American University in Cairo
AUC Knowledge Fountain

Theses and Dissertations | Student Research

2-1-2018

The interplay of genetic variation and regulation of long noncoding RNAs in colorectal cancer

Amany Abdel-Motaleb

Follow this and additional works at: https://fount.aucegypt.edu/etds

Recommended Citation

APA Citation
Abdel-Motaleb, A. (2018). *The interplay of genetic variation and regulation of long noncoding RNAs in colorectal cancer* [Master's Thesis, the American University in Cairo]. AUC Knowledge Fountain. https://fount.aucegypt.edu/etds/634

MLA Citation
Abdel-Motaleb, Amany. *The interplay of genetic variation and regulation of long noncoding RNAs in colorectal cancer*. 2018. American University in Cairo, Master's Thesis. *AUC Knowledge Fountain*. https://fount.aucegypt.edu/etds/634

This Master's Thesis is brought to you for free and open access by the Student Research at AUC Knowledge Fountain. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of AUC Knowledge Fountain. For more information, please contact thesisadmin@aucegypt.edu.

THE AMERICAN UNIVERSITY IN CAIRO
School of Sciences and Engineering

The Interplay of Genetic Variation and Regulation of Long Noncoding RNAs in Colorectal Cancer

A Thesis Submitted to the Biotechnology Program
In Partial Fulfillment of the Requirements for
The Degree of Masters of Science in Biotechnology

By: Amany Abdel-Motaleb
Bachelor of Science in Biology, Northern Illinois University

Under the Supervision of Dr. Hassan Azzazy
Professor and Chairman
Department of Chemistry
The American University in Cairo
and
Dr. Ahmed Moustafa
Professor and Director of the Biotechnology Program
Department of Biology
The American University in Cairo

December 2017

Dedication

I dedicate this research to my mother, father, and husband, Ali. These three people have played an integral role in accomplishing this feat. My mother and father have instilled in me the values and discipline to work hard and to accomplish my goals. Ali has been indispensable in the love and support he has given me over the course of this research. Surely, this research would have been impossible to complete without these people in my life. For that, I dedicate this work to them.

Acknowledgements

I would like to acknowledge the Department of Biology for providing the resources and facilities in order to complete this research. In particular, I would like to acknowledge Mr. Ahmed Elhosseiny for his long hours and support. His efforts were vital in this project.

Abstract

Background: Colorectal cancer (CRC) is the third leading cause of death worldwide comprising ~8% of cancer deaths per year. The survival rates of metastatic CRC is 13% because of the lack of successful treatment due to a lack of understanding of the scope and environment of stage IV CRC.

Materials and Methods: The transcripts of five normal colon mucosa tissue samples and their matched five stage IV CRC samples were chosen and analyzed from the dataset with the GEO Accession GSE50760. The Tuxedo Suite pipeline was used to determine the differentially expressed genes (DEGs) with a fold change cut off of ≥ 0.5 and ≤ -0.5 and a p value cut off of ≤ 0.05. Using the the DEG list, PANTHER database was used for pathway enrichment. LncRNA2Target database was used to find associated long non coding RNAs (lncRNAs) of the genes of interest. The Integrated Genome Viewer (IGV) was used to visualize any mutations or variations among the genes of interest. DeepSEA was used to functionally predict any potential novel SNPs. Using literature along with the results from PANTHER, lncRNA2Target, and IGV a novel connection was deduced.

Results: There were 5,303 DEGs. The Wnt pathway had the greatest portion of DEGs indicating pathway activity. Interestingly, a number of inhibitors of the Wnt pathway were also upregulated including WIF1 and SFRP4. LncRNA2Target analysis showed that HOTAIR, a lncRNA, has a number of target genes and effectively silenced all of its targets except for WIF1 and CD82 in this dataset. WIF1, CD82, and SFRP4 has increased fold change values of 5.165, 1.05, 2.121, respectively. Additionally, lncRNAs, UCA1 and CRNDE, were found to positively regulate WNT5A, WNT2 and WNT3 and were upregulated. Using the integrative genome viewer, 10 SNPs were found in WIF1, SFRP4, CD82, WNT5A, and UCA1 of which one was novel. The potentially novel SNP in CD82 was functionally predicted to create a binding site with ZBTB7A. Additionally, CDKN2A and CDKN2B were found to have decreased expression with a fold change value of -2.266.

Discussion: One synonymous SNP was in WIF1 and CD82. The missense SNPs in SFRP4 and CD82 are likely causing protein dysfunction resulting in ill-inhibition of the WNT pathway and metastasis, respectively. The novel SNP was found in CD82 at the location chr11:44,619,242 in the 3' untranslated region. Functional prediction showed that this SNP may create a binding site with ZBTB7A which may be repressing CD82 function. Although most of the SNPs found were recorded to result in synonymous codons, the prevalence and frequency of these SNPs in these vital genes requires further investigation to confirm whether if these SNPs are coincidental or if they are damaging. Moreover, it is also probable that WIF1 and SFRP4 may be competing with UCA1 to exert their effects on WNT5A. CRNDE may also be competing to ultimately positively regulate WNT2 and WNT3. Together with the SNPs, HOTAIR may not be able to silence WIF1 and CD82, WIF1 and SFRP4 are ineffective in inhibiting the WNT ligands, the missense and the potentially novel SNPs in CD82 may be the cause for the lack of metastasis suppression. Finally, the downreglulation of CDKN2A and CDKN2B may be due to environmental and/or ethnic causes as shown in previous studies with Egyptian and Chinese CRC patients.

Table of Contents

Table of Figures

List of Tables

List of Abbreviations

5-FU	5-Fluorouracil
CDS	Coding DNA Sequence
CRC	Colorectal Cancer
dbSNP	SNP Database
dscDNA	Double Stranded Complementary DNA
EGFR	Epidermal Growth Factor Receptor
EMT	Epithelial Mesenchymal Transition
FAP	Familial Adenomatous Polyposis
GEO	Gene Expression Omnibus
GIST	Gastrointestinal Stromal Tumors
IBD	Irritable Bowel Disease
IGRT	Image Guided Radiation Therapy
IGV	Integrative Genomics Viewer
IMRT	Intensity Modulated Radiation Therapy
miRNAs	microRNA
ncRNAs	NonCoding RNA
NSAIDs	Nonsteroidal Anti-Inflammatory Drugs
OSCC	Oral Squamous Cell Carcinoma
PCP	Planar Cell Polarity
SBRT	Stereotactic Body Radiation Therapy
siRNA	Small Interfering RNA
snoRNA	Small Nuclear RNA
SNP	Single Nucleotide Polymorphism
SRS	Stereotactic Radiosurgery
sscDNA	Single Stranded Complementary DNA
TGFB	Transforming Growth Factor Beta
TSS	Transcription Start Site
UTR	Untranslated Region

CHAPTER 1: INTRODUCTION

1.1 Colorectal Cancer

1.1.1 Epidemiology and Statistics

Colorectal cancer is the third leading cause of death worldwide and the fourth leading cause of death in the United States with 50,000 expected deaths in 2017 [1,2]. There is a lifetime risk of developing CRC of 4.7% for men and 4.4% for women. About 95% of CRC start as an adenocarcinomas. The remaining 5% of cases begin as carcinoids, sarcomas, lymphomas, or gastrointestinal stromal tumors (GIST). There is a 90.1% survival rate in patients that have localized CRC (Figure 1). This rate drops significantly to 71.2% when the cancer becomes regionalized and further drops to 13.5% when the cancer has metastasized.

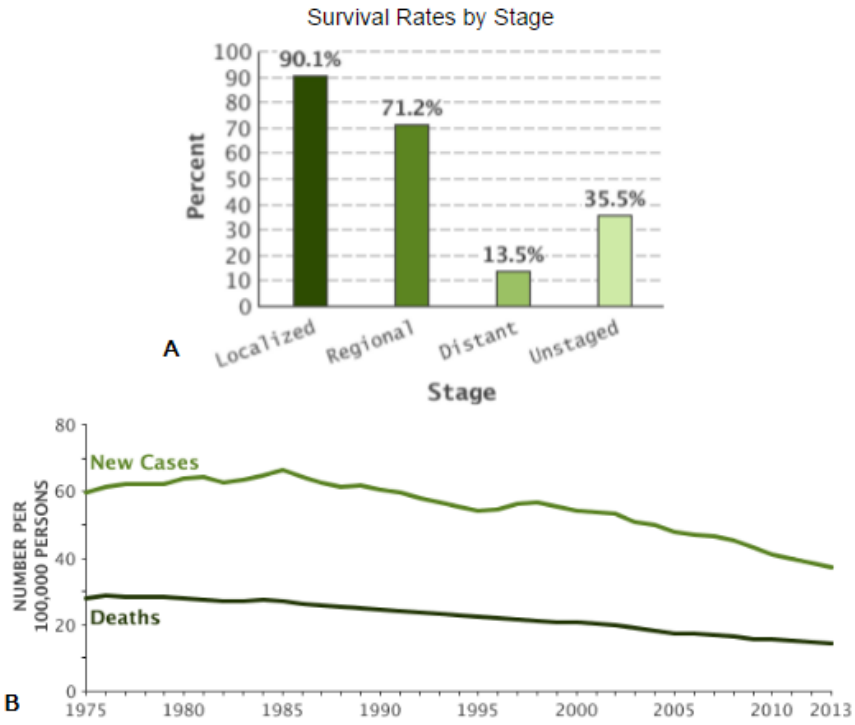


Figure 1: CRC Statistics. A) Five-Year Survival Rates by Stage. The cancers that are caught within the first two stages generally have the greatest chance of survival with an average survival rate of 90.1%. As the cancer progresses, the survival rates decreases. When the cancer becomes regional or when the cancer spreads to the surrounding lymph nodes, the chance of survival drops to 71.2%. When the cancer metastasizes, the chance of survival drops dramatically to 13.5%. (Image Source Reference [3], Copyright: see appendix)

1.1.2 Risk Factors and Prevention

Some risk factors include lack of physical activity, a diet high in meats and low in fruits and vegetables, obesity, smoking, heavy alcohol use, old age, family history of CRC or inflammatory bowel disease (IBD) among others [1]. Prevention measures include regular colonoscopies for polyp screenings, healthy lifestyle with a balanced diet and exercise, no smoking, limiting alcohol consumption, consuming adequate vitamins and minerals, and taking non-steroidal anti-inflammatory drugs (NSAIDs) like aspirin or ibuprofen on a regular basis.

1.1.3 Symptoms and Diagnosis

The symptoms of CRC can vary according to the location of the tumor. Tumors located on the right side of the colon typically have more space to grow so the cancer may be present for some time before symptoms appear [4]. Additionally, right-sided tumors cause iron deficiency anemia which is the main cause of the symptoms. Symptoms can include fatigue, and shortness of breath. The left side of the colon is narrower than the right. Oftentimes a tumor occurring in the left side can cause an obstruction resulting in constipation, narrow stools, cramps, bloating, and blood in the stool. Other symptoms include changes in bowel habits, abdominal discomfort, unexplained weight loss, and others.

Colorectal cancer is diagnosed with a colonoscopy followed by a confirmation via biopsy [5]. Following diagnosis a series of blood tests, diagnostic scans, and a physical exam are conducted for staging purposes. There are several biomarkers that are used to aid diagnosis in addition to lab tests and imaging [6]. These biomarkers include carcinoembryonic antigen (CEA), microsatellite instability (MSI), K-RAS, PIK3CA, and BRAF. CEA is a protein marker that is often elevated in CRC and metastasis, however, an elevated CEA does not always indicate CRC[7]. A positive, elevated CEA is found in about 60% of CRC with metastasis cases. It has been shown that 30-50% of CRC cases have a mutated K-RAS gene. This is an important marker because a positive result indicates that the cancer is less likely to respond to anti-epidermal growth factor receptor (anti-EGFR) therapy. Additionally, high levels of K-RAS is associated with poor survival and high cancer aggressiveness.

1.1.4 Colorectal Cancer Staging Guidelines

There are five stages of CRC: Stage 0, I, II, III, IV, illustrated in Figure 2 [8]. Each one of

these stages has a certain protocol of treatment. Stage 0 is characterized by the limited growth of the cancer, in other words when the cancer has not grown outside of the lining of the colon. Treatment for stage 0 CRC is to remove any cancerous tissue usually during a colonoscopy. Surgery is typically not needed unless the tissue cannot be excised during the colonoscopy. Stage I cancer is characterized by the growth of the cancer to the inner layers of the colon. Treatment for stage I is to excise any cancerous tissue and if necessary, a partial colectomy may be performed. Stage II cancer is characterized by the growth of the cancer into the walls of the colon and to neighboring tissue but not to the lymph nodes. Treatment for stage II cancer is to excise the cancerous tissue and in most cases a colectomy. In addition to surgical excision, adjuvant chemotherapy may also be necessary. Stage III cancer is characterized by the invasion of the cancer to the surrounding lymph nodes. Treatment includes surgical resection and adjuvant chemotherapy and/or radiotherapy. The commonly used chemotherapies for stage III CRC are FOLFOX (5-FU, leucovorin, and oxaliplatin) or CapeOx (capecitabine and oxaliplatin). Stage IV cancer is characterized by the metastasis of the cancer to another organ. In many cases, CRC metastasizes to the lungs, liver, or distant lymph nodes. In some cases, surgery cannot be performed due to the large size of the tumors or the number of tumors. In this case, chemotherapy and radiotherapy are administered to extend the patient's life. In other cases, the tumors in the colon and metastasized region (lung, liver, distant lymph nodes) can be resected and following resection, chemotherapy and radiotherapy are administered.

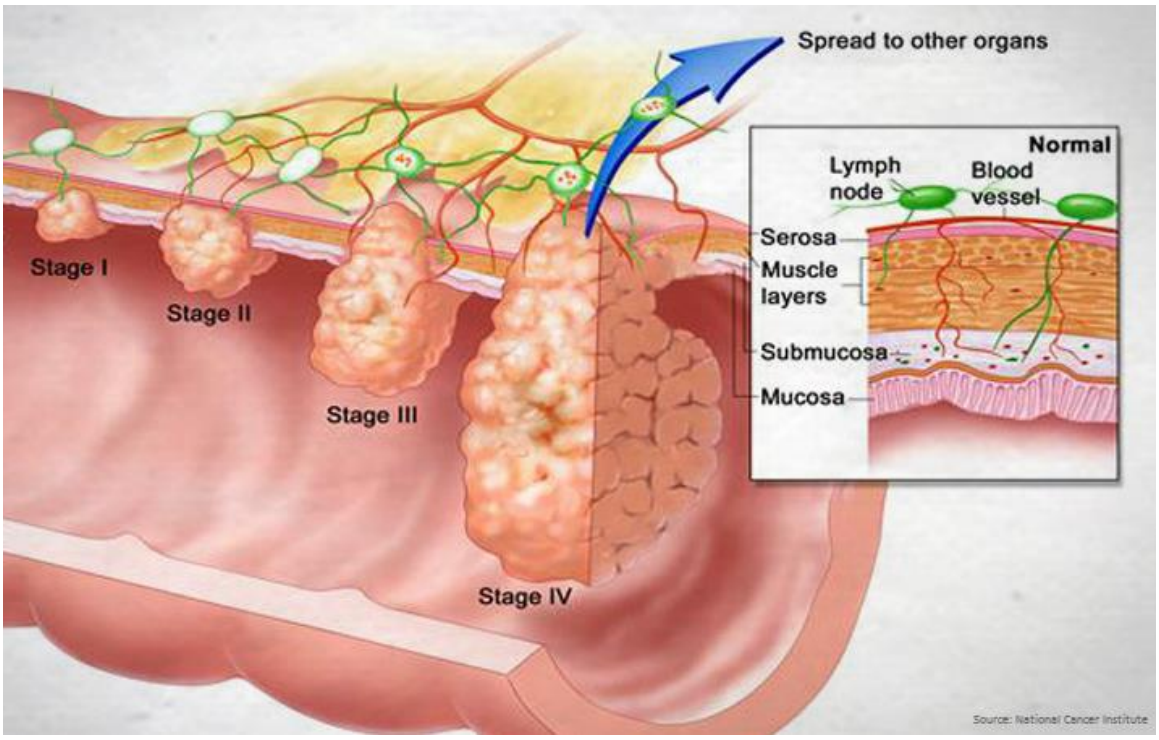


Figure 2: Stages of CRC. In stage I, the tumor is contained within the mucosa, submucosa, and muscle layers of the colon but does not penetrate the serosa layer. In stage II, the tumor grows and invades the serosa layer. In stage III, the tumor penetrates the serosa layer is is now growing outside of the colon and has spread to the surrounding lymph nodes. Finally, in stage IV, the tumor increases in size and metastasizes to distant organs such as the liver or lung. (Image Source Reference [9]; Copyright: see appendix)

1.1.5 Treatment of Colorectal Cancer

In most cases, treatment of CRC starts with surgery to remove the polyps and affected colon, rectum, and lymph nodes [8,10]. In addition to surgical resection, some cases require additional treatments such as colostomies, radiofrequency ablation, or cryoablation. Figure 3 shows the course of action for each stage. A colostomy is a procedure in which a portion of the colon is attached to an exit point in the abdomen where a pouch resides for waste to exit. This is often temporary to allow for the rectum to heal. In some cases where the cancer has metastasized to the liver or lungs, an ablation may be necessary. The two types of ablation currently in practice are radiofrequency ablation and cryoablation. Radiofrequency ablation is performed by

emitting a radio frequency to the tumor in order to heat it; cryoablation is the opposite in which the tumor is frozen. These techniques can be done during surgery or through the skin outside of surgery. The use of these techniques allow for minimal normal tissue damage, however, there is a risk of leaving some tumor behind.

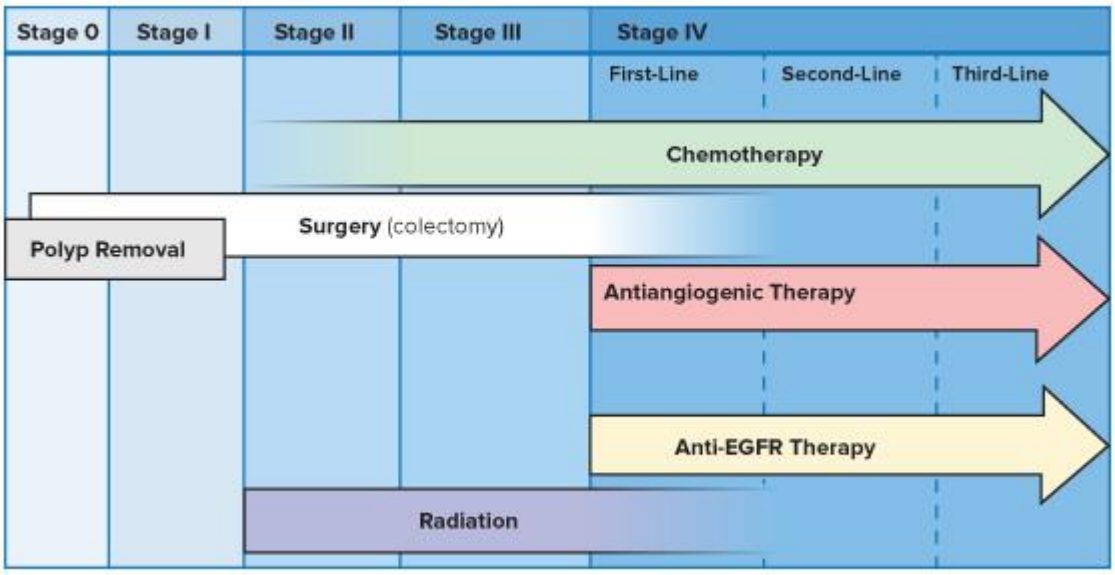


Figure 3: Treatment by CRC Stage. In stage 0 and I, the main line of defense is simply to remove the polyps. If the polyps are too numerous, then a colectomy may be considered. For stages II and III, a colectomy is the first step to remove the cancerous polyps and/or tumors. Following surgical removal, radiation and/or chemotherapy may be administered depending on the apparent aggressiveness of the cancer. In stage IV, a colectomy is performed to remove any tumors. Following surgery, as a first line of defense, radiation is administered along with chemotherapy. There are two general types of chemotherapy used for stage IV CRC, antiangiogenic and anti-EGFR. Depending on the biomarkers exhibited by the cancer, one or both of these types of chemotherapies will be administered. (Image Source Reference [11]; Copyright: see appendix)

Radiation therapy utilizes X-rays, gamma rays, and charged particles to kill cancer cells. Radiation may be used before surgery to reduce tumor size or after surgery to kill any remaining

cancer cells [12,13]. There are two main ways of delivering radiation - external-beam or internal-beam radiation. There are six main types of external-beam radiation therapy. 1) Intensity-modulated radiation therapy (IMRT) aims to deliver different doses of radiation to different parts of the tumor. This technique reduces normal cell damage as the areas on the edges of the tumor where normal cells reside receive a lower dose of radiation than the middle of the tumor. This has proven effective in reducing the side effects. 2) Image guided radiation therapy (IGRT) uses CT scans during radiation delivery. These CT scans allow for precision of dosing according to the position and exact current size of the tumor. 3) Tomotherapy is a more advanced version of IGRT in which the imaging and radiation is from a single rotating device. This allows for even more control and precision than the IGRT. 4) Stereotactic radiosurgery (SRS) is used for small tumors with a well-defined edge. This technique delivers high doses of radiation to a small tumor. This requires very accurate imaging in order to reduce normal tissue damage. This technique is commonly used for brain or spinal tumors. 5) Stereotactic body radiation therapy (SBRT) is used for tumors residing outside the brain or spine. 6) Proton therapy uses protons instead of X-rays or gamma rays. Protons deliver their energy according to the Bragg peak in which the highest amount of energy is delivered at a certain point in their path rather than evenly distributing energy throughout the path as electromagnetic rays . In theory, proton therapy should reduce normal tissue damage.

