Investigating the role of WiNTRLINC3 in intestinal homeostasis and carcinogenesis

Alexandros Galaras

Master of Molecular Biomedicine

Institute of Molecular Biology and Genetics, BSRC “Alexander Fleming”, Pantelis Hatzis Group

Abstract

The canonical Wnt pathway is an essential regulator of intestinal self-renewal, affecting the transcriptional profile of a plethora of genes involved in development, proliferation and stem cell maintenance. Functional perturbations of the Wnt pathway that result in aberrant activity of β-catenin, the main modulator of this signaling cascade, are among the leading causes of Colorectal Cancer (CRC). Interestingly, there is strong evidence that multiple targets of the effector β-catenin/TCF4 transcriptional complex reside in non-coding regions with several long non-coding RNAs (lncRNAs) operating either as regulators or as targets of the Wnt pathway, ultimately affecting intestinal physiology.

In our study we characterize WiNTRLINC3, a long intergenic non-coding RNA (lincRNA) which is negatively regulated by the β-catenin/TCF4 pathway and the low expression of which is associated with Colorectal Cancer. Cis-overexpression of WiNTRLINC3 with the use of CRISPRa system in the intestinal cell line LS174 showed that WiNTRLINC3 mainly increases the expression of genes that control cell adhesion facilitating cell differentiation, while it down-regulates genes that promote cell proliferation. Among the up-regulated genes, we distinguished IGSF9 as a putative mediator of the function of WiNTRLINC3. IGSF9 is a cell adhesion protein that is encoded by a gene located in the vicinity of WiNTRLINC3 and reduced expression of which also correlates with Colorectal Cancer. We report that WiNTRLINC3 and IGSF9 display a similar regulatory pattern, with WiNTRLINC3 controlling the expression of IGSF9, but not vice versa.

Moreover, we investigate the role of CDX2 transcription factor, a protein that was identified as interactor of WiNTRLINC3 in RNA pull-down experiments. in regulating WiNTRLINC3. Re-analysis of publicly available data in LS174 cells helped us determine the global binding pattern of CDX2. Among others, CDX2 strongly binds to both the promoter and an intronic enhancer of WiNTRLINC3, pointing to direct regulation of WiNTRLINC3 expression by CDX2. We are currently evaluating if the activity of those regulatory regions is elevated upon overexpression of CDX2. We speculate that CDX2 activates WiNTLINC3 by interacting with it, ultimately affecting the expression of IGSF9.

We propose that WiNTRLINC3 may be a biomarker and therapeutic target in Colorectal Cancer.

Acknowledgements

Since I enrolled to the Master program of Molecular Biomedicine, I had the chance to be surrounded by inspiring people that helped me to develop my personal and scientific skills.

First of all, I would like to thank my supervisor, Dr Pantelis Hatzis for always motivating me to combine the expertise of Molecular Biology with Bioinformatics, enriching my classic perspective of Genomics with modern trends. From our first encounter at his office, he was willing to discuss with me, guiding me through the whole scientific procedure. He supported my goals and insisted on teaching me to be open minded in my way of thinking.

I would also like to thank Dr. George Kollias and Dr. Antigone Dimas for their supervision and the creative discussions that we had during the last two years.

With profound gratitude, I would like to thank my two lab partners, Vasiliki Zarkou for the instructive guidance and the constructive collaboration and Antonis Giakountis for his excellent mentorship. I would also like to acknowledge the help of Panagiotis Moulos, who always found time to answer even my simplest questions in his field of expertise. Moreover, I would like to thank all the members from the Fousteri and Verykokakis laboratories, for the daily interaction and their positive feedback.

Finally, I am grateful to my parents who were always by my side, supporting and believing every single step. The continuous encouragement of all those people led to the accomplishment of this thesis. Thank you all.

Alexandros Galaras

Supervisor: Pantelis Hatzis

Research committee:

Pantelis Hatzis, Researcher B', Division of Molecular Biology and Genetics, BSRC "Alexander Fleming"

George Kollias, Professor of Experimental Physiology, Medical School of Athens & Researcher A', Division of Immunology, BSRC "Alexander Fleming"

Antigone Dimas, Researcher C', Division of Molecular Biology and Genetics, BSRC "Alexander Fleming"

Table of Contents

1 Introduction

1.1 The Wnt/β-catenin signaling pathway

The Wnt/β-catenin signaling cascade is an essential regulator of major biological processes that affects organismal development and stem cell maintenance[1]. Binding of Wnt ligands to their membrane receptors, activates a series of events that finally orchestrate gene regulation, cell proliferation and other cellular activities[2].

The Wnt pathway is highly conserved in metazoans, playing a crucial role in organizing among others, the developmental axis of patterned organisms, but is absent in unicellular organisms[3]. Wnt signaling comprises three distinct pathways, the canonical or Wnt/β-catenin dependent pathway (which will be mainly described in the next sections) and the non-canonical or β-catenin-independent pathway, which is further sub-divided into two branches, the Planar Cell Polarity (PCP) and the Wnt/Ca^{2+} pathways. While both the canonical and the non-canonical pathways share common effector molecules (such as Wnt, Fz, etc.), their biological outcomes are different. The PCP transduction cascade regulates the actin cytoskeleton dictating the polarization and the migration of the cells during gastrulation. The Wnt/Ca^{2+} pathway induces Ca^{2+} release from the endoplasmic reticulum, affecting dorsal axis formation, promoting ventral cell fate, gastrulation, and organ formation[4].

1.1.1 Wnt proteins: Production, Modifications and Secretion

Wnt proteins are signaling molecules that influence a multitude of developmental pathways through regulating cell-to-cell communication, cell proliferation and stem cell fates[5]. The Wnt family consists of 19 genes, in human and mice, which encode secreted, lipid-modified, hydrophobic proteins[5,6]. Genetic deletions of different Wnt molecules manifest distinct phenotypes, mainly involving developmental defects of the organ/tissue where each individual Wnt molecule is expressed. Their high level of conservation between different species, indicates their fundamental role in the developmental patterning of multi-cellular organisms[2,3].

As the structure of most Wnt ligands have not been described yet, their biochemical properties have been defined based on their primary amino acid sequence. Wnt proteins are 40 kDa in size and are characterized by 22 cysteine residues which form intra-molecular disulfide bridges that are necessary for proper protein folding. They also consist of a signal sequence of 20 hydrophobic amino acids, which is responsible for their secretion[6].

After translation, Wnt proteins interact with multiple enzymes and incorporate into vesicles to achieve their movement to the extracellular space. Along this process, Wnt proteins might be further modified, with the two major modifications being glycosylation and acylation[6]. Glycosylation occurs in the endoplasmatic reticulum, where the oligosaccharyl transferase complex (OST) rapidly and efficiently attaches N-linked oligosaccharide chains to the appropriate residues on the peptide backbone[7]. Glycosylation of Wnts is necessary for their folding and secretion in the extracellular matrix, but is not necessary for their function[6].

On the other hand, lipid modification (acylation) has been shown to be critical for the proper function of Wnt molecules[8]. Active Wnts contain a palmitic acid attached to a conserved cysteine[8] or serine[9] residue and mutagenesis of those sites leads to the dysfunction and poor secretion of Wnt molecules. Porcupine (Porcn), a protein located in the endoplasmatic reticulum of Wnt producing cells, catalyzes the transfer of the lipid group to the Wnt molecules. Palmitoylation of Wnts, targets them to membrane regions that are composed of sphingolipids and cholesterol, the latter acting as scaffolds during signal transduction[5]. Moreover, acylation is closely related with glycosylation of Wnts, as the lipid-modified Wnts can attach more efficiently to the endoplasmatic reticulum[6]. Other putative roles of lipid modification could be the strengthening of Wnt-Frizzled interaction or anchoring the Wnts on the membrane, resulting in increased signaling[5].

Lipid modified Wnts can be further transferred and secreted through the aid of a transmembrane protein, Wntless/Evi (WIs). Even though it is incompletely understood how Wnts are transferred to the cell membrane and then secreted to the extracellular environment, studies suggest that vesicles or exosomes are formed to achieve Wnt secretion. These vesicles contain WIs and mature Wnt proteins which protrude towards the cell exterior to bind to their corresponding receptors[10]. According to an alternative model, Wnt3 transfer is mediated through the direct contact between cells and is regulated through the expression of Frizzled receptors and the transmembrane E3 ligases Rnf43/Znrf3[11]. Because Wnt proteins play a pivotal role in developmental axis specification, they are commonly considered as

long distance morphogens. Controversially, recent discoveries highlight that Wnt molecules exert their signaling activity between neighboring contacting cells, while long distance activity occurs either due to the capacity of vesicles containing WIs to move over longer distances[12] or because of sequential Wnt signaling between neighboring cells[13]. Upon their secretion, Wnt proteins navigate the extracellular space by interacting with other molecules[6]. For instance, the interactions with glycosaminoglycan-modified proteins assist Wnts-receptor binding and modulate the strength of the signaling[14].

1.1.2 Wnt receptors, agonists and antagonists

Wnt proteins bind on an heterodimeric receptor complex, consisting of Frizzled (Fz) and low density lipoprotein receptor-related protein 5/6 (LRP5/6), located on the surface of the target cells[1,2]. Mammalian Fz proteins are 7-transmembrane receptors with a N-terminal cysteine-rich domain (CRD)[15] where the Wnt molecules can bind to multiple locations[16]. Structure analysis of the Wnt-CRD complexes indicates that each Wnt molecule can bind to numerous Fz receptors and vice versa[16].

Binding of a Wnt protein on a Fz receptor triggers its conformational change resulting in the formation of a dimer between Fz and LRP5/6[17]. This structural change makes the cytoplasmic tail of LRP6 accessible for phosphorylation by the kinases Glycogen synthase kinase 3 (GSK3) and Casein kinase 1 (CK1γ)[18]. Consequently, the Axin protein binds on the phosphorylated tail of LRP6 initiating the main series of events of the Wnt signaling cascade[19]. This interaction is further supported by the fact that the cytoplasmic part of Fz interacts with Dishevelled (Dsh)[20] through a common DIX domain that both proteins contain[21,22].

Different alternative ligands can also bind on the Fz receptors, competing with Wnt molecules. One of those agonists is Norrin, a protein encoded by the *NDP* gene, which binds with high affinity to FZD-4[23]. In vitro co-expression of Norrin, FZD-4 and LRP5 showed increased activity of the Wnt/β-catenin signaling pathway[24]. Another group of molecules that enhance the Wnt pathway are the R-spondin proteins (Rspo) which bind the Lgr4, Lgr5 and Lgr6 receptors with high affinity[25–27] and ultimately prevent the membrane clearance of Wnt receptors by the negative regulators Rnf43/Znrf3[28]. On the other hand, several secreted proteins, such as the secreted Frizzled-related proteins (sFRPs) and Wnt inhibitory protein (WIF), function as inhibitors that bind Wnts, preventing them from interacting with Fz receptors[29]. An alternative mechanism of Wnt inhibition is achieved by proteins of the Dickkopf

(DKK) and the WISE/SOST families which bind to the LRP5/6 receptor, disrupting the formation of the Fz-LRP5/6 complex which is necessary for Wnt signaling[30].

1.1.3 The Wnt/β-catenin signaling pathway

The Wnt/β-catenin pathway is characterized by an active and inactive state. The regulated activity of a destruction complex (DC) and accumulation of β-catenin in the nucleus are the two main events that determine the two states. The destruction complex consists of Axin2 which functions as a scaffold where Casein Kinase1 (CK1) and Glycogen Synthase Kinase 3 (GSK3) bind to, along with two tumor suppressor proteins, Adenomatous Polyposis Coli (APC) and WTX1. In the absence of Wnt, CK1 and GSK3 phosphorylate Axin-bound β-catenin at a series of Ser/Thr residues[31]. Phosphorylated β-catenin is recognized by the F box/WD repeat protein β-TrCP[32], a component of an E3 ubiquitin ligase complex, which binds β-catenin inducing ubiquitination and subsequent proteasomal degradation[33]. The low cytosolic levels of β-catenin keep the Wnt pathway inactive (Figure 1).

Interaction between Wnt molecules (or their agonists) leads to conformational change of the receptor complex; the DC moves to the cell membrane with Axin2 being recruited to the phosphorylated tail of LRP5/6 and GSK3 being inactived by phosphorylated LRP6[34]. The disassembly of the DC leads to β-catenin stabilization, translocation to the nucleus and interaction with T-cell factor/lymphoid enhancer factor (TCF/LEF) transcription factors, rendering them transcriptionally active (Figure 1).[35]

TCF proteins are the main effectors of the Wnt signaling output, regulating a plethora of target genes[36]. All the members of the family contain a HMG box and a small peptide motif of basic residues, which both comprise the DNA-binding domain that recognizes specific consensus sequences, called Wnt responsive elements (WREs)[37,38]. In the inactive state, TCF transcription factors interact with Groucho transcriptional repressors, which mediate histone deacetylation and chromatin packaging, repressing Wnt target activity[39,40]. In the active state of the pathway, β-catenin translocates to the nucleus and interacts with a conserved motif in the amino terminus of TCF proteins, displacing Groucho and recruiting other co-activators that enhance gene transcription[41,42]. A genome-wide mapping of TCF4 chromatin occupancy in the intestine revealed that the target genes are characterized by multiple TCF4 binding sites, which are usually located distally from the transcription start site (TSS)[43]. Interestingly, many TCF4 binding regions are located in genomic regions at large distances

from the closest annotated genes or TSSs, indicating that there are still unknown targets and functional elements that remain elusive[43].

