT GATTTAc
lncIGSF9 sg5 KO	TTTAGAACAAGCCAA CAAGA	caccgTTTAGAACAAGCCAA CAAGA	aaacTCTTGTTGGCTTGTTC TAAAc

Target sequence sites:



The complementary oligos forming the single guides were annealed and the Px333 vector was digested first with BbsI restriction enzyme and purified. In order to insert the first single guides (sg1 and sg4 respectively), we performed ligation with T4 (NEB) ligase enzyme. DH5a competent bacterial cells were transformed with the recombinant constructs and grown on agar plates supplemented with ampicillin. Single colonies were selected and resuspended in 20 µl ddH20. Next, we screened colonies with the sense primers of sg2 and sg4, respectively and a reverse primer complementary to CMV (part of the vector). Colonies that produced the expected PCR product of 700 bp were selected, cultured and plasmid DNA was extracted. Next, recombinant constructs were digested with BsaI and the second single guides (single guide2 and single guide 5,

respectively), were ligated with T4 (NEB) ligase. DH5a competent bacterial cells were transformed again with the new recombinant constructs and grown on agar plates supplemented with ampicillin. Single colonies were selected and resuspended in 20 µl ddH20. Multiple PCR screenings followed to ensure that both single guides were inserted in each construct.

For the PCR screening we used the sense or antisense oligos of the single guides and the reverse CMV sequence (CTATTGGCGTTACTATTGACG), which is a part of the pX333 plasmid.

2.8.2 Generation of knock-out cell lines

In order to generate stable cell lines that lack the promoter and the enhancer of lnc-IGSF9, LS174T cells were tranfected with 1,8 µg recombinant contructs [1.pX333 plasmid containing single guide1 + singleguide2 (targeting the promoter of lnc-IGSF9) and pX333 containing sg4+ sg5 (targeting the enhancer of lnc-IGSF9)] as well as 0,2 µg pcDNA 6TR plasmid in order to select clones later. Lipofectamine 3000 was used as a transfection reagent in order to achieve optimal transfection efficiency. 8 h after transfection the medium was replaced with fresh DMEM medium to avoid the toxic effects of lipofectamine on the cells. 24 h after transfection the medium was replaced with selection medium (DMEM supplemented with blasticidin at 10 µg / mL); this was removed after 3 days and replaced with DMEM. For the following two weeks the formation of colonies was tracked regularly and 24 clones were selected when they reached the appropriate size. Clones were expanded and DNA was extracted to test for successful KO clones. We used conventional PCR and suitable primers to distinguish between KO, heterozygous and wild-type clones.

The primer sequences and the expected PCR fragments according to the nature of the clone (KO, WT, HET) are listed below:

enhancer PCR screening		Primers	expected bands
lncIgsf9 enh KO screen F	forward	TAGGAAAGAAGCGCACAGAG	
lncIgsf9 enh KO screen R	reverse	CAGGATTGAGGAGTCCAGTTT	
KO			1 at 260bp
HET			2 at 500bp and 260
WT			1 at 500bp
promoter PCR screening		Primers	expected bands
lncIgsf9 prom KO screen F	forward	CTTGCGATAGTTTGCTGAGAATG	
lncIgsf9 prom KO screen R	reverse	CAGATCCAAAGGGAAGCTAAGG	
KO			1 at 290
HET			2 at 290 and 1100
WT			1 at 1100

Downstream analysis of Intestinal Organoids

W utilized an RNA-seq dataset obtained from patient-derived intestinal organoids- in which the WNT pathway was activated in two different ways (through WNT3a/R-spondin treatment or genetic ablating of the APC gene).Differential expression analysis was performed using the DESeq2 tool (versions 1.12.4 and 1.18.1), and paired analysis was performed as well. Differential gene expression was considered significant if log2-fold change was ≥1 or ≤−1 and the adjusted P value was <0.05. Data were annotated with HGNC symbols using biomaRt (version 2.28.0). Differentially expressed genes were filtered again with stricter Basemean threshold (75% of the higher observations was considered). Scatter plot visualization showing the ranking of the differentially expressed genes was performed in the R programming language. Pearson correlation analysis according to log2-fold change of differentially expressed genes was performed with the cor() function in R. Plots were produced again in the R programming environment.

Enrichment analysis

GSEA analysis was performed using the preranked tool (version 2.2.3, Broad Institute). Data visualized using the replotGSEA function from the Rtoolbox package. We generated a ranked list file from differentially expressed genes upon lnc-IGSF9 overexpression. We built a new molecular signature from the organoid dataset: intrinsic (WT + Wnt vs. *APC*-KO + Wnt) regulation, which was used to run GSEA analysis. The customized molecular signature was divided and marked as “UPREGULATED GENES UPON APC KO” and DOWNREGULATED UPON APC KO” in order to create gene sets that contain genes either upregulated or downregulated upon APC KO.

Results

Lnc-IGSF9 and the WNT Pathway

The canonical WNT pathway activates gene expression via the transcription complex β-catenin/ TCF4 which binds to DNA and modulates the expression of its target genes. However, RNA sequencing experiments from genetically modified LS174T cells which express an shRNA against β-catenin in an inducible manner (WNT OFF) and control cells (WNT ON) showed that the expression of the *lnc-IGSF9* gene is reduced upon WNT activation[Figure 3-1]. Accordingly, subsequent ChIP-sequencing experiments in the same cells before and after WNT abrogation, with an antibody against RNA-Poll II demonstrated that its binding on the promoter region of *lnc-IGSF9* is diminished upon WNT activation.

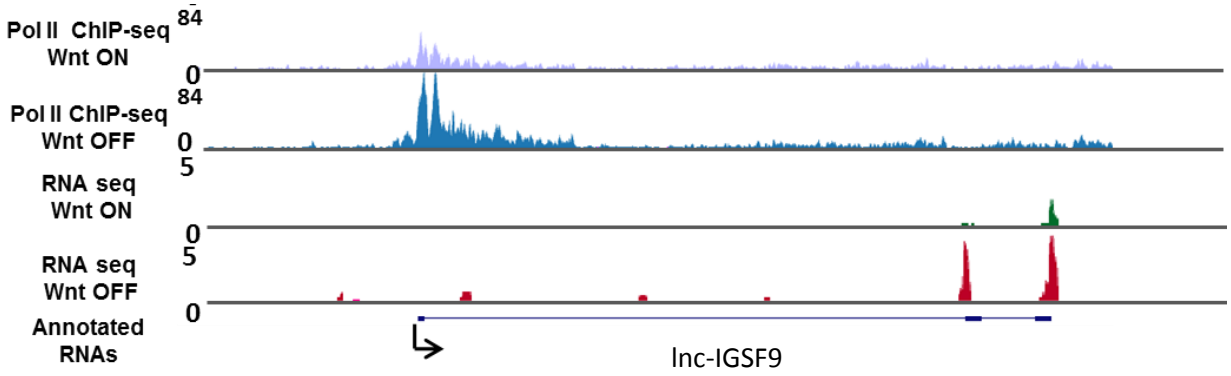


Figure 3-1 Snapshot from UCSC browser depicting RNA-seq and ChIP-seq signals focused on lnc-IGSF9 locus

For WNT pathway abrogation we used a genetically engineered LS174T derived cell line that inducibly overexpresses an shRNA against β-catenin when cells are cultured in the presence of doxycycline. The RNA-seq and ChIP-seq experiments were performed 72h after doxycycline treatment in LS174T cells. The *lnc-IGSF9* locus is illustrated in the bottom of the picture. Bold blue boxes represent the exons and the light blue the introns of the lncRNA. The arrow shows the orientation of its transcription. Signals have been previously normalized.

To confirm that the expression of lnc-IGSF9 is suppressed by Wnt signaling, we performed RT-qPCR experiments on cDNA samples prepared from LS174T cells expressing an shRNA against β-catenin or a mutant form of TCF4, that lacks the N-terminal β-catenin-interaction region (dNTCF4): this mutant form of TCF4 can bind DNA but does not interact with β-catenin, acting as a dominant negative mutant. In both cases, the expression levels of *lnc-IGSF9* increased significantly after WNT pathway abrogation [Figure 3-3,A]. The above pattern of regulation was also confirmed by transducing LS174 cells with lentiviruses overexpressing an shRNA against β-catenin .The levels of lnc-IGSF9 transcript increased markedly in cells overexpressing the shRNA against β-catenin compared to control cells expressing a scrambled sequence [Figure 3-3,B].

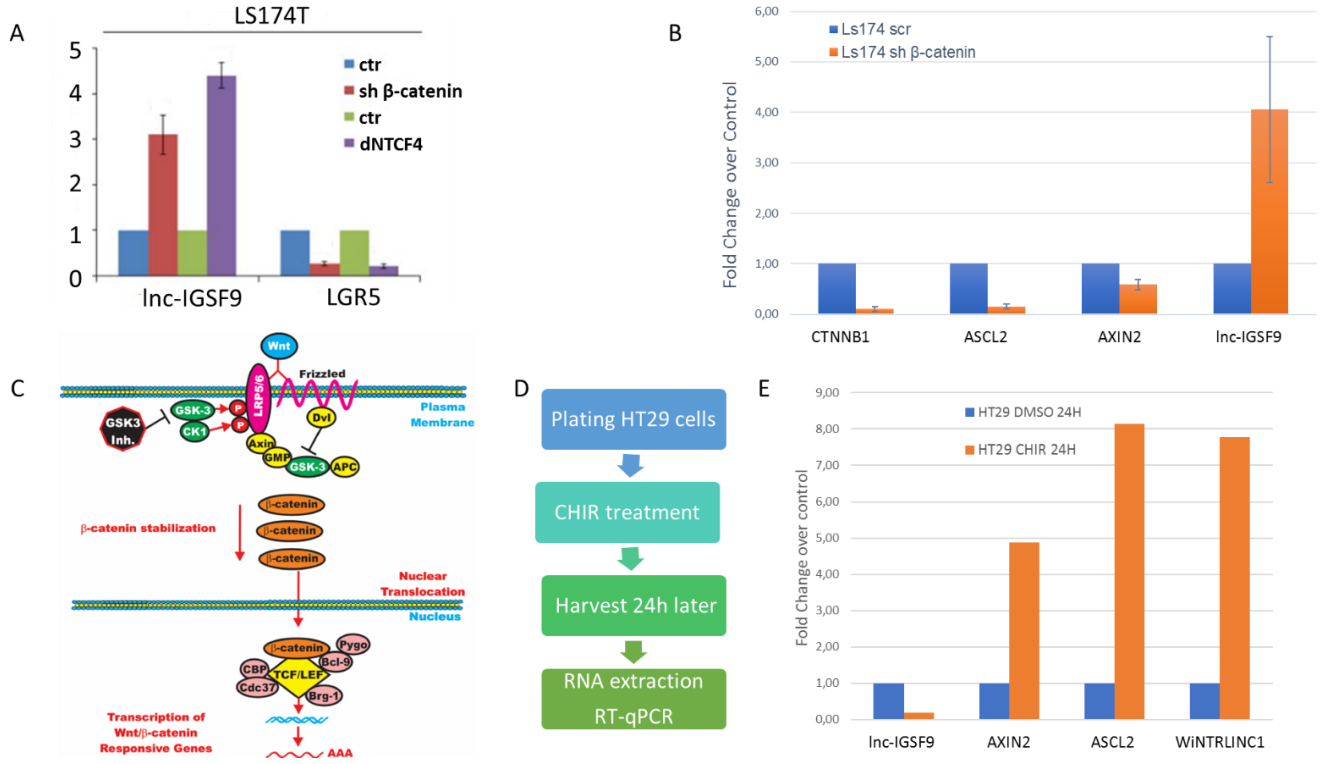


Figure 3-3 lnc-IGSF9 is negatively regulated by the WNT pathway

A. qPCR expression analysis of lnc-IGSF9 in control cells and cells with inducible overexpression of an shRNA against β-catenin or a dominant negative mutant form of TCF4 (dNTCF4). The *LGR5* gene is known to be positively regulated by the WNT pathway and was used as a control. **B.** qPCR expression analysis of lnc-IGSF9 in LS174T cells transduced with lentiviruses which overexpress an shRNA against β-catenin. CTNNB1 expression levels have been checked along with the known WNT-targets AXIN and ASCL2 levels. **C.** Mechanism of function of CHIR99021, which acts to inhibit GSK-3 kinase, ultimately leading to β-catenin accumulation and overall WNT pathway transcriptional program activation (WNT ON state) **D.** Schematic representation of the experimental process that we used for GSK-3 inhibition. 0.5nM CHIR concentration was used for a 24 h. **E.** qPCR analysis of lnc-IGSF9 expression in HT29 cell after WNT activation. AXIN2, ASCL2 and lncRNA WiNTRLINC1 expression levels were inspected to confirm the efficiency of WNT activation.

In addition, we conducted experiments where we activated the WNT pathway, by using the GSK-3 inhibitor, CHIR99021. For these experiments, we took advantage of the human colorectal cancer cell line HT29 which carries a heterozygous mutation in the *APC* gene, rendering the WNT pathway less activated compared to other homozygous *APC* mutant cell lines. Therefore, induction of HT29 cells with CHIR09921 results in further activation of the WNT pathway. HT29 cells were treated with CHIR09921 and harvested 24h later [Figure 3-3,C-D]. RT-qPCR analysis showed increased expression levels of known WNT target genes, such as *AXIN2*, *ASCL2* and *WiNTRLINC1* confirming pathway activation. Moreover, *lnc-IGSF9* expression levels were considerably decreased, demonstrating its negative regulation by the WNT pathway[Figure 3-3,E].