Internal-beam radiation or brachytherapy involves the delivery of radiation through radioactive seeds or pellets placed in the body with catheters, needles, or wires [12,13]. This treatment can be temporary or permanent. There are two main types of internal-beam radiation. 1) Interstitial brachytherapy involves the placement of the radioactive particles within the tumor. 2) Intracavitary brachytherapy is the placement of the radioactive particles in a body cavity that

is near the tumor.

There are several types and classes of chemotherapy that treat CRC. Some of these chemotherapy drugs can be used alone or in combination with a targeted chemotherapeutic. Some common treatment regimens are fluorouracil (5-FU) with leucovorin (together named Wellcovorin), Wellcovorin with oxaliplatin and Wellcovorin with irinotecan [10,12,13]. Some known targeted therapies are cetuximab, bevacizumab, or panitumumab; these targeted therapies can be used alone or in combination with the drugs mentioned. In general there are two types of targeted therapies: anti-angiogenesis and anti-epidermal growth factor receptor therapies. Anti-angiogenesis therapies aim to inhibit the growth of blood vessels to the tumor to limit nutrients and starve the tumor. Bevacizumab or Avastin is used with patients with advanced CRC. This anti-angiogenesis therapy inhibits vascular endothelial growth factor (VEGF) to inhibit blood vessel growth. Regorafenib or Stivarga is another anti-angiogenesis therapy that is approved for patients with metastasized CRC. Anti-epidermal growth factor receptors aim to inhibit cell growth. Cetuximab or Erbitux is a chimeric antibody composed of mouse and human components. This drug is given to patients with metastasized CRC with wild type KRAS. Panitumumab or Vectibix is a human based monoclonal antibody. This anti-EGFR is also used with patients with metastasized CRC with wild type KRAS.

1.2 RNA SEQUENCING

RNA sequencing (RNA-seq) is a method of high throughput, deep sequencing RNA either directly or indirectly through cDNA. RNA-seq aims to analyze the transcriptome at a given period of time, ie the time of RNA extraction [14]. RNA-seq can detect all types of RNA include miRNA, lncRNA, snoRNA, rRNA, tRNA, etc. Additionally, it can detect alternative

splicing arrangements, differential expression of genes between two conditions, post-transcriptional modifications, mutations, single nucleotide polymorphisms (SNPs), and gene fusions. Prior to the development of RNA-seq, microarray was the standard in studying transcriptomic elements. The microarray involves hybridization of the cDNA to a chip with immobilized DNA probes of known sequences. This was used to measure the expression of a sets of genes. The limitations of microarray are the chances of cross hybridization artifacts, poor quantification of low expressed genes, and prior knowledge of the sequences. In RNA-seq, the sequences do not need to be known, so novel genes could be identified with RNA-seq.

To prepare the cDNA for RNA-seq, first total RNA is extracted, purified, quantified, and measured for integrity [14,15]. Following, the RNA can be filtered to only include mRNA or it can be reverse transcribed as total RNA. Following cDNA synthesis, the cDNA is fragmented according to the sequencing machine ideals. Afterwards, the cDNA is sequenced by a next generation sequencer such as Illumina or Roche 454. This analysis is illustrated in Figure 4.

In general, there are two approaches to analyze RNA-seq data which are genome-guided and *de novo* synthesis [14,15]. The genome-guided approach uses a reference genome to map the transcripts to. This process starts by grouping the overlapping reads from each locus followed by creating a graph of all possible isoforms, and then settling on the individual isoforms. *De novo* assembly approach creates a transcriptome without a reference genome. This is done by primarily by locating overlapping sequences in the fragments, then forming the overlapping sequences into contiguous sequences (contigs), then by using the contigs, a scaffold is formed, and finally the gaps are filled within the scaffold and a unique gene or transcript is formed.

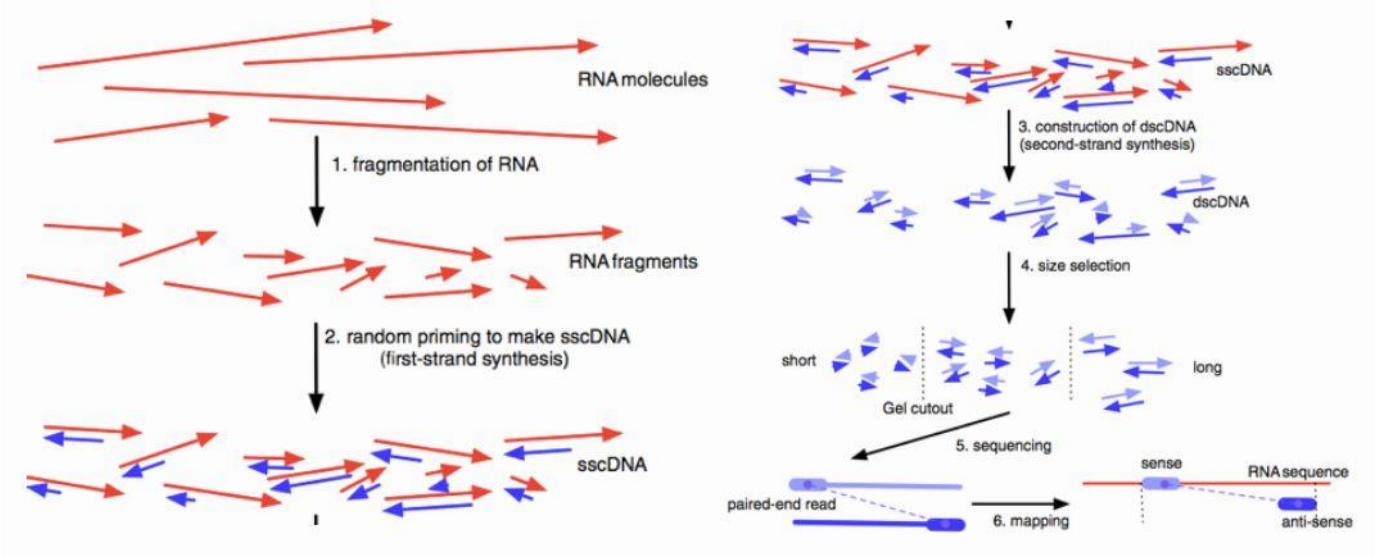


Figure 4: RNA Sequencing Schematic. 1) After RNA extraction and purification, the RNA molecules are fragmented. 2) The fragments are subjected to random priming to produce single stranded complementary DNA (sscDNA). 3) The sscDNA fragments are subjected to a second round of random priming to produce double stranded cDNA (dscDNA). 4) These fragments are selected based on size or number of base pairs. 5) The dscDNAs are sequenced in a sequencer. 6) The sequences are converted to a fasta or fastq file for RNA Seq analysis by Tuxedo Suite Pipeline or similar pipelines. (Image Source Reference [16]; Copyright: see appendix)

1.3 Wnt Signaling and TGFβ Pathways

1.3.1 Wnt Signaling Pathway Overview

The Wnt signaling pathway is a vital process that is involved in every stage of development. The Wnt pathway controls cell proliferation, cell adhesion, and cytoskeletal arrangement [17–20]. There are three variations of the Wnt pathway that controls each of these phenotypes: the canonical, noncanonical planar cell polarity (PCP), and noncanonical Wnt/Ca^{2+} pathways. The canonical pathway was first observed in mice in 1982 and five years later it was observed in *Drosophila melanogaster*. The Wnt pathways are crucial in the development of many organs such as, brain, eye, spinal cord, skin, teeth, heart, etc. In adulthood, the pathways are needed for tissue homeostasis, regeneration, and repair. Given its crucial nature, the Wnt

pathways are well studied in the instances of cancers. The canonical pathway is known to be dysregulated in a variety of cancers including CRC.

The important players of the Wnt canonical pathway include: Wnt ligands, β-catenin (CTNNB1), G-protein Frizzled receptor (Fzd), low density lipoprotein receptor related proteins (LRP), Dishevelled (Dvl), and T-cell factor-lymphoid enhancer factor complex (TCF/LEF) [17,18,20,21]. The noncanonical pathways are also termed β-catenin independent pathways. Important players in the noncanonical PCP pathway are: Wnt ligands, Fzd, JNK, Dvl, RhoA, and ROCK2. The important players in the Wnt/Ca^{2+} pathway are: Wnt ligands, Fzd, Dvl, PLC, Calcium/Calmodulin Dependent Protein Kinase II (CAMKII), PKC, and Calmodulin. Shown in Figure 5, is a simplified version of the Wnt pathways. In the canonical pathway, Wnt is recognized by the intermembrane complex Fzd/LRP complex which recruits the destruction complex comprising of several proteins including Dvl, axin, APC, and GSK3B [17,18,20]. This destruction complex, when recruited by the activation of Wnt/Fzd/LRP complex, can no longer target β-catenin for degradation. In turn, β-catenin accumulates and translocates to the nucleus and interacts with TCF/LEF complex to activate transcription of its target genes. In the PCP pathway, Wnt activates Fzd which activates a signaling cascade that starts with the activation of Dvl which activates RhoA and Ras. RhoA and Ras independently activate ROCK2 and JNK, respectively. ROCK2 and JNK are responsible for cytoskeletal rearrangement and cell polarity. Finally, in the Wnt/Ca^{2+} pathway, Wnt activates Fzd which activates Dvl. The activation of Dvl and its downstream targets (PLC, PLG) allow for the influx of calcium. The high concentration of calcium and the active form of Dvl activate CAMKII, PKC, and Calmodulin. CAMKII inhibits β-catenin action onto TCF/LEF in the nucleus, PKC activates CDC42 which promotes actin polymerization, Calmodulin promotes the translocation of NFAT to the nucleus which acts

as a transcription factor for genes important for the determination of cell fate and migration.

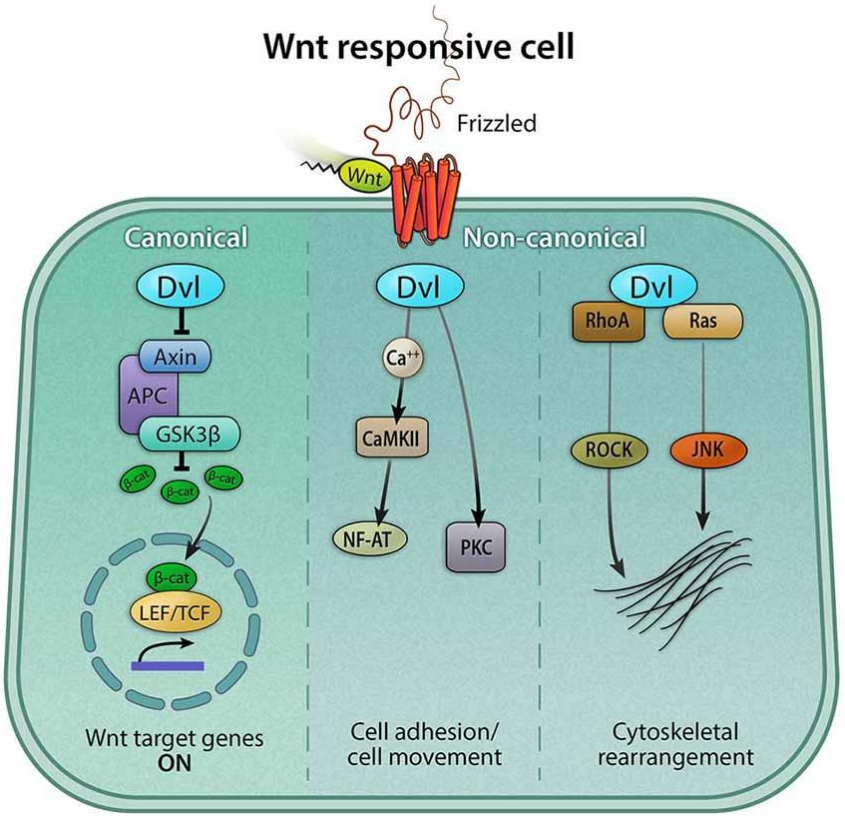


Figure 5: Wnt Pathway Schematic. In the Canonical Pathway, the WNT ligand binds to the Frizzled receptor and causes a signal cascade which activates Dvl to inhibit the β-catenin destruction complex. This inhibition allows β-catenin to accumulate in the cytoplasm which is then translocates to the nucleus to activate the LEF/TCF transcription factor complex. This complex then activates transcription of the target genes. In the Wnt/Ca^{2+} pathway displayed in the middle, upon the binding of the WNT ligand to Frizzled, Dvl is activated. Dvl activates PKC and also promotes the influx of calcium. PKC then activates CDC42 and promotes migration. The influx of calcium activates CAMKII which activates the transcription factor family NFAT. NFAT promotes cell migration and cell fate. The PCP pathway displayed on the right, starts with the WNT ligand binding to the Frizzled receptor in which Dvl is activated and recruits RhoA and Ras. RhoA activates ROCK2 and Ras activates JNK. These two proteins are responsible for cytoskeletal rearrangement. (Image Source Reference [22]; Copyright: see appendix)

1.3.2 Wnt Signaling Pathway in Colorectal Cancer

The Wnt pathway in CRC is known to have aberrant expression in all stages of CRC [19,21,23–25]. There are various mutations that account for the upregulation of Wnt in CRC. Two of

the most common mutations occur in the adenomatous polyposis coli gene (APC) and β-catenin [20,23,25]. Mutations in APC cause a loss of function of the protein. APC is a key component in the destruction complex which targets β-catenin. Mutations in APC can arise sporadically or from a hereditary illness, familial adenomatous polyposis (FAP) (Najdi Rani et al). Mutations in β-catenin cause a loss of phosphorylation site in the N-terminus which is pivotal for the destruction complex to recognize, phosphorylate, and degrade β-catenin. The loss of this site, in turn, allows for constitutive signaling of the canonical Wnt pathway (Morin, Patrice J., et al).

Other mutations that are common in CRC include mutations in other components of the destruction complex including Axin 1 and 2 [26,27]. Axin 1 and 2 are negative regulators of the Wnt pathway and are considered tumor suppressors. When mutated, these proteins cannot effectively degrade β-catenin, resulting in constitutive signaling. There are multiple reports of somatic mutations in Axin 1 and many reports of germ-line mutations in Axin 2. There are three germ-line mutations in Axin 2: two of them cause a premature stop codon in exon 7 creating a truncated, nonfunctional protein, and the other mutation causes a missense in exon 5. Moreover, not only have mutation been reported, but epigenetic silencing of Axin 1 and 2 are reported in CRC.

Wnt inhibitory factor 1 (WIF1) is a Wnt antagonist and acts by binding to the Wnt ligands to prevent binding to the Fzd/LRP receptor complex [28,29]. A report in 2015 in the *Familial Cancer Journal*, noted a mutation in WIF1 at amino acid 294 in which the mutation caused a cysteine to become a phenylalanine and was associated with higher cancer risk by promoting NAD synthesis [30]. Additionally, a WIF1 mutation has been reported in Nail-Patella-like Disorder in which a mutation in WIF1 may be a novel cause of this disorder [31]. In this study, a nonsense mutation was observed which resulted in the complete loss of the N terminus which

inhibited its ability to be secreted.

Secreted Frizzled-Related Protein 4 (SFRP4) is a member of the SFRPs and is an inhibitor of the Wnt pathway and targets the WNT ligands [32,33]. SFRPs contain a cysteine rich binding domain which gives SFRPs the capability to bind to Fzds. This results in a competitive binding with WNT ligands to bind to Fzd, additionally, SFRPs can bind directly with WNT ligands to inhibit their binding to Fzd, thus interfering with the Wnt pathway.

In addition to common mutations in tumor suppressors, there are countless genes that are epigenetically modified to be expressed or repressed. In most cases, these genes are tumor suppressors such as WIF1, Dickkopf 1 (DKK1), CDKN2B, or transcription factors such as CDX2 [34].

1.3.3 Transforming Growth Factor β Pathway Overview

The transforming growth factor β pathway is involved in cell proliferation, cell differentiation, apoptosis, extracellular matrix remodeling, immune functions, invasion, metastasis, and cellular homeostasis [35–38]. The TGFβ pathway has tumor suppressor roles and is often dysregulated in cancers due to mutations, deletions, or dysregulated regulatory factors of the components involved. There are more than 40 members that belong to the TGFβ pathway including activin/inhibin subfamily, bone morphogenetic proteins (BMPs), and others [39]. The TGFβ superfamily of proteins contain three members: TGFβ1, TGFβ2, and TGFβ3. These three proteins share ~70% homology.

There are two types of TGFβ pathways: Smad-dependent and Smad-independent [35,38,39]. In the Smad-dependent pathway, the TGFβ ligands interact with the TGFβ receptors (TGFβR), activin receptor, or BMP receptor which then causes a signal transduction within the receptor

bearing cell. This signal transduction cascade results in the recruitment and phosphorylation of Smad2 and Smad3. Phosphorylated Smads 2 and 3 bind with Smad4 and translocate to the nucleus to interact with a number of transcription factors to regulate target gene expression.

In the Smad-independent pathways, the activation of TGFβ causes activation in a number of other related pathways such as the p38 kinase pathway, MAPK pathway, Wnt PCP pathway, mTOR,pathway and PI3K/Akt pathway (Figure 6) [35,38,39].

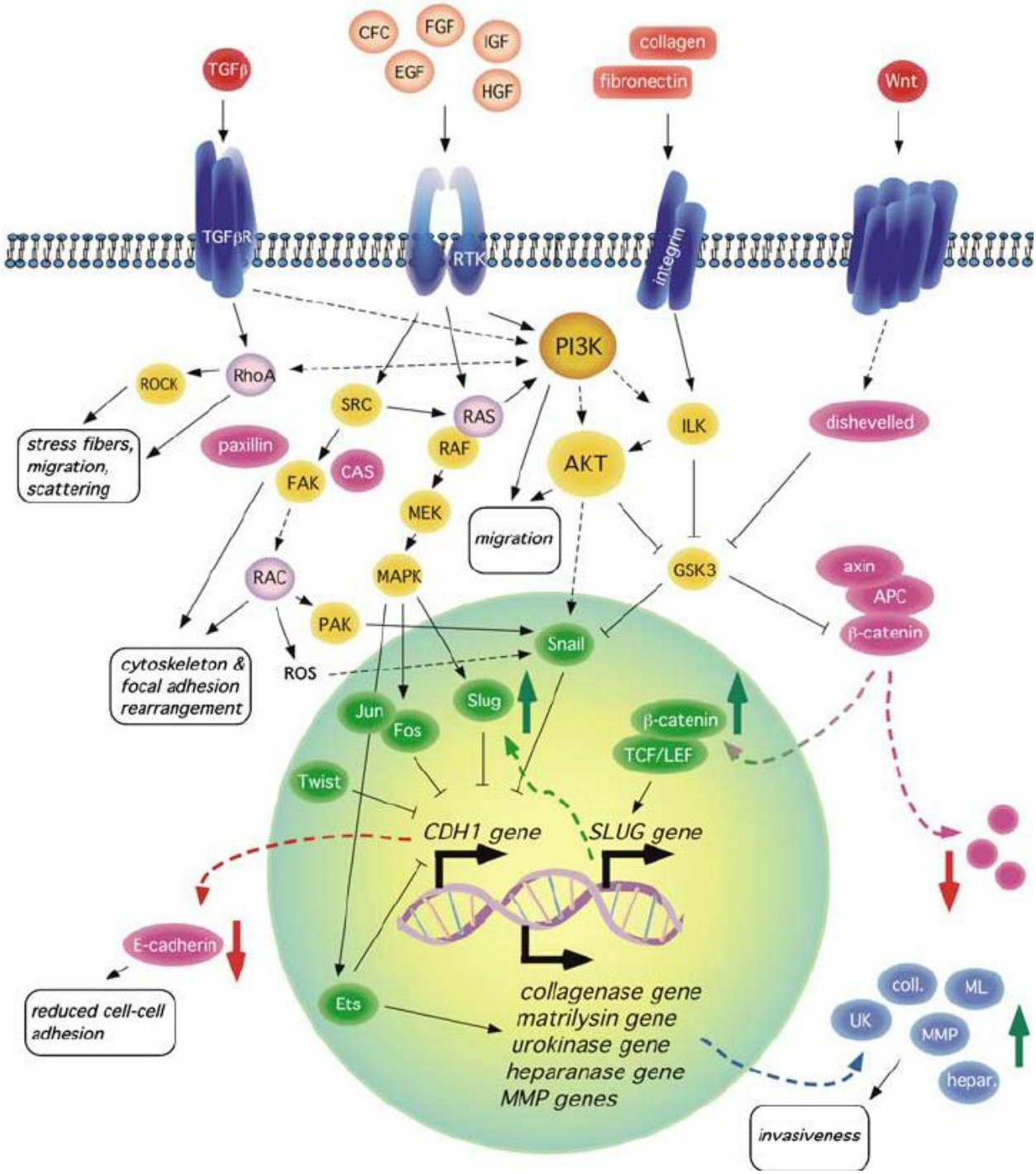


Figure 6. The Various TGFβ Pathways. TGFβ and its receptor starts various signal cascades. Overall, the TGFβ pathway signals for EMT through various intermediate molecules. Through PI3K/Akt, TGFβ promotes cell growth and EMT. Through RhoGTPases, TGFβ promotes EMT, migration, and invasion. Through Smads, TGFβ promotes tumor suppression and tumor prevention. Through p38/JNK, TGFβ promotes EMT and apoptosis. Through MAPK, TGFβ promotes EMT. (Image Source Reference[2]; Copyright: see appendix)

One of the many outcomes of the TGFβ pathway cell cycle arrest (Figure 7) [35,38,39]. TGFβ inhibits the cell cycle at G1 phase by inducing cyclin D kinase inhibitors: CDKN2B, CDKN1A, CDKN1B via Smad and and transcription factors Sp1 and FoxO. CDKN2B binds to the CDK4/6 complex which displaces CDKN1A and CDKN1B which allows CDKN1A to bind to CDK2-cyclin A/E complexes and CDKN1B to directly inhibit G1 progression. Additionally, TGFβ downregulates the oncogene, c-MYC and ID transcription factors through Smad and E2F4 and E2F5. The inhibition of c-MYC allows for the upregulation and activity of CDKN2B and CDKN1A to induce G1 phase arrest. The inhibition of the ID transcription factors allow for the upregulation of CDKN2A which also induce G1 phase arrest.

