Functional Characterization of the Candidate Colorectal Cancer Gene *CNOT1*

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE UNIVERSITY OF MINNESOTA

By

Lei Zhao

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Thesis Advisors: Robert Cormier, Ph.D. and Patricia Scott, Ph.D.

December 2013
Copyright © 2013

Lei Zhao
Acknowledgements

First and foremost, I would like to acknowledge my thesis advisors, Drs. Robert Cormier and Patricia Scott, for their continuous support of my PhD study and mentorship on this thesis research project. I could not have imagined having better advisors and mentors for my PhD studies than them.

I gratefully acknowledge my committee chair Dr. Lester Drewes for the instructions on my research and the organization of my thesis defense. I also would like to thank other members of my committee, Dr. David Largaespada, Dr. Matthew Andrews, Dr. Kendall Wallace, and Dr. Lynne Bemis for their helpful guidance and support throughout the course of this research.

I also would like to thank my lab members: Annette Rod, Elaine Than, and Mike Madden for their friendship and help during the past four years.

My special thanks go to my wife Yanjun Wu for her dedication, love and persistent confidence in me. I could not have done this without her.

Finally, I want to thank the NIH and Whiteside Institute for Clinical Research for providing the financial support for this study.
Dedication

This thesis is dedicated to my family who has always been supporting me. This dissertation is simply impossible without the support from my wife Yanjun Wu and my parents. To them I dedicate this dissertation.
Abstract

Cnot1 was identified as a candidate colorectal cancer (CRC) gene in multiple Sleeping Beauty-based screens. Multiple somatic mutations in CNOT1 in CRC have been documented. Depletion of CNOT1 in the SW480 CRC cell line resulted in a dramatic decrease in cell viability, suggesting that CNOT1 may act as an oncogene. CNOT1 is a scaffold of the CCR4-NOT complex, which regulates mRNA expression at several levels. Examination of global gene expression in CNOT1-depleted SW480 cells revealed dramatically increased expression of TXNIP, which can initiate apoptotic pathways in response to extreme endoplasmic reticulum (ER) stress. We hypothesized that CNOT1 might protect cell viability by limiting TXNIP expression. In support of this idea that upregulation of TXNIP was responsible for decreased cell viability seen in CNOT1-depleted cells, we found that enforced expression of TXNIP in HEK293 cells resulted in decreased cell viability. Also, analysis of TXNIP expression in parallel with cell apoptosis in CNOT1 depleted cells revealed that upregulation of TXNIP occurs well before a measurable increase in cell apoptosis, suggesting that elevated TXNIP may contribute to apoptosis seen in CNOT1-depleted cells. To substantiate this connection we investigated the mechanism by which CNOT1 regulates TXNIP levels. We found that depletion of CNOT8, a deadenylase subunit of the CCR4-NOT complex, also results in decreased cell viability and increased TXNIP expression. Further, CNOT1 depletion results in increased TXNIP mRNA stability. These results argue that CNOT1 limits TXNIP expression via regulation of mRNA deadenylation and stability. TXNIP is activated to promote apoptosis by extreme ER stress in part by means of relief of miR-17-mediated degradation of TXNIP mRNA. In support of this, we found that miR-17 overexpression downregulates TXNIP mRNA in SW480 cells. We also found that limiting TXNIP expression increases cell viability under condition of ER stress. CNOT1 acts as an adapter between the miRNA RISC and deadenylase subunits of the CNOT complex. Thus, we hypothesized that increased activity of CNOT1 may facilitate the degradation of TXNIP by miR-17. Further analyses suggested that CNOT1 might limit ER stress and activation of the unfolded protein response (UPR) pathways. Based on the results above, we proposed the following model. As cancer cells are often subject to extreme stresses, increased activity of CNOT1 may protect against the effects of severe ER stress through directly limiting TXNIP transcript levels and preventing ER stress leading to apoptosis, thus conferring a growth advantage to cancer cells.
# Table of Contents

ACKNOWLEDGEMENTS.................................................................i  
DEDICATION..................................................................................ii  
ABSTRACT....................................................................................iii  
TABLE OF CONTENTS..................................................................iv  
LIST OF TABLES..........................................................................vi  
LIST OF FIGURES.......................................................................viii  
ABBREVIATIONS.........................................................................xii  

## CHAPTER

1. Introduction  
Epidemiology and Challenges in Colorectal Cancer Treatment..................2  
The Biology of the Human Intestinal Tract..............................................4  
Important Signaling Pathways in Intestinal Crypt Renewal.......................8  
  WNT Signaling Pathway.....................................................................8  
  NOTCH Signaling Pathway...............................................................9  
  BMP Signaling Pathway...................................................................10  
Hereditary and Sporadic CRC..............................................................11  
  Hereditary CRC.............................................................................11  
  Genetic Alterations in Sporadic CRC.................................................12  
Sleeping Beauty Transposon-Based Forward Genetic Screens....................18  
The CCR4-NOT (CNOT) Complex and CNOT1.......................................22  
Mechanisms of mRNA Decay................................................................27  
ER Stress and the Unfolded Protein Response.........................................33  
Thioredoxin-interacting Protein.............................................................37  
Summary..........................................................................................39

2. Results.........................................................................................41

3. Discussion....................................................................................86
4. Material and methods........................................................................................................92

BIBLIOGRAPHY.................................................................................................................102

APPENDICES

1. A *Sleeping Beauty* transposon-mediated screen identifies murine susceptibility genes for adenomatous polyposis coli 
   
   *(Apc)-dependent intestinal tumorigenesis........................................................................125

2. The role of *KCNQ1* in mouse and human GI cancer....................................................157

3. The roles of *CNOT1* and *CNOT2* in regulating mRNA levels of *p27kip* and *CDX2*..................................................................................................................188

4. Construction of *CNOT1* mutant plasmids.................................................................200

5. Identification of fusion mRNA transcripts in human colorectal cancer and normal colon epithelial tissues using RNA seq.........................................................209
List of Tables

Chapter 1

Table 1. Cnot1 was identified as CIS candidate cancer gene by multiple SB screens.....24
Table 2. Splicing variants and protein isoforms of CNOT1........................................25

Chapter 2

Table 3. Time course study of a correlation between CNOT1 depletion, upregulated
TXNIP and cell apoptosis.................................................................53
Table 4. List of 266 genes showing significant upregulated (A) and downregulated (B)
expression levels in SW480 cells depleted for CNOT1. ........................67
Table 5. Core genes enriched in CNOT1 HI/CNOT1 LO......................................74
Table 6. Core RIDD genes enriched in CNOT1-depleted SW480 cells..................81
Table 7. Plasma membrane localized components (PLASMA_MEMBRANE
_PART) enriched in CNOT1-depleted SW480 cells..............................82

Chapter 4

Table 8. Primers used in qRT-PCR..........................................................95

Appendices

Table 9. Polyp number and age of death for transgenic mice............................129
Table 10. List of 33 CIS........................................................................133
Table 11. LOH and MOH in ApcMin tumors based on the ratio of T:A trace peaks......154
Table 12. Sequence read overlap between duplicate regions of a single GS FLX
sequencing run...............................................................................154
Table 13. Human orthologous regions to the mouse CIS with recurrent chromosomal
copy number changes based on published data.................................155
Table 14. Knockdown of ApcMin CIS candidate genes affects viability of human colon
cancer cells...............................................................................156
Table 1. Mapped transposon insertions in 96 tumors..........................................................156
Table 16. Loss of Kcnq1 enhances tumor multiplicity in ApcMin mice..........................161
Table 17. ApcMin Kcnq1−/− tumor phenotype is strongest in the proximal quarter of the
  small intestine...........................................................................................................161
Table 18. List of top known genes 1.5 fold (A) up-regulated and (B) down-regulated in
  Kcnq1−/− mouse colons..........................................................................................186
Table 19. List of top known genes 1.5 fold (A) up-regulated and (B) down-regulated in
  Kcnq1−/− mouse small intestines.........................................................................186
Table 20. IPA Analysis of Colon Microarray..................................................................186
Table 21. IPA Analysis of Proximal Small Intestine Microarray..............................186
Table 22. Core genes enriched in Kcnq1 KO..............................................................186
Table 23. Expression of Areg ......................................................................................186
Table 24. The 9 selected mutations in CNOT1............................................................200
Table 25. The sequence of selection and mutagenic primers......................................202
Table 26. The sequencing primers used for validation of mutations............................206
Table 27. Summary of fusion transcripts discovered in 20 CRC samples...............220
Table 28. Summary of fusion transcripts discovered in 5 normal colon samples.....220
Table 29. Lists of the fusions that have been validated by PCR and
  Sanger Sequencing..................................................................................................221
Table 30. Summary of clinico-pathological profile of the 20 CRC patients..............224
Table 31. Summary of clinico-pathological profile of the 5 normal samples...........224
Table 32. RIN of the total 25 RNA samples...............................................................225
Table 33. List of primers used for validation of candidate fusion transcripts............226
Table 34. List of sequencing primers for validation of candidate fusion transcripts...227
Table 35. Summary of fusion transcripts discovered in each CRC sample.............227
Table 36. Summary of fusion transcripts discovered in the 20 CRC samples..........227
Table 37. Summary of fusion transcripts discovered in each normal sample...........234
Table 38. Summary of fusion transcripts discovered in 5 normal samples............234
Table 39. The overlapping fusion transcripts in CRC and normal samples...........235
List of Figures

Chapter 1

Figure 1. The anatomy of the gastrointestinal tract.............................................................6
Figure 2. The multi-layered organization of the mature GI tract...........................................6
Figure 3. The organization of the small intestinal crypt-villus and the colon crypt.............7
Figure 4. Organization of the intestinal epithelium and the crypts of Lieberkühn and cell lineage determination in the intestinal epithelium.................................7
Figure 5. The fraction of colon cancer cases that arise in various family risk settings ..........................................................11
Figure 6. Schematic representation of the Vogelstein model of colonic carcinogenesis..........................................................13
Figure 7. Human colorectal cancer genome landscape..........................................................17
Figure 8. Mechanism of SB-mediated transposition..........................................................20
Figure 9. Using SB system for cancer gene screens in mice..................................................21
Figure 10. A proposed model for the human CCR4-NOT complex.......................................24
Figure 11. Map of SB insertions in Cnot1 in ApcMin screen..................................................24
Figure 12. Mutations in CNOT1 identified in CRC samples by TCGA and Seshagiri et al..........................................................25
Figure 13. Diagrams of CNOT1 with N-terminal, middle and C-terminal regions.............26
Figure 14. Mechanisms of normal mRNA degradation.......................................................29
Figure 15. The canonical pathway of microRNA processing..................................................32
Figure 16. The unfolded protein response in mammalian cells.............................................36

Chapter 2

Figure 17. CNOT1 depletion leads to decreased cell viability in transformed cell lines SW480 and HEK293..........................................................42
Figure 18. CNOT1 depletion leads to increased apoptosis in SW480 cells.........................43
Figure 19. Depletion of CNOT1 in SW480 cells causes an
altered gene expression profile...............................................................45

Figure 20. IPA analysis of gene expression in CNOT1-depleted SW480 cells........46
Figure 21. Geneset Enrichment Analysis of human metastasis CRC samples.........50
Figure 22. CNOT1 limits TXNIP mRNA levels in SW480 and HEK293 cell lines......51
Figure 23. Overexpression of TXNIP reduces SW480 cell viability..........................53
Figure 24. Identification of CNOT1 isoforms and CCR4-NOT subunits involved in TXNIP regulation..............................................................55
Figure 25. CNOT1 limits TXNIP mRNA levels through decreasing mRNA stability.....57
Figure 26. 3’UTR binding proteins ZFP36, PUM1, PUM2 and CELF1 are not involved in CNOT1-mediated regulation of TXNIP..........................58
Figure 27. miR-17 downregulates TXNIP mRNA levels....................................59
Figure 28. Tunicamycin induces decreased cell viability and activation of UPR signaling in SW480 cells...............................................................61
Figure 29. TXNIP plays a role in the UPR in SW480 cells.....................................62
Figure 30. Depletion of CNOT1 results in ER stress and activation of UPR pathways..........................................................................................63
Figure 31. CNOT1 is involved in ER stress and activation of UPR pathways..........64
Figure 32. A model for roles of CNOT1 in regulating cell viability..........................65
Figure 33. Molecules in top three gene networks in IPA shown in Figure 20..........74

Appendices

Figure 34. A pedunculated adenoma stained with H&E (A) or immunostained for β-catenin (B and C)............................................130
Figure 35. Three alleles used to target SB mutagenesis to the intestinal tract..........153
Figure 36. PCR technique for detecting Apc LOH...............................................154
Figure 37. Loss of Kcnq1 promotes tumor progression......................................162
Figure 38. qRT-PCR gene expression analysis of mouse colon............................164
Figure 39. qRT-PCR gene expression analysis of mouse proximal small intestine....165
Figure 40. Results of GSEA showing enrichment scores (ES) of Cfr and Muc2 gene sets with respect to the ranked Kcnq1 expression data set........168
Figure 41. Colon organoids........................................................................................................170
Figure 42. Representative human samples showing the expression
pattern of KCNQ1 in the epithelium of CRCLM..........................................................171
Figure 43. Kaplan-Meier graph depicting OS in months, stratified by
intensity of KCNQ1 expression in CRCLMs.................................................................171
Figure 44. Rectal Adenomatous Hyperplasia.......................................................................179
Figure 45. Stomach of Apc^{+/+} Kcnq1^{-/-} mice.................................................................180
Figure 46. Pancreas of Apc^{Min} Kcnq1^{-/-} and Kcnq1^{wt}..................................................181
Figure 47. Immunohistochemistry for cellular markers in mouse colon...............................182
Figure 48. Colon Microarray-Top Functional Networks Identified by IPA............................183
Figure 49. Proximal Small Intestine Microarray-Top Gene Networks
Identified by IPA..................................................................................................................184
Figure 50. Concordance between expression of KCNQ1 and CFTR in
colorectal cancer liver metastases......................................................................................184
Figure 51. p27kip and CDX2 mRNA expression levels in SW480 cells
depleted for CNOT subunits..............................................................................................191
Figure 52. Cell viability and p27kip and CDX2 mRNA expression levels
in SW480 cells depleted for RBPs.......................................................................................192
Figure 53. Depletion of CDX2 did not result in changes in cell viability
in SW480 cells.....................................................................................................................194
Figure 54. CNOT1 does not regulate p27kip and CDX2 mRNA levels
through 3'UTR-mediated deadenylation..............................................................................195
Figure 55. The distribution of selected mutations in the CNOT1 protein...............................201
Figure 56. The general procedure of making site-directed mutations using a
Transforme Site-Directed Mutagenesis Kit.........................................................................202
Figure 57. The map of pBluescriptR-CNOT1 plasmid and the location of
mutagenic and selection primers on the plasmid.............................................................203
Figure 58. Validation of introduced mutations by Sanger sequencing.....................................206
Figure 59. RNA splicing mechanisms in eukaryotes.............................................................211
Figure 60. The procedure of RNA-seq..................................................................................214
Figure 61. The map of fusion transcripts in 20 CRC liver metastasis.................................219
Figure 62. The map of fusion transcripts in 5 normal colon samples.................................220
Figure 63. The overlapping fusion transcripts discovered in
 both CRC and normal samples.........................................................................................221
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3’UTR</td>
<td>3’ untranslated region</td>
</tr>
<tr>
<td>APC</td>
<td>Adenomatous polyposis coli</td>
</tr>
<tr>
<td>ARE</td>
<td>AU-rich elements</td>
</tr>
<tr>
<td>ASK-1</td>
<td>apoptosis signaling kinase-1</td>
</tr>
<tr>
<td>ATF6</td>
<td>activating transcription factor-6</td>
</tr>
<tr>
<td>BMP</td>
<td>Bone morphogenetic proteins</td>
</tr>
<tr>
<td>CBC</td>
<td>crypt-base-columnar</td>
</tr>
<tr>
<td>CHOP</td>
<td>C/EBP homologous protein</td>
</tr>
<tr>
<td>CIS</td>
<td>Common Insertion Site</td>
</tr>
<tr>
<td>CNOT1</td>
<td>CCR4-NOT transcription complex, subunit 1</td>
</tr>
<tr>
<td>CNOT8</td>
<td>CCR4-NOT transcription complex, subunit 8</td>
</tr>
<tr>
<td>CRC</td>
<td>Colorectal cancer</td>
</tr>
<tr>
<td>CRCLM</td>
<td>CRC liver metastasis</td>
</tr>
<tr>
<td>DDT</td>
<td>Dichloro-diphenyl-trichloroethane</td>
</tr>
<tr>
<td>EMT</td>
<td>Epithelial-mesenchymal transition</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic reticulum</td>
</tr>
<tr>
<td>Estrogen receptor</td>
<td></td>
</tr>
<tr>
<td>FAP</td>
<td>Familial adenomatous polyposis</td>
</tr>
<tr>
<td>FFPE</td>
<td>Formalin-fixed and paraffin-embedded</td>
</tr>
<tr>
<td>FKMF</td>
<td>Fragments per kilobase of exon per Million Fragments mapped</td>
</tr>
<tr>
<td>GEO</td>
<td>Gene expression omnibus</td>
</tr>
<tr>
<td>GI</td>
<td>Gastrointestinal</td>
</tr>
<tr>
<td>GSEA</td>
<td>Gene set enrichment analysis</td>
</tr>
<tr>
<td>HNPCC</td>
<td>Hereditary non-polyposis colorectal cancer</td>
</tr>
<tr>
<td>IL-1β</td>
<td>Interleukin 1β</td>
</tr>
<tr>
<td>IPA</td>
<td>Ingenuity pathway analysis</td>
</tr>
<tr>
<td>IR</td>
<td>Inverted repeats</td>
</tr>
<tr>
<td>IRE1</td>
<td>Inositol requiring enzyme 1</td>
</tr>
<tr>
<td>ISC</td>
<td>Intestinal stem cell</td>
</tr>
<tr>
<td>K-RAS</td>
<td>Kirsten rat sarcoma viral oncogene homolog</td>
</tr>
<tr>
<td>KCNQ1</td>
<td>Potassium voltage-gated channel, KQT-like subfamily, member1</td>
</tr>
<tr>
<td>KO</td>
<td>Knockout</td>
</tr>
<tr>
<td>LM-PCR</td>
<td>Ligation-mediated pcr</td>
</tr>
<tr>
<td>LOH</td>
<td>Loss of heterozygosity</td>
</tr>
<tr>
<td>LQTS</td>
<td>Long QT Syndrome</td>
</tr>
<tr>
<td>LRP5</td>
<td>Low-density lipoprotein receptor-related protein 5</td>
</tr>
<tr>
<td>LSL</td>
<td>Lox-STOP-Lox</td>
</tr>
<tr>
<td>LTR</td>
<td>Long terminal repeat</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Min</td>
<td>Multiple intestinal neoplasia</td>
</tr>
<tr>
<td>MMR</td>
<td>Mis-match repair</td>
</tr>
<tr>
<td>MMTV</td>
<td>Mouse mammary tumor virus</td>
</tr>
<tr>
<td>MSCV</td>
<td>Murine stem cell virus</td>
</tr>
<tr>
<td>MSI</td>
<td>Microsatellite instability</td>
</tr>
<tr>
<td>MTT</td>
<td>(3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, a yellow tetrazole</td>
</tr>
<tr>
<td>NCCN</td>
<td>National comprehensive cancer network</td>
</tr>
<tr>
<td>NES</td>
<td>Normalized enrichment score</td>
</tr>
<tr>
<td>NOTCH</td>
<td>NOTCH homolog 1, translocation-associated (Drosophila)</td>
</tr>
<tr>
<td>OS</td>
<td>Overall survival</td>
</tr>
<tr>
<td>PABP</td>
<td>Poly(A)-binding proteins</td>
</tr>
<tr>
<td>PARN</td>
<td>Poly(A)-specific ribonuclease</td>
</tr>
<tr>
<td>PERK</td>
<td>PKR-like eukaryotic initiation factor 2a kinase</td>
</tr>
<tr>
<td>PI3K</td>
<td>Phosphatidylinositol 3-Kinase</td>
</tr>
<tr>
<td>PTEN</td>
<td>Phosphatase and tensin homologue</td>
</tr>
<tr>
<td>RBP</td>
<td>RNA binding protein</td>
</tr>
<tr>
<td>RIDD</td>
<td>Regulated IRE1-dependent decay</td>
</tr>
<tr>
<td>RISC</td>
<td>RNA-induced silencing complex</td>
</tr>
<tr>
<td>RTK</td>
<td>Receptor tyrosine kinase</td>
</tr>
<tr>
<td>SB</td>
<td>Sleeping beauty</td>
</tr>
<tr>
<td>SD</td>
<td>Splice donor</td>
</tr>
<tr>
<td>sXBP1</td>
<td>Spliced X-box binding protein-1</td>
</tr>
<tr>
<td>TA</td>
<td>Transit-amplifying</td>
</tr>
<tr>
<td>TCGA</td>
<td>The cancer genome atlas</td>
</tr>
<tr>
<td>TLR</td>
<td>Toll-like receptor</td>
</tr>
<tr>
<td>TXNIP</td>
<td>Thioredoxin-interacting protein</td>
</tr>
<tr>
<td>UPR</td>
<td>Unfolded protein response</td>
</tr>
<tr>
<td>WNT</td>
<td>Wingless-type</td>
</tr>
<tr>
<td>WT</td>
<td>Wild-type</td>
</tr>
<tr>
<td>XBP1</td>
<td>X-box binding protein-1</td>
</tr>
</tbody>
</table>
Chapters 1-4 are the basis of the following manuscript in preparation:


Personal Contribution. Zhao, L contributed to design of research. Zhao, L performed the cellular and molecular biology studies including studies of: CNOT1 role in cell viability, role of TXNIP in cell viability, regulation of *TXNIP* by CNOT1, and role of the Unfolded Protein Response. In addition Zhao, L isolated RNA samples for Illumina microarrays and RNA Sequencing. Zhao, L carried out analysis of these studies, including carrying out IPA and GSEA analyses and statistical analysis.
Chapter 1

Introduction

Epidemiology and Challenges in Colorectal Cancer Treatment

Colorectal cancer (CRC) is the third most common cancer in men (663,000 cases) and the second in women (571,000 cases) worldwide. About 608,000 deaths from CRC were estimated to occur worldwide, accounting for 8% of all cancer deaths, making it the fourth most common cause of death from cancer (Ferlay et al., 2010). In the United States, CRC is the third leading cause of cancer death accounting for 142,820 new cases and 50,830 deaths in the year 2012 (Siegel et al., 2013). In China, the incidence rate of CRC is relatively lower than western countries. However, the rate has increased dramatically in recent years partially due to the changes in lifestyle and environmental pollution. In 2008, China reported 220,000 new cases and 109,000 deaths from CRC, the highest number to date in each category of CRC. These statistics highlight the significant impact of CRC on our society and thus the importance of new diagnostic and therapeutic tools for this cancer.

Up to 95% of all CRCs are gland-forming adenocarcinomas derived from the crypt epithelium (Parkin et al., 2001). Other types of malignancy account for the remaining 5% of CRC, including mucinous tumors, signet ring tumors, adenosquamous carcinoma, and undifferentiated carcinoma (Jessurun et al., 1999). CRC often causes no symptoms until it reaches an advanced stage. Symptoms may be present, but are nonspecific and vague, including a change in bowel habits, weight loss, abdominal pain, hematochezia, melena, fatigue and weakness (Benedix et al., 2010). The diagnosis of CRC is usually made on the basis of histopathology of surgical or biopsy specimens. Imaging, including CT and MRI scans of the abdomen and pelvis and lung, is used to accurately stage patients. CRC tissues are usually graded as well-, moderately- and poorly-differentiated based on how closely the tumor cells resemble normal epithelial cells. Once the diagnosis of CRC is made, the disease should be staged. Accurate staging
leads to both proper management and prognostic significance for the patient. The two most common staging systems to define the extent of invasion of CRC are the Dukes' and the TNM staging system. In the TNM staging system, patients are stratified according to the depth of tumor penetration (T), number of positive lymph nodes (N) and extent of distant metastasis (M) (Edge et al., 2009). Although pathological staging is useful in determining the treatment strategy and predicting the prognosis of CRC patients, it has some disadvantages. For example it cannot distinguish stage II patients who will relapse from those who will be cured by surgery alone. A novel staging system integrating the genetic changes in specific cancers may be more effective in selecting therapy and predicting prognosis (Eschrich et al., 2005).

The last decade has witnessed new strategies for the diagnosis and treatment of CRC which have improved patient survival. All treatment decisions are determined on an individual basis based on the stage, differentiation, pathological type of cancer as well as patient’s age, general health, and other medical conditions. Surgery, radiation, chemotherapy or a combination of these approaches are among the options available to treat patients with CRC. Surgery is the most common treatment for all stages of CRC and is the only potentially curative treatment (Lee et al., 2007). Chemotherapy is widely used for treatment of stage III CRC post-surgically to prevent the recurrence of disease. It can also be used as the primary therapy for advanced CRC when surgery is not indicated (NCCN 2012). The use of adjuvant chemotherapy in stage II disease remains unclear and controversial (André et al., 2004). Surgery alone is the predominant therapy for stage II CRC patients as large clinical trails have not demonstrated a significant benefit to a surgery plus chemotherapy regimen in treating stage II disease. However, up to 25% of stage II CRC patients rapidly relapse following surgery alone therapy. Thus, there is a subset of stage II patients who would benefit by more aggressive therapy. This highlights the importance of genetic studies to distinguish this group of high-risk patients from low risk patients. The options for treatment of advanced disease are still limited. Even given the most effective regimen, 50% of stage III patients relapse (NCCN 2012). For stage IV patients, the situation is even worse. The overall survival is only around two years following treatment with surgery combined with standard FOLFOX chemotherapy and
bevacizumab (NCCN 2012). These limitations and challenges in CRC management highlight the importance of research, especially genetic research, in improving outcomes in CRC patients, for example, in the identification of biomarkers for further stratifying patients and as targets for emerging novel therapeutic strategies.

**The Biology of the Human Intestinal Tract**

The human intestinal tract extends from the stomach, the small intestine, and proceeds to the large intestine, and terminates at the anus. The small intestine is composed of the duodenum, jejunum and ileum, and is about 20-25 feet long on average, and is the site where most of the digestion and absorption of nutrients is carried out. The large intestine, about 5 feet long, consists of the caecum with appendix extension; the ascending, transverse, descending and sigmoid colon; and the rectum. The large intestine is horseshoe shaped and extends around the small intestine. A major physiological function of the large intestine is to absorb water from the remaining indigestible food matter and temporarily store waste matter. Besides nutrient digestion and absorption, the intestinal tract, in particular the large intestine, also plays important roles in immunological defense, water homeostasis, and entero-endocrine signaling, and other functions.

The intestinal wall is composed of four layers, which are the serosa, muscularis, submucosa and mucosa from outside to inside (intestinal lumen). Each layer has different tissues and functions. The serosa layer consists of connective tissue and acts as a covering for the digestive tube. The muscularis layer surrounds the submucosa and is responsible for intestinal peristalsis. Rich in vessels like arteries, veins, lymphatics and nerves, the submucosa layer provides the intestine with blood supply and is the location of a varaiety of non-epithelial cells (*e.g.*, fibroblasts, cells of the innate and adaptive immune systems) that comprise the intestinal stroma. The mucosa is the innermost layer of the intestinal tract which is lined by absorptive and secretory epithelium. The epithelium consists of a simple columnar epithelium layer with millions of “crypts of Lieberkuhn” that are finger-like invaginations of the epithelium into the underlying connective tissue (Figure 3). There are four classes of terminally differentiated cells in
small intestinal crypts. The enterocytes, the most abundant cell type in the intestinal epithelium, are absorptive cells while the other three classes are all secretory: goblet cells secrete mucus; enteroendocrine cells are smaller and secrete various gut hormones (peptides and catecholamines) that regulate numerous important biological processes; and Paneth cells secrete digestive enzymes, growth factors and especially diverse host defense (antimicrobial) proteins such as α-defensins, secretory phospholipase A2 and various proinflammatory cytokines (e.g. TNF-α, IL-1β) to protect the constantly renewing crypt from harmful invaders. The composition of cell types in the colon is similar to that of the small intestine, except for the absence of villi and Paneth cells.

All types of terminally differentiated cells originate from the differentiation of progeny of intestinal stem cells (ISCs) which are located at the base of the crypt in both the small and large intestine. In the small intestine, ISCs are interspersed with Paneth cells (Sato et al., 2011). Under normal conditions, the intestinal epithelial cells are renewed every 2-7 days, which is fueled by stem cells (Figure 4, Medema et al., 2011). ISCs are responsible for generating the entire crypt-villus structure. In the small intestine, Paneth cells migrate downward to reside at the bottom of the crypt and live for 6-8 weeks. Slightly higher up the crypt are the transit amplifying (TA) cells and committed progenitor cells that undergo differentiation into the enterocytes, goblet cells, and enteroendocrine cells as they migrate from the base of the crypt towards the lumen. When the migration of these differentiated cells reaches the villus tip (small intestine) or intercrypt table (colon), they undergo apoptosis and are shed into the intestinal lumen.

There are two hypothetical models of stem cells in the intestine (Clevers et al., 2013). The “+4 model” refers to a common intestinal stem cell origin at position 4-5, directly above the Paneth cell compartment. Potten et al. showed that the radiation-sensitive, label-retaining +4 cells were actively proliferating with 24 hour cell cycles. The other model is termed as “the stem cell zone model” which is based on the finding that the crypt base is not solely populated by postmitotic Paneth cells. Instead, a population of slender crypt-base-columnar (CBC) cells was wedged between the Paneth cells located at position 1-5, which were radiation-sensitive but not label-retaining and were called
intestinal stem cells (Cheng and Leblond, 1974). Although the identity of intestinal stem cells is still controversial, it is agreed that intestinal stem cells reside in a niche that provides them with essential signals from important morphogenetic pathways such as WNT, NOTCH, and BMP (Figure 5, Heath JK 2010, Crosnier \textit{et al}., 2006, Reya and Clevers 2005).

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{figure1}
\caption{The anatomy of the gastrointestinal tract. (Nucleus Medical Art, Inc. 2008)}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{figure2}
\caption{The multi-layered organization of the mature GI tract. (Heath JK 2010).}
\end{figure}
Figure 3. The organization of the small intestinal crypt-villus and the colon crypt. Both the colon crypt (left) and the small intestinal crypt (right) are shown. (Medema and Vermeulen, 2011).