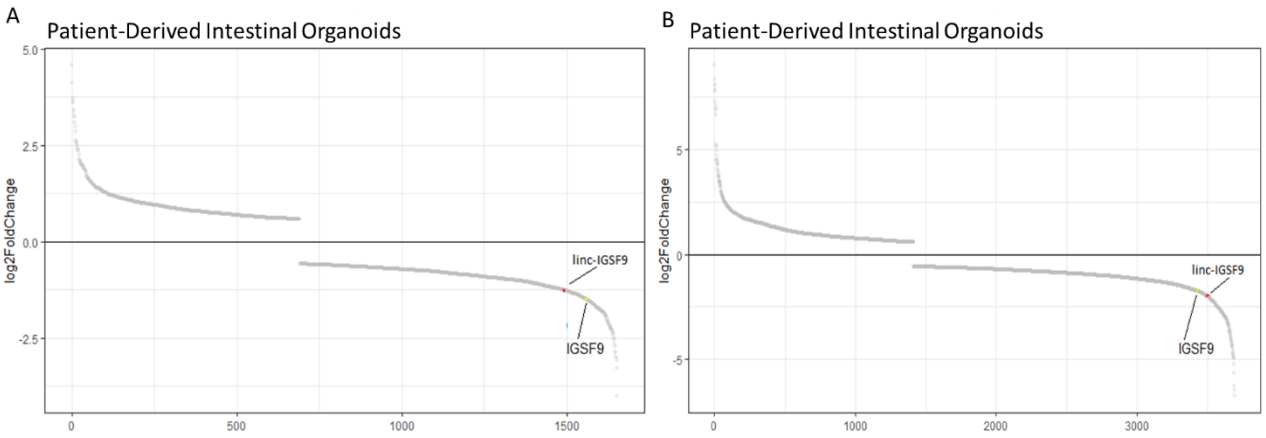


Figure 3-2 Scatter plots that depict the ranking of genes based on differential expression in patient-derived intestinal organoids.

A. Ranking of genes upon WNT activation via the receptor **B.** Ranking of genes upon APC gene deletion (APC KO). With red dot we highlight the *linc-IGSF9* gene.

Moreover, paired differential analysis of RNA-seq data obtained from patient-derived intestinal organoids- in which the WNT pathway was activated in two different ways (through WNT3a/R-spondin treatment or genetic ablating of the APC gene) – revealed lnc-IGSF9 as one of the most down regulated genes upon WNT activation, supporting our previous observations in a model that to some extend recapitulates the in vivo conditions of the intestinal epithelium. Specifically, lnc-IGSF9 is ranked in position 1451 among the 1654 differentially expressed genes that arise after p-value adjusted filtering set at < 0.05 in the WNT3a/Rspondin treatment dataset and in position 3494 among the 3692 differentially expressed genes in the dataset generated after APC deletion organoids**.**[Figure 3-2].

Functional characterization of *lnc-IGSF9*

Because *lnc-IGSF9* is poorly expressed in colorectal cancer cell lines, in order to study its functional role in cancer cells, we performed a gain of function analysis by taking advantage of CRISPR/Cas9 technology. Specifically, we used the CRISPRa system to boost the expression levels of *lnc-IGSF9 in cis* and observe differences at the transcriptional level (qPCR and 3' mRNA-seq). I*n cis* overexpression of lncRNAs is widely used to investigate lncRNAs which act locally by affecting gene expression in cis. Thus, we utilized a cell line derived from LS174T cells that has been genetically engineered in our laboratory to constitutively express a mutant form of Cas9 (dead Cas9, dCas9), which is catalytically inactive in terms of nuclease activity. This dCas9 is coupled with the transcriptional activator VP64 (viral protein 64) and interacts with the transcription factors p65 and HSF1. [Figure 3-4,A] The recruitment of the latter two transcription factors at the target sequence is achieved through their interaction with MS2 aptamers added to the sgRNA sequence. The resulting cell line (LdC10 cells) was transduced with lentiviruses that overexpress either a control sgRNA (Non-Targeting-Guide) or two sgRNAs that each targeted different regions of the *lnc-IGSF9* promoter (sgRNA1, sgRNA2). [Figure 3-4,B]. Therefore, the expression of the sgRNAs led to the targeted guidance of the ribonucleoprotein complex (dCas9-VP64-MS2-p65-HSF1) on the promoter of *lnc-IGSF9.*

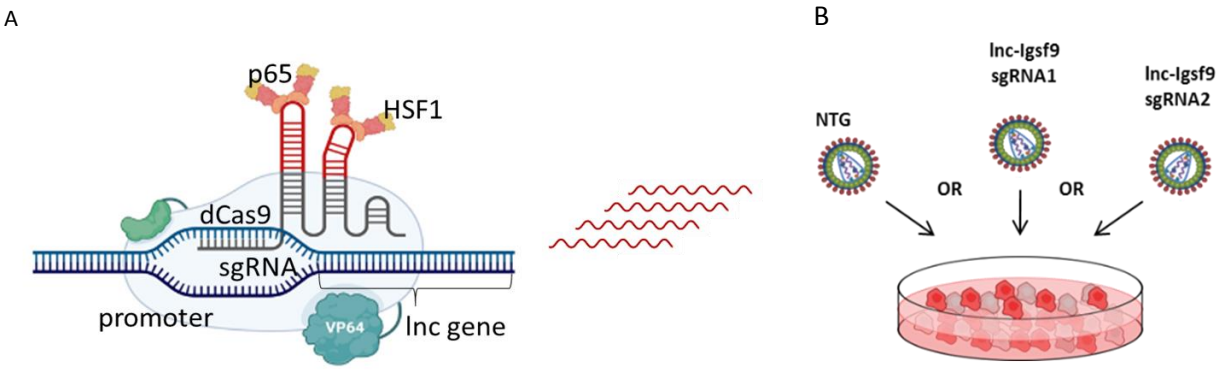


Figure 3-4 CRISPRa mediated in cis overexpression of lnc-IGSF9

A. Outline of the CRISPRa system with the fusion protein stratified in the long non-coding RNA gene locus with the aid of a specific sgRNA. **B.** Transduction of cells with lentiviruses expressing NTG, sgRNA1 and sgRNA2 respectively. 24h after transduction the medium is replaced with fresh one to avoid strong toxic effects from the viruses. Finally, cells are harvested 72h after transduction and whole RNA is extracted.

RT-qPCR analysis showed that lnc-IGSF9 expression levels were increased by approximately 120-180 fold (depending on which guide was used and on the titer of the viruses) compared to the cells that expressed the control sgRNA (NTG) [Figure 3-5,A]. In order to delineate how lnc-IGSF9 expression impacts global gene expression in colorectal cancer cell lines we investigated the transcriptomic profile of LS174T cells that overexpress lncIGSF9 *in cis*, using 3' mRNA-seq[183]. In this method only the 3'UTR of each read is sequenced and as a result signals are observed mainly in the 3' UTR region of each transcript. Differential expression analysis of these cells revealed statistically significant changes in the expression of 1093 genes upon lnc-IGSF9 overexpression. Specifically, we identified 727 upregulated and 366 downregulated genes, respectively, upon *lnc-IGSF9* overexpression (threshold for natural Fold Change >2x, p value< 0.05). Surprisingly, among the differentially expressed genes, we identified a gene located in the vicinity of the *lnc-IGSF9* locus, IGSF9 (from whom the lncRNA was named), whose expression was significantly increased upon *lnc-IGSF9* overexpression. Interestingly, it is the only upregulated gene within a 1MB distance from the *lnc-IGSF9* locus, upon *lnc-IGSF9* overexpression, indicating that its expression may be regulated by *lnc-IGSF9*. [Figure 3-5,B-C]. IGSF9 is a cell adhesion molecule whose function has not been characterized in the intestine. However, there are some studies which have investigated its role in the brain, specifically in the development of dendrites and in the maturation of neuronal synapses.

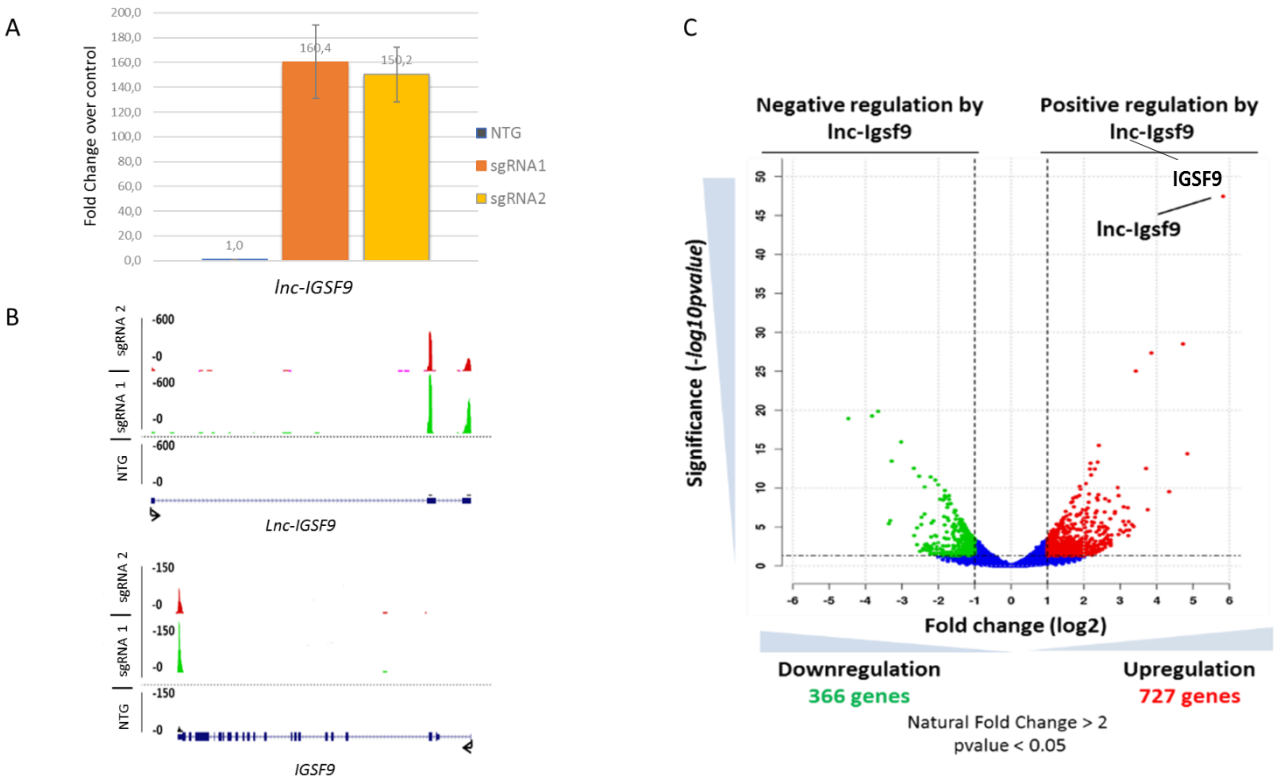


Figure 3-5 *In cis* overexpression of *lnc-IGSF9* impacts gene expression in the LS174T colorectal cell line

A. RT-qPCR analysis from LS174T cells overexpressing dCas9 transduced with control (NTG) sgRNA or two sgRNAs targeting different regions of *lnc-IGSF9* promoter. The bar plot summarizes the average fold change in the expression of *lnc-IGSF9* over control from two independent biological replicates, standard deviation has been also estimated and is illustrated using error bars. **B.** Snapshot from UCSC genome browser, shows Quant-seq signals concentrated on *lnc-IGSF9(up)* and *IGSF9(down) gene locus*. **C.** Volcano plot presenting the de-regulated genes upon *lnc-IGSF9* overexpression in LS174T cells [log2fold change > |1|, pvalue < 0.05 (reported by Deseq algorithm used for the differential expression analysis)]. *lnc-IGSF9* is as expected the most upregulated gene and is pointed in the plot, IGSF9 is also displayed. Down-regulated genes are labelled with green dots, while upregulated genes are reported as red dots. Dots with blue colour represent genes that do not pass the threshold to be regarded as differentially expressed.

To investigate the relationship between *lnc-IGSF9* and *IGSF9* and its connection to the WNT pathway, we performed loss of function experiments. Thus, we designed experiments in which we overexpressed an shRNA against β-catenin in LS174T cells, shutting down the WNT pathway, and we measured the expression levels of *lnc-IGSF9* and IGSF9. The expression of both transcripts was found to be increased about 10 fold and 8 fold compared to control cells, respectively [Figure 3-6, A]. Interestingly, the regulation of IGSF9 by the WNT pathway can also be observed, similar to lnc-IGSF9, after reanalysis of the patient-derived intestinal organoids dataset. IGSF9 is downregulated not only upon WNT activation through the receptor but also upon APC deletion [Figure 3-2,A-B].

To investigate whether is the lnc-IGSF9 transcript itself or the act of its transcription that activates IGSF9 expression, we transduced LS174T cells with two different lentiviruses at the same time, one that overexpressed an shRNA against β-catenin (to increase lnc-IGSF9 transcript levels) and one that overexpressed either an shRNA against the lnc-IGSF9 transcript or an shRNA against the IGSF9 transcript. RT-qPCR analysis demonstrated that both lnc-IGSF9 and IGSF9 expression levels decreased upon lnc-IGSF9 shRNA overexpression, indicating that the expression of IGSF9 is dependent on the lnc-IGSF9 transcript. On the contrary, synchronous transduction of cells with lentiviruses expressing shRNAs against β-catenin and the IGSF9 transcripts showed a significant drop in the expression of IGSF9, while lnc-IGSF9 transcript levels were not decreased [Figure 3-6,B]. This result indicates that IGSF9 is not necessary for lnc-IGSF9 expression.

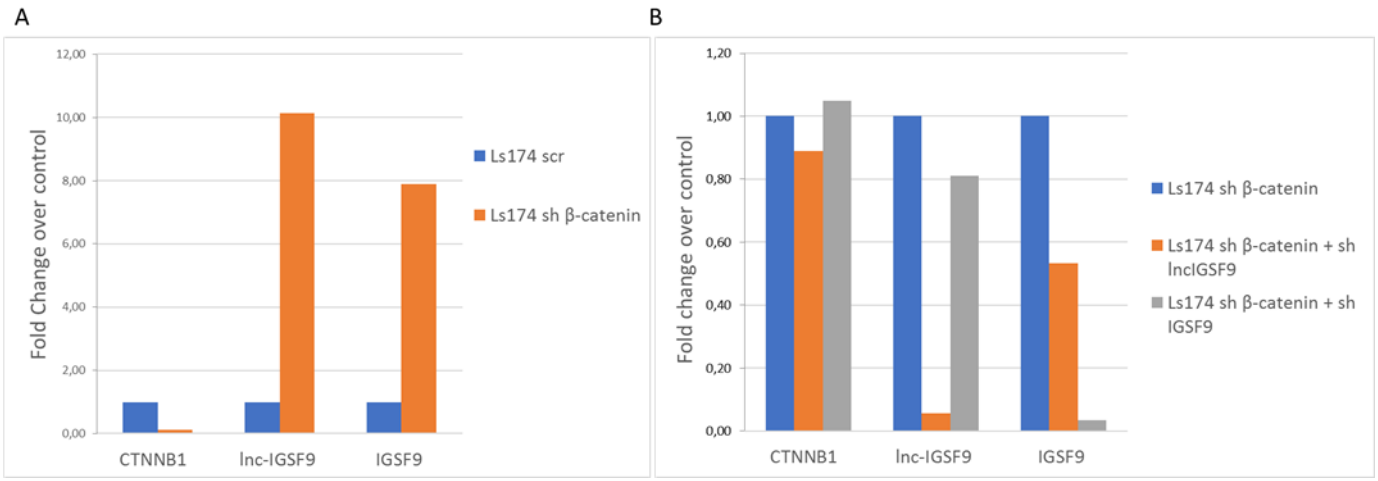


Figure 3-6 lnc-IGSF9 regulates IGSF9 expression but not vice versa.
A. qPCR analysis, bar plots. Both lnc-IGSF9 and IGSF9 are upregulated upon β-catenin-mediated pathway abrogation. *CTNNB1* levels are demonstrated to validate WNT abrogation**. B.** qPCR analysis. Expression levels of each gene are displayed as fold change over control. All samples are normalized over the sample LS174 shRNA β-catenin.