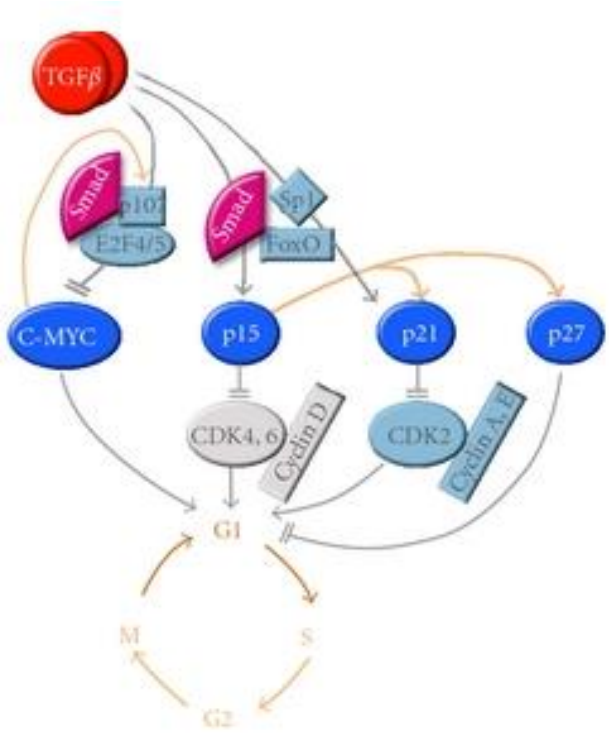


Figure 7: TGFβ and Cell Cycle Inhibition Pathway. The induction of TGFβ signals for cell cycle arrest by inhibiting c-MYC and inducing CDKN2B (p15), CDKN1A (p21), and CDKN1B (p27) through the Smad-dependent pathway. The inhibition of c-MYC prevented its oncogenic capabilities of cell cycle progression. The induction of p15 allowed for the removal of

p21 and p27 from CDK4 and 6. This allows for p15 to bind to CDK4 and 6 to induct G1 arrest. The displacement of p21 and p27 frees p21 to bind to CDK2-Cyclin A/E complex which inhibits its ability to promote G1 progression. The freedom of p27 allows it to directly inhibit G1 progression. (Image Source Reference [39]; Copyright: see appendix)

1.3.4 TGFβ Pathway in Colorectal Cancer

TGFβ exhibits a dual role in cancer as it shows tumor suppressive and tumor promoting capabilities. In colorectal cancer, several plays of the TGFβ pathway are mutated. TGFβR3 is often mutated and results in a truncated final protein [39]. Additionally, it was found that Smad2 mutations are common in CRC which may be due to deletions, frameshift mutations, nonsense or missense mutations. The mutation of Smad2 results in its inability to form a complex with Smad4, inhibit the activation of Smad-mediated transcription, or increase the expression of Smad repressors. Moreover, the increased expression of Smad7, reported in pancreatic, endometrial, and thyroid follicular cancers, antagonize the TGFβ pathway.

In the case of epithelial to mesenchymal transition (EMT), the increased expression of the TGFβ pathway leads to the promotion of EMT [36,37,39,40]. TGFβ represses its downstream target, microRNA-200 (miR-200), which leads to the increased expression of the target of miR-200, ZEB2. The expression of ZEB2 decreases the expression of E-cadherin which leads to cell migration and invasion.

The TGFβ pathway has been shown to be an indicator of prognosis [39]. TGFβ pathway has been shown to regulate telomerase reverse transcriptase (TERT), responsible for immortalization of cancer, by negatively regulating its transcription through Smad3 and the transcription factor E2F1.

1.3.5 CD82

CD82 belongs to the tetraspanin family of proteins and is a tumor metastasis suppressor

that has been found to inversely correlate with cell invasion, motility, metastasis, and cell survival [41,42]. CD82 is involved in various pathways including the TGFβ and Src kinase-dependent pathways. CD82 was first discovered in 1995 as a tumor metastasis suppressor in rats for the treatment of prostate cancer [43]. Since then, recent studies have reported metastasis suppression capabilities in bladder, breast, hepatocellular, and pancreatic cancers [41,42,44,45]. CD82 acts with accordance with integrins and receptor tyrosine kinases, however, its exact mechanism is unclear.

CD82 is regulated on three levels: epigenetic, transcriptional, and post translational [44,46,47]. On the epigenetic level, HOTAIR, a lncRNA, targets CD82 for methylation and thus silencing. On the transcription level, there are various factors that induce the transcription of CD82 at the promoter region including, JunB, ATF3, AP1, AP2, NF-kB and p53 [46]. Additionally, CD82 is upregulated in hypoxic conditions via HIF-1a to promote metastasis in the innermost layer of the tumor micro environmet[47].On the post translational level, CD82 is recruited to the endoplasmic reticulum for N-glycosylation and degradation by ubiquitin ligase gp78. A summary of the factors affecting CD82 function is illustrated in Figure 8.

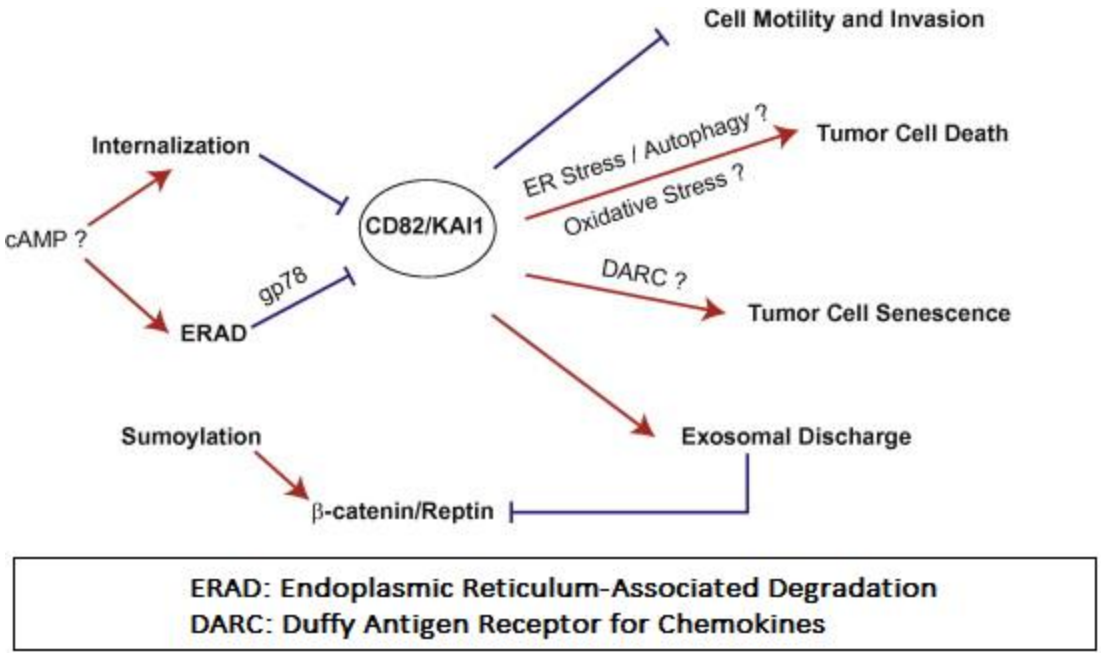


Figure 8: The Regulation and Outcomes of CD82. CD82 has several regulatory factors and can regulate a variety of processes. The increase of cAMP in the cell signals for ERAD which signals for gp78 to post translationally degrade CD82. When CD82 is expressed it can inhibit cell motility and invasion, activate tumor cell death, activate tumor cell senescence, and activate exosomal discharge. In turn, exosomal discharge inhibits β-catenin/reptin. (Image Source Reference [47]; Copyright: see appendix)

1.4 Long Non Coding RNA

1.4.1 Introduction to lncRNAs

Long noncoding RNA (lncRNAs) is a new discovery in the field of biology and medicine [48–50]. LncRNAs are RNA molecules that are greater than 200 bp, polyadenylated, and lack an open reading frame. They are a part of the noncoding RNA (ncRNA) class that includes microRNAs (miRNAs), small nuclear RNA (snoRNA), small interfering RNA (siRNA), etc. The non-protein coding region of DNA is ~70-90% of the total genome, however, very little is known about ncRNAs. The current research being done is investigating the ncRNAs to show that they not "junk" as once believed. These ncRNAs have an immense effect on regulation, epigenetic modifications, splicing, etc.

There are roughly 32,000 ncRNAs transcribed versus 21,000 proteins transcribed [50]. LncRNAs are of interest since these RNAs can fold into 3-dimensional structures in which may interact with transcription factors, histones, DNA, RNA, and other biomolecules to create duplex or triplex structures. LncRNAs have numerous functions including: chromatin modification, direct transcriptional regulation, RNA processing events such as splicing, post-translational modifications, modulation of microRNA, and gene silencing through endogenous siRNA production. These functions are illustrated in Figure 9 below.

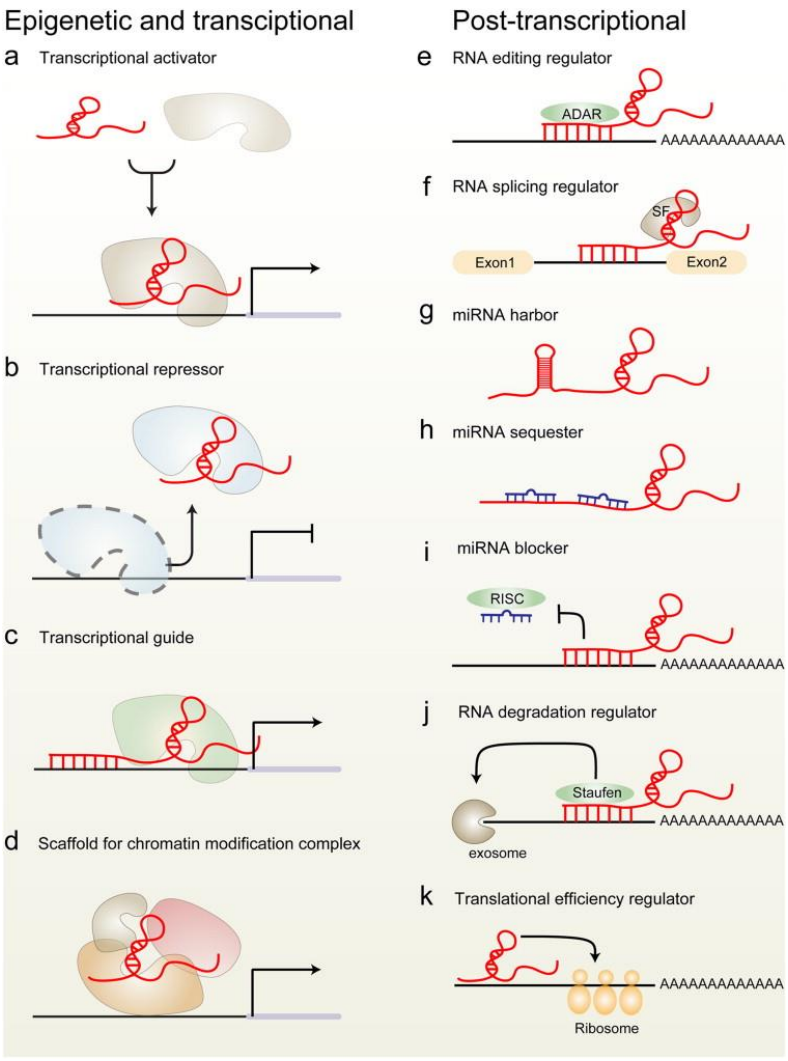


Figure 9: Long Non Coding RNA Functions. A) LncRNAs can act as a transcriptional activator by recruiting RNA polymerase. B) LncRNAs can also act as a transcriptional repressor by sequestering RNA polymerase which prevents RNA polymerase from transcribing DNA. C) LncRNAs can act as a transcriptional guide in which the lncRNA can guide the RNA polymerase to the lncRNA target gene. D) LncRNAs can act as a scaffold by recruiting proteins to perform chromatin modifications. E) LncRNAs can edit RNA via the ADAR deaminase. F) LncRNAs can promote splicing in conjunction with splicing factors. G) Some microRNAs are coded within a lncRNA and therefore some lncRNAs can harbor microRNAs for later use. H) Some lncRNAs can have complementary sequences to microRNAs so that the microRNAs bind to the lncRNAs instead of their mRNA targets, therefore these lncRNAs effectively sequester microRNAs. I) LncRNAs can have complementary binding sites similar to microRNAs; however, instead of sequestering microRNAs, these lncRNAs bind directly to the mRNA target, thus blocking the action of the microRNAs. J) LncRNAs can degrade RNA with the use of Staufen. The lncRNA-Staufen complex recruits an exosome which degrades the target RNA. K) LncRNAs can increase transcription efficiency by increasing ribosomal efficiency. (Image Source Reference [51]; Copyright: see appendix)

It is difficult to characterize lncRNAs because of their range in size, structure, location, and function. Based on the location of the lncRNA genomic target, lncRNAs can be classified as *cis* or *trans* shown in Figure 10 [49,52]. *Cis* acting lncRNAs are act on neighboring genomic locations and *trans* acting lncRNAs are act on a variety of genomic locations. LncRNAs can be classified based on location: sense, antisense, bidirectional, intronic, and intergenic, shown in Figure 11 [48,49]. Sense lncRNAs overlap one or more exons on the same strand. Antisense is the opposite of sense; it overlaps one or more exons on the opposite strand. Bidirectional transcripts have the same start site as a protein coding gene. Intronic lncRNAs originate from the introns of a gene. Intergenic lncRNAs arise from regions between two genes.

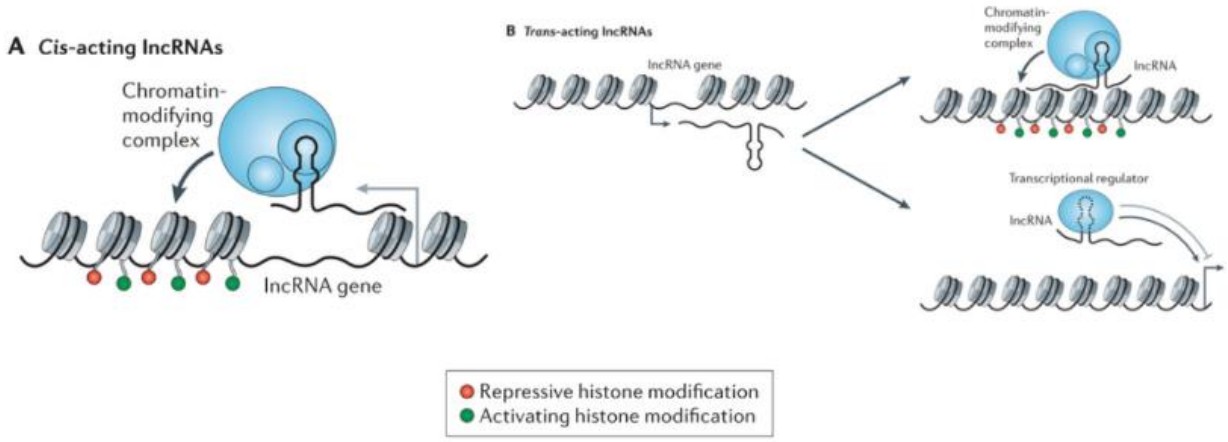


Figure 10: Cis and Trans acting lncRNAs. A) Cis acting lncRNAs are lncRNAs that are transcribed and act in the same region. B) Trans acting lncRNAs are transcribed and act in a distinct genomic locations. (Image Source Reference [51]; Copyright: see appendix)

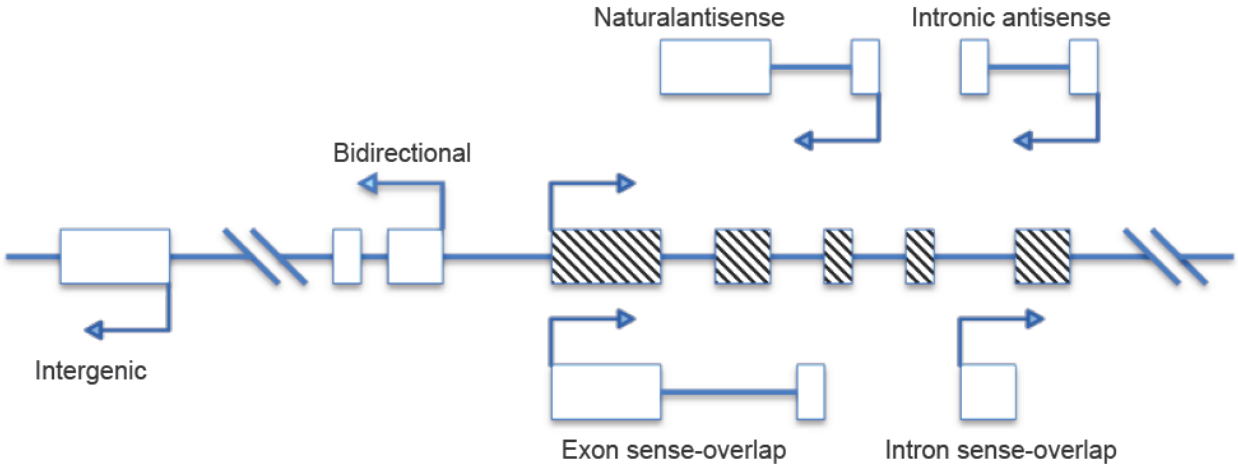


Figure 11: The origins of lncRNAs. There are five types of lncRNAs based on their origin in the genome: Intergenic, bidirectional, sense, natural antisense, and intronic. Intergenic lncRNAs arise from regions between two genes. Bidirectional lncRNAs arise from bidirectional promoters. Sense lncRNAs are lncRNAs that arise in the sense direction of the gene and contain exons of the parent gene. Natural antisense lncRNAs arise from the antisense direction of the parent gene and contains exons. Intronic lncRNAs can be categorized into two types: intronic antisense or intronic sense. Intronic antisense lncRNAs arise from the introns in the antisense direction. Intronic sense lncRNAs are sense lncRNAs from the intronic regions. (Image Source Reference [53]; Copyright: see appendix)

Most importantly, lncRNAs show a great potential in the field of cancer pathology. As more information concerning lncRNAs develops, the better cancer pathology, dysregulated pathways, and progression can be understood.

The most notable lncRNAs have a genome wide effect which include Xist, HOTAIR, MALAT-1, ANRIL, and ZEB1 and ZEB2 [3,49,50,54,55]. These lncRNAs, when dysregulated, can cause hallmark cancer phenotypes such as epigenetic modifications, dysregulation of apoptotic genes, dysregulation of cell cycle genes, alternative splicing, protein localization, stem cell pluripotency, etc.

1.4.2 Molecular Mechanisms

The molecular mechanisms of lncRNAs can be classified into four groups according to what the lncRNA acts as: as a signal, as a decoy, as a guide, and as a scaffold as shown in Figure 12 [56,57]. Since lncRNAs are translated by RNA polymerase II like mRNA, lncRNAs can show cell type specific expression and respond to stimuli to echo the changing dynamic of the cell environment. Since most lncRNAs alter chromatin state, if the lncRNAs are known, the chromatin state can be deduced.

One example of a signal lncRNA is Xist [56,57]. Xist enacts on the X chromosome for gene imprinting in order for the cell to only transcribe one of the two X's. Xist is polyadenylated and capped and is only transcribed by the X chromosome to be inactivated [58]. Xist inactivates its target by coating the X chromosomes. This acts as a signal for recruitment of chromatin remodeling factors. Other lncRNAs may function as signals for DNA damage, pluripotency, reprogramming, enhancing transcription, and repeated or abundant transcription.

When lncRNAs act as decoys, they act as an effector that results in no effect on its target,

inhibiting function of the target[56–58] . LncRNAs can act as decoys for microRNAs, splicing factors, and proteins. In the case of lncRNA PANDA, binds to NF-YA to inhibit the expression of apoptotic genes. PANDA has been shown to be expressed during DNA damage and influences cell survival. When PANDA is depleted or knocked out, a high frequency of apoptosis was observed. In order to repress apoptosis, some cancers overexpress PANDA.

The third mechanism lncRNAs may undergo is guiding. The lncRNA will bind to its protein target and guide the protein to a location either in *cis* or in *trans* [57]. In *cis,* lncRNAs can stimulate chromatin changes cotranscriptionally as they are bound to RNA polymerase II or lncRNAs can bind complementarily to small regulatory RNAs. In *trans,* lncRNAs bind to DNA to create a heteroduplex (RNA: DNA) or triplex (RNA:DNA:DNA). Both mechanisms result in changing the chromatin structure. HOTAIR acts in *trans* by targeting, recruiting, and guiding PRC2 to act on the histones to methylate the target genes [3,57,59].

The last mechanism type is lncRNAs acting as scaffolds. These lncRNAs often possess a variety of domains to bind to different targets [57]. TERC, a telomerase RNA component, is a lncRNA that acts as a scaffold for telomerase. TERC works in combination with TERT (telomerase reverse transcriptase) and TERRA (lncRNA with repeated telomeric sequences) [48,56,57,60]. TERC has two functions: 1) to act as a template for TERT and 2) to act as a scaffold for protein subunits to bind. TERC is associated with sustaining and regulating the length of the telomeres. Telomere length is directly related to cellular aging and senescence.