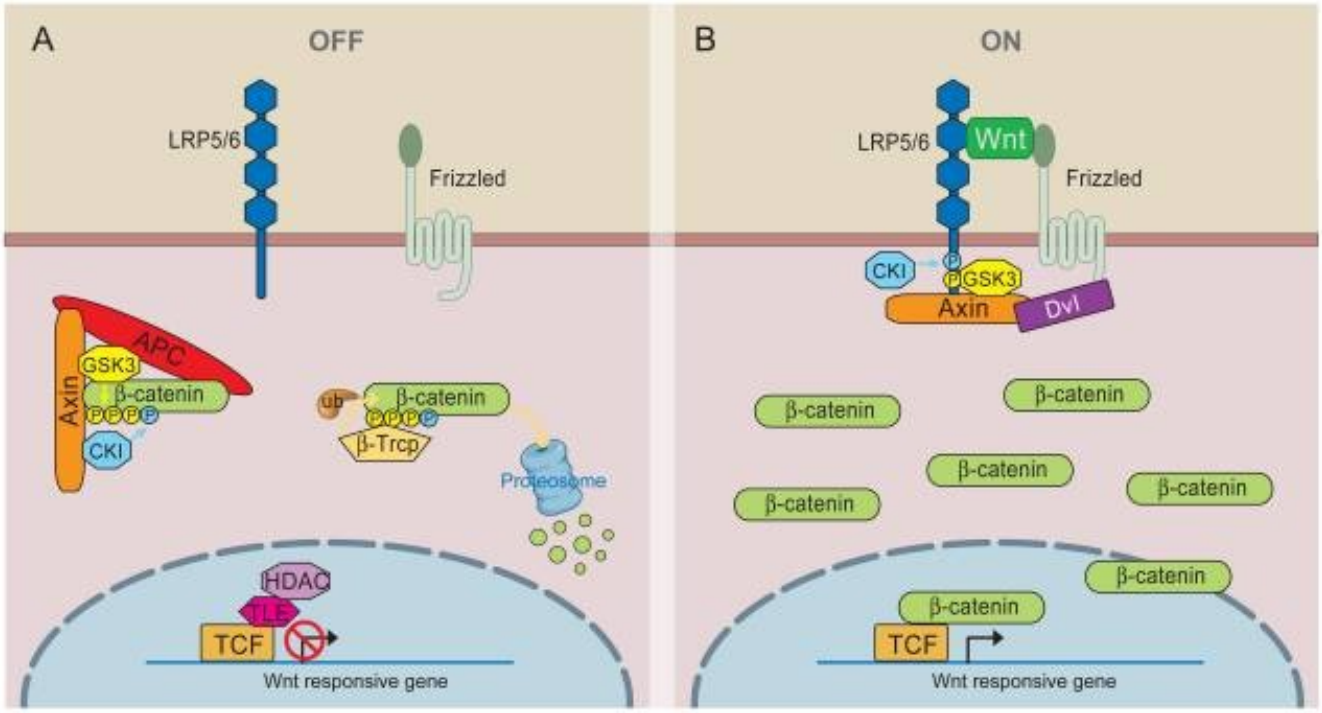

Figure 1: Graphical illustration of the canonical Wnt pathway. During the "OFF" state, the DC degrades β-catenin, while GROUCHO proteins repress TCF4. On the other hand, in the "ON" state, disassembly of the DC leads to stabilization of cytosolic β-catenin, which in turn, translocates to the nucleus activating TCF4 and initiating the Wnt transcriptional program (MacDonald et. al. 2009, Developmental Cell).

1.1.4 Wnt signaling in stem cell maintenance and diseases

The Wnt signaling cascade is among the most essential pathways that dictate stem cell fate and development[1,2,41]. Stem cells possess the unique capacity to self-renew and differentiate into specialized cells, playing a pivotal role in tissue/organ development, architecture and homeostasis. Wnt proteins are responsible for maintaining stem cell identity in all stem cell types, while other growth factors promote their proliferation[44]. The first insights into the significance of Wnt/β-catenin pathway in lineage fate were revealed in the intestinal epithelium, where deletion of TCF4 in mice, resulted in

tissue breakdown due to loss of the intestinal stem cells[44]. In addition, inhibition of Wnt signaling through the overexpression of DKK (a Wnt antagonist) blocked the initiation of hair follicle formation and the growth of the mammary gland[45]. In the haematopoietic system, overexpression of Axin reduces the numbers of haematopoietic stem cells and constrains their growth[46].

In the opposite context, constitutively activated Wnt pathway results in stem cell expansion and increased self-renewal. Mice expressing truncated, non degradable forms of β-catenin in keratinocytes are characterized by de novo hair follicle growth[47]. Other studies have shown that upon Wnt treatment, mammary stem cells can continue to self-renew and evade differentiation for long periods of time[48]. In similar approaches, Lgr5 and Axin2 are found significantly up-regulated due to Wnt activation, making them exquisite markers for studying stem cells in lineage tracing experiments[49–51].

Because the Wnt pathway plays a central role in intestinal stem cell maintenance, it becomes evident that deregulation of this signaling cascade might lead to colorectal cancer. The intestinal track is divided to the small intestine and the colon. The luminal surface of the small intestine consists of a layer of intestinal epithelial cells. The intestinal epithelium is the most rapidly renewing tissue in the human body and is anatomically divided into small pits, the crypts, which are located between the villi, epithelial extrusions in the lumen of the small intestine.

Intestinal stem cells reside in the crypts (stem cell niche) and generate the transient amplifying cells (or progenitors) which in turn, migrate to the villi for terminal differentiation. Continuous tissue renewal depends on the migration of these intestinal stem cell progenitors to the villi where they can differentiate to epithelial cells. After the process of differentiation, enterocytes, goblet cells and enteroendocrine cells continue migrating upwards until they become apoptotic. Only the Paneth cells do not follow this cycle, but they differentiate and migrate to the bottom of the crypts[52].

While stem cell self-renewal is essential for the development and physiology of the organism, in the case of genetic or epigenetic “mistakes”, cells might proliferate uncontrollably leading to tumor growth. According to the Cancer Stem Cell model, it has been proposed that intestinal stem cells operate as the cell of origin for cancer initiation forming the basis of a tumor hierarchy and giving rise to other progenitor cells[52].

Mutations in the APC gene are causative for a type of hereditary colon cancer which is called Familiar Adenomatous Polyposis (FAP)[53,54]. Loss of APC function does not allow the assembly of the DC, stabilizing β-catenin in the nucleus and keeping the Wnt pathway constitutively active[55]. Additionally, mutations in Axin2 underlie a genetic predisposition for hereditary colorectal cancer[56].

Moreover, inactivating mutations can occur in a plethora of other crucial molecules of the pathway, such as β-catenin[57], TCF[58] family and RNF43[59] . Such mutations have been also discovered in other types of cancers and diseases, such as hepatocellular carcinoma[60], melanoma[56], pancreatic cancer[59], metabolic disorders[61–63] and degenerative diseases[60].

The numerous implications of the Wnt pathway in cancer, make it a promising therapeutic target. While the strongest candidate for blocking the pathway is the β-catenin-TCF4 complex, which is the main mediator of the cascade, there have not been successful efforts due to its' complex structure[64]. IAlternative approaches try to i) target upstream effectors of the pathway (e.g Axin2 stabilization in the cytoplasm)[65,66], ii) block Porcupine[65] or iii) intervene with ligand-receptor interaction[67].

1.2 Long non coding RNAs

By the time the Human Genome project was completed, it became obvious that only a small proportion of the genome (~2%) encodes proteins. At the same time, the fact that evolutionary less complex organisms contain similar number of protein-coding genes as humans, points to the existence of underlying organismal complexity in the non-coding genome[68]. The advancement of the sequencing technologies coupled with the elevated interest of the researchers to uncover the putative functions of "junk DNA" led to the discovery of non-coding RNAs and their implication in organismal development and physiology[68,69].

Long non-coding RNAs (lncRNAs) are transcripts of more than 200 nucleotides that do not code for proteins. The group of lncRNAs is characterized by great variety, as it consists of intergenic transcripts, enhancer RNAs (eRNAs) and sense or antisense intragenic transcripts that co-reside in genomic loci with other protein-coding genes[70]. Even though a great number of lncRNAs have been identified, their functions still remain insufficiently characterized. They orchestrate gene expression by acting *in cis* or *in trans*, organize chromatin architecture or even regulate other proteins and transcripts[69].

1.2.1 Identification and characterization of lncRNAs

One of the major challenges of the field is the fact that lncRNAs are loosely defined and comprise a plethora of transcripts with diverse characteristics and actions. It is a challenging task to distinguish functional lncRNAs from "transcriptional noise" which consists of annotated transcripts that are

expressed in low levels, have minimal sequence conservation or do not exert a significant biological output[68–71].

LncRNAs are classified based on their genomic location relative to a protein coding gene. Antisense lncRNAs are defined as having a transcriptional start site (TSS) in the 3' UTR or inside a protein-coding sequence and are transcribed in the opposite direction, while overlapping at least one coding exon. Intronic lncRNAs are transcribed from an intronic sequence of a protein-coding gene and do not overlap any exons. Bidirectional lncRNAs are transcripts that initiate divergently from a protein coding promoter. Long intergenic non coding RNAs (lincRNAs) are lncRNAs with independent transcription that do not overlap other protein coding genes (Figure 2)[70]. These transcripts are polyadenylated, capped and frequently spliced, similar to classical mRNAs[69]. Due to their putative functionalities and their interesting tissue-specific expression pattern, lincRNAs have drawn stronger attention[72].

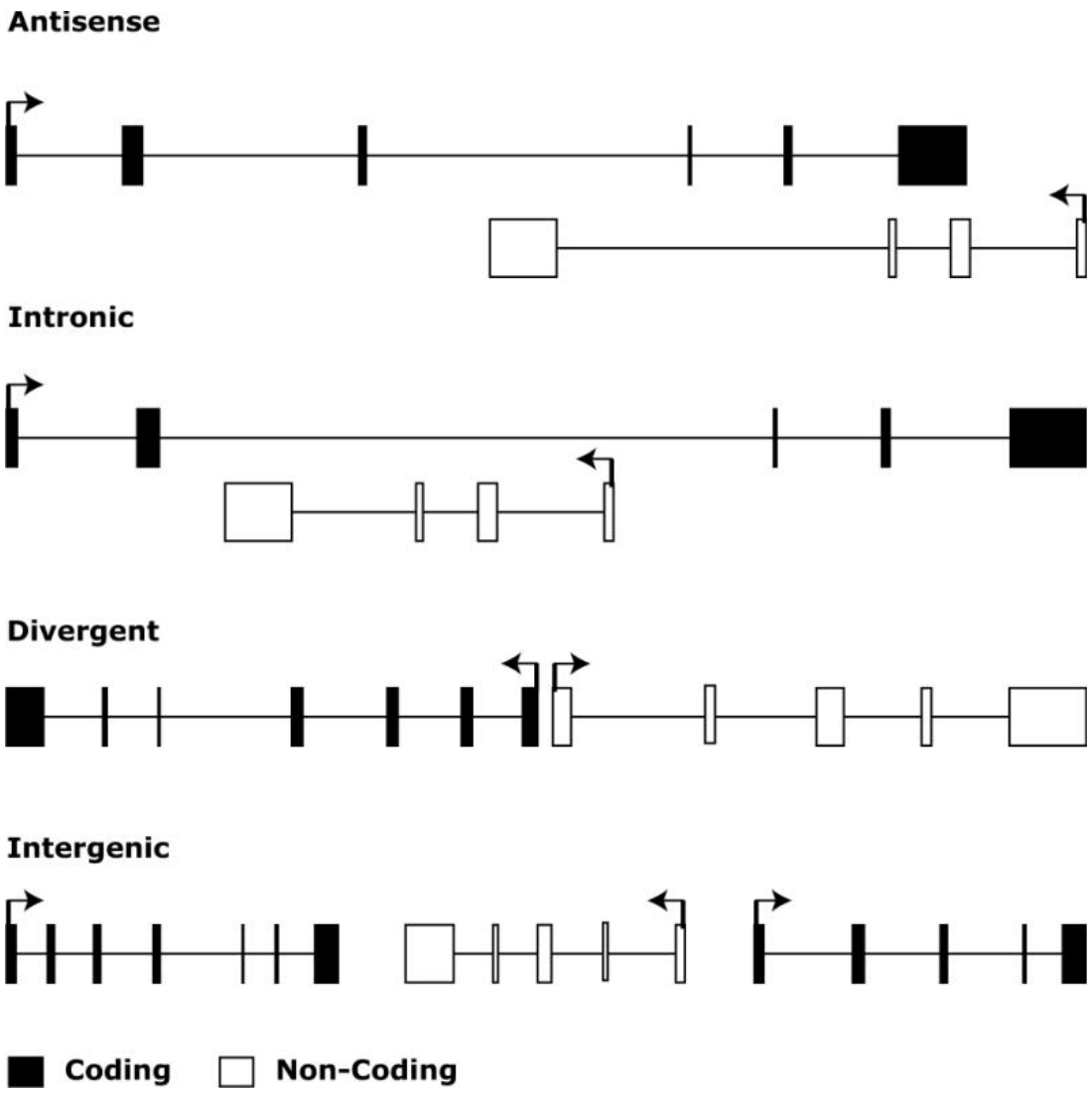

Figure 2: Anatomy of lncRNA loci. Antisense lncRNAs initiate inside a protein coding gene and are transcribed in the opposite direction. Intronic lncRNAs initiate in a protein coding gene's intron and do not overlap any exon. Divergent lncRNAs initiate in a bidirectional manner from the promoter of a protein coding gene. Intergenic lncRNAs are transcribed from genomic regions that are separate from protein coding genes.

The identification of functional lncRNA genes requires the characterization of the accurate TSS, genomic coordinates, splice sites and polyadenylation sites of the transcripts. A combination of sequencing technologies is crucial for defining long non-coding transcripts correctly. The FANTOM project was the first effort to massively report the number of lncRNAs in the mouse, using cDNA cloning followed by Sanger Sequencing[73]. Afterwards, the catalog of lncRNAs was redefined by taking into account information from chromatin marks and the mRNA expression levels. Chromatin

immunoprecipitation followed by sequencing (ChIP-seq) revealed that large numbers of novel transcribed genes were occupied by polymerase II, with histone H3 lysine 4 trimethylation (H3K4m3) marking active promoters and H3 lysine 36 trimethylation (K4-K36 domains) defining the active transcriptional unit[74–76]. Additionally, progress in RNA sequencing (both experimentally and computationally) allowed the identification of the active transcripts and the precise quantification of their abundance[77–80]. To sum up, RNA-sequencing enabled the detection of novel transcripts (even the less abundant ones), while chromatin modifications pointed out the exact genomic coordinates where the transcript is mapped.

Another great task in characterizing lncRNAs is to evaluate whether the transcript is translated or not. It is common for lncRNAs to contain putative open reading frames (ORF) because of their considerable length, making it necessary to use other criteria for distinguishing non coding transcripts[69]. Scientists have tried to define protein-coding capacity by performing homology queries of the ORFs against protein domain databases (or by determining the putative higher conservation of the codons for amino acids)[81]. An alternative methodology is ribosome profiling through which the RNA regions that are associated with translating ribosomes can be mapped[82]. The latter method is important for discovering lncRNAs with small ORFs, which cannot be identified with other computational methods, but might encode small peptides[83].