Figure 4. Organization of the intestinal epithelium and the crypts of Lieberkühn and cell lineage determination in the intestinal epithelium. (Heath JK 2010).
Important Signaling Pathways in Intestinal Crypt Renewal

Specifically, several key regulatory signals have been reported to be involved in intestinal stem cell renewal and differentiation, including the WNT, bone morphogenetic protein (BMP), and NOTCH pathways (Heath JK, 2010, Reya and Clevers, 2005). Furthermore, evidence suggests there is a close interaction between several key pathways in directing intestinal epithelial stem cell renewal and differentiation (Figure 4).

**WNT Signaling Pathway**

WNT signaling is crucial in controlling cell proliferation/regeneration in the intestinal epithelium and the maintenance of intestinal stem cells (Polakis, 2007). WNT proteins are extracellular glycoproteins produced by epithelial cells in the crypt, which bind the Frizzled and the low density lipoprotein receptor-related protein 5 (Lrp5) receptors on the surface of intestinal epithelial cells. β-catenin is the key effector of the canonical WNT signaling pathway. In the absence of WNT binding, cytoplasmic β-catenin levels are normally kept low through continuous degradation, which is controlled by the APC/Axin/CK1/GSK3 destruction complex. This complex binds to newly synthesized β-catenin and phosphorylates it, thereby targeting phosphorylated β-catenin for ubiquitination-mediated proteolysis. When WNT proteins bind to the Frizzled/LRP receptor complex at the cell surface, the destruction complex is inactivated, resulting in the accumulation of β-catenin in the cytoplasm. The accumulated β-catenin then translocates into the nucleus to bind with Tcf/Lef transcription factors, leading to transcriptional activation of important target genes involved in cell proliferation and oncogenesis such as *Cyclin D1* and *c-MYC* (Komiya et al., 2008, Cadigan et al., 2012). The expression level of components of the WNT signaling is highest at the bottom of intestinal crypts. Canonical WNT signaling plays a major role in maintaining intestinal stem cell fate and stimulating progenitor cell proliferation (Reya and Clevers, 2005). For example, in *Tcf4*−/− mice, the proliferating compartment between the villi was entirely absent, suggesting an essential role for Wnt in maintenance of crypt progenitors (Korinek et al., 1998). Dysregulation of the WNT signaling pathway was shown to cause aberrant proliferation of colonic crypts and is the rate-limiting step in CRC development.
example, in 80% of CRCs both APC alleles are inactivated. Furthermore, mutations in other canonical WNT pathway components have also been observed in CRC. For example, somatic mutations in AXIN1 (Jin et al., 2003) and the TCF4 transcription factor have been observed in a subset of CRCs (Angus-Hill et al., 2011). Activating mutations in the β-catenin gene (CTNNB1) that protect the protein from APC-mediated degradation are also observed in CRC and adenomas (Kaler et al., 2012).

**NOTCH Signaling Pathway**

Another important signaling pathway for intestinal homeostasis and differentiation is the NOTCH signaling pathway. NOTCH signaling mediates cell-cell interactions in animal development and is known to regulate the proliferation and differentiation of intestinal stem cells. Mammals express four transmembrane Notch receptors (Notch1-4) and five canonical transmembrane ligands (Delta-like [DLL] 1,3, and 4, Jagged-1, 2) (Espinoza et al., 2013). In canonical NOTCH signaling, when NOTCH receptors bind to their ligands expressed on neighboring cells, this initiates proteolytic cleavage of the receptors by α-secretase and γ-secretase to release NOTCH receptors’ intracellular domain (NICD). The cleaved receptors then translocate into the nucleus and form complexes with RBP-jκ and induce transcriptional activation of NOTCH target genes such as transcription factor hairy/enhancer of split (Hes1) which is important for NOTCH induced repression of differentiation. Knockout of Hes1 in intestinal tumor cells in Apc-mutant mice was reported to reduce tumor cell proliferation, while promoting differentiation toward epithelial lineages (Ueo et al., 2012). Crucial components of the NOTCH signaling pathway have been shown to localize to the proliferative zone of the intestinal crypts but only weakly in the intestinal villi (Jensen et al., 2000, Riccio et al., 2008). NOTCH signaling promotes progenitor cells’ differentiation into the absorptive cell lineage rather than the secretory cell lineage (Jensen et al., 2000, Riccio et al., 2008) through inhibition of the transcription factor atonal homolog 1 (Atoh1) which appears to be required for secretory cell differentiation (VanDussen and Samuelson, 2010). Thus, disruption of NOTCH signaling results in increased Atoh1 expression and secretory cell hyperplasia, whereas hyperactive NOTCH
signaling results in decreased *Atoh1* expression and increased numbers of absorptive enterocytes (VanDussen *et al.*, 2012). Importantly, NOTCH signaling does not act alone but interacts with other key signaling pathways, such as WNT signaling, to promote functional signaling networks required for development and tissue homeostasis. In addition, there may be a cross-talk between the Notch and RAS signaling pathways, in which RAS activating mutations can activate Notch signaling while Notch activation is required for RAS-mediated transformation (Saif and Chu, 2010). NOTCH signaling also has been shown to be a key participant in epithelial-to-mesenchymal transition (EMT) that is an important process for metastatic disease (Saad *et al.*, 2010).

**BMP Signaling Pathway**

The bone morphogenetic protein (BMP) pathway, originally identified as factors inducing cartilage and bone formation, also has been shown to play a key role in gastrointestinal cell development and in maintaining tissue homeostasis (Itoh *et al.*, 2000). BMPs belong to the transforming growth factor beta (TGFβ) superfamily, which share intracellular signaling through the SMAD proteins. The action of BMP is mediated by type 1 (BMPRI) and type 2 (BMPR2) serine/threonine kinase receptors. In the canonical BMP pathway, the binding of BMP to BMPRI results in activation of BMPRI, leading to phosphorylation of SMAD proteins (SMAD1, SMAD5 and SMAD8). The phosphorylated BMP-SMADs form a complex with a co-SMAD (SMAD4) and translocate into the nucleus. Once in the nucleus, the SMAD complex cooperates with other transcription factors to modulate target gene expression. BMP signaling within the intestine is quite complex as BMPs and their receptors are expressed in both epithelial and mesenchymal compartments. It has been suggested that BMP2 and BMP4, expressed by the mesenchymal cells, can halt proliferation at the crypt-villus border, as BMPs are active at the top of the crypt where differentiation occurs. Though also produced at the crypt base, they are kept in check by BMP inhibitors such as Noggin. BMP signaling has been reported to antagonize WNT signaling along the crypt-villus axis to inhibit intestinal stem cell activation and expansion. This signaling pathway has also been shown to be involved in lineage fate decisions of secretory cells (He *et al.*, 2004, Derynck *et al.*, 2003).
The human adult colonic epithelium undergoes perpetual regeneration fueled by intestinal epithelial stem cells and progenitor cells, which is a highly coordinated process regulated by different signaling pathways. Disruption of the signaling pathways critical for stem cell renewal may result in a series of diseases in GI tract, including CRC.

**Hereditary and Sporadic CRC**

CRC can be subdivided into hereditary or sporadic cancers depending on the presence or absence of familial genetic predisposition. Around 30% of diagnosed CRC are considered to be hereditary, while the majority of the CRC cases (70%) occur in sporadic forms (Figure 5, Lynch *et al.*, 2003).

**Figure 5. The fraction of colon cancer cases that arise in various family risk settings.** (Burt, 2000)

**Hereditary CRC**

Around 30% of the diagnosed CRC cases are considered to be hereditary, including dominantly inherited CRC syndromes, such as Familial Adenomatous Polyposis (FAP), Hereditary Non Polyposis Colorectal Cancer (HNPCC), and MUTYH Associated Polyposis (MAP), which together account for 5-10% of cases and CRC in patients from families with familial susceptibility (~25-30% of cases) (Chapelle *et al.*, 2004).

FAP is an autosomal dominant syndrome which causes 0.5-1.0% of CRCs and affects approximately 1 in 10,000 births (Fearon *et al.*, 2011, Rustgi *et al.*, 2007). FAP
patients usually carry germline mutations in *APC*, typically frameshift or nonsense mutations. FAP is characterized by hundreds to thousands of adenomas in the colon and rectum, which start to appear in the patient’s second decade of life. Some of these adenomas inevitably progress to adenocarcinoma and metastatic disease unless the patient undergoes prophylactic colectomy. Hence, FAP patients are recommended to have endoscopic screening from 10 years old. Genetic studies of FAP led to understanding the role of *APC* somatic mutations in also causing sporadic CRCs.

Also known as Lynch syndrome, HNPCC is an autosomal dominant condition which causes ~2-5% of CRCs and is a genetically heterogeneous disease (Fearon, *et al*., 2011). HNPCC patients have germline mutations in genes involved in DNA mismatch repair (MMR), with the majority of known HNPCC mutations being found in the genes *MSH2* and *MLH1*. Tumors from HNPCC patients also often contain microsatellite instability (MSI). There is a more rapid transition from adenomatous lesion to carcinoma compared to sporadic cancers, as affected cells accumulate mutations swiftly due to impaired DNA repair mechanisms. HNPCC patients have an 80% lifetime risk of developing CRC that typically develop in patients in their 40’s, compared with an average incidence age for sporadic CRC in the 60’s. HNPCC also increases the risk of other cancers such as gastric, ovarian, endometrial, renal and hepatobiliary cancers (Johns Hopkins Medicine CRC).

**Genetic Alterations in Sporadic CRC**

The majority of CRC (~70%) is considered sporadic, reflecting no (or very little) family history of the disease. Sporadic CRC are different from dominant hereditary syndrome forms in that they arise from an accumulation of somatic mutations in rate-limiting tumor suppressor genes and oncogenes rather than inheriting germline mutations (Rustgi *et al*., 2007). Both sporadic and inherited CRC emerge from normal colonic epithelium as a result of a sequence of genetic mutations and genomic alterations (including epigenetic changes) that lead to uncontrolled cell division and tumor formation. Identification of the genetic abnormalities that accumulate in a step-wise manner has led to the well-known model of the adenoma-carcinoma sequence initially
characterized by Fearon and Vogelstein (Figure 6, Fearon et al., 1990). This model describes the stepwise progression from normal glandular epithelial cells to benign neoplasia (adenoma) to carcinoma and eventually metastatic cancer and correlates specific genetic events with evolving tissue morphology. In the model that accounts for ~80% of CRC, mutations in the APC gene result in the activation of WNT signaling, which promotes the formation of small adenomas in the form of polyps. Subsequent mutations in KRAS, BRAF, P53, and other genes promote the progression to carcinomas, and finally additional mutations drive tumor metastasis (Walther et al., 2009). As discussed below, many additional mutations that confer a selective advantage for tumor growth are required for the development of a CRC. The progression from adenoma to invasive carcinoma can take years or even decades (Walther et al., 2009). This concept not only provides an excellent model to study the genetics of invasive cancer, but also affords a means of preventing CRC by endoscopic removal of precursor lesions (Jass et al., 2007).

Figure 6. Schematic representation of the Vogelstein model of colonic carcinogenesis. (Johns Hopkins Medicine CRC)

Identification and characterization of genes that drive cancer processes are critical for the development of cancer diagnostic and therapeutic modalities. Over the past three decades, molecular genetic studies have revealed some critical mutations at high frequency underlying the pathogenesis of the sporadic and inherited forms of CRC, including APC, K-RAS, and TP53.
APC has been reported to be involved in multiple critical cellular functions. The best-established role for APC in CRC is as a major negative regulator of the canonical WNT signaling pathway. Germline mutations in APC play a key role in FAP. More than 95% of known mutations in APC in FAP patients are frameshift or nonsense mutations that lead to premature truncation of protein synthesis (Segditsas et al., 2006). FAP individuals who inherit germline mutations in APC develop thousands of colorectal adenomas by their teens with the inevitable progression of some of these to carcinoma, unless the intestine is surgically resected (Kinzler and Vogelstein, 1996). In addition to the role in FAP, APC plays a very prominent role in sporadic CRC. Loss of function mutations in APC occur in over 80% of sporadic CRC (Kinzler and Vogelstein, 1996). Nearly all somatic mutations lead to premature truncation of the APC protein (Segditsas et al., 2006). Somatic APC mutations are an early and rate-limiting event in the development of most adenomas (Powell et al., 1992). The inactivation of both alleles of the APC gene can be detected in most of the intestinal tumors at early stages of development (Powell et al., 1992), in agreement with Knudson's two hit hypothesis. However, accumulated mutations in additional genes are required for progression of the early lesions to cancer.

Activating mutations in K-RAS (Kirsten rat sarcoma viral oncogene homolog), a member of the RAS family of proto-oncogenes, represent a second genetic alteration which occurs early in the evolution towards intestinal cancer (Plesec et al., 2009). The RAS family of small-G proteins function as molecular switches downstream of multiple receptor tyrosine kinase (RTK) growth factor receptors, such as the epidermal growth factor receptor (EGFR) family (Malumbres et al., 2003, Zhao et al., 2008). The product of the proto-oncogene K-RAS is involved in transduction of the MAPK (mitogen activated protein kinase) signaling pathway. K-RAS mutations in specific codons occur in approximately 40% of CRC and lead to the constitutive activation of the RAS/RAF/MEK/ERK signaling pathway, resulting in the transduction of signals involved in cell proliferation and apoptosis inhibition and thus, contributing to malignant transformation (Kim and Lance, 1997). K-RAS mutations contribute to colorectal adenoma progression but are not required for adenoma initiation, which is supported by
the fact that $K$-RAS mutations are found in at least 50% of colorectal adenomas larger than 1 cm and in carcinomas (Vogelstein et al., 1988). $K$-RAS mutations occur after APC mutations in the adenoma-to-carcinoma progression sequence, but are still a relatively early event in tumorigenesis. Acquired $K$-RAS mutations are maintained throughout carcinogenesis, as evidenced by the concordance of $K$-RAS mutation status in primary and metastatic colorectal cancer (Artale et al., 2008).

The phosphatase and tensin homologue (PTEN) gene codes for a lipid phosphatase which functions as a negative regulator of the phosphoinositide-3-kinase (PI3K)/$\gamma$-akt murine thymoma viral oncogene homolog 1 (AKT) pathway. The loss of PTEN, which occurs through different mechanisms, probably acts to enhance effects downstream of the K-RAS protein via PIP3-mediated activation of AKT that acts on downstream anti-apoptotic factors and the mTOR pathway, resulting in persistent oncogenic signaling (Cully et al., 2006, Chalhoub et al., 2009). Germline PTEN mutations are found in Cowden syndrome, leading to hyperactivity of the mTOR pathway, and an increased risk of some benign GI tumors but not an elevated risk of CRC (Hobert et al., 2009). Some studies suggest that PTEN protein expression may be lost in 15-20% of CRCs (Laurent-Puig et al., 2009). Other studies indicate that loss of PTEN protein expression measured by IHC occurs in 30% to 45% of CRCs (Loupakis et al., 2009).

Loss of heterozygosity (LOH) is a key mechanism for inactivating one allele of certain tumor suppressor genes in cancer. Approximately 70% of CRCs are characterized by loss of the short arm of chromosome 17 (17p) (Fearon et al., 1990, Vogelstein et al., 1988). LOH at this locus correlates with the transition from benign adenoma to invasive cancer (Vogelstein et al., 1988, Baker et al., 1990). The TP53 gene is thought to be the main target of 17p LOH (Baker et al., 1990). In addition to 17p LOH, the function of TP53 may be lost due to intrachromosomal somatic mutations. Approximately 85% of TP53 mutations in CRC are missense defects, which most often occur at codons 175, 245, 248, 273, and 282 (Wood et al., 2007, Baker et al., 1990). The few nonsense and frameshift mutants encode a truncated TP53 product. TP53 is regarded as the "guardian
of the genome" due to its functions in cell-cycle checkpoints at the G1/S and G2/M boundaries, in promoting apoptosis, and in restricting angiogenesis (Vousden et al., 2009). TP53 mutations and LOH of TP53 alterations have been reported in 4-26% of adenomas, approximate 50% of in situ carcinomas, and in 50-75% of adenocarcinoma (Kaserer et al., 2000). This result is indicative of the central role of loss of TP53 function in the adenoma to carcinoma transition. Selection for TP53 mutations in CRC may be due in part to the loss of TP53 function in regulating key target genes, including p21, BAX, and MDM2 (Vousden et al., 2009). Loss of function in TP53 may facilitate continued growth and the acquisition of invasive properties in the face of stresses, including DNA-strand breakage, telomere erosion, hypoxia, and reduced nutrient exposure, that might otherwise severely limit tumor cell survival and maintenance at the adenoma-carcinoma transition.

However, these documented high frequency mutations do not completely explain all CRCs, since more mutations need to be acquired in addition to these high frequency mutations in the progression to highly invasive carcinomas. Moreover, some CRCs lack these high frequency mutations. Therefore, genome-wide analyses of somatic mutations in CRC were performed to obtain a deeper understanding of genetic alterations in CRC. The first genome-scale study of somatic mutations in CRC was conducted by the Vogelstein group at Johns Hopkins University (Wood et al., 2007). They identified a small population of genes that were frequently mutated in CRCs. These were called “mountains”, including genes such as APC, KRAS, and TP53. Surprisingly, their study identified a much larger population of genes mutated at a lower frequency, hence the term “hills” (Figure 7, Wood et al. 2007). They then analyzed the influence of the low frequency mutations in signaling pathways by determining whether the genes within specific pathways were mutated more often than predicted by chance. They found that 38 pathways were preferentially mutated in CRC. Within the 38 pathways, many of them were involved in the PI3K signaling. Additional pathways altered in CRC were related to cell adhesion, the cytoskeleton, and the extracellular matrix. These findings suggest that tumor growth in individual human CRCs may be driven by some genes that are mutated at a low frequency though affecting important signaling pathways.
Figure 7. Human colorectal cancer genome landscape. (Wood et al. 2007).

To identify a more complete list of genes dysregulated in CRC, large-scale exome and genome sequencing studies of human CRCs have been conducted by multiple groups (Cancer Genome Atlas Network 2012, Han et al., 2013, Schee et al., 2013). These studies have revealed that there are hundreds to thousands of DNA alterations in the average CRC genome, and that there are roughly 80–100 genes that are commonly altered by non-synonymous mutations, and roughly 15 mutations in candidate ‘driver’ genes per cancer genome. Somatic mutations can be classified as either driver mutations or passenger mutations. Driver mutations confer a growth advantage to the cells carrying them in the neoplastic process and are positively selected for during tumorigenesis. They reside, by definition, in the subset of genes known as ‘cancer genes’ that are causally implicated in cancer development. Passenger mutations provide no positive or negative selective advantage to the tumor but are retained by chance during repeated rounds of cell division and clonal expansion (Grady et al., 2013, Wood et al., 2007). However, the ‘passengers’ can be converted into ‘drivers’ under certain selective environments, such as chemotherapy (Roche-Lestienne et al., 2002, Mullighan et al., 2008). The number of driver mutations in an individual cancer is a central conceptual parameter of cancer development, but is not well established. On the basis of age-incidence statistics it has been suggested that common adult epithelial cancers such as breast, colorectal and prostate cancers require 5-7 rate-limiting events, possibly equating to drivers, whereas
cancers of the haematological system may require fewer (Stratton et al., 2009). These estimates are supported by experimental studies which show that engineering changes in the functions of at least 5 or 6 genes in normal primary human cells are necessary to convert them into cancer cells (Schinzel et al., 2008). However, recent analyses of somatic mutation data from cancers indicate that the number of drivers might be as many as 20 driver mutations in individual cancers (Beerenwinkel et al., 2007). Ultimately, direct estimates of the number of drivers in individual cancers will be provided by identifying all the cancer genes and systematically measuring the prevalence of genomic mutations and epigenetic alterations in them.

While identifying the whole constellation of genomic abnormalities is significant, identifying those that are cancer drivers is even more meaningful. It appears that most somatic mutations in cancer genomes are likely passenger events (Greenman et al., 2007). A major challenge to all systematic mutation screens will, therefore, be to distinguish the driver mutations hidden within a much larger number of passenger mutations within the human cancer genome (Voidonikolas et al., 2009). Researchers have been tackling this challenge by improving experimental conditions and developing more sophisticated statistical models and functional analysis strategies. Performing an unbiased forward genetic screen in mice provides the tools to generate tumors and analyze their genetic composition, while reducing the background of passenger mutations. A successful strategy for such a genetic screen has been developed using the *Sleeping Beauty* transposon system to induce tumors in transgenic mice (Starr et al., 2011).

**Sleeping Beauty Transposon-Based Forward Genetic Screens**

The *Sleeping Beauty (SB)* transposon system is composed of a Sleeping Beauty transposase and a synthetic DNA transposon. The resurrected transposase gene was named "Sleeping Beauty" because it was brought back to activity from a long evolutionary silencing (Ivics et al., 1997). It was re-constructed from extinct (fossil) transposase sequences belonging to the Tc1/mariner class of transposons found in the genomes of salmonid fish, which have been inactive for more than 10 million years due to accumulated mutations (Doak et al., 1994, Radice et al., 1994). The transposon
component of the $SB$ system was isolated from the genome of another salmonid fish, *Tanichthys albonubes*. The transposon was designed with an expression cassette of interest flanked by inverted repeat/direct repeat (IR/DR) sequences of about 340 base pairs each. Similar to other Tcl/mariner-type transposases, the $SB$ transposase catalyzes the excision of DNA sequences within the IRs and the reintegration of this sequence into a recipient DNA in a cut-and-paste manner. $SB$ transposons integrate nearly randomly into chromosomes at TA-dinucleotide base pairs, making it act as an effective tool for random mutagenesis, although the sequences flanking the TAs may influence the probability of integration at a given site (Figure 8, Geurts *et al*., 2006). As a result, DNA cargo in a chromosomally integrated transposon can be expressed over the lifetime of a cell.

The $SB$ transposon system has several advantages over other mutagenesis systems. First, transposition efficiency of the $SB$ transposase is greatly enhanced, enabling highly efficient transgenic integration and expression (Mates *et al*., 2009, Grabundzija *et al*., 2010). Second, $SB$-based vectors are likely superior in safety for two main reasons. $SB$ shows a random genomic insertion profile without overt preference for integrating into genes and their transcriptional regulatory regions (Grabundzija *et al*., 2010, Ammar *et al*., 2012). In addition, the IRs sequence of $SB$ vectors has negligible enhancer/promoter activity (Moldt *et al*., 2007, Walisko *et al*., 2008). Third, there are no $SB$-related sequences in mammalian genomes, thereby precluding potential cross-mobilization between endogenous and exogenously introduced transposons (Newman *et al*., 2008).

Since the $SB$ transposon system can mediate efficient integration and stable long-term expression of a precisely defined DNA sequence, over the past decade, this system has been widely used in studying functional genomics, insertional mutagenesis, transgenesis, and somatic gene therapy.
Large scale sequencing of cancer genomes has revealed a vast array of somatic mutations in breast, colon, lung and pancreatic cancers (Wood et al., 2007, Ding et al., 2008, Jones et al., 2008). Distinguishing driver mutations from passenger mutations which are in a much larger number within the human cancer genome has been a major task for biologists. The SB transposon-based unbiased forward genetic screen in mice provides a tool to generate tumors and identify potential driver mutations, while reducing the background of passenger mutations, because the tumor formation process could select and enrich for the driver mutations.

An SB-based screen is initiated by breeding mice carrying SB transposon and transposase transgenes together. Further, tissue-specific expression of the transposon can be achieved using conditional expression alleles of the transposase that are activated by a tissue specific Cre recombinase. In addition, the system can be used in mice of differential susceptibility to sporadic cancers. For example, this permits the identification of genes important for tumor development in wildtype mice but also permits the identification of genes that cooperate with known cancer genes, such as APC. The mutagenic SB transposon was designed to create both loss- and gain-of-function mutations similar to those that drive tumorigenesis in sporadic human cancers. Recent modifications to the system using an inducible, tissue-specific transposase under the
control of a tissue-specific Cre recombinase allele have been shown to produce a wide range of epithelial-derived carcinomas (Dupuy et al., 2009) and tissue-specific cancers including GI neoplasms (Starr et al., 2009). Tumors from these mice are isolated and genomic DNA is extracted. Next generation sequencing is then used to identify transposon insertion sites in the tumors and bioinformatics tools are employed to identify high frequency common insertion sites (CIS) (Starr et al., 2009). A CIS is defined by analyzing transposon insertions in many tumors and identifying genomic loci that contain transposon insertions at a higher rate than would be expected by chance. The presence of a CIS indicates that a transposon-mediated mutation in that locus probably has contributed to tumor development. By analyzing the genes within the CIS, one can identify candidate cancer genes which need to be further validated by biological investigations. Cancer genetic fingerprints are obtained that can be characterized by networks of interacting cancer-gene mutations. These genes can be queried in relevant human cancers (Hackett et al., 2013). The process of devising and executing a forward genetic screen using the SB transposon system to identify candidate cancer genes is illustrated in Figure 9.

Figure 9. Using SB system for cancer gene screens in mice. (Hackett et al., 2013)
Due to the characteristics of the SB system, there are several advantages of using the SB system to identify candidate cancer genes. First, the transposon can easily be located in the genome because its sequence is known. Second, transposition can be directed to almost any cell type. Third, the transposon is capable of introducing both gain- and loss-of-function mutations (Bergemann et al., 2012).

The SB transposon-based screens have been successful in a number of studies that yielded lists of CIS genes which are potential cancer genes in colon (Starr et al., 2009, March et al., 2011), liver (O'Donnell et al., 2012), pancreatic (Pérez-Mancera et al., 2012) and skin (Quintana et al., 2013) cancers. Following functional characterization will be helpful in understanding the roles of these novel candidate cancer genes in different types of cancers. In addition, the use of this technique may lead to the eventual design of new therapies aimed to eliminate the results of these specific genetic mutations.

**The CCR4-NOT (CNOT) Complex and CNOT1**

*Cnot1* (Table 1) and other subunits of the CNOT complex, including *Cnot2, Cnot4, Cnot6, Cnot6l, and Cnot8*, were identified as CIS associated candidate cancer genes in CRC and other types of cancers in multiple SB screens in mice (Koso et al., 2012, March et al., 2011, Pérez–Mancera et al., 2012, Rahrmann et al., 2013, Starr et al., 2009, Starr et al., 2011, Wu et al., 2012). The updated information of CIS candidate genes in all published SB-based screens in different cancer types is available on the website of the Candidate Cancer Gene Database (CCGD) which is maintained by Dr. Starr’s group at the University of Minnesota (http://ccgd-starrlab.oit.umn.edu/search.php).

The CNOT complex is a multicomponent complex evolutionarily conserved from yeast to human, which has a crucial role in regulation of gene expression at multiple levels (Collart et al., 2004). To date, the structural components and dynamics of this complex are still not completely known. The human CNOT complex is thought to be composed of 11 subunits (Figure 10, Boland et al., 2013). CNOT1 functions as a modular scaffold to provide binding sites for other CNOT subunits. Deadenylation is one of the main functions of the CNOT complex. The CNOT complex contains four deadenylases,
CNOT6 and its paralog CNOT6L, and CNOT7 and its paralog CNOT8, which are categorized into the exonuclease–endonuclease–phosphatase (EEP) and Asp-Glu-Asp-Asp (DEDD) families of nucleases based on their nuclease domains (Lau et al., 2009). All four deadenylases have deadenylation activity. Different deadenylase subunits have been shown to have various target genes and deadenylases of the same category have partial redundancy in function (Lau et al., 2009). The conserved core of the CNOT complex consists of two major modules: a catalytic module comprising the middle region of CNOT1 and two deadenylases (CNOT6 or CNOT6L, and CNOT7 or CNOT8), and the NOT module which minimally consists of the CNOT1 C-terminal region, CNOT2 and CNOT3. The integration of additional subunits has been shown, including CNOT4, CNOT5, CNOT9, CNOT10 and CNOT11 (Boland et al., 2013). The catalytic module is responsible for the deadenylation of specific mRNAs. The precise function of the NOT module remains largely unclear. It has been shown to regulate the stability and activity of the catalytic module and plays a crucial role in the recruitment of the CNOT complex to mRNAs by microRNAs (Wahle et al., 2013). In addition to the role in mRNA deadenylation/degradation, the CNOT complex is also involved in the transcriptional repression of nuclear receptor-mediated transcription in a ligand-dependent manner via CNOT1 and CNOT2 (Winkler et al., 2006). The CNOT complex also plays a role in protein ubiquitination/degradation via the CNOT4 subunit with an E3 ligase activity (Collart et al., 2004). Furthermore, the CNOT complex is associated with various other functions in cells, such as translational repression (Cooke et al., 2010), DNA repair (Mulder et al., 2005) and histone methylation (Mulder et al., 2007). Given the diverse activities associated with the CNOT complex, it plays a role in a wide range of biological processes, including cell proliferation, apoptosis, oogenesis and embryogenesis, spermatogenesis, bone formation and energy metabolism (Collart et al., 2012).
Figure 10. A proposed model for the human CCR4-NOT complex. (Bartlam et al., 2010)

*Cnot1* (Table 1, Figure 11) was identified as a CIS candidate cancer gene in CRC and other types of cancers in multiple SB screens in mice (Koso et al., 2012, March et al., 2011, Pérez–Mancera et al., 2012, Rahrmann et al., 2013, Starr et al., 2011).

<table>
<thead>
<tr>
<th>Cancer Type</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorectal Cancer</td>
<td>Starr et al., 2011</td>
</tr>
<tr>
<td>Colorectal Cancer</td>
<td>March et al., 2011</td>
</tr>
<tr>
<td>Pancreatic Cancer</td>
<td>Pérez–Mancera et al., 2012</td>
</tr>
<tr>
<td>Nervous System Cancer</td>
<td>Koso et al., 2012</td>
</tr>
<tr>
<td>Nervous System Cancer</td>
<td>Rahrmann et al., 2013</td>
</tr>
</tbody>
</table>

Table 1. *Cnot1* was identified as CIS candidate cancer gene by multiple SB screens.

Figure 11. Map of SB insertions in *Cnot1* in *Apc<sup>Min</sup>* screen. (Starr et al., 2011) Yellow and blue markers represent sites of transposon insertions on the same and opposite direction of *Cnot1* gene, respectively, in GI tumors.

