ALTERNATIVE POLYADENYLATION IN A DIFFERENTIATION MODEL: CACO-2

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

OĞUZHAN BEĞİK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
MOLECULAR BIOLOGY AND GENETICS

JUNE 2017

Approval of the thesis:

ALTERNATIVE POLYADENYLATION IN A DIFFERENTIATION MODEL: CACO-2

submitted by **Oğuzhan BEĞIK** in partial fulfilment of the requirements for the degree of **Master of Science in Department of Molecular Biology and Genetics, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Orhan Adalı _____
Head of Department**, Biological Sciences**

Assoc.Prof.Dr. A. Elif Erson-Bensan _____
Supervisor, **Biology Dept., METU**

Assoc.Prof.Dr. Sreeparna Banerjee _____
Co-Supervisor, **Biology Dept., METU**

Examining Committee Members:

Prof. Dr. Mesut Muyan _____
Head of Committee, Biology Dept., METU

Assoc. Prof. Dr. A. Elif Erson Bensan _____
Biology Dept., METU

Assoc. Prof. Dr Sreeparna Banerjee _____
Biology Dept., METU

Prof. Dr. Çetin Kocaefe _____
Medical Biology Dept., Hacettepe Uni.

Prof. Dr. Tolga Can _____
Computer Engineering Dept., METU

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Oğuzhan Beğik

Signature :

ABSTRACT

ALTERNATIVE POLYADENYLATION IN A DIFFERENTIATION MODEL: CACO-2

Beğik, Oğuzhan

M. S., Department of Molecular Biology and Genetics
Supervisor: Assoc. Prof. Dr. A. Elif Erson Bensan

June 2017, 75 pages

Alternative polyadenylation (APA) is the selection of proximal or distal poly(A) signals on pre-mRNAs. APA has been implicated in many cellular processes, including differentiation. Resulting APA isoforms may have different stability or localization, which may eventually alter the protein function. Therefore, it is important to reveal APA isoforms to better understand post-transcriptional mechanisms in development. In this study, we aimed to investigate APA isoforms in an enterocyte differentiation model, Caco-2 cells. Enterocyte differentiation take place on the axis from colon crypt to villus to produce enterocytes from the intestinal stem cells. Caco-2 cells are derived from colon adenocarcinoma and are able to undergo spontaneous enterocyte differentiation upon confluency. Earlier, we have developed APADetect tool which uses microarray gene expression data to analyze APA events. We used APADetect in order to analyze the APA events in differentiating Caco2 cells. We identified 91 3'UTR lengthening and 43 3'UTR shortening events in differentiated Caco2 cells compared to proliferating Caco-2 cells. APA events were mostly enriched for biological processes such as enzyme binding, endocytosis and RNA processing.

To begin investigating the functional significance of APA isoforms, we have looked into availability or loss of conserved miRNA binding sites on APA isoforms. Interestingly, we found an enrichment of miRNA binding sites close to the active poly(A) sites at the end of the mRNAs, which may allow easier access to miRNAs. Next, we began confirming the *in silico* results by real time RT-PCR (RT-qPCR) using proliferating and differentiated Caco-2 cells. Overall our approach serves as a platform for novel gene discovery in differentiation studies where conventional gene expression analysis may have overlooked 3'UTR isoforms.

Keywords: Caco-2, APADetect, miRNA

ÖZ

BİR FARKLILAŞMA MODELİNDE ALTERNATİF POLİADENİLASYON: CACO-2 HÜCRELERİ

Beğik, Oğuzhan

Yüksek Lisans, Moleküler Biyoloji ve Genetik Bölümü
Tez Yöneticisi: Doç. Dr.A. Elif Erson Bensan

Haziran 2017, 75 sayfa

Alternatif poliadenilasyon (APA), pre-mRNA'da bulunan yakın veya uzaktaki poly(A) sinyallerinin seçimidir. APA, farklılaşma dahil olmak üzere çoğu hücre süreçlerinde önemli rol oynamaktadır. APA sonucunda oluşan izoformlar, farklı stabiliteye ve lokalizasyona sahip olabilirler. Farklı lokalizasyon, proteinin fonksiyonunu da değiştirebilir. Bu yüzden, gelişim sürecinde gerçekleşen transkripsiyon sonrası mekanizmaları anlamak için, APA isoformlarını belirlemek bir önem arz etmektedir. Bu çalışmada, enterosit farklılaşma modeli olan Caco-2 hücrelerindeki APA izorformlarını incelemeyi amaçladık. Caco-2 hücreleri kolon adenokarsinoma'dan türemiş olup, konflüent olmaları sonucunda spontane enterosit farklılaşması sürecine girebilmektedirler. Önceden geliştirdiğimiz ve ekspresyon datasını kullanarak APA olaylarını inceleyen APADetect algoritmasını kullanarak, Caco-2 farklılaşmasında gerçekleşen APA olaylarını analiz ettik. Farklılaşmış Caco-2 hücrelerini, çoğalan Caco-2 hücreleri ile karşılaştırdığımızda, 91 3'UTR uzaması, 43 3'UTR kısalması belirledik. 3'UTR uzunluğunun, miRNA'ya bağlı regulasyondaki önemini göz önünde bulundurarak, kısa ve uzun izoformlardaki miRNA bağlanma

bölgelerini inceledik. İlginç bir şekilde, kullanlılan poly(A) bölgesine yakın (mRNA'nın bitiminde) bölgede bir zenginleşme gördük. Burada gözlemlediğimiz zenginleşme olayının sonucu, daha erişilebilir miRNA bağlanma bölgesi oluşması olabilir. Sonrasında, analiz sonucunda bulduğumuz APA olaylarını doğrulamak için, çoğalan ve farklılaşmış Caco-2 hücrelerini gerçek zamanlı PCR kullanarak (RT-qPCR) inceledik. Sonrasında, farklılaşma olayında farklı düzenlenen bir transkriptin üzerine yoğunlaşarak, bu transkript ile fonsiyonel analizler yaptık. Genel olarak yaklaşımımız, farklılaşma ile ilgili çalışmalarda fazla kullanılan gen ekpresyon analizlerinde gözden kaçırılmış olan 3'UTR izoformlarından yeni gen bulunması için bir platform sağlamaktadır.

Anahtar kelimeler: Caco-2, APADetect, miRNA

To the fantastic Mr. Feynman and the source of my enthusiasm; my family

ACKNOWLEDGEMENTS

I would like to express my endless gratitude to my supervisor Assoc. Prof. Dr. A. Elif Erson-Bensan and my mentor Prof. Dr. Mesut Muyan for they have showed me the ways of becoming a great scientist by forcing me to be better all the time. I also would like to thank my co-supervisor Assoc. Prof. Dr. Sreeparna Banerjee and Dr. Sinem Tuncer for their contributions to my research.

I would like to thank the rest of my thesis committee members, Prof. Dr. Çetin Kocaefe and Prof. Dr. Tolga Can.

I thank all current and previous lab members. Especially Dr. Begüm Akman for giving me a place by her side in the lab and giving me the most precious advices. Very special thanks go to my friend and lab mate Harun, who helped me ease my burden during the thesis work.

My warmest thanks go to my family, who have always trusted my decisions and encouraged me to be who I want to be.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ACTB	Beta-actin
Ago 2	Argonaute 2
bp	Base pair
CDC6	Cell Division cycle 6
CSTF2	Cleavage stimulation factor subunit 2
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GSEA	Gene Set Enrichment Analysis
miRNA	microRNA
MSigDB	Molecular Signature Database
MYC	c-Myc
Poly(A)	Polyadenylation
RBP	RNA-Binding Protein
RT-qPCR	Real-Time quantitative Polymerase Chain Reaction
SI	Sucrase Isomaltase
sh-RNA	Small heterogenous RNA
SLR	Short: Long Ratio
SNX3	Sorting Nexin-3
TCF3	Trancription factor 3
UTR	Untranslated region
WLS	Wntless

CHAPTER 1

INTRODUCTION

1.1 Alternative Polyadenylation and Its Role in Development

1.1.1 Alternative Polyadenylation Mechanism

Transcription by the RNA Polymerase II enzyme is coupled to polyadenylation by which the transcript is cleaved off at the 3'end. With the exception of histone mRNAs, almost all of the RNA Polymerase II transcripts go through this process, which is an essential step in the post-transcriptional regulatory mechanisms and protection against nucleases [1]. A protein complex defined as the polyadenylation complex recognizes a sequence called polyadenylation (poly(A)) signal located on the pre-mRNA. The most frequently used poly(A) signal is AAUAAA; however, other less frequent signal variants also do exist [2]. Several cis-regulatory elements located upstream and downstream of the poly(A) signal also contribute to the cleavage process. For example, U-rich elements, UGUA elements and CA sequence usually found just upstream of the cleavage site. GU-rich and U-rich elements are found downstream of the cleavage site [3]. The recognition is followed by the cleavage of the transcript from 15-20 nt downstream of the signal and addition of a poly(A) tail by the enzyme called poly(A) polymerase [4]. The poly(A) tail length could be extended up to 250 adenine nucleotides and it primarily depends on the nuclear poly(A)-binding protein, which destroys the cooperation between polyadenylation machinery and poly(A) polymerase [5].

Poly(A) machinery consists of four major complexes and other auxiliary proteins. The four protein sub-complex are the cleavage and polyadenylation specificity factor (CPSF), including CPSF1 (160 kDa subunit), CPSF2 (100 kDa subunit), CPSF3 (73 kDa subunit), CPSF4 (30 kDa subunit), FIP1 (factor interacting with PAP) and WDR33; cleavage stimulation factor (CSTF), including CSTF1 (50 kDa subunit), CSTF2 (64 kDa subunit), CSTF2T (paralogue of CSTF2), and CSTF77 (77 kDa subunit); cleavage factor I (CFI), including CFI25 (25 kDa subunit), CFI59 (59 kDa subunit), CFI68 (68 kDa subunit) and cleavage factor II (CFII), including PCF11 and CLP1. Each of these proteins have their specific roles in the polyadenylation machinery by interacting with the cis elements or forming bridge structure between the machinery proteins. Polyadenylation signal is targeted by the CPSF4 and WDR33 proteins while the U-rich sequences are targeted by the FIP1; UGUA elements by the CFI25; U/GU-rich elements by the CSTF2 and CSTF2T. Having endonuclease activity, CPSF3 is likely to target the CA nucleotide just before the cleavage site. Symplekin, carboxy-terminal domain (CTD) of RNA-Polymerase II, poly(A) polymerase, nuclear poly(A)- binding protein 1 (PABPN1), and retinoblastoma-binding protein (RBBP6) are not included in any of the protein sub-complexes, yet have essential roles as scaffolding proteins [3]. An illustration of the poly(A) complex shows the subunit protein localization on the RNA template (Figure 1.1)

Recent discoveries showed that around 70% of all known human genes have multiple poly(A) sites and that the usage of these poly(A) sites, referred to as alternative polyadenylation (APA), is tightly regulated [6].

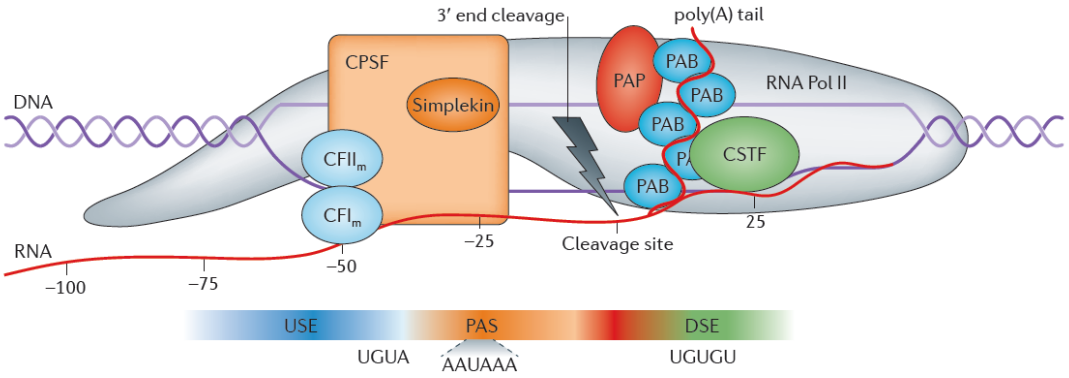


Figure 1.1. Polyadenylation complex subunit proteins and the RNA template with cis-regulatory elements. The cis acting elements are upstream sequence elements (USE) and downstream sequence elements (DSE) increasing cleavage efficiency; poly(A) signal (PAS) indicating the cleavage site with 6 nt motif. The canonical PAS is AAUAAA with at least ten other variants, which is positioned around 15-30 nucleotides upstream of the cleavage site. Cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulating factor protein complexes bind to PAS and DSEs, respectively. Poly(A) polymerase, poly(A) binding protein (PAB), simplekin scaffold protein and cleavage factors Im (CFIm) and IIm(CFIIm) are other protens involved in this complex (Figure taken from Elkon et al., 2013) [4].

APA events can be grouped into four categories based on poly(A) site positions; (1) tandem 3’ untranslated region (3’UTR), (2) alternative terminal exon, (3) intronic and (4) internal exon. The most common type of APA is the tandem 3’UTR APA, which takes place in the last exon and does not change the translated protein. Other three types however, affect the genes in the protein-coding extent and are probably coupled to splicing. Alternative terminal exon type APA causes a different last exon usage in APA isoforms. Intronic APA, on the other hand, happens when there is a choice of intronic poly(A) signal, resulting in a novel terminal exon. Finally, internal exon APA includes the presence of a poly(A) site inside the protein-coding region [4] (Figure 1.2).

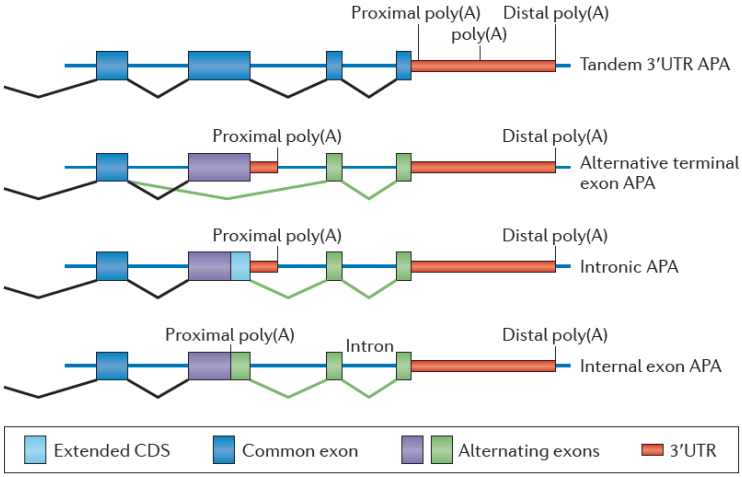


Figure 1.2. APA in four different forms. Tandem 3'UTR APA is the most common APA type, leading to cleavage and polyadenylation on 3'UTR without affecting the protein sequence. The other types raise the possibility for novel protein sequences, in addition to altering the 3'UTR (Figure taken from Elkon et al., 2013) [4].

The choice of different poly(A) signals by the poly(A) machinery has been shown to depend on numerous factors. Histone and DNA modifications are reported to be one of the determinants of the APA, for instance APA events in a newly characterized murine imprinted gene (H13) is affected by the genomic imprinting such that the allele lacking the methylation tends to use its internal poly(A) signal [7]. On the other hand, poly(A) signals used more frequently have histone enrichment on their downstream sequences [8]. As another factor, transcription speed is also a principal factor in APA. For example, when transcriptional elongation is partially blocked, upstream poly(A) sites are favored [9], which is consistent with a finding that suggests slower elongation of RNA polymerase II leads to more proximal poly(A) site selection [10].

Polyadenylation machinery proteins have main regulatory functions in APA and thus, their expression patterns are a major determinant in poly(A) signal choice. During induced pluoripotent stem cell generation, for example, the upregulation of CSTF subunits have lead to a general proximal poly(A) site selection [11].

Additionally, one of the earliest observations of this phenomenon provided evidence for the proximal poly(A) site choice of IgM mRNAs when there is a higher CSTF2 protein level in the mouse primary B cells, leading to a switch from membrane-bound heavy chain to secreted heavy chain[12]. The role of CSTF2 protein is shared by its paralogue CSTF2T and the silencing of both factors leads to a general lengthening of 3'UTRs pattern in HeLa cells [13]. On the contrary to the CPSF factors that preferentially mediates shorthening events, CFIm factors such as CFI-25, CFI-68 have opposing roles in both normal physiological and disease states [14]–[16]. Another RNA-Binding factor, Fip1, plays a vital role in the ESC self-renewal by mediating the choice of proximal poly(A) signals in an ESC-specific group of genes [17]. It is also notable to mention that, another factor PCF11 which does not directly interact with the RNA has also been shown to favor the proximal poly(A) signals [16].

In addition to the polyadenylation complex elements, splicing-related proteins might also regulate APA events via various mechanisms. For instance, although muscleblind-like (MBNL) 1 and 2 are regulators of alternative splicing throughout the muscle and brain development, their silencing in mouse embryo fibroblasts had a very dramatic effect on alternative polyadenylation events. This was a consequence of repression, by direct binding of MBNL proteins to the RNA sequences to block the binding of other core proteins [18]. Another splicing-related protein, U1 snRNP (U1) has been shown to possess a vital role in regulating the transcript lenghts and the absence of U1 leads to truncated isoforms [19]. HnRNPH splicing element, is involved in promoting the proximal poly(A) signal usage, indicated upon the observation of a high abundance of proximal signal usage in hnRNPH knockdown cells [20].

Other proteins have also been implicated in APA. An example is the cytoplasmic polyadenylation element binding protein 1 (CPEB1), an RNA-Binding Protein (RBP) often related to the mRNA translation process. CPEB1 positions itself to the nucleus in association with the splicing elements and modulates the choice of proximal poly(A)

signal in certain set of genes, which has been shown to be related to proliferation and tumorigenicity. Interestingly, this mediation in turn increases the translation efficiency, thereby combining its cytoplasmic and nuclear functions [21]. In addition to that, another RNA-binding protein HUR (Hu antigen R) interferes with its own transcription, so that the resulting mRNA bears a longer transcript. This is indeed due to HUR protein binding to its proximal poly(A) signal and therefore prevention of CSTF2 binding, which then mediates the usage of the distal poly(A) signal that has the shorter half-life. This feedback loop compensates the high amount of HUR protein in cancer cells by decreasing the mRNA stability [22]. The association between the RNA Polymerase II and Embryonic Lethal, Abnormal Vision, Drosophila (ELAV), an RNA-Binding protein, has been suggested to be a determinant factor in the poly(A) signal choice in the lengthened genes. The possible mechanism is explained by ELAV's ability to bind and suppress the RNA processing at the proximal poly(A) sites and thereby promoting the longer transcripts (Figure 1.3) [23].