The RNA-seq dataset from the The RNA-seq dataset produced fro intestinal organoids after WNT activation is an exceptional model for studying the effects of the WNT pathway in intestinal physiology as i) it provides an 'in vitro' model of the intestinal epithelium and ii) both lnc-IGSF9 and IGSF9 are more robustly expressed in normal cells compared with cancer cells. Thus, we generated two custom molecular signatures (MSig) from the RNA-seq data Thus, we generated two custom molecular signatures (MSig) from the RNA-seq data derived upon deletion of *APC* gene, marked as either 'GENES UP-REGULATED UPON APC KO' OR 'GENES DOWN-REGULATED UPON APC KO'. Gene set enrichment analysis (GSEA) of the genes differentially expressed upon lnc-IGSF9 overexpression demonstrates that upregulated genes are enriched in the

molecular signature 'GENES DOWNREGULATED UPON APC KO' whereas genes downregulated upon lnc-IGSF9 overexpression are enriched in the Molecular Signature 'GENES UPREGULATED UPON APC KO' [Figure 3-7 GSEA analysis of differentially expressed genes in cells overexpressing lnc-IGSF9.] The above demonstrates that the overexpression of lnc-IGSF9 deregulates a relevant set of target genes that are divergently regulated by WNT signalling.

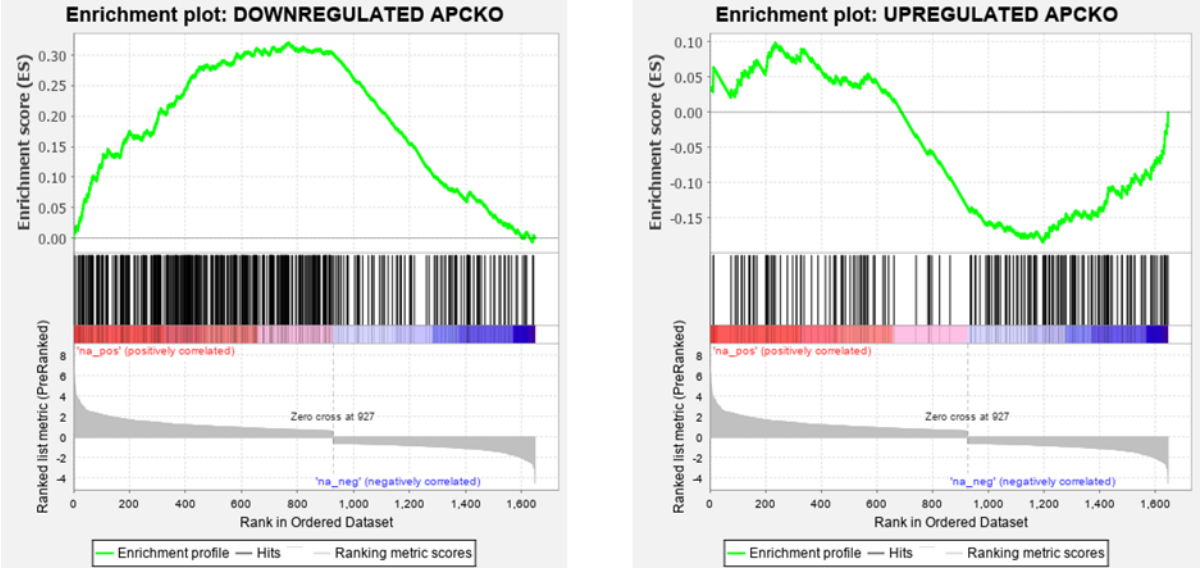
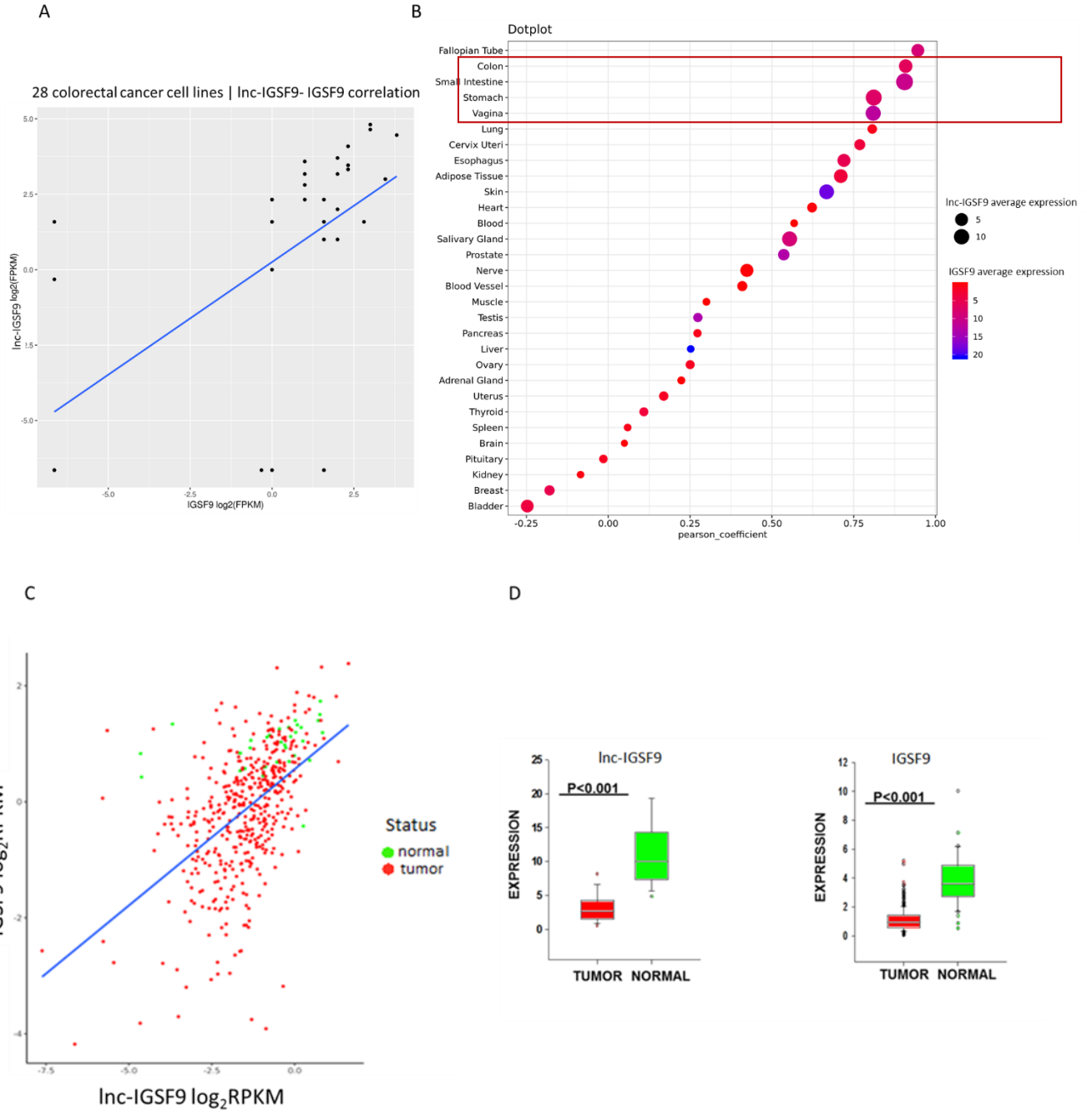


Figure 3-7 GSEA analysis of differentially expressed genes in cells overexpressing lnc-IGSF9.
A. Plot showing the enrichment of the upregulated genes upon lnc-IGSF9 overexpression in the Molecular Signature 'DOWNREGULATED UPON APC KO'. Each black line represents a gene from the dataset. Red shading corresponds to upregulated genes and blue to downregulated **B.** Plot depicting the enrichment of genes down-regulated upon lnc-IGSF9 overexpression in the 'UPREGULATED UPON APC KO' Molecular Signature.

Although, the IGSF9 protein has been mostly studied in the brain, analysis of Genotype-Tissue Expression (GTEX) data showed that it is expressed at higher levels in other tissues, such as in the colon and small intestine [Figure 3-8,B]. We wanted to further elucidate whether IGSF9 is a direct target of lnc-IGSF9, as well as its role in intestinal homeostasis. Therefore, we took advantage of publicly available RNA-seq data from i) different colorectal cancer cell lines ii) the GTEX project, iii) the The Cancer Genome Atlas Network (TCGAN), iv) and expression data generated from patient-derived intestinal organoids in order to examine lnc-IGSF9/IGSF9 expression patterns and study their correlation in terms of expression. Specifically, significant correlation of expression levels of the two genes was detected in 28 colorectal cancer cell lines ($R=0.74$, p-value< 0.01) [Figure 3-8,A]. Furthermore, in order to study the correlation between lnc-IGSF9 and IGSF9 in all normal tissues, we analysed GTEX RNA-seq data. The outcome of the analyses is summarized in the dot plot in [Figure 3-8,B]. It is apparent that the expression of lnc-IGSF9 and IGSF9 is closely linked in most tissues; however, in tissues where the expression of lnc-IGSF9 is low, IGSF9 expression is also limited and correlation is not strong. Importantly, significantly positive correlations in the expression of the two genes can be observed in the gastrointestinal tract (small intestine, colon and stomach) (See highlighted red box in [Figure 3-8,B]. In these tissues, we observed the highest average expression of lnc-IGSF9 and IGSF9, and the most significant correlations ($R^2 = 0.8$-0.9), pointing to a strong association between the two genes. A similar pattern of co-expression is evident from colorectal cancer samples and healthy tissues from the

TCGA Network data (R^2=0.51) [Figure 3-8, C]. It is noticeable that the correlation in expression of lnc-IGSF9-IGSF9 in healthy tissues is stronger than in colorectal cancer (see green dots Figure 3-8 , C). Moreover, the expression of lnc-IGSF9 and IGSF9 is reduced in colorectal cancer biopsies compared to the neighbouring healthy tissues [Figure 3-8,D]. Finally, we observed a strong correlation in expression between lnc-IGSF9 and IGSF9 in patient-derived intestinal organoids. The square correlation coefficient for the two genes was 0.927 and IGSF9. We also observed that REG4, a gene that it is negatively regulated by the WNT pathway, and CDX2, a known tumour suppressor (see below) strongly correlate with lnc-IGSF9 (Pearson Correlation=0.727 and 0.177, respectively) and IGSF9 (Pearson Correlation= 0.484, 0.404, respectively) whereas AXIN2, ASCL2 and CDK6, *which are known positively regulated WNT targets,* are found to be anticorrelated with both lnc-IGSF9 (Pearson Correlation=-0.405, -0.547 and -0.838, respectively) and IGSF9 (Pearson Correlation=-0.306, -0.441, -0.806, respectively)[Figure 3-8].



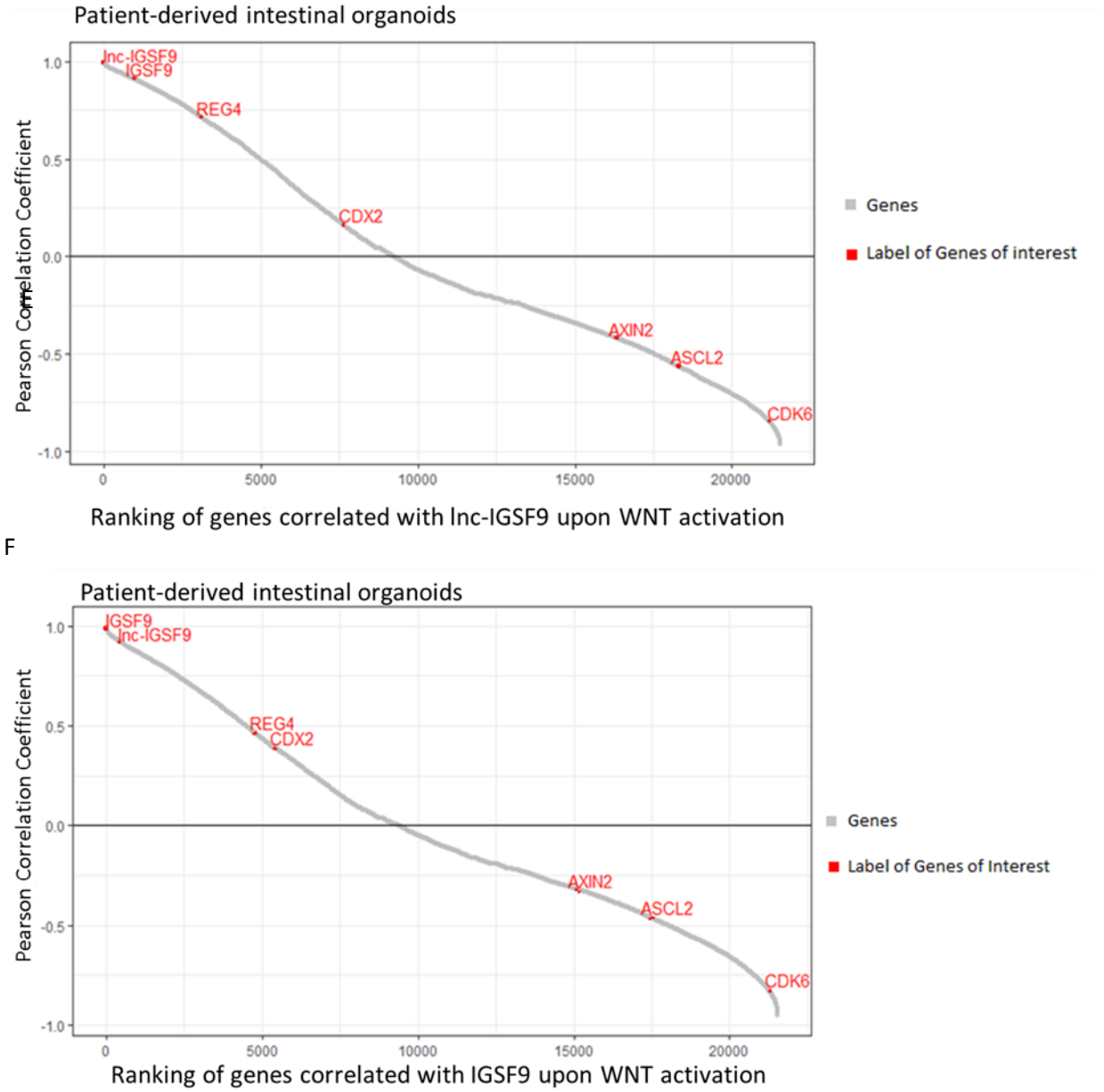


Figure 3-8 lnc-IGSF9 expression is positively correlated with IGSF9 expression

A. Scatter plot illustrating the association in terms of expression of lnc-IGSF9 and IGSF9 in 34 Colorectal Cancer Cell Lines **B.** Dot plot depicting in parallel the average expression of lnc-IGSF9 and IGSF9 in different tissues and their correlation coefficients. Data obtained from GTEX. **C.** Scatter plot that presents the correlation between lnc-IGSF9 and IGSF9 in 434 colorectal cancer biopsies, as well as in 40 biopsies from neighbouring healthy tissue. Data derived from TCGA. **D.** Box plot that gives information on the total expression of lnc-IGSF9 and IGSF9 in 434 colorectal cancer biopsies and in 40 healthy counterpants. Data obtained from TCGAN**. E.** Scatter plot, ranking all expressed genes with respect to the correlation with lnc- IGSF9 upon WNT activation and **F**. Same as E, for IGSF9.