1.2.2 Molecular mechanisms of lncRNAs function

In a broad context, lncRNAs can be classified to those that function *in cis*, affecting the expression or the chromatin conformation/state of nearby genes or *in trans*, influencing other factors directly or gene expression at greater distances [70]. LncRNAs can modulate the expression of neighboring genes by recruiting chromatin modification complexes on their genomic site. A characteristic example is the lncRNA Xist which is pivotal in silencing one of the two X chromosomes for dosage compensation in females. Xist interacts with the Polycomb Repressive Complex 2 (PRC2) which decorates one X chromosome with repressive chromatin marks, leading to its inactivation[84].

Local gene regulation can also be affected by the process of transcription of a lncRNA itself, independently of the produced transcript. Antisense Igfr2 noncoding RNA (Airn) is an antisense

lncRNA that overlaps with the promoter and the gene body of the imprinted *Igf2r* locus. Airn silences the paternal allele by interfering with RNA polymerase II (Pol II) occupancy through its antisense transcription through the *Igf2r* promoter[85]. Similarly, even the transcription of lncRNAs that do not overlap with protein-coding genes, can dictate the local chromatin architecture and the binding of transcription factors (TF) or Pol II on nearby promoters and enhancers[86,87].

DNA elements that are localized in the promoters or gene bodies of lncRNA loci can also exert *cis* activity independently of the produced transcript. For example, lincRNA-p21 which is transcribed from a locus close to *CDKN1A* gene, regulates PRC2 target genes by recruiting heterogenous ribonucleoprotein K (hnRNP-K) and p53 to the *CDKN1A* promoter[88]. The elements of this locus are significant for the regulation of neighboring genes even in tissues where the transcript is not expressed[89].

On the other hand, there are several lncRNAs that move away from their site of transcription and can alter gene expression *in trans* by regulating the state of the chromatin[70]. The HOX antisense intergenic RNA (HOTAIR) is a lncRNA that is transcribed from the *HOXC* locus, but regulates the distant *HOXD* gene family[90]. The family of HOX genes is essential for patterning the developmental axis of organisms and defects in their regulation can result in severe developmental disorders[91]. Knock down of HOTAIR leads to augmented expression in the *HOXD* gene locus and decrease of the repressive chromatin mark H3K27me3. Interestingly, HOTAIR functions as a scaffold which recruits PRC2 and in turn, determines the chromatin state which dictates the expression levels of *HOXD*[91].

Moreover, several lncRNAs have the capacity to affect the nuclear architecture and regulate the transcriptional machinery and the spliceosome in order to control gene expression *in trans*. Metastasis associated lung adenocarcinoma transcript 1 (MALAT-1) is a lncRNA that interacts with multiple splicing components in nuclear speckles, aiding them to correctly position upon active gene bodies and nascent mRNAs[92,93]. Another lncRNA that coordinates nuclear structure is the nuclear enriched abundant transcript 1 (NEAT1) which is essential for the maintenance of paraspeckles and controls mRNA nuclear reterntion[94]. Both MALAT-1 and NEAT1 interact with actively transcribed gene bodies, bridging their corresponding nuclear domains with transcribed loci, subsequently affecting the transcription of nascent RNAs[70,93].

LncRNAs can also bind to RNA or proteins, altering their activity or abundance in a stoichiometric manner. Frequently, a lncRNA can function as a decoy, titrating away or inhibiting the function of other proteins. An example of this class is the non-coding RNA activated by DNA damage (NORAD), a

cytoplasmic lncRNA that binds to RNA binding proteins PUMILIO1 (PUM1) and PUMILIO2 (PUM2) blocking them from degrading their mRNA targets[95,96]. Additionally, other lncRNAs bind to miRNAs, titrating them away from their mRNA targets and increasing the expression of the latter[97].

1.2.3 Crosstalk mechanisms of Wnt/β-catenin signaling pathway and lncRNAs

The Wnt signaling cascade consists of a plethora of molecules which combinatorially can influence global gene expression. Recent evidence suggests that there are unexplored layers of regulation of the Wnt pathway, including both its targets and its mediators. Investigation of lncRNAs and their direct or indirect contribution in the Wnt/β-catenin pathway has shed light into the complexity of the Wnt-dependent target repertoire[98–101]. Moreover, the fact that target sites of β-catenin/TCF4 are located in distal genomic regions, points to the significance of non coding transcripts to Wnt-dependent functions[43,102]. Due to their great heterogeneity, Wnt-regulated lincRNAs can be further classified based on their molecular mechanism.

LncRNAs can physically interact with β-catenin or TCF proteins, ultimately modulating the transcription of Wnt target genes. CCAT2 (Colon Cancer Associated Transcript 2) is a lncRNA that resides in the proximity of the *MYC* gene, encompasses the CRC-associated rs6983267 SNP and promotes colorectal carcinogenesis and metastasis[103,104]. Specifically, TCF4 binds to the consensus TCF4 element in *CCAT* that also contains the rs6983267, increasing the expression of the lncRNA which in turn, binds to TCF4 and activates *MYC*[103]. Augmented activity of *MYC* results in higher expression of MYC target genes which are implicated in genomic instability and tumor growth, completing a positive regulatory feedback loop that explains the elevated association of this SNP with colorectal cancer[103,104]. In an alternative example, RBM5-AS1 lncRNA (RBM5 antisense 1) binds β-catenin, enhancing its interplay with TCF4. Consequently, high levels of RBM5-AS1 cause reinforced occupancy of β-catenin/TCF4 complex on its target genes, boosting Wnt signaling and self-renewal of colon cancer cells[105]. Opposite different approach is taken by CCAL (colorectal cancer-associated lncRNA) that indirectly activates the Wnt pathway by interacting with AP-2a (activator protein 2alpha). AP-2a normally functions as a negative regulator of the pathway; upon interaction with CCAL, AP-2a is ubiquitinated and consequently degraded by the proteasome[106].

As mentioned before, lncRNAs can regulate the expression of other genes by forming lncRNA-protein complexes. CASC11 (Cancer susceptibility candidate 11) forms a ribonucleoprotein complex

with the hnRNP-K protein, stabilizing it in the cytoplasm. Upon stabilization, hnRNP-K disassembles the DC, allowing β-catenin to accumulate and translocate to the nucleus initiating a Wnt-depended transcriptional program[107]. Among the genes activated is MYC which binds to the promoter of *CASC11* and enhances its transcription, generating a positive regulatory feed-forward loop that keeps the Wnt pathway in the "On" state[107]. MYU (c-Myc-upregulated lncRNA) is another lncRNA in the same class that is a direct target of c-Myc and interacts with hnRNP-K. Upon their interaction, they suppress miR-16 transcription, which normally inhibits cyclin-dependent kinase 6 (CDK6), ultimately leading to increased cell proliferation[108].

LncRNAs can also modulate the Wnt signaling pathway by assembling chromatin modifying complexes to the promoters of specific gene loci. Lnc34a recruits DNMT3a/PHB2 and HDAC1 at the *miR-34a* promoter, resulting in the methylation and histone deacetylation of this genomic site[109]. Mir34-a has an important role in the regulation of Wnt pathway, targeting various Wnt-pathway components such as the TCF7 transcription factor and decreasing their mRNA abundance[110]. Therefore, high expression of Lnc34a keeps mir34-a at low levels, enhancing indirectly the activity of the pathway and cancer stem cell renewal[109]. On the other hand, NBAT1 (neuroblastoma associated transcript-1) is a lncRNA that interacts with the catalytic subunit of the PRC2 complex, titrating the latter away from its target genes and reducing the activity of Wnt pathway[111]. Binding of NBAT1, prevents the PRC2 complex from epigenetically modifying the promoter of *DKK1* (Dickkopf-related protein 1) which functions both as target and as negative regulator of the Wnt pathway[111,112].

LncRNAs play a pivotal role in structuring higher order chromatin and controlling the spatiotemporal patterns of gene expression. An example of this class is WiNTRLINC1 (Wnt-regulated lincRNA1) which is a direct target of the β-catenin-TCF4 complex. Knock down of WiNTRLINC1 decreases the expression of the neighboring gene *ASCL2* and leads to increased apoptosis and G2-cell cycle arrest. ASCL2 is a transcription factor that is essential for stem cell maintenance. The two loci interact by forming a chromatin loop that connects the TSS of *WiNTRLINC1* with an enhancer that is located downstream of *ASCL2*. The WiNTRLINC1 transcript is also necessary for the formation of the chromatin loop, functioning as a scaffold for recruiting Pol II and the β-catenin-TCF4 complex at the regulatory regions of *ASCL2* locus. At the same time, ASCL2 binds to the promoter of *WiNTRLINC1*, completing a feedforward regulatory loop that maintains stemness[113].

LncRNAs can indirectly affect the Wnt pathway by competing against endogenous miRNAs, ultimately affecting the expression levels of key proteins of the pathway. Such an example is HNF1-

AS1, a lncRNA that binds to mir34a, blocking it from binding to its' target, Sirtuin 1 (SIRT1)[114]. The SIRT1 protein plays a crucial role in the inactivation of p53. Both miR34a and p53 participate in a regulatory loop that suppresses the Wnt pathway[115,116]. Therefore, the competitive binding of HNF1-AS1 to the microRNA leads to increased expression of SIRT1 and inactivation of p53, resulting in induction of the Wnt pathway[114].

1.3 Identification and characterization of WiNTRLINC3

We have identified Wnt-regulated lncRNAs by taking advantage of CRC cell lines engineered to inducibly overexpress a small hairpin RNA (shRNA) against β-catenin or a dominant negative mutant form of TCF4, shutting down the Wnt pathway in those cells[117,118]. 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 abrogation, reveal sites of Wnt-depended transcription[119].

WiNTRLINC3 was identified as a gene that is negatively regulated by the Wnt pathway (Figure 3). The *WiNTRLINC3* gene is located on the long arm of chromosome 1 (hg19 coordinates: chr1:159931014-159948876), consists of three exons and transcribes two polyadenylated isoforms. *WiNTRLINC3* does not have protein coding potential according to phyloCSF[120] and in vitro transcription/translation experiments. In addition, the lincRNA is mainly nuclear localized as shown by cell fractionation and FISH experiments (unpublished data). Re-analysis of the RNA-seq data of The Cancer Genome Atlas (TCGA) revealed that WiNTRLINC3 is significantly down-regulated in patient-derived adenomas compared to normal tissue, indicating that it could be used as a predictive biomarker for Colorectal Cancer.

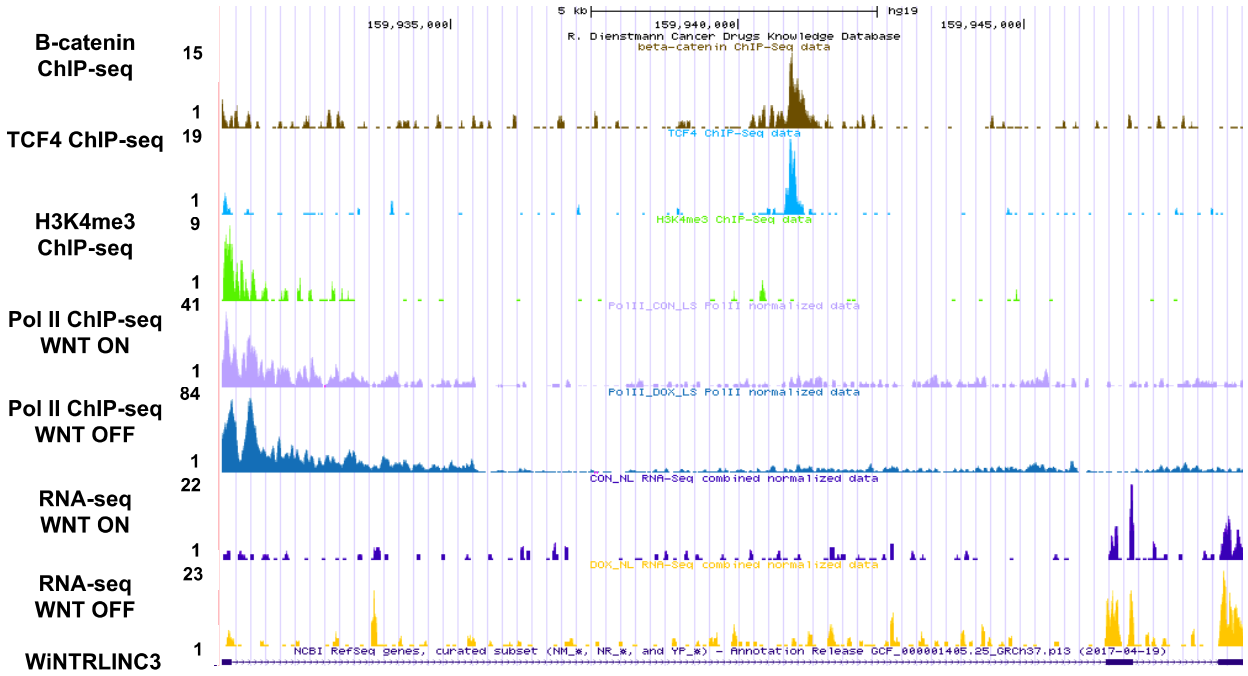

Figure 3: Snapshot showing the β-catenin, TCF4, H3K4me3 and Pol II occupancy along with the RNA-seq reads at the WiNTRLINC3 locus, before and after Wnt abrogation.

In order to elucidate the function of WiNTRLINC3, we conducted RNA pull down experiments in the colorectal cancer cell line LS174 and identified WiNTRLINC3-interacting proteins. 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 that is also necessary for the binding of TCF4 on specific Wnt targets, interacts directly with the WiNTRLINC3 transcript[121,122].

2 Aim of the study

The purpose of this study is to decipher the role of WiNTRLINC3 in intestinal physiology and carcinogenesis. This project is part of our laboratory's attempt to identify and characterize lncRNAs that are regulated by the Wnt pathway. Understanding how those lncRNAs are implicated in Wnt signaling and ultimately affect intestinal carcinogenesis, will help us shed light into unexplored layers of regulation by the Wnt pathway. Moreover, lncRNAs with significant clinical impact could be used as putative biomarkers or therapeutic targets. Our studies should also contribute to further understanding the significance of the vast numbers of non coding RNAs that are expressed from animal genomes.

In order to functionally characterize WiNTRLINC3 in colorectal cancer lines:

I) We evaluate the differentially expressed genes and the main affected pathways upon *cis* overexpression of WiNTRLINC3.

II) We knock down cis-overexpressed WiNTRLINC3 and examine the expression patterns of its predicted targets to determine if the effects produced by cis-overexpression of WiNTRLINC3 are dependent on the WiNTRLINC3 transcript or only the act of transcription of the *WiNTRLINC3* genomic locus.

III) We will investigate whether CDX2 interacts with *WiNTRLINC3* to affect expression of downstream genes.