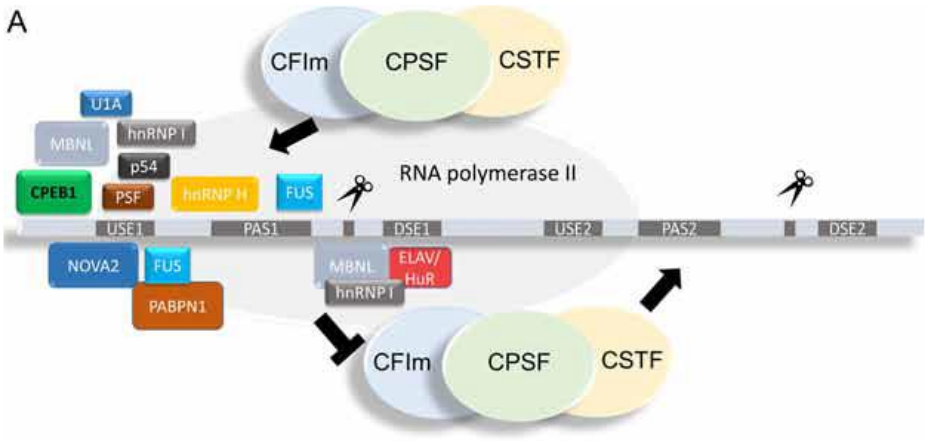


Figure 1.3. Role of RNA-Binding Proteins in APA. (A) Various RBPs related to APA regulation are shown. RBPs act by enhancing or blocking the enrollment of core polyadenylation machinery complexes (CPSF, CTSF, and CFIm) to the cis elements. PAS1 is the proximal poly(A) signal, while PAS2 is the distal poly(A) signal. USE, U-rich/UGUA upstream elements; DSE, U-/GU-rich downstream elements. Scissors indicate cleavage site (poly(A) site) (Figure taken from Erson-Bensan 2016) [24].

1.1.2 Alternative Polyadenylation Detection Platforms

Detecting APA events in the transcriptome is a challenging task. There are several different methods for APA detection which are mainly based on either RNA-Seq or Microarray technology. DaPars (Dynamic analyses of alternative polyadenylation from RNA-seq) was developed to analyze RNA-seq data and characterize dynamic APA events. By applying this bioinformatics algorithm to TCGA Pan-Cancer samples, 1,346 genes with APA events were discovered. Although this technique may provide insights into APA events in different cell types, there are several drawbacks of RNA-Seq technology that could provide a bias when using RNA-Seq data for the analysis of APA events. The most important drawback is the fact that sequencing is based on random priming and differential amplification, which leads to read depletion near 3'ends [6].

Our group has developed a microarray-based method, called APADetect to analyze APA events. Probes are mostly designed from 3'UTRs, which makes it possible to analyze and identify APA events with the information of differential signal intensities around poly(A) sites. A major drawback of this technology is that one can only analyze the transcripts which bear probe sets divided by poly(A) sites (Figure 1.4).

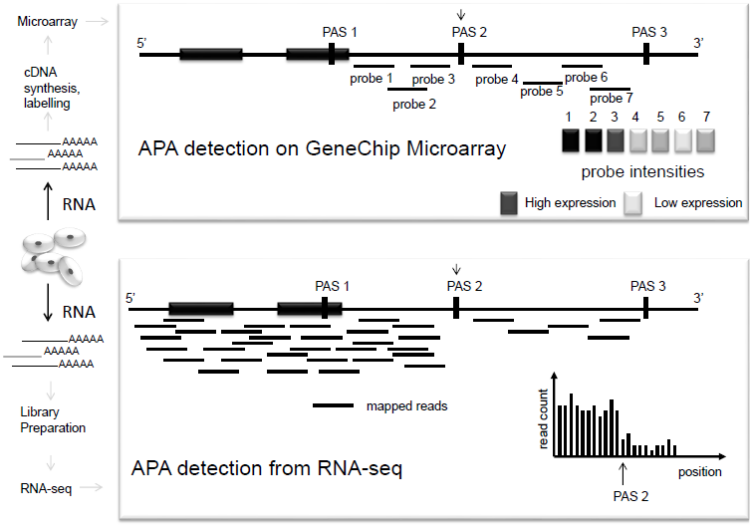


Figure 1.4. Illustration of the two different APA detection platforms. GeneChip microarray platform detection by probe intensity is seen in upper part. RNA molecules are labeled during cDNA synthesis in order to measure probe intensities for expression detection. Probe set is separated into two groups based on their position around the poly(A) site 2 (PAS2). RNA-Seq platform detection by mapped reads is seen in the lower part. Prepared library is sequenced and subsequent reads are aligned to the reference sequence. A significant decrease of the read counts mark the poly(A) site (Taken from Erson-Bensan & Can, 2016) [25].

1.1.3 Consequences of APA

Alternative 3'UTR lengths direct the path of the post-transcriptional and sometimes post-translational regulation of the resulting alternative transcripts. Shorter 3'UTR tend to escape from miRNAs and RBP regulation and therefore have a higher stability and lead to increased protein levels [26]. Interestingly, 3'UTR lengths do not always correlate with protein levels. Specifically, a genome-wide study of proliferating T cells showed that although they had a general 3'UTR shortening event, this was hardly a determinant of the protein output [27]. A study provides evidence for the highly-conserved miRNA binding site in the upstream position of the proximal poly(A) sites, indicating an enhanced and more effective functionality of miRNAs in that position. Therefore, it implies a more robust miRNA binding to target site when the proximal poly(A) site is used and the mRNA is more susceptible to post-transcriptional

regulation (Figure 1.5) [28].

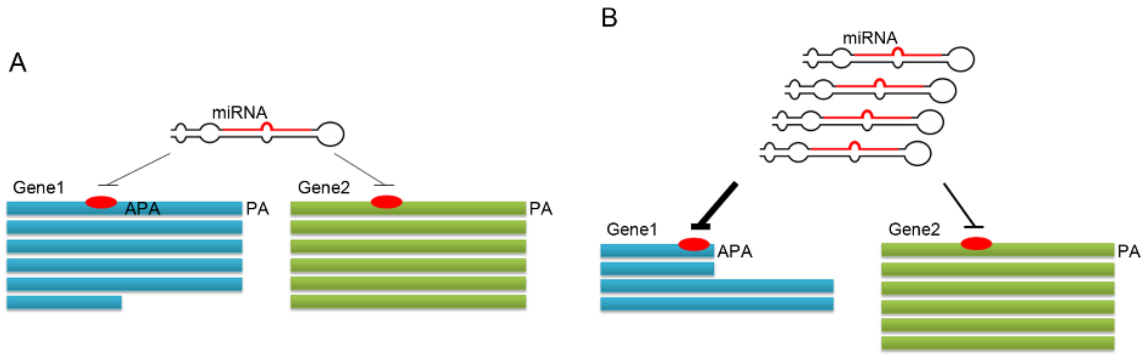


Figure 1.5. Accesibility of miRNA binding site upon alternative polyadenylation. In the first case (A), the miRNA expression is at steady state and there is mostly long 3'UTR choice. In this condition, miRNA functions almost the same on these transcripts since the binding site is not near the cleavage site. (B) Expression of miRNA is upregulated and gene 1, but not gene 2, has shorter isoform, leading to the increased accessibility of miRNA binding site. This makes miRNA targeting to binding site more effectively. This proposed mechanism provides another regulatory layer for the post-transcriptional regulation (Figure taken from Hoffman et al., 2016) (Hoffman et al., 2016).

Besides its effect on the mRNA stability and translation, APA has also been indicated as a determinant of where the mRNA will localize within the neurons where there is a high polarity and specific protein synthesis events needs to be localized [29].

Short and long isoforms may have different RBP sites, causing differential recruitment of proteins during translation, which may have eventually change the protein localization. Specifically, the long 3'UTR of CD47 enables the recruitment of its endoplasmic reticulum protein to the cell membrane, via providing a scaffold for the RBP HuR (ELAVL1) and SET, which facilitates the localization [30] (Figure1.6).

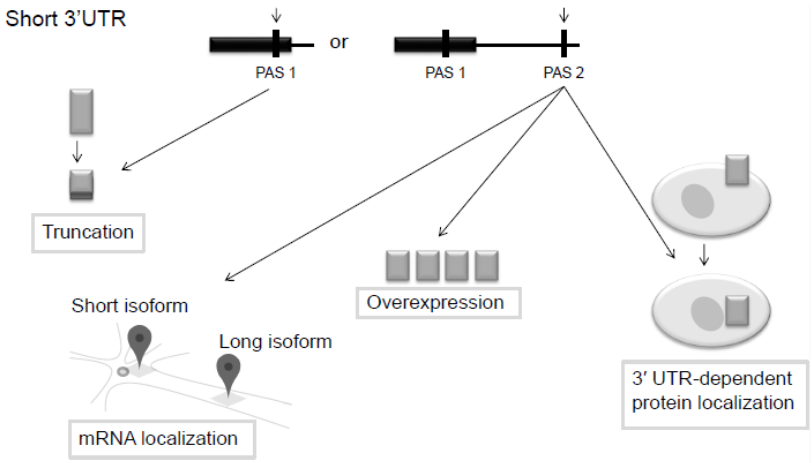


Figure 1.6. Consequences of different poly(A) site (PAS) choice. In case of PAS1 selection, the transcript is cleaved of from its coding region which leads to truncation. PAS2 selection is related to the 3'UTR of the transcript, resulting in different mRNA localization, protein level, and protein localization without altering the protein structure (Figure taken from Erson-Bensan & Can, 2016) [25].

1.1.4 Alternative Polyadenylation and Development

Taken together, APA is considered as a major regulatory mechanism in many cellular processes, including development. Accumulating evidence show existence of APA events during the development [31]–[33], suggesting a functional role of APA in tissue-specific differentiation. A pivotal study in zebrafish tissues at different developmental stages has demonstrated that the most significant 3'UTR length difference is observed in the ovary and brain tissue. Interestingly, same study suggested that shorter transcripts tend to be degraded more than the longer isoforms in pre-maternal to zygotic transition (MZT) embryos [32]. Moreover, 20,000 tissue-specific polyadenylation sites are reported in approximately 30% of transcripts in somatic cells, resulting with 3'UTR isoforms significantly enriched with microRNA targets [34].

When induced pluripotent stem cells (iPSCs) differentiate into somatic cells, 3'UTR lengths tend to be shorter whereas when they differentiate into spermatogonial cells, 3'UTR lengths tend to be longer. The same transcripts that are experiencing alternative poly(A) site preference, on the other hand, show a pattern during embryonic

development that is opposite to what has been observed in iPSCs differentiation [35] . These studies clearly indicate the importance of APA in determining the cell fate during the development and differentiation.

1.2 Intestinal Crypt and Differentiation

Human colon is composed of columnar epithelial cells which make up folded structures to create the crypt. Many crypts reside in a place called stem-cell niche, composed of a population of stem-cell and mesenchymal cells at the crypt base [36]. Gastrointestinal (GI) epithelium homeostasis relies on a constant renewal, which is based on an important differentiation process on the axis from the crypt to the villus [37].

As the stem cells proliferate and subsequently migrate on the axis from the crypt to the villus, they undergo cell cycle arrest and differentiate into four major cell types (Figure 1.7) [38]. Of these cell types, the most abundant cells are colonocytes (also referred as enterocyte or absorptive cells) with a distinct polarized cell architecture [39]. The basal part of the cell is in contact with the extracellular matrix produced by the epithelial and mesenchymal cells. The interaction between the extracellular matrix and cell surface receptors including integrins is an important part in the renewal of the intestinal epithelium via the migration and differentiation of stem cells along the axis. The apical surface, on the other hand is composed of brush border membranes, which have high expression of hydrolases, including Sucrase Isomaltase and Alkaline Phosphatase [37]. The brush border is a structure which broadens the membrane surface for more hydrolysis and transport [39]. The other cell type is the goblet cells, which are responsible for mucus secretion. Peptide-hormone secreting enteroendocrine cells are very few in numbers and they contribute to GI motility. Lastly, paneth cells are localized on the ascending colon and have association with the antimicrobial defense [36], [37]. In addition to these cell types, there are two very rare cell types called tuft cells (related to chemical sensation) and M cells (related to transport of luminal antigens to lymphoid cells). Morphological characteristics and localizations of these

cell types are illustrated in Figure 1.8 [40].

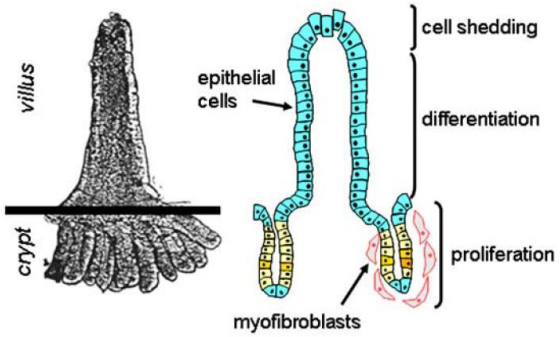


Figure 1.7. A micrograph of the crypt-villus axis (left) and scheme of the lining of cells in different states along the axis (right) (Figure taken from Simon-Assman et al., 2007) [38].

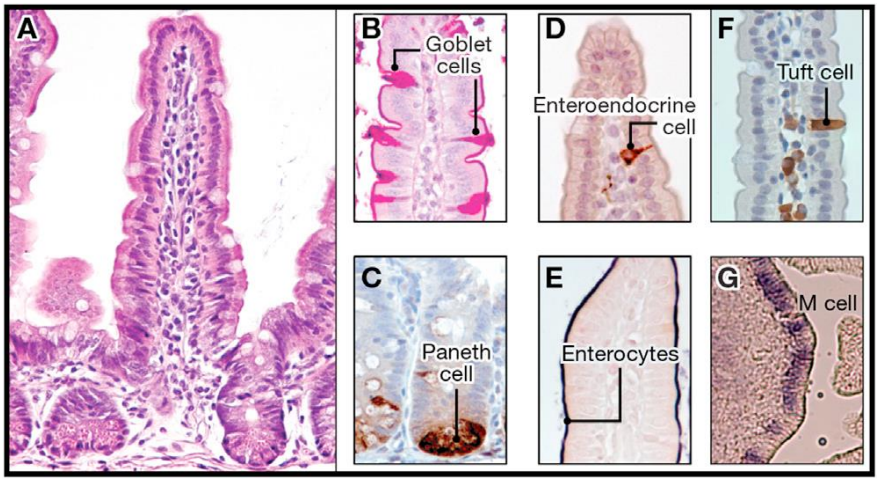


Figure 1.8. Hematoxylin and eosin staining of the intestinal epithelium (A). Periodic Acid-Schiff staining of the goblet cells (purple) located on villus (B). Lysozyme staining of Paneth cells (brown) located at the bottom of crypt (C). Chromogranin-staining of enteroendocrine cells (brown) (D). Alkaline phosphatase staining of colonocytes (blue) (E). DCAMKL1 staining of tuft cell (F). M-Cells (G) (Figure taken from Clevers, 2013) [40].

1.3 *In vitro* Models for Intestinal Differentiation and Colon Cancer Cell Lines

Intestinal differentiation is an attractive model to study molecular dynamics in cell line models. Cell polarity; microvillar membrane assembly and permeability can be modeled in spontaneously differentiating Caco-2 cells and in chemically induced

differentiating HT-29 cells. is done with the human colon adenocarcinoma cell-lines Caco-2 and HT-29 established by J. Fogh [41]. Caco-2 cell lines can be grown in a proliferating state for long time, however upon cell-to-cell contact increase at the confluent phase, they undergo a differentiation phase. At the differentiated state, Caco-2 cells attach each other with tight junctions and develop polarized cell body and microvillar membrane (Figure 1.9) [38]. During differentiation, many molecular changes take place, including the up-regulation of Lactase, Sucrase Isomaltase and down-regulation of c-Myc [41], [42]. Another important characteristic of differentiated Caco-2 cells is that, they transport ions and water from apical to basolateral membrane, leading to formation of dome structures in culture [38].

In addition to Caco-2, HT-29 cell lines are also used to study colonocyte differentiation and unlike Caco-2 they are rather differentiated under the influence of culture medium changes or differentiation inducers. Although they might show colonocyte characteristics upon differentiation, we must note that they still bare mutations due to their cancer cell line background [38].

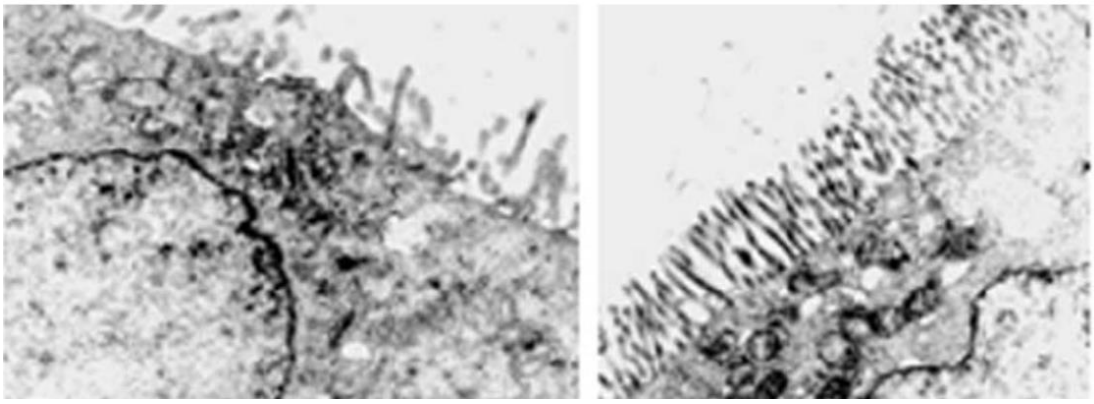


Figure 1.9. Microvilli structure on the membrane of proliferating (left) and differentiated (right) Caco-2 cells as illustrated by electron micrographs (Taken from Simon-Assman et al., 2007) [38].

1.4 Aim of the Study

APA is emerging as a novel mechanism that results with 3'UTR isoforms that may alter the stability, localization and function of the resulting proteins. Given that APA is detected during differentiation and development, we hypothesized APA to take place during enterocyte differentiation, which may have implications in colon cancers. Therefore, we chose the enterocyte differentiation model cell line, Caco-2 to begin investigating APA events in this model.

CHAPTER 2

MATERIALS AND METHODS

2.1 Datasets

GSE7745 (GEO, www.ncbi.nlm.nih.gov/geo/) was used to compare differentiated and undifferentiated Caco-2 cells for APA events (data accessible at NCBI GEO database [43], accession GSE7745). Dataset contains 3 replicates for pre-confluent (proliferating or undifferentiated) and day 10 post-confluent (differentiated) Caco-2 cells. (Table A.1)

2.2 Detection and quantification of APA events

We used APADetect tool [44], [45] for the detection and quantification of APA events in spontaneously differentiating Caco-2 cells. CEL files of Human Genome U133 Plus 2.0 arrays (HGU133Plus2, GPL570) were processed by APADetect to detect differential probe intensities. Poly(A) positions in PolyA_DB were used [46]. APADetect calculates SLR (Short to Long Ratio) values which are based on short and long isoforms and their corresponding probe intensities.