Investigating the involvement of CDX2 in lnc-IGSF9/ IGSF9 regulatory axis

To better elucidate the mechanisms of lnc-IGSF9 function we investigated the proteins that interact with this transcript in cancer cells. For this purpose, we performed RNA pull-down experiments, using lnc-IGSF9 as a bait, followed by mass-spectrometry (mentioned in the introduction). Among the most prominent and consistent interactors was the transcription factor CDX2, which has been shown to function as a negative regulator of the β-catenin/TCF4 pathway, and is an intestine-specific regulator of cellular differentiation, limiting the proliferation of colon cancer cells and promoting cell differentiation[184].

In order to define the global binding pattern of CDX2 in intestinal cancer cells, we re-analyzed previously published data from a CDX2 ChIP-seq experiment in the LS174T cell line[185]. Interestingly, we identified CDX2 binding sites in the promoter as well as in a putative intronic enhancer of lnc-IGSF9, pointing to a putative role for the transcription factor in regulating lnc-IGSF9 expression [Figure 3-9,A]. In addition, another binding site, but with lower affinity, is locate in the promoter of the *IGSF9 gene*. We additionally identified a great number of prominent peaks close to genes of the *HOX* family, which are known CDX2 targets in intestinal epithelial cells[186], which validates the quality of the ChIP-seq dataset and its analysis.

Next, we tested whether lnc-IGSF9 expression is affected by CDX2 in LS174T cells. Firstly, we overexpressed *in-trans* the CDX2 protein in our cell system by transducing LS174T cells with lentiviruses constructed to overexpress CDX2. RT-qPCR analysis showed increased expression levels of both lnc-IGSF9 and IGSF9 genes, pointing to CDX2 being implicated in their regulation [Figure 3-9,B].

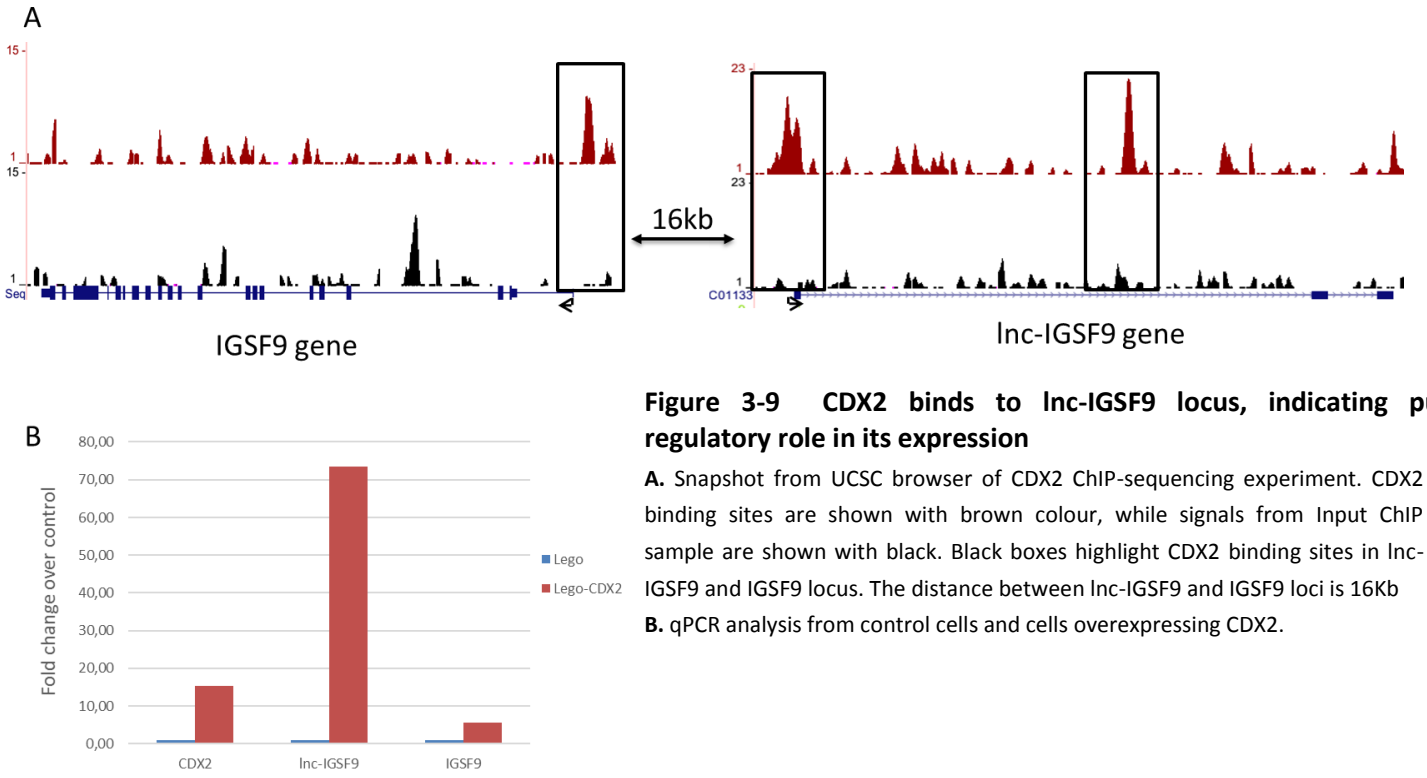


Figure 3-9 CDX2 binds to lnc-IGSF9 locus, indicating putative regulatory role in its expression

A. Snapshot from UCSC browser of CDX2 ChIP-sequencing experiment. CDX2 binding sites are shown with brown colour, while signals from Input ChIP sample are shown with black. Black boxes highlight CDX2 binding sites in lnc-IGSF9 and IGSF9 locus. The distance between lnc-IGSF9 and IGSF9 loci is 16Kb

B. qPCR analysis from control cells and cells overexpressing CDX2.

To examine the effects of the lnc-IGSF9-CDX2 interaction on IGSF9 expression, we overexpressed lnc-IGSF9 *in cis* by utilizing the CRISPRa system and synchronously overexpressed CDX2 *in trans* in LS174T cells. Surprisingly, we observed synergistic activation of IGSF9 by lnc-IGSF9 and CDX2: they increased IGSF9 levels far more than either factor alone [Figure 3-10]. We also observed that lnc-IGSF9 expression levels increased robustly after its *in cis* overexpression and simultaneous *in trans* overexpression of CDX2, indicating that its expression is highly dependent on CDX2. HOXC5 gene expression levels, a known target of CDX2, were quantified as a control.

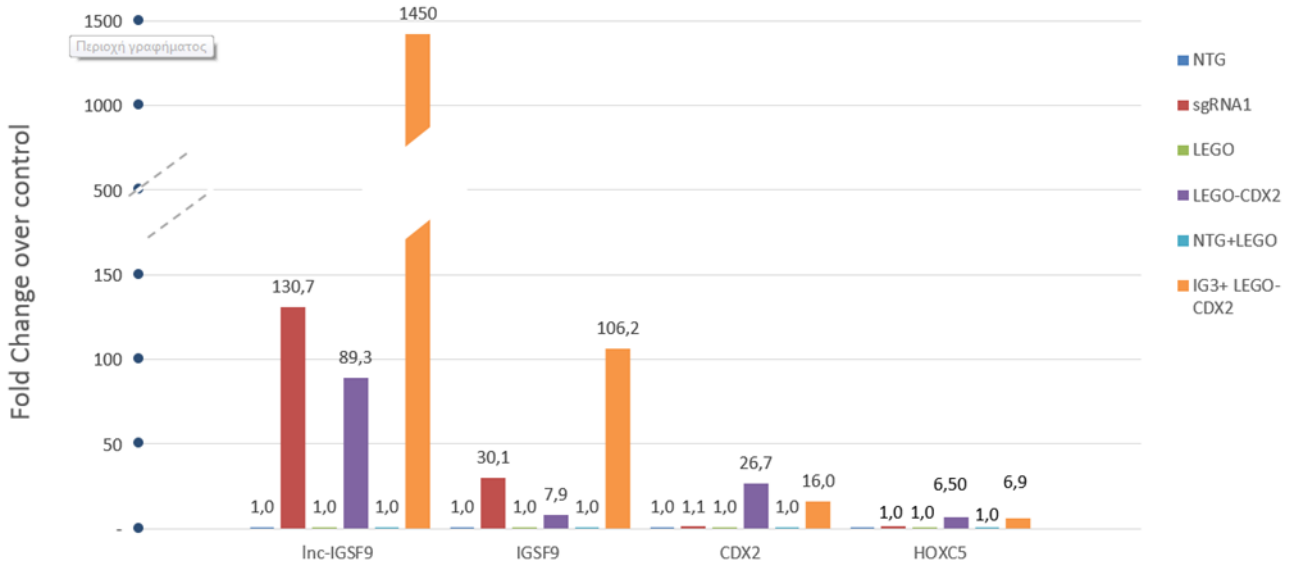


Figure 3-10 lnc-IGSF9 and CDX2 act synergistically to regulate IGSF9 expression
RT-qPCR analysis from LS174T cells overexpressing CRISPRa transduced with control (NTG) sgRNA or sgRNA1 targeting the promoter of lnc-GSF9.

The identification of the two CDX2 binding sites on lnc-IGSF9 gene locus prompted us to hypothesize that the CDX2 protein might participate in the lnc-IGSF9/IGSF9 regulatory axis by binding at these regions. To test this hypothesis, we firstly knocked-out the putative intronic enhancer of lnc-IGSF9, where CDX2 binds, using CRISPR/Cas9. Therefore, we transfected LS174T cells with the pX333 plasmid vector. The latter carries expresses Cas9 as well as two U6 cloning cassettes for the expression of two single guide RNAs. These sgRNAs are designed to target the region to be excised [Figure 3-11,A]. DNA extraction from the resulting cell clones, coupled with a conventional genotyping PCR (using a primer set flanking the region of interest), revealed 2 knock-out clones that lack the 'enhancer' region of *lnc-IGSF9* [Figure 3-11].

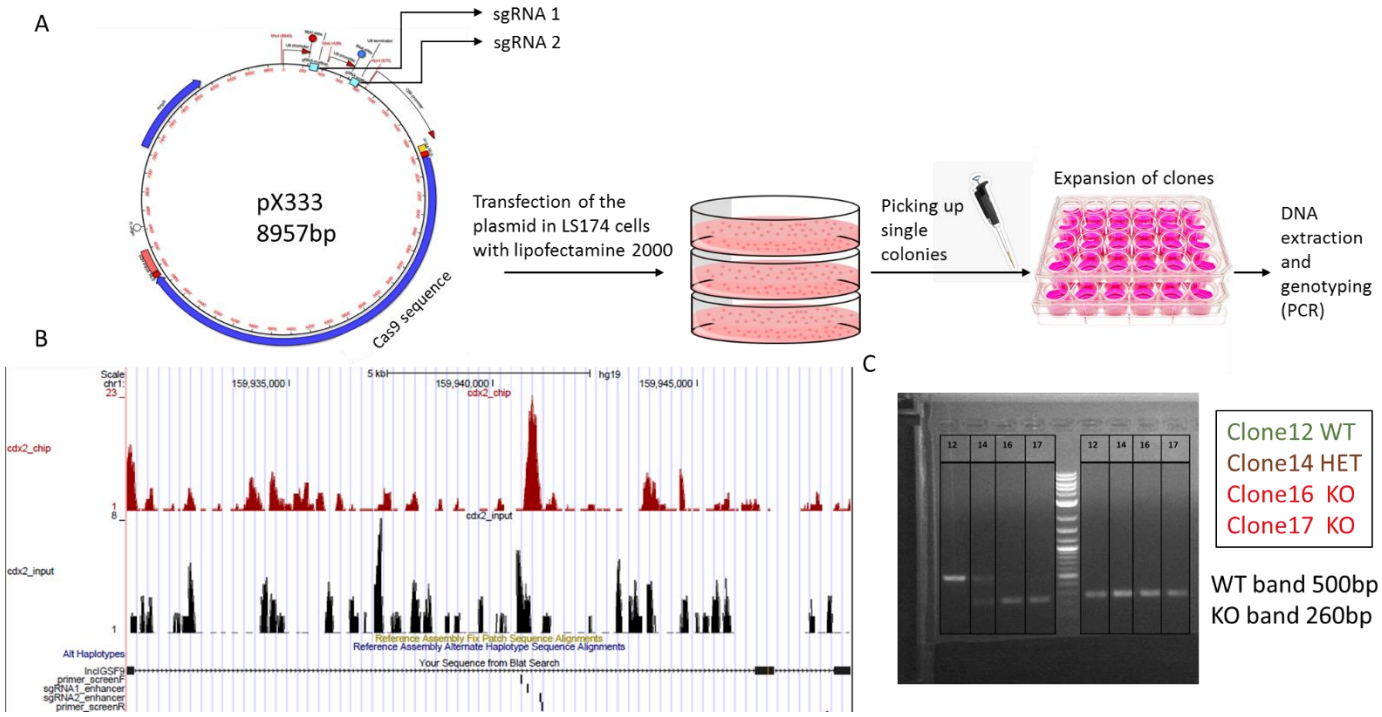


Figure 3-11 Generation of a stable cell line that lacks the internal enhancer region of lnc-IGSF9

A. Graphical representation of the experimental set-up used to obtain knock-out clones. Two different guide RNAs flanking the region of interest were cloned into the pX333 plasmid. Cells were plated and transfected synchronously with pX333 plasmid and a vector carrying a blasticidin resistance gene. 24 h after transfection the medium was refreshed with culture medium containing 10 ug/mL blasticidin. 72 h later blasticidin was removed. Clones were picked approximately 2 weeks later and expanded for 1-2 more weeks. Clones were screened by conventional genotyping PCR **B.** Snapshot from the UCSC Genome Browser depicting the binding site of CDX2 on the 'enhancer' region of lnc-IGSF9, the target sites of the designed sgRNAs and the screening primers that were used for the genotyping of KO, WT and HET clones **C.** Agarose gel showing the expected DNA bands. For the WT clones we expect a band at 500 bp, for HET clones two bands at 500 bp and 260 bp respectively, as the region of interest is deleted only in one allele, and for KO clones one band at 260 bp (left). The quality of the extracted DNA was also tested by using a primer set targeting a control genomic region that is not affected by the CRISPR/Cas9 manipulation (right).