3 Materials & Methods

3.1 *In cis* overexpression of *WiNTRLINC3* with CRISPR-Cas9

In order to overexpress the WiNTRLINC3 transcript in cis, we used a LS174T-derived colon cancer cell line that has been engineered to constitutively overexpress nuclease-deficient Cas9 (dCas9) fused to VP64 and HSF1 transactivating domains (LDC10 cell line). dCas9 can precisely recognize target sites through hybridization of a Cas9 interacting guide RNA (gRNA) to precise 20 bp DNA sequences, but cannot cleavage them. Instead, upon binding, it functions as a scaffold that recruits the transcriptional machinery and chromatin remodellers[123]. To achieve *WiNTRLINC3* overexpression, we designed two single guide RNAs (sgRNA 1 and sgRNA 2) that target the promoter of the gene , one sgRNA (sgRNA 3) that targets the enhancer and one non targeting sgRNA as control (NTG). The sgRNAs were cloned into the pLV-U6-gRNA-diMS2-EF1Alpha-Puro-T2A-BFP vector.

To achieve efficient delivery of the sgRNA-transcribing vectors into the LDC10 cells, we used a lentiviral approach. For lentiviral production, we plated 4000000 lentiX-293T cells (293T cell line that is easily transfectable and supports high level expression of viral proteins) in p10 cell culture dishes that have been previously coated with 1mg/ml collagen. The next day we transfected the cells with a DNA mix consisting of 3.75 ug plasmid vector containing the gene to be delivered, 1.32 ug of pMD2.G envelope vector and 2.43 ug of psPAX packaging vector. LentiX-293T cells were transfected with Polyethylenimine (PEI)[123]. After testing the transfection efficacy, we collected the virus supernatant and filtered it through a 0.45μm sterile filter.

Based on previous titration experiments, we transduced 250000 LDC10 cells with 1 ml of the 14 ml produced virus. The transduction was assisted with the addition of the cationic polymer polybrene (Hexadimethrine bromide, 8μg/ml) which significantly enhances the infection efficacy.

The sequences of the sgRNAs that were used in this study are shown in the following table:

Primer name	Target site	Sense oligo	Antisense Oligo
WiNTRLINC3 sgRNA1	GGAAAATGAGACAAGGAA	TTGGGGAAAATGAGACAAG GAAGTTTAAGAGC	TTAGCTCTTAAACTTCCTTG TCTCATTTTCCCCAACAAG
WiNTRLINC3 sgRNA2	GGAGGAAAGCCAAAGAAA	TTGGGGAGGAAAGCCAAA GAAAGTTTAAGAGC	TTAGCTCTTAAACTTTCTTT GGCTTTCCTCCCCAACAAG

WiNTRLINC3 sgRNA3	CAGAGGGACTGGAGAACAA	TTGGCAGAGGGACTGGAGAACAAGTTTAAGAGC	TTAGCTCTTAAACTTGTTCTCCAGTCCCTCTGCCAACAAG
Non targeting guide (NTG)	GACCAGGATGGGCACCACCC	TTGGGACCAGGATGGGCACCACCCGTTTAAGAGC	TTAGCTCTTAAACGGGTGGTGCCCATCCTGGTCCCAACAAG

3.2 Knock down of WiNTRLINC3 and IGSF9

In order to perform knock down experiments, we utilized a lentiviral system for stably overexpressing shRNAs against target-transcripts. We designed one shRNA oligo that targets the second exon of *WiNTRLINC3* and three shRNAs that target multiple exons of IGSF9. The shRNA oligos were cloned into the pSICOR PGK Puro vector (Addgene plasmid # 11579)[123]. The annealed oligos were ligated with 50 ng of HpaI/SlaI (Minotech) digested pSICOR vector by using T4 Ligase (NEB) overnight (O/N) at 16 C°. DH5a competent bacterial cells were transformed with the recombinant constructs and grew in agar plates with ampicillin. Single colonies were selected and after growing them in LB (Luria Broth) medium at 37ºC O/N, we extracted the plasmid DNA (Miniprep). Successful recombinant constructs were distinguished from the negative ones, after double digestion with SlaI and XbaI restriction enzymes, generating a fragment of 373 bp compared to a 333 bp fragment of the empty vector.

Production of lentiviruses and transduction of LDC10 cells was conducted as described in the previous section.

The sequences of the shRNAs used are shown in the following table:

Primer name	PSICOR sense oligo	PSICOR antisense oligo
IGSF9 sh1	TTGGAATTGCTGGAGACTTTTTTTCAAGAGAAAAAAGTCTCCAGCAATTCCATTTTTTC	TCGAGAAAAAATGGAATTGCTGGAGACTTTTTTCTCTTGAAAAAAAGTCTCCAGCAATTCCAA
IGSF9 sh4	TCAGAAGAGTGAGAGCTGAATTCAAGAGATTCAGCTCTCACTCTTCTGTTTTTTC	TCGAGAAAAAACAGAAGAGTGAGAGCTGAATCTCTTGAATTCAGCTCTCACTCTTCTGA
IGSF9 sh5	TCGTGGAAGCTCCGAGGAAATTCAAGAGATTTCCTCGGAGCTTCCACGTTTTTTC	TCGAGAAAAAACGTGGAAGCTCCGAGGAAATCTCTTGAATTTCCTCGGAGCTTCCACGA
WiNTRLINC3 sh1	TGGGAGGAGGTAAAGAGTAGTTCAAGAGACTACTCTTTACCTCCTCCCTTTTTTC	TCGAGAAAAAAGGGAGGAGGTAAAGAGTAGTCTCTTGAACTACTCTTTACCTCCTCCCA

Scrambled (scr)	TGTACAGCCGCCTCAATTCTTTCAAG AGAAGAATTGAGGCGGCTGTACTTTT TTC	TCGAGAAAAAAGTACAGCCGCCTCA ATTCTTCTCTTGAAAGAATTGAGGCG GCTGTACA

3.3 RNA extraction and Quantitative PCR (qPCR) – 3' mRNA Quant-seq

Total RNA was isolated with TRI reagent (Molecular Research Center) and 10 μg were treated with DNase I (Promega). Dnase I treated RNA was purified with phenol/chloroform and ethanol precipitation. 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 of 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.

The primer sequences we used are shown in the following table:

Primer name	Sequence
IGSF9 F1	CCGAGAGATGAATGTGGATGG
IGSF9 R1	GGAGAACGAAGGAAAGATGAGG
IGSF9 F3	CTGGCTTTGATGGTGGTTATCT
IGSF9 R3	ACCCAGTCATGGTGCATTC
WiNTRLINC3 F4	GAATGGTTGGGAGGAGGTAAAG
WiNTRLINC3 R4	CTGGGCTCAAGGAATCTGAATAG
RPLP1 F1	AAGCAGCCGGTGTAAATGTTGAGC
RPLP1 R1	CATTGCAGATGAGGCTCCCAATGT
GAPDH F1	ACATCAAGAAGGTGGTGAAGCAGG
GAPDH R1	TGTCGCTGTTGAAGTCAGAGGAGA

Next Generation Sequencing of the extracted RNA was conducted by the Fleming Genomic Facility, using the Quantseq 3' mRNA-sequencing (Quantseq) methodology[124]. For each condition (LDC10-sgRNA 1, LDC10-sgRNA 2, LDC10-NTG) we used three replicates for reproducibility.

Quantseq is a novel protocol of RNA sequencing with high level performance in differential expression analysis. In Quantseq, the libraries are generated in a multistep workflow. Firstly, oligodT primers bind on the poly(A) tail of the 3' segment of mRNAs and a complementary cDNA strand is synthesized with reverse transcription. Then, the RNA template is removed and a second strand synthesis follows, based on random primers. The double stranded cDNA can be purified with magnetic beads and amplified with PCR. Next generation sequencing is the final step of the process, before the data processing.

The major feature of Quantseq is the generation of libraries of sequences that mostly cover the 3' end of polyadenylated RNAs. Moreover, the protocol is simple and produces datasets of smaller sizes, making the analysis much quicker. As only one fragment is generated for each transcript, the expression of a gene can be quantified proportionally from the number of reads that are generated, without the need for complex normalization procedures[124].

3.4 Luciferase reporter assay

In order to clone the promoter and enhancer regions of WiNTRLINC3 we performed qPCRs with Taq DNA-polymerase (Minotech) using as template cDNA synthetized from LS174T-extracted RNA. The PCR-amplified products were then extracted from an agarose gel with the Nucleospin Gel and PCR clean up kit (Macherey Nagel). The extracted fragments were ligated into the pGEM-T vector using T4 Ligase (NEB), transformed into competent DH5a bacterial cells, followed by plasmid extractions (Minipreps).

The recombinant vectors were digested with EcoRI High Fidelity restriction enzyme (NEB) and the excised promoter and enhancer fragments were blunt-ended by Klenow (NEB) and isolated with PCR clean up (as described above). The promoter and enhancer fragments were then ligated into EcoRV-digested Pgl4.10 and Pgl4.10-ADML vectors respectively, followed by transformation of DH5a competent cells, Miniprep plasmid extraction and diagnostic digests.

We also cloned the CDX2 coding region into the pcDNA3 expression vector. For this purpose, RNA was extracted from CaCO2 cells and the synthesized cDNA (as described before) was used as a template for CDX2 PCR amplification. We used primers containing a flag tag before the CDX2 specific targeting sequences and performed the PCR with Q5 High Fidelity DNA Polymerase (NEB).

The PCR product was cleaned by gel extraction (as described before). The pcDNA3 vector was digested with EcoRV restriction enzyme and then dephosphorylated by Shrimp Alkaline Phosphatase (SAP). The PCR product was ligated into pcDNA3, using T4 DNA Ligase. We transformed DH5a competent bacterial cells, cultured single colonies and extracted the plasmid DNA (minipreps). We distinguished the recombinant products after PstI digestion, generating two fragments at ~2180 bp and 180 bp.

The sequences of the primers we used are shown in the following table:

Primer name	Sequence
WiNTRLINC3 promoter F1	GTCCTTGCGATAGTTTGCTGAG
WiNTRLINC3 promoter R2	CAGTACCATGGCTCTGCCGTG
WiNTRLINC3 enhancer F1	GAAGTACTAAGTACGTCCTTTCC
WiNTRLINC3 enhancer R1	GCAGGATTGAGGAGTCCAGTTTG
CDX2 Clon F	ATCATGGACTACAAAGACGATGACGACAAGTACGTGAGC TACCTCCTGGACAAGGACGTG
CDX2 Clon R	GATCTATCACTGGGTGACGGTGGGGTTTAGCACC

Luciferase Reporter Assay was conducted with the Promega Dual Glo Luciferase assay system. We plated 150000 LDC10 cells in 24-well plates. The next day we transfected the cells with a DNA mix, using PEI. For each plate, we mixed 50 ng of reporter plasmid, 10 ng of Renilla plasmid (as normalizer), 100 ng of pcDNA3-CDX2 vector (if needed) and stuffer DNA (pcDNA3) to reach 460 ng of total DNA mix. We generated duplicates for each transfection. In the cases that we transduced the cells with lentiviral vectors, we performed the plating and the transduction of the cells at the same time and we transfected them the next day. 24h post-transfection, we performed the Luciferase assay by lysing the cells with 100 μl Passive Lysis Buffer for 30 minutes, while rotating. For each sample, we mixed 20 μl of cell lysate with 20 μl Dual Glo substrate, estimated the luminescence of the Firefly luciferase and then added 20 μl Dual Glo Stop & Glow to estimate the luminescence of Renilla luciferase.

3.5 Colony Formation Assays

For colony formation assay, 2x103 Ls174T cells that were transduced with sgRNA1 or sgRNA2 were grown with and without the addition of doxycycline for 2 weeks in 100 mm tissue culture dishes (Sarstedt). Colony formation was visualized after staining with crystal violet which stains only viable cells as it binds to proteins and DNA. All experiments were performed in three independent biological replicates.

3.6 Sequencing analysis

3.6.1 Quant-seq 3' mRNA-sequencing analysis

Sequencing data were aligned to the human reference genome (build hg19) with topHat[78]. Differential Expression Analysis was conducted with bioconductor package metaseqR[125]. A gene model was assembled consisting of the 3'UTR sites of the genes and used for the counting of the reads. Prior to statistical procedures, the gene read counts were filtered for possible artifacts that could affect the subsequent statistical testing procedures. Genes presenting any of the following were excluded from further analysis: i) genes with length less than 500 bp, ii) genes whose average reads per 100 bp were fewer than the 25th quantile of the total normalized distribution of average reads per 100bp, iii) genes with read counts below the median read counts of the total normalized count distribution. The final table with the remaining genes (after filtering) was subjected to Differential Expression Analysis with the DESEQ algorithm[126]. We considered as significantly differentially expressed genes (DEG) those with an absolute natural fold change ≥ 1.5 and p value < 0.05.

3.6.2 ChIP-sequencing analysis

We re-analysed the ChIP-sequencing data from the study with accession number GSE97273[127]. We checked the quality of the sequencing with the FASTQC program (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were aligned to the hg19 build with bowtie2 aligner[128]. Peak calling was performed with MACS2, after down-sampling the input's library size to the library size of the treatment file (ChIP sample)[129].

Distribution of the ChIP-seq reads on important genomic features was calculated with Cis-regulatory Element Annotation System (Ceas)[130]. We used hg19 build as a reference genome and

selected to find the distribution in intervals of 666 bp 1333 bp and 2000 bp distance from gene promoters. For the other options we used the default settings.

Annotation of peaks was performed with a custom made script. Gene coordinates were retrieved from the Ensembl database for the hg19 build and peaks were assigned to genes based on the distance of the peak from gene TSS.

Average genomic profile was calculated and generated using the bioconductor package Recoup (https://bioconductor.org/packages/release/bioc/html/recoup.html). The genomic coverage at the transcription start site (TSS) was calculated based on the number and distribution of reads at -2000bp and +2000 bp around the TSS. The TSSs were split in 200 bins of dynamic length each (so that all genes are split into the same number of bins). Then, the ChIP-seq signal was calculated as normalized (down-sampling as described above) coverage per base-pair for each bin and averaged per bin, resulting in 200 data points for each TSS, comprising the binding pattern for each corresponding gene.

3.6.3 Pathway Enrichment Analysis-Gene Set Enrichment Analysis-TCGA clinical data

We conducted pathway enrichment analysis with the online tool Genecodis, selecting GO biological pathway analysis and GOSlim analysis[131]. We conducted gene set enrichment analysis with GSEA, with the gene sets being retrieved from the Molecular Signature Database (Hallmark gene sets). A list of the DEGs ranked according to fold change (log2 scale) was used as input to the GSEAPreranked method and the analysis was done with 1000 permutations by gene set[132]. For IGSF9, we plotted survival analysis based on the TCGA data for Colon Adenocarcinoma using the UALCAN platform. UALCAN uses TCGA level 3 RNA-seq and clinical data from 33 cancer types. The expression level was normalized as transcripts per million reads (TMP) for comparison across groups and individuals[133].