SLR values of differentiated cells were compared to that of undifferentiated Caco-2 cells. Significant APA events were determined by a fold change filter (SLR >1.5 for shortening events or SLR<0.66 for lengthening events) (Figure 2.1).

Graphpad Prism (California, USA) was used as a graph plotting tool. Scatter plots were drawn illustrate the individual SLRs for each APA event. Means of the pre-confluent and differentiated Caco-2 sample SLRs were compared using unpaired t-test. The workflow is shown in Figure 2.1.

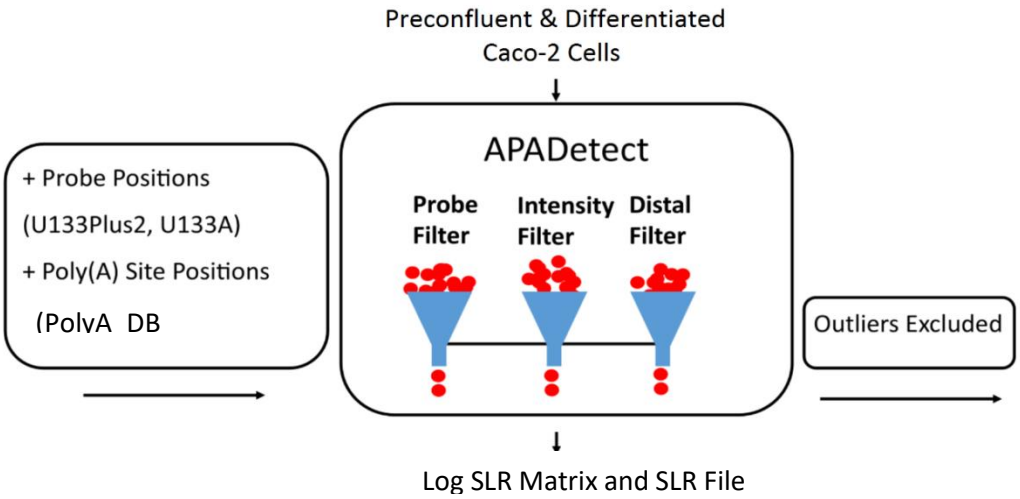


Figure 2.1. APADetect Pipeline. CEL files of microarray data for undifferentiated (3) and differentiated samples (3) were processed with the previously developed APADetect tool (Akman et al., 2012; Akman et al., 2015). PolyA_DB is the source for the Poly(A) genomic position information. Probe intensities grouped by poly(A) site positions were processed through probe, intensity and distal filters, which exclude outliers. LogSLR matrix file was the output of APADetect where individual APA events were assigned a short/long isoform ratio (SLR) value.

2.3 Gene Set Enrichment Analysis (GSEA)

For the APADetect output of Caco-2, the functional enrichment analysis was done using the web interface of Gene Set Enrichment Analysis (GSEA) (http://software.broadinstitute.org/gsea). Significant APA transcripts were listed and used as input to analyze the enriched biological processes. Processes having the largest -log(P) value were selected for illustration via Graphpad Prism (California, USA).

2.4 Cancer Cell Lines, Cell culture, Differentiation

Caco-2 cells were kind gifts from Dr. Sreeparna Banerjee and grown in Earle's minimum essential medium containing 1.5 g/L sodium bicarbonate, 1 mM sodium pyruvate, 2 mM L-glutamine, 0.1 mM non-essential amino acids, 20 % Fetal Bovine Serum (FBS) and 1 % penicillin-streptomycin. All cell culture supplements were obtained from Biochrom (Berlin, Germany). Spontaneous Caco-2 differentiation was induced by growing cells until confluency and full confluent cells were considered as Day 0. The cells were grown ten days after confluency and harvested at various intervals [47], [48]. Cell lines were cultured as monolayers and incubated at 37 ºC with 5% CO2 and 95% humidified air. Cells were frozen in liquid nitrogen at 70-80% confluency with 5% DMSO (dimethylsulfoxide) (Sigma, Cat# 154938) in order to store cells for long term. Cell pellets were obtained with by 1400 rpm centrifugation for 5 minutes. Cell thawing was done at 37°C.

2.5 Transfection

SNX3-sh in pSUPER (designed and cloned by Merve Öyken, Erson Lab) was used to generate stably transfected Caco-2 cells. Stable cell lines were maintained with 0.8 mg/ml Geneticin (Cat# 108321-42-2; Sigma-Aldrich). Transfections were performed with Lipofectamine LTX Reagent (ThermoFisher Scientific, CAT#15338100) according to manufacturer's manual.

2.6 Actinomycin D Treatment

Caco-2 cells were plated in 6-well plates. Actinomycin D (Abcam, CAT#Ab141058) solution was dissolved in DMSO (Sigma, Cat# 154938) at (1mg/mL) stock concentration. Treatment was done with a final concentration of 10 µg/mL in 2mL medium (specified above). Proliferating and differentiated cells were collected at 1, 2, 4, 8 and 12 hour intervals as treated and untreated samples.

2.7 Alkaline Phosphatase Staining

SNX3-sh Caco-2 cells were plated on a 6-well plate. When they reached 70% confluency, cells were washed twice with TBS buffer. Cells were fixed with 1mL of 70% molecular grade EtOHand incubated at RT for 10 minutes. Cells were washed twice with TBS. 500 µL NBT/BCIP was added to each well and cells were incubatedfor 2 hours. cells were washed 3 times with TBS. 1.5 mL TBS was added into each well and cells were visualized under phase contrast microscope. Recipes are in the appendix E.

2.8 RNA isolation and real time RT-PCR (RT-qPCR)

Total cellular RNA was isolated with Zymo Quick-RNA MiniPrep (CAT#R1055) and further processed with an overnight DNAse I enzyme treatment (Thermo Fisher Scientific, CAT#EN0521) according to the manufacturer's manual (Table 2.1). DNA contamination was checked by PCR with GAPDH primers (F: 5'GGGAGCCAAAAGGGTCATCA-3', R: 5'-TTTCTAGACGGCAGGTCAGGT-3'). 0.5-1 µg RNA was reverse transcribed by RevertAid First Strand cDNA Synthesis kit (Thermo Fisher Scientific, CAT# EP0441) using oligo-dT primers (Table 2.2) and stored at -20 ºC. Quantitative Real-Time PCR (RT-qPCR) reaction was performed using BioRAD SYBR Green Supermix (CAT#172-5270) with 0.5 µM forward and reverse primers and 1 µl cDNA. RT-qPCR Machine BioRAD CFX-Connect was used. Ct values were calculated using relative standard curve method and the fold change was calculated by Pfaffl method (Pfaffl, 2001). Human colon total RNA was purchased from Clontech (CAT#636553). The primers used in the study are given in Table 2.3.

Table 2.1. DNase I reaction mixture

RNA (4-5 ug/µl)	10 µl
10 X Reaction Buffer	10 µl
DNase I (1u/µl)	2 µl
DEPC water	variable
TOTAL	100 µl

Table 2.2. Reverse Transcription Reaction Conditions

RNA	500 ng (1-2 µl
Primer (oligodT or random hexamer)	1 µl
MG water	variable
TOTAL	12 µl
Briefly centrifuged and incubated for 5 minutes at 70 °C.	
5X Reaction Buffer	4 µl
Ribolock RNAse inhibitor	1 µl
dNTP mix	2 µl
RevertAid RT enzyme	1 µl
TOTAL	20 µl
Tubes were incubated for 60 minutes at 42°C; reaction was stopped by heating at 70°C for 5 minutes.	

Table 2.3. Primers used in PCR and RT-qPCR

Gene name	Primer Sequence	Product Length	Annealing Temperature
SI	5'-CAAATGGCCAAACACCAATG-3' 5'-CCACCACTCTGCTGTGGAAG-3'	160	59°C
MYC	5'-CAGCTGCTTAGACGCTGGATT-3' 5'-GTAGAAATACGGCTGCACCGA-3'	131	59°C
GAPDH qPCR	5'-CGACCACTTTGTCAAGCTCA-3' 5'-CCCCTCTTCAAGGGGTCTAC-3'	212	59°C
GAPDH PCR	5'-TGCCTTCTTGCCTCTTGTCT-3' 5'-TTGATTTTGGAGGGATCTCG-3'	472	59°C
SNX3 FS-RS	5'-GCCTGAAATTTGGCAAGAAG-3' 5'-TCTTGTCAACTGCCAAAACAA-3'	165	59°C
SNX3 FL2-RL2	5'-TCATTCCTGTAACTCCATTCCCT-3' 5'-GCAGTTTTCAAATACACAAAGTGCT-3'	165	59°C
TCF3 FS-RS	5'-CAAAACCTGAAAGCAAGCA-3' 5'-TTAGGCACAATTTGCTGGTG-3'	154	56°C
TCF3 FL2-RL2	5'-TTGCCTCTCCCTCTTGTTTT-3' 5'- CCCCCATAATTGTGGTTCC-3'	160	56°C
PNRC1 FS-RS	5'-GCTGGGGCAAAGTTTAGTGA-3' 5'-GAGTCCAGGGATATGGGAAAA-3'	201	61°C
PNRC1 FL-RL	5'-GTGTGTGCTAAGGCACATGGA-3' 5'-GAGAAACAAACCCCATGCTT-3'	172	61°C
CDC6 3′-UTR-short	5′-TTCAGCTGGCATTTAGAGAGC-3′ 5′-AAGGGTCTACCTGGTCACTTTT-3′	185	59°C
CDC6 3′-UTR-long	5′-TTCAGCTGGCATTTAGAGAGC-3′ 5′-CGCCTCAAAAACAACAACAA-3′	349	59°C
WNT5A	5′-AGGGCTCCTACGAGAGTGCT-3′ 5′-CTTCTCCTTCAGGGCATCAC-3′	185	59°C
WNT3	5'-TTCTTGGTCCACTCCCATTC-3' 5'-GAACACATGGCTGCTCTTCA-3'	178	59°C
TCF4	5′-CCTGGCTATGCAGGAATGTT-3′ 5′-CAGGAGGCGTACAGGAAGAG-3′	193	59°C
LEF	5′-ATATGATTCCCGGTCCTCCT-3′ 5′-TGAGGCTTCACGTGCATTAG-3′	121	59°C

2.9 RNA quantification

RNAs were quantified via BioDrop Duo (Isogen Life Science). RNA sample purity was, A260/A280 ratio was between 1.8 and 2 and A260/A230 ratio was higher detected by A260/A280 and A260/A230 ratios. RNA concentration was calculated using the following formula: RNA (μg/mL) = 40 X Dilution Factor X OD260. For all RNA samples than 1.8.

2.10 Protein Isolation and Western Blotting

Total cellular proteins were isolated with M-PER Mammalian Protein Extraction Reagent (Thermo Fisher Scientific, #78501) containing phosSTOP (Roche-CAT#04906837001) and protease inhibitor coctail (Roche-CAT#11873580001). Protein concentrations were determined with the BCA kit assay. Total protein extracts were denaturated with 6X Laemmli buffer (Appendix E) at 100°C for 10 minutes. The electrophoresis of the proteins was applied by using 5% stacking and 8-12% separating SDS-PAGE and subsequent electroblot was done onto PVDF membrane (Roche). 5% skim milk or 5% bovine serum albumin (BSA) in PBS-T (Phosphate Buffer Saline-Tween) (Appendix E) were used as blocking reagent at room temperature for 1 hour. Blocking was followed by overnight incubation with the primary antibodies: B-actin (1:2000) (Santa-Cruz, CAT# sc-47778), SNX3 (1:500) (PTG Lab, CAT# 10772-1-AP), CSTF2 (1:1000) (Abnova, CAT#H00001478-PW2), WLS (1:200) (Santa Cruz, CAT#655902) with a subsequent 1 hour incubation of secondary antibody (Santa Cruz, anti-mouse CAT#sc-2005; anti-rabbit CAT#sc-2030, 1:2000). Membranes were visualized by BioRAD Clarity™ Western ECL Blotting Substrates (CAT#1705060) according to the manufacturer's instructions.

CHAPTER 3

RESULTS AND DISCUSSION

3.1. Alternative polyadenylation isoforms in proliferating and differentiated Caco-2

To investigate whether alternative polyadenylation takes place in Caco-2 differentiation, we analyzed microarray datasets for pre-confluent (proliferating) and differentiated Caco-2 cells (data accessible at NCBI GEO database [43], accession GSE7745) via APADetect (Figure 2.1). The analysis resulted with 43 APA events with an SLR>1.5 (shortening) and 91 events with an SLR<0.66 (lengthening) (Figure 3.1A, B).

The APA events were then analyzed for functional enrichment via MSigDB-GSEA (Molecular Signature Database of Gene Set Enrichment Analysis) [49]. The significant events with the highest -log(P) value were selected for illustration. Interestingly, enriched biological processes included enzyme binding, endosomes and RNA processing (Figure 3.1 C).

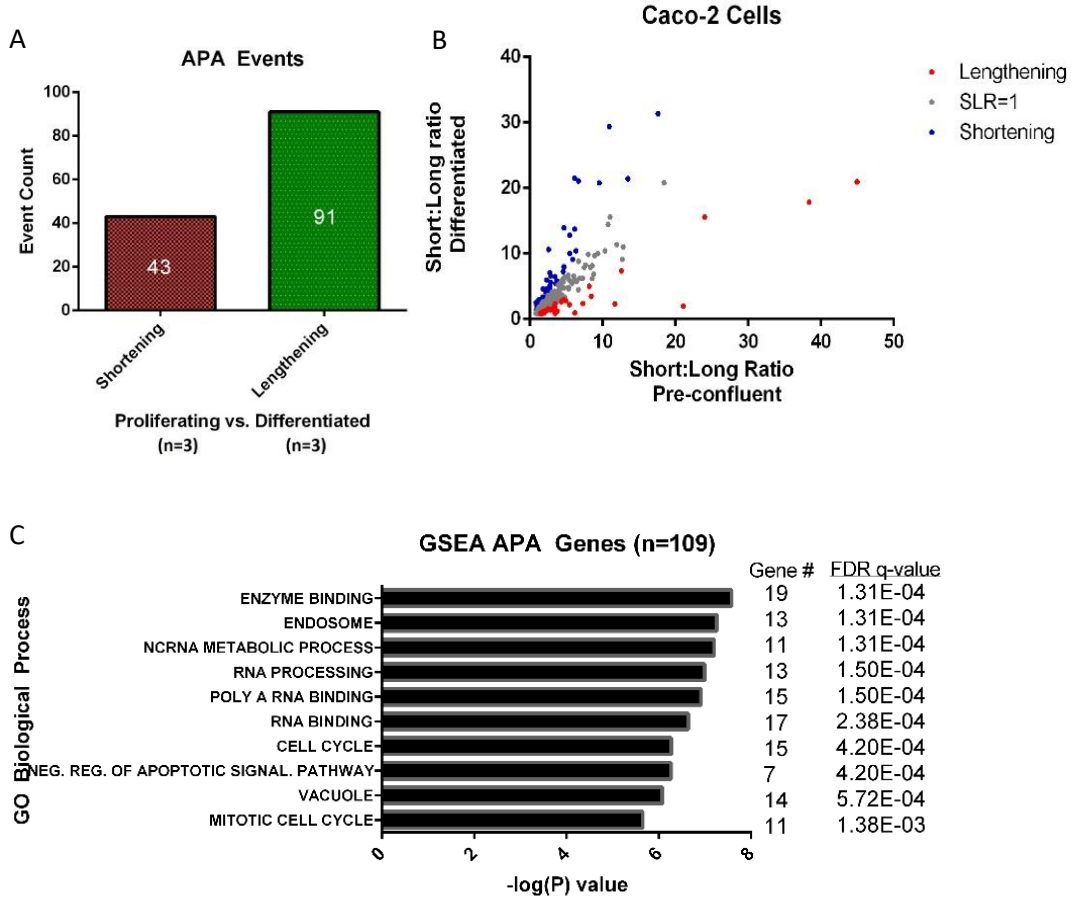


Figure 3.1. (A) A total of 134 APA events were detected in differentiation of Caco-2 cells; 43 for shortening and 91 for lengthening events when compared pre-confluent Caco-2 cells (3) with differentiated Caco-2 cells (3). (B) Volcano plot shows the distribution of all APA events with respect to their SLR ratios in pre-confluent and differentiated cells. X and Y axes represents Short: Long ratios of APA events in pre-confluent and differentiated samples, respectively. Red dots represent lengthening events; blue dots represent shortening events and gray dots represent insignificant events. (C) Functional Enrichment analysis by using MSigDB of Gene Set Enrichment Analysis (GSEA) showed enrichment in biological processes including enzyme binding, endosome and RNA processing with highest log(P) values. FDR q-values indicate the false-discovery rate for each biological process.

3.2. miRNA Binding Site Positions and Their Potential Importance

Next, we wanted to further investigate these APA events in terms of miRNA dependent regulation. Common form of APA is the shortening or the lengthening of 3'UTRs. Hence, altered 3'UTRs may either retain or lose cis-regulatory sequences such as miRNA binding sites. To investigate positions of potential miRNA binding sites on the APA isoforms, we used TargetScan [50] prediction program and calculated the relative distances of miRNAs to the active poly(A) sites. (Figure1.5) [28]. Specifically, for all 3'UTR shortening or lengthening events, we performed miRNA binding prediction in relation to the position of the selected poly(A) sites. According to our results; we observed an enrichment of 111 (among shortening) and 192 (among lengthening) miRNA sites at around 300 bases upstream of poly(A) sites (Figure 3.2). When we further investigated this enrichment, we detected three subgroups. Group "a" transcripts had an enrichment in miRNA binding sites within 300 bp upstream of the selected proximal poly(A) site. Group "b", on the other hand, had miRNA binding sites outside of the within 300 bp upstream region. Finally, group "c" did not have any miRNA binding site with 1000 bp upstream and downstream region of poly(A) site (Figure 3.3). Results indicate that miRNA binding site enrichment in the 300 bp upstream region is specific to a group of genes.

According to Hoffman et al., 2016 [28], miRNA binding sites are enriched within 300 bases upstream of the selected poly(A) site. Authors argue that the 3'UTR end of the transcript is more accessible to miRNAs. Our findings are in agreement with Hoffman et al., 2016. Enrichment of conserved miRNA binding sites upstream of the used poly(A) site means that, this region has an evolutionary conserved importance for

the miRNA mediated regulation. upon 3'UTR shortening by the selection of the indicated poly(A) site, miRNA binding site might become more accessible to miRNA mediated regulation. Therefore, even though miRNA level does not change between

two different states, 3'UTR shortening could affect miRNA function by changing the accessibility of the binding site on the target mRNA isoform.