The human *CNOT1* gene is located over 100kb on the long arm of chromosome 16 at position 16q21. *CNOT1* mRNA has three splicing variants, NM_016284.4,
NM_206999.2, and NM_001265612.1, encoding three protein isoforms NP_057368.3, NP_996882.1 and NP_001252541.1, respectively (Table 2). The variant NM_016284.4 (*CNOT1L*) encodes for the longest isoform of 2376 amino acids (aa) (Figure 13). The variant NM_206999.2 (*CNOT1S*) differs in the 3’ UTR compared to variant NM_016284.4. The encoded isoform NP_996882.1 is 1551aa and has a distinct C-terminus. The variant NM_001265612.1 (*CNOT1M*) uses an alternate splice site in the coding region, but maintains the reading frame, encoding an isoform NP_001252541.1 that is 2371aa.

<table>
<thead>
<tr>
<th>Transcript variant</th>
<th>mRNA length (nt)</th>
<th>Protein isoform</th>
<th>Protein length (aa)</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_016284.4</td>
<td>8476</td>
<td>NP_057368.3</td>
<td>2376</td>
<td>CNOT1L</td>
</tr>
<tr>
<td>NM_206999.2</td>
<td>5221</td>
<td>NP_996882.1</td>
<td>1551</td>
<td>CNOT1S</td>
</tr>
<tr>
<td>NM_001265612.1</td>
<td>8461</td>
<td>NP_001252541.1</td>
<td>2371</td>
<td>CNOT1M</td>
</tr>
</tbody>
</table>

**Table 2. Splicing variants and protein isoforms of CNOT1.**

Multiple somatic mutations in *CNOT1* in different types of cancers, in particular CRC, have been documented by several projects. The COSMIC database documented 49 mutations in *CNOT1* among 610 cases (8%). In the TCGA database, *CNOT1* mutations were identified in 11 out of 212 cases (5.2%) (Cancer Genome Atlas Network, 2012). In a study by Seshagiri *et al* (Seshagiri *et al.*, 2012), *CNOT1* mutations were identified in 9 out of 72 cases (12.5%) (Figure 12). Some mutations, such as R1733H and R1640Q, occurred multiple times. For all three databases, *CNOT1* is mutated most frequently in CRC. Most of these mutations are missense mutations and no specific cluster or hotspot was observed. The high *CNOT1* mutation rate in CRC suggests a correlation of *CNOT1* genetic alteration with CRC development.
Figure 12. Mutations in CNOT1 identified in CRC samples by TCGA and Seshagiri et al.

The longest CNOT1 isoform is composed of 2376 amino acids which can be divided into N-terminal, middle and C-terminal regions. The CNOT1 N terminus consists of two HEAT-repeat domains. The CNOT1 middle region contains a MIF4G domain and a domain of unknown function (DUF3819). The CNOT1 C terminus contains the NOT1 superfamily homology (SH) domain (Figure 13).

Figure 13. Diagrams of CNOT1 with N-terminal, middle and C-terminal regions. (Boland et al., 2013)

CNOT1 is essential for the assembly of the complete CNOT complex (Boland et al., 2013). The N terminus of CNOT1 provides binding sites for the CNOT10–CNOT11 module with functions that are largely unknown (Boland et al., 2013). The catalytic module comprising deadenylases centers on the middle region of CNOT1, which is critical for deadenylation of specific mRNAs. For example, the MIF4G domain (residues 1093-1317) in middle region of CNOT1 binds to CNOT7 directly through a pre-formed interface and leaves the CNOT7 catalytic site fully accessible to RNA substrates (Petit et al., 2012). The the middle region of CNOT1 is also important for interaction with RNA binding proteins. For example, residues 800-1015 of CNOT1 are critical for binding with the ZFP36 ring finger protein (ZFP36). Disruption of the ZFP36-CNOT1 interaction impairs mRNA deadenylation in vitro and in vivo (Fabian et al., 2013). The C terminus of CNOT1 is the scaffold of the NOT module which minimally consists of CNOT2 and CNOT3. The NOT module has been shown to regulate the stability and activity of the catalytic module and play a crucial role in the recruitment of the CNOT complex to mRNAs by microRNAs (Wahle et al., 2013). GW182, a RISC complex component, directly binds to the CNOT1 C terminal domain to recruit the CNOT deadenylase complex (Fabian et al., 2011).
Due to the central role of CNOT1 in the integrity and function of the CNOT complex, disruption of CNOT1 has been shown to attenuate deadenylation mediated by the CNOT complex (Ito et al., 2011, Fabian et al., 2013, Boland et al., 2013). Biological functions of CNOT1 have not been well studied, in particular in cancer. Some studies showed that depletion of CNOT1 could cause significant apoptotic cell death in a caspase-4 dependent manner in Hela cells (Ito et al., 2011). Depletion of another CNOT subunit, CNOT2, also has been shown to cause cell apoptosis (Sonnichsen et al., 2005, Ito et al., 2011). Deadenylase CNOT6L is required for cell proliferation via regulation of mRNA levels of the cell cycle inhibitor $p27^{kip}$ (Morita et al., 2007). CNOT7 and CNOT8 are important for cell proliferation in MCF7 breast cancer cells (Aslam et al., 2009). These results suggest that the integrity and the deadenylation function of the CNOT complex are critical for cell survival. Furthermore, CNOT1 mediates repression of estrogen receptor $\alpha$ (ER$\alpha$)- and retinoid X receptor (RXR)-mediated transcription via direct recruitment of the CNOT complex to hormone-regulated promoters. In support of this notion, knockdown of CNOT1 by siRNA results in induction of ER$\alpha$ target genes, TFF1 and c-MYC, in MCF-7 cells (Winkler et al., 2006).

Moreover, the disrupted deadenylation function of the CNOT complex can cause up-regulated mRNA levels of a large group of genes (Aslam et al., 2009). When these mRNAs are translated in the endoplasmic reticulum (ER), it may cause overproduction of proteins, inducing a cellular stress condition called ER stress. This is supported by an Ito et al study showing that depletion of CNOT1 and CNOT2 in Hela cells can cause activation of unfolded protein response (UPR) pathways, which is a cellular response to ER stress (Ito et al., 2011).

**Mechanisms of mRNA Decay**

One of the main functions of the CCR4-NOT complex and CNOT1 is regulation of mRNA deadenylation and turnover. Control of mRNA decay is a fundamentally important step in determining the amount of protein produced from the mRNA by translation. Microarray analyses have revealed that 40-50% of changes in gene expression in response to cellular signals occurs at the level of mRNA stability, allowing
rapid regulation of gene expression in response to different cellular conditions (Fan et al., 2002, Cheadle et al., 2005). Polyadenylation and deadenylation of an mRNA transcript are key processes that regulate its stability and degradation.

During the process of maturation, almost all eukaryotic precursor mRNAs undergo cleavage/polyadenylation at their 3' ends, with the exception of histone mRNAs, which are cleaved but not polyadenylated (Gilmartin et al., 2005). During transcription, the mRNA transcript is cleaved at a site 10-30 nucleotides downstream of the poly(A) signal (AAUAAA), and then a poly(A) tail of 200-300 nucleotides long is added to the 3' end by polyadenylate polymerase. The poly(A) tail and the poly(A)-binding proteins (PABPs) bound to it aid in protecting mRNAs from degradation by exonucleases (Mandel et al., 2008).

In eukaryotes the 5’ cap and the 3’ poly(A) tail interact with the cytoplasmic proteins eIF4E and PABPs, respectively, to protect the transcript from exonucleases and to enhance translation initiation. To initiate decay, either one of these two structures must be compromised. Decay of most mRNAs is initiated by shortening of the poly(A) tail at the 3’ end, referred to as deadenylation (Shyu et al., 1991). The deadenylation process is considered as the rate-limiting step and is also the most efficient step in controlling mRNA decay (Decker and Parker, 1993). Many cellular factors and mechanisms are devoted to modulating the rate of deadenylation, including deadenylases which carry out the deadenylation of target mRNAs. There are three main categories of eukaryotic deadenylases, including the PAN2-PAN3 complex, the CNOT complex and the PARN (poly(A)-specific ribonuclease), each with unique properties (Yamashita et al. 2005). The CNOT complex has been described above. PAN is a heterodimeric complex consisting of catalytic PAN2 and regulatory PAN3 subunits (Uchida et al., 2004). It was reported that mammalian PAN2-PAN3 carries out the initial shortening of the tail of a β-globin reporter transcript from the usual 200 nucleotides to a length of ~80 nucleotides (Yamashita et al. 2005). At this point, the deadenylation is handed over to the CNOT complex. Whether PAN2-PAN3 has a role in deadenylating all mRNAs in higher eukaryotes is still unclear as it seems likely that this function can be carried out by other
deadenylases (Yamashita et al. 2005). PARN has cap-dependent deadenylase activity, i.e., its processivity is enhanced by the presence of a 5’ cap and inhibited by cap-binding proteins (Balatsos et al., 2006). PARN is essential for embryogenesis in many higher eukaryotes including mammals. For example, PARN was reported to be implicated in the mass deadenylation of maternal mRNAs in *Xenopus laevis* oocytes during maturation (Korner et al., 1998). Following deadenylation, two mechanisms can degrade an mRNA: the 5’-to-3’ and 3’-to-5’ exonucleolytic degradation pathways. In the 5’-to-3’ pathway, the 5’ cap is removed by the DCP1–DCP2 complex which is involved in early steps of poly(A) tail metabolism. Following decapping, the 5’-3’ exoribonuclease 1 (XRN1) degrades the mRNA transcript (Fenger-Gron et al., 2005). Alternatively, the deadenylated mRNA can be degraded in the 3’-to-5’ direction by the exosome, a multi-subunit complex that consists of 6 core subunits proteins with RNase activity. Following 3’-to-5’ decay, the 5’ cap on the remaining oligomer is hydrolyzed by the scavenger decapping enzyme DcpS (Garneau et al., 2007). These two pathways are not mutually exclusive and both are involved in the decay of unstable mRNAs in mammalian cells (Figure 14).

![Figure 14. Mechanisms of normal mRNA degradation.](Garneau et al., 2007)
mRNA deadenylation/decay represents an important checkpoint for regulation of post-transcriptional gene expression. The mRNA deadenylation determinants are found predominantly in the 3’UTR, such as AU-rich elements (ARE) and microRNA target sites (Sandberg et al., 2008), but they are also found in the 5’ UTR and coding regions.

About 12% of mammalian mRNAs, especially those encoding cytokines, proto-oncogenes and transcription factors, bear important regulatory ARE signals in their 3’UTRs (Khabar et al., 2005). The AREs typically contain one or several AUUUA pentamer repeats within a U-rich region of the 3’UTR. Both the AREs sequences and the flanking sequences can influence the overall effect on mRNA stability (Stoecklin et al., 2003). RNA binding proteins (RBPs) can bind to the ARE element to regulate ARE-mediated mRNA decay by either inhibiting or activating deadenylation (Barreau et al., 2006). RBPs are key components in RNA metabolism, regulating the temporal, spatial and functional dynamics of RNAs (Barreau et al., 2006). Most mRNAs are associated with multiple RBPs, while a RBP can bind to multiple mRNA targets, suggesting the potential for a multidimensional network of post-transcriptional regulation of RNA metabolism. RBPs play diverse roles in mRNA decay. Some RBPs accelerate decay of target mRNA transcripts through promoting deadenylation, including ZFP36, butyrate response factor 1 (BRF1), AU-rich binding factor 1 (AUF1) and KH-type splicing regulatory protein (KSRP). To promote deadenylation of target mRNA transcripts RBPs can recruit the CNOT complex to target transcripts to carry out the removal of the poly(A) tail. In this process, CNOT1 provides a platform that recruits other subunits of the CNOT complex, including deadenylase subunits, to target transcripts (Sandler et al., 2011). For example, Nanos2 plays a key role in the sexual differentiation of germ cells through regulating the expression levels of several important genes, such as Sycep3, Stra8 and Dazl. It can bind to the CNOT1 protein directly, in turn recruiting the deadenylase subunits to the Nanos2-RNA complex to execute deadenylation of these mRNAs (Suzuki et al., 2010). PUM1 can promote HMOX1 mRNA deadenylation through binding to its ARE and recruiting the CNOT complex (Miller et al., 2011). ZFP36 is an RNA binding zinc-finger protein which binds the ARE of TNF-α mRNA and accelerates its degradation through recruiting the CNOT complex (Sandler et al., 2011). The disruption of the
interaction between RBPs and the CNOT complex may interrupt RBP-mediated mRNA degradation. For example, the depletion of CNOT1 by siRNA could result in a dramatically decreased decay of globin-ARE reporter mRNA by ZFP36 (Sandler et al., 2011). The defects in RNA metabolism caused by aberrations in RBP-mediated mRNA decay contribute to a broad spectrum of human disorders including cancer (Dean et al., 2004). Besides the CNOT complex the deadenylase PARN is also involved in ZFP36-directed deadenylation in vitro (Lin et al., 2007). On the other hand, ELAV like RNA binding protein 1 (ELAVL1) can stabilize mRNAs containing AREs in their 3’UTR, such as c-fos, GM-CSF and EGF, by blocking the recruitment of deadenylases or the exosome (Fan et al., 1998).

MicroRNAs (miRNAs) are a large family of single-stranded small non-coding RNA molecules (~22 nucleotides in length) found in plants and animals, which function in transcriptional and post-transcriptional regulation of gene expression. The human genome may encode over 1000 miRNAs, which may target about 60% of mammalian genes (Friedman et al., 2009). miRNAs are encoded by the eukaryotic genome from either their own genes or from introns or even exons of other genes. miRNA genes can be transcribed as independent units, or, in some cases, together with its host gene, providing a means for coupled regulation of both the miRNA and protein-coding gene (Lee et al., 2004). In the canonical maturation pathways, miRNA genes are transcribed by either RNA polymerase II or RNA polymerase III into primary miRNA transcripts (pri-miRNA) which are usually capped and polyadenylated and spliced (Lee et al., 2004). The pri-miRNA is next endonucleolytically cleaved by the nuclear microprocessor complex Drosha–DGCR8 in the nucleus to produce a hairpin structure of about 70 nucleotides, termed a precursor-miRNA (pre-miRNA). The pre-miRNA is then exported from the nucleus by the Exportin-5–Ran-GTP complex to the cytoplasm. In the cytoplasm, the pre-miRNA hairpin is cleaved by the RNase III enzyme Dicer in a complex with the double-stranded RNA-binding protein TRBP, yielding an imperfect miRNA:miRNA* duplex about 22 nucleotides in length (Lund et al., 2006). Although either strand of the duplex may potentially act as a functional miRNA, only the functional strand of the mature
miRNA is loaded together with Argonaute ( Ago ) proteins into the RNA-induced silencing complex ( RISC ), whereas the passenger strand is degraded ( Figure 15 ).

\[ \text{Figure 15. The canonical pathway of microRNA processing. (Winter et al., 2009)} \]

miRNAs result in gene silencing either through translational repression and/or degradation of target mRNAs (Eulalio et al., 2008). These two mechanisms may represent two distinct independent pathways: the ability of miRNAs to expedite deadenylation does not depend on decreased translation; in some cases translational repression by miRNAs require a poly(A) tail (Wu et al., 2006). miRNAs accomplish their regulatory function through the RISC complex. miRISCs facilitate degradation of miRNA targeted mRNAs via deadenylation followed by Dcp1-Dcp2 complex-directed decapping, and subsequent exonucleolytic digestion (Chen et al., 2009). The core RISC contains at least a specificity-determining miRNA, an Ago protein bound to miRNA, and a GW182 protein. Target recognition by a RISC is determined mainly by the base-pairing between the 5′ portion of the miRNA and the 3′UTR of target mRNAs (Bartel et al., 2009). Ago proteins bind the mature miRNA and facilitates its interaction with a target mRNA. Some Ago proteins, such as Ago2, can cleave target transcripts directly by their PIWI domain when the match-up of miRNA with target mRNA is
perfect, such as in the case of miR-196 downregulating the HOXB8 mRNA (Soraya et al., 2004). However, the vast majority of miRNAs are only partially complementarity with the 3’UTR of their targets, so in most cases, Ago has to recruit additional proteins to achieve translational repression and mRNA decay (Bartel et al., 2009). Generally, Ago proteins recruit GW182 proteins to miRNA-targeted mRNAs in which the C-terminal region of Ago directly binds to the GW-rich regions of GW182 in the N terminal (Takimoto et al., 2009). Three GW182 paralogs exist in vertebrates (TNRC6A/GW182, TNRC6B, and TNRC6C) (Eulalio et al., 2009). The binding of the RISC to the target transcript can result in the recruitment of deadenylation factors mediated by GW182 (Fabian et al., 2011). Human GW182 proteins recruit the PAN2-PAN3 and the CNOT complexes through direct interactions with PAN3 and CNOT1 by the middle and C-terminal regions of the silencing domains, respectively. Therefore, GW182 proteins provide a docking platform through which deadenylase complexes gain access to the poly (A) tail of miRNA targets to promote their deadenylation. In this process, CNOT1 mediates the interactions between GW182 and the additional subunits of the CNOT complex that provides a major contribution to miRNA-mediated mRNA degradation (Piao et al., 2010). The CNOT complex recruited by GW182 can also act to repress translation initiation (Pasquinelli et al., 2012). Thus we hypothesized that the genetic alterations in CNOT1 may impair the miRNA-mediated post-transcriptional regulation of genes, contributing to the development of CRC. Supporting this idea, knockdown of CNOT8 and CNOT1 expression and the overexpression of CCR4 or CAF1 mutants could significantly reduce miRNA-mediated deadenylation and mRNA decay, but not translational repression (Piao et al., 2010, Fabian et al., 2009). The PAN2-PAN3 complex has also been shown to be involved in but not essential for degradation of miRNA targets, as inactivation of the PAN2-PAN3 complex slows down the initial deadenylation, but deadenylation and decay still occur, minimally affecting miRNA target levels at steady state (Chen et al., 2009). How and if PARN is involved in miRNA-mediated deadenylation still remains to be established.

**ER Stress and the Unfolded Protein Response**
The CNOT complex and CNOT1 are widely involved in RBP- and miRNA-mediated mRNA deadenylation/decay. Therefore, the disruption of the function of the CNOT complex and CNOT1 may lead to an altered gene expression profile. In addition, an enlarged pool of mRNAs may be translated in the endoplasmic reticulum (ER). In the ER the nascent polypeptides are folded into specific tertiary structures essential to their functions. When the requirement for proper folding of nascent polypeptides exceeds the protein folding capacity of the ER, unfolded or misfolded proteins will accumulate in the ER lumen, inducing a situation termed ER stress. When ER stress occurs, cells will activate an elaborate adaptive response known as the unfolded protein response (UPR) (Ron, et al., 2007). In eukaryotic cells, canonical branches of the UPR are mediated by three ER membrane-associated proteins, PKR-like eukaryotic initiation factor 2α kinase (PERK), inositol requiring enzyme 1α (IRE1α), and activating transcription factor-6 (ATF6) (Tsai et al., 2010) (Figure16). In a “stress-free” ER, these three trans-membrane proteins are bound by a chaperone, glucose-regulated protein, 78kDa (GRP78/BiP), in their intraluminal domains and rendered inactive. Accumulation of improperly folded proteins in the ER results in the dissociation of GRP78/BiP away from these UPR sensors (Bertolotti et al., 2000). This event leads to oligomerization and activation of PERK and IRE1α and translocation of ATF6 to the Golgi apparatus where it is processed into an active transcription factor, resulting in the activation of a series of downstream signaling events (Ron et al, 2007). Together these three arms of the UPR can mitigate ER stress by reducing protein synthesis, facilitating protein degradation, and increasing production of protein folding chaperones, which helps cells to restore homeostasis.

The IRE1α pathway is the oldest branch of the UPR in an evolutionary sense. The endoribonuclease activity of IRE1α cleaves a 26 base-pair segment from the mRNA of the X-box binding protein-1 (XBP1), creating an alternative message that is translated into the spliced active form of the transcription factor sXBP1 (Calfon et al., 2002). sXBP1, alone or in conjunction with ATF6, launches a transcriptional program to produce chaperones and proteins involved in ER biogenesis, phospholipid synthesis, and ER-associated protein degradation (ERAD). Thus, sXBP1 activates one of the major pathways for enhancing the folding capacity of the ER to deal with ER stress.
Hyperactivated IRE1α has been reported to increase TXNIP mRNA stability by reducing levels of miR-17, a TXNIP destabilizing microRNA. In turn, elevated TXNIP protein activates the NLRP3 inflammasome, causing procaspase-1 cleavage and interleukin 1β (IL-1β) secretion (Lerner et al., 2012).

Under ER stress, ATF6 translocates to the Golgi apparatus, where it is processed by Golgi proteases (S1P, S2P) to release the active transcription factor. Activated ATF6 then moves to the nucleus and functions as a transcription factor to promote the transcription of genes containing ER stress elements (ERSE), UPR elements (UPRE), and cAMP response elements (CRE) in their promoters. As a result, ERAD and production of the ER degradation-enhancing mannosidase-like protein (EDEM) are boosted by these events, facilitating clearance and degradation of misfolded proteins from the ER lumen. Furthermore, ATF6 can upregulate XBP1 mRNA transcription levels. Currently, ATF6 and sXBP1 are viewed as the predominant regulators of the transcriptional response programs triggered by the activation of the UPR (Tsai, et al., 2010).

Activation of the third arm of the UPR through PERK results in phosphorylation of eIF2α (eukaryotic translational initiation factor 2α) at serine 51, which converts eIF2α to a competitor of eIF2b, resulting in reduced global protein synthesis. In addition, this branch of the UPR can also lead to activation of activating transcription factor-4 (ATF4), nuclear erythroid 2 p45-related factor 2 (NRF2), and nuclear factor kappa b (NF-κB) through several distinct mechanisms. ATF4 is produced through enhanced translation and induces expression of genes involved in apoptosis (CHOP, C/EBP homologous protein), ER redox control (ERO1, endoplasmic reticulum oxidoreductin), and glucose metabolism (GCK, glucokinase). One consequence of ER stress is the accumulation of reactive oxygen species (ROS) that promotes a state of oxidative stress. The UPR has incorporated an antioxidant defense system via activation of the NRF2 and ATF4 transcription factors, coordinating the convergence of ER stress with oxidative stress signaling, and orchestrating the transcription program of the antioxidant response element-dependent genes, including hemeoxygenase-1 (HMOX1), thioredoxin reductase 1 (TXNRD1), and the glutathione S-transferases (GSTP1).
Figure 16. Unfolded protein response in mammalian cells. (Schöenthal et al., 2011)

Through activation of these three UPR pathways, cells can deal with a moderate level of ER stress and restore homeostasis. However, the UPR does not always result in successful alleviation of ER stress and reestablishment of a functional equilibrium in the ER. When driven by severe or prolonged stress signals, the UPR can also induce cell apoptosis through several mechanisms (Tabas et al., 2011). The IRE1α mediated UPR pathway is a critical lever in the UPR that controls commitment to cell death or promotes survival (Ron et al., 2007). IRE1α leads to activation of the apoptosis signaling kinase 1 (ASK1) and downstream kinases through binding TRAF2, leading to activation of the c-Jun N-terminal kinase (JNK) pro-apoptotic signaling pathway that promotes apoptosis by switching on the pro-apoptotic B-cell CLL/lymphoma 2 (BCL-2) family proteins BCL2-associated X protein (BAX) and BCL2-antagonist/killer 1 (BAK) (Zong et al., 2003). IRE1α also has been reported to upregulate thioredoxin-interacting protein (TXNIP) levels by reducing levels of miR-17. In turn, elevated TXNIP protein activates a series of pro-apoptotic pathways, such as activation of interleukin-1 beta (IL-1β) (Lerner et al., 2012). In addition, IRE1α can also induce the preferential degradation of a pool of mRNAs through a mechanism known as regulated IRE1-dependent decay (RIDD) (Hollien et al., 2009). Hyper activation of the PERK pathway can induce expression of CHOP, a pro-apoptotic transcription factor, which regulates the transcription of BCL2 family members and other pro-apoptotic genes (Zinszner et al., 1998). In addition, activation of both the PERK and IRE1α pathways lead to activation of the NF-κB signaling pathway (Tam et al., 2012).
There is evidence that chronic ER stress and activation of the UPR are important in the pathogenesis of a number of diseases, including cancer, neurodegenerative diseases, diabetes, and inflammatory disease (Hosoi et al., 2010, Lerner et al., 2012). Some cancer cells, unlike normal cells, harbor a chronically activated protective module of ER stress components in order to manage intensified protein synthesis or to adapt to hostile microenvironmental conditions, such as hypoglycemia, hypoxia, acidosis, or chemotherapy, which are known triggers of ER stress (Luo et al., 2013). The genetic alterations resulting in better management of ER stress in tumor cells may confer growth advantage to cancer cells and contribute to cancer progression. Therefore, we hypothesize that the increased activity of CNOT1 may help cancer cells to limit the production of proteins in the ER, thus reducing the levels of ER stress.

In accordance with the previously described balance of this cellular system, two opposing approaches are offered in order to therapeutically target ER stress. On the one hand, experimental efforts are aimed at supporting pro-survival modules and/or blocking the pro-apoptotic components in order to increase cellular survival. For example, in type 2 diabetes where chronic ER stress may lead to the destruction of pancreatic β-cells, controlling ER stress may help to protect β-cells and delay the development of diabetes (Lerner et al., 2012). On the other hand, efforts to further aggravate pre-existing ER stress and enhance pro-apoptotic processes could be beneficial in the case of malignant neoplasms. For example, deterioration of chronic ER stress in cancer cells by therapeutic reagents may induce ER stress mediated cell apoptosis selectively in tumor cells which have higher levels of ER stress than normal cells (Schönthal et al., 2011).

**Thioredoxin-Interacting Protein (TXNIP)**

When unfolded proteins accumulate to irremediably high levels within the ER, the UPR becomes hyperactivated to cause programmed cell death. TXNIP that is rapidly induced by IRE1α is a critical node in this “terminal UPR.” TXNIP was originally identified in HL-60 leukemia cells treated with 1,25-dihydroxyvitamin D3 and was termed as *vitamin D up-regulated protein 1 (VDUP1)* (Chen et al. 1995). This gene encodes for an inhibitor of the antioxidant protein Thioredoxin (TRX), which directly
interacts with TRX and blocks its reducing activity and inhibits the interaction between TRX and other factors, such as ASK-1 and PAG (Junn, et al., 2000). TXNIP is also reported to exert functions through redox independent mechanisms. TXNIP has been reported to be involved in multiple cellular processes, including cell growth and apoptosis (Jeon, et al., 2005), metabolic regulation (Patwari, et al., 2009), and inflammatory response (Zhou, et al., 2010). TXNIP expression is regulated by multiple mechanisms, such as glucose, HDACs, KLFs, and PPARs. Its expression is also upregulated by various cellular stresses, including ER stress, oxidative stress, and irradiation (Han et al., 2003). Further, NRF2, an important transcription factor known to regulate anti-oxidant responses, also upregulates TXNIP expression levels (Adair-Kirk et al., 2008). As described before, TXNIP levels are also upregulated by activation of IRE1α that destabilizes miR-17.

As a tumor suppressor, TXNIP has been reported to cause cell cycle arrest and apoptosis. To induce cell growth arrest, TXNIP translocates to the nucleus where it regulates cell proliferation by multiple mechanisms, including increasing the stability of the cell cycle inhibitor p27kip, upregulating levels of the CDK inhibitor p16, and reducing RB phosphorylation (Jeon et al., 2005). TXNIP can also promote apoptosis through redox-dependent and -independent mechanisms. Thioredoxin (TRX) inhibits apoptosis via inhibiting ASK1 and PTEN (Saitoh et al., 1998). TXNIP inhibits the interaction between TRX and ASK1, relieving the constitutive inhibition of ASK1 and the released ASK1 induces apoptosis through activation of the JNK and p38 signaling pathways (Tobiume et al., 2001). A recent report described how TXNIP promotes NLR family, pyrin domain containing 3 (NLRP3) inflammasome activation and IL-1β maturation that can lead to apoptosis in response to different cellular stresses (Zhou, et al., 2010).

TXNIP has been identified as a novel tumor suppressor based on its effects on regulating both cell growth and death (Han et al., 2003, Goldberg et al., 2003). TXNIP has dramatically reduced expression in multiple tumors including renal, breast, lung, gastric, colon, and hepatocellular carcinoma (Cadenas et al., 2010, Dutta et al., 2005, Takahashi et al., 2002). The enforced expression of TXNIP inhibited the proliferation of
cancer cells including gastric cancer, and promyelocytic leukemia, and suppressed tumor growth and metastasis in transplantation models (Han et al., 2003, Wang et al., 2002, Goldberg et al., 2003). TXNIP has also been shown to play an important role in several pathologic conditions besides cancer, including diabetes, cardiovascular disease, and chronic inflammation.

It has been shown that miR-17 destabilizes TXNIP mRNA (Lerner et al., 2012). miRNAs recruit the CNOT complex through direct binding to CNOT1 to deadenylate target mRNAs. Therefore, we hypothesize that increased activity of CNOT1 as well as the CNOT complex may contribute to the regulation of TXNIP levels by miR-17.