4 Results

4.1 cis-Overexpression of *WiNTRLINC3* affects the expression of multiple genes

In order to understand the functional role of WiNTRLINC3, we overexpressed it *in cis*, from its own locus, in LS174 cells engineered to overexpress the CRISPRa system. Specifically, transduction of LDC10 cells with two sgRNAs (sgRNA1 and sgRNA2), led to robust increase of the transcript's expression levels in both cases. In these cells, we performed 3' mRNA Quant-seq to evaluate the transcriptional programs that are mainly affected by WiNTRLINC3 *in cis* overexpression. A multidimensional scaling plot revealed complete separation between LDC10 that overexpress WiNTRLINC3 and LDC10 that were transduced with a non-targeting guide (NTG), showing that WiNTRLINC3 globally alters the transcriptional profile of colorectal cancer cells (Figure 4B).

Differential expression analysis revealed 1649 genes to be deregulated (928 up-regulated and 721 down-regulated) between LDC10-sgRNA1 and LDC10-NTG conditions and 2351 genes (1256 up-regulated and 1095 down-regulated) between LDC10-sgRNA2 and LDC10-NTG conditions (Figure 4A). For further molecular interpretation of the results we included only the protein-coding genes in our analysis, which comprised the majority of the deregulated genes. We calculated that LDC10-sgRNA1 and LDC10-sgRNA2 conditions shared 730 common protein-coding genes, which is a significant overlap, with nearly all the overlapping genes being either up- or down-regulated in both conditions (Figure 4C,D).

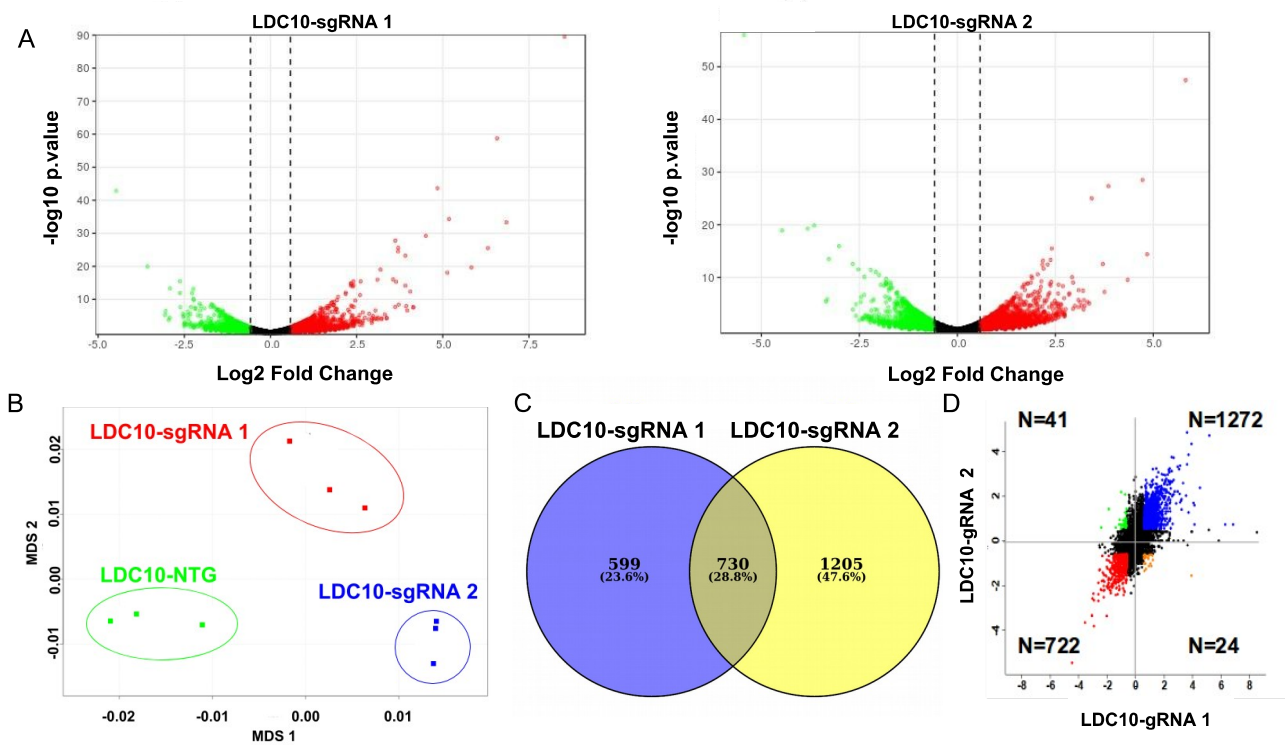

Figure 4: A. Volcano plot showing the differentially expressed genes upon WiNTRLINC3 cis-overexpression, upon transduction of LDC10 with lentis expressing sgRNA 1 and sgRNA 2 respectively, B. Multidimensional Scaling plot with the dots representing the replicates of each experimental condition, C. Venn Diagram showing the overlap of protein-coding differentially expressed genes between LDC10-sgRNA1 and LDC10-sgRNA2 conditions, D. Scatter plot of the fold changes of the differentially expressed genes in the 2 conditions of cis-overexpression.

Because WiNTRLINC3 was identified as a lincRNA that is negatively regulated by the Wnt pathway, we compared the DEGs upon WiNTRLINC3 overexpression with the Wnt target gene signature in the same cells. From previous experiments, we had conducted conditional knock-down of β-catenin in LS174 cells, followed by Quant-seq and Differential Expression Analysis with the DESEQ algorithm (unpublished data). Knock-down of β-catenin is one of the major ways to identify the genomic targets of the Wnt pathway, so we compared the lists of deregulated genes of the two experiments. As expected, we found significant overlaps between the LDC10-sgRNA1, LDC10-sgRNA2 and β-catenin knock-down datasets (Figure 5A,B) with 155 genes being common in all three conditions.

4.2 cis-Overexpression of WiNTRLINC3 inhibits Colorectal Carcinogenesis

We performed pathway enrichment analysis with the aim of discovering the biological pathways that are significantly affected upon WiNTRLINC3 overexpression. The major pathways that were overrepresented in the lists of up-regulated genes were signal transduction, cell adhesion, actin cytoskeleton, apoptosis and negative regulation of cell proliferation (Figure 5C). On the other hand, down-regulated genes were overrepresented in GO terms involving metabolic processes, mitosis, cell proliferation and transcription (Figure 5D). The same analysis was also performed for the common deregulated genes between WiNTRLINC3 overexpression and β-catenin knock-down. In this case, the most overrepresented processes involved again cell adhesion and cell proliferation. Thus, pathway analysis showed that WiNTRLINC3 plays an important role in regulating, the expression of proteins that are responsible for cell-cell interaction and proteins that inhibit cell proliferation.

To further discover more specific biological processes that are altered, upon cis-overexpression of WiNTRLINC3, we compared our lists of DEGs with multiple publicly available molecular signatures, using gene set enrichment analysis (GSEA). By using the GSEA Preranked method, a molecular signature is classified as statistically significant when it consists of genes that are present in the extremes of our gene sets, taking into account that larger differences in expression might be associated with phenotypic differences[132]. In both LDC10-sgRNA1 and LDC10-sgRNA2 conditions, the most overrepresented signature was epithelial to mesenchymal transition, involving genes such as DKK1, members of TIMP family (metallopeptidase inhibitors) and collagens (Figure 6B). Other signatures that were mainly overrepresented in the list of down-regulated genes are a G2M checkpoint signature, Myc targets, E2F targets and MTORC1 signaling, pointing to the regulatory role of WiNTRLINC3 in cell proliferation (Figure 6B).

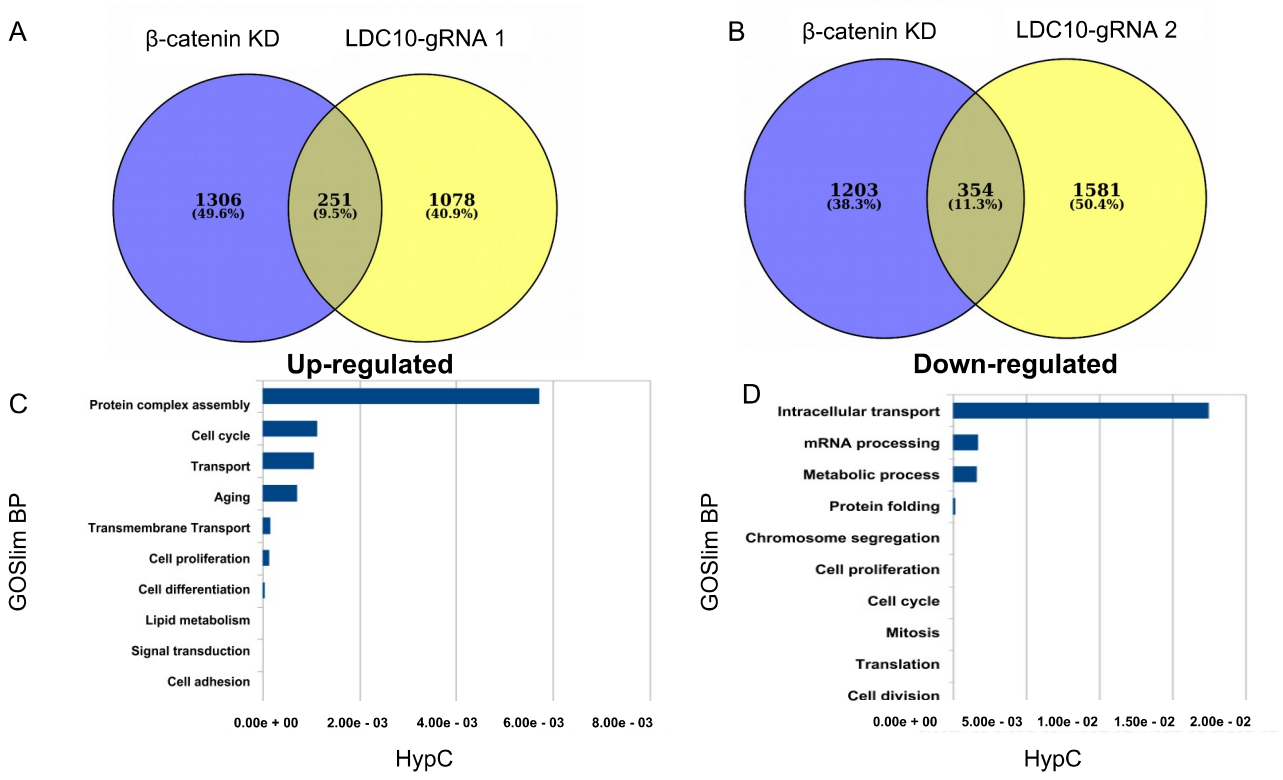

Figure 5: A,B. Venn diagrams showing the number of common differentially expressed genes between β-catenin knock down and cis-overexpression of WiNTRLINC3, C. Top 10 statistically significant pathways that are overrepresented in the up-regulated genes and D. the down-regulated genes upon WiNTRLINC3 overexpression.

With the aim of validating whether these effects can be recapitulated in vitro, we conducted colony formation assays, a common assay for evaluating cell proliferation and survival. As expected, we detected high reduction of the number and size of the colonies upon cis-overexpression of WiNTRLINC3, compared to the control (Figure 6A). These results show that the expression of WiNTRLINC3 is critical for cellular physiology, with high levels of expression leading to significant reduction in cell proliferation.

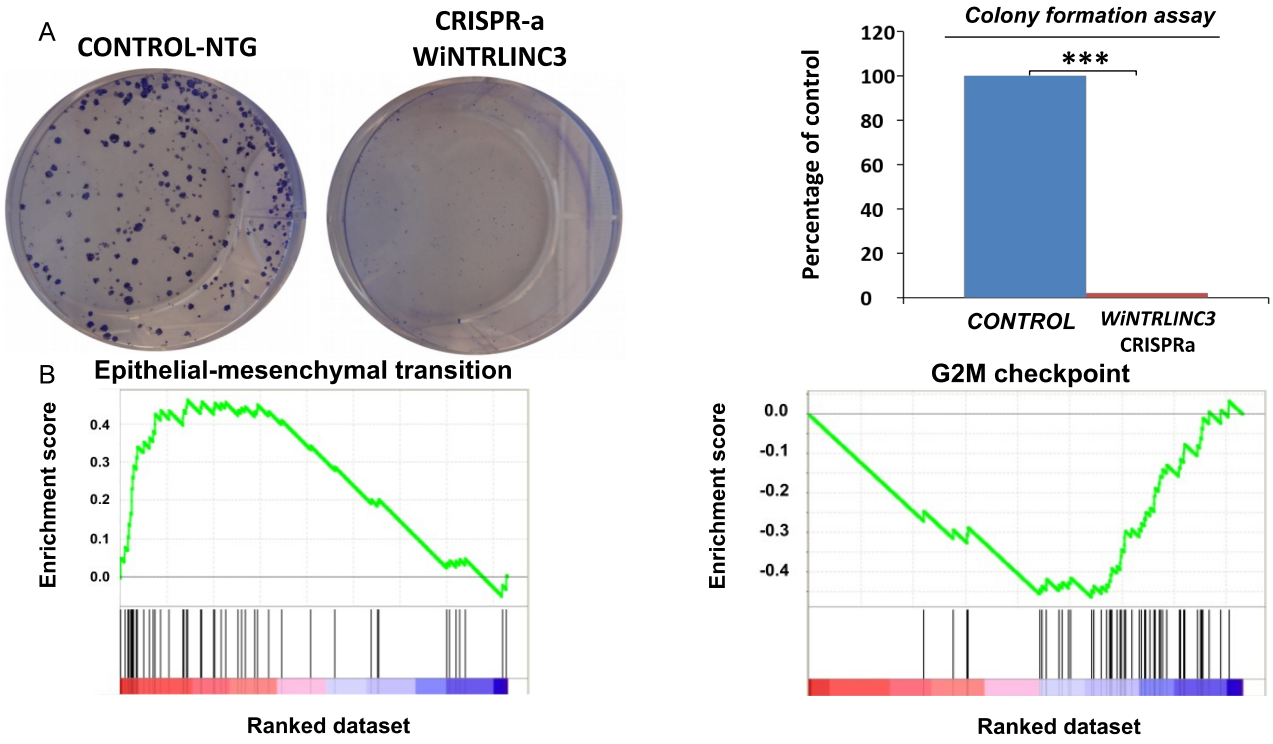

Figure 6: A. Colony formation assay for estimating cell survival in LDC10 control cells and in LDC10 cells that overexpress WiNTRLINC3 with the CRISPRa system, B. The top overrepresented molecular signatures for positively and negatively regulated genes of WiNTRLINC3.