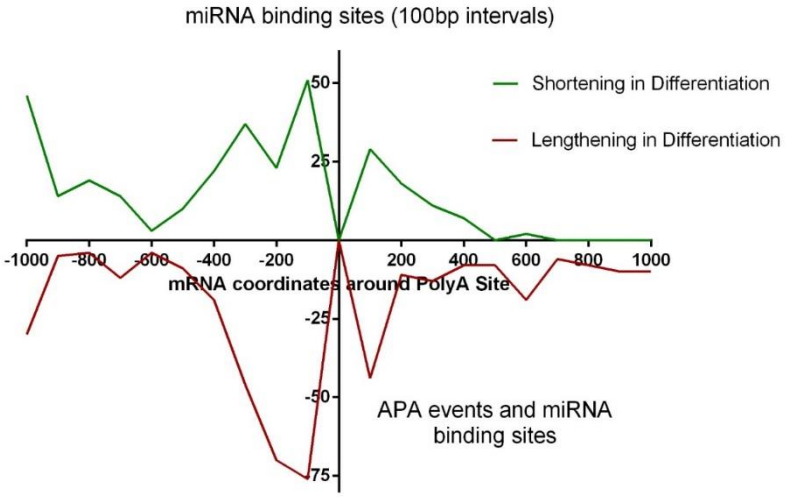


Figure 3.2. Conserved miRNA binding site predictios by TargetScan (Agarwal et al.,2015) given as a cumulative of all predictions within hundred bases upstream/downstream of poly(A) sites. The active poly(A) sites were detected using APADetect. A noticeable enrichment of miRNA binding sites was detected within 300 bases region upstream of the poly(A) site. Green line indicates miRNA binding sites on shortened transcripts and red line indicates miRNA binding sites on lengthened transcripts in differentiation. GraphPad software is used.

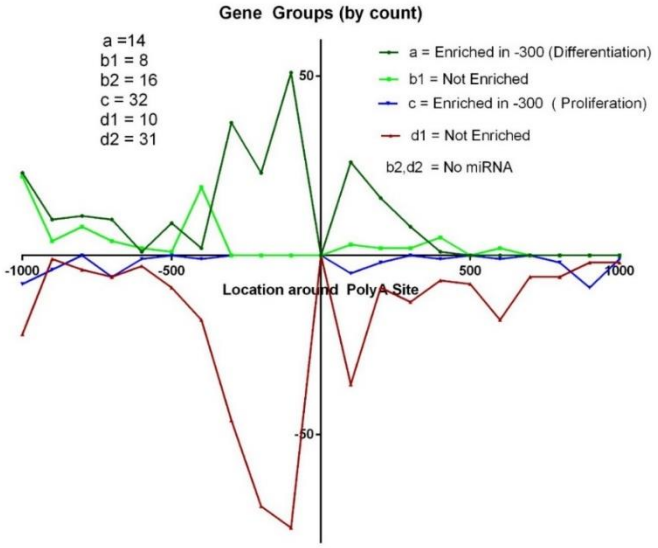


Figure 3.3. Gene groups based on miRNA binding site locations in their sequence. Genes with enrichment in their 0/-300 region are separated from genes with no enrichment. Another group of genes are with no enrichment in 1000/-1000 region. Generally, they are separated based on whether they are shortened in proliferation or differentiation.

Overall, APADetect results clearly indicated that major APA events are taking place in differentiation of Caco-2 cells. This finding has encouraged us to investigate these events *in vitro*, in order to confirm significant APA events, understand their implications and APA dynamics.

3.3. Differentiation of Caco-2 Cells

In order to confirm *in silico* APADetect analysis, we used Caco-2 cells. Caco-2 cells differentiated spontaneously approximately 10 days after reaching 100% confluency. These cells were examined during this time interval with an inverted microscope. As an indication of liquid accumulation and tight junctions between the cells, dome structures began to form starting from second day until the full differentiation [51]. The time-course microscope images and markings of dome structures are seen in Fig. 3.4.A

Besides morphological confirmation, differentiation was also confirmed by gene expression analysis of enterocyte differentiation markers. Sucrase isomaltase (SI) is

known to be transcriptionally up-regulated, whereas MYC (c-Myc) is known to be transcriptionally down-regulated with differentiation [41], [42]. We have confirmed this observation via RT-qPCR in proliferating and differentiated Caco-2 samples (Figure 3.4B-C). Dome structures indicate the colonocyte-like phenotype and SI indicate the upregulated enzymatic activity in brush borders of differentiating Caco-2 cells.

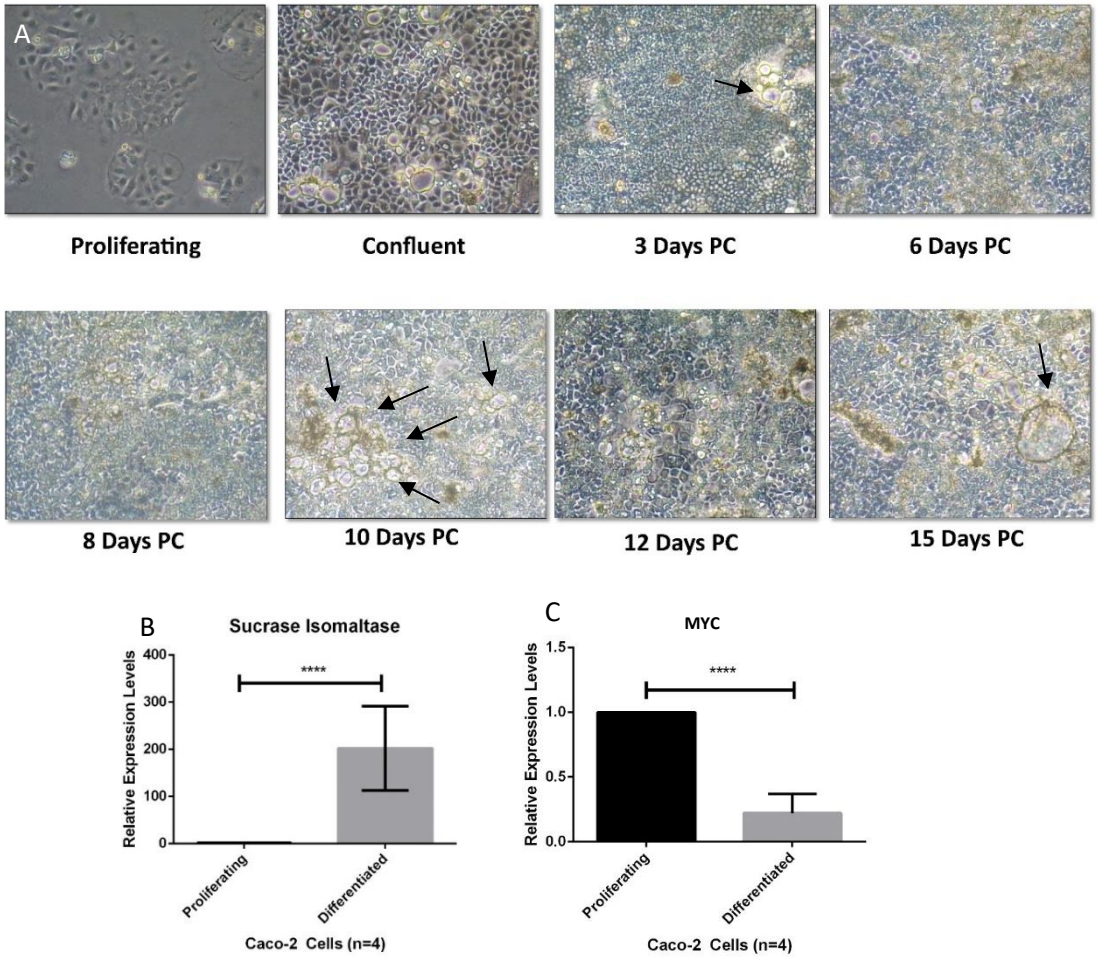


Figure 3.4. Differentiation experiment of Caco-2 cells. Microscope images of Caco-2 cell in enterocyte differentiation time-course experiment (40X). Black arrows indicate the formation of dome structures, formed as a result of water intake of enterocyte-like differentiated Caco-2 cells. PC implies post-confluency (A). Differentiation is confirmed by up-regulation of Sucrase Isomaltase (SI) mRNA (B) and down-regulation of MYC(c-Myc) mRNA (C), known differentiation markers. The difference between mRNA levels was analyzed by two-tailed non-parametric Mann Whitney test. **** indicates statistical significance ($p<0.0001$). Differentiation experiment was repeated with 4 independent biological replicates.

3.4. *In vitro* Confirmation of Identified APA Events

Once we confirmed upregulation of known differentiation markers in our model, we wanted to confirm the *in silico* data using this model. Therefore, we selected significant shortening and lengthening events to experimentally confirm *in vitro*.

PNRC1 (Proline-rich nuclear receptor coactivator1) was a candidate mRNA that had 3'UTR shortening in differentiating Caco-2 cells detected by APADetect. Scatter plot graph of *PNRC1* was drawn using log Short: Long (SLR) values of 3 preconfluent and 3 differentiated Caco-2 samples, detected by APADetect (Figure 3.5. A). In order to confirm this *in silico* analysis*in vitro*, we have used our Caco-2 differentiation model for RT-qPCR analysis. In APADetect, the SLR was 10.64 in differentiating cells, and we detected 1.6-fold increase in SLR of *PNRC1* gene in Caco-2 cells detected by RT-qPCR (Figure 3.5. B). Interestingly, in addition to confirming the APA event in vitro ($p<0.0006$), we have seen a dramatic upregulation in the short and long *PNRC1* mRNA levels ($p<0.0001$) (Figure 3.5. C, D).

PNRC1 is known to be a coactivator of nuclear receptors including SF1 (Steroidogenic Factor 1) and ERRa1 (Estrogen Related Receptor a-1) [52]. Besides its reported roles in the nucleus, it was also shown to interact with Grb2 (Growth factor receptor-bound protein 2) in the cytosol and consequently inhibit the further activation of growth factor/Ras-mediated MAP kinase pathways [52].

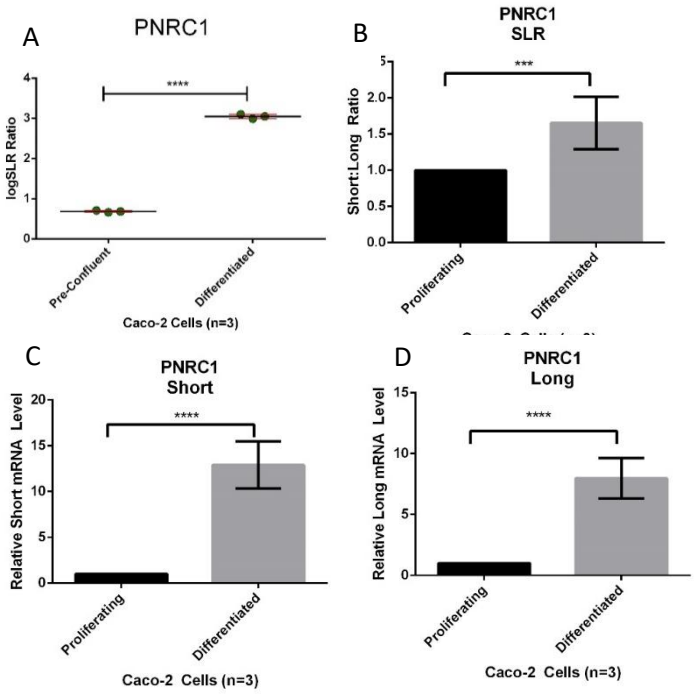


Figure 3.5. APADetect analysis output for *PNRC1* and *in vitro* confirmation. (A) LogSLR values of PNRC-1 for preconfluent and differentiated Caco-2 samples. (B) RT-qPCR results Short:Long ratio, (C) short mRNA level and (D) long mRNA level. *** ($p<0.0006$) and ***** ($p<0.0001$) indicate statistical significance analyzed with two-tailed non-parametric Mann Whitney test.

TCF3 was another signature APA regulated mRNA in the *in silico* analysis. Scatter plot graph of *TCF3* was drawn using log Short: Long (SLR) values of 3 preconfluent and 3 differentiated Caco-2 samples, detected by APADetect tool. The logSLR values significantly ($p <0.0001$) decreased in differentiation, an indication of lengthening event (Figure 3.6. A). RT-qPCR analysis of *TCF3* indeed resulted with 0.7-fold lengthening event in differentiation, compared to proliferating Caco-2 cells (Figure 3.6. B). Short mRNA level decreased with a low significance, while long mRNA level did not change significantly (Figure 3.6. C, D).

TCF3 is a transcription factor involved in WNT/B-catenin pathway, which also controls *MYC* expression in colorectal cancer cells. A potential APA mediated regulation might affect the WNT/B-catenin pathway and MYC expression, which is known to be involved in enterocyte differentiation [42].

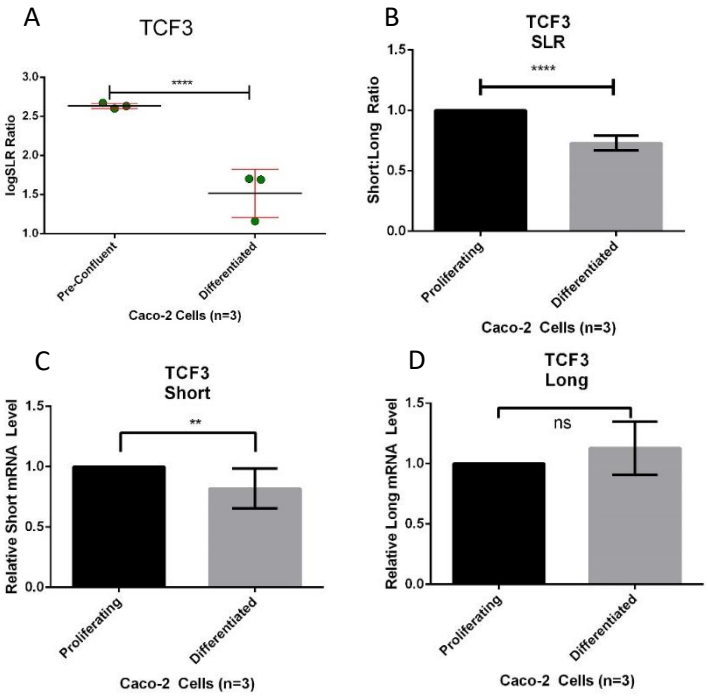


Figure 3.6. APADetect analysis output for *TCF3* and *in vitro* confirmation. (A) LogSLR values of *TCF-3* for preconfluent and differentiated Caco-2 samples. (B) RT-qPCR results of Short: Long ratio, (C) short mRNA level and (D) long mRNA level. ***** ($p<0.0001$) and ** ($p<0.0019$) indicate statistical significance and ns indicates not-significant analyzed with two-tailed non-parametric Mann Whitney test.

The third test gene was *SNX3*, which was determined to undergo lengthening in its 3'UTR in differentiation, compared to proliferating Caco-2 cells. Scatter plot graph of *SNX3* was drawn using log Short: Long (SLR) values of 3 preconfluent and 3 differentiated Caco-2 samples, detected by APADetect tool (Figure 3.7. A). RT-qPCR analysis of *SNX3* indeed resulted with 0.5-fold lengthening event in differentiation, compared to proliferating Caco-2 cells (Figure 3.7. B). Furthermore, there was a significant increase of *SNX3* short and long mRNA levels (Figure 3.7. C, D).

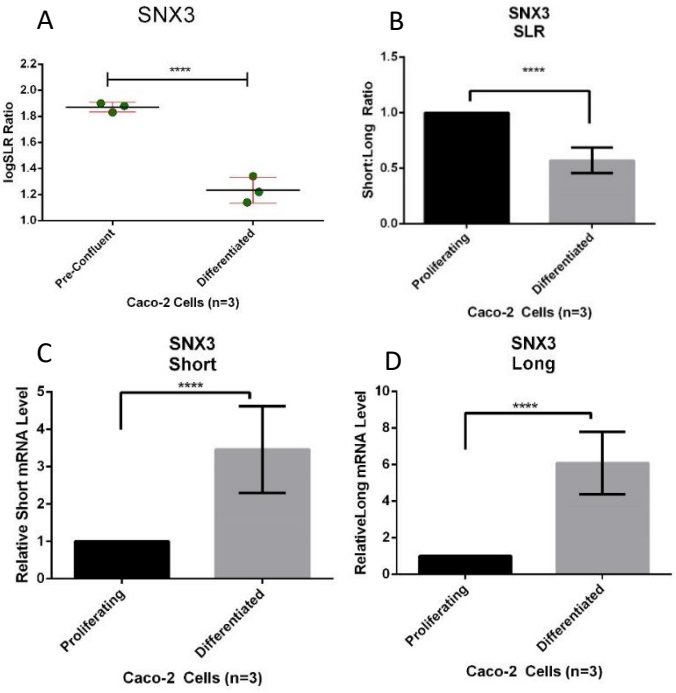


Figure 3.7. APADetect analysis output for *SNX3* and *in vitro* confirmation. (A) LogSLR values of *SNX3* for preconfluent and differentiated Caco-2 samples. (B) RT-qPCR results of Short: Long ratio, (C) short mRNA level and (D) long mRNA level. ***** ($p<0.0001$) indicates statistical significance analyzed with two-tailed non-parametric Mann Whitney test.

Finally, as a negative control, we picked *CDC6* (Cell division cycle 6) gene that was not regulated by APA, according to APADetect tool (Figure 3.8. A). The RT-qPCR experiment showed the same pattern, confirming the *in silico* result (Figure 3.8. B). It is also important to note that there was a significant down-regulation of CDC6 short and long RNA levels (Figure 3.8. C-D).

Following the confirmation of the *in silico* analysis in a small group of test genes, we decided to focus on SNX3. *SNX3* is involved in the c group, according to the miRNA binding sites located on its 3'UTR, meaning that it has miRNA binding sites at the upstream region of its detected poly(A) site.

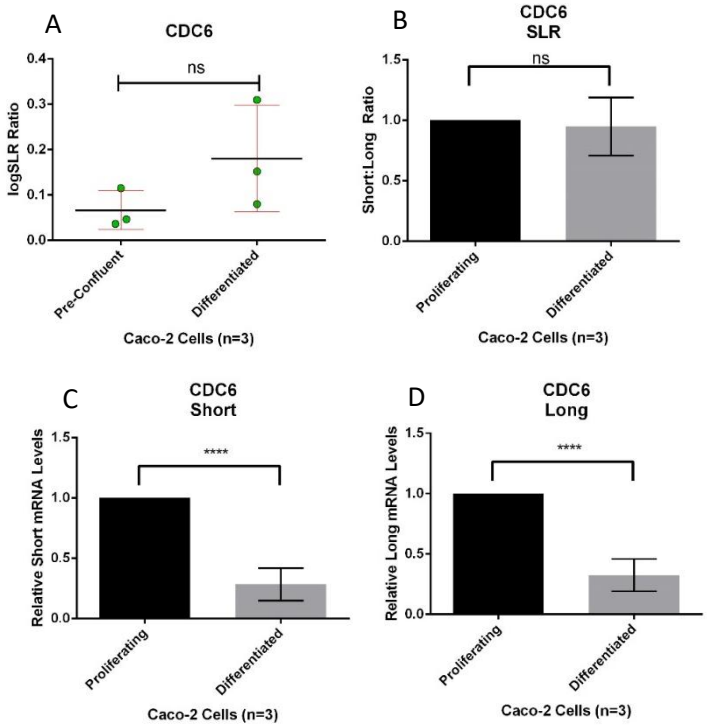


Figure 3.8. APADetect analysis output for *CDC6* and *in vitro* confirmation. (A) LogSLR values of *CDC6* for preconfluent and differentiated Caco-2 samples. (B) RT-qPCR results of Short: Long ratio, (C) short mRNA level and (D) long mRNA level. ***** ($p<0.0001$) indicates statistical significance and ns indicates not significant, analyzed with two-tailed non-parametric Mann Whitney test.