Summary

Identification of driver mutations in CRC is important in overcoming the challenges in CRC management and improving patients’ outcome. In the hundreds of DNA mutations in each CRC genome, only a small portion are driver mutations, whereas most of them are passenger mutations. Therefore, a major challenge in CRC genetic studies is to distinguish driver mutations hidden within a much larger number of passenger mutations. To identify candidate CRC driver mutations, we performed a Sleeping Beauty transposon-based forward genetic screen in mice and we found a list of CIS candidate CRC genes. One of the top CIS candidates is Cnot1 which encodes for the scaffold factor of the CNOT complex that is a multicomponent complex widely involved in post-transcriptional gene expression regulation. One of the main functions of the CNOT complex is regulation of mRNA deadenylation which is the rate-limiting step and the most efficient step in controlling mRNA decay. RBPs and miRNAs can recruit the CNOT complex to target transcripts to promote deadenylation of target mRNA through direct binding to CNOT1. Aberrations in the deadenylation function of the CNOT complex have been shown to cause altered gene expression profiles and reduced cell viability. Furthermore, multiple somatic mutations have been documented in CRC by the TCGA and COSMIC databases. Based on the SB-based screens, somatic mutations and function of the CNOT complex, we hypothesize that genetic alterations in CNOT1 may contribute to CRC development through affecting deadenylation of mRNA transcripts.
Notably, the functions of CNOT1 and the CNOT complex in cancer development have not been adequately studied. Therefore the objective of this study is to characterize CNOT1 in the development of CRC and investigate the underlying mechanisms.
Chapter 2

Results

Loss of CNOT1 Results in Decreased Cell Viability

Cnot1 was identified as a candidate colorectal cancer (CRC) gene in the Apc\textsuperscript{Min} (Starr \textit{et al}., 2011) and LSL-p53\textsuperscript{R270H} (Starr T.K., unpublished data) \textit{Sleeping Beauty} (SB)-based screens. To determine the contribution of genetic alterations in CNOT1 to CRC we evaluated cell viability in CNOT1-depleted SW480 human CRC cell line and a non-intestinal cell line HEK293. The SW480 cell line is APC deficient as it lacks one allele and has a truncating mutation at position 1338. This APC status is similar to the status of Apc in most adenomas formed in the Apc\textsuperscript{Min} mouse strain used in the original SB screens that identified Cnot1 as a CIS associated candidate CRC gene. Apc\textsuperscript{Min} mice carry an inactivating germline mutation in one allele and the second allele is inactivated in most adenomas. For these studies SW480 and HEK293 cells were transfected twice, 48 hours apart, with either a pool of four siRNA oligos targeting distinct regions in the CNOT1 mRNAs, or a non-targeting control siRNA, or transfection reagent only (mock transfected), or not transfected. The effectiveness of depletion was confirmed by qRT-PCR (Figure 17A). CNOT1 expression was reduced to 35% and 33% of the expression levels in control SW480 cells and HEK293 cells, respectively. Cell viability was evaluated daily for 6 days after the second transfection using the MTT assay. Over the 6-day interval, depletion of CNOT1 by siRNA resulted in a dramatic decrease in cell viability in SW480 cells (Figure 17B) as well as HEK293 cells (Figure 17C).

To examine the means by which CNOT1 depletion causes decreased cell viability we examined the levels of apoptosis in CNOT1-depleted SW480 cells using an Annexin V fluorescein isothiocyanate (AV-FITC) and propidium iodide (PI) double staining flow cytometry assay, where AV-FITC staining identifies cells in the early stages of apoptosis which have undergone translocation of phosphatidylserine to the external surface of the plasma membrane lipid but still have intact plasma membranes and thus are impervious
to PI staining. AV-FITC-PI staining was determined on day 6 (3 days after the first transfection) and day 7 (1 day after the second transfection). No changes in apoptosis were seen on day 6 but on day 7 CNOT1 depletion resulted in a greater than two fold increase in apoptosis, from 7 to 19% of total cell population, with a corresponding decrease in unstained live cells from 78 to 64% (Figure 18A, B and C). These results show that depletion of CNOT1 causes decreased cell viability in SW480 and HEK293 cells and increased apoptotic cell death in SW480 cells.

A.

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Relative CNOT1 mRNA levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>SW480</td>
<td>35.89± 3.09%</td>
</tr>
<tr>
<td>HEK293</td>
<td>33.81± 0.99%</td>
</tr>
</tbody>
</table>

B.

C.

Figure 17. CNOT1 depletion leads to decreased cell viability in transformed cell lines SW480 and HEK293. (A) Cells at 70% confluence were transfected with siRNA oligos targeting CNOT1, non-targeting siRNA, transfection reagent only (mock transfected), or not transfected, twice, 48 hours apart, at a final concentration of 50 nM. CNOT1 levels were determined by qRT-PCR. Relative CNOT1 mRNA levels represent the expression of CNOT1 in CNOT1 siRNA-treated cells relative to control siRNA treated. Values represent the average of three (SW480) or two (HEK293) experiments ± SD. (B and C) Cells were plated at 5000 (SW480) or 2500 (HEK293) cells per well in triplicate in 96 well plates 24 hours after the second transfection. Cell
viability was measured by MTT assay on days 1-6 after the second transfection. (B) SW480 cells. (C) HEK293 cells. Each data point represents mean ± SD of three replicates from a representative experiment. Significance of differences in cell viability on day 6 was determined by one-way ANOVA. * = p < 0.05 between CNOT1 and all controls.
C.

Figure 18. CNOT1 depletion leads to increased apoptosis in SW480 cells. SW480 cells were transfected with siRNA oligos targeting CNOT1 or non-targeting siRNA as in Figure 17. Cell apoptosis was determined by an annexin V fluorescein isothiocyanate (AV-FITC) and propidium iodide (PI) double staining flow cytometry assay at 72 hours (day 6) after the first transfection and at 24 hours (day 7) after the second transfection. (A and B) Representative dot plots of AV-FITC-PI staining in SW480 cells on day 6 (A) and day 7 (B). The number in each quadrant indicates the percentage of cells stained with specific combination of dyes. Lower right quadrant represents cells stained by AV-FITC alone that are early apoptotic cells. (C) Quantitative analysis showing AV-FITC stained apoptotic cells as percent of total cells. Values represent mean ± SD of three independent experiments. Significance of differences was determined by one-way ANOVA. ***=p < 0.001, *=p < 0.05.

Loss of CNOT1 Results in Extensive Changes in Gene Expression

CNOT1 is a member of the CCR4-NOT complex, which regulates mRNA expression at several levels. Therefore, to begin to determine the means by which CNOT1 depletion might promote decreased cell viability we examined global gene expression in CNOT1-depleted SW480 cells using an Illumina Human HT-12 v4 BeadChip array. High quality total RNA was isolated from CNOT1-depleted and control SW480 cells. RNA Integrity Number (RIN) values were ≥ 9.8 (Figure 19A). qRT-PCR confirmed 90% knockdown of CNOT1 mRNA levels (Figure 19B). Analysis of gene expression identified 266 genes with statistically significant, greater than or equal to two-fold,
change in gene expression. Of these 183 (69%) were up-regulated and 83 (31%) were down-regulated (Figure 19C, Table 4). Expression levels of five top changing genes, TXNIP, DKK1, EMP1, BTG1 and IGFBP3, were confirmed by qRT-PCR to validate the results of the microarray (Figure 19D). Ingenuity Pathway Analysis (IPA) was used to identify pathways and processes associated with changes in gene expression in CNOT1-depleted cells. IPA analysis generated three networks with scores of 10 or more. These networks represent molecules interacting with protein products of genes that change expression in CNOT1-depleted cells. The functions associated with these networks are associated with tumor morphology, cellular growth and proliferation, cellular death and survival and cell development consistent with an effect of CNOT1 depletion on cell viability (Figure 20A, B, C and D, Figure 33).
Figure 19. Depletion of CNOT1 in SW480 cells causes an altered gene expression profile. SW480 cells were transfected with either CNOT1 siRNA or a non-targeting siRNA control at 50 nM final concentrations, twice, 48 hours apart. Total RNA was extracted 72 hours after the second transfection from three pairs of biological replicates of CNOT1-depleted and non-targeting control treated SW480 cells and hybridized to an Illumina Human HT-12 v4 BeadChip array to determine the gene expression profile. To improve reproducibility each of the 6 independent RNA samples was hybridized twice. (A) The integrity of RNA samples was determined using an Agilent Bioanalyzer Lab-on-a-chip system to generate an RNA Integrity Number (RIN) representing the ratio of 28S/18S RNA as well as other features. All samples met the standard of quality with a RIN > 8. OT3 = non-targeting control siRNA. (B) Validation of knockdown of CNOT1 in RNA samples by qRT-PCR. (C) Heatmap presentation of the 266 genes showing significant differences in expression (fold change ≥ 2, p < 0.05) between CNOT1-depleted and control cells. Relative gene expression is shown as Z-score. Genes are shown on the Y-axis and siRNA treatments on the X-axis. The quality control and normalization were performed using the R-based package: lumi and limma. (D) The results of the array analysis were validated via measuring the mRNA levels of five of the genes with highest fold changes in the microarray by qRT-PCR. For all qRT-PCR assays, values represent mean ± SD of three independent experiments. Statistical significance was determined by a t-test. ** = p < 0.01, * = p < 0.05.
Figure 20. IPA analysis of gene expression in CNOT1-depleted SW480 cells. Networks based on genes that change expression in CNOT1-depleted cells (Network Eligible molecules) were generated by Ingenuity Pathway Analysis (IPA). The three top scoring networks are shown (A, B and C). The score is calculated using the Fishers Exact Test to determine the likelihood of the network containing a given number of Network Eligible molecules by chance. The score = -log (Fisher's Exact test result). (D) The functions and scores of the three networks are shown.

CRC Liver Metastases With High CNOT1 Expression Levels Show Upregulation Of Proliferation Related Genes

To determine if these changes are relevant to human cancers we analyzed gene expression in 20 human CRC liver metastasis samples and 5 unrelated normal samples. High quality RNA was isolated from flash frozen tissue using RNAIceLater (Ambion) and RNeasy (Qiagen) kits. This RNA was sequenced by the University of Minnesota BioMedical Genomics Center using the Illumina HiSeq 2000 platform. Sequence was assembled and analyzed by Dr. Ying Zhang at the University of Minnesota Supercomputing Institute. Fragments per kilobase of exon per million fragments (FPKM) gene expression values were obtained. Levels of CNOT1 expression vary by 3-fold among the metastasis samples. Levels in the normal samples fall around the average of the normal samples with all normal samples falling within two standard deviations. To identify genes whose expression correlated with highest levels and lowest levels of CNOT1 expression we used Geneset Enrichment Analysis (GSEA) (Subramanian et al., 2005). The three samples with highest CNOT1 expression were designated CNOT1-Hi and the three samples with lowest CNOT1 expression were designated CNOT1-Lo.
complete ranked expression dataset was generated in which genes were ranked according to ratio of expression in CNOT1-Hi samples to CNOT1-Lo samples (CNOT1-Hi/CNOT1-Lo). This ranked dataset was compared to genesets derived from Merlos-Suárez (Merlos-Suárez et al., 2011). Merlos-Suárez et al sorted mouse intestinal crypt cells based on expression of EphB2 to define gene programs for normal crypt cell populations. Genes enriched in EphB2<sup>Hi</sup> cells defined the intestinal stem cell (ISC) signature. Genes expressed at equal levels in EphB2<sup>Hi</sup> and EphB2<sup>Med</sup> defined a proliferation signature. Genes differentially expressed in EphB2<sup>Lo</sup> cells defined a late progenitor, transit amplifying population. We found that genes in the proliferation geneset were enriched in the CNOT1-Hi samples while CNOT1-Lo genes were not enriched in any of the three crypt populations (Figure 21A, Table 5A). The CNOT1-Hi/CNOT1-Lo dataset was also compared to genesets of candidate CRC genes identified by SB transposon mutagenesis of mouse intestine (Figure 21B an C, Table 5B and C). It was found that genes up-regulated in CNOT1-Hi tumors, but not in CNOT1-Lo tumors were enriched in candidate cancer genes (CIS genes) identified by Sleeping Beauty mutagenesis. These results are consistent with in vitro results in which loss of CNOT1 resulted in decreased cell viability. However, three metastasis samples had CNOT1 expression levels significantly lower than normal samples. As the anchor protein of a multifunctional complex CNOT1 may potentially regulate many cellular processes. The fact that both CNOT1-Hi and CNOT1-Lo are found in metastasis samples but associated with different expression patterns suggests that CNOT1 affect on CRC may differently dependent on genetic context.
Figure 21. Geneset Enrichment Analysis of human metastasis CRC samples. *CNOT1*-Hi/CNOT1-Lo expression dataset (see text for details) was analyzed for enrichment in crypt proliferation genes and CRC candidate cancer genes identified by *SB* screens. Genesets consist of: (A) genes differentially expressed in proliferating crypt cells (Merlos-Suárez *et al*., 2011); (B) top CIS-associated genes identified in a wild-type background (Starr *et al*., 2009), in an *Apc*<sup>Min</sup> background (Starr *et al*., 2011), and a p53-mutant background (Starr T.K., unpublished data); (C) top 200 CIS-associated genes identified in *SB* screen using a 120 k window (March *et al*., 2011). (D) Normalized enrichment score (NES) and nominal p value (Nom. p value) are shown for each comparison.

**CNOT1 Limits the mRNA Expression Levels of TXNIP in SW480 and HEK293 Cell Lines.**

Because of connections seen between *CNOT1*-Hi expression and cancer-related phenotypes, we chose to investigate the mechanism by which *CNOT1* expression protects cell viability. Microarray analysis revealed dramatically increased expression of *Thioredoxin-interacting protein (TXNIP)* (7.45-fold). *TXNIP* has been identified as a tumor suppressor in human breast, colon, lung and liver cancers through inhibition of cell proliferation and arrest of the cell cycle at *G*<sub>0</sub>/*G*<sub>1</sub> phase and its expression is decreased in several cancers including CRC (Jeon, *et al*., 2005, Han *et al*., 2003, Goldberg *et al*., 2003,
Zhou, et al., 2010, Sheth et al., 2006). In addition, in pancreatic beta cells, activation of TXNIP by oxidative or ER stress initiates apoptotic pathways (Lerner et al., 2012, Zhou et al., 2010, Oslowski et al., 2012). As cancer cells are often subject to extreme stresses we hypothesized that CNOT1 might protect cell viability by limiting TXNIP expression. In accordance with the array results CNOT1 depletion results in increased TXNIP expression by qRT-PCR in CNOT1-depleted SW480 cells and HEK293 cells (8.8 and 2.85 fold, respectively) (Figure 22 A and B). Increased expression of TXNIP in SW480 cells was detected as early as 2 days after the first siRNA transfection (Table 3). In addition, overexpression of the longest CNOT1 transcript variant (NM_00126561, CNOT1L) in HEK293 cells results in a decrease in TXNIP expression to about 70% of controls (Figure 22C). These results indicate that CNOT1 limits TXNIP mRNA levels in these two cell lines.
Figure 22. CNOT1 limits TXNIP mRNA levels in SW480 and HEK293 cell lines. (A) TXNIP mRNA expression levels in RNA samples used for the microarray assay were determined by qRT-PCR. (B) TXNIP mRNA levels in HEK293 cells depleted for CNOT1 were determined by qRT-PCR. HEK293 cells were transfected with either CNOT1 siRNA or a non-targeting siRNA control as described in Figure 17. Total RNA was extracted 48 hours after the second transfection and subjected to qRT-PCR to determine TXNIP levels. (C) HEK293 cells were transfected with plasmids encoding CNOT1L (NM_016284.4) or a control plasmid encoding green fluorescent protein (GFP). 48 hours after transfection RNA samples were isolated and TXNIP mRNA levels were determined by qRT-PCR. Values in (A), (B), (C) represent mean ± SD of 3, 2, 4 independent experiments, respectively. Significance of differences was calculated by a t-test. * = p < 0.05.

Increased TXNIP Levels are Linked to CNOT1-Mediated Apoptosis

To determine if CNOT1-mediated up-regulation of TXNIP might be responsible for decreased cell viability seen in CNOT1-depleted cells, we examined the effect of overexpression of TXNIP on cell viability. For these experiments a plasmid encoding TXNIP was transiently overexpressed in HEK293 cells. Control cells were transfected with a GFP-encoding plasmid in the same vector. Transfection efficiency was estimated to be 60%. TXNIP mRNA expression was increased by 7.5-fold 48 hours after transfection (Figure 23A). In a parallel experiment, cells were evaluated for cell viability by MTT assay. Viability was decreased to about 50% that of control cells at 72 hours (Figure 23B). To test this relationship further, we determined whether elevated TXNIP levels occurred before apoptosis in CNOT1-depleted cells. Compilation of data shown in Figures 18A, 18B and 22A indicates that upregulation of TXNIP mRNA expression is detected as early as 2 days (day 5) after the first transfection. However, detectable increases in apoptosis do not occur until the first day (day 7) after the second transfection (Table 3). This result is consistent with CNOT1 mediating cell viability by regulating TXNIP levels. To summarize, depletion of CNOT1 results in increased apoptosis, decreased cell viability and increased levels of TXNIP in SW480 and HEK293 cells. Increased levels of TXNIP result in decreased cell viability in HEK293 cells. Finally, CNOT1 depletion in SW480 cells results in increased TXNIP levels before increased
apoptosis and decreased cell viability. Taken together, these results suggest that CNOT1 protects cell viability at least in part by limiting TXNIP expression.

Figure 23. Overexpression of TXNIP reduces SW480 cell viability. HEK 293 cells were transiently transfected with plasmids encoding either TXNIP or GFP as a control. (A) Total RNA was isolated 48 hours after transfection and TXNIP mRNA levels were determined by qRT-PCR. (B) Transfected cells were plated at $2 \times 10^4$ cells per well in triplicate in 96 well plates. Cell viability was measured by MTT assay on 6 hours to 72 hours after transfection. Viability is shown as absorbance at 595 nm. Each data point represents mean ± SD of three replicates from one representative experiment of two. Significance of differences in cell viability at 72 hours was determined by a $t$-test. * = $p < 0.05$.

<table>
<thead>
<tr>
<th>Transfection</th>
<th>d3</th>
<th>d4</th>
<th>d5</th>
<th>d6</th>
<th>d7</th>
<th>d9</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNOT1 fold change</td>
<td>0.11</td>
<td>0.13</td>
<td>0.17</td>
<td>0.17</td>
<td>4.37</td>
<td>8.8</td>
</tr>
<tr>
<td>TXNIP fold change</td>
<td>0.81</td>
<td>6.04</td>
<td>2.9</td>
<td>2.9</td>
<td>2.9</td>
<td>8.8</td>
</tr>
<tr>
<td>Apoptosis fold change</td>
<td>No change</td>
<td>Increase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Time course study of a correlation between CNOT1 depletion, upregulated TXNIP and cell apoptosis. Compilation of results of CNOT1 depletion in SW480 cells taken from Figures 17A, CNOT1 depletion; 17B, cell viability; 18C, apoptosis; and 22A, TXNIP mRNA levels. Fold change = level in CNOT1-depleted cells/level in non-targeting control treated cells. Results are shown relative to a timeline of the CNOT1 siRNA depletion protocol.
Identification of CNOT1 Isoforms and CCR4-NOT Subunits Involved in TXNIP Regulation

We then investigated the mechanism by which CNOT1 regulates TXNIP levels. Human cells express three CNOT1 transcript variants with translated protein products (See details in Introduction). The two longest forms (CNOT1L, NM_016284.4 and CNOT1M, NM_001265612.1) contain a C-terminal NOT domain with binding sites for several potentially biologically relevant interactors, such as CNOT2, CNOT3, and CNOT5. To determine which of these variants contribute to cell viability in SW480 cells, primers were designed to specifically distinguish between the two long forms (CNOT1L and M) and the short form (CNOT1S). qRT-PCR analysis of SW480 cells revealed that one or both of the long forms are virtually the only form expressed in these cells (98%), while the short form makes up no more than 2% of overall CNOT1 mRNA expression (Figure 24A). In addition, the short form is almost undetectable in CRC liver metastasis and normal colon samples, suggesting that one or both of the long forms are the biologically relevant form in the colon (Figure 24B). Importantly, expression of the shortest isoform (CNOT1S) resulted in increased TXNIP expression (Figure 24C). This result suggests that some part of the C-terminal region unique to the CNOT1L and/or CNOT1M isoforms contributes to downregulation of TXNIP. As CNOT1 acts as part of a larger complex CNOT1S may act as a dominant negative by sequestering essential subunits.

Because CNOT1 is part of the CCR4-NOT complex and because the full range of configurations of this complex is not known (Chen et al., 2011) we determined if knockdown of other CCR4-NOT subunits would also decrease cell viability and increase TXNIP levels. We focused on CNOT2, which is also a core member of the CCR4-NOT and which has also been identified as a candidate CRC gene in SB screens, and on the four deadenylase subunits because deadenylation leading to mRNA degradation is the best characterized mechanism of CCR4-NOT mediated gene regulation. Analysis of SW480 cells depleted for CNOT2, and the deadenylase subunits CNOT6, 6L, 7 and 8 (Figure 24D) indicated that while cell viability was decreased in cells depleted for each of these subunits (CNOT6 not measured), TXNIP levels were elevated only in CNOT8-depleted cells (7.07 fold) (Figure 24E). These results suggest that CNOT1 works with
CNOT8 to regulate *TXNIP* levels. In addition, they suggest that CNOT1 will be found to regulate cell viability by additional pathways involving interactions with other CCR4-NOT subunits.

![Graph A](image1.png)  ![Graph B](image2.png)  ![Graph C](image3.png)  ![Graph D](image4.png)  ![Graph E](image5.png)

**Figure 24.** Identification of CNOT1 isoforms and CCR4-NOT subunits involved in *TXNIP* regulation. (A) *CNOT1* variants expressed in untreated SW480 cells were determined by qRT-
PCR using primers designed to distinguish between the two long isoforms, \textit{CNOTIL} and \textit{CNOTIM} (NM_016284.4 and NM_001265612.1) and the short isoform \textit{CNOTIS} (NM_206999.2). Values represent mean ± SD of three independent experiments. Significance of differences was calculated by a \textit{t}-test. ** = \textit{p} < 0.01. (B) FPKM (fragments per kilobase of exon per million fragments mapped) expression levels of \textit{CNOTIL} and \textit{CNOTIS} in 20 liver metastasis of stage IV human CRCs and 5 normal colonic tissues. Values represent mean ± SD of each group of samples. Significance of differences was calculated by one-way ANOVA. *** = \textit{p} < 0.001. (C) Expression of \textit{TXNIP} was determined in cells overexpressing \textit{CNOTIS}. HEK 293 cells were transfected with plasmids encoding either CNOT1S or GFP as a control. Total RNA was isolated at 48 hours after transfection and probed for \textit{TXNIP} levels by qRT-PCR. Values represent result of one experiment. (D) SW480 cells were transfected with siRNA oligos targeting \textit{CNOT2}, 6, 6L, 7, 8, or non-targeting siRNA, twice, 48 hours apart, at a final concentration of 50 nM. Total RNA was extracted 72 hours after the second transfection and depletion of target genes was determined by qRT-PCR. Relative mRNA level = level of indicated subunit in subunit-specific siRNA-treated cells/control siRNA treated. On day 1 after the second transfection SW480 cells were plated at 2,500 cells per well in triplicate in 96 well plates. Cell viability was measured by MTT assay on day 6 after the second transfection. Relative cell viability represents = absorbance at 595 nm of cells treated with indicated siRNA/cells treated with non-targeting control. Results represent at least two experiments except for CNOT6 (one experiment, no MTT assay). (E) \textit{TXNIP} expression levels were determined by qRT-PCR in SW480 cells depleted for \textit{CNOT2}, and the deadenylase subunits \textit{CNOT6}, 6L, 7, and 8. Values represent mean ± SD of at least two experiments except for CNOT6 (one experiment). Significance of difference was calculated by a \textit{t}-test.

\textbf{CNOT1 Limits \textit{TXNIP} mRNA Levels Through Decreasing mRNA Stability}

As CCR4-NOT subunits with deadenylase activity, such as CNOT8, can initiate mRNA degradation, we pursued the idea that CNOT1 regulates \textit{TXNIP} mRNA stability. \textit{CNOT1}-depleted and control SW480 cells were treated with Actinomycin D to block transcription and \textit{TXNIP} mRNA levels were determined by qRT-PCR for 2 hours. This study showed that \textit{TXNIP} levels were significantly higher in \textit{CNOT1}-depleted cells after 2-hour exposure to Actinomycin D (Figure 25) indicating that \textit{TXNIP} mRNA stability is enhanced in the absence of CNOT1.
Figure 25. CNOT1 limits TXNIP mRNA levels through decreasing mRNA stability. SW480 cells were transfected with either CNOT1 siRNA or control siRNA. Actinomycin D was added at a final concentration of 5 µg/ml 72 hours after one siRNA transfection. Total mRNA was isolated at indicated intervals from 0 to 120 minutes after treatment. TXNIP mRNA levels were determined by qRT-PCR. The TXNIP mRNA levels at each time point are shown as the percentage of TXNIP mRNA levels at 0 minute. Values represent mean ± SD of three independent experiments. The best-fit line is shown as a dotted line. Significance of difference was calculated by a t-test at 60 minutes and 120 minutes. * = p < 0.05.

3’UTR Binding Proteins ZFP36 (TTP), PUM1, PUM2 and CELF1 (CUGBP1) Are Not Involved in CNOT1-Mediated Regulation of TXNIP.

CNOT1 commonly regulates mRNA stability by acting as an adapter between 3’UTR binding factors and deadenylase subunits of the CCR4-NOT complex. CNOT1 has been reported to interact with the 3’UTR binding proteins ZFP36, PUM1 and PUM2 (Fabian et al., 2013, Van Etten et al., 2012), which mediate deadenylation and degradation of targeted mRNAs. In addition Pum1, Pum2 and Celf1, another 3’UTR binding protein, were identified as candidate colorectal cancer genes in SB screens (Starr et al., 2009, Starr et al., 2011). To determine if these 3’UTR binding proteins were involved in CNOT1 regulation of TXNIP, TXNIP levels were determined in SW480 cells depleted for each protein. Depletion of each of these proteins results in less than two-fold increase in TXNIP (Figure 26A and B) suggesting that these 3’UTR binding proteins are
not involved in CNOT1-mediated regulation of TXNIP. 3’UTR binding proteins often bind AU rich motifs in the 3’UTR to target mRNAs for deadenylation and decay. Examination of the number and locations of conserved ARE motifs in the TXNIP mRNA transcript using a bioinformatics tool (AREsite 1.0 http://rna.tbi.univie.ac.at/cgi-bin/AREsite.cgi) identified only one ARE motif (ATTTA) in the 3’UTR of TXNIP mRNA (1934-1938 nt). As mRNAs targeted by RBPs usually contain multiple sites, the lack of ARE sites may account for the result that TXNIP mRNA is not regulated by these 3’UTR binding proteins.

<table>
<thead>
<tr>
<th>Gene</th>
<th>mRNA levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZFP36</td>
<td>74.07±17.04%</td>
</tr>
<tr>
<td>PUM1</td>
<td>60.03%</td>
</tr>
<tr>
<td>PUM2</td>
<td>15.76%</td>
</tr>
<tr>
<td>CELF1</td>
<td>46.16%</td>
</tr>
</tbody>
</table>

Figure 26. 3’UTR binding proteins ZFP36, PUM1, PUM2 and CELF1 are not involved in CNOT1-mediated regulation of TXNIP. SW480 cells were transfected with siRNA oligos targeting indicated 3’UTR binding proteins or non-targeting siRNA, twice, 48 hours apart, at a final concentration of 50 nM. (A) Relative mRNA levels of targeted genes. Relative mRNA level = levels of indicated RBPs in RBP siRNA-treated cells/control siRNA treated. (B) TXNIP mRNA levels in SW480 cells depleted for indicated RNA binding proteins. The values represent result of one experiment, except for ZNF36, which represents mean ± SD of two experiments. Significance of differences was calculated by a t-test.

miR-17 Downregulates TXNIP mRNA Levels

CNOT1 also interacts as an adapter between TNRC6A (GW182) of the microRNA RISC complex and CCR4-NOT deadenylase subunits, thus facilitating microRNA-mediated mRNA degradation. TXNIP has been reported to be degraded by
miR-17 suggesting that CNOT1 may facilitate miR-17-mediated degradation of TXNIP. However, miR-17 regulates TXNIP expression in a cell specific fashion. For instance, it is active in pancreatic beta cells but not in macrophages. The role of microRNAs in TXNIP mRNA stability in CRC cells is unknown so we first sought to determine if miR-17 regulates TXNIP in SW480 cells. In initial studies miR-17 was found to be present in SW480 cells (Figure 27A) and overexpression of a miR-17 mimic in SW480 cells resulted in a 31% decrease in levels of TXNIP mRNA. mRNA levels of two other miR-17 targets, which like TXNIP are upregulated in CNOT1 depleted cells, HBP1 (Li et al., 2011) and TGFBR2 (Dews et al., 2010), were decreased by 30% and 49%, respectively (Figure 27B). In addition, TXNIP levels were reduced by more than 50% in HEK293 cells transfected with the miR-17 mimic (Figure 27C). Results of these studies suggested that miR-17 downregulates TXNIP levels in SW480 and HEK293 cell lines.

These results support a model in which CNOT1 regulates TXNIP expression post-transcriptionally by facilitating the deadenylase activity of CNOT8. This CCR4-NOT complex may be recruited to the 3’UTR of TXNIP mRNA via interaction with a miRNA RISC complex containing miR-17.

A.

B.                                                 C.
**Figure 27.** miR-17 downregulates TXNIP mRNA levels. (A) Amplification curve showing amplification of endogenous miR-17 and U6 snRNA from SW480 cells using qRT-PCR. Ct (cycle threshold) = the number of cycles required for the fluorescent signal to cross the threshold (0.1 in this case). Total RNA including small RNAs was isolated using a miRNeasy Mini Kit (Qiagen). RNA was then transcribed to cDNA and ligated with universal adaptors using a miScript II RT Kit (Qiagen). Expression levels of mature miR-17 and U6 snRNA (internal control, abundant in most type of cells) were measured by qRT-PCR using a miScript SYBR Green PCR Kit with a miScript universal primer and specific primer for miR-17 and U6 from the miScript Primer Assay Plate (Qiagen). (B and C) SW480 and HEK293 cells were transfected with either miR-17 mimic oligos or non-targeting siRNA control at 5nM final concentration. Total RNA was extracted 24 hours after transfection and probed for known targets of miR-17 by qRT-PCR. Values represent ratio of mRNA expression in miR-17 mimic treated cells relative to non-targeting control treated cells. (B) Expression levels of three known mirR-17 targets, TXNIP, HBP1 and TGBR2, in SW480 cells. Values represent mean ± SD of two independent experiments. Significance of difference was calculated by a t-test. *** = p < 0.001, ** = p < 0.01. (C) Expression levels of TXNIP in HEK293 cells. Values represent result from one experiment.

**TXNIP and CNOT1 Are Associated with the Unfolded Protein Response (UPR).**

**TXNIP and the UPR.** TXNIP is activated to promote apoptosis by cellular conditions such as oxidative stress and extreme endoplasmic reticulum (ER) stress. Accumulation of unfolded proteins in the ER results in ER stress and activation of the unfolded protein response (UPR). Moderate ER stress leads to activation of an adaptive UPR, which initiates several pathways that decrease general translation and increase transcription and translation of factors to increase ER folding capacity. However, under
conditions of extreme stress the UPR initiates pathways leading to cell death, a process known as terminal UPR. It has recently been reported that activation of TXNIP can promote the switch from adaptive to terminal UPR. Further, this activation takes place in part by means of relief of miR-17-mediated degradation of TXNIP mRNA. Thus, CNOT1-mediated downregulation of TXNIP expression may contribute to protection from terminal UPR pathways.