To better characterize these clones, we firstly examined lnc-IGSF9 and IGSF9 expression profiles. RT-qPCR analysis in cDNAs derived from lnc-IGSF9 enhancer WT and KO clones revealed that there are no significant changes either in lnc-IGSF9 or in IGSF9 transcript levels in KO clones compared to the WT clones [Figure 3-12,A]. Subsequently, we investigated whether this region is implicated in CDX2-mediated regulation of the *lnc-IGSF9* gene. Therefore, we *trans* overexpressed CDX2 by using lentiviruses in lnc-IGSF9 enhancer KO and WT clones. The subsequent RT-qPCR analysis showed that CDX2 overexpression did not result to significant differences in *lnc-IGSF9* and *IGSF9* expression levels between WT and KO clones, suggesting that the enhancer is not necessary for the CDX2-mediated regulation of the *lnc-IGSF9 and IGSF9 genes* [Figure 3-12,B]

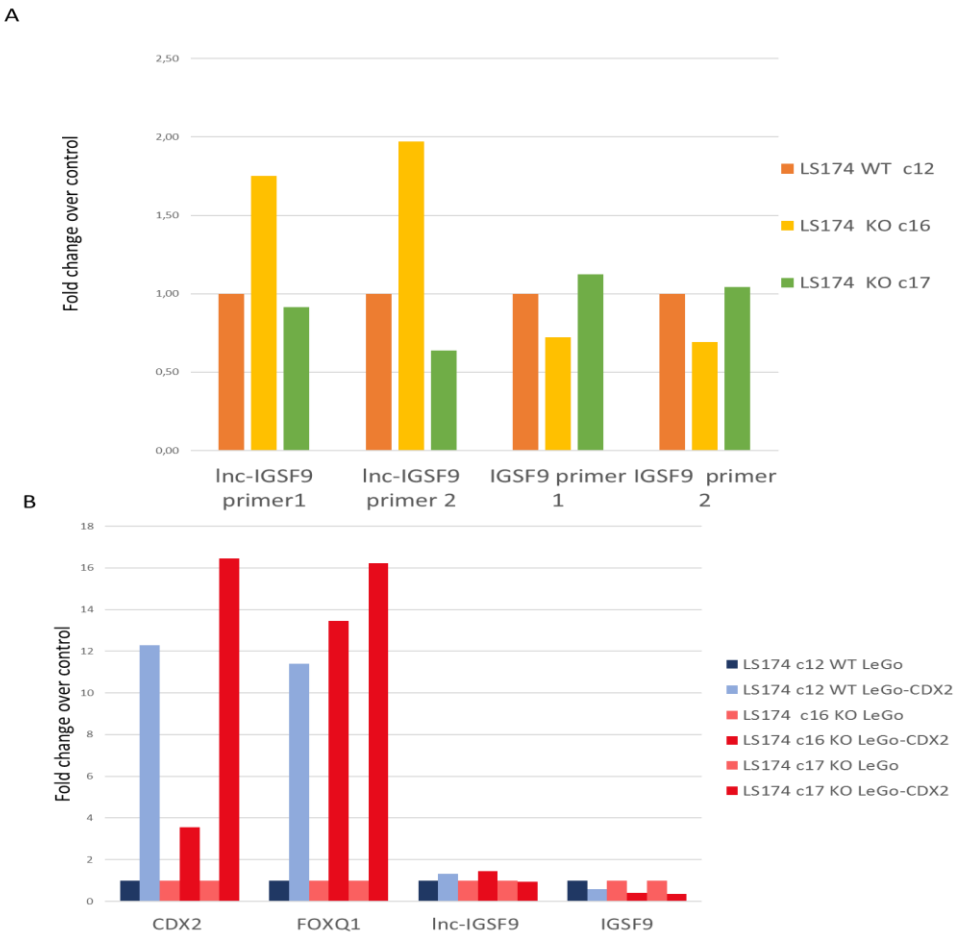


Figure 3-12 The lnc-IGSF9 enhancer is not necessary for lnc-IGSF9/IGSF9 regulation via CDX2

A. RT-qPCR analysis of that demonstrates the expression profile of KO clones. Expression was checked with two different primers for both lnc-IGSF9 and IGSF9. All samples have been normalized over the WT clone. **B.** RT-qPCR analysis in which LS174T WT and KO cells for the enhancer of lnc-IGSF9 were transduced with LeGo-IG2 and LeGo-IG2-CDX2 lentiviruses. CDX2 and FOXQ1 (a known CDX2 target gene) levels were examined to confirm CDX2 overexpression.

Moreover, we deleted a part of the promoter region of *lnc-IGSF9* to examine whether the CDX2-mediated effect on lnc-IGSF9 expression is achieved through CDX2 binding on its promoter region. For this purpose, we followed the same experimental approach as in the case of the 'enhancer' knock-out by utilizing different sgRNAs designed to specifically target and delete a part of the promoter region of *lnc-IGSF9*. DNA extraction from each clone and subsequent screening by PCR resulted in 3 KO clones that lack a part of the promoter sequence of lnc-IGSF9[Figure 3-13].Furthermore, we tested with RT-qPCR the expression levels of *lnc-IGSF9* in promoter WT and KO clones. We observed minimal to no expression of lnc-IGSF9 in all KO clones compared to the WT ones. IGSF9 transcript levels were not affected significantly, probably because *lnc-IGSF9* is already expressed at low levels in the LS174T cell line. Further experiments are designed to elucidate the importance of the promoter in CDX2-mediated regulation, as well as to delineate the mechanism by which the CDX2—lnc-IGSF9 complex impacts IGSF9 expression.

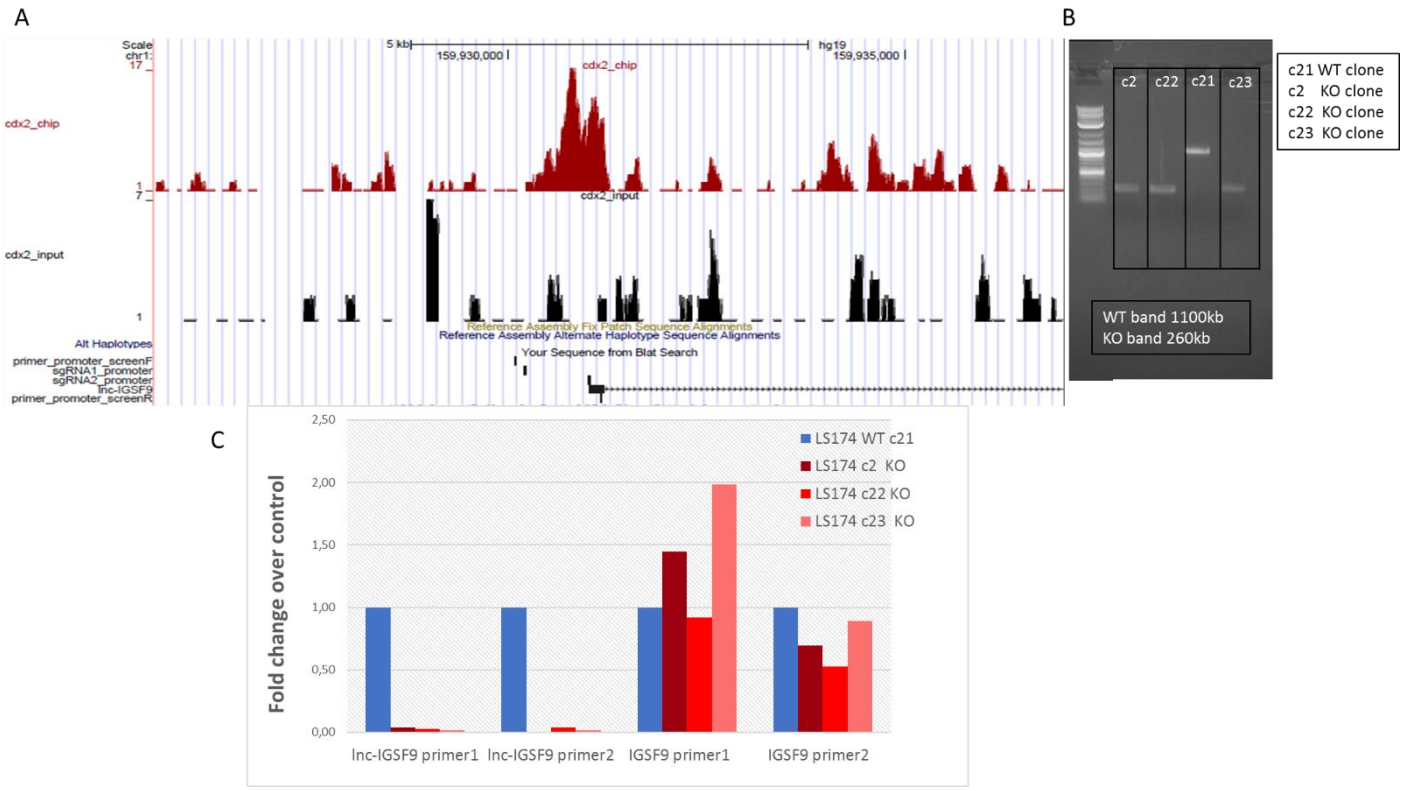


Figure 3-13 Excision of the promoter of lnc-IGSF9

A. Snapshot from the UCSC genome browser illustrating the position of sgRNAs that were used for targeted deletion of *lnc-IGSF9*. The primers used for the screening of promoter KO clones are also depicted. Tracks from a CDX2 ChIP-seq experiment illustrate the position where CDX2 binds to the promoter of lnc-IGSF9. **B.** Picture of an agarose gel showing the output of genotyping PCR in LS174T clones. 3 KO clones are depicted and 1 WT clone. **C.** qPCR analysis showing the expression of lnc-IGSF9 and IGSF9 in one WT (blue) and 3 KO clones (different shades of red). 2 different primers were used for both lnc-IGSF9 and IGSF9 for validation of the expression levels.

Discussion

The first cases of functional lncRNAs were described two decades ago, and since then a plethora of studies have identified lncRNAs that play a pivotal role in the regulation of cell proliferation, cell differentiation and apoptosis[187]. LncRNAs have been associated with disease, and most notably with cancer. Moreover, lncRNAs have been implicated in several signaling pathways, ultimately affecting their outputs.

The purpose of our study was to discover and characterize WNT-regulated lncRNAs with a functional role in intestinal homeostasis and carcinogenesis. Therefore, we utilized as a model cell lines in which we reduced the expression of β-catenin, the main transcriptional effector of the WNT cascade. ChIP-seq using antibodies against β-catenin and TCF4 and genome-wide transcriptional profiling upon WNT abrogation, revealed many unknown sites of WNT dependent-transcription and a series of lnc-RNAs that are deregulated upon β-catenin knock-down. Among them we identified lnc-IGSF9, a lncRNA that seems to be negatively regulated by the WNT pathway Interestingly, previous studies, which refer to lnc-IGSF9 by its alternative annotation LINC01133, have reported its implication in distinct cancer types.

The first reference of lnc-IGSF9 in the literature originates from a study which highlights its role as an oncogene in lung squamous cell cancer (LSCC). Since then, lnc-IGSF9 has been characterized in a spectrum of cancers, e.g., cervical squamous cell carcinoma, osteosarcoma, pancreatic ductal adenocarcinoma, and hepatocellular carcinoma, where it acts as an oncogene[176]. On the other hand, it has been characterized as a tumor suppressor in cancers such as oral squamous cell carcinoma, gastric cancer, breast cancer and colorectal cancer. Specifically, in a recent study, Zhang et al. demonstrated that lnc-IGSF9 functions as a tumour suppressor, titrating the SFSF6 protein away from its RNA targets by directly binding to its critical domain, ultimately blocking the induction of EMT[188]. This particular study demonstrates that lnc-IGSF9 acts as a molecular decoy in the cytoplasm to suppress colorectal carcinogenesis. However, work in our laboratory has shown that lnc-IGSF9 is almost exclusively expressed in the nucleus in our CRC cell line system, LS174T cells, as well as in other colorectal cell lines.

Lnc-IGSF9 drew our attention as an example of a lncRNA that is negatively regulated by the WNT pathway. As is generally accepted in the field, the β-catenin/TCF4 complex mainly activates gene expression. Therefore, lnc-IGSF9 is upregulated upon WNT abrogation and also features β-catenin and TCF4 binding sites in its locus, implying a novel putative repressive role for the β-catenin/TCF4 complex.