SNX3 is a sorting nexin protein involved in retromer complex, responsible for endosomal traficking [53]. It has also been implicated in the regulation Wnt ligand secretion through Wls (Wntless) recycling pathway. Wls is known to be responsible for the transportation of Wnt ligands from Golgi network to the membrane for the secretion [54]. In addition to its role in recycling the Wls, SNX3 has also been shown to recycle other receptors, including EGFR [55].

Caco-2 cells form a polarized epithelium structure when differentiated. This process is affected by directed endosomal trafficking [56], [57]. It has also been shown that this process is mediated by Wnt pathway, which provides cell polarity and cytoskeletal

changes [58]. Therefore, considering that SNX3 might regulate WNT signaling, we checked protein levels of SNX3 protein levels in proliferating and differentiated Caco-2 cells. Interestingly, we saw an increase (1.9-fold) in the protein levels in differentiation, despite of the lengthening event we detected in differentiated cells, which is usually attributed to negative regulation [33], [45] (Figure 3.9. A). The mRNA level of SNX3 increases during the course of differentiation. (Figure 3.9. B).

Next, we wanted to examine whether 3'UTR short and long isoforms have different stabilities to possibly explain increased protein levels despite 3'UTR lengthening. Actinomycin treatment in proliferating vs. differentiated cells did not reveal significant difference in the stabilities of the isoforms (Figure 3.9. C, D). Our current explanation is that, although there is 3'UTR lengthening, increased transcriptional upregulation seems to be causing the increase in protein levels.

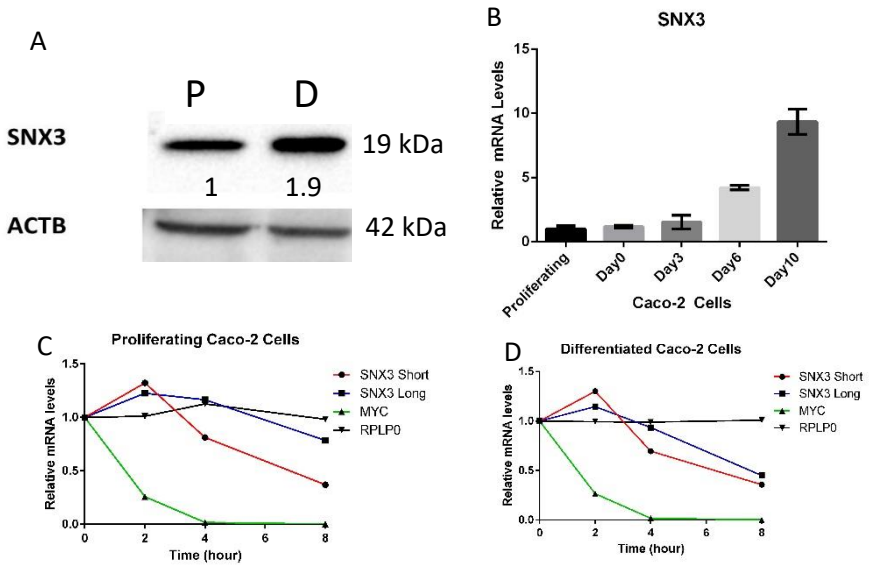


Figure 3.9. (A)Protein levels determined by Western Blot and subsequent densitometry analysis in Caco-2 differentiation model. (B) RT-qPCR results of SNX3 mRNA in time-course differentiation. Relative mRNA stabilities of MYC, SNX3 short and SNX3 long mRNAs in proliferating (C) and differentiated (D) Caco-2 cells upon actinomycin-D treatment. MYC is used as a positive control for the actinomycin-D treatment, while RPLP0 was used as a reference gene, whose expression did not change with actinomycing treatment. The treatment was repeated for once (n=1), while RT-qPCR was repeated three times.

We also checked other colon cancer cell lines for short (Figure 3.10. A), long (Figure 3.10. B) mRNA levels, SLR values (Figure 3.10. C) and protein expression (Figure 3.10. D) of SNX3. Interestingly, while SLR was generally high in 6 of 9 colon cancer cell lines, when we detected SNX3 protein levels, we saw variable expression. Perhaps, most interestingly Caco-2 proliferating vs. differentiated and SW480 vs. SW620 cell lines had opposite patterns of SLR and protein levels (Figure 3.10. E). SW480 is a nonmetastatic cell lines, while SW620 is a metastatic cell line, taken from the same patient. When SLR values were compared in SW480 and SW620, there was lengthening in the metastatic cell line, compared to non-metastatic one, however with a decreased SNX3 protein levels.

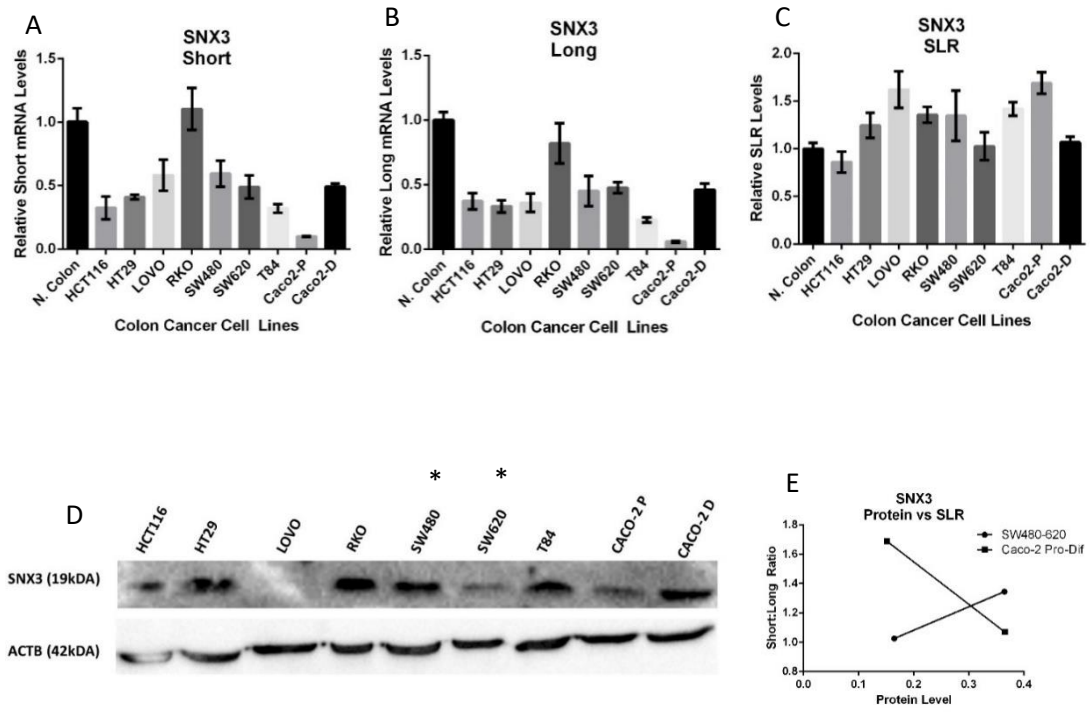


Figure 3.10. RT-qPCR results of (A) short, (B) long mRNA level and (C) Short: Long ratio in colon cancer cell lines. (D) Protein levels as determined by Western Blot, beta-actin is used as loading control. (E) Protein level vs SLR in SW480-620 and Caco-2 proliferating-differentiated states show opposite patterns. The protein level positively correlates with the SLR value in SW480-SW480, while negatively correlates with in proliferating-differentiated Caco-2. * shows the different SNX3 protein levels in SW480 and SW620 cell lines.

The discordany between decreased SLR values and increased protein levels suggest existence of as of yet unknown post-transcriptional or post-translations regulations on SNX3.

3.5. SNX3 Silencing Experiments

To understand the role of SNX3 gene in the Caco-2 cells, we silenced *SNX3* mRNA with shRNA in proliferating Caco-2 cells and checked the cells for the specific mRNA and protein levels, as well as the alkaline phosphatase staining. In our previous results, we have shown that SNX3 mRNA and protein levels increase with differentiation. First, we have confirmed the silencing at mRNA level with a 20-fold decrease (Figure 3.11. A), as well as at protein level (Figure 3.11. B). Interestingly, Wls protein levels increased significantly in SNX3 silenced cells (Figure 3.11. C). However, we have speculated that this could be due to an upregulation of *Wls* mRNA, which might be caused by a feedback regulatory mechanism (Figure 3.11. D).

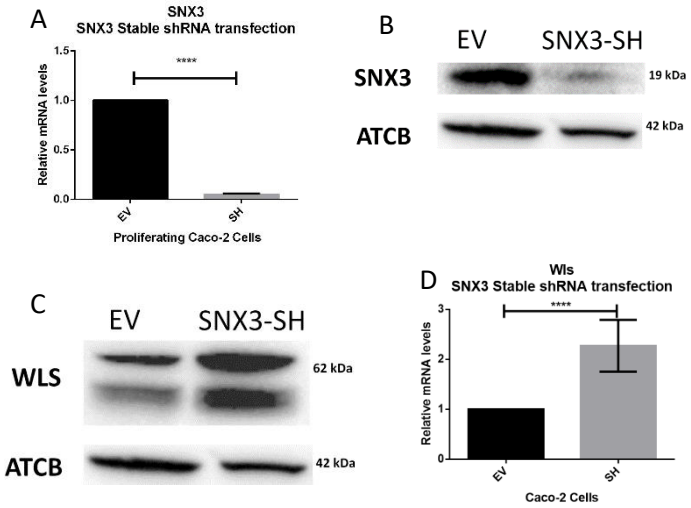


Figure 3.11. (A)RT-qPCR results of *SNX3* mRNA in EV and SNX3-shRNA transfected Caco-2 cells. (B) SNX3 and ACTB protein levels detected by immunoblotting. (C) Wls and ACTB protein levels detected by immunoblotting. (D) RT-qPCR results of *Wls* mRNA in EV and SNX3-shRNA transfected Caco-2 cells). ***** ($p<0.0001$) indicates statistical significance, analyzed with two-tailed non-parametric Mann Whitney test.

Next, we wanted to see the expression levels of representative WNT signaling related genes to understand whether SNX3 silencing indeed has an effect on WNT signaling.

Interestingly, differentiation related genes such as *SI* [59]and *WNT5A* [60] showed an upregulation in SNX3 silenced Caco-2 cells (Figure 3.12. A, B). On the other hand, when we investigated the WNT/β-catenin related genes such as *TCF4*, *WNT3* and *LEF1* [61], [62], we did not see a significant change in the mRNA of *TCF4* and *WNT3,* whereas we saw an upregulation of *LEF1* mRNA (Figure 3.12. C, D, E).

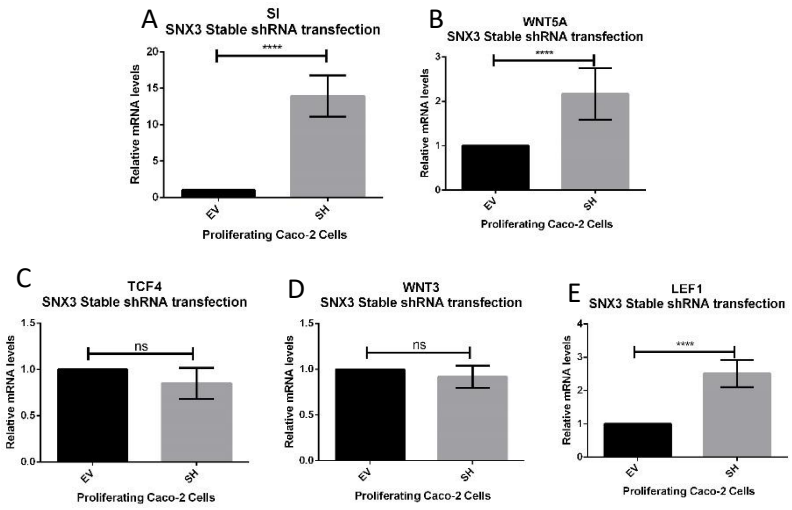


Figure 3.12. RT-qPCR results (A) *SI*, (B) *WNT5A*, (C) *TCF4*, (D) *WNT4*, (E) *LEF1* mRNA levels in EV and SNX3-shRNA transfected Caco-2 cells. ***** ($p<0.0001$) indicates statistical significance and ns indicates not significant, analyzed with two-tailed non-parametric Mann Whitney test.

In order to further examine any differentiation related change in the SNX3 silenced Caco-2 cells, we used alkaline phosphatase (AP) staining for the detection of enzyme presence. Intriguingly, SNX3 silenced cells had a more intensive AP staining than EV cells (Figure 3.13).

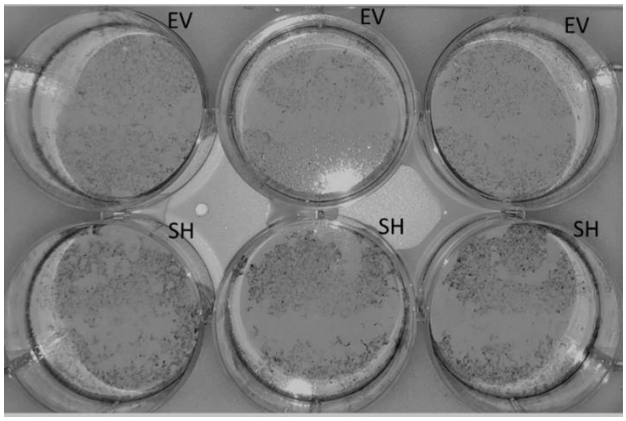


Figure 3.13. Alkaline phosphatase staining of the EV and SNX3-shRNA transfected Caco-2 cells seeded in 6-well. Proliferating cells were plated to become 60 % confluent at the day of staining. This experiment was applied once (n=1).

Based on these observations, we could suggest that SNX3 could be potentially cancer regulator of WNT signaling in this enterocyte differentiation model. Functional assays are underway to better understand how SNX3 might contribute to this how protein levels of SNX3 might be regulated.

3.6. APA Machinery gene expression

Finally, to begin understanding how APA may operate in Caco-2 differentiation model, we initially screened mRNA levels of the polyadenylation machinery proteins in proliferating and differentiated Caco-2 cells. Some polyA machinery proteins are known to influence the selection of proximal or distal polyA sites. For example, CSTF2 protein and CPSF factors have been implicated in 3'UTR shortening by inducing the selection of proximal sites [12], [15]. Interestingly, *CSTF1*, *CSTF2* and *CPSF73* had decreased expression in differentiated Caco-2 cells. However when we checked protein levels of CSTF2, we did not observe a significant change in the protein levels in differentiated compared to proliferating Caco-2 cells, suggesting that other

post-transcriptional and post-translational mechanisms to be highly active in differentiation. ROther mechanisms including epigenetics or involvement of auxillary proteins might be contributing to APA regulation in differentiation process (Figure 3. 14)

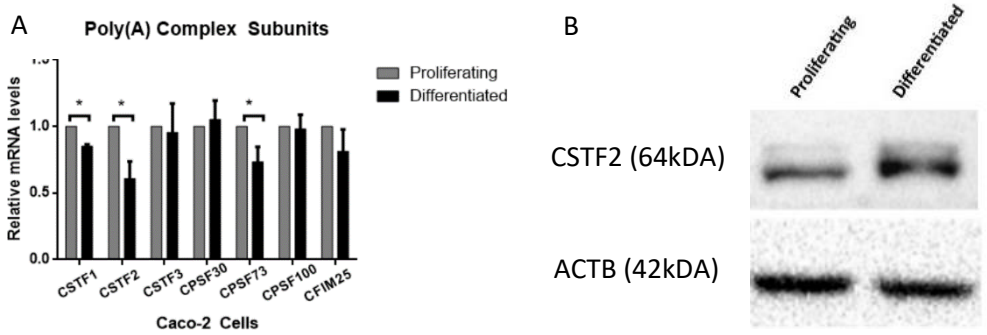


Figure 3.14. (A) mRNA expression of polyadenylation complex subunit proteins determined by quantitative real-time PCR (qRT-PCR). * indicates statistical significance analyzed with t-test (<0.0001 for CSTF1/2, 0.00039 for CPSF73). (B) CSTF2 protein levels of proliferating and differentiated Caco-2 cells, where ACTB is the loading control shown with Western-Blot.

CHAPTER 4

CONCLUSION

In this study, we have aimed to understand the importance and consequence of APA events in differentiation, By using microarray data for pre-confluent and differentiated Caco-2 cells, we have analyzed APA events via APADetect tool. As a result of APADetect analysis, we have observed 43 APA events with an SLR>1.5 (shortening) and 91 events with an SLR<0.66, among which are genes involved in enzyme binding, endocytosis and RNA processing, implicating that APA might be an important event in differentiation process. Having more lengthening events in differentiation is also consistent with the recent experiments, indicating the proliferative phenotypes show a higher proximal poly(A) site usage [63], while differentiation and development favor lengthening events transcriptome-wide [11].

Furthermore, we have analyzed our APA genes for the conserved miRNA binding sites around the detected poly(A) sites via TargetScan [50]. Interestingly, consistent with a previous study [28], we have identified an enrichment at around the 300 bp upstream region of the detected poly(A) site for a group of genes. This finding might implicate a conserved functionality of the miRNA binding sites on this region, which possibly changes its accesibility upon cleavage from poly(A) site. Observation of the secondary structure short and long isoforms for each APA gene might be necessary to validate this hypothesis.

In order to validate our *in silico* analysis, we used *in vitro* model Caco-2 cell lines. As a first step, we have confirmed differentiation by the characteristic dome structures and analyzing the mRNA levels of the marker genes, *SI* and *MYC*.

Following the confirmation of differentiation, we selected and confirmed three significant APA events (*PNRC1*, *TCF3*, and SNX3), as well as one insignificant event (*CDC6*). Confirming the *in silico* APADetect events with the *in vitro* model, we have demonstrated the validity of APADetect tool, developed by our group.