4.3 WiNTRLINC3 positively regulates the expression of IGSF9

Because lincRNAs frequently function *in cis,* we searched for genes in the close proximity of *WiNTRLINC3* locus and evaluated whether their expression is affected by WiNTRLINC3 cis-overexpression. Interestingly, we discovered that IGSF9, a gene located 16 kb from the *WiNTRLINC3* locus, is the closest neighboring gene with significantly elevated expression (Figure 4A), upon WiNTRLINC3 overexpression. IGSF9 encodes an adhesion protein which is essential for maintaining synapses in mice[134], while the Drosophila homolog Turtle, affects dendrite differentiation in sensory neurons[135] and also promotes cell-cell interaction, ultimately regulating the R7 photoreceptor terminals[136]. Even though there have been no studies involving IGSF9 in the colon, expression levels are high in the adult duodenum and the small intestine in mice (Mouse ENCODE project, https://www.ncbi.nlm.nih.gov/gene/93842#gene-expression). Utilizing the publicly available

platform UALCAN, we found that IGSF9 is significantly down-regulated in colon adenocarcinomas in the TCGA cohort[133], following a similar pattern with WiNTRLINC3 (Figure 7B).

Furthermore, we validated by RT-qPCR that the expression levels of both WiNTRLINC3 and IGSF9 are significantly elevated after WiNTRLINC3 cis-overexpression (Figure 7A,C). To evaluate whether there is a common pattern of regulation that is controlled by WiNTRLINC3, we generated shRNAs that target either WiNTRLINC3 or IGSF9. Because of the low expression of WiNTRLINC3 in colorectal cancer cell lines, we generated a cellular system of knocking down WiNTRLINC3 and IGSF9 after cis-overexpression of WiNTRLINC3, in order to test the efficacy of the designed shRNAs. Knock-down of WiNTRLINC3, resulted in nearly 10 times lower expression and reduced in half the expression of IGSF9. Similarly, loss of IGSF9 (nearly 70%) led to a small decrease of WiNTRLINC3's expression (Figure 7D).

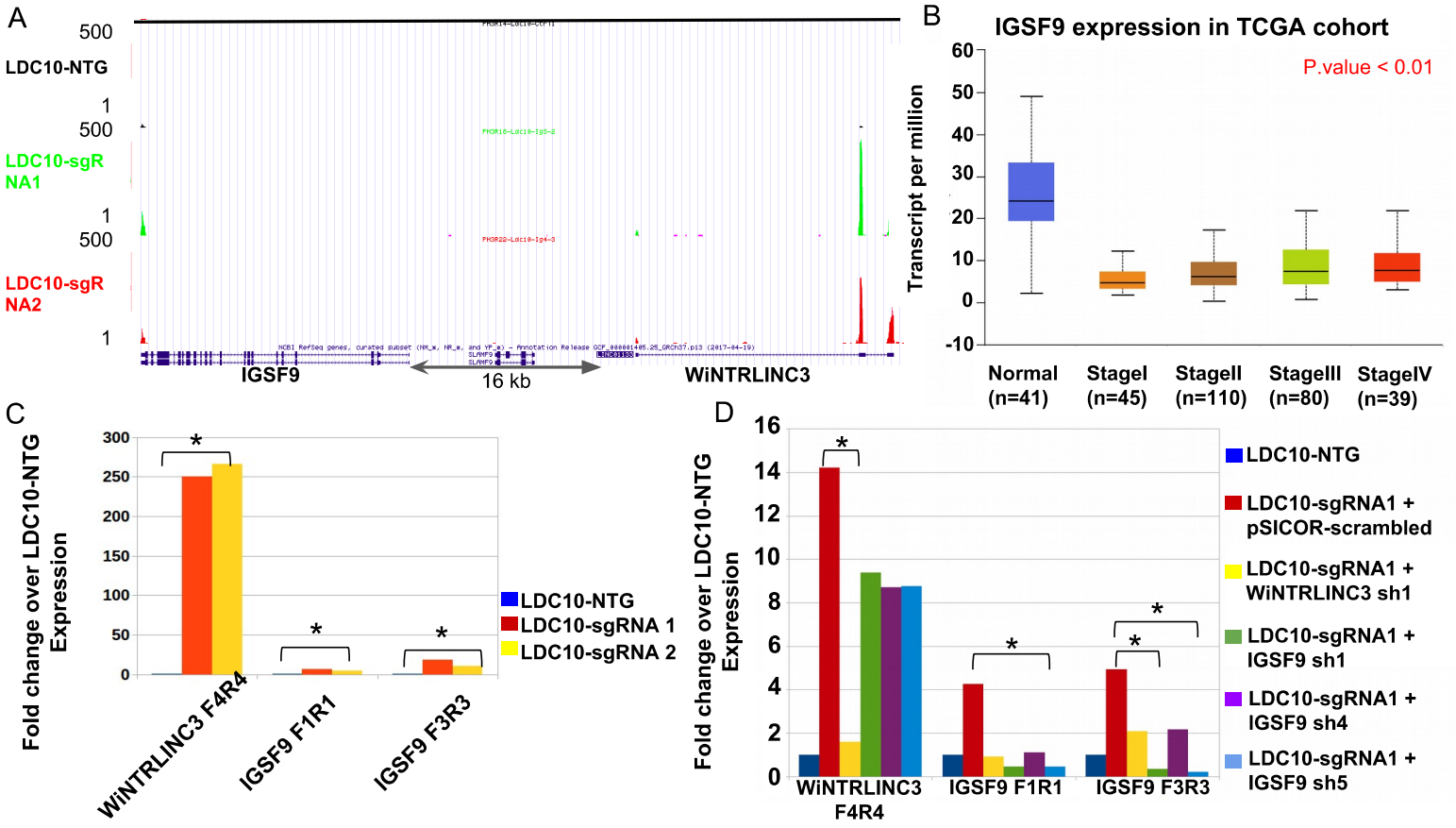

Figure 7: A. Snapshot of the expression levels of WiNTRLINC3 and IGSF9 in LDC10-NTG, LDC10-sgRNA1 and LDC10-sgRNA2, B. Expression of IGSF9 in colorectal adenocarcinoma based on individual cancer stages, from the TCGA cohort, C. qPCR of WiNTRLINC3 and IGSF9 in LDC10-sgRNA 1, D. qPCR of WiNTRLINC3 and IGSF9 after knocking down either WiNTRLINC3 or IGSF9, in cells that are cis-overexpressing WiNTRLINC3.

4.4 CDX2 binds to the locus of WiNTRLINC3

To better understand how WiNTRLINC3 is regulated, we investigated the proteins that interact with the transcript based on our previous RNA pull-down experiments followed by mass-spectrometry (described in the introduction). CDX2 was the transcription factor that caught our attention, as it has been shown to function as a negative regulator of the β-catenin/TCF4 pathway, limiting the proliferation of colon cancer cells and promoting cell differentiation[137].

We took advantage of previously published data of a CDX2 ChIP-seq experiment in the LS174 cell line and re-analyzed it in order to define CDX2 binding sites[127]. We identified 3175 binding sites of CDX2 that are statistically significant. Average genomic profile plot and ceas analysis revealed that CDX2 binds mainly to the promoter region of genes (< 1000bp from the TSS), while there are multiple binding sites in distal genomic regions away from annotated genes (Figure 8A, B). Among those sites, CDX2 also binds to the promoter and enhancer of *WiNTRLINC3*, indicating a putative regulatory role of the transcription factor in WiNTRLINC3 expression. Another binding site, but with lower affinity, resides at the promoter of *IGSF9*. Among the other binding sites, we identified a great number of abundant peaks close to Hox genes, correlating positively with previously known CDX2 targets in intestinal epithelial cells[121].

We investigated whether CDX2 could activate WiNTRLINC3 expression in vitro. For this reason, we cloned the promoter or the enhancer sequences of *WiNTRLINC3* in plasmid vectors, upstream of the luciferase reporter gene, followed by luciferase assays to evaluate whether CDX2 affects the activity of the promoter or/and enhancer of WiNTRLINC3. With the aim of validating that the cloned fragments were functional, we transduced LDC10 cells with the lentiviruses expressing WiNTLINC3 sgRNAs that target the sequence of the promoter (sgRNA 1 & 2) or the enhancer (sgRNA 3) region. As expected, there was high increase in the promoter and enhancer activity in the corresponding transduction experiments (Figure 8C). Additionally, we tested whether their activity was altered upon the presence and absence of CDX2. To achieve overexpression of CDX2, we transfected the cells with the pcDNA 3-CDX2 expressing plasmid. Our first results show that CDX2 alone does not significantly activate neither the promoter or the enhancer of *WiNTRLINC3* (Figure 8D).

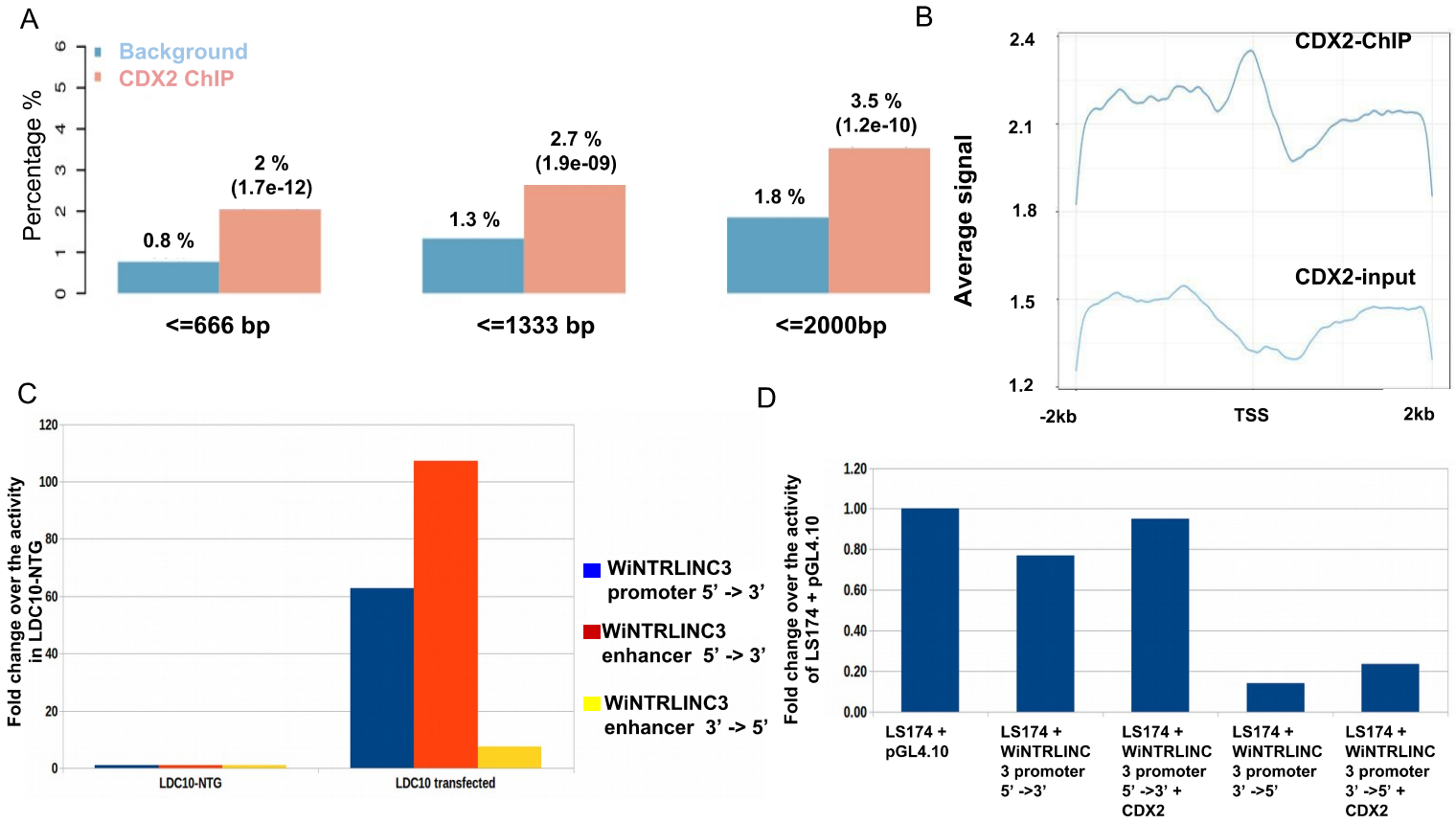

Figure 8: A. Genomic distribution of the CDX2 ChIP-seq peaks around the promoters, B. Average Genomic profile of CDX2 Chip-seq and Input 2 kb upstream and downstream of the TSS, C. Luciferase assays showing the activity of the WiNTRLINC3 promoter 5'->3' , enhancer 5'->3' and enhancer 3'->5' before and after transduction with sgRNA 1 (for promoter) and sgRNA 3 (for enhancer), D. Luciferase reporter assay showing the activity of WiNTRLINC3 promoter before and after transfection with pcDNA-CDX2.

5 Discussion

In our study, we gained further insights in how WiNTRLINC3, a lincRNA that is negatively regulated by the Wnt pathway, functions in intestinal carcinogenesis. For this purpose, we generated cell lines that overexpress WiNTRLINC3 *in cis* with the use of a CRISPRa system, a technique which enables the cis-overexpression of an endogenous gene with higher efficiency.

Cis-overexpression of WiNTRLINC3, followed by RNA-sequencing, allowed us to unravel the transcriptional effects of WiNTRLINC3 on target genes *in cis*. We showed that a great proportion of WiNTRLINC3 DEGs are also convergently deregulated by Wnt pathway abrogation. Pathway and gene set enrichment analysis of the deregulated genes revealed that WiNTRLINC3 regulates the expression of proteins that are critical for cell adhesion and cell proliferation, a fact that may underlie the prognostic value of WiNTRLINC3 in Colorectal Cancer. These effects were also shown with in vitro experiments with the cells overexpressing WiNTRLINC3 being unable to proliferate properly. We propose that aberrant activity of the Wnt pathway reduces WiNTRLINC3 expression, leading to down-regulation of a group of cell adhesion molecules, consisting of laminins, filamin binding proteins, cadherins, collagens and integrins, ultimately preventing the intestinal cells from differentiating while promoting the proliferation of tumor cells (Figure 9).