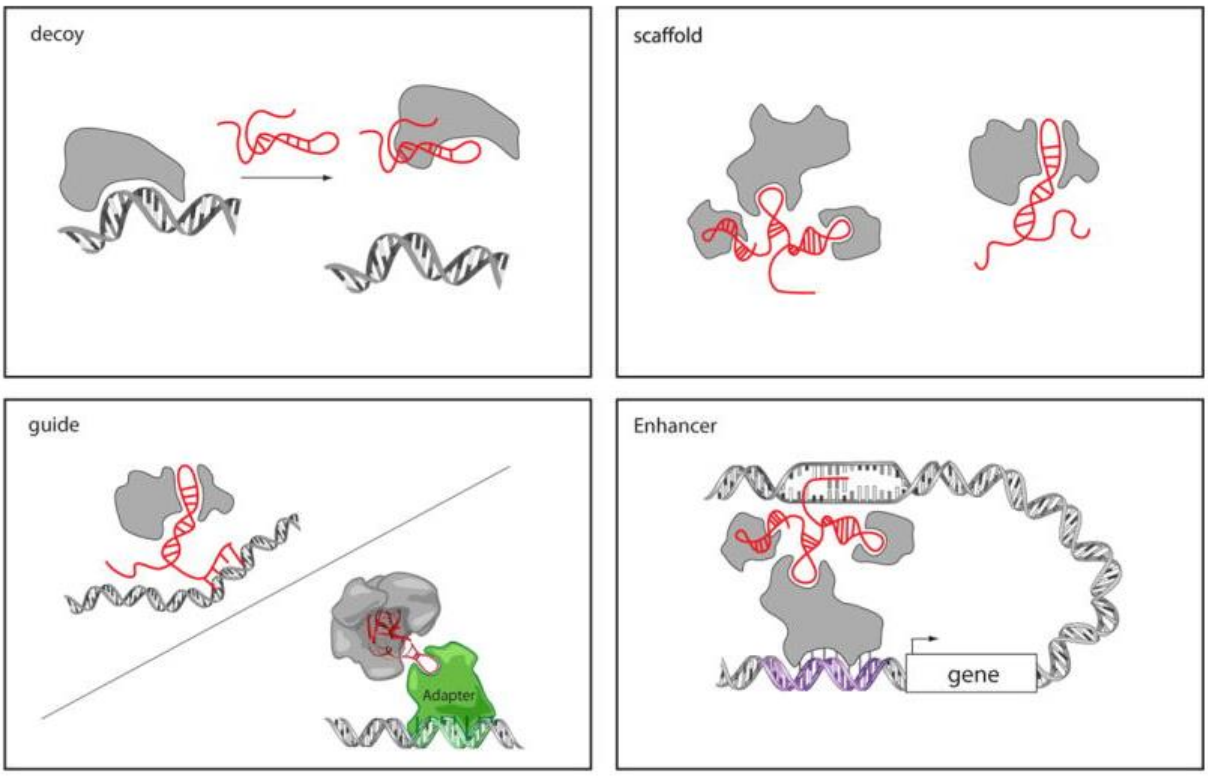


Figure 12: Molecular mechanisms of lncRNAs. There are four main molecular mechanisms for lncRNAs. LncRNAs can act as a decoy, scaffold, guide, or enhancer. LncRNAs can act as a decoy by binding to a protein which prevents the protein from performing its original action. As a scaffold, lncRNAs can recruit proteins which together perform a distinct function such as chromatin remodeling. As a guide, lncRNAs can guide transcription factors or RNA polymerases to the target gene. As an enhancer, lncRNAs can use its scaffolding abilities to perform gene enhancement such as demethylation of a gene to promote transcription. (Image Source Reference [61]; Copyright: see appendix)

1.4.3 HOX Transcript Antisense RNA

HOX transcript antisense RNA (HOTAIR) is one of the first known lncRNAs. HOTAIR is comprised of ~2 kb and 6 exons [3,50,57,59]. It originates from the antisense strand of the *HOXC* gene and is a *trans*-acting lncRNA and has a variety of targets (Figure 13). In general, HOTAIR alters the epigenetic modifications on chromatin shown in Figure 9. HOTAIR targets, recruits, and guides the Polycomb Repressive Complex 2 (PRC2). PRC2 silences its targets through its histone methyltransferase activity. Additionally, HOTAIR can recruit the lysine specific demethylase 1 (LSD1) which demethylates histone H3 at lysine 4 to un-silence or activate its

target. HOTAIR has the capability to silence and un-silence genes concurrently by binding to PRC2 at the 5' end and LSD1 at the 3' end.

Recent studies have found that HOTAIR is a major component in many cancers due to its ability to promote mobility, proliferation, apoptosis, invasion, aggression, metastasis, and stemness of the cells [3,50,57,59]. HOTAIR may be a likely biomarker. A meta-analysis was done in 2014 by Cai *et al,* in order to evaluate the association between HOTAIR and lymph node metastasis [59]. They found that higher levels of HOTAIR correlated with high incidence of metastasis.

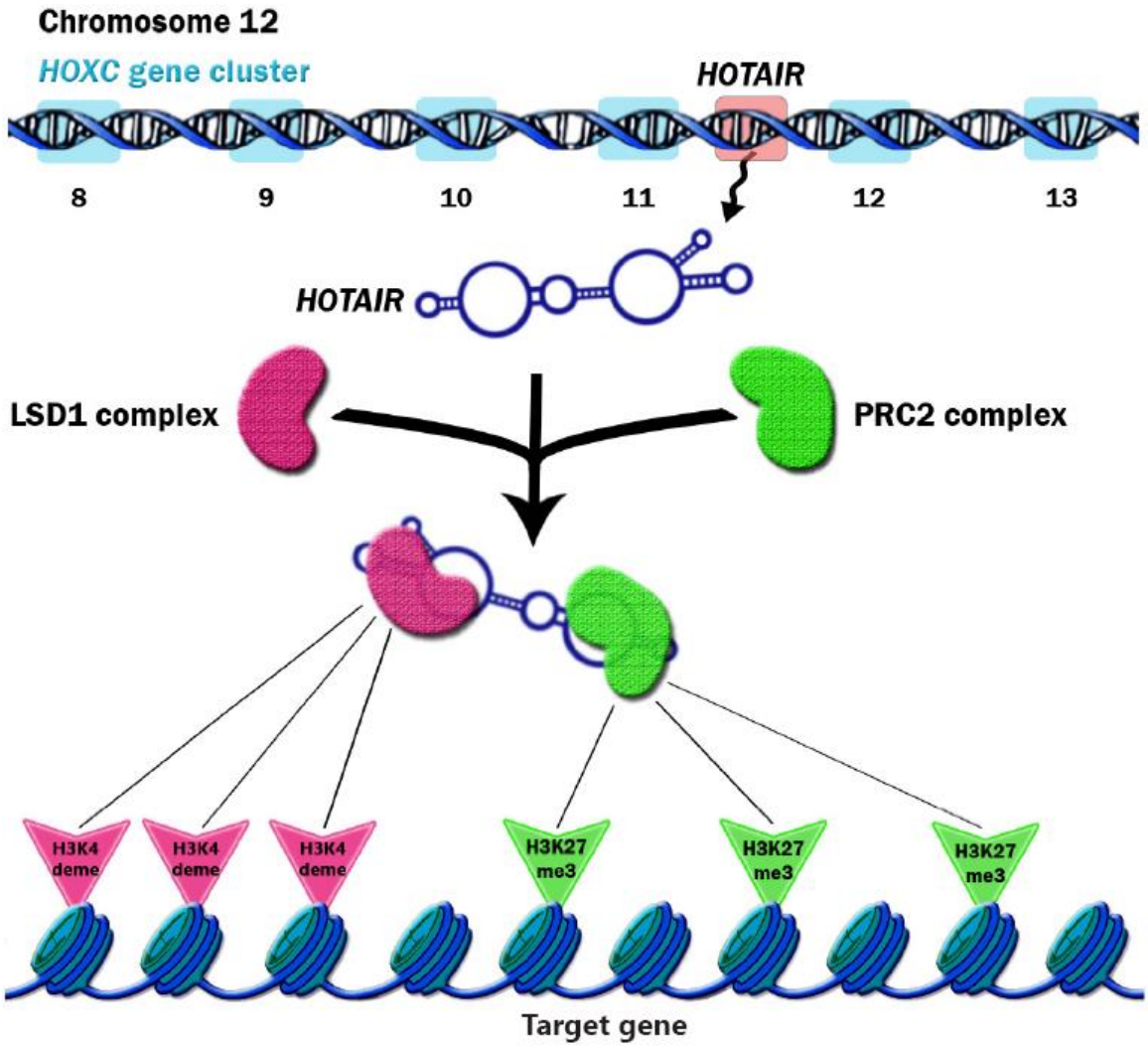


Figure 13:HOTAIR mechanism of action. HOTAIR is located on chromosome 12 in the HOXC gene cluster between exons 11 and 12 and is in the antisense direction. HOTAIR acts as a scaffold for the recruitment of the Polycomb repressive complex 2 (PRC2) and the lysine-specific histone demethylase 1 (LSD1) complex. These two complexes have opposing roles: PRC2 methylates target gene promoters and silences expression while the LSD1 demethylates previously silenced genes to promote their transcription. (Image Source Reference [3]; Copyright: see appendix)

1.4.4 Urothelial Carcinoma Associated 1

Urothelial Carcinoma Associated 1 lncRNA is 1.4 kb in length and was first discovered in bladder cancer was found to be associated with tumorigenesis and invasion [62–64]. It was also found as a tumor promoter in breast, colorectal, gastric, and esophageal squamous cell carcinoma. A recent study completed in 2016, found that in CRC cells, the upregulation of UCA1 was associated with the uncontrolled growth through the p53/p21 signaling pathway [65]. In oral squamous cell carcinoma (OSCC), it was found that UCA1 may also have a relationship with the Wnt signaling pathways [65]. In UCA1 knockdown experiments, it was found that the Wnt pathway was significant decreased as well as its downstream targets were downregulated, however, the mechanism of how UCA1 modulates the Wnt pathway is unresolved. One report shows that the overexpression of UCA1 was correlated with at least 2-fold expression of two Wnt pathway members: WNT5A and FZD5 [64], however, the mechanism of action is unknown.

There is little information on the structure or mechanism of action of UCA1. What is known is that UCA1 is strongly associated with cancer progression, cell proliferation, apoptosis, and chemoresistance [66–70]. In CRC, UCA1 has been shown to be markedly upregulated shown in many studies suggesting that UCA1 may be a candidate for the use of a prognostic marker, diagnosis, or as a therapeutic target [71].

1.4.5 Colorectal Neoplasia Differentially Expressed

Colorectal Neoplasia Differentially Expressed (CRNDE) is a lncRNA that is known to be differentially expressed in colorectal cancer [54,72,73]. This lncRNA was originally found in CRC with a high expression. It is located on chromosome 16 on the antisense strand of the gene IRX5. CRNDE contains 10 kb, 5 exons, and has been shown to be alternatively spliced [54]. CRNDE is

also found to be overexpressed in various cancers including, ovarian, pancreatic, liver, kidney, and leukemia. CRNDE has been shown to regulate neuronal differentiation, gametogenesis, and cancer progression. Some recent studies have suggested that CRNDE may also be involved in chromatin remodeling; however, more investigation is needed for confirmation [72]. Since CRNDE is normally involved in brain and gamete development, it is found to be tissue-specific and temporally regulated [73]. In adulthood, CRNDE expression levels are little to none in most tissue types.

The exact mechanism of CRNDE is not yet elucidated. One study as shown that CRNDE may negatively regulate microRNA 186 (miR-186) in human glioma stem cells (GSCs) [74]. In this study, the overexpression of CRNDE was shown to decrease miR-186 and increase cell proliferation, migration, invasion, and decrease apoptosis. While the overexpression of miR-186 had the opposite effects of decreasing cell proliferation, migration, invasion, and increasing apoptosis rates. The mechanism of action of these lncRNAs are not yet elucidated, however, there are several correlations that have been made. In addition to establishing correlations between lncRNAs and vital cancer promoting genes, the investigation of genetic variations in lncRNAs will facilitate understanding of the mechanisms of actions and downstream effects.

1.5 Single Nucleotide Polymorphisms

Single nucleotide polymorphisms (SNPs) are mutations, insertions, or deletion of a single nucleotide [75–77]. SNPs have been known to be the underlying reason for various cancers and diseases. SNPs can be categorized by the location: protein-coding, non-protein coding, and intergenic. The consequence of a SNP can be categorized as synonymous or nonsynonymous with nonsynonymous SNPs having two subcategories: missense and nonsense mutations. Missense mutations result in a different amino acid than the original sequence and nonsense

mutations result in an early stop codon which creates a truncated protein. SNPs in the non-protein coding and intergenic regions do not cause amino acid changes, however, these SNPs may cause changes in gene splicing, transcription factor binding, and regulatory element detection. There are various known SNPs linked with colorectal cancer including SNPs in microRNAs and genes in the interferon signaling pathway [75–77].

CHAPTER 2: HYPOTHESIS AND OBJECTIVES

2.1 Hypothesis

Stage IV colorectal cancer exhibits the standard metastasized characteristics of invasion, uncontrolled cell proliferation, deletions, mutations, and dysregulation of various pathways. However, more information needs to be uncovered about the underlying mechanisms and the role of lncRNAs to fully understand the scope and environment of metastasized colorectal cancers.

Therefore, it is hypothesized that the interplay of lncRNAs, inhibitors, activators and mutations in these players that are involved in cell proliferation pathways may have a greater role in metastasized colorectal cancer than thought before and may be therapeutic targets.

2.2 Objectives

1. To determine differentially expressed genes in colorectal cancer
2. To determine a vital differentially expressed pathway in colorectal cancer
3. To predict a novel dysregulation between differentially expressed protein coding genes and long non coding RNAs in the differentially expressed pathway
4. To determine possible SNPs in the genes involved.
5. To suggest a pathway between long non coding RNAs and protein coding genes

CHAPTER 3: MATERIALS AND METHODS

3.1 Data Source and Sample Selection

The data was obtained from the Gene Expression Omnibus (GEO) where the data is stored with the accession number GSE560760. The data was produced by Seon-Kyu Kim et al, a research group in South Korea in 2014 [78]. Their published paper is entitled, "A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients" and was published in the *Molecular Oncology Journal*, Volume 8, Issue 8, December 2014.

The study used matched tissue samples from 18 patients who had primary CRC cancer and liver metastasis. Three sample tissue types were extracted from each patient: primary CRC tissue, liver metastatic tissue, and normal colon tissue. The authors extracted total RNA, assembled a sequencing library, and ran RNA sequencing with Illumina 2000. All patients were treated in the Asan Medical Centre in Seoul, South Korea between May 2011 and February 2012.

3.2 RNA-seq Data Processing

In this current study, five paired normal and cancer RNA datasets were analyzed. The reads were paired-end reads and were processed by splitting the ends and trimming and filtering poor quality reads. The processed reads were then assessed for quality using FastQC v 0.11.2 (available online at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) [79].

3.3 RNA-seq Alignment and Differential Expression

After the initial processing, the Tuxedo Suite Package was used in order to determine the differentially expressed genes from the processed RNA sequences from paired normal and cancerous tissues (Figure 14). First, the processed reads were aligned to the human genome using

Bowtie v.2.2.6.0 (http://bowtie-bio.sourceforge.net/index.shtml) and TopHat v.2.1.0 (https://ccb.jhu.edu/software/tophat/index.html) [80]. Then the sequences were assembled into transcripts using CuffLinks v2.2.1 (http://cole-trapnell-lab.github.io/cufflinks/install), then the transcripts for all replicates are merged together by CuffMerge v2.0.2 (http://cole-trapnell-lab.github.io/cufflinks/cuffmerge/) [81,82]. Differential expression of the transcripts was determined by using CuffDiff v2.2.2 (http://cole-trapnell-lab.github.io/cufflinks/cuffdiff/), and finally, the differential expression was visualized using the R package, CummeRbund v3.2.3 (http://bioconductor.org) [83].

Bowtie is a short sequence aligner that is used for quick read alignment to the reference genome. Additionally, Bowtie indexes the genome using the Burrows-Wheeler transform (BWT) algorithm [84]. Indexing allows for faster analysis in TopHat.

TopHat uses the index created by Bowtie to align the reads on the genome with splice junctions [80]. This algorithm is optimized to align short reads with little error and is able to identify novel splice sites.

Cufflinks is used for transcript assembly, abundance estimation, and differential expression [85]. Cufflinks estimates the abundance of the transcripts based on the number of supporting transcripts while excluding any biases from sequence library preparation.

Cuffmerge is used to merge the Cufflinks transcripts from each replicate into one composite file [82]. This step is not vital for the pipeline, however, it allows for fluid final differential expression analysis.

Cuffdiff is used to assembly transcripts and to find the differential expression of genes, transcription start sites, splicing, and promoters [83]. This is the final step that renders the differential expression of the two conditions. These results are visualized in CummeRbund, an R

program.

CummeRbund uses the output from Cuffdiff and creates a database of the results and describes their relationship and significance [85]. This is then used to visualize the data in the form of boxplots, scatterplots, volcano plots, heatmaps, and others.

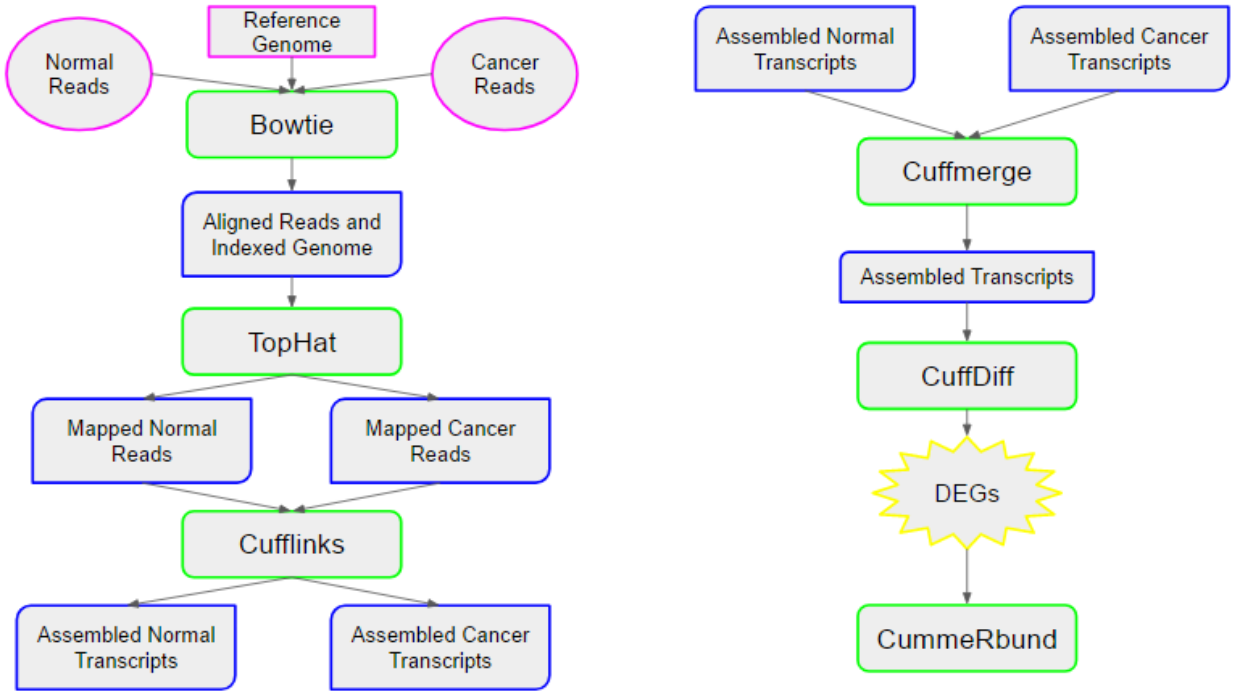


Figure 14: Tuxedo Suite Pipeline Workflow. The starting components for the Tuxedo Suite are the raw sequences or reads of the two conditions in question and a reference genome. These three components are entered to Bowtie which aligns the reads to the genome and also indexes the genome for faster computing. TopHat uses the aligned reads and maps these to the genome. Cufflinks uses these mapped reads and assembles the associated transcripts. Cuffmerge combines the normal and cancer transcripts to one composite file. Cuffdiff measures the differential expression between the two conditions and gives the final output of complete list of differentially expressed genes (DEGs). Finally, CummeRbund, an R program, is used to visualize these results.

3.4 Differential Gene Analysis

The workflow to analyze the differentially expressed genes is shown in Figure 15. To classify the DEGs, the PANTHER database (v 11.1) was used [86]. PANTHER is available online at http://pantherdb.org/about.jsp. PANTHER is an acronym for Protein ANalysis THrough Evolutionary Relationships and is a classification tool designed for analysis of high-throughput results. PANTHER is a part of the Gene Ontology Phylogenetic Annotation Project and contains 13,096 protein families, divided into 78,442 functionally distinct protein subfamilies. The genes are classified in four ways: family and subfamily, molecular function, biochemical process, and pathway. Three sets of genes were analyzed: all DEGs, upregulated genes, and downregulated genes.

lncRNA2Target (v 1.2) is a webtool which finds the lncRNA that targets the gene of interest or it can find what genes a particular lncRNA targets [34]. LncRNA2Target is available online at https://www.lncrna2target.org/. The evidence supporting the results are from literature. Each result has the link to the supporting report. Using the results from PANTHER, a pathway and its DEG were identified. Focusing on the WNT proteins and WNT inhibitors, a table of associated lncRNAs was established.

The Integrative Genomics Viewer (IGV) (v 2.3.97) was used to visualize transcripts and sequences of the data [87]. IGV is a desktop tool and is available for download at http://software.broadinstitute.org/software/igv/. The input data was derived from TopHat which is the aligned and mapped transcripts. IGV can be used to search for specific genes to visualize the number of transcripts in a locus and also to visualize mutations such as deletions, insertions, and SNPs.

Once a SNP was found in IGV, the database of SNPs (dbSNP) was used to determine if

the SNP was previously recorded. dbSNP is a public online database maintained by NCBI and is available at https://www.ncbi.nlm.nih.gov/snp/ [88].

The SNPs were then further evaluated for function and consequence with FuncPred. FuncPred is a SNP prediction webtool provided by the NIH and is available at https://snpinfo.niehs.nih.gov/snpinfo/snpfunc.html [89].