Furthermore, we have aimed to specifically investigate the SNX3 gene in the differentiation model, since it is one of the significant APA events and is known to be involved in the endocytosis, which is an important process in differentiation (REF). For this, we first have observed that SNX3 protein level increased significantly in differentiated Caco-2 cells, although the lengthening process took place in its 3'UTR upon differentiation. Since this is an interesting result, considering that 3'UTR lengthening is usually implicated with a decreased protein output, we have investigated its short and long mRNA levels in proliferating and differentiated Caco-2 cells via Actinomycin-D treatment. This has provided evidence that, mRNA stability is not affected significantly and therefore transcription level, rather than APA might be the determinant of the SNX3 protein levels in Caco-2 differentiation model. Next, we have checked the short and long mRNA isoform levels and protein levels of SNX3 in colon cancer cell lines to determine the patterns in distinct cell lines. We have observed interesting correlations between SLR value and protein level in these cell lines, indicating that SNX3 is regulated distinctly in different colon cancer cell lines. Specifically, we have observed an opposite SLR vs. protein level in SW cells and Caco-2 cells, showing that SLR value could result in a significantly different protein levels, depending on the cellular background of the cancer cells.

In order to investigate the role of SNX3 protein in the Caco-2 differentiation model, we have silenced *SNX3* mRNA with an shRNA, targeting its coding sequence. When we have checked the protein levels of the Wls, a protein known to be recycled by SNX3, we have observed a pattern that is contradictory to what has been reported before. Upon silencing the *SNX3*, WLS protein had increased. Furthermore, when we have observed an upregulation on *WLS* mRNA, we speculated that increased protein

level could be due to a transcriptional upregulation, rather than a protein level regulation. Furthermore, we have checked mRNA levels of differentiation related genes *SI* and *WNT5A* and observed an upregulation. On the other hand, WNT/B-catenin related genes *TCF4*, *WNT3* were unchanged at mRNA level, while *LEF1* was upregulated in the *SNX3* silenced cells. Overall, these results indicated that SNX3 might play a role in differentiation related pathways.

Further experiments are needed to understand how SNX3 protein levels are regulated and the significance of SNX3 in WNT-related differentiation pathways.

REFERENCES

[1] N. J. Proudfoot, “Ending the message : poly (A) signals then and now,” *Genes Dev.*, vol. 25, pp.a 1770–1782, 2011.

[2] E. Beaudoing, “Patterns of Variant Polyadenylation Signal Usage in Human Genes,” vol. 10, pp. 1001–1010, 2000.

[3] B. Tian and J. L. Manley, “Alternative cleavage and polyadenylation: The long and short of it,” *Trends Biochem. Sci.*, vol. 38, no. 6, pp. 312–320, 2013.

[4] R. Elkon, A. P. Ugalde, and R. Agami, “Alternative cleavage and polyadenylation: extent, regulation and function,” *Nat. Rev. Genet.*, vol. 14, no. 7, pp. 496–506, 2013.

[5] U. Kühn, M. Gündel, A. Knoth, Y. Kerwitz, S. Rüdel, and E. Wahle, “Poly(A) tail length is controlled by the nuclear Poly(A)-binding protein regulating the interaction between Poly(A) polymerase and the cleavage and polyadenylation specificity factor,” *J. Biol. Chem.*, vol. 284, no. 34, pp. 22803–22814, 2009.

[6] Z. Xia *et al.*, “Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3’-UTR landscape across seven tumour types.,” *Nat. Commun.*, vol. 5, p. 5274, 2014.

[7] A. J. Wood *et al.*, “Regulation of alternative polyadenylation by genomic imprinting,” *Genes Dev.*, vol. 22, pp. 1141–1146, 2008.

[8] and C. B. B. Noah Spies, Cydney B. Nielsen, Richard A. Padgett, “Biased Chromatin Signatures Around Polyadenylation Sites and Exons,” *October*, vol. 36, no. 2, pp. 245–254, 2010.

[9] Y. Cui and C. L. Denis, “In Vivo Evidence that Defects in the Transcriptional Elongation Factors RPB2, TFIIS, and SPT5 Enhance Upstream Poly(A) Site Utilization,” *Mol. Cell. Biol.*, vol. 23, no. 21, pp. 7887–7901, 2003.

[10] P. A. B. Pinto *et al.*, “RNA polymerase II kinetics in *polo* polyadenylation signal selection,” *EMBO J.*, vol. 30, no. 12, pp. 2431–2444, 2011.

[11] Z. Ji, J. Y. Lee, Z. Pan, B. Jiang, and B. Tian, “Progressive lengthening of 3’ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development,” *Proc. Natl. Acad. Sci.*, vol. 106, no. 17, pp. 7028–7033, 2009.

[12] Y. Takagaki, R. L. Seipelt, M. L. Peterson, and J. L. Manley, “The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation,” *Cell*, vol. 87, no. 5, pp. 941–952, 1996.

[13] C. Yao *et al.*, “Overlapping and distinct functions of CstF64 and CstF64τ in mammalian mRNA 3’ processing.,” *RNA*, vol. 19, no. 12, pp. 1781–90, 2013.

[14] G. Martin, A. R. Gruber, W. Keller, and M. Zavolan, “Genome-wide Analysis of Pre-mRNA 3’ End Processing Reveals a Decisive Role of Human Cleavage Factor I in the Regulation of 3’ UTR Length,” *Cell Rep.*, vol. 1, no. 6, pp. 753–763, 2012.

[15] J. G. Hardy and C. J. Norbury, “Cleavage factor Im (CFIm) as a regulator of alternative polyadenylation,” *Biochem. Soc. Trans.*, vol. 44, no. 4, pp. 1051–1057, 2016.

[16] W. Li *et al.*, “Systematic Profiling of Poly(A)+ Transcripts Modulated by Core 3’ End Processing and Splicing Factors Reveals Regulatory Rules of Alternative Cleavage and Polyadenylation,” *PLoS Genet.*, vol. 11, no. 4, pp. 1–28, 2015.

[17] B. Lackford *et al.*, “Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal,” *EMBO J.*, vol. 33, no. 8, pp. 878–889, 2014.

[18] R. Batra *et al.*, “Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease,” *Mol. Cell*, vol. 56, no. 2, pp. 311–322, 2014.

[19] M. G. Berg *et al.*, “U1 snRNP determines mRNA length and regulates isoform expression,” *Cell*, vol. 150, no. 1, pp. 53–64, 2012.

[20] Y. Katz, E. T. Wang, E. M. Airoldi, and C. B. Burge, “Analysis and design of RNA sequencing experiments for identifying isoform regulation.,” *Nat. Methods*, vol. 7, no. 12, pp. 1009–15, 2010.

[21] F.-A. Bava *et al.*, “CPEB1 coordinates alternative 3’-UTR formation with translational regulation.,” *Nature*, vol. 495, no. 7439, pp. 121–5, 2013.

[22] W. Dai, G. Zhang, and E. V. Makeyev, “RNA-binding protein HuR autoregulates its expression by promoting alternative polyadenylation site usage,” *Nucleic Acids Res.*, vol. 40, no. 2, pp. 787–800, 2012.

[23] K. Oktaba *et al.*, “ELAV Links Paused Pol II to Alternative Polyadenylation in the Drosophila Nervous System,” *Mol. Cell*, vol. 57, no. 2, pp. 341–348, 2014.

[24] A. E. Erson-Bensan, “Alternative polyadenylation and RNA-binding proteins,” *J. Mol. Endocrinol.*, vol. 57, no. 2, pp. F29–F34, 2016.

[25] A. E. Erson-Bensan and T. Can, “Alternative Polyadenylation: Another Foe in Cancer,” *Mol. Cancer Res.*, vol. 14, no. 6, pp. 507–517, 2016.

[26] C. Mayr and D. P. Bartel, “Widespread shortening of 3’UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells,” vol. 138, no. 4, 2010.

[27] A. R. Gruber *et al.*, “Global 3’ UTR shortening has a limited effect on protein abundance in proliferating T cells,” *Nat Commun*, vol. 5, p. 5465, 2014.

[28] Y. Hoffman *et al.*, “3’UTR Shortening Potentiates MicroRNA-Based Repression of Pro-differentiation Genes in Proliferating Human Cells,” *PLoS Genet.*, vol. 12, no. 2, pp. 1–15, 2016.

[29] C. Andreassi and A. Riccio, “To localize or not to localize: mRNA fate is in 3’UTR ends,” *Trends Cell Biol.*, vol. 19, no. 9, pp. 465–474, 2009.

[30] B. D. Berkovits and C. Mayr, “Alternative 3’ UTRs act as scaffolds to regulate membrane protein localization.,” *Nature*, vol. 522, no. 7556, pp. 363–367, 2015.

[31] P. Miura *et al.*, “Widespread and extensive lengthening of 3 ′ UTRs in the mammalian brain Widespread and extensive lengthening of 3 9 UTRs in the mammalian brain,” pp. 812–825, 2013.

[32] I. Ulitsky *et al.*, “Extensive alternative polyadenylation during zebrafish development,” *Genome Res.*, vol. 22, no. 10, pp. 2054–2066, 2012.

[33] W. Li, J. Y. Park, D. Zheng, M. Hoque, G. Yehia, and B. Tian, “Alternative cleavage and polyadenylation in spermatogenesis connects chromatin regulation with post-transcriptional control,” *BMC Biol.*, vol. 14, no. 1, p. 6, 2016.

[34] S. M. Blazie, C. Babb, H. Wilky, A. Rawls, J. G. Park, and M. Mangone, “Comparative RNA-Seq analysis reveals pervasive tissue-specific alternative polyadenylation in Caenorhabditis elegans intestine and muscles.,” *BMC Biol.*, vol. 13, no. 1, p. 4, 2015.

[35] Z. Ji and B. Tian, “Reprogramming of 3’ untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types.,” *PLoS One*, vol. 4, no. 12, p. e8419, 2009.

[36] A. Humphries and N. a Wright, “Colonic crypt organization and tumorigenesis.,” *Nat. Rev. Cancer*, vol. 8, no. 6, pp. 415–24, 2008.

[37] P. De Santa Barbara, G. R. Van Den Brink, and D. J. Roberts, “Development and differentiation of the intestinal epithelium,” *Cell. Mol. Life Sci.*, vol. 60, no. 7, pp. 1322–1332, 2003.

[38] P. Simon-Assmann, N. Turck, M. Sidhoum-Jenny, G. Gradwohl, and M. Kedinger, “In vitro models of intestinal epithelial cell differentiation,” pp. 241–256, 2007.

[39] D. Louvard, M. Kedinger, and H. P. Hauri, “The Differentiating Intestinal Epithelial Cell: Establishment and Maintenance of Functions Cellular Structures,” 1992.

[40] H. Clevers, “The intestinal crypt, a prototype stem cell compartment,” *Cell*, vol. 154, no. 2, pp. 274–284, 2013.

[41] E. H. Van Beers, R. H. Al, E. H. Rings, a W. Einerhand, J. Dekker, and H. a Büller, "Lactase and sucrase-isomaltase gene expression during Caco-2 cell differentiation.," *Biochem. J.*, vol. 308 (Pt 3), pp. 769–775, 1995.

[42] B. D. Leoni *et al.*, "Differentiation of Caco-2 cells requires both transcriptional and post-translational down-regulation of Myc," *Differentiation*, vol. 83, no. 3, pp. 116–127, 2012.

[43] R. Edgar, "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository," *Nucleic Acids Res.*, vol. 30, no. 1, pp. 207–210, 2002.

[44] B. H. Akman, M. Oyken, T. Tuncer, T. Can, and A. E. Erson-Bensan, "3' UTR Shortening and EGF signaling: Implications for breast cancer," vol. 90, no. 312, pp. 1–28, 2015.

[45] B. H. Akman, T. Can, and A. Elif Erson-Bensan, "Estrogen-induced upregulation and 3'-UTR shortening of CDC6," *Nucleic Acids Res.*, vol. 40, no. 21, pp. 10679–10688, 2012.

[46] J. Y. Lee, I. Yeh, J. Y. Park, and B. Tian, "PolyA_DB 2: mRNA polyadenylation sites in vertebrate genes," *Nucleic Acids Res.*, vol. 35, no. SUPPL. 1, pp. 165–168, 2007.

[47] E. Astarci, A. E. Erson-Bensan, and S. Banerjee, "Matrix metalloprotease 16 expression is downregulated by microRNA-146a in spontaneously differentiating Caco-2 cells," *Dev. Growth Differ.*, vol. 54, no. 2, pp. 216–226, 2012.

[48] E. Astarci, A. Sade, I. ??imen, B. Sava??, and S. Banerjee, "The NF-??B target genes ICAM-1 and VCAM-1 are differentially regulated during spontaneous differentiation of Caco-2 cells," *FEBS J.*, vol. 279, no. 16, pp. 2966–2986, 2012.

[49] A. Subramanian *et al.*, "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 43, pp. 15545–50, 2005.

[50] V. Agarwal, G. W. Bell, J. W. Nam, and D. P. Bartel, “Predicting effective microRNA target sites in mammalian mRNAs,” *Elife*, vol. 4, no. AUGUST2015, pp. 1–38, 2015.

[51] J. Fantini, B. Verrier, J. Marvaldi, and J. Mauchamp, “In vitro differentiated HT 29-D4 clonal cell line generates leakproof and electrically active monolayers when cultured in porous-bottom culture dishes,” *Biol Cell*, vol. 65, no. 2, pp. 163–169, 1989.

[52] D. Zhou, B. Chen, J.-J. Ye, and S. Chen, “A novel crosstalk mechanism between nuclear receptor-mediated and growth factor/Ras-mediated pathways through PNRC-Grb2 interaction.,” *Oncogene*, vol. 23, no. 31, pp. 5394–404, 2004.

[53] Y. Xu, H. Hortsman, L. Seet, S. H. Wong, and W. Hong, “SNX3 regulates endosomal function through its PX-domain-mediated interaction with PtdIns(3)P.,” *Nat. Cell Biol.*, vol. 3, no. 7, pp. 658–666, 2001.

[54] M. Harterink *et al.*, “A SNX3-dependent retromer pathway mediates retrograde transport of the Wnt sorting receptor Wntless and is required for Wnt secretion,” *Nat. Cell Biol.*, vol. 13, no. 8, pp. 914–923, 2011.

[55] K. H. Chiow *et al.*, “SNX3-dependent regulation of epidermal growth factor receptor (EGFR) trafficking and degradation by aspirin in epidermoid carcinoma (A-431) cells,” *Cell. Mol. Life Sci.*, vol. 69, no. 9, pp. 1505–1521, 2012.

[56] A. Knight, E. Hughson, C. R. Hopkins, and D. F. Cutler, “Membrane protein trafficking through the common apical endosome compartment of polarized Caco-2 cells.,” *Mol. Biol. Cell*, vol. 6, no. 5, pp. 597–610, 1995.

[57] E. J. Hughson and C. R. Hopkins, “Endocytic pathways in polarized Caco-2 cells: identification of an endosomal compartment accessible from both apical and basolateral surfaces,” *J. Cell Biol.*, vol. 110, no. 2, pp. 337–348, 1990.

[58] I. I. Pacheco and R. J. Macleod, “CaSR stimulates secretion of Wnt5a from colonic myofibroblasts to stimulate CDX2 and sucrase-isomaltase using Ror2 on intestinal epithelia.,” *Am. J. Physiol. Gastrointest. Liver Physiol.*, vol. 295, no. 4, pp. G748–G759, 2008.

[59] I. Chantret *et al.*, “Differential expression of sucrase-isomaltase in clones isolated from early and late passages of the cell line Caco-2: evidence for glucose-dependent negative regulation.,” *J. Cell Sci.*, vol. 107 (Pt 1, pp. 213–225, 1994.

[60] L. M. Mehdawi, C. P. Prasad, R. Ehrnstrem, T. Andersson, and A. Slander, “Non-canonical WNT5A signaling up-regulates the expression of the tumor suppressor 15-PGDH and induces differentiation of colon cancer cells,” *Mol. Oncol.*, vol. 10, no. 9, pp. 1415–1429, 2016.

[61] M. Shah, S. a Rennoll, W. M. Raup-Konsavage, and G. S. Yochum, “A dynamic exchange of TCF3 and TCF4 transcription factors controls MYC expression in colorectal cancer cells,” *Cell Cycle*, vol. 14, no. 3, pp. 323–332, 2015.

[62] O. Voloshanenko *et al.*, “Wnt secretion is required to maintain high levels of Wnt activity in colon cancer cells.,” *Nat. Commun.*, vol. 4, no. May, p. 2610, 2013.

[63] R. Sandberg, J. R. Neilson, A. Sarma, P. a Sharp, and C. B. Burge, “Proliferating cells express mRNAs with shortened 3' UTRs and fewer microRNA target sites,” *Science (80-.).*, vol. 320, no. 5883, pp. 1643–1647, 2008.