To test this idea we first determined if the UPR contributed to apoptosis in SW480 cells. Cells were treated with different concentrations of tunicamycin, a potent pharmacological inducer of ER stress and the UPR, from 0.05 µg/ml to 1.0 µg/ml over a time period of 96 hours and cell viability was evaluated by MTT assay at intervals of 24 hours (Figure 28A). This experiment revealed that treatment with tunicamycin results in a dose and time dependent decrease in cell viability in SW480 cells. To confirm that treatment with tunicamycin can induce ER stress and activate UPR pathways, we also measured mRNA levels of genes involved in UPR signaling. We found that GRP78/Bip, sXBP1, ATF6, PERK and XBP1 mRNA levels showed at least two fold upregulation in SW480 cells treated with tunicamycin at 0.25 µg/ml for 6 hours (Figure 28B). This result suggested that tunicamycin could induce decreased cell viability through activation of UPR signaling.

Figure 28. Tunicamycin induces decreased cell viability and activation of UPR signaling in SW480 cells. (A) SW480 cells were plated at 2 X 10^4 cells per well in triplicate in 96 well plates. Cells were treated with tunicamycin at 0.05 µg/ml to 1.0 µg/ml and cell viability was measured
by MTT assay over a time period of 96 hours at intervals of 24 hours. Each data point represents mean ± SD of three replicates from an MTT assay. (B) SW480 cells at 70% confluence were treated with tunicamycin at 0.25 µg/ml for 6 hours. Total RNA was extracted and measured for mRNA expression levels of indicated genes. Values represent mean ± SD of two experiments. Significance of differences was determined by a t-test. *** = p < 0.001, * = p < 0.05.

To determine if TXNIP is involved in the response to UPR in SW480 cells we first examined the effect of tunicamycin treatment on TXNIP expression. Treatment with tunicamycin at 0.25 µg/ml for 6 hours resulted in upregulated TXNIP levels (Figure 29A) consistent with a role for TXNIP in the UPR in these cells. We then examined cell viability in cells treated with a moderate dose (0.1 µg/ml) of tunicamycin. TXNIP-depleted cells showed increased viability compared to control cells (Figure 29B) suggesting that TXNIP promotes the terminal UPR pathway in these cells.

Figure 29. TXNIP plays a role in the UPR in SW480 cells. (A) SW480 cells were treated with tunicamycin of 0.25 µg/ml for 6 hours. Total RNA was isolated and analyzed by RT-qPCR to determine TXNIP expression levels. Values represent mean ± SD of two experiments. Significance of differences was calculated by a t-test. (B) SW480 cells were transfected with either siRNA targeting TXNIP or control siRNA twice 48 hours apart. Cells were plated at 5000 cells per well in triplicate in 96 well plates 24 hours after the second transfection. Cells were treated with tunicamycin at 0.1 µg/ml and cell viability was measured by MTT assay over a time period of 96 hours as indicated. Each data point represents mean ± SD of three replicates from
one experiment. Significance of differences in cell viability was determined at 72 hour and 96 hour by a t-test. ** = p < 0.01.

**CNOT1 and the UPR.** To determine if CNOT1 also plays a role in the UPR in SW480 cells we examined expression of mRNAs upregulated during the UPR. qRT-PCR analysis revealed that mRNA levels of genes involved in all three pathways of the UPR are upregulated in CNOT1-depleted SW480 cells (Figure 30A), such as sXBPI (IRE1α pathway), PERK, ATF4 and DDIT3 (CHOP) (PERK pathway), and ATF6, XBPI (ATF6 pathway). It is important to note that only changes in mRNA levels were examined. To gain a full picture of the effect of CNOT1 on these pathways we will need to examine translational and post-translational regulation.

To determine if CNOT1 might be involved in apoptotic pathways initiated by TXNIP activation we examined expression of IL8 and NFKB1 and found that their levels were increased in CNOT1-depleted cells consistent with a model in which CNOT1 acts through TXNIP (Figure 30B).

**Figure 30. Depletion of CNOT1 results in ER stress and activation of UPR pathways.** (A) Quantification of mRNA expression levels of genes involved in UPR pathways in SW480 depleted for CNOT1 by qRT-PCR. (B) mRNA expression levels of NFKB1 and IL8 in SW480 depleted for CNOT1. Values represent mean ± SD of three independent experiments. Significance of differences was calculated by a t-test. *** = p < 0.001, * = p < 0.05.
To further characterize the role of CNOT1 in the unfolded protein response we performed GSEA to analyze the altered gene expression profile following CNOT1 depletion. A complete ranked expression dataset was generated in which genes were ranked according to ratio of expression in CNOT1-depleted cells to control treated cells. Genesets were compiled to represent gene expression in adaptive UPR or terminal UPR. GSEA revealed significant overlap with a group of mRNAs associated with response to extreme UPR known as Regulated IRE-1 Dependent Decay (RIDD) genes. These included RIDD genes degraded in mouse liver (So et al., 2012) and in pancreatic beta cells (Han et al., 2009) in response to extreme ER stress. Transcripts down-regulated in CNOT1-depleted cells were enriched in genes down-regulated in response to extreme ER stress, suggesting that downregulation of CNOT1 promotes RIDD and therefore CNOT1 expression is protective (Figure 31A, Table 6). In addition, comparison with the Molecular Signatures Database geneset collection of cellular components (MsigDB C5 CC, http://www.broadinstitute.org/gsea/msigdb/collections.jsp#C5) revealed that four of the five most enriched genesets consisted of plasma membrane localized components and thus transcripts that would be translated and processed in the ER (Figure 31B, Table 7). This result suggests that CNOT1 may also protect the cell from the UPR by limiting transcript levels. Together these results suggest that CNOT1 may protect against the effects of ER stress by mechanisms in addition to TXNIP regulation.

<table>
<thead>
<tr>
<th>GENESET</th>
<th>vs. CNOT1-depleted array</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NES</td>
</tr>
<tr>
<td>RIDD genes (So et al)</td>
<td>-1.99</td>
</tr>
<tr>
<td>Plasma membrane protein genes</td>
<td>2.02</td>
</tr>
</tbody>
</table>
**Figure 31. CNOT1 is involved in ER stress and activation of UPR pathways.** Gene set enrichment analysis (GSEA) of gene expression profile in SW480 cells depleted for CNOT1. Ranked CNOT1 expression data set (see text) was analyzed for enrichment in UPR-related genesets. (A) Representative enrichment plot of genesets compiled from gene expression studies of adaptive and terminal responses to the UPR. Shown is RIDD geneset from So et al (So et al., 2012). (B) Representative enrichment plot analysis of genesets found in MsigDB C5 CC collection, consisting of genes expressed in specific cellular components. Shown is enrichment plot for the GO geneset ‘PLASMA_MEMBRANE_PART’. (C) Normalized enrichment score (NES) and nominal p value (Nom. p value) are shown for each comparison.

**Figure 32. A model for roles of CNOT1 in regulating cell viability.**

To summarize, these studies demonstrate that depletion of CNOT1 in the human CRC cell line SW480 results in decreased cell viability, increased apoptosis and increased TXNIP levels. Increased TXNIP levels in HEK293 cells resulted in decreased cell viability. In addition, in CNOT1-depleted SW480 cells, elevation of TXNIP levels occurs before an increase in apoptosis. These results suggest that decreased cell viability in CNOT1-depleted cells is due to increased TXNIP levels. To further test this idea we will examine the effect of co-expression and co-depletion of CNOT1 and TXNIP. Further, our data indicate that CNOT1 destabilizes TXNIP mRNA possibly through miR-17-mediated degradation involving CNOT8 deadenylase. To test this idea we will examine 3’UTR reporter activity.
In pancreatic beta cells elevation of TXNIP levels occurs in response to extreme ER stress and leads to apoptosis. As increased TXNIP levels occur through post-transcriptional stabilization, CNOT1 expression may protect from and CNOT1 depletion may exacerbate the effects of extreme ER stress and the terminal unfolded protein response in SW480 cells. In support of this idea, depletion of TXNIP in SW480 cells protects against tunicamycin-induced cell death suggesting that CNOT1-mediated limitation of TXNIP may be protective. In support of a role for CNOT1, mRNAs from each of the major UPR pathways are upregulated in CNOT1-depleted cells. In addition, analysis of gene expression in CNOT1-depleted cells suggests that CNOT1 may limit the UPR by other mechanisms as well.

A model based on these results is shown in Figure 32. This model predicts that CNOT1 activity protects cell viability by limiting the effects of the terminal UPR. Severe cellular conditions, such as protein overproduction and oxidative stress, cause ER stress and activate UPR pathways. Under terminal, non-remediable ER stress, hyperactivation of IRE1α results in unregulated TXNIP expression, which in turn, activates apoptotic signaling pathways. CNOT1 activity may protect against the effects of severe ER stress through directly limiting TXNIP transcript levels. In addition, CNOT1 activity may limit the pool of mRNA to be translated in the ER.

However, two major gaps need to be filled. (1) Evidence that increased activity of CNOT1 can prevent cell apoptosis under severe ER stress, such as treatment with tunicamycin, needs to be provided. (2) We also need to show increased TXNIP activity is the cause of increased apoptosis in CNOT1-depleted cells and that limiting TXNIP activity can improve cell viability in these cells. As cancer cells are often subject to extreme stresses due to harsh growth conditions or enhanced protein production, increased activity of CNOT1 may protect cancer cells from apoptosis through this mechanism and confer growth advantage to cancer cells.
Supplemental Information:

Table 4. List of 266 genes showing significant upregulated (A) and downregulated (B) expression levels in SW480 cells depleted for CNOT1.

(A) Upregulated Genes:

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>log fold change</th>
<th>n fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>TXNIP</td>
<td>2.896508483</td>
<td>7.44622</td>
</tr>
<tr>
<td>DKK1</td>
<td>2.35158454</td>
<td>5.10385</td>
</tr>
<tr>
<td>DUSP5</td>
<td>2.150944437</td>
<td>4.44118</td>
</tr>
<tr>
<td>EMP1</td>
<td>2.096157218</td>
<td>4.27569</td>
</tr>
<tr>
<td>NA</td>
<td>1.973349402</td>
<td>3.92679</td>
</tr>
<tr>
<td>NRP1</td>
<td>1.96182098</td>
<td>3.89553</td>
</tr>
<tr>
<td>TNFRSF6B</td>
<td>1.923383113</td>
<td>3.79312</td>
</tr>
<tr>
<td>IL1RAP1</td>
<td>1.918700003</td>
<td>3.78082</td>
</tr>
<tr>
<td>LYPD6B</td>
<td>1.915699299</td>
<td>3.77297</td>
</tr>
<tr>
<td>SGK1</td>
<td>1.88472168</td>
<td>3.69282</td>
</tr>
<tr>
<td>S100P</td>
<td>1.863698467</td>
<td>3.63939</td>
</tr>
<tr>
<td>STEAP1</td>
<td>1.812791333</td>
<td>3.51321</td>
</tr>
<tr>
<td>BTG1</td>
<td>1.771000055</td>
<td>3.41290</td>
</tr>
<tr>
<td>PLNUR</td>
<td>1.762770283</td>
<td>3.39349</td>
</tr>
<tr>
<td>ZNF365</td>
<td>1.7489493</td>
<td>3.36114</td>
</tr>
<tr>
<td>MUC13</td>
<td>1.714169268</td>
<td>3.28108</td>
</tr>
<tr>
<td>PLIN2</td>
<td>1.6600699</td>
<td>3.16032</td>
</tr>
<tr>
<td>PROCR</td>
<td>1.654243386</td>
<td>3.14758</td>
</tr>
<tr>
<td>SERPINE1</td>
<td>1.648272383</td>
<td>3.13458</td>
</tr>
<tr>
<td>GPRC5A</td>
<td>1.643272601</td>
<td>3.12374</td>
</tr>
<tr>
<td>IGFBP3</td>
<td>1.637889705</td>
<td>3.11210</td>
</tr>
<tr>
<td>HAVCR2</td>
<td>1.629860019</td>
<td>3.09483</td>
</tr>
<tr>
<td>EFNB2</td>
<td>1.624880988</td>
<td>3.08417</td>
</tr>
<tr>
<td>LYPD6B</td>
<td>1.6078959</td>
<td>3.04807</td>
</tr>
<tr>
<td>IGFBP3</td>
<td>1.574230479</td>
<td>2.97777</td>
</tr>
<tr>
<td>ADAM19</td>
<td>1.567745266</td>
<td>2.96441</td>
</tr>
<tr>
<td>TM4SF1</td>
<td>1.565727169</td>
<td>2.96027</td>
</tr>
<tr>
<td>TNFRSF6B</td>
<td>1.563680091</td>
<td>2.95607</td>
</tr>
<tr>
<td>TGM2</td>
<td>1.56321117</td>
<td>2.95511</td>
</tr>
<tr>
<td>LPAR5</td>
<td>1.547263062</td>
<td>2.92262</td>
</tr>
<tr>
<td>ZBED2</td>
<td>1.522691159</td>
<td>2.87327</td>
</tr>
<tr>
<td>AKAP12</td>
<td>1.52109659</td>
<td>2.87009</td>
</tr>
<tr>
<td>Gene</td>
<td>Log2 Fold Change</td>
<td>p-Value</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------</td>
<td>---------</td>
</tr>
<tr>
<td>POLR1C</td>
<td>1.5002564</td>
<td>2.82848</td>
</tr>
<tr>
<td>PLAUR</td>
<td>1.498237703</td>
<td>2.82497</td>
</tr>
<tr>
<td>FZD9</td>
<td>1.48279479</td>
<td>2.79490</td>
</tr>
<tr>
<td>TMEM22</td>
<td>1.482594424</td>
<td>2.79451</td>
</tr>
<tr>
<td>SGK1</td>
<td>1.476720245</td>
<td>2.78315</td>
</tr>
<tr>
<td>C2orf49</td>
<td>1.471895329</td>
<td>2.77386</td>
</tr>
<tr>
<td>CD58</td>
<td>1.453009577</td>
<td>2.73779</td>
</tr>
<tr>
<td>NA</td>
<td>1.448775551</td>
<td>2.72995</td>
</tr>
<tr>
<td>NRP1</td>
<td>1.43272506</td>
<td>2.69956</td>
</tr>
<tr>
<td>COL7A1</td>
<td>1.419252403</td>
<td>2.67447</td>
</tr>
<tr>
<td>MMD</td>
<td>1.411019172</td>
<td>2.65925</td>
</tr>
<tr>
<td>SERPINB5</td>
<td>1.410881657</td>
<td>2.65900</td>
</tr>
<tr>
<td>SGK1</td>
<td>1.405235672</td>
<td>2.64861</td>
</tr>
<tr>
<td>POLR1C</td>
<td>1.404352218</td>
<td>2.64699</td>
</tr>
<tr>
<td>F2RL1</td>
<td>1.386263215</td>
<td>2.61401</td>
</tr>
<tr>
<td>UCN2</td>
<td>1.368591711</td>
<td>2.58218</td>
</tr>
<tr>
<td>BCAR3</td>
<td>1.368519302</td>
<td>2.58205</td>
</tr>
<tr>
<td>PRDM1</td>
<td>1.367655949</td>
<td>2.58051</td>
</tr>
<tr>
<td>KLF9</td>
<td>1.359572662</td>
<td>2.56609</td>
</tr>
<tr>
<td>CD58</td>
<td>1.354648683</td>
<td>2.55735</td>
</tr>
<tr>
<td>NA</td>
<td>1.349074792</td>
<td>2.54749</td>
</tr>
<tr>
<td>F2RL1</td>
<td>1.341613134</td>
<td>2.53435</td>
</tr>
<tr>
<td>KLF2</td>
<td>1.328770538</td>
<td>2.51189</td>
</tr>
<tr>
<td>CDKN2B</td>
<td>1.327136363</td>
<td>2.50904</td>
</tr>
<tr>
<td>MYC</td>
<td>1.322388951</td>
<td>2.50080</td>
</tr>
<tr>
<td>EDAR</td>
<td>1.322030806</td>
<td>2.50018</td>
</tr>
<tr>
<td>ZNF296</td>
<td>1.31673858</td>
<td>2.49126</td>
</tr>
<tr>
<td>PEL0</td>
<td>1.313247411</td>
<td>2.48500</td>
</tr>
<tr>
<td>PTRH2</td>
<td>1.312743744</td>
<td>2.48414</td>
</tr>
<tr>
<td>DNAJB9</td>
<td>1.309275207</td>
<td>2.47817</td>
</tr>
<tr>
<td>DHR52</td>
<td>1.300894478</td>
<td>2.46382</td>
</tr>
<tr>
<td>MYC</td>
<td>1.300779098</td>
<td>2.46362</td>
</tr>
<tr>
<td>DEPDC7</td>
<td>1.299956801</td>
<td>2.46222</td>
</tr>
<tr>
<td>IFRD1</td>
<td>1.292732494</td>
<td>2.44992</td>
</tr>
<tr>
<td>UBE2H</td>
<td>1.276342691</td>
<td>2.42224</td>
</tr>
<tr>
<td>TRIB1</td>
<td>1.266021197</td>
<td>2.40497</td>
</tr>
<tr>
<td>IFRD3</td>
<td>1.263630497</td>
<td>2.40099</td>
</tr>
<tr>
<td>PLIN2</td>
<td>1.261534137</td>
<td>2.39751</td>
</tr>
<tr>
<td>HIST1H2BD</td>
<td>1.261314556</td>
<td>2.39714</td>
</tr>
<tr>
<td>LMO2</td>
<td>1.257296455</td>
<td>2.39047</td>
</tr>
<tr>
<td>Gene</td>
<td>Log2 Fold Change</td>
<td>p-Value</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------</td>
<td>---------</td>
</tr>
<tr>
<td>ICAM2</td>
<td>1.25432976</td>
<td>2.38562</td>
</tr>
<tr>
<td>SERPINE2</td>
<td>1.253807761</td>
<td>2.38470</td>
</tr>
<tr>
<td>DKK4</td>
<td>1.244804631</td>
<td>2.36986</td>
</tr>
<tr>
<td>NOP2</td>
<td>1.244625355</td>
<td>2.36957</td>
</tr>
<tr>
<td>TNFRSF6B</td>
<td>1.243115008</td>
<td>2.36709</td>
</tr>
<tr>
<td>DEPDC7</td>
<td>1.23948181</td>
<td>2.36114</td>
</tr>
<tr>
<td>HEY2</td>
<td>1.235810152</td>
<td>2.35514</td>
</tr>
<tr>
<td>KRT81</td>
<td>1.231054712</td>
<td>2.34739</td>
</tr>
<tr>
<td>FRAT2</td>
<td>1.225722218</td>
<td>2.33872</td>
</tr>
<tr>
<td>SLC40A1</td>
<td>1.219612857</td>
<td>2.32884</td>
</tr>
<tr>
<td>TNFRSF11B</td>
<td>1.21452103</td>
<td>2.32064</td>
</tr>
<tr>
<td>TANC1</td>
<td>1.213628042</td>
<td>2.31920</td>
</tr>
<tr>
<td>AKIRIN2</td>
<td>1.212223406</td>
<td>2.31694</td>
</tr>
<tr>
<td>KANSL2</td>
<td>1.211242518</td>
<td>2.31537</td>
</tr>
<tr>
<td>VASN</td>
<td>1.211055295</td>
<td>2.31507</td>
</tr>
<tr>
<td>RGL1</td>
<td>1.210689082</td>
<td>2.31448</td>
</tr>
<tr>
<td>GZMB</td>
<td>1.202988495</td>
<td>2.30216</td>
</tr>
<tr>
<td>GNL3</td>
<td>1.198410379</td>
<td>2.29487</td>
</tr>
<tr>
<td>FAM216A</td>
<td>1.197542091</td>
<td>2.29349</td>
</tr>
<tr>
<td>HES7</td>
<td>1.195248472</td>
<td>2.28984</td>
</tr>
<tr>
<td>HS3ST3A1</td>
<td>1.192675668</td>
<td>2.28576</td>
</tr>
<tr>
<td>YRDC</td>
<td>1.191016461</td>
<td>2.28314</td>
</tr>
<tr>
<td>CST1</td>
<td>1.18858886</td>
<td>2.27930</td>
</tr>
<tr>
<td>KIAA0895</td>
<td>1.185541336</td>
<td>2.27449</td>
</tr>
<tr>
<td>KCTD14</td>
<td>1.184888387</td>
<td>2.27346</td>
</tr>
<tr>
<td>ARHGEF3</td>
<td>1.184740405</td>
<td>2.27322</td>
</tr>
<tr>
<td>PHLDA1</td>
<td>1.177206928</td>
<td>2.26139</td>
</tr>
<tr>
<td>TMEM177</td>
<td>1.175174863</td>
<td>2.25820</td>
</tr>
<tr>
<td>TM4SF19</td>
<td>1.174174787</td>
<td>2.25664</td>
</tr>
<tr>
<td>KRT6A</td>
<td>1.167536766</td>
<td>2.24628</td>
</tr>
<tr>
<td>RNFI49</td>
<td>1.165157634</td>
<td>2.24258</td>
</tr>
<tr>
<td>RNASE1</td>
<td>1.164472224</td>
<td>2.24151</td>
</tr>
<tr>
<td>ANXA1</td>
<td>1.160708624</td>
<td>2.23567</td>
</tr>
<tr>
<td>ITGA2</td>
<td>1.15583959</td>
<td>2.22814</td>
</tr>
<tr>
<td>NDRG1</td>
<td>1.155667332</td>
<td>2.22787</td>
</tr>
<tr>
<td>TSPAN5</td>
<td>1.155568505</td>
<td>2.22772</td>
</tr>
<tr>
<td>NOP16</td>
<td>1.152421033</td>
<td>2.22287</td>
</tr>
<tr>
<td>SLC30A1</td>
<td>1.151151774</td>
<td>2.22091</td>
</tr>
<tr>
<td>ATG16L1</td>
<td>1.144311522</td>
<td>2.21041</td>
</tr>
<tr>
<td>NOP16</td>
<td>1.139920452</td>
<td>2.20369</td>
</tr>
<tr>
<td>Gene</td>
<td>Fold Change</td>
<td>p-value</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>----------</td>
</tr>
<tr>
<td>ITGA5</td>
<td>1.137479692</td>
<td>2.19996</td>
</tr>
<tr>
<td>IRF1</td>
<td>1.135474794</td>
<td>2.19691</td>
</tr>
<tr>
<td>GTF2B</td>
<td>1.132604331</td>
<td>2.19254</td>
</tr>
<tr>
<td>RN28S1</td>
<td>1.132534635</td>
<td>2.19244</td>
</tr>
<tr>
<td>GJB3</td>
<td>1.13250918</td>
<td>2.19240</td>
</tr>
<tr>
<td>S100A3</td>
<td>1.12691402</td>
<td>2.18391</td>
</tr>
<tr>
<td>GJB4</td>
<td>1.126536321</td>
<td>2.18334</td>
</tr>
<tr>
<td>ARHGGEF2</td>
<td>1.12360666</td>
<td>2.17899</td>
</tr>
<tr>
<td>LGR6</td>
<td>1.123615779</td>
<td>2.17892</td>
</tr>
<tr>
<td>ADCK3</td>
<td>1.121741223</td>
<td>2.17609</td>
</tr>
<tr>
<td>GPRC5C</td>
<td>1.117589583</td>
<td>2.16984</td>
</tr>
<tr>
<td>TNFRSF21</td>
<td>1.110188935</td>
<td>2.15874</td>
</tr>
<tr>
<td>SRPK2</td>
<td>1.110075426</td>
<td>2.15857</td>
</tr>
<tr>
<td>LRRN4</td>
<td>1.109654821</td>
<td>2.15794</td>
</tr>
<tr>
<td>RCL1</td>
<td>1.107295885</td>
<td>2.15441</td>
</tr>
<tr>
<td>ECE2</td>
<td>1.107189669</td>
<td>2.15426</td>
</tr>
<tr>
<td>WFDC3</td>
<td>1.105573451</td>
<td>2.15184</td>
</tr>
<tr>
<td>PDE8B</td>
<td>1.104913843</td>
<td>2.15086</td>
</tr>
<tr>
<td>BRIX1</td>
<td>1.103104365</td>
<td>2.14816</td>
</tr>
<tr>
<td>MRPS30</td>
<td>1.102410059</td>
<td>2.14713</td>
</tr>
<tr>
<td>NGDN</td>
<td>1.098735412</td>
<td>2.14167</td>
</tr>
<tr>
<td>GTPBP4</td>
<td>1.097972613</td>
<td>2.14054</td>
</tr>
<tr>
<td>CXorf38</td>
<td>1.093019097</td>
<td>2.13320</td>
</tr>
<tr>
<td>IRF9</td>
<td>1.092852582</td>
<td>2.13295</td>
</tr>
<tr>
<td>ELF1</td>
<td>1.088877751</td>
<td>2.12709</td>
</tr>
<tr>
<td>EFNB1</td>
<td>1.085619757</td>
<td>2.12229</td>
</tr>
<tr>
<td>NA</td>
<td>1.080680745</td>
<td>2.11503</td>
</tr>
<tr>
<td>HBEGF</td>
<td>1.077663741</td>
<td>2.11062</td>
</tr>
<tr>
<td>NA</td>
<td>1.0726737</td>
<td>2.10333</td>
</tr>
<tr>
<td>BRIX1</td>
<td>1.072631123</td>
<td>2.10327</td>
</tr>
<tr>
<td>BMP4</td>
<td>1.066868563</td>
<td>2.09488</td>
</tr>
<tr>
<td>GNE</td>
<td>1.064478369</td>
<td>2.09141</td>
</tr>
<tr>
<td>TNFRSF1B</td>
<td>1.062264226</td>
<td>2.08821</td>
</tr>
<tr>
<td>PHLDB2</td>
<td>1.062150174</td>
<td>2.08804</td>
</tr>
<tr>
<td>C15orf39</td>
<td>1.061062307</td>
<td>2.08647</td>
</tr>
<tr>
<td>SLC52A3</td>
<td>1.058176403</td>
<td>2.08230</td>
</tr>
<tr>
<td>TPST2</td>
<td>1.053237269</td>
<td>2.07518</td>
</tr>
<tr>
<td>BIK</td>
<td>1.052793952</td>
<td>2.07454</td>
</tr>
<tr>
<td>PCGF1</td>
<td>1.050910242</td>
<td>2.07184</td>
</tr>
<tr>
<td>ERRFI1</td>
<td>1.049601191</td>
<td>2.06996</td>
</tr>
</tbody>
</table>
## (B) Downregulated Genes:

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>log fold change</th>
<th>n fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOTUM</td>
<td>-1.000398452</td>
<td>0.49986</td>
</tr>
<tr>
<td>CAPNS1</td>
<td>-1.001472128</td>
<td>0.49949</td>
</tr>
<tr>
<td>DPM3</td>
<td>-1.00613552</td>
<td>0.49788</td>
</tr>
<tr>
<td>MYO1D</td>
<td>-1.007438581</td>
<td>0.49743</td>
</tr>
<tr>
<td>PNPLA7</td>
<td>-1.008982297</td>
<td>0.49690</td>
</tr>
<tr>
<td>Gene</td>
<td>Log2 Fold Change</td>
<td>p-value</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------</td>
<td>----------</td>
</tr>
<tr>
<td>NA</td>
<td>-1.009959309</td>
<td>0.49656</td>
</tr>
<tr>
<td>ACAT2</td>
<td>-1.011950446</td>
<td>0.49588</td>
</tr>
<tr>
<td>FRG1B</td>
<td>-1.016104068</td>
<td>0.49445</td>
</tr>
<tr>
<td>PRDX5</td>
<td>-1.018125981</td>
<td>0.49376</td>
</tr>
<tr>
<td>ACTA2</td>
<td>-1.019750865</td>
<td>0.49320</td>
</tr>
<tr>
<td>LRRC26</td>
<td>-1.022477331</td>
<td>0.49227</td>
</tr>
<tr>
<td>GUSB</td>
<td>-1.024492829</td>
<td>0.49158</td>
</tr>
<tr>
<td>NBPF10</td>
<td>-1.026236051</td>
<td>0.49099</td>
</tr>
<tr>
<td>C9orf86</td>
<td>-1.02854056</td>
<td>0.49021</td>
</tr>
<tr>
<td>LARP1</td>
<td>-1.028968038</td>
<td>0.49006</td>
</tr>
<tr>
<td>PRDX2</td>
<td>-1.029784963</td>
<td>0.48978</td>
</tr>
<tr>
<td>NA</td>
<td>-1.034871243</td>
<td>0.48806</td>
</tr>
<tr>
<td>ATP5G2</td>
<td>-1.039293116</td>
<td>0.48657</td>
</tr>
<tr>
<td>STMN1</td>
<td>-1.039295822</td>
<td>0.48656</td>
</tr>
<tr>
<td>MT2A</td>
<td>-1.039511988</td>
<td>0.48649</td>
</tr>
<tr>
<td>ZNF358</td>
<td>-1.041108951</td>
<td>0.48595</td>
</tr>
<tr>
<td>CCDC3</td>
<td>-1.042282209</td>
<td>0.48556</td>
</tr>
<tr>
<td>BLOC1S1</td>
<td>-1.044234985</td>
<td>0.48490</td>
</tr>
<tr>
<td>TMEM139</td>
<td>-1.046066621</td>
<td>0.48411</td>
</tr>
<tr>
<td>PHPT1</td>
<td>-1.049034528</td>
<td>0.48329</td>
</tr>
<tr>
<td>SW15</td>
<td>-1.054176634</td>
<td>0.48157</td>
</tr>
<tr>
<td>RRM2</td>
<td>-1.055679283</td>
<td>0.48107</td>
</tr>
<tr>
<td>NA</td>
<td>-1.05646049</td>
<td>0.48081</td>
</tr>
<tr>
<td>DUT</td>
<td>-1.062637917</td>
<td>0.47876</td>
</tr>
<tr>
<td>DGCR6L</td>
<td>-1.06420164</td>
<td>0.47824</td>
</tr>
<tr>
<td>PRDX5</td>
<td>-1.065571077</td>
<td>0.47778</td>
</tr>
<tr>
<td>ID2</td>
<td>-1.084708887</td>
<td>0.47149</td>
</tr>
<tr>
<td>EEF1A1</td>
<td>-1.085389656</td>
<td>0.47126</td>
</tr>
<tr>
<td>HES6</td>
<td>-1.086841283</td>
<td>0.47079</td>
</tr>
<tr>
<td>COL18A1</td>
<td>-1.088102123</td>
<td>0.47038</td>
</tr>
<tr>
<td>NA</td>
<td>-1.089487748</td>
<td>0.46993</td>
</tr>
<tr>
<td>NA</td>
<td>-1.09305909</td>
<td>0.46877</td>
</tr>
<tr>
<td>CHMP3</td>
<td>-1.095295069</td>
<td>0.46804</td>
</tr>
<tr>
<td>NOTUM</td>
<td>-1.095386954</td>
<td>0.46801</td>
</tr>
<tr>
<td>ESPN</td>
<td>-1.097498842</td>
<td>0.46733</td>
</tr>
<tr>
<td>LYPD5</td>
<td>-1.100455961</td>
<td>0.46637</td>
</tr>
<tr>
<td>NEBL</td>
<td>-1.1127466</td>
<td>0.46241</td>
</tr>
<tr>
<td>ANAPC11</td>
<td>-1.113088303</td>
<td>0.46230</td>
</tr>
<tr>
<td>NA</td>
<td>-1.126632316</td>
<td>0.45798</td>
</tr>
<tr>
<td>Gene</td>
<td>Log2 Fold Change</td>
<td>p-value</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------</td>
<td>---------</td>
</tr>
<tr>
<td>TNNC1</td>
<td>-1.127174225</td>
<td>0.45781</td>
</tr>
<tr>
<td>SLC48A1</td>
<td>-1.129260008</td>
<td>0.45715</td>
</tr>
<tr>
<td>CHMP3</td>
<td>-1.135397638</td>
<td>0.45521</td>
</tr>
<tr>
<td>BCL7C</td>
<td>-1.14512003</td>
<td>0.45215</td>
</tr>
<tr>
<td>DGCR6</td>
<td>-1.147594581</td>
<td>0.45138</td>
</tr>
<tr>
<td>ASPSCR1</td>
<td>-1.151948655</td>
<td>0.45002</td>
</tr>
<tr>
<td>LAMA5</td>
<td>-1.153491384</td>
<td>0.44954</td>
</tr>
<tr>
<td>TMSB15B</td>
<td>-1.156822818</td>
<td>0.44850</td>
</tr>
<tr>
<td>NA</td>
<td>-1.172094204</td>
<td>0.44378</td>
</tr>
<tr>
<td>ITGAE</td>
<td>-1.177191077</td>
<td>0.44221</td>
</tr>
<tr>
<td>DUT</td>
<td>-1.17874924</td>
<td>0.44173</td>
</tr>
<tr>
<td>LSM4</td>
<td>-1.180022421</td>
<td>0.44134</td>
</tr>
<tr>
<td>RARRES2</td>
<td>-1.18033437</td>
<td>0.44125</td>
</tr>
<tr>
<td>TMSB15B</td>
<td>-1.181507537</td>
<td>0.44089</td>
</tr>
<tr>
<td>TSTD1</td>
<td>-1.183928672</td>
<td>0.44015</td>
</tr>
<tr>
<td>NA</td>
<td>-1.187084934</td>
<td>0.43919</td>
</tr>
<tr>
<td>SPC24</td>
<td>-1.187984996</td>
<td>0.43892</td>
</tr>
<tr>
<td>SLC25A23</td>
<td>-1.197016162</td>
<td>0.43618</td>
</tr>
<tr>
<td>C1orf86</td>
<td>-1.205394913</td>
<td>0.43365</td>
</tr>
<tr>
<td>C11orf51</td>
<td>-1.215371146</td>
<td>0.43066</td>
</tr>
<tr>
<td>C9orf142</td>
<td>-1.221651257</td>
<td>0.42879</td>
</tr>
<tr>
<td>NA</td>
<td>-1.225679708</td>
<td>0.42760</td>
</tr>
<tr>
<td>HMGB3</td>
<td>-1.230032617</td>
<td>0.42631</td>
</tr>
<tr>
<td>DGCR6</td>
<td>-1.259765955</td>
<td>0.41761</td>
</tr>
<tr>
<td>ANAPC11</td>
<td>-1.262579478</td>
<td>0.41680</td>
</tr>
<tr>
<td>RPS23</td>
<td>-1.279880514</td>
<td>0.41183</td>
</tr>
<tr>
<td>PDXK</td>
<td>-1.287642434</td>
<td>0.40962</td>
</tr>
<tr>
<td>GCHFR</td>
<td>-1.298126329</td>
<td>0.40665</td>
</tr>
<tr>
<td>NA</td>
<td>-1.305968633</td>
<td>0.40445</td>
</tr>
<tr>
<td>MAL2</td>
<td>-1.323249611</td>
<td>0.39963</td>
</tr>
<tr>
<td>AGRN</td>
<td>-1.341449303</td>
<td>0.39462</td>
</tr>
<tr>
<td>SKA2</td>
<td>-1.349357812</td>
<td>0.39247</td>
</tr>
<tr>
<td>C1orf56</td>
<td>-1.368946242</td>
<td>0.38717</td>
</tr>
<tr>
<td>NUDT14</td>
<td>-1.378213468</td>
<td>0.38469</td>
</tr>
<tr>
<td>GAS6</td>
<td>-1.503109748</td>
<td>0.35279</td>
</tr>
<tr>
<td>GAS6</td>
<td>-1.508886578</td>
<td>0.35138</td>
</tr>
<tr>
<td>TCEA3</td>
<td>-1.57567094</td>
<td>0.33549</td>
</tr>
<tr>
<td>PXMP2</td>
<td>-1.636131507</td>
<td>0.32172</td>
</tr>
<tr>
<td>CNOT1</td>
<td>-2.549125461</td>
<td>0.17086</td>
</tr>
</tbody>
</table>
Figure 33. Molecules in top three gene networks in IPA shown in Figure 20. Names in bold represent genes that change expression in CNOT1-depleted cells. Unbold represent interacting genes. Arrows show direction of change.

Table 5. Core genes enriched in CNOT1-HI/CNOT1-LO.

(A) Genes differentially expressed in proliferating intestinal crypts (Merlos-Suarez et al., 2011).
<table>
<thead>
<tr>
<th>row</th>
<th>Gene</th>
<th>Id</th>
<th>Fold Change</th>
<th>p-Value</th>
<th>Significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>KIF24</td>
<td>1073</td>
<td>0.7982</td>
<td>0.1120</td>
<td>Yes</td>
</tr>
<tr>
<td>19</td>
<td>TXNDC16</td>
<td>1118</td>
<td>0.7851</td>
<td>0.1165</td>
<td>Yes</td>
</tr>
<tr>
<td>20</td>
<td>SGOL2</td>
<td>1176</td>
<td>0.7697</td>
<td>0.1200</td>
<td>Yes</td>
</tr>
<tr>
<td>21</td>
<td>PUS7</td>
<td>1181</td>
<td>0.7683</td>
<td>0.1269</td>
<td>Yes</td>
</tr>
<tr>
<td>22</td>
<td>GLS2</td>
<td>1219</td>
<td>0.7601</td>
<td>0.1316</td>
<td>Yes</td>
</tr>
<tr>
<td>23</td>
<td>CKAP2L</td>
<td>1239</td>
<td>0.7536</td>
<td>0.1374</td>
<td>Yes</td>
</tr>
<tr>
<td>24</td>
<td>TBC1D4</td>
<td>1250</td>
<td>0.7518</td>
<td>0.1438</td>
<td>Yes</td>
</tr>
<tr>
<td>25</td>
<td>DACH1</td>
<td>1289</td>
<td>0.7422</td>
<td>0.1482</td>
<td>Yes</td>
</tr>
<tr>
<td>26</td>
<td>DNA2</td>
<td>1309</td>
<td>0.7364</td>
<td>0.1539</td>
<td>Yes</td>
</tr>
<tr>
<td>27</td>
<td>GEMIN4</td>
<td>1312</td>
<td>0.7354</td>
<td>0.1606</td>
<td>Yes</td>
</tr>
<tr>
<td>28</td>
<td>CELSR4</td>
<td>1317</td>
<td>0.7346</td>
<td>0.1672</td>
<td>Yes</td>
</tr>
<tr>
<td>29</td>
<td>FTSJD1</td>
<td>1329</td>
<td>0.7314</td>
<td>0.1734</td>
<td>Yes</td>
</tr>
<tr>
<td>30</td>
<td>MCM5</td>
<td>1332</td>
<td>0.7304</td>
<td>0.1801</td>
<td>Yes</td>
</tr>
<tr>
<td>31</td>
<td>TTPA</td>
<td>1369</td>
<td>0.7198</td>
<td>0.1844</td>
<td>Yes</td>
</tr>
<tr>
<td>32</td>
<td>CASC5</td>
<td>1420</td>
<td>0.7098</td>
<td>0.1878</td>
<td>Yes</td>
</tr>
<tr>
<td>33</td>
<td>LIG1</td>
<td>1633</td>
<td>0.6709</td>
<td>0.1802</td>
<td>Yes</td>
</tr>
<tr>
<td>34</td>
<td>SCML2</td>
<td>1649</td>
<td>0.6692</td>
<td>0.1854</td>
<td>Yes</td>
</tr>
<tr>
<td>35</td>
<td>CEP76</td>
<td>1654</td>
<td>0.6672</td>
<td>0.1914</td>
<td>Yes</td>
</tr>
<tr>
<td>36</td>
<td>CENPI</td>
<td>1728</td>
<td>0.6526</td>
<td>0.1927</td>
<td>Yes</td>
</tr>
<tr>
<td>37</td>
<td>PCK2</td>
<td>1768</td>
<td>0.6454</td>
<td>0.1962</td>
<td>Yes</td>
</tr>
<tr>
<td>38</td>
<td>ALMS1</td>
<td>1808</td>
<td>0.6355</td>
<td>0.1996</td>
<td>Yes</td>
</tr>
<tr>
<td>39</td>
<td>MCM8</td>
<td>1851</td>
<td>0.6288</td>
<td>0.2027</td>
<td>Yes</td>
</tr>
<tr>
<td>40</td>
<td>RRM2</td>
<td>1875</td>
<td>0.6239</td>
<td>0.2071</td>
<td>Yes</td>
</tr>
<tr>
<td>41</td>
<td>PIP4K2B</td>
<td>1904</td>
<td>0.6188</td>
<td>0.2110</td>
<td>Yes</td>
</tr>
<tr>
<td>42</td>
<td>CP110</td>
<td>1955</td>
<td>0.6120</td>
<td>0.2135</td>
<td>Yes</td>
</tr>
<tr>
<td>43</td>
<td>RPUSD2</td>
<td>2032</td>
<td>0.6002</td>
<td>0.2141</td>
<td>Yes</td>
</tr>
<tr>
<td>44</td>
<td>MYC</td>
<td>2034</td>
<td>0.6002</td>
<td>0.2196</td>
<td>Yes</td>
</tr>
<tr>
<td>45</td>
<td>MCM7</td>
<td>2082</td>
<td>0.5933</td>
<td>0.2221</td>
<td>Yes</td>
</tr>
<tr>
<td>46</td>
<td>DUS4L</td>
<td>2180</td>
<td>0.5806</td>
<td>0.2212</td>
<td>Yes</td>
</tr>
<tr>
<td>47</td>
<td>CCNF</td>
<td>2274</td>
<td>0.5656</td>
<td>0.2204</td>
<td>Yes</td>
</tr>
<tr>
<td>48</td>
<td>DSN1</td>
<td>2339</td>
<td>0.5547</td>
<td>0.2214</td>
<td>Yes</td>
</tr>
<tr>
<td>49</td>
<td>CDCA2</td>
<td>2344</td>
<td>0.5537</td>
<td>0.2263</td>
<td>Yes</td>
</tr>
<tr>
<td>50</td>
<td>MPHOSPH9</td>
<td>2356</td>
<td>0.5514</td>
<td>0.2307</td>
<td>Yes</td>
</tr>
<tr>
<td>51</td>
<td>ZBTB16</td>
<td>2381</td>
<td>0.5480</td>
<td>0.2343</td>
<td>Yes</td>
</tr>
<tr>
<td>52</td>
<td>BRCA1</td>
<td>2467</td>
<td>0.5356</td>
<td>0.2337</td>
<td>Yes</td>
</tr>
<tr>
<td>53</td>
<td>MELK</td>
<td>2485</td>
<td>0.5320</td>
<td>0.2376</td>
<td>Yes</td>
</tr>
<tr>
<td>54</td>
<td>DCK</td>
<td>2548</td>
<td>0.5229</td>
<td>0.2384</td>
<td>Yes</td>
</tr>
<tr>
<td>55</td>
<td>KIF18B</td>
<td>2557</td>
<td>0.5214</td>
<td>0.2428</td>
<td>Yes</td>
</tr>
<tr>
<td>56</td>
<td>KLHL23</td>
<td>2596</td>
<td>0.5176</td>
<td>0.2451</td>
<td>Yes</td>
</tr>
<tr>
<td>57</td>
<td>KBTBD6</td>
<td>2612</td>
<td>0.5153</td>
<td>0.2489</td>
<td>Yes</td>
</tr>
<tr>
<td>row</td>
<td>Gene</td>
<td>GeneID</td>
<td>Entrez</td>
<td>Freq</td>
<td>Entrez</td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>------</td>
<td>--------</td>
</tr>
<tr>
<td>row_58</td>
<td>TUBE1</td>
<td>2723</td>
<td>0.4975</td>
<td>0.2464</td>
<td>Yes</td>
</tr>
<tr>
<td>row_59</td>
<td>RPP40</td>
<td>2736</td>
<td>0.4959</td>
<td>0.2502</td>
<td>Yes</td>
</tr>
<tr>
<td>row_60</td>
<td>ZNF367</td>
<td>2747</td>
<td>0.4943</td>
<td>0.2542</td>
<td>Yes</td>
</tr>
<tr>
<td>row_61</td>
<td>TPX2</td>
<td>2771</td>
<td>0.4912</td>
<td>0.2573</td>
<td>Yes</td>
</tr>
<tr>
<td>row_62</td>
<td>INCENP</td>
<td>2829</td>
<td>0.4828</td>
<td>0.2581</td>
<td>Yes</td>
</tr>
<tr>
<td>row_63</td>
<td>KIF20B</td>
<td>2874</td>
<td>0.4770</td>
<td>0.2596</td>
<td>Yes</td>
</tr>
<tr>
<td>row_64</td>
<td>PFAS</td>
<td>2961</td>
<td>0.4659</td>
<td>0.2584</td>
<td>Yes</td>
</tr>
<tr>
<td>row_65</td>
<td>XRCC2</td>
<td>3117</td>
<td>0.4461</td>
<td>0.2524</td>
<td>Yes</td>
</tr>
<tr>
<td>row_66</td>
<td>SMC2</td>
<td>3130</td>
<td>0.4442</td>
<td>0.2557</td>
<td>Yes</td>
</tr>
<tr>
<td>row_67</td>
<td>ZNF275</td>
<td>3132</td>
<td>0.4438</td>
<td>0.2598</td>
<td>Yes</td>
</tr>
<tr>
<td>row_68</td>
<td>WDR35</td>
<td>3136</td>
<td>0.4428</td>
<td>0.2638</td>
<td>Yes</td>
</tr>
<tr>
<td>row_69</td>
<td>TSGA14</td>
<td>3145</td>
<td>0.4418</td>
<td>0.2674</td>
<td>Yes</td>
</tr>
<tr>
<td>row_70</td>
<td>SGOL1</td>
<td>3160</td>
<td>0.4397</td>
<td>0.2706</td>
<td>Yes</td>
</tr>
<tr>
<td>row_71</td>
<td>PRR11</td>
<td>3181</td>
<td>0.4373</td>
<td>0.2734</td>
<td>Yes</td>
</tr>
<tr>
<td>row_72</td>
<td>WDR76</td>
<td>3184</td>
<td>0.4368</td>
<td>0.2773</td>
<td>Yes</td>
</tr>
<tr>
<td>row_73</td>
<td>KIFC1</td>
<td>3220</td>
<td>0.4319</td>
<td>0.2791</td>
<td>Yes</td>
</tr>
<tr>
<td>row_74</td>
<td>DIAPH3</td>
<td>3228</td>
<td>0.4303</td>
<td>0.2826</td>
<td>Yes</td>
</tr>
<tr>
<td>row_75</td>
<td>KIF11</td>
<td>3293</td>
<td>0.4220</td>
<td>0.2824</td>
<td>Yes</td>
</tr>
<tr>
<td>row_76</td>
<td>KIAA0649</td>
<td>3301</td>
<td>0.4211</td>
<td>0.2859</td>
<td>Yes</td>
</tr>
<tr>
<td>row_77</td>
<td>NFATC2</td>
<td>3319</td>
<td>0.4188</td>
<td>0.2887</td>
<td>Yes</td>
</tr>
<tr>
<td>row_78</td>
<td>RFC3</td>
<td>3322</td>
<td>0.4187</td>
<td>0.2925</td>
<td>Yes</td>
</tr>
<tr>
<td>row_79</td>
<td>SHCBP1</td>
<td>3327</td>
<td>0.4184</td>
<td>0.2961</td>
<td>Yes</td>
</tr>
<tr>
<td>row_80</td>
<td>ASPM</td>
<td>3335</td>
<td>0.4170</td>
<td>0.2996</td>
<td>Yes</td>
</tr>
<tr>
<td>row_81</td>
<td>SLC7A2</td>
<td>3386</td>
<td>0.4117</td>
<td>0.3001</td>
<td>Yes</td>
</tr>
<tr>
<td>row_82</td>
<td>CLSPN</td>
<td>3433</td>
<td>0.4043</td>
<td>0.3009</td>
<td>Yes</td>
</tr>
<tr>
<td>row_83</td>
<td>HMMR</td>
<td>3462</td>
<td>0.4005</td>
<td>0.3028</td>
<td>Yes</td>
</tr>
<tr>
<td>row_84</td>
<td>TOPBP1</td>
<td>3492</td>
<td>0.3967</td>
<td>0.3046</td>
<td>Yes</td>
</tr>
<tr>
<td>row_85</td>
<td>CHEK1</td>
<td>3604</td>
<td>0.3841</td>
<td>0.3009</td>
<td>Yes</td>
</tr>
<tr>
<td>row_86</td>
<td>GSTCD</td>
<td>3622</td>
<td>0.3809</td>
<td>0.3034</td>
<td>Yes</td>
</tr>
<tr>
<td>row_87</td>
<td>POLE2</td>
<td>3733</td>
<td>0.3679</td>
<td>0.2996</td>
<td>Yes</td>
</tr>
<tr>
<td>row_88</td>
<td>TRIM37</td>
<td>3745</td>
<td>0.3661</td>
<td>0.3023</td>
<td>Yes</td>
</tr>
<tr>
<td>row_89</td>
<td>NRM</td>
<td>3779</td>
<td>0.3621</td>
<td>0.3035</td>
<td>Yes</td>
</tr>
<tr>
<td>row_90</td>
<td>FIGNL1</td>
<td>3820</td>
<td>0.3570</td>
<td>0.3042</td>
<td>Yes</td>
</tr>
<tr>
<td>row_91</td>
<td>RAD18</td>
<td>3826</td>
<td>0.3562</td>
<td>0.3072</td>
<td>Yes</td>
</tr>
<tr>
<td>row_92</td>
<td>NCAPH</td>
<td>3831</td>
<td>0.3550</td>
<td>0.3103</td>
<td>Yes</td>
</tr>
<tr>
<td>row_93</td>
<td>BLM</td>
<td>3866</td>
<td>0.3523</td>
<td>0.3114</td>
<td>Yes</td>
</tr>
<tr>
<td>row_94</td>
<td>KIF4A</td>
<td>4061</td>
<td>0.3300</td>
<td>0.3017</td>
<td>Yes</td>
</tr>
<tr>
<td>row_95</td>
<td>LMNB1</td>
<td>4075</td>
<td>0.3290</td>
<td>0.3039</td>
<td>Yes</td>
</tr>
<tr>
<td>row_96</td>
<td>HAUS5</td>
<td>4102</td>
<td>0.3263</td>
<td>0.3053</td>
<td>Yes</td>
</tr>
<tr>
<td>row_97</td>
<td>TAF4B</td>
<td>4154</td>
<td>0.3209</td>
<td>0.3049</td>
<td>Yes</td>
</tr>
<tr>
<td>Row</td>
<td>Name</td>
<td>Probe</td>
<td>Rank in Gene List</td>
<td>Rank Metric Score</td>
<td>Running Score</td>
</tr>
<tr>
<td>-----</td>
<td>-----------</td>
<td>-------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>row_98</td>
<td>MYBL2</td>
<td>4161</td>
<td>0.3203</td>
<td>0.3075</td>
<td>Yes</td>
</tr>
<tr>
<td>row_99</td>
<td>PLK4</td>
<td>4168</td>
<td>0.3191</td>
<td>0.3101</td>
<td>Yes</td>
</tr>
<tr>
<td>row_100</td>
<td>ESPL1</td>
<td>4199</td>
<td>0.3153</td>
<td>0.3111</td>
<td>Yes</td>
</tr>
<tr>
<td>row_101</td>
<td>ZNF473</td>
<td>4210</td>
<td>0.3141</td>
<td>0.3134</td>
<td>Yes</td>
</tr>
<tr>
<td>row_102</td>
<td>ESCO2</td>
<td>4265</td>
<td>0.3083</td>
<td>0.3127</td>
<td>Yes</td>
</tr>
<tr>
<td>row_103</td>
<td>ATAD5</td>
<td>4349</td>
<td>0.2978</td>
<td>0.3101</td>
<td>Yes</td>
</tr>
<tr>
<td>row_104</td>
<td>TCF19</td>
<td>4371</td>
<td>0.2956</td>
<td>0.3115</td>
<td>Yes</td>
</tr>
<tr>
<td>row_105</td>
<td>BCL7A</td>
<td>4383</td>
<td>0.2944</td>
<td>0.3135</td>
<td>Yes</td>
</tr>
<tr>
<td>row_106</td>
<td>MCM4</td>
<td>4458</td>
<td>0.2851</td>
<td>0.3113</td>
<td>Yes</td>
</tr>
<tr>
<td>row_107</td>
<td>NEIL3</td>
<td>4518</td>
<td>0.2788</td>
<td>0.3101</td>
<td>Yes</td>
</tr>
<tr>
<td>row_108</td>
<td>NCAPG2</td>
<td>4525</td>
<td>0.2782</td>
<td>0.3123</td>
<td>Yes</td>
</tr>
<tr>
<td>row_109</td>
<td>ATAD2</td>
<td>4556</td>
<td>0.2745</td>
<td>0.3129</td>
<td>Yes</td>
</tr>
<tr>
<td>row_110</td>
<td>TMEM107</td>
<td>4560</td>
<td>0.2743</td>
<td>0.3152</td>
<td>Yes</td>
</tr>
<tr>
<td>row_111</td>
<td>AURKA</td>
<td>4580</td>
<td>0.2672</td>
<td>0.3165</td>
<td>Yes</td>
</tr>
<tr>
<td>row_112</td>
<td>FAM54A</td>
<td>4603</td>
<td>0.2687</td>
<td>0.3176</td>
<td>Yes</td>
</tr>
<tr>
<td>row_113</td>
<td>BUB1</td>
<td>4677</td>
<td>0.2613</td>
<td>0.3153</td>
<td>Yes</td>
</tr>
<tr>
<td>row_114</td>
<td>KIF18A</td>
<td>4682</td>
<td>0.2607</td>
<td>0.3174</td>
<td>Yes</td>
</tr>
<tr>
<td>row_115</td>
<td>PALB2</td>
<td>4720</td>
<td>0.2571</td>
<td>0.3174</td>
<td>Yes</td>
</tr>
<tr>
<td>row_116</td>
<td>ANLN</td>
<td>4777</td>
<td>0.2505</td>
<td>0.3161</td>
<td>Yes</td>
</tr>
<tr>
<td>row_117</td>
<td>DKC1</td>
<td>4796</td>
<td>0.2648</td>
<td>0.3172</td>
<td>Yes</td>
</tr>
<tr>
<td>row_118</td>
<td>PSIP1</td>
<td>4809</td>
<td>0.2468</td>
<td>0.3187</td>
<td>Yes</td>
</tr>
<tr>
<td>row_119</td>
<td>NUSAP1</td>
<td>4832</td>
<td>0.2433</td>
<td>0.3196</td>
<td>Yes</td>
</tr>
<tr>
<td>row_120</td>
<td>POLD1</td>
<td>4833</td>
<td>0.2433</td>
<td>0.3219</td>
<td>Yes</td>
</tr>
</tbody>
</table>

(B) CIS-associated genes identified in SB screens by Starr et al. (Starr et al., 2009, Starr et al., 2011, Starr T.K., unpublished data).
<table>
<thead>
<tr>
<th>row</th>
<th>Gene</th>
<th>Value1</th>
<th>Value2</th>
<th>Value3</th>
<th>Value4</th>
</tr>
</thead>
<tbody>
<tr>
<td>row_11</td>
<td>SND1</td>
<td>452</td>
<td>1.0393</td>
<td>0.1654</td>
<td>Yes</td>
</tr>
<tr>
<td>row_12</td>
<td>PUM1</td>
<td>454</td>
<td>1.0387</td>
<td>0.1791</td>
<td>Yes</td>
</tr>
<tr>
<td>row_13</td>
<td>EXOC2</td>
<td>499</td>
<td>1.0067</td>
<td>0.1896</td>
<td>Yes</td>
</tr>
<tr>
<td>row_14</td>
<td>PPP4R2</td>
<td>526</td>
<td>0.9901</td>
<td>0.2011</td>
<td>Yes</td>
</tr>
<tr>
<td>row_15</td>
<td>UGGT1</td>
<td>550</td>
<td>0.9791</td>
<td>0.2126</td>
<td>Yes</td>
</tr>
<tr>
<td>row_16</td>
<td>PBRM1</td>
<td>569</td>
<td>0.9711</td>
<td>0.2243</td>
<td>Yes</td>
</tr>
<tr>
<td>row_17</td>
<td>GMDS</td>
<td>580</td>
<td>0.9652</td>
<td>0.2365</td>
<td>Yes</td>
</tr>
<tr>
<td>row_18</td>
<td>INPP5D</td>
<td>616</td>
<td>0.9466</td>
<td>0.2468</td>
<td>Yes</td>
</tr>
<tr>
<td>row_19</td>
<td>ATP8B1</td>
<td>649</td>
<td>0.9321</td>
<td>0.2571</td>
<td>Yes</td>
</tr>
<tr>
<td>row_20</td>
<td>STAG1</td>
<td>696</td>
<td>0.9174</td>
<td>0.2663</td>
<td>Yes</td>
</tr>
<tr>
<td>row_21</td>
<td>ROCK2</td>
<td>698</td>
<td>0.9170</td>
<td>0.2784</td>
<td>Yes</td>
</tr>
<tr>
<td>row_22</td>
<td>CNOT4</td>
<td>725</td>
<td>0.9033</td>
<td>0.2887</td>
<td>Yes</td>
</tr>
<tr>
<td>row_23</td>
<td>MKL2</td>
<td>791</td>
<td>0.8793</td>
<td>0.2962</td>
<td>Yes</td>
</tr>
<tr>
<td>row_24</td>
<td>UBAP2</td>
<td>876</td>
<td>0.8506</td>
<td>0.3020</td>
<td>Yes</td>
</tr>
<tr>
<td>row_25</td>
<td>UBN2</td>
<td>1049</td>
<td>0.8030</td>
<td>0.3015</td>
<td>Yes</td>
</tr>
<tr>
<td>row_26</td>
<td>TM9SF2</td>
<td>1130</td>
<td>0.7812</td>
<td>0.3067</td>
<td>Yes</td>
</tr>
<tr>
<td>row_27</td>
<td>ANKRD11</td>
<td>1197</td>
<td>0.7652</td>
<td>0.3125</td>
<td>Yes</td>
</tr>
<tr>
<td>row_28</td>
<td>MLL3</td>
<td>1222</td>
<td>0.7590</td>
<td>0.3211</td>
<td>Yes</td>
</tr>
<tr>
<td>row_29</td>
<td>KCTD20</td>
<td>1283</td>
<td>0.7440</td>
<td>0.3270</td>
<td>Yes</td>
</tr>
<tr>
<td>row_30</td>
<td>PDZD8</td>
<td>1285</td>
<td>0.7434</td>
<td>0.3369</td>
<td>Yes</td>
</tr>
<tr>
<td>row_31</td>
<td>DACH1</td>
<td>1289</td>
<td>0.7422</td>
<td>0.3465</td>
<td>Yes</td>
</tr>
<tr>
<td>row_32</td>
<td>SETD5</td>
<td>1337</td>
<td>0.7293</td>
<td>0.3532</td>
<td>Yes</td>
</tr>
<tr>
<td>row_33</td>
<td>RSFI1</td>
<td>1346</td>
<td>0.7278</td>
<td>0.3623</td>
<td>Yes</td>
</tr>
<tr>
<td>row_34</td>
<td>STXBP5</td>
<td>1497</td>
<td>0.6961</td>
<td>0.3618</td>
<td>Yes</td>
</tr>
<tr>
<td>row_35</td>
<td>CDK8</td>
<td>1519</td>
<td>0.6913</td>
<td>0.3696</td>
<td>Yes</td>
</tr>
<tr>
<td>row_36</td>
<td>NSD1</td>
<td>1522</td>
<td>0.6907</td>
<td>0.3787</td>
<td>Yes</td>
</tr>
<tr>
<td>row_37</td>
<td>RASGRF1</td>
<td>1546</td>
<td>0.6868</td>
<td>0.3863</td>
<td>Yes</td>
</tr>
<tr>
<td>row_38</td>
<td>XPO7</td>
<td>1591</td>
<td>0.6768</td>
<td>0.3924</td>
<td>Yes</td>
</tr>
<tr>
<td>row_39</td>
<td>ZNRF2</td>
<td>1699</td>
<td>0.6589</td>
<td>0.3942</td>
<td>Yes</td>
</tr>
<tr>
<td>row_40</td>
<td>MYO5B</td>
<td>1726</td>
<td>0.6528</td>
<td>0.4012</td>
<td>Yes</td>
</tr>
<tr>
<td>row_41</td>
<td>ZNF48</td>
<td>1795</td>
<td>0.6386</td>
<td>0.4052</td>
<td>Yes</td>
</tr>
<tr>
<td>row_42</td>
<td>TPP2</td>
<td>1892</td>
<td>0.6202</td>
<td>0.4072</td>
<td>Yes</td>
</tr>
<tr>
<td>row_43</td>
<td>PDCD6IP</td>
<td>1930</td>
<td>0.6150</td>
<td>0.4130</td>
<td>Yes</td>
</tr>
<tr>
<td>row_44</td>
<td>MAPK8</td>
<td>1954</td>
<td>0.6121</td>
<td>0.4196</td>
<td>Yes</td>
</tr>
<tr>
<td>row_45</td>
<td>ARID2</td>
<td>1973</td>
<td>0.6097</td>
<td>0.4266</td>
<td>Yes</td>
</tr>
<tr>
<td>row_46</td>
<td>CASP8</td>
<td>2004</td>
<td>0.6042</td>
<td>0.4326</td>
<td>Yes</td>
</tr>
<tr>
<td>row_47</td>
<td>GSK3B</td>
<td>2008</td>
<td>0.6039</td>
<td>0.4405</td>
<td>Yes</td>
</tr>
<tr>
<td>row_48</td>
<td>ZCCHC7</td>
<td>2052</td>
<td>0.5985</td>
<td>0.4456</td>
<td>Yes</td>
</tr>
<tr>
<td>row_49</td>
<td>ESCO1</td>
<td>2120</td>
<td>0.5883</td>
<td>0.4491</td>
<td>Yes</td>
</tr>
<tr>
<td>row_50</td>
<td>PAN3</td>
<td>2221</td>
<td>0.5740</td>
<td>0.4502</td>
<td>Yes</td>
</tr>
</tbody>
</table>
(C) CIS-associated genes identified in a SB screen by March et al (March et al., 2011).