Any potentially novel SNPs were examined for predicted function via DeepSEA [90]. DeepSEA uses deep learning algorithms to determine predicted binding sites in variant sequences.

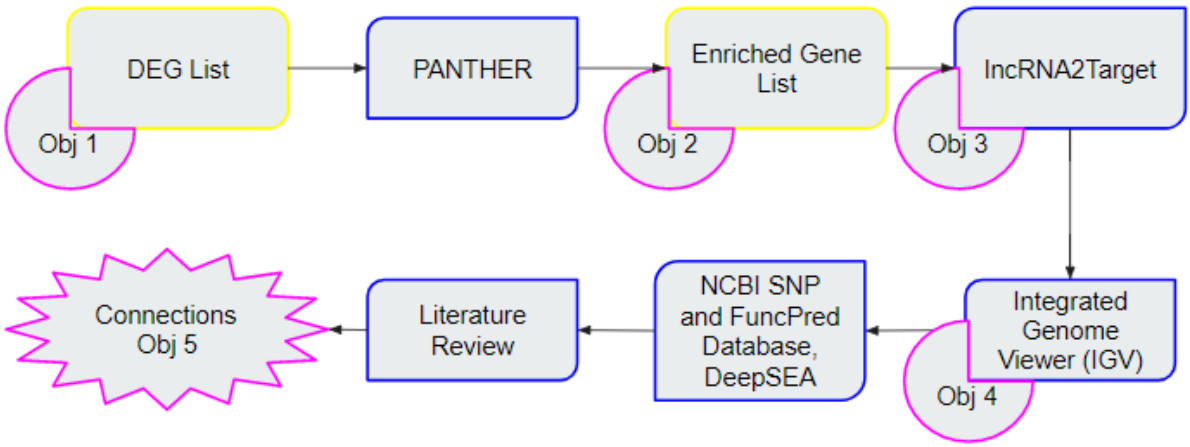


Figure 15: DEG analysis workflow. After obtaining the DEGs from the Tuxedo Suite Pipeline, the genes were entered into PANTHER for pathway enrichment. Activators and inhibitors from the most enriched set of genes were entered into the lncRNA2Target database to determine if these genes had associated lncRNAs. Once a complete set of activators, inhibitors, and associated lncRNAs were gathered, they were visualized in IGV to determine any genetic variants. Any genetic variants or SNPs were further investigated in NCBI SNP database, FuncPred database, and if there were any potentially novel SNPs, DeepSEA predicted software was used to obtain information on the consequence. With this information combined with literature reviews, a novel connection could be made.

CHAPTER 4: RESULTS

4.1 Metadata and Data Statistics

Prior to starting the Tuxedo Suite pipeline analysis, each of the samples was analyzed for its number of reads, GC content, and gender. These results are shown in Table 1. The samples analyzed were normal colon mucosa and paired primary CRC from male patients from South Korea [78].

Table 1: Description of Normal and Cancer Datasets. Each sequence dataset has an identifier prefix of 'N' or 'C' signifying Normal or Cancer samples, respectively. All normal samples are paired with cancer samples. Each sample has its corresponding number of reads.

Sample Identifier	Tissue Type	Gender	Stage	Number of Reads
N1	Normal Colon Mucosa	Male	-	54,992,947
N2	Normal Colon Mucosa	Male	-	23,653,357
N3	Normal Colon Mucosa	Male	-	67,611,184
N4	Normal Colon Mucosa	Male	-	50,662,612
N5	Normal Colon Mucosa	Male	-	38,477,372
C1	Primary CRC	Male	IV	64,924,130
C2	Primary CRC	Male	IV	56,433,491
C3	Primary CRC	Male	IV	43,648,154
C4	Primary CRC	Male	IV	58,551,634
C5	Primary CRC	Male	IV	58,368,923

4.2 Differentially Expressed Genes

After analyzing the datasets with the Tuxedo Suite pipeline analysis, the DEGs were visualized via CummeRbund. Differentially expressed genes were identified with two filters: p-value and fold change. Genes with a p-value of 0.05 or less and a fold change of less than -0.5 or

greater than +0.5 were considered significant. The summarized output of the Tuxedo Suite pipeline analysis is shown in Table 2. Table 2 shows several variables identified in the data including the total number of open reading frames, number of promoters, number of transcription starts sites, number of splicing sites, and others. The number of differentially expressed genes was 5,303 (Table 3). Also included in Table 3 is the range of fold change values and the number of up- and downregulated genes. The scatter plot shows the distribution of the genes under both conditions, normal and cancer conditions (Figure 16). The volcano plot is used to visualize the number of significant versus insignificant genes (Figure 17).

Table 2: Tuxedo Suite Pipeline Output. This table shows the overall differential expression output. There were 57,176 total genes, 335,486 isoforms, 142,942 TSS, 82,459 CDS, 57,176 promoters, and 142,942 splicing sites.

Total Open Reading Frames (ORFs)	57,176
Isoforms	335,486
Transcription Start Sites (TSS)	142,942
Coding DNA Sequence (CDS)	82,459
Promoters	57,176

Table 3: Differentially Expressed Gene Summary. The number of differentially expressed genes with a p-value of less than or equal to 0.05 and with a fold change value less than or equal to -0.5 and greater than or equal to +0.5 were 5,303. The fold change values ranged from 5.676 to +10.738. The number of upregulated genes were 3,070. The number of downregulated genes were 2,233.

Differentially Expressed Genes, P value ≤ 0.05	5,303
Range of Log2 Fold Values	-5.676 to +10.738
Number of Upregulated Genes	3,070
Number of Downregulated Genes	2,233

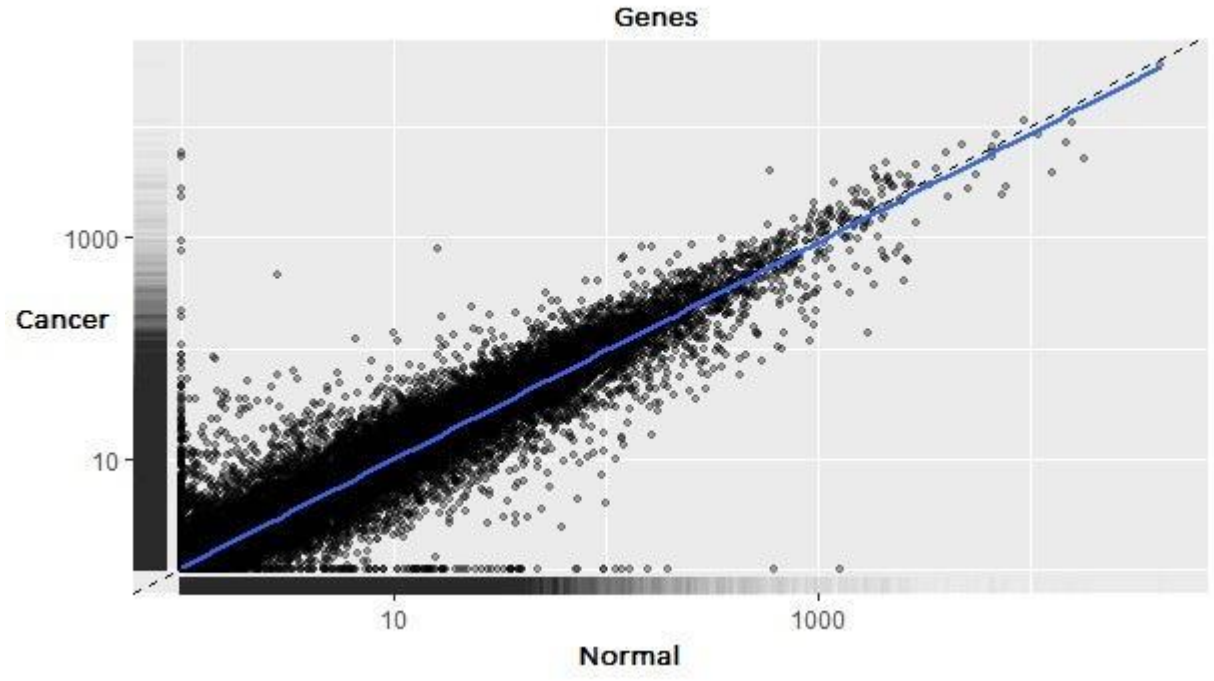


Figure 16: Scatter Plot of Genes Expressed in Normal Colon Mucosa vs Genes Expressed in CRC. The scatter plot maps the genes expressed in both conditions and in this scatter plot there is a roughly linear relationship between most of the genes expressed in both conditions. The data points that are outside of the linear portion are the genes that are differentially expressed.

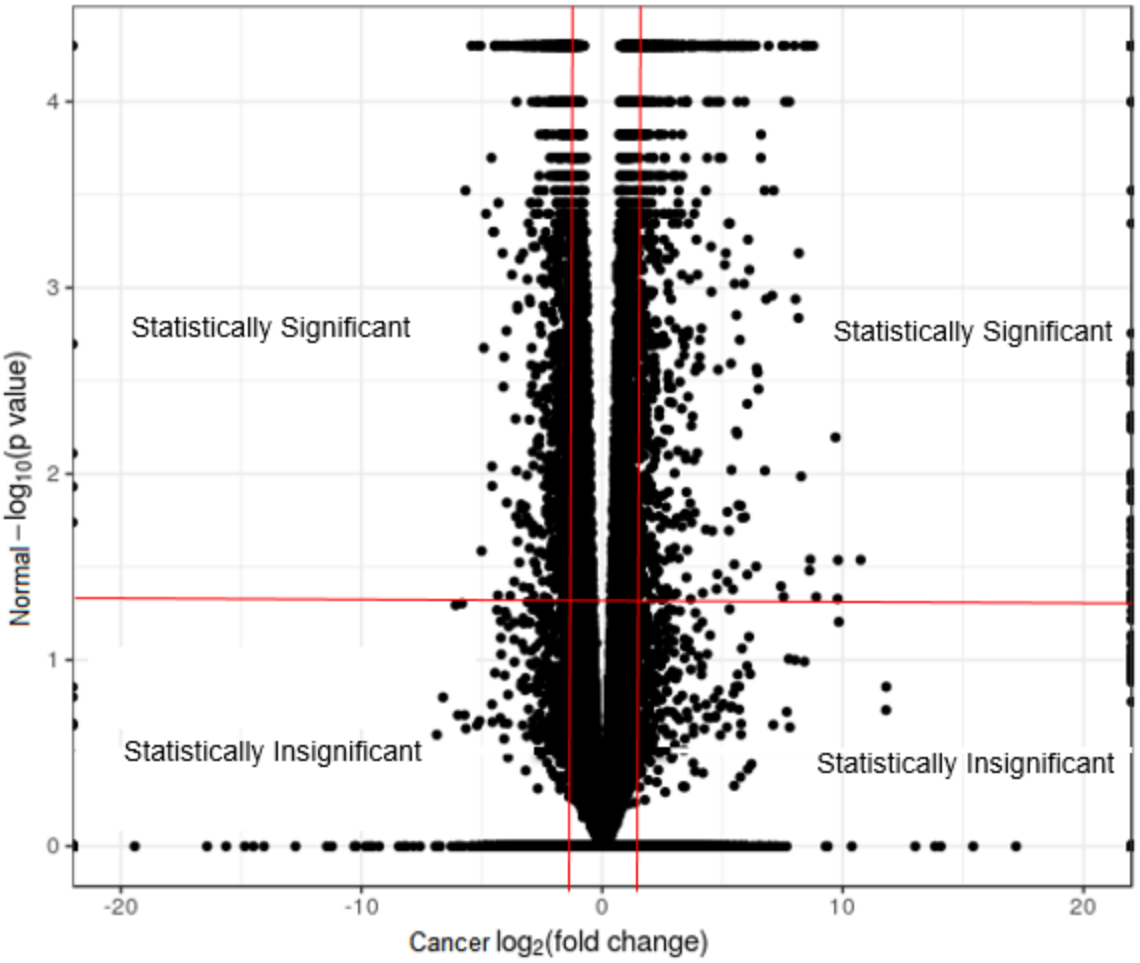


Figure 17: Volcano Plot showing the Relationship between the Normal and Cancer Samples by cummeRbund. Using the p-value cut off of less than or equal to 0.05 and the fold change cut off of greater than or equal to 0.5 and less than or equal to -0.5, the biological and statistical significance were determined. The data points above the red line show the biological and statistical significant genes.

4.3 Pathway Enrichment Analysis

Pathway enrichment analysis was done using the Panther database. This analysis generated a pie chart with the names of the pathways that the DEGs were involved in and the percentage of genes involved in each pathway (Figure 18). Pathway enrichment was done three times, once with all of the DEGs, once with only the upregulated genes, and once with only the downregulated genes. In all of the DEGs, the Wnt pathway was the most enriched set representing 4.9% of the total genes. In the upregulated genes/ORFs, the Wnt pathway represented most enriched set with 7.2%. Finally, when analyzing the downregulated genes/ORFs, the Wnt pathway was the fifth most enriched set with only representing 2.5% of the genes.

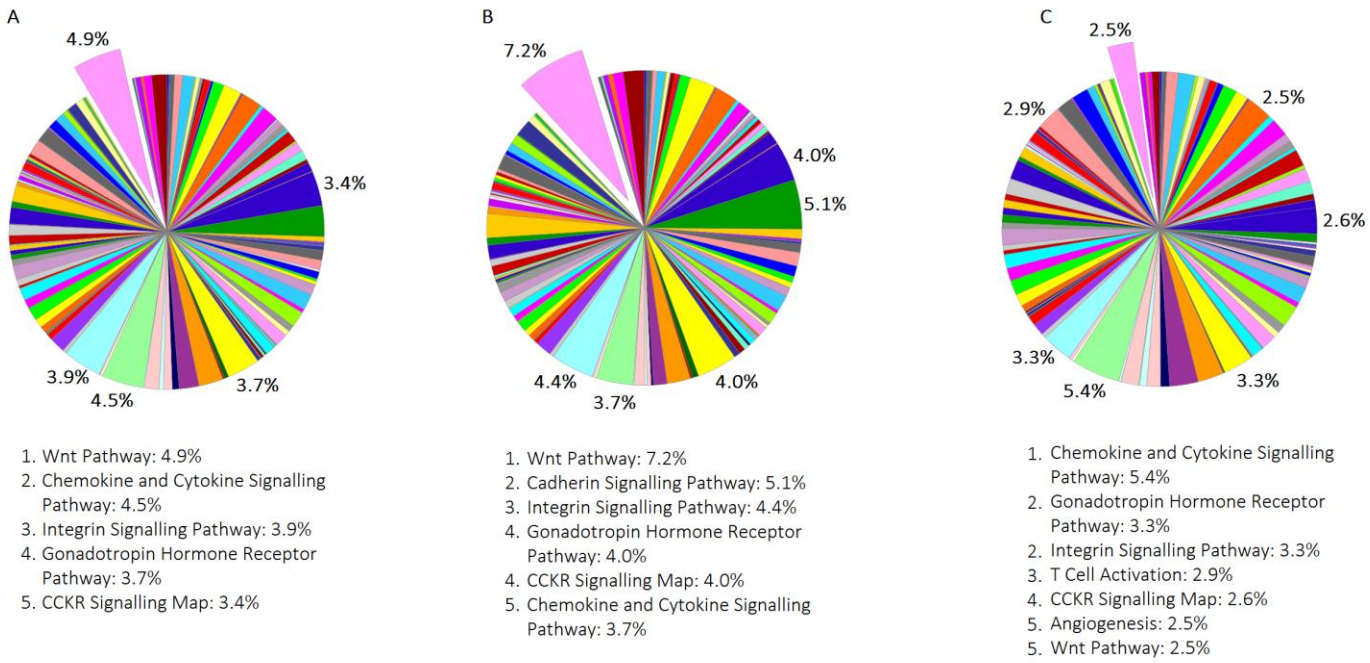


Figure 18: Pathway enrichment analysis by Panther DB. A) This pie chart represents the pathway enrichment for all the differentially expressed genes and ORFs, up- and downregulated. The top 5 enriched pathways are listed with their corresponding percentages. The Wnt pathway was the pathway the most enriched with 4.9% of the genes representing it. B) This pie chart represents the pathway enrichment for only the upregulated genes. The top 5

enriched pathways are listed with their corresponding percentages. The Wnt Pathway is at the top and represents 7.2% of the genes. C) This pie chart represents the pathway enrichment for the downregulated genes. The top 5 enriched pathways are listed with their corresponding percentages. The Wnt Pathway ranks at the fifth most enriched pathway with its genes representing 2.5% of the downregulated differentially expressed genes.

The differentially expressed Wnt pathway genes were identified and a literature review was done on the Wnt pathway and the Wnt pathway DEGs. Tables 4, 5, and 6 show the Wnt pathway genes, the inhibitors, and the targets of the Wnt pathway and their associated fold change, respectively, which confirms that the Wnt pathway is active.

Table 4: Differentially Expressed Wnt pathway genes. The table is organized from lowest level to highest fold change. A majority of the genes are upregulated as expected from the Panther analysis. Several Wnt and Frizzled (Fzd) proteins are upregulated which initiate the Wnt pathway.

Wnt Pathway Genes	**Fold Change**
PRKACB	-1.497
PRKCBA	-1.387
TCF21	-1.383
WNT9A	-1.383
MAPK10	-1.247
FZD1	-0.862
RILP	-0.796
CSNK2A1	0.593

AXIN1	0.615
CTNNB1	0.626
PRICKLE4	0.833
TCF3	0.860
FZD6	0.943
LRP11	0.974
LRP3	1.041
GPC4	1.058
PPP2R3B	1.279
CCND1	1.429
WNT3	1.503
AXIN2	1.650
WNT11	1.996
CDK8	2.237
WNT5A	2.261
FZD10	2.677
LEF1	2.804

LRP8	3.319
NKD1	3.353
NKD2	4.913
WNT2	5.722
MMP7	6.032

Table 5: Differentially Expressed Wnt pathway inhibitors. This table is organized from lowest to high fold change. Note: all inhibitors are upregulated with DKK4 having an infinite fold change, this is due to the undetected expression in the normal samples.

Wnt Inhibitors	**Fold Change**
SFRP4	2.121
APCDD1	2.384
DKK2	4.342
WIF1	5.165
NOTUM	6.812
DKK4	INF/Undetected in Normal Samples

Table 6: Differentially Expressed Target Genes of the Wnt Pathway: The table is organized from lowest to highest fold change. All of the target genes are upregulated except for ATOH1 which is expected to be downregulated when the Wnt pathway is promoted.

Gene	**Fold Change**
ATOH1	-1.114

PLAUR	0.842
VEGF	1.069
BIRC5	1.266
CD44	1.311
CCND1	1.429
MET	1.512
AXIN2	2.050
MYC	2.355
NRCAM	2.580
LEF1	2.804
CLDN	4.631
MMP7	6.032

4.4 Long Non Coding RNA Target Analysis

In order to find what lncRNAs target a particular gene and vice versa, a database named lncRNA2Target was utilized. The inhibitors and WNT ligands were focused on (Tables 7 and 8). WIF1 and SFRP4 specifically target the WNT ligands. By target analysis, HOTAIR was found to target and silence WIF1. This result is supported by a study completed in 2013 by Xiao-Song Ge et al in which they found that HOTAIR directly methylates and silences WIF1 in esophageal squamous cell carcinoma and other groups have verified this relationship in bladder, gastric, and pancreatic cancers [91–93].

Table 7: lncRNA2Target results for DE Wnt ligands. The differentially expressed Wnt ligands were entered into lncRNA2Target db in order to find associated lncRNAs. WNT9A has no known associated lncRNAs. lncRNA CRNDE is known to activate WNT2 and WNT3. TUG1 is known to target WNT11, however TUG1 is not DE in this dataset. UCA1 is known to target and activate WNT5A.

WIF1 Target	Fold Change	P value	Associated lncRNA	Fold Change	P value	lncRNA Action
WNT9A	-1.503	0.00075	None	-	-	-
WNT3	***1.503***	***5.00E-05***	***CRNDE***	***3.227***	***0.0032***	***Positively Regulates***
WNT11	1.996	5.00E-05	TUG1	-	-	Positively Regulates
WNT5A	***2.261***	***5.00E-05***	***UCA1***	***5.542***	***5.00E-05***	***Positively Regulates***
WNT2	***5.722***	***0.0019***	***CRNDE***	***3.227***	***0.0032***	***Positively Regulates***

Table 8: lncRNA2Target Database results for DE Wnt inhibitors. The differentially expressed Wnt inhibitors were entered into lncRNA2Target database to find associated lncRNAs. CRNDE is known to negatively regulate SFRP4. MALAT1 is known to negatively regulate APCDD1, however MALAT1 is not differentially expressed in this dataset. TUG1 is known to positively regulate DKK2, however TUG1 is not differentially expressed. HOTAIR is known to negatively regulate WIF1. There are no known lncRNAs associated with NOTUM or DKK4.

Wnt Inhibitors	Fold Change	P value	Associated lncRNA	Fold Change	P Value	lncRNA Action
SFRP4	***2.121***	***0.00185***	***CRNDE***	***3.227***	***0.0032***	***Negatively Regulates***
APCDD1	2.384	5.00E-05	MALAT1	-	-	Negatively Regulates
DKK2	4.342	5.00E-05	TUG1	-	-	Positively Regulates
WIF1	***5.165***	***5.00E-05***	***HOTAIR***	***5.198***	***0.03785***	***Negatively Regulates***
NOTUM	6.812	0.00115	None	-	-	-
DKK4	INF	0.00175	None	-	-	-

4.5 Differentially Expressed HOTAIR Target Genes

HOTAIR has many known target genes, and its effect can either be to unsilence or to silence the target gene. There are 11 known targets for HOTAIR in which HOTAIR un-silences five and silences six. Table 9 shows a list of the target genes that were differentially expressed with their associated expected and observed effect, along with the expression and p values. There were two genes that acted opposite to the expected effect, WIF1 and CD82.