APPENDIX A

DATASETS AND ANALYSIS OUTPUTS

Table A.1. Microarray experiment samples GSE7745

GSM Accesion Number	Status of Caco-2 Cells
GSM187694	Differentiated
GSM187695	Differentiated
GSM187696	Differentiated
GSM187459	Pre-Confluent
GSM187460	Pre-Confluent
GSM187461	Pre-Confluent

Table A.2. APADetect Result of APA Genes

Gene Symbol	Probeset ID	Poly(A) Site ID	Poly(A) Site Location	# of Valid Probes	# of Invalid Probes	SLR for Treated (Avg)	SLR for Control (Avg)	Treated /Control SLR Ratio (RT/RC)
PNRC1	209034_at	Hs.75969.1.7	89794154	4	7	21.11	1.98	10.64
SLC46A3	214719_at	Hs.117167.1.1	29274218	9	2	6.19	0.96	6.45
TMEM92	235245_at	Hs.224630.1.10	48357150	8	3	11.68	2.31	5.06
CUL4A	201424_s_at	Hs.339735.1.60	113919171	7	3	3.47	0.89	3.9
CORO2A	205538_at	Hs.113094.1.5	100886745	8	2	7.29	2.37	3.08
FGB	204988_at	Hs.300774.1.15	155491898	7	3	3.75	1.22	3.06
CYP3A7	211843_x_at	Hs.111944.1.4	99301705	9	2	5.43	2.21	2.46
CYP3A7	211843_x_at	Hs.111944.1.5	99302661	9	2	5.43	2.21	2.46
CCDC120	239403_at	Hs.522643.1.11	48927506	3	7	8.42	3.47	2.43
LGR4	218326_s_at	Hs.502176.1.4	27387818	8	3	3.29	1.36	2.42
PAM	214620_x_at	Hs.369430.1.59	102365209	7	4	38.37	17.85	2.15
PAM	214620_x_at	Hs.369430.1.60	102365295	7	4	38.37	17.85	2.15
PAM	214620_x_at	Hs.369430.1.61	102365416	7	3	44.97	20.96	2.15
MAF	1566324_a_at	Hs.134859.1.5	79629609	4	7	1.92	0.92	2.08
KLHL24	242088_at	Hs.407709.1.19	183398730	2	8	2.83	1.39	2.04

Table A.2. Cont'd

WDTC1	40829_at	Hs.469154.1.38	27634733	11	5	3.04	1.55	1.97
EPS15	217886_at	Hs.83722.1.3	51820379	3	8	1.69	0.87	1.93
METT5D1	238773_at	Hs.243326.1.20	28354863	6	5	3.39	1.81	1.87
C3orf23	241666_at	Hs.55131.1.18	44401293	8	3	2.24	1.25	1.79
WDTC1	215497_s_at	Hs.469154.1.38	27634733	3	8	2.68	1.53	1.75
ZNF764	222120_at	Hs.132227.1.3	30565086	6	5	4.93	2.84	1.74
SDF4	217855_x_at	Hs.42806.1.2	1152459	9	2	12.56	7.36	1.71
DCAF6	232349_x_at	Hs.435741.1.38	168015014	6	5	2.36	1.38	1.71
OSBP2	221237_s_at	Hs.517546.1.20	31290183	4	6	1.59	0.95	1.67
OSBP2	221237_s_at	Hs.517546.1.21	31290456	4	6	1.59	0.95	1.67
LTBP3	227308_x_at	Hs.289019.1.4	65306270	6	3	2.15	1.3	1.65
ANP32A	201038_s_at	Hs.458747.1.3	69071243	4	7	8.16	5	1.63
BIRC6	233093_s_at	Hs.150107.1.118	32843481	5	6	4.32	2.64	1.63
SOCS5	209648_x_at	Hs.468426.1.6	46989453	8	3	3.38	2.09	1.62
CALCOCO2	235076_at	Hs.514920.1.38	46942227	7	4	4.42	2.77	1.6
HIST1H3E	214616_at	Hs.443021.1.4	26225911	2	9	1.42	0.88	1.6
TREM1	219434_at	Hs.283022.1.8	41247277	6	4	2.51	1.57	1.6
COL4A1	211980_at	Hs.17441.1.3	110801755	2	9	4.68	2.95	1.58
INTS8	218905_at	Hs.521693.1.33	95885977	3	8	1.66	1.06	1.56
C14orf93	219009_at	Hs.255874.1.4	23456406	4	7	1.76	1.14	1.55
FDX1	203646_at	Hs.744.1.11	110333791	8	3	24	15.57	1.54
SLC25A17	211754_s_at	Hs.474938.1.9	41166421	6	5	1.47	0.95	1.54
TPT1	214327_x_at	Hs.374596.1.15	45912810	7	4	2.03	1.32	1.53
PRLR	231981_at	Hs.368587.1.6	35058689	3	8	1.42	0.94	1.52
ZNF226	219603_s_at	Hs.145956.1.12	44679403	7	4	1.76	1.16	1.51
DCTD	201571_s_at	Hs.183850.1.16	183815722	3	8	1.7	1.12	1.51
ANKRD12	216550_x_at	Hs.464585.1.21	9255713	7	4	3.47	2.31	1.5
UBXN6	220757_s_at	Hs.435255.1.5	4445459	8	3	1.45	0.97	1.5
LIPH	235871_at	Hs.68864.1.2	185226022	2	8	1.4	0.93	1.5
RGS10	204316_at	Hs.501200.1.2	121259337	4	7	1.93	2.94	0.66
APIP	218698_at	Hs.447794.1.3	34904044	8	2	3.58	5.4	0.66
PTPLAD1	217777_s_at	Hs.512973.1.19	65868838	3	7	1.13	1.71	0.66
PTRF	208790_s_at	Hs.437191.1.5	40556277	3	7	1.61	2.44	0.66
CFLAR	210564_x_at	Hs.390736.1.14	202001314	3	8	1.72	2.59	0.66
CYP20A1	219565_at	Hs.446065.1.24	204161650	8	3	2.65	4	0.66
CYP20A1	219565_at	Hs.446065.1.25	204161738	8	3	2.65	4	0.66
CSTF1	32723_at	Hs.172865.1.23	54979106	9	7	1.61	2.45	0.66
CCNB1	228729_at	Hs.23960.1.20	68473875	7	4	1.43	2.16	0.66
VTI1A	235034_at	Hs.194554.1.28	114577728	2	7	1.79	2.76	0.65

Table A.2. Cont'd

DENND5B	243613_at	Hs.118166.1.12	31604636	9	2	1.14	1.75	0.65
ARF6	203312_x_at	Hs.525330.1.12	50361483	9	2	1.88	2.89	0.65
METTL9	217868_s_at	Hs.279583.1.14	21667150	6	4	5.89	9.13	0.65
SIAH1	202981_x_at	Hs.295923.1.4	48394963	3	8	1.21	1.85	0.65
C17orf85	218896_s_at	Hs.120963.1.2	3715002	7	3	1.71	2.63	0.65
GTPBP8	223486_at	Hs.127496.1.12	112720058	8	2	1.89	2.92	0.65
PACRGL	235517_at	Hs.479298.1.22	20729973	5	6	1.04	1.59	0.65
GSTM4	210912_x_at	Hs.348387.1.24	110204323	2	8	1.97	3.07	0.64
CCAR1	224737_x_at	Hs.49853.1.56	70551237	5	6	3.71	5.82	0.64
UBASH3B	228353_x_at	Hs.444075.1.31	122682531	7	4	1.25	1.97	0.64
OSBP2	1569617_at	Hs.517546.1.20	31290183	4	5	1.26	1.97	0.64
PTPN12	202006_at	Hs.61812.1.40	77268924	7	4	0.9	1.41	0.64
VTI1A	235034_at	Hs.194554.1.29	114577752	4	7	2.37	3.76	0.63
TPI1	210050_at	Hs.524219.1.15	6980108	4	6	13.47	21.43	0.63
CSTF1	32723_at	Hs.172865.1.25	54979312	13	3	4.58	7.24	0.63
TUBB2B	214023_x_at	Hs.300701.1.19	3224731	6	4	1.17	1.85	0.63
ZNF100	238791_at	Hs.365142.1.4	21906844	5	6	2.04	3.29	0.62
CCDC93	219774_at	Hs.107845.1.3	118677572	6	4	1.22	1.96	0.62
CLDN22	222738_at	Hs.333179.1.39	184239795	8	3	1.3	2.12	0.62
WDHD1	204727_at	Hs.385998.1.1	55406680	2	9	1.03	1.68	0.61
METTL9	217868_s_at	Hs.279583.1.15	21667257	8	3	6.31	10.43	0.61
FOXA2	40284_at	Hs.155651.1.1	22561813	11	5	2.62	4.3	0.61
CLDN22	222738_at	Hs.333179.1.37	184239588	2	8	1.01	1.67	0.61
LARS	222428_s_at	Hs.432674.1.6	145493460	7	4	1.6	2.64	0.61
C6orf57	238504_at	Hs.418495.1.5	71298578	8	3	0.9	1.48	0.61
IL17RA	205707_at	Hs.129751.1.20	17590871	4	7	0.99	1.65	0.6
MED14	202610_s_at	Hs.407604.1.7	40510760	8	3	0.98	1.63	0.6
CCAR1	224737_x_at	Hs.49853.1.57	70551310	7	4	2.75	4.66	0.59
GAN	220124_at	Hs.112569.1.10	81413789	6	3	1.18	2.01	0.59
MGC45800	232446_at	Hs.175465.1.2	183063040	7	3	1.53	2.6	0.59
CLDN22	222738_at	Hs.333179.1.38	184239682	4	5	1.94	3.29	0.59
CCNB1	228729_at	Hs.23960.1.22	68473928	7	3	1.33	2.25	0.59
SNX3	213545_x_at	Hs.12102.1.3	108532714	4	6	4.67	7.98	0.59
MAGOHB	222776_at	Hs.104650.1.1	10756792	9	2	1.25	2.16	0.58
METTL1	204027_s_at	Hs.42957.1.3	58162395	7	3	1.14	1.99	0.58
LMAN2	200805_at	Hs.75864.1.4	176758951	6	5	2.58	4.44	0.58
TRUB1	235447_at	Hs.21187.1.14	116735333	8	2	1.56	2.72	0.57
PA2G4	214794_at	Hs.524498.1.17	56506882	3	8	2.39	4.16	0.57
USP6	206405_x_at	Hs.448851.1.9	5078190	7	4	0.87	1.53	0.57

Table A.2. Cont'd

MAP1D	1569029_at	Hs.298250.1.17	172945588	2	9	1.24	2.19	0.57
C6orf57	238504_at	Hs.418495.1.4	71298486	5	5	0.91	1.61	0.57
CFTR	215703_at	Hs.489786.1.10	117267962	8	3	1.35	2.38	0.57
DKC1	216212_s_at	Hs.4747.1.24	154004524	8	2	1.44	2.54	0.57
UNKNOWN	216304_x_at	Hs.499145.1.10	27400662	7	4	17.6	31.36	0.56
TSR1	239042_at	Hs.388170.1.1	2225689	2	9	1.58	2.82	0.56
C10orf119	217905_at	Hs.124246.1.4	121589303	3	7	5.46	10	0.55
FBXO17	220233_at	Hs.531770.1.10	39435465	3	8	0.89	1.63	0.55
CCDC112	235208_at	Hs.436121.1.2	114602890	6	4	1.3	2.38	0.55
CA5BP	238435_at	Hs.532326.1.10	15716202	3	7	1.26	2.32	0.55
PFDN4	205362_s_at	Hs.91161.1.9	52835951	2	8	2.72	5.02	0.54
DNAL1	223958_s_at	Hs.525445.1.6	74162727	3	8	1.78	3.38	0.53
DNAL1	223958_s_at	Hs.525445.1.7	74162779	3	8	1.78	3.38	0.53
CDC6	203967_at	Hs.405958.1.26	38458596	9	2	1.16	2.19	0.53
SNX3	200067_x_at	Hs.12102.1.3	108532714	7	4	3.44	6.49	0.53
CHST5	64900_at	Hs.156784.1.13	75572474	7	9	1.2	2.3	0.52
SMC2	213253_at	Hs.119023.1.48	106903093	3	8	1.19	2.31	0.52
RPS6KB1	204171_at	Hs.463642.1.34	58024689	9	2	2.33	4.58	0.51
BPTF	231953_at	Hs.444200.1.67	65956180	4	7	2.14	4.24	0.51
DNAJC21	235032_at	Hs.131887.1.30	34955306	6	5	2.84	5.58	0.51
GINS3	45633_at	Hs.47125.1.11	58439943	13	3	1.44	2.98	0.48
METTL1	204027_s_at	Hs.42957.1.2	58162365	8	3	1.16	2.49	0.47
TNFSF10	214329_x_at	Hs.478275.1.2	172223465	2	7	9.54	20.8	0.46
TCF3	213730_x_at	Hs.371282.1.8	1610954	7	4	6.15	13.76	0.45
GLRX3	214205_x_at	Hs.42644.1.14	131977680	3	7	2.86	6.62	0.43
GLRX3	214205_x_at	Hs.42644.1.16	131977870	3	7	2.86	6.62	0.43
GLRX3	214205_x_at	Hs.42644.1.17	131977933	3	7	2.86	6.62	0.43
WARS	200628_s_at	Hs.497599.1.4	100801050	8	2	1.26	2.95	0.43
CLU	222043_at	Hs.436657.1.7	27455510	2	7	5.47	12.78	0.43
HNRNPA1	214280_x_at	Hs.546261.1.20	54678663	5	6	1.23	2.93	0.42
HNRNPA1	214280_x_at	Hs.546261.1.23	54678975	5	6	1.23	2.93	0.42
HNRNPA1	214280_x_at	Hs.546261.1.24	54679502	5	6	1.23	2.93	0.42
INCENP	219769_at	Hs.142179.1.21	61919799	2	8	0.91	2.24	0.41
SNHG10	238691_at	Hs.448753.1.2	95999250	3	8	1.81	4.62	0.39
CEP76	52285_f_at	Hs.236940.1.5	12673095	11	3	2.74	7.06	0.39
TCF3	210776_x_at	Hs.371282.1.10	1611038	6	5	2.28	5.98	0.38
DAZAP2	214334_x_at	Hs.369761.1.9	51637544	5	6	10.93	29.38	0.37
MCM10	222962_s_at	Hs.198363.1.34	13251901	4	6	0.86	2.55	0.34
TCF3	213730_x_at	Hs.371282.1.10	1611038	5	6	4.69	13.91	0.34

YARS2	218470_at	Hs.505231.1.6	32900087	2	7	6.67	21.08	0.32
YARS2	218470_at	Hs.505231.1.4	32899958	6	5	6.12	21.51	0.28
NCAPH2	40640_at	Hs.180903.1.42	50961900	2	14	2.58	10.63	0.24

Table A.3. logSLR values of chosen genes

Gene	Probeset ID	PolyA Site ID	GSM187459	GSM187460	GSM187461	GSM187694	GSM187695	GSM187696
SNX3	200067_x_at	Hs.12102.1.3	1.88	1.83	2	1.13	1.22	1.33
TCF3	213730_x_at	Hs.371282.1.10	2.66	2.62	2.6	1.69	1.16	1.69
PNRC1	209034_at	Hs.75969.1.7	0.68	0.66	0.7	3.1	2.98	3.05
CDC6	203968_s_at	Hs.405958.1.22	0.15	0.05	0.04	0.08	0.15	0.31

Table A.4. Gene groups by their miRNA binding sites

ANP32A	a
BIRC6	a
CCDC120	a
COL4A1	a
CORO2A	a
CUL4A	a
EPS15	a
PNRC1	a
LGR4	a
SLC25A17	a
LIPH	a
SOCS5	a
PAM	a
TMEM92	a
FDX1	b1
FGB	b1
KLHL24	b1
MAF	b1
METT5D1	b1
PRLR	b1

Table A.4. Cont'd

SLC46A3	b1
WDTC1	b1
ANKRD12	b2
C14orf93	b2
C3orf23	b2
CALCOCO2	b2
CYP3A7	b2
DCAF6	b2
DCTD	b2
HIST1H3E	b2
INTS8	b2
LTBP3	b2
OSBP2	b2
SDF4	b2
TPT1	b2
TREM1	b2
ZNF226	b2
ZNF764	b2
ARF6	c
C10orf119	c
CCAR1	c
CCDC93	c
CCNB1	c
CDC6	c
CEP76	c
CLDN22	c
CLU	c
DAZAP2	c
DNAJC21	c
FOXA2	c
GAN	c
LMAN2	c
HNRNPA1	c
MED14	c
METTL9	c
PACRGL	c
PFDN4	c
PTPLAD1	c

Table A.4. Cont'd

PTPN12	c
RGS10	c
RPS6KB1	c
SIAH1	c
SNX3	c
TCF3	c
TRUB1	c
TUBB2B	c
USP6	c
VTI1A	c
C17orf85	d1
CSTF1	d1
CYP20A1	d1
DNAL1	d1
GLRX3	d1
MAGOHB	d1
PTRF	d1
UBASH3B	d1
WARS	d1
WDHD1	d1
APIP	d2
BPTF	d2
C6orf57	d2
CA5BP	d2
CCDC112	d2
CFLAR	d2
CFTR	d2
CHST5	d2
DENND5B	d2
DKC1	d2
FBXO17	d2
GINS3	d2
GSTM4	d2
GTPBP8	d2
IL17RA	d2
LARS	d2
LMAN2	d2
MAP1D	d2

Table A.4. Cont'd

MCM10	d2
METTL1	d2
MGC45800	d2
NCAPH2	d2
OSBP2	d2
PA2G4	d2
POLR2K	d2
SMC2	d2
SNHG10	d2
TNFSF10	d2
TPI1	d2
TSR1	d2
ZNF100	d2

APPENDIX B

LACK OF DNA CONTAMINATION AND CDNA SYNTHESIS CONFIRMATION

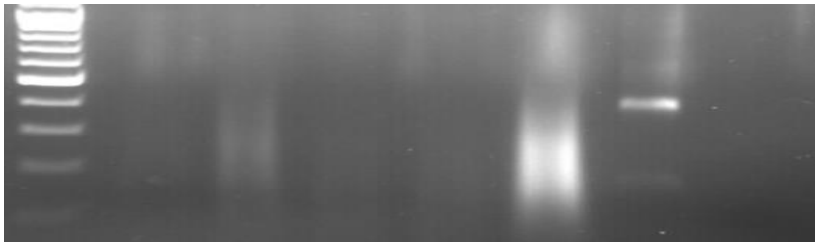


Figure B.1. Confirmation of lack of DNA contamination in RNA samples. PCR was performed using GAPDH specific primers. GAPDH_F: 5'-GGGAGCCAAAAGGGTCATCA3' and GAPDH_R: 5'-TTTCTAGACGGCAGGTCA GGT-3' (product size: 409 bp). Following conditions were used for the PCR reactions: incubation at 94°C for 10 minutes, 40 cycles of 94°C for 30 seconds, 56°C for 30 seconds, and 72°C for 30 seconds, and final extension at 72°C for 5 minutes. Genomic DNA was used as a positive control.

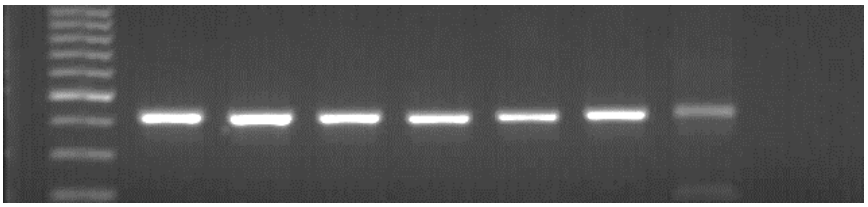


Figure B.2. Confirmation of cDNA synthesis. PCR was performed using GAPDH specific primers. GAPDH_F: 5'-GGGAGCCAAAAGGGTCATCA3' and GAPDH_R: 5'-TTTCTAGACGGCAGGTCA GGT-3' (product size: 409 bp). Following conditions were used for the PCR reactions: incubation at 94°C for 10 minutes, 40 cycles of 94°C for 30 seconds, 56°C for 30 seconds, and 72°C for 30 seconds, and final extension at 72°C for 5 minutes. Genomic DNA was used as a positive control.