Supporting further this theory, we discovered an underlying regulatory axis between WiNTRLINC3 and IGSF9. We showed that the increased expression of WiNTRLINC3 transcript is essential for the expression of IGSF9, with IGSF9 following the same pattern of expression as the lincRNA. We propose that WiNTRLINC3 mediates its phenotypic effects potentially by solely regulating the expression of IGSF9 *in cis*. The IGSF9 protein functions as a cell adhesion molecule. While IGSF9 has been thoroughly studied in synapses development, there is no clear evidence for its role in other tissues, although the human orthologue is expressed in various epithelial compartments much more than in the nervous system. Interestingly, the expression of IGSF9 has been found significantly altered in various types of cancer[138]. Specifically, IGSF9 was found down-regulated in advanced-stage melanomas and in APC[139,140]. On the other hand, it was among the most up-regulated genes in gallbladder and endometrial cancer, revealing potential differential roles in carcinogenesis[141,142]. This can be explained by the fact

that cell adhesion molecules can either facilitate or reduce the migration and invasion of tumor cells which depends both on their expression levels and the proteins they interact with[138]. Surprisingly, an interaction network derived from the STRING database[143] reveals a predicted interaction (based on homologs in other species) between IGSF9 and LRP5/6, the receptors of Wnt signaling. Experimentally validating this interaction could possibly explain the mutual interplay between the Wnt pathway and the WiNTRLINC3-IGSF9 axis.

Even though we discovered a regulatory axis between WiNTRLINC3 and IGSF9, the underlying mechanism still remains elusive. IGSF9 was not among the interacting proteins of WiNTRLINC3, so we will investigate whether WiNTRLINC3 regulates the expression of IGSF9 by binding to chromatin or by facilitating chromatin looping. By utilizing chromatin isolation by RNA purification (ChIRP) followed by sequencing we will be able to identify sites of WiNTRLINC3 interaction with chromatin[144]. In our future plans, we will also establish a circular chromosome conformation capture (4C) protocol, which will enable us to investigate the chromatin architecture of LDC10 cells with cis-overexpression of WiNTRLINC3[145]. Putative interactions between the regulatory regions of WiNTRLINC3 and IGSF9, as well as with other loci, will reveal regions that are commonly regulated due to spatial proximity, such as in the case of WiNTRLINC1 and ASCL2[113].

Moreover, we reported a potential role of CDX2 in affecting WiNTRLINC3 transcription by binding to its promoter and enhancer. Lack of CDX2 has been previously characterized as a biomarker for stages II and III of colorectal cancer and has been associated with lower survival rates[146]. In addition, absence of CDX2 is mainly found in the MSI-immune tumors (CMS 1) subtype, based on the consensus molecular subtype classification of colon cancer, but is associated with poor prognosis only in the mesenchymal subgroup (CMS4)[147]. In a meta-analysis it was found, that there was lower risk of death for patients with higher CDX2 expression and the latter was highly reduced in cases of stage II and stage III CRC[148]. These clinical findings agree with our results about WiNTRLINC3 and IGSF9 expression in colorectal cancer according to the TCGA cohort, showing a common pattern of operation of all three factors in carcinogenesis.

Additionally, there is an interesting interplay between CDX2 and the β-catenin-TCF4 pathway. It has been shown that CDX2 can bind to β-catenin and disrupt its interaction with TCF4, shutting down the Wnt pathway and inhibiting cell proliferation in Caco2 and 293T cells[146]. Moreover, CDX2 was found to co-occupy several cis-regulatory regions of TCF4 with CDX2 being necessary for TCF4

binding to some of those sites in intestinal epithelial cells. As expected, the common binding sites mainly involved intestine-specific genes[121]. Another case where the Wnt pathway cooperates with CDX2 occurs in the development of the large and small intestine where they affect global gene expression[149].

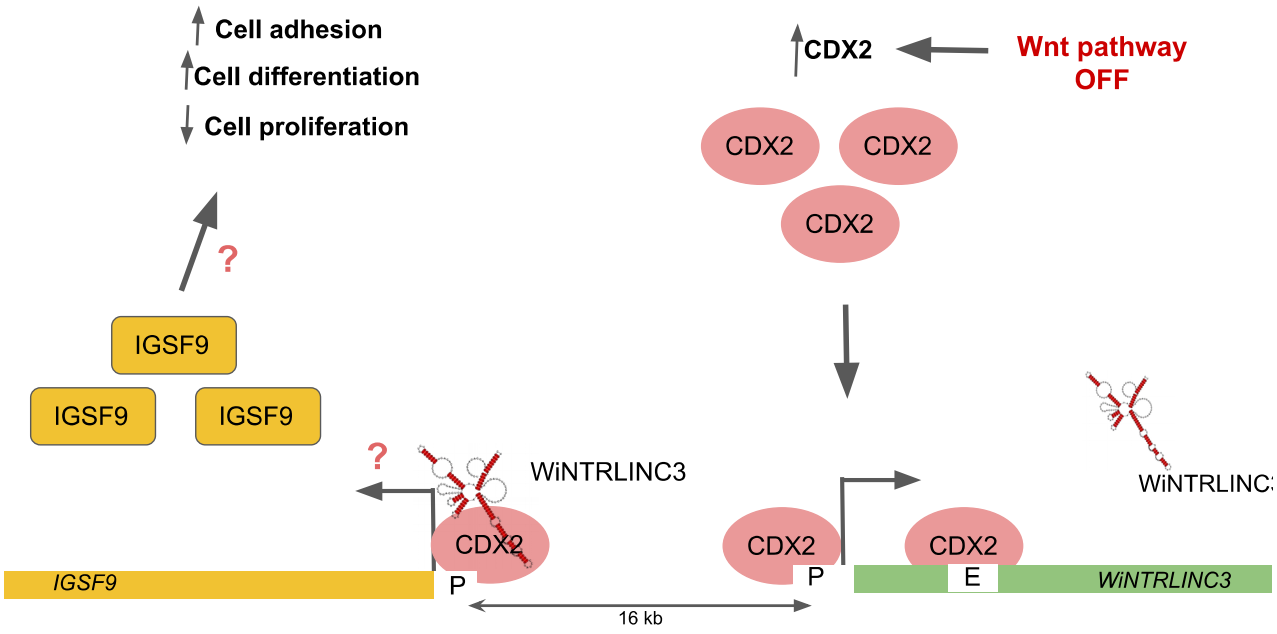

Figure 9: Graphical illustration of the proposed mechanism of WiNTRLINC3 function

CDX2 binding on the WiNTRLINC3 locus and transcript make it a prominent player in the regulation of this particular lincRNA. Cis-overexpression of WiNTRLINC3 did not affect the expression of CDX2, indicating that the lincRNA does not regulate the expression of this transcription factor. Even though CDX2 did not enhance in vitro the activity of either the promoter or the enhancer of WiNTRLINC3, we hypothesize that CDX2 needs to interact with other proteins in order to activate gene expression. Based on the motifs that were found within the central 100 bp of CDX2-occupied regions, GATA, FOXA1, HNF1 and HNF4A have been proposed as cofactors of CDX2 in intestinal cells[121]. With RNA immunoprecipitation (RIP) followed by Mass spectrometry we will be able to validate and identify the proteins that interact with WiNTRLINC3 and form a complex with CDX2.

To conclude, we have identified a novel lincRNA that is negatively regulated by the Wnt pathway and the reduced expression of which is associated with colorectal cancer. We propose that silencing of the Wnt pathway enhances the binding of CDX2 at the WiNTRLINC3 regulatory regions, with WiNTRLINC3 being a key mediator of CDX2 transcriptional program, ultimately affecting a plethora of cell adhesion molecules which promote cell differentiation (Figure 9). We believe that further investigation of the WiNTRLINC3 mechanisms of function might help us define a novel prognostic marker and therapeutic for colorectal cancer.

6 References

1. Clevers, H. & Nusse, R. Wnt/β-catenin signaling and disease. *Cell* **149,** 1192–1205 (2012).
2. Nusse, R. & Clevers, H. Wnt/β-Catenin Signaling, Disease, and Emerging Therapeutic Modalities. *Cell* **169,** 985–999 (2017).
3. Loh, K. M., van Amerongen, R. & Nusse, R. Generating Cellular Diversity and Spatial Form: Wnt Signaling and the Evolution of Multicellular Animals. *Dev. Cell* **38,** 643–655 (2016).
4. Bryan T. MacDonald; Keiko Tamai and Xi He. NIH Public Access. *Dev. Biol.* **17,** 9–26 (2010).
5. Mikels, A. J. & Nusse, R. Wnts as ligands: processing, secretion and reception. *Oncogene* **25,** 7461 (2006).
6. Willert, K. & Nusse, R. Wnt Proteins. *Cold Spring Harb. Perspect. Biol.* **4,** a007864 (2012).
7. Coudreuse, D. & Korswagen, H. C. The making of Wnt: new insights into Wnt maturation, sorting and secretion. *Development* **134,** 3 LP-12 (2007).

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

9. Takada, R. *et al.* Monounsaturated fatty acid modification of Wnt protein: its role in Wnt secretion. *Dev. Cell* **11,** 791–801 (2006).

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

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

12. Stanganello, E. *et al.* Filopodia-based Wnt transport during vertebrate tissue patterning. *Nat. Commun.* **6,** 5846 (2015).

13. Kusserow, A. *et al.* Unexpected complexity of the Wnt gene family in a sea anemone. *Nature* **433,** 156 (2005).

14. Capurro, M. I., Xiang, Y.-Y., Lobe, C. & Filmus, J. Glypican-3 promotes the growth of hepatocellular carcinoma by stimulating canonical Wnt signaling. *Cancer Res.* **65,** 6245–6254 (2005).

15. Bhanot, P. *et al.* A new member of the frizzled family from Drosophila functions as a Wingless receptor. *Nature* **382,** 225–230 (1996).

16. Janda, C. Y., Waghray, D., Levin, A. M., Thomas, C. & Garcia, K. C. Structural basis of Wnt recognition by Frizzled. *Science* **337,** 59–64 (2012).

17. Janda, C. Y. *et al.* Surrogate Wnt agonists that phenocopy canonical Wnt and β-catenin signalling. *Nature* **545,** 234 (2017).

18. Tamai, K. *et al.* A mechanism for Wnt coreceptor activation. *Mol. Cell* **13,** 149–156 (2004).

19. Mao, J. *et al.* Low-density lipoprotein receptor-related protein-5 binds to Axin and regulates the canonical Wnt signaling pathway. *Mol. Cell* **7,** 801–809 (2001).

20. Chen, W. *et al.* Dishevelled 2 recruits beta-arrestin 2 to mediate Wnt5A-stimulated endocytosis of Frizzled 4. *Science* **301,** 1391–1394 (2003).

21. Fiedler, M., Mendoza-Topaz, C., Rutherford, T. J., Mieszczanek, J. & Bienz, M. Dishevelled interacts with the DIX domain polymerization interface of Axin to interfere with its function in down-regulating beta-catenin. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 1937–1942 (2011).

22. Schwarz-Romond, T. *et al.* The DIX domain of Dishevelled confers Wnt signaling by dynamic polymerization. *Nat. Struct. Mol. Biol.* **14,** 484–492 (2007).

23. Ke, J. *et al.* Structure and function of Norrin in assembly and activation of a Frizzled 4-Lrp5/6 complex. *Genes Dev.* **27,** 2305–2319 (2013).

24. Xu, Q. *et al.* Vascular development in the retina and inner ear: control by Norrin and Frizzled-4, a high-affinity ligand-receptor pair. *Cell* **116,** 883–895 (2004).

25. Glinka, A. *et al.* LGR4 and LGR5 are R-spondin receptors mediating Wnt/β-catenin and Wnt/PCP signalling. *EMBO Rep.* **12,** 1055–1061 (2011).

26. de Lau, W. *et al.* Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. *Nature* **476,** 293 (2011).

27. Carmon, K. S., Gong, X., Lin, Q., Thomas, A. & Liu, Q. R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/beta-catenin signaling. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 11452–11457 (2011).

28. Hao, H.-X. *et al.* ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. *Nature* **485,** 195–200 (2012).

29. 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. *J. Cell Sci.* **121,** 737–746 (2008).

30. Semënov, M. V, Zhang, X. & He, X. DKK1 Antagonizes Wnt Signaling without Promotion of LRP6 Internalization and Degradation. *J. Biol. Chem.* **283,** 21427–21432 (2008).

31. Liu, C. *et al.* Control of beta-catenin phosphorylation/degradation by a dual-kinase mechanism. *Cell* **108,** 837–847 (2002).

32. Kitagawa, M. *et al.* An F-box protein, FWD1, mediates ubiquitin-dependent proteolysis of beta-catenin. *EMBO J.* **18,** 2401–2410 (1999).

33. Aberle, H., Bauer, A., Stappert, J., Kispert, A. & Kemler, R. beta-catenin is a target for the ubiquitin-proteasome pathway. *EMBO J.* **16,** 3797–3804 (1997).

34. Stamos, J. L., Chu, M. L.-H., Enos, M. D., Shah, N. & Weis, W. I. Structural basis of GSK-3 inhibition by N-terminal phosphorylation and by the Wnt receptor LRP6. *Elife* **3,** e01998 (2014).

35. Behrens, J. *et al.* Functional interaction of beta-catenin with the transcription factor LEF-1. *Nature* **382,** 638–642 (1996).

36. Cadigan, K. M. & Waterman, M. L. TCF/LEFs and Wnt Signaling in the Nucleus. *Cold Spring Harb. Perspect. Biol.* **4,** a007906 (2012).

37. Love, J. J. *et al.* Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376,** 791–795 (1995).

38. Giese, K., Amsterdam, A. & Grosschedl, R. DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1. *Genes Dev.* **5,** 2567–2578 (1991).

39. Roose, J. *et al.* The Xenopus Wnt effector XTcf-3 interacts with Groucho-related transcriptional repressors. *Nature* **395,** 608–612 (1998).

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

41. MacDonald, B. T., Tamai, K. & He, X. Wnt/β-catenin signaling: components, mechanisms, and diseases. *Dev. Cell* **17,** 9–26 (2009).

42. Daniels, D. L. & Weis, W. I. Beta-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nat. Struct. Mol. Biol.* **12,** 364–371 (2005).

43. Hatzis, P. *et al.* Genome-wide pattern of TCF7L2/TCF4 chromatin occupancy in colorectal cancer cells. *Mol. Cell. Biol.* **28,** 2732–2744 (2008).