Table 9. Differential Expression of HOTAIR Target Genes. A list of HOTAIR target genes was obtained from lncRNA2Target database. This list was cross referenced with the DEGs and a table was generated showing the expected and observed effect with the corresponding expression value and p value. WIF1 and CD82 are two targets that did not show the expected effect.

Gene	Expected Effect	Observed Effect	Log Fold Change	P Value
MMP1	↑	↑	4.706	5.0E-5
MMP3	↑	↑	5.54	5.0E-5
MMP9	↑	↑	1.425	5.0E-5
VEGFA	↑	↑	1.06	0.0004
WIF1	↓	↑	***5.165***	***5.0E-5***
CD82	↓	↑	***1.05***	***0.00075***

4.6 IGV SNP Analysis

Normal and cancer datasets were loaded into IGV and the following genes were investigated: WIF1, SFRP4, WNT2, WNT3, WNT5A, CRNDE, UCA1, HOTAIR, and CD82. Eleven SNPs were found in five of the genes: WIF1, SFRP4, WNT5A, CD82, and UCA1; no consistent abnormalities were found in the others. Ten out of eleven SNPs were previously recorded and one appears to be novel in CD82 since this SNP was not found in SNPdb. Of the nine known SNPs, four were synonymous mutations, two were 3' UTR variants, two were missense variants, and one was an intron variant (Table 10).

Table 10: Observed SNPs in CD82, SFRP4, UCA1, WIF1, and WNT5A. All genes of interest were analyzed by IGV for abnormalities in the transcripts, CD82, SFRP4, UCA1, WIF1, and WNT5A showed consistent SNPs throughout the samples. Nine out of the ten observed SNPs were previously recorded and found in dbSNP, one SNP in CD82 is novel at location chr11:44,619,242 with a frequency in 100% of the samples.

Gene	Observed SNP Location	Accession No.	SNP Name	Consequence	Cancer Replicates Containing SNP (/10)
CD82	*chr11:44,619,242*	*NA*	*CD82-N*		*10*
	chr11:44,619,090	rs7107335	CD82-S1	Synonymous	4
	chr11:44,618,718	rs11541053	CD82-M1	Missense	10
SFRP4	chr7:37,907,562	rs1802073	SFRP4-M1	Missense	7
	chr7:37,906,899	rs1052981	SFRP4-3	3'UTR	5
	chr7:37,912,124	rs1132553	SFRP4-S1	Synonymous	8
	chr7:37,914,238	rs1132552	SFRP4-S2	Synonymous	10
UCA1	chr19:15,834,304	rs7258210	UCA1-IV	Intron Variant	7
WIF1	chr12:65,120,486	rs7301320	WIF1-S1	Synonymous	4

WNT5A	chr3:55,466,974	rs669889	WNT5A-3	3'UTR	7

4.7 FuncPred SNP Results

FuncPred prediction software was used to determine the function of the SNPs. The resulting consequence of the SNP including changes in transcription factor binding sites (TFBS), splicing, microRNA binding sites, and possible phenotype (Table 11).

Table 11: FuncPred Results on Previously Known Observed SNPs. The nine known SNPs were analyzed by FuncPred for determination of possible consequences. UCA1-IV affects a transcription factor binding site (TFBS), CD82-M1, CD82-S1, SFRP4-S1, SFRP4-S2 affect a splicing mechanism, SFRP4-3 and WNT5A-3 is in a miRNA binding region, and SFRP4-M1 shows that it is possibly damaging but without a reason, and WIF1-S1 shows no adverse consequence.

SNP	rs	TFBS	Splicing	miRNA	nsSNP	Stop Codon	Polyphen	Reg Potential	Conservation
CD82-M1	rs11541053	--	Y	--	Y	--	Benign	0.494976	0.99
CD82-S1	rs7107335	--	Y	--	--	--	--	0.446336	0.94
UCA1-3	rs7258210	Y	--	--	--	--	--	NA	0
SFRP4-3	rs1052981	--	--	Y	--	--	--	0.032177	0.002
SFRP4-S1	rs1132552	--	Y	--	--	--	--	0.227926	0.801
SFRP4-S2	rs1132553	--	Y	--	--	--	--	0.350342	1
SFRP4-M1	rs1802073	--	--	--	Y	--	Possibly damaging	0.405909	0.028
WIF1-S1	rs7301320	--	--	--	--	--	--	0.294625	0.979
WNT5A-3	rs669889	--	--	Y	--	--	--	0.122852	0

4.8 Prediction of Possible SNP Consequence using DeepSEA

Using DeepSEA prediction software, the coordinates of the variant were entered along with the nucleotide change in a VCF file. A graph was generated comparing the possible chromatin features and log fold change values. The chromatin feature with the highest log fold change value was ZBTB7A.

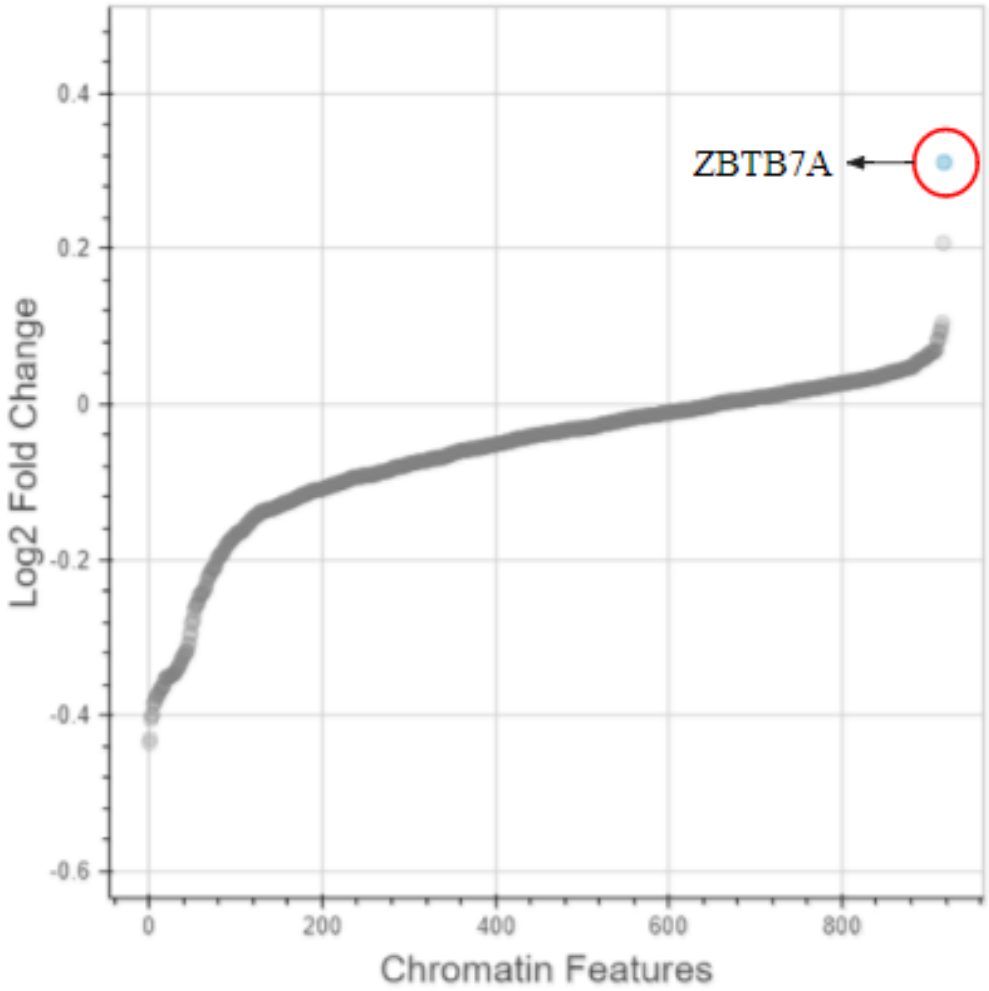


Figure 19: Prediction Graph for Potentially Novel CD82-N. Generated by DeepSEA, this graph represents all possible chromatin features that may bind to the site of the SNP. Ranking by log2 fold change, the blue dot on the far right and top most represents the chromatin feature with the highest fold change, ZBTB7A. ZBTB7A was predicted to have a log2 fold change value of 0.311.

4.9 CDKN2A and CDKN2B Downregulation

By DEG analysis and literature review, the TGFβ pathway has been shown to be implicated in colorectal carcinoma. In this dataset, the TGFβ pathway is shown to be upregulated by the increased fold change in the members of the pathway (Table 12).

Table 12: Differentially Expressed TGFβ Pathway Genes. Of the three known TGFβ receptors, only TGFβR1 is differentially expressed with a fold change of 0.575. The TGFβ inducer (TGFβI) is upregulated with a fold change of 1.532. miR-200 is upregulated with a fold change of 1.429 and its target, ZEB2 is repressed with a fold change of -1.131. CDKN2A and CDKN2B are downregulated. Most of the BMP proteins are downregulated with the exception of BMP7 which is upregulated. BMPER is a BMP inhibitor which may account for the BMP downregulation. Finally, All of the Smad proteins are downregulated and differentially expressed except for Smad2. Smad2 did not meet the cutoff stipulations to be considered as biologically and statistically significant.

TGFβ Pathway Genes	Fold Change	P Value	Significant?
TGFβR1	0.575	0.02455	Yes
TGFβR2	-	-	-
TGFβR3	-	-	-
TGFβI	1.532	5.00E-05	Yes
miR-200	1.429	0.0001	Yes
ZEB2	-1.131	5.00E-05	Yes
CDKN2A	-2.266	5.00E-05	Yes
CDKN2B	-2.266	5.00E-05	Yes
BMP2	-1.102	5.00E-05	Yes
BMP3	-2.355	5.00E-05	Yes
BMP5	-1.629	0.00105	Yes
BMP6	-1.801	0.0389	Yes
BMP7	1.296	0.00085	Yes
BMPER	2.11899	0.00105	Yes

SMAD2	-0.5799	0.0858	No
SMAD3	-0.5475	0.04255	Yes
SMAD4	-0.6879	0.01085	Yes
SMAD7	-0.8065	0.0032	Yes

The CDK inhibitors CDKN2A and CDKN2B are known as a tumor suppressor and is frequently deleted in cancers and colorectal cancer. In this dataset, CDKN2A and CDKN2B fold change was -2.266. The downstream targets are upregulated (Table 13) indicating a regulation mechanism to negatively regulate CDKN2A and CDKN2B.

Table 13: Differential Expression of the Downstream Targets of CDKN2A and CDKN2B. When upregulated, the downstream targets of CDKN2A and CDKN2B are downregulated. With the downregulation of CDKN2A and CDKN2B transcripts, it is expected for the downstream targets are upregulated which is demonstrated in this dataset.

Downstream Target	**Log2 Fold**
CDK4	1.183
CDK6	0.666
E2F1	1.898
E2F3	1.170
CCND1	1.429

CHAPTER 5: DISCUSSION

5.1 The Wnt Pathway has the Greatest Number of Differentially Expressed Genes

The Wnt pathway can be divided into three subtypes: Wnt/β-catenin or canonical pathway, non-canonical planar cell polarity pathway (PCP), and non-canonical Wnt/Calcium pathways [18,20,32]. These subtypes promote cell proliferation, cell polarity, and cell adhesion and migration, respectively. The Wnt pathway is upregulated in several cancers including colorectal, breast, liver, gastric cancers among others.

The canonical pathway also referred to as the Wnt/β-catenin pathway is dependent on the accumulation β-catenin in the nucleus which activates the transcription factor complex, TCF/LEF [18,20,32]. This complex promotes the transcription of the Wnt target genes: MYC, CCND1, MMP7, AXIN2, VEGF, MET, and others which were upregulated in this dataset. These genes promote the cell cycle progression and proliferation.

The PCP pathway promotes cell polarity and cytoskeletal rearrangement [18,20,32]. It is β-catenin independent and is thus categorized as a non-canonical pathway. This pathway activates ROCK2 which is responsible for cell adhesion and cytoskeletal rearrangement. Additionally, this pathway also targets the JNK pathway. ROCK2 targets and phosphorylates several genes including ADD1, BRCA2, CNN1, EZR, and others in which some of these targets are differentially expressed. The JNK pathway genes include MAPK8, MAPK9, and MAPK10 [94]. This pathway plays an essential role in apoptosis and cell proliferation. The target genes of this pathway include MMP3, CCND1, JUNB, FOSL1, FOSL2, FOSB, cFOS, and ATF2 [18].

The Wnt/Calcium pathway promotes cell adhesion and migration by targeting NFAT, NEMO, and CDC42 [20,25,32]. NFAT is activated through the activation of PLC which allows for

the influx of calcium. Calcium activates calmodulin which in turn activates NFAT. NFAT is a family of transcription factors which regulate cell survival, cell proliferation, migration, invasion, and angiogenesis [95]. NEMO which inhibits β-catenin/TCF function. CDC42 is a GTPase and is responsible for tissue separation which in turn promotes cell migration [20].

Pathway enrichment showed that the Wnt pathway was the most enriched gene set in all of the DEGs. In the downregulated pathway enrichment results, the Wnt pathway genes was ranked number five; the pathway with the greatest number of genes expressed is cytokine and chemokine mediated inflammation or the NFκB pathway. Previous studies showed a crosstalk between the Wnt and NFκB signaling pathways [96–99]. Furthermore, the relationship between the Wnt pathway and the NFκB pathway is shown to be highly dependent on cellular context. In the dataset, it appears that the Wnt pathway is negatively regulating the NFκB pathway. In the dataset, the Wnt pathway represents the most genes in the downregulated gene set. It has been shown that the Wnt pathway can negatively regulate the NFκB pathway through the accumulation of β-catenin. β-catenin forms a complex with RelA and p50 which is responsible for facilitating DNA binding with NFκB. In turn, NFκB is unable to transcribe its target genes. This relationship is consistent with research done on colon cancer cell lines [97–100].

5.2 Lack of Epigenetic Silencing of WIF1 and CD82 by HOTAIR

When lncRNA2Target was used, the targets of HOTAIR were found and crossed referenced with the DEG list. Of the differentially expressed targets, all showed the expected expression pattern when HOTAIR is expressed except for WIF1 and CD82. WIF1 and CD82 are expected to be downregulated in the presence of HOTAIR; however, in this dataset WIF1 and CD82 are upregulated at 5.165 and 1.05, respectively.

HOTAIR epigenetically silences its targets through the binding, recruitment, and guiding

of Polycomb repressive complex 2 (PRC2) [59,101,102]. The catalytically active subunit of the PRC2 is differentially expressed; however, the two remaining subunits are not differentially expressed. They are, however, expressed in both conditions.

A recent study showed that HOTAIR is able to work independently of the PRC2 in a breast cancer cell line [101]. By the use of forceful overexpression of HOTAIR, subtle transcriptomic changes were observed. Additionally, the study found that by the use of artificial tethering of HOTAIR to chromatin, it showed transcriptional repression was observed but not with the use of the PRC2.

WIF1 is a proliferation inhibitor and CD82 is a metastasis inhibitor [17,25,28,44,45]. When WIF1 and CD82 are expressed, the expected phenotypes are that of normal tissue homeostatic growth. These two genes, WIF1 and CD82, were found to carry SNPs. WIF1 has a single synonymous SNP in 40% of the replicates while CD82 has three SNPs, one novel (CD82-N), one synonymous (CD82-S1), and one missense (CD82-M1) in 10/10, 4/10, and 10/10 of the replicates, respectively. According to FuncPred, SNPs CD82-S1 and CD82-M1 of CD82 cause a change in splicing site that causes either a splice site enhancer or silencer region. Additionally, CD82-M1 is a nonsynonymous SNP and by FuncPred, it is also a benign mutation. These SNPs may be the cause of which may play a role in its inability to be silenced by HOTAIR.

Whether HOTAIR utilizes the PRC2 to silence its targets or not, it would be expected with such a high expression of HOTAIR that WIF1 and CD82 would be silenced, however, the SNPs in WIF1 and CD82 may be preventing their silencing.

5.3 Functional Prediction of Potentially Novel SNP in CD82

The potentially novel SNP, CD82-N (location: chr11:44,619,242) was predicted to create a binding site with ZBTB7A. ZBTB7A or Pokemon has two domains, a zinc finger at the C

terminus and a BTB domain at the N terminus [103,104]. The zinc finger is used for DNA recognition and the BTB domain is used for homo- or heterodimerization to interact with proteins. ZBTB7A is known to be deleted or mutated in a large subset of cancers as well as amplified in small cell lung cancer [104–106]. It is a known oncogene and is involved in adipogenesis, osteogenesis, and oncogenesis. The primary functions are to act as a transcription factor to suppress glycolysis through repressing genes in the glycolytic cycle and to suppress p53 expression through the repression of ARF. Additionally, ZBTB7A has also been reported to regulate transcription, cytoskeletal dynamics, ion channel assembly and gating, and protein targeting for ubiquitination.

ZBTB7A was not differentially expressed in this dataset, however, this may be due to any loss of RNA during the RNA extraction process and/or filtering low quality reads. If ZBTB7A were to be expressed in the cellular context studied, then ZBTB7A may be repressing CD82 function through this SNP. If so, then metastasis would not be suppressed and CRC would be able to progress and metastasize to the liver as known by the patient histories. Since it was not differentially expressed in this dataset, further investigation is required to determine binding and role.

5.4 Interplay Between SFRP4 and CRNDE

By the use of lncRNA2Target, it was found that lncRNA CRNDE negatively affects SFRP4, the mechanism by which CRNDE accomplishes this is unknown. In this dataset, CRNDE and SFRP4 are expressed at 3.228 and 2.121, respectively, clearly indicating another factor disrupting this relationship. By IGV analysis, it was found that SFRP4 carries four SNPs in the cancer samples. SFRP4-S1 and SFRP4-S2 were synonymous (rs1132553, rs1132552), SFRP4-M1 was nonsynonymous (rs1802073), and SFRP4-3 was in the 3’UTR (rs1052981).

SFRP4-S1 and SFRP4-S2 were found to cause changes in splicing site that may cause a splice site enhancer or silencer, SFRP4-M1 was found to be possible damaging but the mechanism of how this could be is not elucidated, and SFRP4-3 was found to have a miRNA binding site. By this analysis, it can be assumed that the SNPs play a role in the inhibitory function of SFRP4, its regulation by CRNDE and/or miRNAs, and splicing in which contributes to the tumorigenesis of CRC.

5.5 The Interplay Between WIF1 and UCA1 to Inhibit or Activate WNT5A

Since, WIF1-S does not cause any structural abnormalities that would inhibit the function and there are no known post-translational modifications that would degrade WIF1, WIF1 in theory would still be active to inhibit its Wnt ligand targets. The lncRNA2Target database, showed that lncRNA UCA1 positively regulates WNT5A (Table 6). A possible defense mechanism of cancer would be to upregulate a competitor of WIF1 to ultimately activate the Wnt pathway. The mechanism by which UCA1 positively regulates WNT5A is unknown; however, it is possible that WIF1 and UCA1 may compete with each other to inhibit or activate WNT5A, respectively (Figure 20). This mechanism would explain why although WIF1 is highly expressed, the Wnt pathway is able to transcribe and activate its targets.

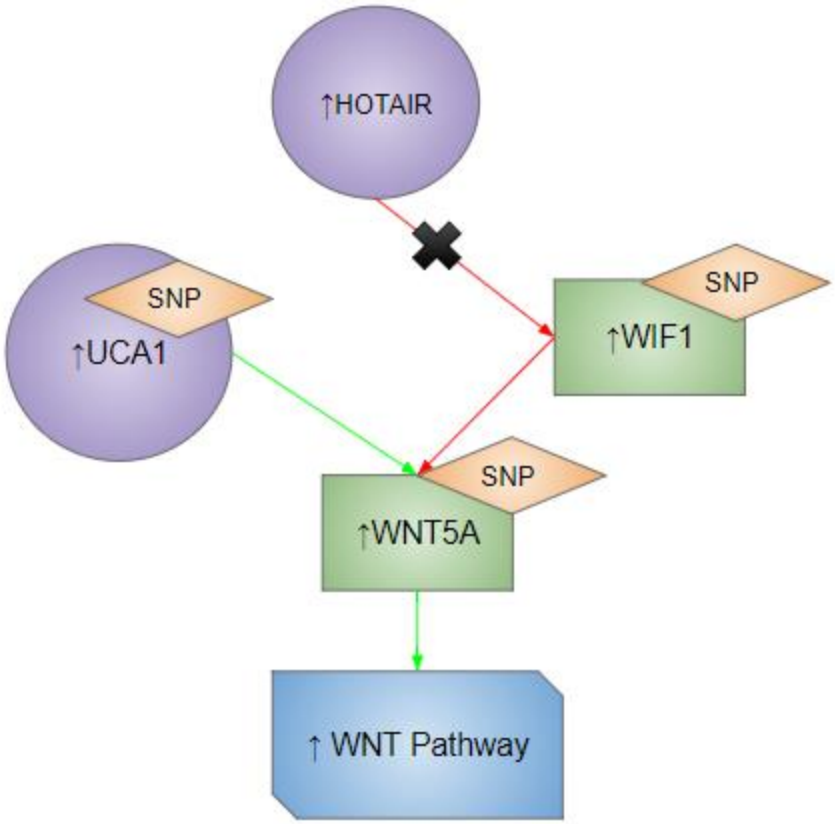


Figure 20. Possible Competition Mechanism between WIF1 and UCA1 to Inhibit or Activate WNT5A. With the upregulation of HOTAIR and its inability to silence WIF1, it is possible that there may be a mutation in the WIF1 promoter that is preventing the recognition of WIF1 by HOTAIR, which is observed in this dataset. This would subsequently cause an upregulation of WIF1 which would be able to inhibit its targets, WNT5A. WNT5A and its associated lncRNA UCA1 are upregulated. There may be a competition between WIF1 and WNT5A associated lncRNA, UCA1. UCA1 is known to positively regulate WNT5A. So, with the upregulation of HOTAIR, WIF1, WNT5A, UCA1, and the Wnt pathway as a whole, UCA1 may be competing with WIF1 to positively regulate WNT5A and subsequently the Wnt pathway.