WNT pathway abrogation, after inducible or lentiviral-mediated knock-down of β-catenin or overexpression of dominant-negative TCF4, resulted in up-regulation of lnc-IGSF9. Then, we examined whether the activation of the WNT pathway could conversely repress the expression of lnc-IGSF9. To study this, we exploited publicly available data from patient-derived intestinal organoids, after extrinsic (activation after genetic ablating of the APC gene) or intrinsic(activation through WNT3a/R-spondin treatment) WNT activation[189]. Interestingly, the expression of lnc-IGSF9 was significantly down-regulated by both modes of WNT activation, corroborating the idea that the WNT pathway negatively regulates lnc-IGSF9. Data derived from organoids provides valuable information because they mimic in more deatail the in vivo environment compared to two-dimensional cellular models. The in vitro recapitulation of some aspects of native physiology of the adult intestinal epithelium and stem cell function makes the intestinal organoid technology an excellent tool that can recapitulate disease pathways ore accurately. In our case, it was also important that lnc-IGSF9 is robustly expressed in organoids compared to our colorectal cancer cell lines, where the low

expression of lnc-IGSF9 is a limiting factor in its investigation. Taking all these into consideration, we propose that lnc-IGSF9 expression is regulated by the WNT pathway in the intestine.

Due to the fact that lnc-IGSF9 is downregulated in colorectal cancer, the majority of CRC cell lines express low levels of lnc-IGSF9. The low basal expression of lnc-IGSF9 in our CRC cancer cell lines taken together with previous in trans overexpression experiments performed in our laboratory that demonstrated little effect of lnc-IGSF9 in the phenotype of CRC cells, may suggest that this lncRNA acts in cis.

To overcome the limitations of low expression of linc-IGSF9 in our cellular systems, we overexpressed it in cis with the use of the CRISPRa system to identify putative lnc-IGSF9 target genes. The CRISPRa system is a powerful technique that enables overexpression of endogenous genes by targeting an activating Cas9 complex close to their transcription sites; it acts with high efficiency and considerably lower off-target effects and is widely used to study the *in cis* effects of lncRNAs[190]. In cis overexpression of lnc-IGSF9, followed by 3' mRNA-sequencing revealed its transcriptional effects on target genes. We discovered that lnc-IGSF9 orchestrates a gene expression program that induces cell adhesion and differentiation, while it restricts cell proliferation, highlighting its prognostic value in colorectal cancer.

Supporting further this theory, we discovered an underlying regulatory axis between lnc-IGSF9 and IGSF9, a gene that is located in the vicinity of the lnc-IGSF9 gene locus. Specifically, IGSF9 was the only gene that was found to be affected within a 1MB distance from the *lnc-IGSF9* locus upon *lnc-IGSF9* overexpression Moreover, we showed that the lnc-IGSF9 transcript is essential for the expression of IGSF9, with the latter following the same pattern of expression as the lncRNA, after WNT abrogation/activation and in colorectal carcinogenesis. Therefore, we propose that lnc-IGSF9 potentially regulates IGSF9 expression through a cis-acting mechanism.

The IGSF9 protein is a cell adhesion molecule that has been thoroughly studied in synapse development; however, there is no clear evidence for its role in other tissues, where it is more abundantly expressed[191]. Indeed, profiling using GTEX data clearly shows minimal expression of IGSF9 in the human brain, while its expression in the intestine and colon is much higher. Because of the low expression levels of this protein in intestinal cancers, to further investigate its role in colorectal carcinogenesis, we are planning to perform gain of function studies for IGSF9 in our CRC cell lines by using an IGSF9 ORF expressing plasmid. By performing immunofluorescence experiments with an antibody against IGSF9 and cell scratch/wound healing assays we will decipher the contribution of the above protein in cell migration and/or cell-cell communication as an adhesion molecule.

Although we unveiled a regulatory axis between lnc-IGSF9 and IGSF9, the exact underlying mechanism of regulation is incompletely understood. RNA pull-down experiments coupled with mass-spectrometry using lnc-IGSF9 as a bait revealed lnc-IGSF9 interacting proteins. CDX2 is a protein whose interaction with lnc-IGSF9 was highly reproducible and, in addition, is an intestine-specific transcription factor, responsible for cell differentiation[182,184]. In concordance with the RNA pull-down experiments, meta-analysis of ChIP-seq experiments using an antibody against CDX2 in LS174T cells[192], revealed binding sites in the promoter and the 'enhancer' region of lnc-IGSF9. Taking all these into consideration we propose that CDX2 as a potential mediator of lnc-IGSF9 regulation. Moreover, *in trans* overexpression of CDX2, showed that it regulates the expression of both lnc-IGSF9 and IGSF9. Further experiments will be designed to overexpress CDX2 and synchronously knockdown lnc-IGSF9 to investigate whether the CDX2-mediated regulation of IGSF9 depends on lnc-IGSF9. Such a result will demonstrate the direct dependence of IGSF9 on the lnc-IGSF9 transcript.

We also found that, based on double transduction experiments (CDX2 and lnc-IGSF9 simultaneous overexpression), the lnc-IGSF9 – CDX2 interaction can affect IGSF9 expression in a synergistic manner. Finally, we report that the successful deletion of the lnc-IGSF9 promoter region with CRISPR/Cas9 technolog, showed loss of lnc-IGSF9 expression in the resulting knock-out clones. On the other hand, this did not occur after the excision of the putative enhancer, implying that this region might have less importance for the lnc-IGSF9/IGSF9 regulatory axis. Finally, lnc-IGSF9 promoter KO clones could be further utilized as a tool to elucidate the impact of CDX2 on IGSF9 in the absence of lnc-IGSF9 expression.

From a broad perspective lncRNAs is a novel class of transcripts, whose expression is more tissue-specific than that of mRNAs[193]. Only a small number of lncRNAs is expressed in a particular cell type, indicating that transcripts themselves might regulate tissue and cell-type specific transcriptional programs driving for example differentiation or cell proliferation[194]. It is obvious that whole transcriptome analysis of single cells, in combination with deep RNA-seq of tissues, will assist to identify abundant cell type-specific lncRNAs. Due to their specificity lncRNAs can be potent targets for lncRNA-centered therapies. Indeed, lnc-IGSF9 is a lncRNA, whose expression is barely detected in many tissues but robustly expressed in others such as the small intestine and the colon. Moreover, it is known that a particular lncRNA can act through a different mechanism at different tissues and/or different developmental stages[195]. For example, Malat1 a well-known lncRNA abundant in most tissues has been found to act differently according to the tissue where it is investigated[196,197]. In our case, lnc-IGSF9 has been characterized by different groups to act either as an oncogene or as tumor suppressor in different cancers[198], indicating that its molecular function is highly dependent on the tissue and the context where it is expressed. Here we report an in cis mechanism for lnc-IGSF9 that appears to impact the expression of a gene in its close neighborhood. Such in cis mechanisms have been previously reported for a plethora of nuclear lncRNAs, yet it is still difficult to distinguish how frequently this regulation might result from the lncRNA transcript or its regulatory regions or even the process of transcription itself. Numerous examples in the literature have demonstrated that a transcriptional effect previously attributed to a lncRNA transcript was in fact mediated through the promoter of the lncRNA[199]. It goes without saying that new methods and technologies such as CRISPR-KRAB, CRISPR-SAM, Road-block dCas9, CRISPR/Cas9 knock-out, shRNA technology[200,201] and others should be applied to distinguish the real cause of changes in the transcriptome. As far as, lnc-IGSF9 is concerned there is strong evidence arising from our experiments that it is the transcript that regulates the expression of IGSF9, but further experiments should be performed to reinforce our working hypothesis. In line with this, the fact that we have discovered that CDX2 — an intestinal-specific factor which also binds to lnc-IGSF9 locus— interacts with lnc-IGSF9 suggests that lnc-IGSF9 might be an example of lncRNA that regulates its neighbouring gene through direct interaction with CDX2. Yet this direct interaction remains to be proved as well as the exact mechanism through which those two molecules regulate the expression of IGSF9. Finally, from a systems biology perspective, the lncRNAs generally perform their biological functions together with other molecules, rather than individually. Many examples of lncRNAs have been discovered that interact with general transcription factors to regulated gene expression. For instance, noncoding RNAs UI and 7SK directly interact with general transcription factor TFIID or transcription elongation factor PTEF-b, respectively, to modulate transcription initiation or pause release[202]. However, in our study we have identified a tissue-specific transcription factor that can regulate expression by interacting with the lnc-IGSF9 transcript. Taken into consideration the tissue specificity of lncRNAs and the fact that they can interact with transcription factors we can suggest that the interaction between CDX2 and lnc-IGSF9 is a potent mechanism that mediates IGSF9 expression and highlights the importance of lncRNAs in tissue-specific and disease-specific molecular mechanisms.

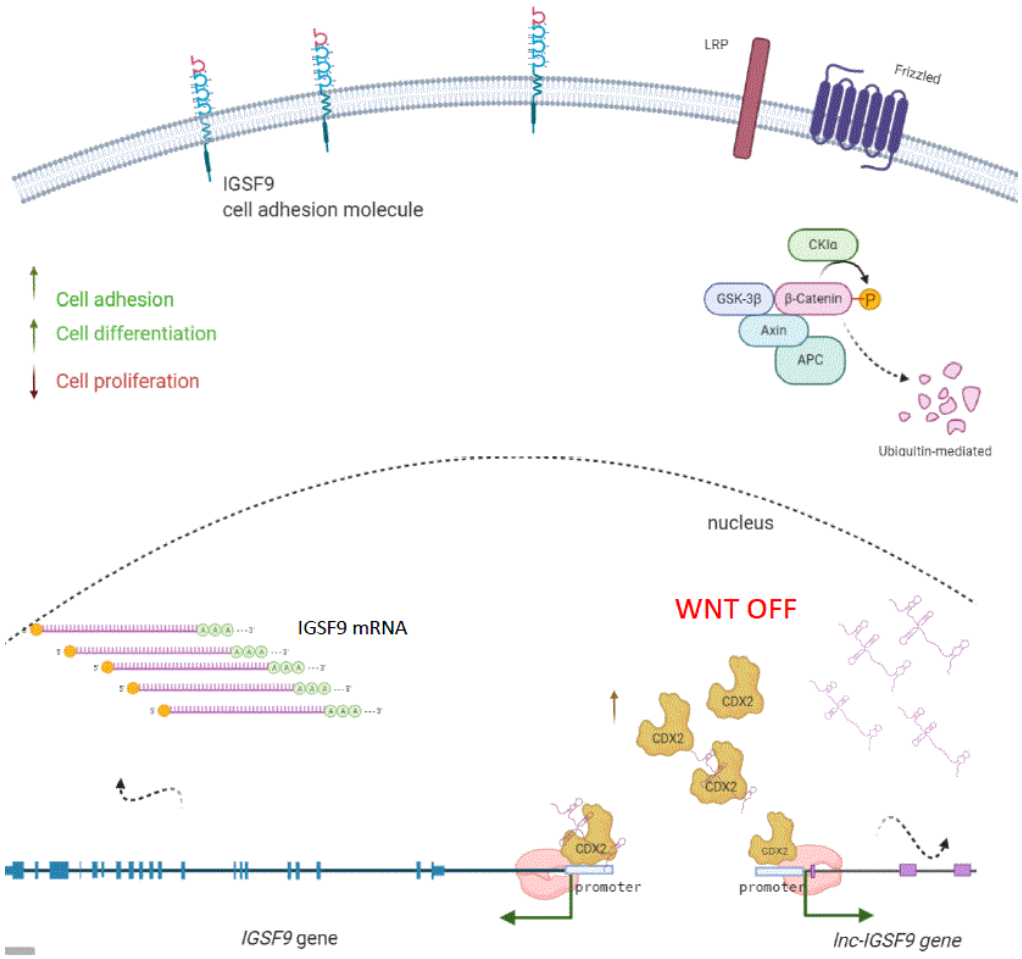


Figure 4-1 Schematic representation of the proposed molecular mechanism

When the WNT pathway is in active, lnc-IGSF9 is robustly expressed, interacts with CDX2 transcription factor to activate the expression of IGSF9. IGSF9 protein is expressed and translocated to the cell membrane, promoting cell adhesion, cell-to-cell communication and differentiation.

Bibliography

1. van der Heijden, M. & Vermeulen, L. Stem cells in homeostasis and cancer of the gut. *Molecular Cancer* vol. 18 (2019).
2. Sancho, E., Batlle, E. & Clevers, H. Live and let die in the intestinal epithelium. *Current Opinion in Cell Biology* vol. 15 763–770 (2003).
3. van der Flier, L. G. & Clevers, H. Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annual Review of Physiology* vol. 71 241–260 (2009).
4. Bjerknes, M. & Cheng, H. Clonal analysis of mouse intestinal epithelial progenitors. *Gastroenterology* **116**, 7–14 (1999).
5. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003–1007 (2007).
6. Cheng, H. & Leblond, C. P. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine V. Unitarian theory of the origin of the four epithelial cell types. *American Journal of Anatomy* **141**, 537–561 (1974).
7. Metcalfe, C., Kljavin, N. M., Ybarra, R. & de Sauvage, F. J. Lgr5+ stem cells are indispensable for radiation-induced intestinal regeneration. *Cell Stem Cell* **14**, 149–159 (2014).
8. Tetteh, P. W. *et al.* Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters. *Cell Stem Cell* **18**, 203–213 (2016).
9. Tomic, G. *et al.* Phospho-regulation of ATOH1 Is Required for Plasticity of Secretory Progenitors and Tissue Regeneration. *Cell Stem Cell* **23**, 436-443.e7 (2018).
10. Jadhav, U. *et al.* Dynamic Reorganization of Chromatin Accessibility Signatures during Dedifferentiation of Secretory Precursors into Lgr5+ Intestinal Stem Cells. *Cell Stem Cell* **21**, 65-77.e5 (2017).
11. Sato, T. *et al.* Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469**, 415–418 (2011).
12. Snippert, H. J. *et al.* Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**, 134–144 (2010).
13. Lopez-Garcia, C., Klein, A. M., Simons, B. D. & Winton, D. J. Intestinal stem cell replacement follows a pattern of neutral drift. *Science* **330**, 822–825 (2010).
14. Vermeulen, S. J. *et al.* Addressing uncertainty in adaptation planning for agriculture. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 8357–8362 (2013).
15. Jemal, A. *et al.* Cancer Statistics, 2006. *CA: A Cancer Journal for Clinicians* **56**, 106–130 (2006).
16. Rustgi, A. K. The genetics of hereditary colon cancer. *Genes and Development* vol. 21 2525–2538 (2007).
17. Brannon, A. R. *et al.* Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic

colorectal cancer lesions. *Genome Biology* **15**, 454 (2014).

18. Wei, E. K. *et al.* Comparison of risk factors for colon and rectal cancer. *International Journal of Cancer* **108**, 433–442 (2004).

19. Fodde, R. & Brabletz, T. Wnt/β-catenin signaling in cancer stemness and malignant behavior. *Current Opinion in Cell Biology* vol. 19 150–158 (2007).

20. Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459**, 262–265 (2009).

21. Jenny, F. H. & Basler, K. Powerful Drosophila screens that paved the wingless pathway. *Fly* vol. 8 218–225 (2014).

22. Rijsewijk, F. *et al.* The Drosophila homology of the mouse mammary oncogene int-1 is identical to the segment polarity gene wingless. *Cell* **50**, 649–657 (1987).

23. Kikuchi, A., Yamamoto, H., Sato, A. & Matsumoto, S. New Insights into the Mechanism of Wnt Signaling Pathway Activation. in *International Review of Cell and Molecular Biology* vol. 291 21–71 (Elsevier Inc., 2011).

24. MacDonald, B. T., Tamai, K. & He, X. Wnt/β-Catenin Signaling: Components, Mechanisms, and Diseases. *Developmental Cell* **17**, 9–26 (2009).

25. MacDonald, B. T., Tamai, K. & He, X. Wnt/β-Catenin Signaling: Components, Mechanisms, and Diseases. *Developmental Cell* vol. 17 9–26 (2009).

26. Clevers, H. Wnt/β-Catenin Signaling in Development and Disease. *Cell* vol. 127 469–480 (2006).

27. Niehrs, C. The complex world of WNT receptor signalling. *Nature Reviews Molecular Cell Biology* vol. 13 767–779 (2012).

28. Katoh, M. Canonical and non-canonical WNT signaling in cancer stem cells and their niches: Cellular heterogeneity, omics reprogramming, targeted therapy and tumor plasticity (Review). *International Journal of Oncology* vol. 51 1357–1369 (2017).

29. Katoh, M. WNT/PCP signaling pathway and human cancer (Review). *Oncology Reports* vol. 14 1583–1588 (2005).

30. Willert, K. *et al.* Wnt proteins are lipid-modified and can act as stem cell growth factors. *Nature* **423**, 448–452 (2003).

31. Rios-Esteves, J. & Resh, M. D. Stearoyl CoA Desaturase Is Required to Produce Active, Lipid-Modified Wnt Proteins. *Cell Reports* **4**, 1072–1081 (2013).

32. Takada, R. *et al.* Monounsaturated Fatty Acid Modification of Wnt Protein: Its Role in Wnt Secretion. *Developmental Cell* **11**, 791–801 (2006).

33. Rios-Esteves, J. & Resh, M. D. Stearoyl CoA Desaturase Is Required to Produce Active, Lipid-Modified Wnt Proteins. *Cell Reports* **4**, 1072–1081 (2013).

34. Janda, C. Y., Waghray, D., Levin, A. M., Thomas, C. & Garcia, K. C. Structural basis of Wnt recognition by frizzled. *Science* **336**, 59–64 (2012).

35. Bartscherer, K., Pelte, N., Ingelfinger, D. & Boutros, M. Secretion of Wnt Ligands Requires

Evi, a Conserved Transmembrane Protein. *Cell* **125**, 523–533 (2006).

36. Bänziger, C. *et al.* Wntless, a Conserved Membrane Protein Dedicated to the Secretion of Wnt Proteins from Signaling Cells. *Cell* **125**, 509–522 (2006).

37. Najdi, R. *et al.* A uniform human Wnt expression library reveals a shared secretory pathway and unique signaling activities. *Differentiation* **84**, 203–213 (2012).

38. Herr, P. & Basler, K. Porcupine-mediated lipidation is required for Wnt recognition by Wls. *Developmental Biology* **361**, 392–402 (2012).

39. Saha, S. *et al.* Macrophage-derived extracellular vesicle-packaged WNTs rescue intestinal stem cells and enhance survival after radiation injury. *Nature Communications* **7**, 1–16 (2016).

40. Korkut, C. *et al.* Trans-Synaptic Transmission of Vesicular Wnt Signals through Evi/Wntless. *Cell* **139**, 393–404 (2009).

41. Farin, H. F. *et al.* Visualization of a short-range Wnt gradient in the intestinal stem-cell niche. *Nature* **530**, 340–343 (2016).

42. He, X., Semenov, M., Tamai, K. & Zeng, X. LDL receptor-related proteins 5 and 6 in Wnt/β-catenin signaling: Arrows point the way. *Development* vol. 131 1663–1677 (2004).

43. Cavallo, R. A. *et al.* Drosophila Tcf and Groucho interact to repress wingless signalling activity. *Nature* **395**, 604–608 (1998).

44. van de Wetering, M. *et al.* Armadillo coactivates transcription driven by the product of the Drosophila segment polarity gene dTCF. *Cell* **88**, 789–799 (1997).

45. Logan, C. Y. & Nusse, R. The Wnt signaling pathway in development and disease. *Annual Review of Cell and Developmental Biology* **20**, 781–810 (2004).

46. Binnerts, M. E. *et al.* R-Spondin1 regulates Wnt signaling by inhibiting internalization of LRP6. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 14700–14705 (2007).

47. Cong, F., Schweizer, L. & Varmus, H. Wnt signals across the plasma membrane to activate the β-catenin pathway by forming oligomers containing its receptors, Frizzled and LRP. *Development* **131**, 5103–5115 (2004).

48. Tolwinski, N. S. *et al.* Wg/Wnt signal can be transmitted through arrow/LRP5,6 and axin independently of Zw3/Gsk3β activity. *Developmental Cell* **4**, 407–418 (2003).

49. Bryja V et al. (2009), The extracellular domain of Lrp5/6 inhibits non... - Paper. http://www.xenbase.org/literature/article.do?method=display&articleId=38620.

50. van Amerongen, R., Mikels, A. & Nusse, R. Alternative Wnt signaling is initiated by distinct receptors. *Science Signaling* vol. 1 (2008).

51. Bovolenta, P., Esteve, P., Ruiz, J. M., Cisneros, E. & Lopez-Rios, J. Beyond Wnt inhibition: New functions of secreted Frizzled-related proteins in development and disease. *Journal of Cell Science* vol. 121 737–746 (2008).

52. Satoh, D. *et al.* Spatial control of branching within dendritic arbors by dynein-dependent transport of Rab5-endosomes. *Nature Cell Biology* **10**, 1164–1171 (2008).

53. Mao, B. *et al.* Kremen proteins are Dickkopf receptors that regulate Wnt/β-catenin signalling. *Nature* **417**, 664–667 (2002).

54. Semënov, M. v. *et al.* Head inducer dickkopf-1 is a ligand for Wnt coreceptor LRP6. *Current Biology* **11**, 951–961 (2001).

55. Ellwanger, K. *et al.* Targeted Disruption of the Wnt Regulator Kremen Induces Limb Defects and High Bone Density. *Molecular and Cellular Biology* **28**, 4875–4882 (2008).

56. Wang, K. *et al.* Characterization of the Kremen-binding site on Dkk1 and elucidation of the role of Kremen in Dkk-mediated Wnt antagonism. *Journal of Biological Chemistry* **283**, 23371–23375 (2008).

57. Semënov, M. v. *et al.* Head inducer dickkopf-1 is a ligand for Wnt coreceptor LRP6. *Current Biology* **11**, 951–961 (2001).

58. Yamamoto, H. *et al.* Basolateral secretion of Wnt5a in polarized epithelial cells is required for apical lumen formation. *Journal of Cell Science* doi:10.1242/jcs.163683.

59. Kazanskaya O et al. (2004), R-Spondin2 is a secreted activator of Wnt/beta-... - Paper. http://www.xenbase.org/literature/article.do?method=display&articleId=2907.

60. Daniels, D. L. & Weis, W. I. β-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nature Structural and Molecular Biology* **12**, 364–371 (2005).

61. Arce, L., Yokoyama, N. N. & Waterman, M. L. Diversity of LEF/TCF action in development and disease. *Oncogene* vol. 25 7492–7504 (2006).

62. Hatzis, P. *et al.* Genome-Wide Pattern of TCF7L2/TCF4 Chromatin Occupancy in Colorectal Cancer Cells. *Molecular and Cellular Biology* **28**, 2732–2744 (2008).

63. Atcha, F. A. *et al.* A Unique DNA Binding Domain Converts T-Cell Factors into Strong Wnt Effectors. *Molecular and Cellular Biology* **27**, 8352–8363 (2007).

64. Molenaar, M. *et al.* XTcf-3 transcription factor mediates β-catenin-induced axis formation in xenopus embryos. *Cell* **86**, 391–399 (1996).

65. Huber, O. *et al.* Nuclear localization of β-catenin by interaction with transcription factor LEF-1. *Mechanisms of Development* **59**, 3–10 (1996).

66. Mosimann, C., Hausmann, G. & Basler, K. β-Catenin hits chromatin: Regulation of Wnt target gene activation. *Nature Reviews Molecular Cell Biology* vol. 10 276–286 (2009).

67. Kinzler, K. W. *et al.* Identification of FAP locus genes from chromosome 5q21. *Science* **253**, 661–665 (1991).

68. Nishisho, I. *et al.* Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. *Science* **253**, 665–669 (1991).

69. Ilyas, M., Tomlinson, I. P. M., Rowan, A., Pignatelli, M. & Bodmer, W. F. β-Catenin mutations in cell lines established from human colorectal cancers. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 10330–10334 (1997).

70. Morin, P. J. *et al.* Activation of β-catenin-Tcf signaling in colon cancer by mutations in β-catenin or APC. *Science* **275**, 1787–1790 (1997).

71. Webster, K. E., Ferree, P. M., Holmes, R. P. & Cramer, S. D. Identification of missense, nonsense, and deletion mutations in the GRHPR gene in patients with primary hyperoxaluria type II (PH2). *Human Genetics* **107**, 176–185 (2000).

72. Shimizu, Y. *et al.* Frequent alterations in the Wnt signaling pathway in colorectal cancer with microsatellite instability. *Genes, Chromosomes and Cancer* **33**, 73–81 (2002).

73. Liu, W. *et al.* Axin is a scaffold protein in TGF-β signaling that promotes degradation of Smad7 by Arkadia. *EMBO Journal* **25**, 1646–1658 (2006).

74. Lammi, L. *et al.* Mutations in AXIN2 Cause Familial Tooth Agenesis and Predispose to Colorectal Cancer. *American Journal of Human Genetics* **74**, 1043–1050 (2004).

75. Duval, A. *et al.* The human T-cell transcription factor-4 gene: Structure, extensive characterization of alternative splicings, and mutational analysis in colorectal cancer cell lines. *Cancer Research* **60**, 3872–3879 (2000).

76. Aguilera, O. *et al.* Epigenetic inactivation of the Wnt antagonist DICKKOPF-1 (DKK-1) gene in human colorectal cancer. *Oncogene* **25**, 4116–4121 (2006).

77. Li, X. & Jope, R. S. Is glycogen synthase kinase-3 a central modulator in mood regulation. *Neuropsychopharmacology* vol. 35 2143–2154 (2010).

78. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).

79. Matsumoto, A., Clohessy, J. G. & Pandolfi, P. P. SPAR, a lncRNA encoded mTORC1 inhibitor. *Cell Cycle* vol. 16 815–816 (2017).

80. Anderson, D. M. *et al.* A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **160**, 595–606 (2015).

81. Wang, K. C. & Chang, H. Y. Molecular Mechanisms of Long Noncoding RNAs. *Molecular Cell* vol. 43 904–914 (2011).

82. Ma, L., Bajic, V. B. & Zhang, Z. On the classification of long non-coding RNAs. *RNA Biology* vol. 10 924–933 (2013).

83. Rutenberg-Schoenberg, M., Sexton, A. N. & Simon, M. D. The Properties of Long Noncoding RNAs That Regulate Chromatin. *Annual Review of Genomics and Human Genetics* **17**, 69–94 (2016).

84. Zhao, R. H. *et al.* BC200 lncRNA a potential predictive marker of poor prognosis in esophageal squamous cell carcinoma patients. *OncoTargets and Therapy* **9**, 2221–2226 (2016).

85. Harrow, J. *et al.* GENCODE: The reference human genome annotation for the ENCODE project. *Genome Research* **22**, 1760–1774 (2012).

86. Brockdorff, N. *et al.* The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71**, 515–526 (1992).

87. Brown, C. J. *et al.* The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**, 527–542 (1992).

88. Kanduri, C. Long noncoding RNAs: Lessons from genomic imprinting. *Biochimica et*

Biophysica Acta - Gene Regulatory Mechanisms vol. 1859 102–111 (2016).

89. Fatica, A. & Bozzoni, I. Long non-coding RNAs: New players in cell differentiation and development. *Nature Reviews Genetics* vol. 15 7–21 (2014).

90. Fortes, P. & Morris, K. v. Long noncoding RNAs in viral infections. *Virus Research* **212**, 1–11 (2016).

91. Heo, J. B. & Sung, S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* **331**, 76–79 (2011).

92. Okazaki, Y. *et al.* Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**, 563–573 (2002).

93. Carninci, P. *et al.* Molecular biology: The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).

94. Ravasi, T. *et al.* Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Research* **16**, 11–19 (2006).

95. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research* **22**, 1775–1789 (2012).

96. Pruitt, K. D., Tatusova, T., Brown, G. R. & Maglott, D. R. NCBI Reference Sequences (RefSeq): Current status, new features and genome annotation policy. *Nucleic Acids Research* **40**, (2012).

97. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).

98. Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 11667–11672 (2009).

99. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* **28**, 511–515 (2010).

100. Cabili, M. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes and Development* **25**, 1915–1927 (2011).

101. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644–652 (2011).

102. Pauli, W. M., O'Reilly, R. C., Yarkoni, T. & Wager, T. D. Regional specialization within the human striatum for diverse psychological functions. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 1907–1912 (2016).

103. Hutchinson, S. A., Cheesman, S. E., Hale, L. A., Boone, J. Q. & Eisen, J. S. Nkx6 proteins specify one zebrafish primary motoneuron subtype by regulating late islet1 expression. *Development* **134**, 1671–1677 (2007).

104. Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H. & Bartel, D. P. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**, 1537–1550 (2011).

105. Church, D. M. *et al.* Lineage-Specific Biology Revealed by a Finished Genome Assembly of the Mouse. *PLoS Biology* **7**, e1000112 (2009).

106. Johnsson, P. *et al.* A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. *Nature Structural and Molecular Biology* **20**, 440–446 (2013).

107. Quinn, J. J. & Chang, H. Y. Unique features of long non-coding RNA biogenesis and function. *Nature Reviews Genetics* vol. 17 47–62 (2016).