44. Nusse, R. Wnt signaling and stem cell control. *Cell Res.* **18,** 523 (2008).

45. Andl, T., Reddy, S. T., Gaddapara, T. & Millar, S. E. WNT Signals Are Required for the Initiation of Hair Follicle Development. *Dev. Cell* **2,** 643–653 (2002).

46. Reya, T. *et al.* A role for Wnt signalling in self-renewal of haematopoietic stem cells. *Nature* **423,** 409–414 (2003).

47. Gat, U., DasGupta, R., Degenstein, L. & Fuchs, E. De Novo hair follicle morphogenesis and hair tumors in mice expressing a truncated beta-catenin in skin. *Cell* **95,** 605–614 (1998).

48. Zeng, Y. A. & Nusse, R. Wnt Proteins Are Self-Renewal Factors for Mammary Stem Cells and Promote Their Long-Term Expansion in Culture. *Cell Stem Cell* **6,** 568–577 (2010).

49. Huch, M. *et al.* Unlimited in vitro expansion of adult bi-potent pancreas progenitors through the Lgr5/R-spondin axis. *EMBO J.* **32,** 2708–2721 (2013).

50. Huch, M. *et al.* In vitro expansion of single Lgr5+ liver stem cells induced by Wnt-driven regeneration. *Nature* **494,** 247–250 (2013).

51. Wang, B., Zhao, L., Fish, M., Logan, C. Y. & Nusse, R. Self-renewing diploid Axin2(+) cells fuel homeostatic renewal of the liver. *Nature* **524,** 180–185 (2015).

52. Vries, R. G. J., Huch, M. & Clevers, H. Stem cells and cancer of the stomach and intestine. *Mol. Oncol.* **4,** 373–384 (2010).

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

54. Kinzler, K. W. *et al.* Identification of a gene located at chromosome 5q21 that is mutated in colorectal cancers. *Science* **251,** 1366–1370 (1991).

55. Korinek, V. *et al.* Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. *Science* **275,** 1784–1787 (1997).

56. Lammi, L. *et al.* Mutations in AXIN2 cause familial tooth agenesis and predispose to colorectal cancer. *Am. J. Hum. Genet.* **74,** 1043–1050 (2004).

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

58. Takeda, H. *et al.* Human sebaceous tumors harbor inactivating mutations in LEF1. *Nat. Med.* **12,** 395–397 (2006).

59. Wu, J. *et al.* Whole-exome sequencing of neoplastic cysts of the pancreas reveals recurrent mutations in components of ubiquitin-dependent pathways. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 21188–21193 (2011).

60. Satoh, S. *et al.* AXIN1 mutations in hepatocellular carcinomas, and growth suppression in cancer cells by virus-mediated transfer of AXIN1. *Nat. Genet.* **24,** 245–250 (2000).

61. Grant, S. F. A. *et al.* Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat. Genet.* **38,** 320–323 (2006).

62. Christodoulides, C. *et al.* WNT10B mutations in human obesity. *Diabetologia* **49,** 678–684 (2006).

63. Kanazawa, A. *et al.* Association of the gene encoding wingless-type mammary tumor virus integration-site family member 5B (WNT5B) with type 2 diabetes. *Am. J. Hum. Genet.* **75,** 832–843 (2004).

64. Huber, A. H., Nelson, W. J. & Weis, W. I. Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell* **90,** 871–882 (1997).

65. Lu, J. *et al.* Structure-activity relationship studies of small-molecule inhibitors of Wnt response. *Bioorg. Med. Chem. Lett.* **19,** 3825–3827 (2009).

66. Huang, S.-M. A. *et al.* Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. *Nature* **461,** 614 (2009).

67. Steinhart, Z. *et al.* Genome-wide CRISPR screens reveal a Wnt-FZD5 signaling circuit as a druggable vulnerability of RNF43-mutant pancreatic tumors. *Nat. Med.* **23,** 60–68 (2017).

68. Wilusz, J. E., Sunwoo, H. & Spector, D. L. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* **23,** 1494–1504 (2009).

69. Ulitsky, I. & Bartel, D. P. lincRNAs: Genomics, Evolution, and Mechanisms. *Cell* **154,** 26–46 (2013).

70. Kopp, F. & Mendell, J. T. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* **172,** 393–407 (2018).

71. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447,** 799–816 (2007).

72. Rinn, J. L. & Chang, H. Y. Genome Regulation by Long Noncoding RNAs. *Annu. Rev. Biochem.* **81,** 145–166 (2012).

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

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

75. Marson, A. *et al.* Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134,** 521–533 (2008).

76. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448,** 553–560 (2007).

77. Guttman, M. *et al.* Ab initio reconstruction of cell type–specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* **28,** 503 (2010).

78. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).

79. Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18,** 1509–1517 (2008).

80. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5,** 621–628 (2008).

81. Lin, M. F., Deoras, A. N., Rasmussen, M. D. & Kellis, M. Performance and Scalability of Discriminative Metrics for Comparative Gene Identification in 12 Drosophila Genomes. *PLoS Comput. Biol.* **4,** e1000067 (2008).

82. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147,** 789–802 (2011).

83. Tupy, J. L. *et al.* Identification of putative noncoding polyadenylated transcripts in Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 5495–5500 (2005).

84. Cerase, A., Pintacuda, G., Tattermusch, A. & Avner, P. Xist localization and function: new insights from multiple levels. *Genome Biol.* **16,** 166 (2015).

85. Latos, P. A. *et al.* Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* **338,** 1469–1472 (2012).

86. Engreitz, J. M. *et al.* Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539,** 452 (2016).

87. Anderson, K. M. *et al.* Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. *Nature* **539,** 433–436 (2016).

88. Dimitrova, N. *et al.* LincRNA-p21 activates p21 in cis to promote Polycomb target gene expression and to enforce the G1/S checkpoint. *Mol. Cell* **54,** 777–790 (2014).

89. Groff, A. F. *et al.* In vivo characterization of Linc-p21 reveals functional cis-regulatory DNA elements. *Cell Rep.* **16,** 2178–2186 (2016).

90. 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).

91. Mallo, M. & Alonso, C. R. The regulation of Hox gene expression during animal development. *Development* **140,** 3951 LP-3963 (2013).

92. Engreitz, J. M. *et al.* RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell* **159,** 188–199 (2014).

93. West, J. A. *et al.* The Long Noncoding RNAs NEAT1 and MALAT1 Bind Active Chromatin Sites. *Mol. Cell* **55,** 791–802 (2014).

94. Clemson, C. M. *et al.* An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol. Cell* **33,** 717–726 (2009).

95. Tichon, A. *et al.* A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nat. Commun.* **7,** 12209 (2016).

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

97. Tay, Y., Rinn, J. & Pandolfi, P. P. The multilayered complexity of ceRNA crosstalk and competition. *Nature* **505,** 344 (2014).

98. Yang, G. *et al.* Crosstalk between long non-coding RNAs and Wnt/beta-catenin signalling in cancer. *J. Cell. Mol. Med.* **22,** 2062–2070 (2018).

99. Ong, M. S. *et al.* 'Lnc'-ing Wnt in female reproductive cancers: therapeutic potential of long non-coding RNAs in Wnt signalling. *Br. J. Pharmacol.* **174,** 4684–4700 (2017).

100. Zarkou, V., Galaras, A., Giakountis, A. & Hatzis, P. Crosstalk mechanisms between the WNT signaling pathway and long non-coding RNAs. *Non-coding RNA Res.* **3,** 42–53 (2018).

101. Shen, P., Pichler, M., Chen, M., Calin, G. A. & Ling, H. To Wnt or Lose: The Missing Non-Coding Linc in Colorectal Cancer. *Int. J. Mol. Sci.* **18,** (2017).

102. Mokry, M. *et al.* Efficient double fragmentation ChIP-seq provides nucleotide resolution protein-DNA binding profiles. *PLoS One* **5,** e15092 (2010).

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

104. Tuupanen, S. *et al.* The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat. Genet.* **41,** 885–890 (2009).

105. Di Cecilia, S. *et al.* RBM5-AS1 Is Critical for Self-Renewal of Colon Cancer Stem-like Cells. *Cancer Res.* **76,** 5615–5627 (2016).

106. Ma, Y. *et al.* Long non-coding RNA CCAL regulates colorectal cancer progression by activating Wnt/beta-catenin signalling pathway via suppression of activator protein 2alpha. *Gut* **65,** 1494–1504 (2016).

107. Zhang, Z. *et al.* Long non-coding RNA CASC11 interacts with hnRNP-K and activates the WNT/beta-catenin pathway to promote growth and metastasis in colorectal cancer. *Cancer Lett.* **376,** 62–73 (2016).

108. Kawasaki, Y. *et al.* MYU, a Target lncRNA for Wnt/c-Myc Signaling, Mediates Induction of CDK6 to Promote Cell Cycle Progression. *Cell Rep.* **16,** 2554–2564 (2016).

109. Wang, L. *et al.* A long non-coding RNA targets microRNA miR-34a to regulate colon cancer stem cell asymmetric division. *Elife* **5,** (2016).

110. Chen, W.-Y. *et al.* MicroRNA-34a regulates WNT/TCF7 signaling and inhibits bone metastasis in Ras-activated prostate cancer. *Oncotarget* **6,** 441–457 (2015).

111. Hu, P. *et al.* NBAT1 suppresses breast cancer metastasis by regulating DKK1 via PRC2. *Oncotarget* **6,** 32410–32425 (2015).

112. Niida, A. *et al.* DKK1, a negative regulator of Wnt signaling, is a target of the beta-catenin/TCF pathway. *Oncogene* **23,** 8520–8526 (2004).

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

114. Fang, C. *et al.* Long non-coding RNA HNF1A-AS1 mediated repression of miR-34a/SIRT1/p53 feedback loop promotes the metastatic progression of colon cancer by functioning as a competing endogenous RNA. *Cancer Lett.* **410,** 50–62 (2017).

115. Lou, G. *et al.* The p53/miR-34a/SIRT1 Positive Feedback Loop in Quercetin-Induced Apoptosis. *Cell. Physiol. Biochem.* **35,** 2192–2202 (2015).

116. Kim, N. H. *et al.* p53 and microRNA-34 are suppressors of canonical Wnt signaling. *Sci. Signal.* **4,** ra71 (2011).

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

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

119. 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 Res.* **40,** 148–158 (2012).

120. Lin, M. F., Jungreis, I. & Kellis, M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27,** i275-82 (2011).

121. Verzi, M. P. *et al.* TCF4 and CDX2, major transcription factors for intestinal function, converge on the same cis-regulatory regions. *Proc. Natl. Acad. Sci.* **107,** 15157 LP-15162 (2010).

122. Saad, R. S., Ghorab, Z., Khalifa, M. A. & Xu, M. CDX2 as a marker for intestinal differentiation: Its utility and limitations. *World J. Gastrointest. Surg.* **3,** 159–166 (2011).

123. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517,** 583 (2014).

124. Moll, P., Ante, M., Seitz, A. & Reda, T. QuantSeq 3′ mRNA sequencing for RNA quantification. *Nat. Methods* **11,** 972 (2014).

125. Moulos, P. & Hatzis, P. Systematic integration of RNA-Seq statistical algorithms for accurate detection of differential gene expression patterns. *Nucleic Acids Res.* **43,** e25–e25 (2015).

126. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).

127. Pinto, R. *et al.* Precise integration of inducible transcriptional elements (PrIITE) enables absolute control of gene expression. *Nucleic Acids Res.* **45,** e123 (2017).

128. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).

129. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).

130. Shin, H., Liu, T., Manrai, A. K. & Liu, X. S. CEAS: cis-regulatory element annotation system. *Bioinformatics* **25,** 2605–2606 (2009).

131. Carmona-Saez, P., Chagoyen, M., Tirado, F., Carazo, J. M. & Pascual-Montano, A. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. *Genome Biol.* **8,** R3 (2007).

132. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102,** 15545 LP-15550 (2005).

133. Chandrashekar, D. S. *et al.* UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* **19,** 649–658 (2017).

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

135. Long, H., Ou, Y., Rao, Y. & van Meyel, D. J. Dendrite branching and self-avoidance are controlled by Turtle, a conserved IgSF protein in Drosophila. *Development* **136,** 3475–3484 (2009).

136. Ferguson, K., Long, H., Cameron, S., Chang, W.-T. & Rao, Y. The conserved Ig superfamily member Turtle mediates axonal tiling in Drosophila. *J. Neurosci.* **29,** 14151–14159 (2009).

137. 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).

138. Hansen, M. & Walmod, P. S. IGSF9 Family Proteins. *Neurochem. Res.* **38,** 1236–1251 (2013).

139. Smith, A. P., Hoek, K. & Becker, D. Whole-genome expression profiling of the melanoma progression pathway reveals marked molecular differences between nevi/melanoma in situ and advanced-stage melanomas. *Cancer Biol. Ther.* **4,** 1018–1029 (2005).

140. Bowden, N. A., Croft, A. & Scott, R. J. Gene Expression Profiling in Familial Adenomatous Polyposis Adenomas and Desmoid Disease. *Hered. Cancer Clin. Pract.* **5,** 79–96 (2007).

141. Shi, Z. *et al.* RNA-seq Reveals the Overexpression of IGSF9 in Endometrial Cancer. *J. Oncol.* **2018,** 1–13 (2018).

142. Kim, J. H. *et al.* Gene expression profiles in gallbladder cancer: the close genetic similarity seen for early and advanced gallbladder cancers may explain the poor prognosis. *Tumour Biol.* **29,** 41–49 (2008).

143. Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45,** D362–D368 (2017).

144. Chu, C., Quinn, J. & Chang, H. Y. Chromatin isolation by RNA purification (ChIRP). *J. Vis. Exp.* (2012). doi:10.3791/3912

145. van de Werken, H. J. G. *et al.* 4C technology: protocols and data analysis. *Methods Enzymol.* **513,** 89–112 (2012).

146. Dalerba, P. *et al.* CDX2 as a Prognostic Biomarker in Stage II and Stage III Colon Cancer. *N. Engl. J. Med.* **374,** 211–222 (2016).

147. Pilati, C. *et al.* CDX2 prognostic value in stage II/III resected colon cancer is related to CMS classification. *Ann. Oncol.* **28,** 1032–1035 (2017).

148. Tomasello, G. *et al.* Association of CDX2 Expression With Survival in Early Colorectal Cancer: A Systematic Review and Meta-analysis. *Clin. Colorectal Cancer* **17,** 97–103 (2018).

149. Sherwood, R. I., Maehr, R., Mazzoni, E. O. & Melton, D. A. Wnt Signaling Specifies and Patterns Intestinal Endoderm. *Mech. Dev.* **128,** 387–400 (2011).