5.6 The Interplay Between WIF1 and CRNDE to Inhibit or Activate WNT2 and WNT3

For the WNT2 and WNT3 ligands, it was found using lncRNA2Target that lncRNA CRNDE positively regulates WNT2 and WNT3; however, the mechanism of which CRNDE positively regulates WNT2 and WNT3 is unknown. Because the WIF1 protein would remain unchanged, it is possible that CRNDE would also compete with WIF1 to activate or inhibit

WNT2 and WNT3 (Figure 21). This mechanism would explain why the Wnt pathway is upregulated. Figure 22 shows a complete overview with all of the genes involved with SNPs denoted.

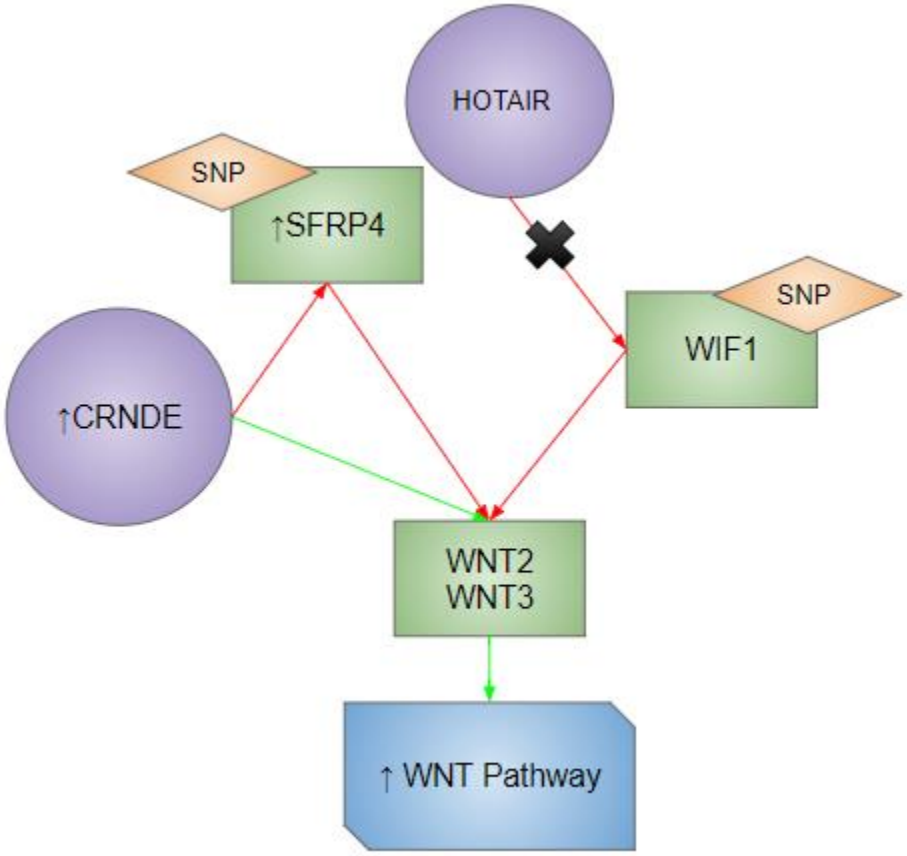


Figure 21: Possible Competition Mechanism between WIF1 and CRNDE to Inhibit or Activate WNT2 and WNT3. With the upregulation of HOTAIR and its inability to silence WIF1, it is possible that there may be a mutation in the WIF1 promoter that is preventing the recognition of WIF1 by HOTAIR, which is observed in this dataset. This would subsequently cause an upregulation of WIF1 which would be able to inhibit its targets, WNT2 and WNT3. WNT2 and WNT3 and its associated lncRNA CRNDE are upregulated. There may be a competition between WIF1 and CRNDE. CRNDE is known to positively regulate WNT2 and WNT3. So, with the upregulation of HOTAIR, WIF1, WNT2, WNT3, CRNDE, and the Wnt pathway as a whole, CRNDE may be competing with WIF1 to positively regulate WNT2 and WNT3 and subsequently the Wnt pathway.

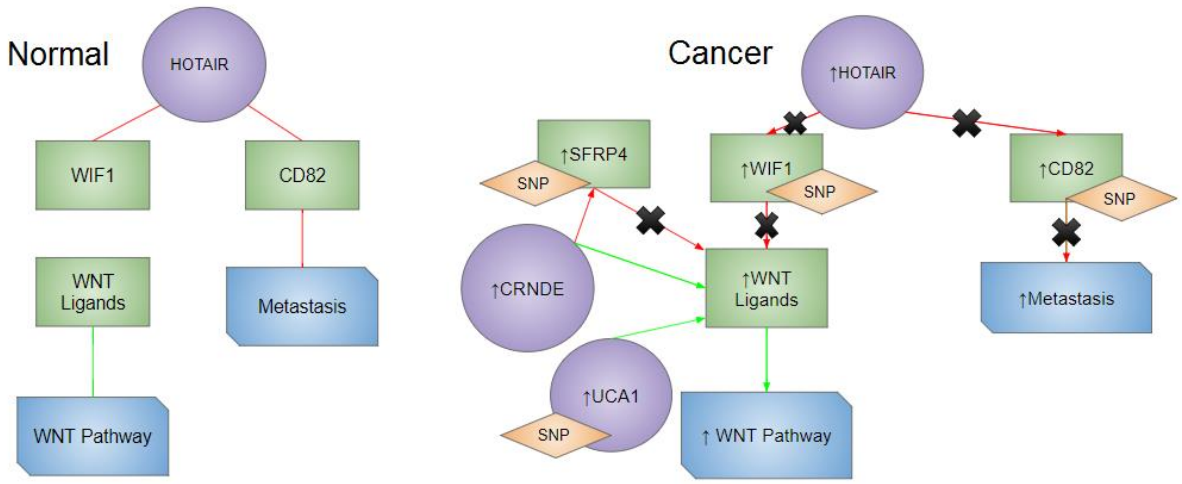


Figure 22: Comparison of Pathway Outcomes between Normal and Cancer Conditions. On the left is a schematic of events when HOTAIR is expressed. When HOTAIR is expressed, WIF1 and CD82 are silenced thus the Wnt Pathway and metastatic pathways are active. In this case, on the right, when HOTAIR is expressed, the SNPs in WIF1 and CD82 may be preventing HOTAIR from silencing. WIF1 and SFRP4 are competing with CRNDE and UCA1 to inhibit or activate the WNT ligands and in this case, CRNDE and UCA1 are ultimately activating the WNT ligands for the Wnt pathway to be active. Additionally, CRNDE may be negatively regulating SFRP4 which may be a contributing factor to the inability of SFRP4 to inhibit the WNT ligands.

5.7 CDKN2A and CDKN2B Methylation in South Korean Colorectal Cancer Patients

A study was conducted to assess the methylation patterns of Finnish and Egyptian colorectal cancer patients [107]. The study showed that there are many features of CRC that are universal across both ethnic types such as the dysregulation of p53. The study also found that there are some features of CRC that are specific to the ethnic origin of the patients. For example, microsatellite instability was highly prevalent in Egyptian CRC patients compared to Finnish patients. This study also found that previous research conducted that focused on CDKN2B was mainly in Western patient cohorts residing in Europe of the United States, and that there was no

strong correlation between the methylation of CDKN2B and colorectal cancer. The methylation of CDKN2B was found to be a distinct feature of the Egyptian population of CRC. In another study done on a Chinese cohort, CDKN2B methylation was found at a rate of 68% [108]. The study hypothesized that the difference may be due to genetic and environmental predisposition of the Egyptian population.

The downregulation of CDKN2A and CDKN2B was observed. It is likely that this downregulation is due to methylation, as this is a common consequence in cancer. This methylation pattern may be specific to Eastern populations as seen in studies in Egyptian and Chinese cohorts. Since the dataset studied is derived from South Korean patients, it is likely that the environmental factors and genetics of the South Koreans may have contributed to the methylation of CDKN2A and CDKN2B.

CHAPTER 6: CONCLUSIONS AND FUTURE PERSPECTIVES

6.1 Conclusions

Using a fold change cutoff of greater than 0.5 and less than -0.5 and a p value cut off of less than 0.05, there are 5,303 differentially expressed genes in South Korean male patients with stage IV colorectal cancer. The Wnt pathway is known to have aberrant expression in colorectal cancer [18,20,25,32]. The Wnt pathway can be divided into three subtypes: Wnt/β-catenin or canonical pathway, non-canonical planar cell polarity pathway, and non-canonical Wnt/Calcium pathways. These subtypes regulate cell proliferation, cell polarity, and cell adhesion and migration, respectively.

Among the upregulated Wnt pathway inhibitors were Wnt Inhibitory Factor 1 (WIF1) and Secreted Frizzled Related Protein 4 (SFRP4). WIF1 is known to bind to the Wnt ligands and inhibits its binding to the Frizzled and Frizzled/LRP cell surface receptors [18,20,25,32]. SFRP4 is known to either bind with the WNT ligands or Frizzled to prevent the downstream cascade and activation of the Wnt Pathway. WIF1 has a fold change value of 5.165 and SFRP4 has a fold expression of 2.121. With an expression this significantly upregulated, it would be expected that the Wnt pathway and the target genes would be downregulated; however, the pathway genes and their targets are also upregulated.

HOTAIR is known to epigenetically affect the expression of many genes utilizing PRC2 and LSD1 to methylate and de-methylate targets, respectively [50,91,101,102]. All of the differentially expressed targets of HOTAIR were expressed as expected when there is a high expression of HOTAIR except for WIF1 and CD82. HOTAIR is known to silence WIF1 and CD82; however, these genes were upregulated in this dataset.

The lack of silencing by HOTAIR may be due to the SNPs observed in WIF1 and CD82. These SNPs may be preventing HOTAIR and the Polycomb repressive complex 2 (PRC2) from recognizing and silencing WIF1 and CD82.

Urothelial cancer associated 1 (UCA1) is a lncRNA where its expression has been correlated with poor prognosis in CRC, increased chemoresistance in urothelial cancer, and promotion of proliferation, metastasis, and tumorigenesis through various pathways including the Wnt signaling pathways [62–67,70]. LncRNA UCA1 was found to positively regulate WNT5A. So, UCA1 may be competing with WIF1 to ultimately activate WNT5A. The mechanism by which UCA1 positively regulates WNT5A is unknown. The SNP found in UCA1, UCA1-3 is in a transcription factor binding site which may enhance UCA1 effectiveness; the exact mechanism and clinical consequence is unknown. The SNP found in WNT5A, WNT5A-3 is in the 3'UTR and was found to alter a microRNA binding site. Because miRNAs target genes for degradation, this specific alteration may be preventing the miRNAs from recognizing WNT5A for degradation.

Colorectal Neoplasia Differentially Expressed (CRNDE) is a lncRNA where its expression is associated with poor prognosis, chemoresistance, cell proliferation, tumorigenesis, and stemness [54,72–74,109]. Along the same terms, lncRNA CRNDE has been found to positively regulate WNT2 and WNT3. A similar mechanism may also be the case; CRNDE may be competing with WIF1 to ultimately activate WNT2 and WNT3. There were no SNPs or abnormalities in these genes and the mechanism of how CRNDE positively regulates WNT2 and WNT3 is unknown.

The Wnt pathway is ultimately upregulated shown by the upregulation in the Wnt

pathway proteins and its targets despite the upregulation of its inhibitors. This incongruity may be due to a competition between WIF1 and CRNDE and UCA1 to positively regulate the WNT ligands. The SNPs in CD82 may be hindering its function as a tumor metastasis suppressor and its ability to be silenced by HOTAIR. The potentially novel SNP, CD82-N, was predicted to create a binding site for ZBTB7A which would repress CD82 function and, in turn, allow for metastasis to be uninhibited by CD82. Additionally, it seems that UCA1 and CRNDE are ultimately activating WNT5A, WNT2, and WNT3 in order for the Wnt pathway to proceed.

The limitations of this study are the small sample size with only 5 matched normal and cancer datasets, all of the sample datasets were from South Korean males, and the limited computational and memory capacity available. Ideally, all 18 available sample datasets would be analyzed and compared with Eastern (ie Egyptian and Chinese) and Western cohorts to validate these results.

6.2 Future Perspectives

Because of the small sample size of this study, future studies are headed to confirm the mutations in the WIF1, SFRP4, WNT5A, UCA1, and CD82 by DNA sequencing of South Korean male stage IV colorectal patients. Additionally, functional assays should be conducted to determine the efficiency of the Wnt pathway in the presence of WIF1, SFRP4, UCA1, and CRNDE. Although these players are highly expressed at the RNA level, there may be post-transcriptional and post-translational modifications that are not shown by this analysis. Little is known about the mechanism of action of lncRNAs UCA1 and CRNDE. Further investigation should delve into the functions of these lncRNAs in order to understand their precise role in the Wnt pathway. Once the roles for UCA1 and CRNDE are elucidated, they may be potential therapeutic or diagnostic targets. Inhibiting UCA1 and CRNDE may allow for WIF1 to

effectively inhibit the Wnt pathways, this would decrease cell proliferation, cell polarity, and migration. Lastly, more investigation is needed for the mechanisms of the remaining differentially expressed inhibitors found in this dataset: DKK2, DKK4, APCDD1, and Notum. The full elucidation of the mechanisms and pathways of the inhibitors in the Wnt pathway would be a powerful tool against stage IV colorectal cancer. This information would be vital for the development of therapeutic targets.

It is clear that the Wnt pathway has a complex mechanism for regulation. The genes discussed may be candidates for therapeutic targets once the detailed mechanism of action is elucidated. LncRNAs are multifaceted and can regulate genes on all levels, so it is important to understand the key lncRNAs in the Wnt pathway and in all stages of colorectal cancer in order to develop personalized and efficient therapy. Rapid screening, diagnosis, and point of care tests are becoming a necessity. Once the mechanisms of actions of UCA1 and CRNDE are elucidated and the discrepancy of the remaining inhibitors is explained, the serum, saliva, and/or urine levels of these players and their correlation to the stages of CRC may lead to rapid screening and diagnosis. With the knowledge of the interplay of lncRNAs and their targets, noninvasive early detection of CRC is in the near future.

REFERENCES

1. Colorectal Cancer. https://www.cancer.org/cancer/colon-rectal-cancer.html. Accessed December 12, 2017.
2. Larue L, Bellacosa A. Epithelial–mesenchymal transition in development and cancer: role of phosphatidylinositol 3′ kinase/AKT pathways. *Oncogene*. 2005;24(50):7443-7454. doi:10.1038/sj.onc.1209091.
3. Hajjari M, Salavaty A. HOTAIR: an oncogenic long non-coding RNA in different cancers. *Cancer Biol Med*. 2015;12(1):1-9. doi:10.7497/j.issn.2095-3941.2015.0006.
4. Colon cancer - Symptoms and causes - Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/colon-cancer/symptoms-causes/syc-20353669. Accessed December 14, 2017.
5. Colon cancer - Diagnosis and treatment - Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/colon-cancer/diagnosis-treatment/drc-20353674. Accessed December 14, 2017.
6. Coppedè F, Lopomo A, Spisni R, Migliore L. Genetic and epigenetic biomarkers for diagnosis, prognosis and treatment of colorectal cancer. *World J Gastroenterol*. 2014;20(4):943-956. doi:10.3748/wjg.v20.i4.943.
7. Lech G, Słotwiński R, Słodkowski M, Krasnodębski IW. Colorectal cancer tumour markers and biomarkers: Recent therapeutic advances. *World J Gastroenterol*. 2016;22(5):1745. doi:10.3748/wjg.v22.i5.1745.
8. Treatment of Colon Cancer by Stage. https://www.cancer.org/cancer/colon-rectal-cancer/treating/by-stage-colon.html. Accessed December 14, 2017.
9. Advances in Colorectal Cancer Research | National Institutes of Health (NIH).

https://www.nih.gov/research-training/advances-colorectal-cancer-research. Accessed December 14, 2017.

10. Gustavsson B, Carlsson G, Machover D, et al. A Review of the Evolution of Systemic Chemotherapy in the Management of Colorectal Cancer. *Clin Colorectal Cancer*. 2015;14(1):1-10. doi:https://doi.org/10.1016/j.clcc.2014.11.002.
11. Treatment | Science Of CRC. https://www.scienceofcrc.org/treat/. Accessed December 14, 2017.
12. Radiation Therapy for Cancer - National Cancer Institute. https://www.cancer.gov/about-cancer/treatment/types/radiation-therapy/radiation-fact-sheet. Accessed December 14, 2017.
13. Mirnezami R, Chang GJ, Das P, et al. Intraoperative radiotherapy in colorectal cancer: Systematic review and meta-analysis of techniques, long-term outcomes, and complications. *Surg Oncol*. 2013;22(1):22-35. doi:https://doi.org/10.1016/j.suronc.2012.11.001.
14. Florea LD, Salzberg SL. Genome-guided transcriptome assembly in the age of next-generation sequencing. *IEEE/ACM Trans Comput Biol Bioinforma*. 2013;10(5):1234-1240. http://www.ncbi.nlm.nih.gov/pubmed/24524156. Accessed December 10, 2017.
15. Mantione KJ, Kream RM, Kuzelova H, et al. Comparing bioinformatic gene expression profiling methods: microarray and RNA-Seq. *Med Sci Monit Basic Res*. 2014;20:138-142. doi:10.12659/MSMBR.892101.
16. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol*. 2011;12(3):R22. doi:10.1186/gb-2011-12-3-r22.

17. Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. *Oncogene*. 2017;36(11):1461-1473. doi:10.1038/onc.2016.304.

18. Pongracz JE, Stockley RA. Wnt signalling in lung development and diseases. *Respir Res*. 2006;7(1):15. doi:10.1186/1465-9921-7-15.

19. Tabatabai R, Linhares Y, Bolos D, Mita M, Mita A. Targeting the Wnt Pathway in Cancer: A Review of Novel Therapeutics. *Target Oncol*. 2017;12(5):623-641. doi:10.1007/s11523-017-0507-4.

20. Komiya Y, Habas R. Wnt signal transduction pathways. *Organogenesis*. 2008;4(2):68-75. http://www.ncbi.nlm.nih.gov/pubmed/19279717. Accessed December 10, 2017.

21. Novellasdemunt L, Antas P, Li VSW. Targeting Wnt signaling in colorectal cancer. A Review in the Theme: Cell Signaling: Proteins, Pathways and Mechanisms. *Am J Physiol - Cell Physiol*. 2015;309(8):C511-C521. doi:10.1152/ajpcell.00117.2015.

22. Jansson L, Kim GS, Cheng AG. Making sense of Wnt signaling—linking hair cell regeneration to development. *Front Cell Neurosci*. 2015;9:66. doi:10.3389/fncel.2015.00066.

23. Tanaka N, Mashima T, Mizutani A, et al. APC Mutations as a Potential Biomarker for Sensitivity to Tankyrase Inhibitors in Colorectal Cancer. *Mol Cancer Ther*. 2017;16(4):752-762. doi:10.1158/1535-7163.MCT-16-0578.

24. Ramachandran I, Ganapathy V, Gillies E, et al. Wnt inhibitory factor 1 suppresses cancer stemness and induces cellular senescence. *Cell Death Dis*. 2014;5(5):e1246. doi:10.1038/cddis.2014.219.

25. Najdi R, Holcombe RF, Waterman ML. Wnt signaling and colon carcinogenesis: beyond APC. *J Carcinog*. 2011;10:5. doi:10.4103/1477-3163.78111.

26. Mazzoni SM, Fearon ER. AXIN1 and AXIN2 variants in gastrointestinal cancers. *Cancer Lett*. 2014;355(1):1-8. doi:https://doi.org/10.1016/j.canlet.2014.09.018.

27. Picco G, Petti C, Centonze A, et al. Loss of AXIN1 drives acquired resistance to WNT pathway blockade in colorectal cancer cells carrying RSPO3 fusions. *EMBO Mol Med*. 2017;9(3):293-303. doi:10.15252/emmm.201606773.

28. Roperch J-P, Incitti R, Forbin S, et al. Aberrant methylation of NPY, PENK, and WIF1 as a promising marker for blood-based diagnosis of colorectal cancer. *BMC Cancer*. 2013;13(1):566. doi:10.1186/1471-2407-13-566.

29. Abdelmaksoud-Dammak R, Miladi-Abdennadher I, Saadallah-Kallel A, et al. Downregulation of WIF-1 and Wnt5a in patients with colorectal carcinoma: clinical significance. *Tumor Biol*. 2014;35(8):7975-7982. doi:10.1007/s13277-014-2015-9.

30. Wei C, Peng B, Han Y, et al. Mutations of HNRNPA0 and WIF1 predispose members of a large family to multiple cancers. *Fam Cancer*. 2015;14(2):297-306. doi:10.1007/s10689-014-9758-8.

31. Jones MC, Topol SE, Rueda M, et al. Mutation of WIF1: a potential novel cause of a Nail-Patella-like disorder. *Genet Med*. 2017;19:1179. http://dx.doi.org/10.1038/gim.2017.20.

32. Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. *Oncogene*. 2017;36(11):1461-1473. doi:10.1038/onc.2016.304.

33. Huang S, Zhong X, Gao J, et al. Coexpression of SFRP1 and WIF1 as a prognostic predictor of favorable outcomes in patients with colorectal carcinoma. *Biomed Res Int*. 2014;2014:256723. doi:10.1155/2014/256723.

34. Jiang Q, Wang J, Wu X, et al. LncRNA2Target: a database for differentially expressed genes after lncRNA knockdown or overexpression. *Nucleic Acids Res*. 2015;43(Database

issue):D193-6. doi:10.1093/nar/gku1173.

35. Guo X, Wang X-F. Signaling cross-talk between TGF-β/BMP and other pathways. *Cell Res*. 2008;19:71. http://dx.doi.org/10.1038/cr.2008.302.

36. Drabsch Y, He S, Zhang L, Snaar-Jagalska BE, ten Dijke P. Transforming growth factor-β signalling controls human breast cancer metastasis in a zebrafish xenograft model. *Breast Cancer Res*. 2013;15(6):R106. doi:10.1186/bcr3573.