APPENDIX C

QUANTIATIVE REAL-TIME PCR REPORT OF GAPDH

Run Setup

Protocol

1: 95,0°C for 10:00
2: 94,0°C for 0:30
3: 59,0°C for 0:30
4: 72,0°C for 0:30
Plate Read
5: GOTO 2, 35 more times
6: 95,0°C for 0:10
7: Melt Curve 50,0°C to 99,0°C: Increment 1,0°C 0:05
Plate Read

Quantification

Step #: 4
Analysis Mode: Fluorophore
Cq Determination: Single Threshold
Baseline Method:
SYBR: Auto Calculated
Threshold Setting:
SYBR: 14,91, Auto Calculated

Figure C.1. Run setup for the RT-qPCR of GAPDH

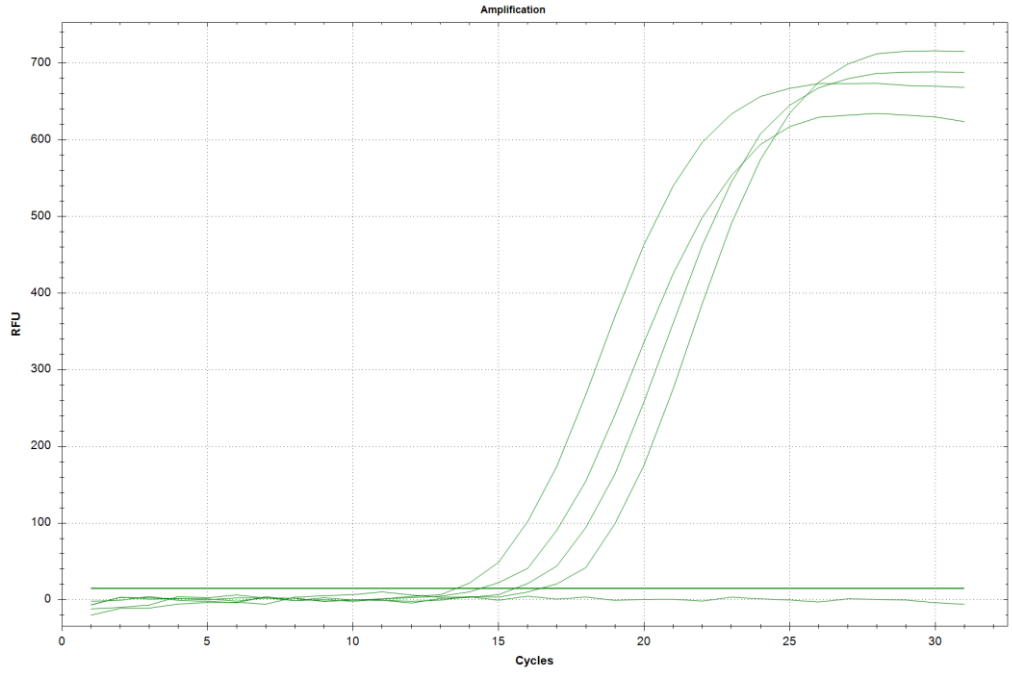


Figure C.2. Quantification graph for Cycling Green

Quantification Data

Well	Fluor	Target	Content	Sample	Cq	Cq Mean	Cq Std. Dev	Starting Quantity (SQ)	Log Starting Quantity	SQ Mean	SQ Std. Dev
B01	SYBR		Std	Std1	13,55	13,55	0,000	1,000E+06	6,000	1,00E+06	0,00E+00
D01	SYBR		Std	Std2	14,39	14,39	0,000	5,000E+05	5,699	5,00E+05	0,00E+00
F01	SYBR		Std	Std3	15,57	15,57	0,000	2,500E+05	5,398	2,50E+05	0,00E+00
H01	SYBR		Std	Std4	16,45	16,45	0,000	1,250E+05	5,097	1,25E+05	0,00E+00
B03	SYBR		NTC		N/A	0,00	0,000	N/A	N/A	0,00E+00	0,00E+00

Figure C.3. Quantification data

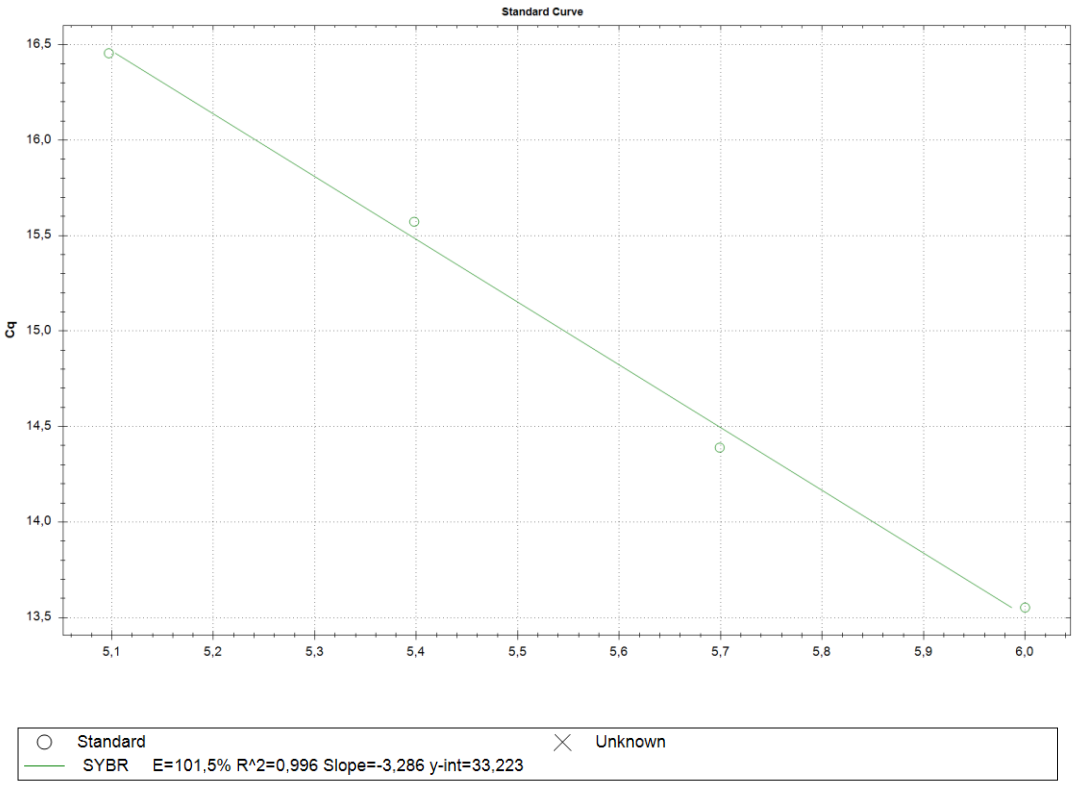
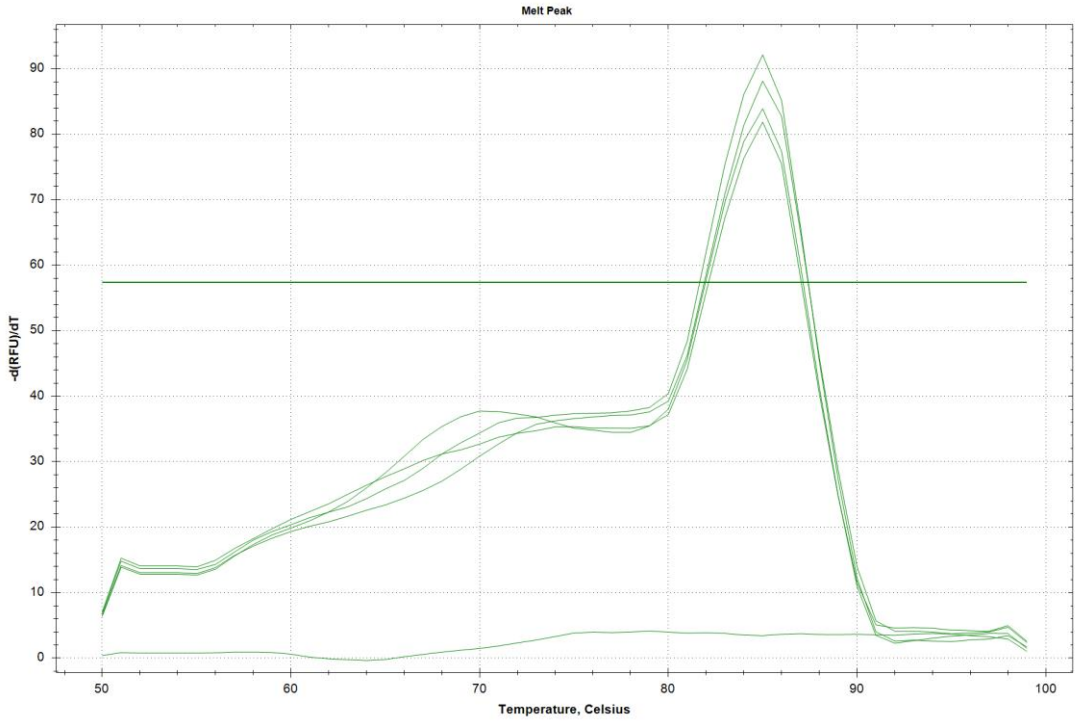


Figure C.4. Standart curve and equation



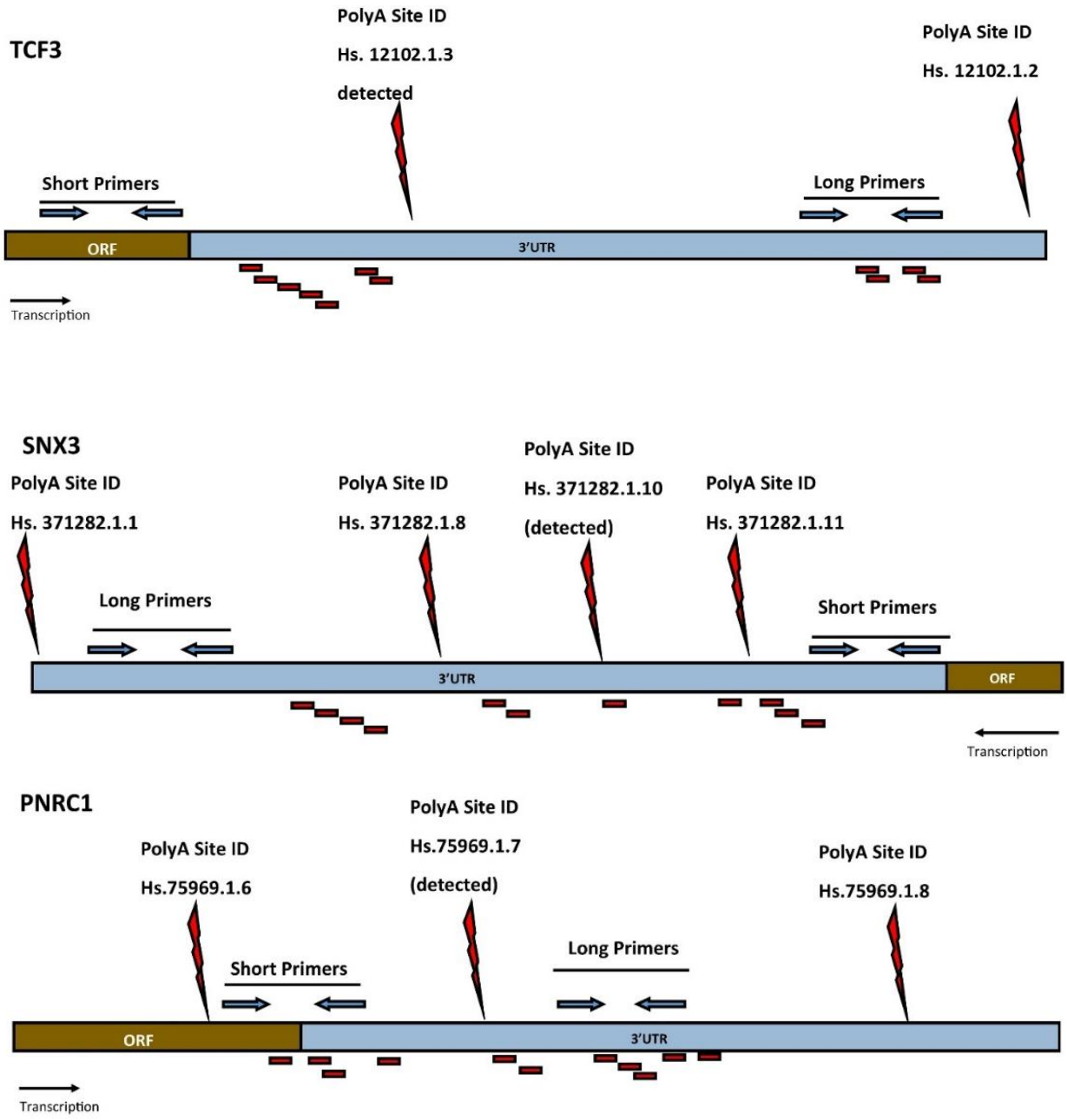
Melt Curve Data

Well	Fluor	Target	Content	Sample	Melt Temp
B01	SYBR		Std	Std1	85,00
B03	SYBR		NTC		None
D01	SYBR		Std	Std2	85,00
F01	SYBR		Std	Std3	85,00
H01	SYBR		Std	Std4	85,00

Figure C.4. Melt curve data

APPENDIX D

GENE DIAGRAM FOR PROBE BINDING POSITIONS



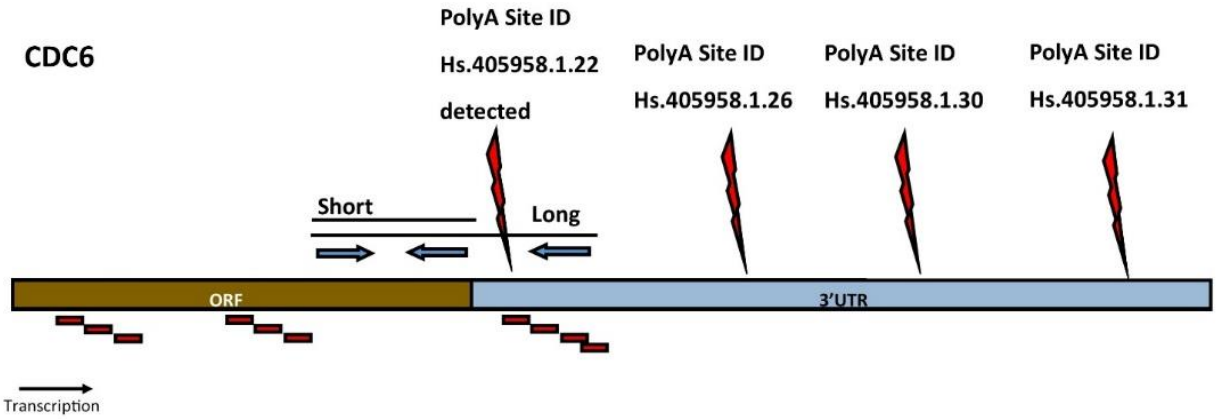


Figure D.1. Probe distribution and primer locations around the poly(A) sites locations gathered from poly(A)_dB. Probes are shown as red squares below the open reading frame (ORF; brown) and 3'UTR (Untranslated region; blue) bars. Each trancript has unique transcription direction. Short and long primers are indicated as blue arrows.

APPENDIX E

BUFFERS FOR EXPERIMENTS

12% Seperating Gel Mix

3.8 ml 10%SDS+1.5M TrisHCl pH 8.8

6.7ml of Acrylamide (30%)

150 µl APS

20 µl TEMED

4.3 ml dH2O

4% Stacking Gel Mix

2 ml 10%SDS+1.5M TrisHCl pH 6.8

1.2 ml of Acrylamide (30%T)

50 µl APS

10 µl TEMED

4.7 ml dH2O

TBS-T:

20 mM Tris

137 mM NaCl

0.1% Tween 20

pH: 7.6

PBS-T:

137mM NaCl

2,7mM KCl

10mM Na2HPO4.2H2O

2mM KH2PO4

0.1% Tween 20

pH: 7.4

6X Laemmli Buffer:

12% SDS

30% 2-mercaptoethanol

60% Glycerol

0.012% bromophenol blue

0.375 M Tris

Mild Stripping Buffer

15g glycine

1 g SDS

10 ml Tween 20

Adjust the pH to 2,2

Complete to 1L with distilled water

Running Buffer

25 mM Tris base

190 mM glycine

0.1% SDS

Transfer Buffer

200 ml Methanol

100 ml 10X Blotting Buffer

700 ml dH2O

10 X Blotting Buffer

30.3 g Trizma Base (0.25M)

144 g Glycine (1.92 M)

pH ˜ 8.3

NBT/BCIP Mix for Alkaline Phosphatase Staining

122.86 mg $MgSO_4.7H_2O$

121 mg Tris

4 mg NBT

1.9 mg BCIP

Dissolve in 5 mL ddH_2O in 50 mL falcon

Adjuct pH to 9.5

Comlete final volume to 10 mL with ddH_2O

APPENDIX F

MARKERS

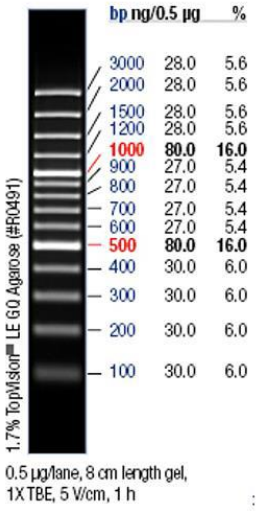


Figure F.1.: GeneRuler 100 bp DNA Ladder Plus

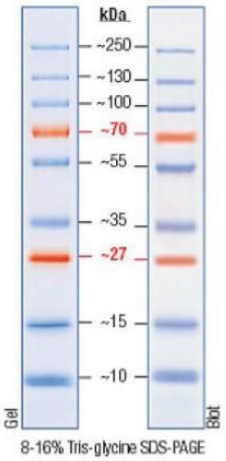


Figure F.2.: PageRuler Pre-stained Protein Ladder. Fermentas SM1811 Protein ladder was used during Western Blot experiments.

APPENDIX G

MAMMALIAN CELL LINE PROPERTIES

Table G.1. The properties of the cell-lines used in this study

Cell Line	Age	Sex	Tissue	Cancer type	Tumor type
Caco-2	77-year old	male	Colon	Colorectal adenocarcinoma	
HCT-116	48-year old	male	Colon	Colorectal carcinoma	
HT-29	44-year-old	female	Colon	Colorectal adenocarcinoma	Dukes' C Primary
LoVo	56-year-old	male	Colon	Colorectal adenocarcinoma	Dukes' C Metastatic
RKO			Colon	Colonic Carcinoma	Primary
SW480	50-year-old	male	Colon	Colorectal adenocarcinoma	Dukes' B Primary
SW620	51-year-old	male	Colon	Colorectal adenocarcinoma	Dukes' C Metastatic
T-84	72-year-old		Colon	Colorectal carcinoma	Metastatic

Figure G.1. Colon cancer cell lines classified by the molecular pathways CIN, MSI and CIMP, and mutation status of cancer critical genes (Figure taken from Ahmed et al., 2013) (Ahmed et al., 2013).

Cell line	*MSI status*	*CIMP panel 1*	*CIMP panel 2*	*CIN*	KRAS	BRAF	PIK3CA	PTEN	TP53
SW480	MSS	–	–	+	G12V	wt	wt	wt	R273H;P309S
SW620	MSS	+	–	+[46]	G12V	wt	wt	wt	R273H;P309S
HT-29	MSS	+	+	+	wt	V600E	P449T[b]	wt	R273H
Caco-2	MSS	+	–	+[48]	wt	wt	wt	wt	E204X
RKO	MSI	+	+	–	wt	V600E	H1047R	wt	wt
LoVo	MSI	–	–	–	G13D;A14V	wt	wt	wt	wt
HCT-116	MSI	+	+	–	G13D	wt	H1047R	wt	wt

Abbreviations: CIN, chromosomal instability pathway; MSI, microsatellite instability; MSS, microsatellite stable; CIMP, CpG island methylator phenotype; X, stop codon; fs, frame shift; wt, wild type. Mutations are annotated at the protein level as described by den Dunnen *et al.*[54] (standard one-letter amino acid abbreviations, X and fs). For further details, see Supplementary Table 1. [a]No publication on WiDr karyotype was found; however, WiDr and HT-29 are identical cell lines.[19] [b]Previously reported mutations not covered by our assays.