<table>
<thead>
<tr>
<th>NAME</th>
<th>PROBE</th>
<th>RANK IN GENE LIST</th>
<th>RANK METRIC SCORE</th>
<th>RUNNING ES</th>
<th>CORE ENRICHMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>row_0</td>
<td>LRP5</td>
<td>22</td>
<td>1.9433</td>
<td>0.0295</td>
<td>Yes</td>
</tr>
<tr>
<td>row_1</td>
<td>CNOT1</td>
<td>31</td>
<td>1.8093</td>
<td>0.0577</td>
<td>Yes</td>
</tr>
<tr>
<td>row_2</td>
<td>BIRC6</td>
<td>116</td>
<td>1.4012</td>
<td>0.0745</td>
<td>Yes</td>
</tr>
<tr>
<td>row_3</td>
<td>NEDD4</td>
<td>132</td>
<td>1.3770</td>
<td>0.0954</td>
<td>Yes</td>
</tr>
<tr>
<td>row_4</td>
<td>LRIG1</td>
<td>148</td>
<td>1.3435</td>
<td>0.1158</td>
<td>Yes</td>
</tr>
<tr>
<td>row_5</td>
<td>PRKCA</td>
<td>202</td>
<td>1.2556</td>
<td>0.1324</td>
<td>Yes</td>
</tr>
<tr>
<td>row_6</td>
<td>PCF11</td>
<td>311</td>
<td>1.1427</td>
<td>0.1435</td>
<td>Yes</td>
</tr>
<tr>
<td>row</td>
<td>Gene</td>
<td>ID</td>
<td>logFC</td>
<td>pVal</td>
<td>Stat</td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>----</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>7</td>
<td>RREB1</td>
<td>314</td>
<td>1.1414</td>
<td>0.1615</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>ITGA6</td>
<td>339</td>
<td>1.1207</td>
<td>0.1778</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>NIPBL</td>
<td>395</td>
<td>1.0834</td>
<td>0.1914</td>
<td>Yes</td>
</tr>
<tr>
<td>10</td>
<td>PUM1</td>
<td>454</td>
<td>1.0387</td>
<td>0.2042</td>
<td>Yes</td>
</tr>
<tr>
<td>11</td>
<td>GMDS</td>
<td>580</td>
<td>0.9652</td>
<td>0.2114</td>
<td>Yes</td>
</tr>
<tr>
<td>12</td>
<td>MLL5</td>
<td>614</td>
<td>0.9470</td>
<td>0.2243</td>
<td>Yes</td>
</tr>
<tr>
<td>13</td>
<td>ATP8B1</td>
<td>649</td>
<td>0.9321</td>
<td>0.2369</td>
<td>Yes</td>
</tr>
<tr>
<td>14</td>
<td>ROCK2</td>
<td>698</td>
<td>0.9170</td>
<td>0.2484</td>
<td>Yes</td>
</tr>
<tr>
<td>15</td>
<td>NRIP1</td>
<td>894</td>
<td>0.8449</td>
<td>0.2856</td>
<td>Yes</td>
</tr>
<tr>
<td>16</td>
<td>ATAD2B</td>
<td>936</td>
<td>0.8326</td>
<td>0.2597</td>
<td>Yes</td>
</tr>
<tr>
<td>17</td>
<td>LGR4</td>
<td>944</td>
<td>0.8304</td>
<td>0.2724</td>
<td>Yes</td>
</tr>
<tr>
<td>18</td>
<td>PIP5K1B</td>
<td>946</td>
<td>0.8303</td>
<td>0.2856</td>
<td>Yes</td>
</tr>
<tr>
<td>19</td>
<td>DENND4C</td>
<td>973</td>
<td>0.8245</td>
<td>0.2970</td>
<td>Yes</td>
</tr>
<tr>
<td>20</td>
<td>ARID1B</td>
<td>1147</td>
<td>0.7759</td>
<td>0.3037</td>
<td>Yes</td>
</tr>
<tr>
<td>21</td>
<td>ATAD2B</td>
<td>936</td>
<td>0.8326</td>
<td>0.2597</td>
<td>Yes</td>
</tr>
<tr>
<td>22</td>
<td>LGR4</td>
<td>944</td>
<td>0.8304</td>
<td>0.2724</td>
<td>Yes</td>
</tr>
<tr>
<td>23</td>
<td>GTF2IRD1</td>
<td>1288</td>
<td>0.7430</td>
<td>0.3248</td>
<td>Yes</td>
</tr>
<tr>
<td>24</td>
<td>BAZ2B</td>
<td>1302</td>
<td>0.7381</td>
<td>0.3457</td>
<td>Yes</td>
</tr>
<tr>
<td>25</td>
<td>TLK1</td>
<td>1381</td>
<td>0.7167</td>
<td>0.3420</td>
<td>Yes</td>
</tr>
<tr>
<td>26</td>
<td>USP6NL</td>
<td>1477</td>
<td>0.7000</td>
<td>0.3470</td>
<td>Yes</td>
</tr>
<tr>
<td>27</td>
<td>NSD1</td>
<td>1522</td>
<td>0.6907</td>
<td>0.3551</td>
<td>Yes</td>
</tr>
<tr>
<td>28</td>
<td>NUP98</td>
<td>1571</td>
<td>0.6818</td>
<td>0.3628</td>
<td>Yes</td>
</tr>
<tr>
<td>29</td>
<td>ASXL1</td>
<td>1735</td>
<td>0.6515</td>
<td>0.3626</td>
<td>Yes</td>
</tr>
<tr>
<td>30</td>
<td>SLC25A20</td>
<td>1790</td>
<td>0.6403</td>
<td>0.3692</td>
<td>Yes</td>
</tr>
<tr>
<td>31</td>
<td>CTNND1</td>
<td>1829</td>
<td>0.6317</td>
<td>0.3768</td>
<td>Yes</td>
</tr>
<tr>
<td>32</td>
<td>SETD2</td>
<td>1918</td>
<td>0.6162</td>
<td>0.3809</td>
<td>Yes</td>
</tr>
<tr>
<td>33</td>
<td>ARID2</td>
<td>1973</td>
<td>0.6097</td>
<td>0.3871</td>
<td>Yes</td>
</tr>
<tr>
<td>34</td>
<td>GSK3B</td>
<td>2008</td>
<td>0.6039</td>
<td>0.3945</td>
<td>Yes</td>
</tr>
<tr>
<td>35</td>
<td>WAPAL</td>
<td>2204</td>
<td>0.5766</td>
<td>0.3910</td>
<td>Yes</td>
</tr>
<tr>
<td>36</td>
<td>PAN3</td>
<td>2221</td>
<td>0.5740</td>
<td>0.3990</td>
<td>Yes</td>
</tr>
<tr>
<td>37</td>
<td>WDR33</td>
<td>2236</td>
<td>0.5715</td>
<td>0.4072</td>
<td>Yes</td>
</tr>
<tr>
<td>38</td>
<td>PDXDC1</td>
<td>2251</td>
<td>0.5689</td>
<td>0.4154</td>
<td>Yes</td>
</tr>
<tr>
<td>39</td>
<td>RBBP1</td>
<td>2314</td>
<td>0.5584</td>
<td>0.4202</td>
<td>Yes</td>
</tr>
<tr>
<td>40</td>
<td>DDX6</td>
<td>2342</td>
<td>0.5540</td>
<td>0.4273</td>
<td>Yes</td>
</tr>
<tr>
<td>41</td>
<td>PTPRK</td>
<td>2355</td>
<td>0.5515</td>
<td>0.4352</td>
<td>Yes</td>
</tr>
<tr>
<td>42</td>
<td>ATRX</td>
<td>2476</td>
<td>0.5342</td>
<td>0.4359</td>
<td>Yes</td>
</tr>
<tr>
<td>43</td>
<td>PIK3R1</td>
<td>2655</td>
<td>0.5081</td>
<td>0.4324</td>
<td>Yes</td>
</tr>
<tr>
<td>44</td>
<td>ANKRD17</td>
<td>2684</td>
<td>0.5044</td>
<td>0.4386</td>
<td>Yes</td>
</tr>
<tr>
<td>45</td>
<td>MPP5</td>
<td>2785</td>
<td>0.4889</td>
<td>0.4399</td>
<td>Yes</td>
</tr>
<tr>
<td>46</td>
<td>MAN1A2</td>
<td>2794</td>
<td>0.4879</td>
<td>0.4472</td>
<td>Yes</td>
</tr>
<tr>
<td>row_47</td>
<td>ATP2A2</td>
<td>2841</td>
<td>0.4813</td>
<td>0.4518</td>
<td>Yes</td>
</tr>
<tr>
<td>row_48</td>
<td>VPS24</td>
<td>2916</td>
<td>0.4727</td>
<td>0.4545</td>
<td>Yes</td>
</tr>
<tr>
<td>row_49</td>
<td>ZFHX3</td>
<td>2948</td>
<td>0.4674</td>
<td>0.4599</td>
<td>Yes</td>
</tr>
<tr>
<td>row_50</td>
<td>SDC1</td>
<td>2967</td>
<td>0.4652</td>
<td>0.4662</td>
<td>Yes</td>
</tr>
<tr>
<td>row_51</td>
<td>APC</td>
<td>2968</td>
<td>0.4648</td>
<td>0.4736</td>
<td>Yes</td>
</tr>
<tr>
<td>row_52</td>
<td>FNBP1L</td>
<td>2971</td>
<td>0.4645</td>
<td>0.4808</td>
<td>Yes</td>
</tr>
<tr>
<td>row_53</td>
<td>CAPRIN1</td>
<td>3004</td>
<td>0.4601</td>
<td>0.4860</td>
<td>Yes</td>
</tr>
<tr>
<td>row_54</td>
<td>ARHGAP5</td>
<td>3009</td>
<td>0.4595</td>
<td>0.4931</td>
<td>Yes</td>
</tr>
<tr>
<td>row_55</td>
<td>KCNQ1</td>
<td>3018</td>
<td>0.4588</td>
<td>0.4999</td>
<td>Yes</td>
</tr>
<tr>
<td>row_56</td>
<td>FAM116A</td>
<td>3043</td>
<td>0.4550</td>
<td>0.5055</td>
<td>Yes</td>
</tr>
<tr>
<td>row_57</td>
<td>ERBB2IP</td>
<td>3118</td>
<td>0.4459</td>
<td>0.5078</td>
<td>Yes</td>
</tr>
<tr>
<td>row_58</td>
<td>GAB1</td>
<td>3194</td>
<td>0.4355</td>
<td>0.5099</td>
<td>Yes</td>
</tr>
<tr>
<td>row_59</td>
<td>NFIB</td>
<td>3211</td>
<td>0.4331</td>
<td>0.5157</td>
<td>Yes</td>
</tr>
<tr>
<td>row_60</td>
<td>ACVR1B</td>
<td>3347</td>
<td>0.4161</td>
<td>0.5136</td>
<td>Yes</td>
</tr>
<tr>
<td>row_61</td>
<td>MYOM1</td>
<td>3383</td>
<td>0.4121</td>
<td>0.5178</td>
<td>Yes</td>
</tr>
<tr>
<td>row_62</td>
<td>SYNRIP</td>
<td>3427</td>
<td>0.4059</td>
<td>0.5215</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 6. Core RIDD genes enriched in CNOT1-depleted SW480 cells (So et al., 2012).

| row_19 | COL18A1 | 5581 | -0.082292341 | -0.4058358 | Yes |
| row_20 | BTBD1   | 5619 | -0.083633006 | -0.38739416 | Yes |
| row_21 | DERA    | 5641 | -0.084373057 | -0.36650765 | Yes |
| row_22 | PAPOLA  | 5705 | -0.087333411 | -0.35065457 | Yes |
| row_23 | FDPS    | 5961 | -0.09705963  | -0.3588861  | Yes |
| row_24 | SLC27A5 | 6177 | -0.108391821 | -0.3583285  | Yes |
| row_25 | ABCA2   | 6183 | -0.108704239 | -0.3283367  | Yes |
| row_26 | FADS1   | 6184 | -0.108723998 | -0.29764056 | Yes |
| row_27 | PTPRF   | 6232 | -0.11189945  | -0.27261585 | Yes |
| row_28 | BOLA3   | 6366 | -0.120483741 | -0.2571854  | Yes |
| row_29 | ID11    | 6518 | -0.130967647 | -0.2413104  | Yes |
| row_30 | RNF130  | 6563 | -0.13481155  | -0.20939767 | Yes |
| row_31 | TM7SF2  | 6566 | -0.135234132 | -0.17149644 | Yes |
| row_32 | PMVK    | 6816 | -0.162154734 | -0.16051117 | Yes |
| row_33 | TPP1    | 7127 | -0.245197475 | -0.13460472 | Yes |
| row_34 | BLOC1S1 | 7129 | -0.246350572 | -0.06519217 | Yes |
| row_35 | PXMP2   | 7145 | -0.261100173 | 0.006428248 | Yes |

Table 7. Plasma membrane localized components (PLASMA_MEMBRANE_PART) enriched in CNOT1-depleted SW480 cells.
<table>
<thead>
<tr>
<th></th>
<th>PROBE</th>
<th>RANK IN GENE LIST</th>
<th>RANK METRIC SCORE</th>
<th>RUNNING ES</th>
<th>CORE ENRICHMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GPRC5A</td>
<td>11</td>
<td>0.466</td>
<td>0.0112</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>PROCR</td>
<td>12</td>
<td>0.463</td>
<td>0.024</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>EFN B2</td>
<td>13</td>
<td>0.46</td>
<td>0.0367</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>STEAP1</td>
<td>16</td>
<td>0.434</td>
<td>0.0484</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>SLC40A1</td>
<td>28</td>
<td>0.378</td>
<td>0.0572</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>F2RL1</td>
<td>29</td>
<td>0.377</td>
<td>0.0676</td>
<td>Yes</td>
</tr>
<tr>
<td>7</td>
<td>GZMB</td>
<td>43</td>
<td>0.354</td>
<td>0.0755</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>ITGA2</td>
<td>53</td>
<td>0.331</td>
<td>0.0833</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>TM4SF1</td>
<td>55</td>
<td>0.328</td>
<td>0.0922</td>
<td>Yes</td>
</tr>
<tr>
<td>10</td>
<td>ITGA5</td>
<td>56</td>
<td>0.327</td>
<td>0.1013</td>
<td>Yes</td>
</tr>
<tr>
<td>11</td>
<td>MMD</td>
<td>62</td>
<td>0.322</td>
<td>0.1094</td>
<td>Yes</td>
</tr>
<tr>
<td>12</td>
<td>ICAM2</td>
<td>70</td>
<td>0.311</td>
<td>0.117</td>
<td>Yes</td>
</tr>
<tr>
<td>13</td>
<td>HBE GF</td>
<td>79</td>
<td>0.306</td>
<td>0.1243</td>
<td>Yes</td>
</tr>
<tr>
<td>14</td>
<td>KCNK6</td>
<td>80</td>
<td>0.306</td>
<td>0.1327</td>
<td>Yes</td>
</tr>
<tr>
<td>15</td>
<td>SLC16A6</td>
<td>103</td>
<td>0.285</td>
<td>0.1374</td>
<td>Yes</td>
</tr>
<tr>
<td>16</td>
<td>TNFSF10</td>
<td>139</td>
<td>0.266</td>
<td>0.1396</td>
<td>Yes</td>
</tr>
<tr>
<td>17</td>
<td>EFN B1</td>
<td>162</td>
<td>0.258</td>
<td>0.1435</td>
<td>Yes</td>
</tr>
<tr>
<td>18</td>
<td>ITGA1</td>
<td>180</td>
<td>0.252</td>
<td>0.148</td>
<td>Yes</td>
</tr>
<tr>
<td>19</td>
<td>SLC20A2</td>
<td>251</td>
<td>0.229</td>
<td>0.1441</td>
<td>Yes</td>
</tr>
<tr>
<td>20</td>
<td>ZAP70</td>
<td>252</td>
<td>0.229</td>
<td>0.1504</td>
<td>Yes</td>
</tr>
<tr>
<td>21</td>
<td>EPHA4</td>
<td>258</td>
<td>0.227</td>
<td>0.156</td>
<td>Yes</td>
</tr>
<tr>
<td>22</td>
<td>EGFR</td>
<td>262</td>
<td>0.226</td>
<td>0.1618</td>
<td>Yes</td>
</tr>
<tr>
<td>23</td>
<td>SLC31A2</td>
<td>274</td>
<td>0.222</td>
<td>0.1663</td>
<td>Yes</td>
</tr>
<tr>
<td>24</td>
<td>GPRC5C</td>
<td>299</td>
<td>0.217</td>
<td>0.1688</td>
<td>Yes</td>
</tr>
<tr>
<td>25</td>
<td>CD24</td>
<td>305</td>
<td>0.215</td>
<td>0.174</td>
<td>Yes</td>
</tr>
<tr>
<td>26</td>
<td>KCNN4</td>
<td>307</td>
<td>0.215</td>
<td>0.1798</td>
<td>Yes</td>
</tr>
<tr>
<td>27</td>
<td>GPER</td>
<td>308</td>
<td>0.215</td>
<td>0.1857</td>
<td>Yes</td>
</tr>
<tr>
<td>28</td>
<td>LAT</td>
<td>322</td>
<td>0.212</td>
<td>0.1896</td>
<td>Yes</td>
</tr>
<tr>
<td>29</td>
<td>CCR7</td>
<td>323</td>
<td>0.212</td>
<td>0.1955</td>
<td>Yes</td>
</tr>
<tr>
<td>30</td>
<td>PRRG1</td>
<td>354</td>
<td>0.207</td>
<td>0.1968</td>
<td>Yes</td>
</tr>
<tr>
<td>31</td>
<td>IFNGR1</td>
<td>358</td>
<td>0.206</td>
<td>0.2021</td>
<td>Yes</td>
</tr>
<tr>
<td>32</td>
<td>CD55</td>
<td>410</td>
<td>0.196</td>
<td>0.2001</td>
<td>Yes</td>
</tr>
<tr>
<td>33</td>
<td>ACVR1</td>
<td>414</td>
<td>0.195</td>
<td>0.205</td>
<td>Yes</td>
</tr>
<tr>
<td>34</td>
<td>GJB2</td>
<td>447</td>
<td>0.191</td>
<td>0.2056</td>
<td>Yes</td>
</tr>
<tr>
<td>35</td>
<td>RCE1</td>
<td>452</td>
<td>0.19</td>
<td>0.2103</td>
<td>Yes</td>
</tr>
<tr>
<td>36</td>
<td>RGMB</td>
<td>459</td>
<td>0.189</td>
<td>0.2146</td>
<td>Yes</td>
</tr>
<tr>
<td>37</td>
<td>CD33</td>
<td>469</td>
<td>0.188</td>
<td>0.2185</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Genes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>--------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>38</td>
<td>TLR4</td>
<td>494</td>
<td>0.185</td>
<td>0.2201</td>
<td>Yes</td>
</tr>
<tr>
<td>39</td>
<td>MALL</td>
<td>521</td>
<td>0.181</td>
<td>0.2213</td>
<td>Yes</td>
</tr>
<tr>
<td>40</td>
<td>ZYX</td>
<td>526</td>
<td>0.181</td>
<td>0.2257</td>
<td>Yes</td>
</tr>
<tr>
<td>41</td>
<td>TRIP6</td>
<td>530</td>
<td>0.18</td>
<td>0.2303</td>
<td>Yes</td>
</tr>
<tr>
<td>42</td>
<td>TNFRSF1A</td>
<td>535</td>
<td>0.18</td>
<td>0.2346</td>
<td>Yes</td>
</tr>
<tr>
<td>43</td>
<td>CCR3</td>
<td>538</td>
<td>0.18</td>
<td>0.2393</td>
<td>Yes</td>
</tr>
<tr>
<td>44</td>
<td>SLC25A13</td>
<td>539</td>
<td>0.18</td>
<td>0.2443</td>
<td>Yes</td>
</tr>
<tr>
<td>45</td>
<td>BCAR1</td>
<td>546</td>
<td>0.179</td>
<td>0.2483</td>
<td>Yes</td>
</tr>
<tr>
<td>46</td>
<td>EFNA4</td>
<td>600</td>
<td>0.172</td>
<td>0.2454</td>
<td>Yes</td>
</tr>
<tr>
<td>47</td>
<td>TRIM27</td>
<td>603</td>
<td>0.172</td>
<td>0.2498</td>
<td>Yes</td>
</tr>
<tr>
<td>48</td>
<td>TACSTD2</td>
<td>617</td>
<td>0.17</td>
<td>0.2526</td>
<td>Yes</td>
</tr>
<tr>
<td>49</td>
<td>IL27RA</td>
<td>634</td>
<td>0.169</td>
<td>0.255</td>
<td>Yes</td>
</tr>
<tr>
<td>50</td>
<td>ACVR1B</td>
<td>644</td>
<td>0.168</td>
<td>0.2583</td>
<td>Yes</td>
</tr>
<tr>
<td>51</td>
<td>CD97</td>
<td>675</td>
<td>0.164</td>
<td>0.2584</td>
<td>Yes</td>
</tr>
<tr>
<td>52</td>
<td>EPHA2</td>
<td>698</td>
<td>0.162</td>
<td>0.2597</td>
<td>Yes</td>
</tr>
<tr>
<td>53</td>
<td>PRNP</td>
<td>718</td>
<td>0.16</td>
<td>0.2613</td>
<td>Yes</td>
</tr>
<tr>
<td>54</td>
<td>SLC7A2</td>
<td>721</td>
<td>0.16</td>
<td>0.2655</td>
<td>Yes</td>
</tr>
<tr>
<td>55</td>
<td>ATP6V0A4</td>
<td>745</td>
<td>0.158</td>
<td>0.2665</td>
<td>Yes</td>
</tr>
<tr>
<td>56</td>
<td>TRPM2</td>
<td>769</td>
<td>0.156</td>
<td>0.2674</td>
<td>Yes</td>
</tr>
<tr>
<td>57</td>
<td>CXCR4</td>
<td>793</td>
<td>0.153</td>
<td>0.2683</td>
<td>Yes</td>
</tr>
<tr>
<td>58</td>
<td>ADORA2B</td>
<td>819</td>
<td>0.151</td>
<td>0.2688</td>
<td>Yes</td>
</tr>
<tr>
<td>59</td>
<td>VIPR1</td>
<td>864</td>
<td>0.148</td>
<td>0.2665</td>
<td>Yes</td>
</tr>
<tr>
<td>60</td>
<td>GPR35</td>
<td>886</td>
<td>0.146</td>
<td>0.2675</td>
<td>Yes</td>
</tr>
<tr>
<td>61</td>
<td>SLC19A1</td>
<td>912</td>
<td>0.144</td>
<td>0.2678</td>
<td>Yes</td>
</tr>
<tr>
<td>62</td>
<td>NTSR1</td>
<td>913</td>
<td>0.144</td>
<td>0.2718</td>
<td>Yes</td>
</tr>
<tr>
<td>63</td>
<td>PDZD2</td>
<td>916</td>
<td>0.144</td>
<td>0.2755</td>
<td>Yes</td>
</tr>
<tr>
<td>64</td>
<td>SLC2A8</td>
<td>930</td>
<td>0.143</td>
<td>0.2775</td>
<td>Yes</td>
</tr>
<tr>
<td>65</td>
<td>IFNGR2</td>
<td>931</td>
<td>0.143</td>
<td>0.2815</td>
<td>Yes</td>
</tr>
<tr>
<td>66</td>
<td>P2RY2</td>
<td>937</td>
<td>0.143</td>
<td>0.2847</td>
<td>Yes</td>
</tr>
<tr>
<td>67</td>
<td>IL28RA</td>
<td>940</td>
<td>0.143</td>
<td>0.2884</td>
<td>Yes</td>
</tr>
<tr>
<td>68</td>
<td>LAT2</td>
<td>949</td>
<td>0.142</td>
<td>0.2911</td>
<td>Yes</td>
</tr>
<tr>
<td>69</td>
<td>ORAI1</td>
<td>953</td>
<td>0.142</td>
<td>0.2946</td>
<td>Yes</td>
</tr>
<tr>
<td>70</td>
<td>CTLA4</td>
<td>958</td>
<td>0.141</td>
<td>0.2979</td>
<td>Yes</td>
</tr>
<tr>
<td>71</td>
<td>DCBLD2</td>
<td>959</td>
<td>0.141</td>
<td>0.3018</td>
<td>Yes</td>
</tr>
<tr>
<td>72</td>
<td>EFNA1</td>
<td>985</td>
<td>0.139</td>
<td>0.302</td>
<td>Yes</td>
</tr>
<tr>
<td>73</td>
<td>ROR2</td>
<td>987</td>
<td>0.139</td>
<td>0.3057</td>
<td>Yes</td>
</tr>
<tr>
<td>74</td>
<td>CCRL2</td>
<td>993</td>
<td>0.138</td>
<td>0.3087</td>
<td>Yes</td>
</tr>
<tr>
<td>75</td>
<td>ABCA1</td>
<td>1027</td>
<td>0.136</td>
<td>0.3077</td>
<td>Yes</td>
</tr>
<tr>
<td>76</td>
<td>SLC7A7</td>
<td>1040</td>
<td>0.135</td>
<td>0.3097</td>
<td>Yes</td>
</tr>
<tr>
<td>77</td>
<td>BST2</td>
<td>1047</td>
<td>0.134</td>
<td>0.3125</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Gene Symbol</td>
<td>Gene ID</td>
<td>log_{10} p-Value</td>
<td>LogfoldChange</td>
<td>Status</td>
</tr>
<tr>
<td>---</td>
<td>-------------</td>
<td>---------</td>
<td>------------------</td>
<td>--------------</td>
<td>--------</td>
</tr>
<tr>
<td>78</td>
<td>CD83</td>
<td>1053</td>
<td>0.134</td>
<td>0.3155</td>
<td>Yes</td>
</tr>
<tr>
<td>79</td>
<td>IL2RB</td>
<td>1061</td>
<td>0.133</td>
<td>0.3181</td>
<td>Yes</td>
</tr>
<tr>
<td>80</td>
<td>MC1R</td>
<td>1068</td>
<td>0.133</td>
<td>0.3209</td>
<td>Yes</td>
</tr>
<tr>
<td>81</td>
<td>GZMA</td>
<td>1069</td>
<td>0.133</td>
<td>0.3246</td>
<td>Yes</td>
</tr>
<tr>
<td>82</td>
<td>CAV1</td>
<td>1076</td>
<td>0.133</td>
<td>0.3274</td>
<td>Yes</td>
</tr>
<tr>
<td>83</td>
<td>STOM</td>
<td>1105</td>
<td>0.131</td>
<td>0.3269</td>
<td>Yes</td>
</tr>
<tr>
<td>84</td>
<td>SYTL2</td>
<td>1116</td>
<td>0.13</td>
<td>0.3291</td>
<td>Yes</td>
</tr>
<tr>
<td>85</td>
<td>SLMAP</td>
<td>1118</td>
<td>0.13</td>
<td>0.3325</td>
<td>Yes</td>
</tr>
<tr>
<td>86</td>
<td>AXIN1</td>
<td>1126</td>
<td>0.13</td>
<td>0.3351</td>
<td>Yes</td>
</tr>
<tr>
<td>87</td>
<td>IL4R</td>
<td>1132</td>
<td>0.13</td>
<td>0.3379</td>
<td>Yes</td>
</tr>
<tr>
<td>88</td>
<td>VCL</td>
<td>1161</td>
<td>0.128</td>
<td>0.3374</td>
<td>Yes</td>
</tr>
<tr>
<td>89</td>
<td>TM4SF5</td>
<td>1193</td>
<td>0.126</td>
<td>0.3363</td>
<td>Yes</td>
</tr>
<tr>
<td>90</td>
<td>PAG1</td>
<td>1199</td>
<td>0.125</td>
<td>0.3391</td>
<td>Yes</td>
</tr>
<tr>
<td>91</td>
<td>LCK</td>
<td>1272</td>
<td>0.12</td>
<td>0.3319</td>
<td>Yes</td>
</tr>
<tr>
<td>92</td>
<td>CLDN3</td>
<td>1273</td>
<td>0.12</td>
<td>0.3352</td>
<td>Yes</td>
</tr>
<tr>
<td>93</td>
<td>ITGB7</td>
<td>1285</td>
<td>0.119</td>
<td>0.3369</td>
<td>Yes</td>
</tr>
<tr>
<td>94</td>
<td>CLDN23</td>
<td>1350</td>
<td>0.116</td>
<td>0.3308</td>
<td>Yes</td>
</tr>
<tr>
<td>95</td>
<td>KCNK1</td>
<td>1362</td>
<td>0.115</td>
<td>0.3324</td>
<td>Yes</td>
</tr>
<tr>
<td>96</td>
<td>SORBS1</td>
<td>1364</td>
<td>0.115</td>
<td>0.3354</td>
<td>Yes</td>
</tr>
<tr>
<td>97</td>
<td>LDLR</td>
<td>1376</td>
<td>0.115</td>
<td>0.337</td>
<td>Yes</td>
</tr>
<tr>
<td>98</td>
<td>SLC46A1</td>
<td>1378</td>
<td>0.114</td>
<td>0.34</td>
<td>Yes</td>
</tr>
<tr>
<td>99</td>
<td>BTN2A1</td>
<td>1388</td>
<td>0.114</td>
<td>0.3418</td>
<td>Yes</td>
</tr>
<tr>
<td>100</td>
<td>PSMG1</td>
<td>1395</td>
<td>0.114</td>
<td>0.3441</td>
<td>Yes</td>
</tr>
<tr>
<td>101</td>
<td>LYN</td>
<td>1399</td>
<td>0.114</td>
<td>0.3468</td>
<td>Yes</td>
</tr>
<tr>
<td>102</td>
<td>SLC4A11</td>
<td>1466</td>
<td>0.11</td>
<td>0.3402</td>
<td>Yes</td>
</tr>
<tr>
<td>103</td>
<td>CXADR</td>
<td>1489</td>
<td>0.109</td>
<td>0.34</td>
<td>Yes</td>
</tr>
<tr>
<td>104</td>
<td>LAPTMT5</td>
<td>1496</td>
<td>0.109</td>
<td>0.3421</td>
<td>Yes</td>
</tr>
<tr>
<td>105</td>
<td>LYPD3</td>
<td>1505</td>
<td>0.108</td>
<td>0.3439</td>
<td>Yes</td>
</tr>
<tr>
<td>106</td>
<td>ABCB4</td>
<td>1519</td>
<td>0.107</td>
<td>0.345</td>
<td>Yes</td>
</tr>
<tr>
<td>107</td>
<td>ARRB1</td>
<td>1608</td>
<td>0.103</td>
<td>0.335</td>
<td>Yes</td>
</tr>
<tr>
<td>108</td>
<td>CD52</td>
<td>1609</td>
<td>0.103</td>
<td>0.3378</td>
<td>Yes</td>
</tr>
<tr>
<td>109</td>
<td>CAPRIN1</td>
<td>1617</td>
<td>0.102</td>
<td>0.3397</td>
<td>Yes</td>
</tr>
<tr>
<td>110</td>
<td>STX4</td>
<td>1620</td>
<td>0.102</td>
<td>0.3422</td>
<td>Yes</td>
</tr>
<tr>
<td>111</td>
<td>CLDN11</td>
<td>1638</td>
<td>0.101</td>
<td>0.3425</td>
<td>Yes</td>
</tr>
<tr>
<td>112</td>
<td>ERBB3</td>
<td>1671</td>
<td>0.1</td>
<td>0.3406</td>
<td>Yes</td>
</tr>
<tr>
<td>113</td>
<td>NCSTN</td>
<td>1677</td>
<td>0.1</td>
<td>0.3426</td>
<td>Yes</td>
</tr>
<tr>
<td>114</td>
<td>MAEA</td>
<td>1681</td>
<td>0.1</td>
<td>0.3449</td>
<td>Yes</td>
</tr>
<tr>
<td>115</td>
<td>LAMB1</td>
<td>1691</td>
<td>0.099</td>
<td>0.3464</td>
<td>Yes</td>
</tr>
<tr>
<td>116</td>
<td>OXTR</td>
<td>1697</td>
<td>0.099</td>
<td>0.3484</td>
<td>Yes</td>
</tr>
<tr>
<td>117</td>
<td>PTPRJ</td>
<td>1708</td>
<td>0.098</td>
<td>0.3496</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>118</td>
<td>KL</td>
<td>1732</td>
<td>0.098</td>
<td>0.349</td>
<td>Yes</td>
</tr>
<tr>
<td>119</td>
<td>TMEM11</td>
<td>1750</td>
<td>0.097</td>
<td>0.3492</td>
<td>Yes</td>
</tr>
<tr>
<td>120</td>
<td>PODXL2</td>
<td>1754</td>
<td>0.097</td>
<td>0.3514</td>
<td>Yes</td>
</tr>
<tr>
<td>121</td>
<td>SLC16A3</td>
<td>1764</td>
<td>0.097</td>
<td>0.3527</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Chapter 3