37. Finnson KW, McLean S, Di Guglielmo GM, Philip A. Dynamics of Transforming Growth Factor Beta Signaling in Wound Healing and Scarring. *Adv Wound Care*. 2013;2(5):195-214. doi:10.1089/wound.2013.0429.

38. Giannelli G, Villa E, Lahn M. Transforming growth factor-β as a therapeutic target in hepatocellular carcinoma. *Cancer Res*. 2014;74(7):1890-1894. doi:10.1158/0008-5472.CAN-14-0243.

39. Lebrun J-J. The Dual Role of TGF β in Human Cancer: From Tumor Suppression to Cancer Metastasis. *ISRN Mol Biol*. 2012;2012:1-28. doi:10.5402/2012/381428.

40. Zhuang J, Lu Q, Shen B, et al. TGFβ1 secreted by cancer-associated fibroblasts induces epithelial-mesenchymal transition of bladder cancer cells through lncRNA-ZEB2NAT. *Sci Rep*. 2015;5(1):11924. doi:10.1038/srep11924.

41. Zhu J, Liang C, Hua Y, et al. The metastasis suppressor CD82/KAI1 regulates cell migration and invasion via inhibiting TGF-β 1/Smad signaling in renal cell carcinoma. *Oncotarget*. 2017;8(31):51559-51568. doi:10.18632/oncotarget.18086.

42. Tsai YC, Weissman AM. Dissecting the diverse functions of the metastasis suppressor CD82/KAI1. *FEBS Lett*. 2011;585(20):3166-3173. doi:10.1016/J.FEBSLET.2011.08.031.

43. Dong JT, Lamb PW, Rinker-Schaeffer CW, et al. KAI1, a metastasis suppressor gene for

prostate cancer on human chromosome 11p11.2. *Science*. 1995;268(5212):884-886. http://www.ncbi.nlm.nih.gov/pubmed/7754374. Accessed December 12, 2017.

44. Challa R, Shoji Y, Conrads KA, et al. Metastasis suppressor CD82/KAI1 expression is dependent on functional SWI/SNF ARID1A/B. *Cancer Res*. 2014;74(19 Supplement):990-990. doi:10.1158/1538-7445.AM2014-990.

45. Aram R, Dotan I, Hotz-Wagenblatt A, Canaani D. Identification of a novel metastasis inducing lncRNA which suppresses the KAI1/CD82 metastasis suppressor gene and is upregulated in triple-negative breast cancer. *Oncotarget*. 2017;8(40):67538-67552. doi:10.18632/oncotarget.18733.

46. Marreiros A, Dudgeon K, Dao V, et al. KAI1 promoter activity is dependent on p53, junB and AP2: evidence for a possible mechanism underlying loss of KAI1 expression in cancer cells. *Oncogene*. 2005;24(4):637-649. doi:10.1038/sj.onc.1208216.

47. Tsai YC, Weissman AM. Dissecting the diverse functions of the metastasis suppressor CD82/KAI1. *FEBS Lett*. 2011;585(20):3166-3173. doi:10.1016/j.febslet.2011.08.031.

48. Mattick JS, Rinn JL. Discovery and annotation of long noncoding RNAs. *Nat Struct Mol Biol*. 2015;22(1):5-7. doi:10.1038/nsmb.2942.

49. Xie X, Tang B, Xiao Y-F, et al. Long non-coding RNAs in colorectal cancer. *Oncotarget*. 2016;7(5):5226-5239. doi:10.18632/oncotarget.6446.

50. Guerrieri F. Long non-coding RNAs era in liver cancer. *World J Hepatol*. 2015;7(16):1971. doi:10.4254/wjh.v7.i16.1971.

51. Yang G, Lu X, Yuan L. LncRNA: A link between RNA and cancer. *Biochim Biophys Acta - Gene Regul Mech*. 2014;1839(11):1097-1109. doi:10.1016/j.bbagrm.2014.08.012.

52. Ma L, Bajic VB, Zhang Z. On the classification of long non-coding RNAs. *RNA Biol*.

2013;10(6):924-933. doi:10.4161/rna.24604.

53. Wang Z, Li X. The role of noncoding RNA in hepatocellular carcinoma. *Gland Surg*. 2013;2(1). http://gs.amegroups.com/article/view/1596/2218. Accessed December 15, 2017.

54. Ellis BC, Molloy PL, Graham LD. CRNDE: A Long Non-Coding RNA Involved in CanceR, Neurobiology, and DEvelopment. *Front Genet*. 2012;3:270. doi:10.3389/fgene.2012.00270.

55. Liz J, Esteller M. lncRNAs and microRNAs with a role in cancer development. *Biochim Biophys Acta - Gene Regul Mech*. 2016;1859(1):169-176. doi:https://doi.org/10.1016/j.bbagrm.2015.06.015.

56. Kung JTY, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. *Genetics*. 2013;193(3):651-669. doi:10.1534/genetics.112.146704.

57. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cell*. 2011;43(6):904-914. doi:10.1016/j.molcel.2011.08.018.

58. Pontier DB, Gribnau J. Xist regulation and function explored. *Hum Genet*. 2011;130(2):223-236. doi:10.1007/s00439-011-1008-7.

59. Cai B, Song XQ, Cai JP, Zhang S. HOTAIR: a cancer-related long non-coding RNA. *Neoplasma*. 2014;61(4):379-391. http://www.ncbi.nlm.nih.gov/pubmed/25027739. Accessed December 12, 2017.

60. Grammatikakis I, Panda AC, Abdelmohsen K, Gorospe M. Long noncoding RNAs (lncRNAs) and the molecular hallmarks of aging. *Aging (Albany NY)*. 2014;6(12):992-1009. doi:10.18632/aging.100710.

61. Rinn JL, Chang HY. Genome Regulation by Long Noncoding RNAs. *Annu Rev Biochem*.

2012;81(1):145-166. doi:10.1146/annurev-biochem-051410-092902.

62. Wang H, Guan Z, He K, Qian J, Cao J, Teng L. LncRNA UCA1 in anti-cancer drug resistance. *Oncotarget*. 2017;8(38):64638-64650. doi:10.18632/oncotarget.18344.

63. Han Y, Yang Y, Yuan H, et al. UCA1, a long non-coding RNA up-regulated in colorectal cancer influences cell proliferation, apoptosis and cell cycle distribution. *Pathology*. 2014;46(5):396-401. doi:10.1097/PAT.0000000000000125.

64. Li X. Long non-coding RNA UCA1a(CUDR) promotes proliferation and tumorigenesis of bladder cancer. *Int J Oncol*. April 2012. doi:10.3892/ijo.2012.1443.

65. Yang Y-T, Wang Y-F, Lai J-Y, et al. Long non-coding RNA UCA1 contributes to the progression of oral squamous cell carcinoma by regulating the WNT/β-catenin signaling pathway. *Cancer Sci*. 2016;107(11):1581-1589. doi:10.1111/cas.13058.

66. Fan Y, Shen B, Tan M, et al. Long non-coding RNA UCA1 increases chemoresistance of bladder cancer cells by regulating Wnt signaling. *FEBS J*. 2014;281(7):1750-1758. doi:10.1111/febs.12737.

67. Pan J, Li X, Wu W, et al. Long non-coding RNA UCA1 promotes cisplatin/gemcitabine resistance through CREB modulating miR-196a-5p in bladder cancer cells. *Cancer Lett*. 2016;382(1):64-76. doi:10.1016/j.canlet.2016.08.015.

68. Fotouhi Ghiam A, Taeb S, Huang X, et al. Long non-coding RNA urothelial carcinoma associated 1 (UCA1) mediates radiation response in prostate cancer. *Oncotarget*. 2017;8(3):4668-4689. doi:10.18632/oncotarget.13576.

69. NI B, YU X, GUO X, et al. Increased urothelial cancer associated 1 is associated with tumor proliferation and metastasis and predicts poor prognosis in colorectal cancer. *Int J Oncol*. 2015;47(4):1329-1338. doi:10.3892/ijo.2015.3109.

70. Jiang H, Chen Y-T, Fu X-G. Tissue expression level of lncRNA UCA1 is a prognostic biomarker for colorectal cancer. *Int J Clin Exp Pathol*. 2016;9(4):4241-4246. www.ijcep.com. Accessed December 15, 2017.

71. Wang F, Ni H, Sun F, Li M, Chen L. Overexpression of lncRNA AFAP1-AS1 correlates with poor prognosis and promotes tumorigenesis in colorectal cancer. *Biomed Pharmacother*. 2016;81(Supplement C):152-159. doi:https://doi.org/10.1016/j.biopha.2016.04.009.

72. Ding J, Li J, Wang H, et al. Long noncoding RNA CRNDE promotes colorectal cancer cell proliferation via epigenetically silencing DUSP5/CDKN1A expression. *Cell Death Dis*. 2017;8(8):e2997. doi:10.1038/cddis.2017.328.

73. Han P, Li J, Zhang B, et al. The lncRNA CRNDE promotes colorectal cancer cell proliferation and chemoresistance via miR-181a-5p-mediated regulation of Wnt/β-catenin signaling. *Mol Cancer*. 2017;16(1):9. doi:10.1186/s12943-017-0583-1.

74. Zheng J, Li X, Wang P, et al. CRNDE affects the malignant biological characteristics of human glioma stem cells by negatively regulating miR-186. *Oncotarget*. 2015;6(28):25339-25355. doi:10.18632/oncotarget.4509.

75. Wei. Associations of Single Nucleotide Polymorphisms in miR-146a, miR-196a, miR-149 and miR-499 with Colorectal Cancer Susceptibility. *J Cancer Prev*. 2014;15. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.829.7344&rep=rep1&type=pdf. Accessed December 15, 2017.

76. Lu S, Pardini B, Cheng B, et al. Single Nucleotide Polymorphisms within Interferon Signaling Pathway Genes Are Associated with Colorectal Cancer Susceptibility and Survival. Lamaze C, ed. *PLoS One*. 2014;9(10):e111061.

doi:10.1371/journal.pone.0111061.

77. Mullany LE, Herrick JS, Wolff RK, Slattery ML. Single nucleotide polymorphisms within MicroRNAs, MicroRNA targets, and MicroRNA biogenesis genes and their impact on colorectal cancer survival. *Genes, Chromosom Cancer*. 2017;56(4):285-295. doi:10.1002/gcc.22434.

78. Kim S-K, Kim S-Y, Kim J-H, et al. A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients. *Mol Oncol*. 2014;8(8):1653-1666. doi:10.1016/j.molonc.2014.06.016.

79. Andrews S. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Published 2010. Accessed December 15, 2017.

80. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25(9):1105-1111. doi:10.1093/bioinformatics/btp120.

81. Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics*. 2011;27(17):2325-2329. doi:10.1093/bioinformatics/btr355.

82. Trapnell, Cole et al. Cuffmerge. http://software.broadinstitute.org/cancer/software/genepattern/modules/docs/Cuffmerge/3?print=yes. Accessed December 13, 2017.

83. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol*. 2012;31(1):46-53. doi:10.1038/nbt.2450.

84. Martin JA, Wang Z. Next-generation transcriptome assembly. *Nat Rev Genet*.

2011;12(10):671-682. doi:10.1038/nrg3068.

85. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;7(3):562-578. doi:10.1038/nprot.2012.016.

86. Thomas PD, Campbell MJ, Kejariwal A, et al. PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Res*. 2003;13(9):2129-2141. doi:10.1101/gr.772403.

87. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24-26. doi:10.1038/nbt.1754.

88. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29(1):308-311. http://www.ncbi.nlm.nih.gov/pubmed/11125122. Accessed December 13, 2017.

89. Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res*. 2009;37(Web Server issue):W600-5. doi:10.1093/nar/gkp290.

90. Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning–based sequence model. *Nat Methods*. 2015;12(10):931-934. doi:10.1038/nmeth.3547.

91. Yan T-H, Lu S-W, Huang Y-Q, et al. Upregulation of the long noncoding RNA HOTAIR predicts recurrence in stage Ta/T1 bladder cancer. *Tumor Biol*. 2014;35(10):10249-10257. doi:10.1007/s13277-014-2344-8.

92. Jiang Y, Li Z, Zheng S, et al. The long non-coding RNA HOTAIR affects the radiosensitivity of pancreatic ductal adenocarcinoma by regulating the expression of Wnt inhibitory factor 1. *Tumor Biol*. 2016;37(3):3957-3967. doi:10.1007/s13277-015-4234-0.

93. Su C, Li J, Cao Y, Xiao L, Shi C. Long noncoding RNA HOTAIR promotes cisplatin resistance in gastric cancer through promoting WIF-1 promoter methylation. *Int J Clin Exp Med*. 2017;10(2):2717-2726. www.ijcem.com. Accessed December 10, 2017.

94. Bubici C, Papa S. JNK signalling in cancer: in need of new, smarter therapeutic targets. *Br J Pharmacol*. 2014;171(1):24-37. doi:10.1111/bph.12432.

95. Qin J-J, Nag S, Wang W, et al. NFAT as cancer target: Mission possible? *Biochim Biophys Acta - Rev Cancer*. 2014;1846(2):297-311. doi:https://doi.org/10.1016/j.bbcan.2014.07.009.

96. Ma B, Hottiger MO. Crosstalk between Wnt/β-Catenin and NF-κB Signaling Pathway during Inflammation. *Front Immunol*. 2016;7:378. doi:10.3389/fimmu.2016.00378.

97. Moreau M, Mourah S, Dosquet C. β-Catenin and NF-κB cooperate to regulate the uPA/uPAR system in cancer cells. *Int J Cancer*. 2011;128(6):1280-1292. doi:10.1002/ijc.25455.

98. Liu J, Liao Y, Ma K, et al. PI3K is required for the physical interaction and functional inhibition of NF-κB by β-catenin in colorectal cancer cells. *Biochem Biophys Res Commun*. 2013;434(4):760-766. doi:10.1016/j.bbrc.2013.03.135.

99. Du Q, Zhang X, Cardinal J, et al. Wnt/beta-catenin signaling regulates cytokine-induced human inducible nitric oxide synthase expression by inhibiting nuclear factor-kappaB activation in cancer cells. *Cancer Res*. 2009;69(9):3764-3771. doi:10.1158/0008-5472.CAN-09-0014.

100. Deng J, Miller SA, Wang H-Y, et al. β-catenin interacts with and inhibits NF-κB in human colon and breast cancer. *Cancer Cell*. 2002;2(4):323-334. doi:10.1016/S1535-6108(02)00154-X.

101. Portoso M, Ragazzini R, Brenčič Ž, et al. PRC2 is dispensable for HOTAIR-mediated transcriptional repression. *EMBO J*. 2017;36(8):981-994. doi:10.15252/embj.201695335.

102. Xue Y, Gu D, Ma G, et al. Genetic variants in lncRNA HOTAIR are associated with risk of colorectal cancer. *Mutagenesis*. 2015;30(2):303-310. doi:10.1093/mutage/geu076.

103. Liu X-S, Liu Z, Gerarduzzi C, et al. Somatic human ZBTB7A zinc finger mutations promote cancer progression. *Oncogene*. 2016;35(23):3071-3078. doi:10.1038/onc.2015.371.

104. Liu X-S, Haines JE, Mehanna EK, et al. ZBTB7A acts as a tumor suppressor through the transcriptional repression of glycolysis. *Genes Dev*. 2014;28(17):1917-1928. doi:10.1101/GAD.245910.114.

105. Molloy ME, Lewinska M, Liu X-S, Wang AA, Yuan Z-M. Abstract 3657: ZBTB7A is a tumor suppressor and novel therapeutic target in triple-negative breast cancer. *Cancer Res*. 2016;76(14 Supplement):3657-3657. doi:10.1158/1538-7445.AM2016-3657.

106. Zu X, Yu L, Sun Q, et al. SP1 enhances Zbtb7A gene expression via direct binding to GC box in HePG2 cells. *BMC Res Notes*. 2009;2(1):175. doi:10.1186/1756-0500-2-175.

107. Nieminen TT, Shoman S, Eissa S, Peltomäki P, Abdel-Rahman WM. Distinct genetic and epigenetic signatures of colorectal cancers according to ethnic origin. *Cancer Epidemiol Biomarkers Prev*. 2012;21(1):202-211. doi:10.1158/1055-9965.EPI-11-0662.

108. Xu X-L, Yu J, Zhang H-Y, et al. Methylation profile of the promoter CpG islands of 31 genes that may contribute to colorectal carcinogenesis. *World J Gastroenterol*. 2004;10(23):3441-3454. doi:10.3748/WJG.V10.I23.3441.

109. Li H, Ma S-Q, Huang J, Chen X-P, Zhou H-H. Roles of long noncoding RNAs in colorectal cancer metastasis. *Oncotarget*. 2017;8(24):39859-39876.

doi:10.18632/oncotarget.16339.

APPENDIX

These images are in the public domain and may be reproduced or copied without permission. However, a citation as to the source is appreciated. The suggested citation is as follows:

SEER Cancer Stat Facts: Colon and Rectum Cancer. National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/statfacts/html/colorect.html.

Figure 1 copyright agreement

These images are in the public domain and may be reproduced or copied without permission. However, a citation as to the source is appreciated. The suggested citation is as follows:

SEER Cancer Stat Facts: Colon and Rectum Cancer. National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/statfacts/html/colorect.html.

Figure 2 copyright agreement

This Agreement between American University in Cairo -- Amany Abdel-Motaleb ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4250310788798
License date	Dec 15, 2017
Licensed Content Publisher	Elsevier
Licensed Content Publication	Clinical Colorectal Cancer
Licensed Content Title	A Review of the Evolution of Systemic Chemotherapy in the Management of Colorectal Cancer
Licensed Content Author	Bengt Gustavsson,Göran Carlsson,David Machover,Nicholas Petrelli,Arnaud Roth,Hans-Joachim Schmoll,Kjell-Magne Tveit,Fernando Gibson

Figure 3 copyright agreement

Copyright

© Roberts et al.; licensee BioMed Central Ltd. 2011

This article is published under license to BioMed Central Ltd. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Figure 4 copyright agreement

Copyright © 2015 Jansson, Kim and Cheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Figure 5 copyright agreement

This Agreement between American University in Cairo -- Amany Abdel-Motaleb ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4265211205960
License date	Jan 10, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Oncogene
Licensed Content Title	Epithelial–mesenchymal transition in development and cancer: role of phosphatidylinositol 3′ kinase/AKT pathways

Figure 6 copyright agreement

Copyright © 2012 Jean-Jacques Lebrun. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Figure 7 copyright agreement

This Agreement between American University in Cairo -- Amany Abdel-Motaleb ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4250690771283
License date	Dec 16, 2017
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	FEBS Letters
Licensed Content Title	Dissecting the diverse functions of the metastasis suppressor CD82/KAI1
Licensed Content Author	Yien Che Tsai,Allan M. Weissman

Figure 8 copyright agreement

This Agreement between American University in Cairo -- Amany Abdel-Motaleb ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4250691093127
License date	Dec 16, 2017
Licensed Content Publisher	Elsevier
Licensed Content Publication	Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms
Licensed Content Title	LncRNA: A link between RNA and cancer
Licensed Content Author	Guodong Yang,Xiaozhao Lu,Lijun Yuan

Figure 9 copyright agreement

This Agreement between American University in Cairo -- Amany Abdel-Motaleb ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4250270932800
License date	Dec 15, 2017
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Genetics
Licensed Content Title	Long non-coding RNAs: new players in cell differentiation and development
Licensed Content Author	Alessandro Fatica, Irene Bozzoni

Figure 10 copyright agreement

This Agreement between American University in Cairo -- Amany Abdel-Motaleb ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4250711219663
License date	Dec 16, 2017
Licensed Content Publisher	Elsevier
Licensed Content Publication	Trends in Biochemical Sciences
Licensed Content Title	Long Noncoding RNAs: A New Regulatory Code in Metabolic Control
Licensed Content Author	Xu-Yun Zhao,Jiandie D. Lin

Figure 11 copyright agreement

ANNUAL REVIEWS

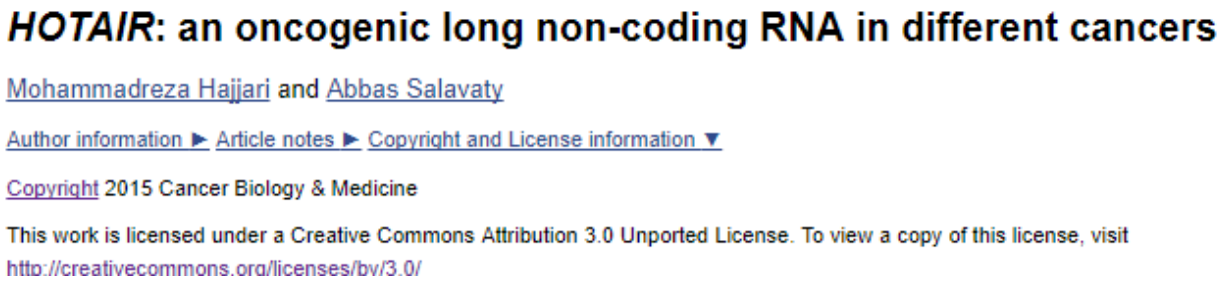
Title:	Genome Regulation by Long Noncoding RNAs
Author:	John L. Rinn, Howard Y. Chang
Publication:	Annual Review of Biochemistry
Publisher:	Annual Reviews
Date:	Jul 7, 2012

Copyright © 2012, Annual Reviews

Permission Not Required

Material may be republished in a thesis / dissertation without obtaining additional permission from Annual Reviews, providing that the author and the original source of publication are fully acknowledged.

Figure 12 copyright agreement



HOTAIR: an oncogenic long non-coding RNA in different cancers

Mohammadreza Hajjari and Abbas Salavaty

Author information ► Article notes ► Copyright and License information ▼

Copyright 2015 Cancer Biology & Medicine

This work is licensed under a Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by/3.0/

Figure 13 copyright agreement