108. Ulitsky, I. & Bartel, D. P. XLincRNAs: Genomics, evolution, and mechanisms. *Cell* **154**, 26 (2013).

109. Guttman, M. & Rinn, J. L. Modular regulatory principles of large non-coding RNAs. *Nature* vol. 482 339–346 (2012).

110. Sigova, A. A. *et al.* Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2876–2881 (2013).

111. Clark, B. S. *et al.* Loss of Llgl1 in retinal neuroepithelia reveals links between apical domain size, Notch activity and neurogenesis. *Development* **139**, 1599–1610 (2012).

112. Tani, H. *et al.* Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Research* **22**, 947–956 (2012).

113. Tichon, A. *et al.* A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nature Communications* **7**, (2016).

114. Lee, S. *et al.* Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell* **164**, 69–80 (2016).

115. Musgrove, C., Jansson, L. I. & Stone, M. D. New perspectives on telomerase RNA structure and function. *Wiley Interdisciplinary Reviews: RNA* vol. 9 (2018).

116. Vendramin, R. *et al.* SAMMSON fosters cancer cell fitness by concertedly enhancing mitochondrial and cytosolic translation. *Nature Structural and Molecular Biology* **25**, 1035–1046 (2018).

117. Grote, P. *et al.* The Tissue-Specific lncRNA Fendrr Is an Essential Regulator of Heart and Body Wall Development in the Mouse. *Developmental Cell* **24**, 206–214 (2013).

118. Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409–419 (2010).

119. Andersen, R. E. *et al.* The Long Noncoding RNA Pnky Is a Trans-acting Regulator of Cortical Development In Vivo. *Developmental Cell* **49**, 632-642.e7 (2019).

120. Werner, M. S. & Ruthenburg, A. J. Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes. *Cell Reports* **12**, 1089–1098 (2015).

121. Balas, M. M. & Johnson, A. M. Exploring the mechanisms behind long noncoding RNAs and cancer. *Non-coding RNA Research* vol. 3 108–117 (2018).

122. Wang, K. C. & Chang, H. Y. Molecular Mechanisms of Long Noncoding RNAs. *Molecular Cell* vol. 43 904–914 (2011).

123. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annual Review of Biochemistry* **81**, 145–166 (2012).

124. Kung, J. T. Y., Colognori, D. & Lee, J. T. Long noncoding RNAs: Past, present, and future. *Genetics* vol. 193 651–669 (2013).

125. Chen, L. L. Linking Long Noncoding RNA Localization and Function. *Trends in Biochemical Sciences* vol. 41 761–772 (2016).

126. Hung, T. *et al.* Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. in *Nature Genetics* vol. 43 621–629 (2011).

127. Baldassarre, A. & Masotti, A. Long non-coding RNAs and p53 regulation. *International Journal of Molecular Sciences* vol. 13 16708–16717 (2012).

128. Schmitt, A. M. & Chang, H. Y. Long Noncoding RNAs in Cancer Pathways. *Cancer Cell* vol. 29 452–463 (2016).

129. Bhan, A., Soleimani, M. & Mandal, S. S. Long noncoding RNA and cancer: A new paradigm. *Cancer Research* vol. 77 3965–3981 (2017).

130. Thomson, D. W. & Dinger, M. E. Endogenous microRNA sponges: Evidence and controversy. *Nature Reviews Genetics* **17**, 272–283 (2016).

131. Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and Functions of Long Noncoding RNAs. *Cell* vol. 136 629–641 (2009).

132. Ebert, M. S. & Sharp, P. A. Emerging roles for natural microRNA sponges. *Current Biology* vol. 20 (2010).

133. Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P. P. A ceRNA hypothesis: The rosetta stone of a hidden RNA language? *Cell* vol. 146 353–358 (2011).

134. Huarte, M. The emerging role of lncRNAs in cancer. *Nature Medicine* vol. 21 1253–1261 (2015).

135. Rasool, M. *et al.* Non-coding RNAs in cancer diagnosis and therapy. *Non-coding RNA Research* vol. 1 69–76 (2016).

136. Bolha, L., Ravnik-Glavač, M. & Glavač, D. Long Noncoding RNAs as Biomarkers in Cancer. *Disease Markers* vol. 2017 (2017).

137. Rinn, J. L. *et al.* Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell* **129**, 1311–1323 (2007).

138. Gupta, R. A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).

139. Tsai, M. C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).

140. Balas, M. M. & Johnson, A. M. Exploring the mechanisms behind long noncoding RNAs and cancer. *Non-coding RNA Research* vol. 3 108–117 (2018).

141. Mondal, T. *et al.* MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA-DNA triplex structures. *Nature Communications* **6**, (2015).

142. Zhang, X. *et al.* Maternally expressed gene 3 (MEG3) noncoding ribonucleic acid: Isoform structure, expression, and functions. *Endocrinology* **151**, 939–947 (2010).

143. Zhou, Y., Zhang, X. & Klibanski, A. MEG3 noncoding RNA: A tumor suppressor. *Journal of Molecular Endocrinology* vol. 48 (2012).

144. Liu, B. *et al.* A Cytoplasmic NF-κB Interacting Long Noncoding RNA Blocks IκB Phosphorylation and Suppresses Breast Cancer Metastasis. *Cancer Cell* **27**, 370–381 (2015).

145. Huang, L. *et al.* Systematic identification of long non-coding RNAs during pollen development and fertilization in *Brassica rapa*. *The Plant Journal* **96**, 203–222 (2018).

146. Maeder, M. L. & Gersbach, C. A. Genome-editing technologies for gene and cell therapy. *Molecular Therapy* vol. 24 430–446 (2016).

147. Durruthy-Durruthy, J. *et al.* The primate-specific noncoding RNA HPAT5 regulates pluripotency during human preimplantation development and nuclear reprogramming. *Nature Genetics* **48**, 44–52 (2015).

148. Zhu, S. *et al.* Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. *Nature Biotechnology* **34**, 1279–1286 (2016).

149. Bassett, A. R. *et al.* Considerations when investigating lncRNA function in vivo. *eLife* **3**, 1–14 (2014).

150. Paralkar, V. R. *et al.* Unlinking an lncRNA from Its Associated cis Element. *Molecular Cell* **62**, 104–110 (2016).

151. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature Biotechnology* **32**, 670–676 (2014).

152. Gilbert, L. A. *et al.* XCRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442 (2013).

153. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).

154. Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647–661 (2014).

155. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).

156. Fellmann, C., Gowen, B. G., Lin, P. C., Doudna, J. A. & Corn, J. E. Cornerstones of CRISPR-Cas in drug discovery and therapy. *Nature Reviews Drug Discovery* vol. 16 89–100 (2017).

157. Veres, A. *et al.* Low incidence of Off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell* **15**, 27–30 (2014).

158. Hu, W. *et al.* RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 11461–11466 (2014).

159. Schaefer, K. A. *et al.* Unexpected mutations after CRISPR-Cas9 editing in vivo. *Nature Methods* vol. 14 547–548 (2017).

160. Yang, J. *et al.* CRISPR/Cas9-mediated noncoding RNA editing in human cancers. *RNA Biology* vol. 15 35–43 (2018).

161. Nusse, R. & Clevers, H. Wnt/β-Catenin Signaling, Disease, and Emerging Therapeutic Modalities. *Cell* **169**, 985–999 (2017).

162. van de Wetering, M. *et al.* The β-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* **111**, 241–250 (2002).

163. van de Wetering, M. *et al.* Specific inhibition of gene expression using a stably integrated, inducible small-interfering-RNA vector. *EMBO Reports* **4**, 609–615 (2003).

164. Shen, P., Pichler, M., Chen, M., Calin, G. A. & Ling, H. To Wnt or Lose: The Missing Non-Coding Linc in Colorectal Cancer. *International journal of molecular sciences* vol. 18 (2017).

165. Ong, M. S. *et al.* 'Lnc'-ing Wnt in female reproductive cancers: therapeutic potential of long non-coding RNAs in Wnt signalling. *British Journal of Pharmacology* vol. 174 4684–4700 (2017).

166. Trimarchi, T. *et al.* Genome-wide mapping and characterization of notch-regulated long noncoding RNAs in acute leukemia. *Cell* **158**, 593–606 (2014).

167. Yuan, J. hang *et al.* A Long Noncoding RNA Activated by TGF-β promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer Cell* **25**, 666–681 (2014).

168. Li, W. & Kang, Y. A New Lnc in metastasis: Long noncoding RNA mediates the prometastatic functions of TGF-β. *Cancer Cell* vol. 25 557–559 (2014).

169. Peng, W. X., Koirala, P. & Mo, Y. Y. LncRNA-mediated regulation of cell signaling in cancer. *Oncogene* **36**, 5661–5667 (2017).

170. Pomerantz, M. M. *et al.* The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nature Genetics* **41**, 882–884 (2009).

171. Ling, H. *et al.* CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. *Genome Research* **23**, 1446–1461 (2013).

172. Mokry, M. *et al.* Integrated genome-wide analysis of transcription factor occupancy, RNA polymerase II binding and steady-state RNA levels identify differentially regulated functional gene classes. *Nucleic Acids Research* **40**, 148–158 (2012).

173. Cheng, L., Ming, H., Zhu, M. & Wen, B. Long noncoding RNAs as Organizers of Nuclear Architecture. *Science China Life Sciences* **59**, 236–244 (2016).

174. Giakountis, A. *et al.* A Positive Regulatory Loop between a Wnt-Regulated Non-coding RNA and ASCL2 Controls Intestinal Stem Cell Fate. *Cell Reports* **15**, 2588–2596 (2016).

175. Huang, C. S. *et al.* The C/EBPβ-LINC01133 axis promotes cell proliferation in pancreatic ductal adenocarcinoma through upregulation of CCNG1. *Cancer Letters* **421**, 63–72 (2018).

176. Zhang, J., Zhu, N. & Chen, X. A novel long noncoding RNA LINC01133 is upregulated in lung squamous cell cancer and predicts survival. *Tumor Biology* **36**, 7465–7471 (2015).

177. Li, Z. *et al.* LINC01133: an emerging tumor-associated long non-coding RNA in tumor and osteosarcoma. *Environmental Science and Pollution Research* vol. 27 32467–32473 (2020).

178. Song, Z. *et al.* LINC01133 inhibits breast cancer invasion and metastasis by negatively regulating SOX4 expression through EZH2. *Journal of Cellular and Molecular Medicine* **23**, 7554–7565 (2019).

179. Zhang, J. H., Li, A. Y. & Wei, N. Downregulation of long non-coding RNA LINC01133 is predictive of poor prognosis in colorectal cancer patients. *European review for medical and pharmacological sciences* **21**, 2103–2107 (2017).

180. Lin, M. F., Jungreis, I. & Kellis, M. PhyloCSF: A comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27**, (2011).

181. Verzi, M. P. *et al.* TCF4 and CDX2, major transcription factors for intestinal function, converge on the same cis-regulatory regions. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 15157–15162 (2010).

182. Saad, R. S. CDX2 as a marker for intestinal differentiation: Its utility and limitations. *World Journal of Gastrointestinal Surgery* **3**, 159 (2011).

183. Moll, P., Ante, M., Seitz, A. & Reda, T. QuantSeq 3′ mRNA sequencing for RNA quantification. *Nature Methods* **11**, i–iii (2014).

184. Guo, R. J., Funakoshi, S., Lee, H. H., Kong, J. & Lynch, J. P. The intestine-specific transcription factor Cdx2 inhibits β-catenin/TCF transcriptional activity by disrupting the β-catenin-TCF protein complex. *Carcinogenesis* **31**, 159–166 (2010).

185. Pinto, R. *et al.* Precise integration of inducible transcriptional elements (PrIITE) enables absolute control of gene expression. *Nucleic Acids Research* **45**, (2017).

186. Verzi, M. P. *et al.* TCF4 and CDX2, major transcription factors for intestinal function, converge on the same cis-regulatory regions. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 15157–15162 (2010).

187. Kung, J. T. Y., Colognori, D. & Lee, J. T. Long noncoding RNAs: Past, present, and future. *Genetics* **193**, 651–669 (2013).

188. Kong, J. *et al.* Long non-coding RNA LINC01133 inhibits epithelial–mesenchymal transition and metastasis in colorectal cancer by interacting with SRSF6. *Cancer Letters* **380**, 476–484 (2016).

189. Michels, B. E. *et al.* Human colon organoids reveal distinct physiologic and oncogenic Wnt responses. *Journal of Experimental Medicine* **216**, 704–720 (2019).

190. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature Methods* **10**, 973–976 (2013).

191. Mishra, A. *et al.* Genetic evidence for the adhesion protein IgSF9/Dasm1 to regulate inhibitory synapse development independent of its intracellular domain. *Journal of Neuroscience* **34**, 4187–4199 (2014).

192. Pinto, R. *et al.* Precise integration of inducible transcriptional elements (PrIITE) enables absolute control of gene expression. *Nucleic Acids Research* **45**, (2017).

193. Kern, C. *et al.* Genome-wide identification of tissue-specific long non-coding RNA in three farm animal species 06 Biological Sciences 0604 Genetics. *BMC Genomics* **19**, 684 (2018).

194. Flynn, R. A. & Chang, H. Y. Long noncoding RNAs in cell-fate programming and reprogramming. *Cell Stem Cell* vol. 14 752–761 (2014).

195. Sarropoulos, I., Marin, R., Cardoso-Moreira, M. & Kaessmann, H. Developmental dynamics of lncRNAs across mammalian organs and species. *Nature* **571**, 510–514 (2019).

196. Sun, Y. & Ma, L. New insights into long non-coding rna malat1 in cancer and metastasis. *Cancers* vol. 11 (2019).

197. Arun, G., Aggarwal, D. & Spector, D. L. MALAT1 long non-coding RNA: Functional implications. *Non-coding RNA* vol. 6 (2020).

198. Li, Z. *et al.* LINC01133: an emerging tumor-associated long non-coding RNA in tumor and osteosarcoma. *Environmental Science and Pollution Research* vol. 27 32467–32473 (2020).

199. Cho, S. W. *et al.* Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell* **173**, 1398-1412.e22 (2018).

200. Goyal, A. *et al.* Challenges of CRISPR/Cas9 applications for long non-coding RNA genes. *Nucleic acids research* **45**, e12 (2017).

201. Pickar-Oliver, A. & Gersbach, C. A. The next generation of CRISPR–Cas technologies and applications. *Nature Reviews Molecular Cell Biology* vol. 20 490–507 (2019).

202. O'Gorman, W., Thomas, B., Kwek, K. Y., Furger, A. & Akoulitchev, A. Analysis of U1 small nuclear RNA interaction with cyclin H. *Journal of Biological Chemistry* **280**, 36920–36925 (2005).