Discussion

Cnot1 was identified as a candidate CRC cancer gene in multiple Sleeping Beauty-based screens. In addition, a high incidence of CNOT1 somatic mutations in CRC has been documented by several projects including the COSMIC database (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/), TCGA (Cancer Genome Atlas Network, 2012) and, Seshagiri et al., 2012, providing evidence for a correlation between genetic alterations in CNOT1 and CRC. To determine the contribution of genetic alterations in CNOT1 to CRC we evaluated cell viability in CNOT1-depleted SW480 CRC cell line and the non-cancerous HEK293 cell line. Depletion of CNOT1 resulted in a dramatic decrease in cell viability in both cell lines (Figure 17), and increased cell apoptosis in SW480 cells (Figure 18), suggesting that CNOT1 might act as an oncogene.

In support of this role, CNOT1 limits expression of TXNIP which has been identified as a tumor suppressor in several cancers, including colorectal cancer. The enforced expression of TXNIP inhibits proliferation of cancer cells and suppresses tumor growth and metastasis in transplantation models (Han et al., 2003, Wang et al., 2002, Goldberg et al., 2003). In addition, activation of TXNIP initiates apoptotic pathways under some cell stress conditions, including extreme ER stress through multiple mechanisms including activation of ASK1-induced apoptotic pathways and maturation of IL-1β (Tobiume et al., 2001, Zhou, et al., 2010). Two results suggest that CNOT1 may protect cell viability by means of limiting TXNIP levels. First, in HEK293 cells depletion of CNOT1 leads to increased TXNIP levels, while overexpression of TXNIP leads to decreased cell viability. Second, in CNOT1-depleted SW480 cells, upregulation of TXNIP precedes increased apoptosis and decreased cell viability (Table 3).

Our data argue that CNOT1 limits TXNIP expression by facilitating post-transcriptional degradation as we found that TXNIP mRNA transcripts were significantly stabilized after CNOT1 knockdown. CNOT1-mediated destabilization likely involves
CNOT8-mediated deadenylation as depletion of CNOT8 caused decreased cell viability and increased expression of TXNIP (Figure 24E). TXNIP mRNA levels are limited by miR-17 in pancreatic beta cells. We have found that miR-17 regulates TXNIP levels in SW480 and HEK293 cells as well. Given that CNOT1 binds GW182 and acts as an adapter between the miRNA RISC complex and CCR4-NOT deadenylase subunits, this result suggests that CNOT1 may limit TXNIP expression by facilitating miR-17-mediated TXNIP mRNA degradation.

TXNIP mRNA levels are upregulated in response to extreme ER stress. Increased TXNIP activity can, in turn, initiate pro-apoptotic signaling pathways. Upregulation of TXNIP following depletion of CNOT1 suggests that CNOT1 may be involved in the unfolded protein response to ER stress as well. In support of this idea depletion of CNOT1 in SW480 cells results in upregulated transcripts of genes involved in all three UPR pathways such as sXBP1 (IRE1α pathway), PERK, ATF4 and DDIT3 (PERK pathway), and ATF6, XBP1 (ATF6 pathway) (Figure 30A) (see Figure 16 in introduction describing the three UPR pathways). It is important to note that only changes in mRNA transcripts levels of these genes were examined. To gain a full picture of the effect of CNOT1 depletion on these pathways we will need to examine translational and post-translational regulation.

Moreover, the disrupted deadenylation function of the CNOT complex can cause upregulated mRNA levels of a large group of genes (Aslam et al., 2009). When these mRNAs are translated in the ER, it may cause overproduction of proteins in the ER, inducing ER stress and activation of UPR pathways. This is supported by an Ito et al study showing that depletion of CNOT1 in Hela cells can cause ER stress and activation of UPR pathways (Ito et al., 2011). We examined global gene expression in CNOT1-depleted SW480 cells using a microarray assay. 266 genes showed a significant alteration in mRNA expression levels and 2/3 of them were upregulated, consistent with the idea that CNOT1-mediated degradation plays a major role in these cells and that CNOT1 depletion could result in an enlarged mRNA pool. In support of this idea, GSEA comparison of microarray data with the Molecular Signatures Database geneset collection
of cellular components (MsigDB C5 CC, http://www.broadinstitute.org/gsea/msigdb/collections.jsp#C5) revealed that four of the five most enriched genesets consisted of plasma membrane localized components and thus transcripts that would be translated and processed in the ER (Figure 31B, Table 7), suggesting that CNOT1 may help in limiting protein production in the ER.

Under moderate ER stress the only known target of IRE1 alpha RNase activity is the XBP1 transcript. However, under extreme ER stress IRE1 alpha RNase becomes less specific. This is thought to lead to degradation of miR-17 and thus upregulation of TXNIP (see introduction for details). In addition, under extreme stress IRE1 alpha RNase activity degrades a pool of RNAs in a process known as Regulated IRE1 alpha Dependent Decay (RIDD). Decay of these mRNAs is a well-characterized part of the terminal UPR. GSEA of global gene expression changes revealed that transcripts downregulated in CNOT1-depleted SW480 cells were enriched in genes downregulated in response to extreme ER stress (RIDD genes), suggesting that downregulation of CNOT1 promotes RIDD and therefore CNOT1 expression is protective against terminal UPR (Figure 31A, Table 6).

Based on these studies we will determine if CNOT1 is required for miR-17-mediated degradation by determining if depletion of CNOT1 can ablate the effect of overexpression of miR-17.

In sum, these results suggest a model in which CNOT1 protects cell viability by protecting cells against ER stress by raising the bar for activation of the apoptotic UPR pathways both by limiting TXNIP expression and possibly by regulating expression of other pools of mRNAs. Cancer cells are commonly subject to increased levels of ER stress. Thus increased CNOT1 activity may provide a selective advantage to cancer cells by making it more difficult to activate the UPR apoptotic pathway under high stress conditions, such as severe ER stress. Thus, this model suggests a mechanism by which CNOT1 may act as an oncogene.

Although our data are supportive of this model, several key experiments remain to be done to fill the gaps. (1) To determine if increased TXNIP activity contributes to
increased apoptosis in CNOT1-depleted cells we will need to examine the effect of co-depletion of CNOT1 and TXNIP on cell viability and apoptosis. If CNOT1-depletion promotes apoptosis by increasing TXNIP activity, then apoptosis will be reduced when TXNIP is depleted together with CNOT1 depletion. (2) Several experiments will be conducted to further investigate the mechanism of CNOT1-mediated destabilization of TXNIP mRNA transcripts. mRNA stability is commonly regulated through interactions of complexes containing 3’UTR binding factors, adapters and deadenylases with the 3’UTR region of the target mRNA transcript. To determine if CNOT1 regulates TXNIP mRNA stability through the TXNIP 3’UTR we will examine the effect of CNOT1 depletion on expression of a luciferase reporter construct containing the 3’UTR of TXNIP. To determine if miR-17-mediated regulation of TXNIP involves CNOT1 we will co-transfect SW480 cells with both miR-17 mimics and CNOT1 siRNA to investigate whether depletion of CNOT1 reduces the effect of overexpression of miR-17 on TXNIP mRNA levels. (3) Finally to determine if CNOT1 is protective against severe ER stress we will look at the response to induction of ER stress on cells depleted for or overexpressing CNOT1 to determine if CNOT1 depletion results in increased cell death and overexpression is protective and improving cell viability under severe ER stress.

Although these in vitro studies point to a role for CNOT1 as an oncogene in CRC, the actual role in in vivo cancers is likely to be more complex. As the scaffold for the CCR4-NOT complex CNOT1 is likely involved in most or all functions of the complex in addition to deadenylation, such as transcriptional regulation, chromatin modification and ubiquitin-mediated proteolysis (Bartlam et al., 2010). In addition, as an adapter mediating mRNA degradation by several 3’UTR binding factors, including miRNAs and RNA binding proteins, CNOT1 can potentially influence expression of a large group of genes, which is supported by the microarray data. Therefore, the roles of CNOT1 in CRC are likely to depend on the context of a specific tumor, including other genetic alterations, stage of tumor development and interactions between the cancer and its microenvironment.
Our data suggest that CNOT1 interaction with additional CNOT subunits also plays a role in regulating TXNIP levels. The shortest isoform of CNOT1, CNOT1S, lacks a NOT domain that is important for interaction with CNOT 2, CNOT3 and CNOT5 (Boland et al., 2013). Unlike the long isoform of CNOT1 (CNOT1L) overexpression of CNOT1S results in increased expression of TXNIP mRNA levels, suggesting that one or more of these subunits may also be important to TXNIP regulation by CNOT1.

In addition, depletion of CNOT 6L and 7 results in decreased cell viability but little or no increase in TXNIP levels, indicating that the CNOT complex can support cell viability by mechanisms other than TXNIP regulation.

Finally, in SB screens insertions into CNOT1 are found in both orientations (Figure 11). This pattern is often associated with inactivation of a tumor suppressor gene. However, it could indicate that either activation or inactivation of a specific gene could contribute to development of CRC depending on site of insertion and the context of other genetic alterations. Two pieces of data from our work suggest that this might be the case. First, overexpression of CNOT1S has the opposite effect as overexpression of CNOT1L (Figure 24C), suggesting that the specific domains affected by a genetic alteration will influence the effect of CNOT1 on gene regulation. Second, analysis of human CRC liver metastases reveals both CNOT1 high expressers and CNOT1 low expressers, with different gene expression profiles (Figure 21), suggesting that genetic context will influence which alterations in CNOT1 provide selective advantage to cancer cells.

Together, although our in vitro studies suggest a role for CNOT1 as an oncogene in CRC through protecting cells from severe ER stress-induced cell apoptosis via limiting TXNIP levels and limiting mRNA transcripts to be translated in the ER, we should note that CNOT1 might contribute to CRC through additional mechanisms. The following experiments are designed to further investigate the role of genetic alteration in CNOT1 in CRC. (1) Perform in vivo studies in a mouse model carrying a conditional knockout allele of CNOT1 in wildtype or ApcMin backgrounds to investigate the effect of CNOT1 deficiency on intestinal tumor phenotypes. (2) Evaluate the effect of somatic mutations in CNOT1 on oncogenic phenotypes, assembly of the CNOT complex and deadenylation of
target mRNA transcripts in in vitro studies. Furthermore, we may carry out in vivo studies by producing transgenic mice bearing the CNOT1 mutations which show strong phenotypes in cell line based studies. (3) Evaluate the role of CNOT1 in maintenance of colonic epithelial stem cells. The CNOT complex has been reported to influence other aspects of cancer development. For example, Cnot1, Cnot2 and Cnot3 were reported to play critical roles in maintaining mouse and human embryonic stem cell (ESC) identity and inhibiting ESC differentiation into the extraembryonic lineages via repressing the expression of early trophectoderm transcription factors such as Cdx2 (Zheng et al., 2012). CRC is thought to arise from stem cells that lie in the base of colonic crypts. Factors that affect normal stem cell activity also affect development of cancer stem cells so that the role of CNOT1 and other members of the CCR4-NOT complex in ES cells may also reflect a role in CRC. In particular, the unfolded protein response is required for differentiation of progenitor cells from stem cells in the normal colon epithelium (Heijmans et al., 2013), suggesting that factors that limit the unfolded protein response may support stem cell activity. Organoid culture can be used to model stem cell activity and study differentiation of colonic crypts. Therefore we will investigate the role of CNOT1 and the CNOT complex in stem cell activity using this system.
Chapter 4

Material and Methods

Reagents

Tunicamycin and Actinomycin D were obtained from Sigma Aldrich (St. Louis, MO, USA). Stocks were prepared in DMSO at 1mg/ml and stored at -20°C. Dimethyl sulfoxide (DMSO), Dithiothreitol (DTT) and β-Mercaptoethanol (β-ME) were purchased from Sigma Chemical Co. (St. Louis, MO, USA). RNasin was obtained from Promega (Madison, WI, USA).

Cell Culture

SW480 and HEK 293 cells were purchased from American Type Culture Collection (ATCC). Cells were maintained in DMEM supplemented with 10% FBS, 2 mmol/L glutamine, and incubated in a humidified atmosphere of 95% air and 5% CO₂ at 37°C.

RNA Isolation

**RNA isolation from cell culture.** Total RNA was extracted from cell cultures using an RNeasy kit (Qiagen) according to manufacturer’s directions for isolation of cytoplasmic RNA with an added DNA digestion step. Cells growing in a monolayer on culture dishes were harvested using Trypsin-EDTA 0.25% at 37°C and centrifuged at 300 x g for 3 minutes. The plasma membrane was lysed by addition of 175 µl ice cold RLN buffer (50 mM Tris-HCl pH 8.0, 140 mM NaCl, 1.5 mM MgCl₂, 0.5% (v/v) NP-40, DEPC H2O) supplemented with 1000 U/ml RNasin and 1 mM DTT on ice for 5 minutes. Lysate was centrifuged at 300 x g for 2 minutes at 4°C to remove nuclei and debris, followed by addition of RLT (guanidine thiocyanate) buffer supplemented with 10 U/ml β-mercaptoethanol and 100% ethanol. Sample was applied to column and washed with RW1 buffer. DNase digestion was carried by incubation with 240 unit of RNase free
DNase (Qiagen) for 15 minutes at room temperature. After washes with RPE buffer, RNA was eluted with 30 µl elution buffer. RNA was stored at -80°C.

**RNA isolation from human fresh frozen tissue.** Samples obtained during resection of human CRC liver metastases or CRC were flash frozen in liquid nitrogen and processed by the University of Minnesota Biological Procurement Network (BioNet). Total RNA was isolated from 20 fresh frozen CRC liver metastasis samples and 5 fresh frozen normal colon tissues. Tissues were transferred from liquid nitrogen to precooled RINAlater-ICE (Ambion, Foster City, CA, USA) at -80°C for 30 minutes followed by transfer to -20°C for at least 48 hours to inactivate endogenous RNase activities. Immediately before RNA isolation, tissue in RINAlaterICE was placed in 3 ml of RLT buffer supplemented with β-mercaptoethanol at 10 µl/ml. The tissue was then homogenized using an IKA ultra-Turrax T25 digital homogenizer (Fisher Scientific, Waltham, MA, USA). Tissue lysates were centrifuged at 300 x g for 5 minutes and supernatants were transferred to new 1.5 ml tubes. RNA isolation was then conducted using an RNeasy Mini Kit (Qiagen) following the manufacturer’s directions with an added DNase digestion step starting after the RLT addition step. RNA was stored at -80°C.

**Plasmid Construction**

Plasmids encoding various open reading frames were obtained from Open Biosystems (Thermo Scientific).

To construct a plasmid expressing the short form of CNOT1 (encoded by NM_206999.2) Open Biosystems clone #5266600 encoding CNOT1S ORF (BC040523) in a pBluescriptR vector was digested with NotI and ApaI and inserted between the corresponding restriction sites on a pcDNA3 vector (Life Technologies) to generate pcDNA3-CNOT1S. To construct a plasmid expressing the longest form of CNOT1 (CNOT1L) (encoded by NM_016284.4) pcDNA3-CNOT1S was combined with a fragment obtained from OpenBiosystems clone #3913601 encoding a partial CNOT1L ORF (BC0234317) in a pCMV-SPORT6 vector. To add the 3’ region of CNOT1L to CNOT1S to generate a full length CNOT1L plasmid both plasmids were digested with
Cla1 and Apa1 and the Cla1-Apa1 fragment from the partial CNOT1L plasmid was inserted into the pcDNA3-CNOT1S plasmid. To construct a plasmid expressing TXNIP (NM_006472.4) Open Biosystems clone #7939547 encoding the TXNIP ORF (BC093702) in a pCR4-TOPO vector was digested with EcoR1 and inserted in the EcoR1 restriction sites of pcDNA3 vector to generate pcDNA3-TXNIP. Orientation was determined by Sanger sequencing. Plasmids were prepared using a QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer’s protocol.

**DNA and RNA Quantification and Integrity Analysis**

DNA and RNA quantity and purity was analyzed using a Nanodrop 2000 spectrophotometer (NanoDrop, ThermoScientific) to measure absorbance (A) at a range of concentrations. DNA concentration = $A_{260} \times 50 \mu g/\mu l$. RNA concentration = $A_{260} \times 40 \mu g/\mu l$. The ratio of the $A_{260}$ nm/$A_{280}$ nm values allows the assessment of nucleic acid purity. Pure preparations of DNA will have a 260/280 ratio > 1.8. RNA samples are considered to be pure if the absorbance ratio is close to 2.0. For Illumina gene expression array analysis and Illumina RNA sequence analysis, RNA integrity was assessed by capillary electrophoresis using an Agilent BioAnalyzer 2100 (Agilent Technologies). An RNA integrity number (RIN) was generated for each sample representing 28S/18S ribosomal RNA detection as well as other parameters. All the samples met the standard of quality with a RIN > 8.

**qRT-PCR**

**cDNA synthesis.** For each experimental sample 1.5 µg RNA was converted to cDNA with random nonomer primers (IDT) and recombinant Omniscript Reverse Transcriptase using the Omniscript RT kit (Qiagen) according to manufacturer’s instructions. Each reaction consisted of 2 µl of 10 X RT buffer, 2 µl of 50 mM dNTPs, 4 µl of random nonomer primers, 0.25 µl of RNasin RNase inhibitor (Promega) and 1 µl of Omniscript Reverse Transcriptase, 1.5 µg RNA and H2O sufficient for final volume of 20 µl. Reactions were incubated at 65°C for 60 minutes, 72°C for 10 minutes, 85°C for 5 minutes. cDNA was stored at -20°C until needed.
**qPCR.** Target cDNAs and 18S ribosomal DNA were amplified from cDNA by PCR. Real-time PCR reactions were carried out in 10 µl using 2 µl cDNA from a 10 X dilution of the 20 µl RT reaction (equivalent to 15 ng reverse transcribed RNA) for target genes and a 100 X dilution for 18S rRNA, 500 nM each primer, 5 µl LightCycler480 SYBR Green 1 Master Mix (Roche Applied Science). PCRs were performed on a LightCycler® 480 System (Roche Diagnostics) in 96 well plates using the amplification protocol: 1 cycle of preincubation, 5 minutes at 95°C; 45 cycles of amplification each consisting of denaturation at 95°C for 5 seconds, annealing at 60°C for 5 seconds and elongation at 72°C for 10 seconds; 1 cycle melting at 95°C for 5 seconds, 65°C for 1 minutes, heating to 97°C; 1 cycle cooling at 40°C for 30 seconds. Water was used as a template for negative control amplifications for each PCR run. All reactions were performed in duplicate. Standards were generated by reverse transcription of total RNA from untreated cells followed by PCR amplification to generate template DNA of the same sequence as target genes or 18S gene PCR product. Serial dilutions of template DNA were amplified in parallel with experimental samples and used to generate a standard curve for each gene. Data were analyzed using Roche LightCycler® 480 software and CP’s calculated using the Absolute Quantification-Second Derivative Maximum method. The standard curve was used to determine efficiency of PCR amplification (E) for each gene. “Relative mRNA levels” represent the relative expression of a target gene in corresponding siRNA-treated cells to expression in control siRNA-treated cells. Target gene expression was normalized to 18S rRNA levels and relative mRNA levels were calculated as described by Pfaffl (Pfaffl M.W., 2001): 

\[
\text{relative mRNA levels} = \frac{(\Delta C_{\text{target (control-treated)}}^{\text{E target}})}{(\Delta C_{\text{ref (control-treated)}}^{\text{E ref}})}
\]

Primers used in RT-PCR are shown below (Table 8).

**Table 8. Primers used in qRT-PCR.**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>18s</td>
<td>CGCCGCTAGAGGTGAAATTCTT</td>
<td>CAGTCGGCATCGTTATGGTC</td>
</tr>
<tr>
<td>AFT6</td>
<td>ATGTCTCCCCTTTCCTTATGGT</td>
<td>AAGGCTTGGGCTGAATGGAA</td>
</tr>
<tr>
<td>ATF4</td>
<td>CCACCATGGCGATTAGGGG</td>
<td>TGCTGAATGCCGTGAGAAGC</td>
</tr>
<tr>
<td>BTG1</td>
<td>TCTCCAAGGTTTCTCCGCACC</td>
<td>GTAAACCCGATCCCTTGCAT</td>
</tr>
</tbody>
</table>
CDX2    ACTACAGTCGCTACATCACCA    GAAGACACCGGACTCAAGGG
CHOP    GGTGGCAGCGACAGAGCAAA    ACTCAGTGCCATCTCTGAGTT
CNOT1   CTTTCA ACCCCAATACAGACC    AGGGTTTCATCTTACTCTGCTGGA
CNOT1 long   GAGGCTCCTCAAAGCTCTCG    TGTGCATACATAGGGGGAAGT
CNOT1 total  TCCCTAGTGAAATAAGGACCC    CAGGACCTGAGCAAGG
CNOT2    CTTCCTACACACATATTGGATA    GGTTAGCCATATGAGGGAAGT
CNOT6    GGTTGGCAGCGACAGAGCAAA    ACCACACCTGGAAACTCGGTGC
CNOT6L   TCCCTGAAGATGACAGGATCAGG   CAGGACCTGAGCAAGG
CNOT7    CCCACACTATGCGACAGG    ACCGACACGACGACGACG
CNOT8    TCCTTCA ACCCCAATACAGACC    CCACACCTGGAAACTCGGTGC
CPEL1    GCCGAGATCTGTGTTGGG    ACCGACGAGGGGCCAGG
DKK1    GGGGTTCTCAGTGGTCTGCTG    CAGCAGGAGGAGGGCCAGG
EMP1    ACCGGAAGAAGGAAACCG    TGGGTTGTCATGACACCTCCCA
GRP78/Bip  AAGCTTCTGACTGATGCTCT    GGGGAGAAGTGAGGCCAG
HBP1    CAGCCGAGCCTCTACTGGCT    GAGAGGCTCGGGCAGGG
IGFBP3   TAAAGACAGGGCGCGCTACA    TCCCTCTACCGGAGGAGG
IL8    TTTTGCCAAGAGGTGCTAAAGA    ACCCCTGCACTCAGTTCTC
p27kip    TAGCGGAGAACATGCGG    ACCGAGATTTGGGAAACCG
PERK    ATGGAACAGAGGGCGCAGG    CGGATGGGCAGCTCGCTG
PUM1    ACCGATTCAGGCGCCAGTCC    GGGACATTTGGGAAACCG
PUM2    AGCTTCCTGAGATTTACACG    CAGACAGGACAGGACAGG
sXBP1   TTAGCAAGCCGCACAGG    TCCCTAGTGGATTTGGACCA
TGFBR2   CCCCTGTCAGAAGGCGATG    TGCACACACACCATGGCC
ZFP36    TCCACTCTCGGCCGACACC    GATGGCAGGGCAGCTCGG
TXNIP    TCCACTCTCGGGCGACACC    GATGGCAGGGCAGCTCGG
XBP1    GGGGAATGAGTGAGGCCAG    TTGCACAGATGCCCAAACGG

Transfection Protocols

DNA plasmid transfection. HEK 293 cells were plated at 7.5 x 10^5 cells/dish into 6 cm dishes or 2 x 10^4 cells/well in 96 well plates. 24 hours later, cells were transfected with pcDNA3-CNOT1S, pcDNA3-CNOT1L, pcDNA3-TXNIP or a pcDNA3-GFP plasmid using Lipofectamine LTX with Plus reagent (Invitrogen) according to the manufacturer’s protocol. For a 6 cm dish, 4 µg of plasmid was added to 4 µl PLUS reagent in 200 µl of OptiMEM. The plasmid solution was mixed with 12 µl Lipofectamine LTX reagent in 200 µl OptiMEM. The 400 µl reaction was then added to the cells.
**siRNA transfection.** SW480 cells and HEK 293 cells at 70% confluence were transfected twice, unless otherwise noted, 48 hours apart, with short interfering RNA oligonucleotides (siRNA oligos) targeting a specific gene or with a non-targeting control siRNA, at a final concentration of 50 nmol/L, using Lipofectamine 2000 transfection reagent (Life Technologies). Oligo targeting *CNOT1* was obtained from Dharmacon: *CNOT1* On-TARGETplus SMARTPool. Control siRNA oligo was also obtained from Dharmacon: OnTARGETplus Non-targeting siRNA # 3. siRNA oligos targeting other genes were obtained from Qiagen: *Hs_CDX2_6_HP* siRNA, *Hs_CNOT2_10_HP* siRNA, *Hs_CNOT6_3_HP* siRNA, *Hs_CNOT6L_4_HP* siRNA, *Hs_CNOT7_4_HP* siRNA, *Hs_CNOT8_5_HP* siRNA, *Hs_CELF1_1_HP* siRNA, *Hs_PUM1_6_HP* siRNA, *Hs_PUM2_4_HP* siRNA, *Hs_ZFP36_6_HP* siRNA, *Hs_TXNIP_6_HP* siRNA.

**Transfection of miR-17 mimic.** SW480 cells and HEK 293 cells at 70% confluence were transfected once with an miR-17 mimic (Qiagen, Syn-hsa-miR-17-5p miScript miRNA Mimic) or with a non-targeting control siRNA (Dharmacon, OnTARGETplus Non-targeting siRNA # 3) at a final concentration of 5 nmol/L, using Lipofectamine 2000 transfection reagent (Life Technologies).

**Illumina Beadchip Microarray**

SW480 cells were transfected twice, 48 hours apart, with *CNOT1* siRNA or with a non-targeting control siRNA. Three experiments were done to achieve three biological replicates for each treatment. Total RNA was extracted from cells as described. Six samples of total RNA, 2 µg per sample, were submitted to the University of Minnesota BioMedical Genomics Center for microarray analysis. All samples had RIN > 8. 1 µg of total RNA was converted to biotinylated, antisense cRNA using the Illumina TotalPrep-96 RNA Amplification Kit (Life Technologies), and 1500 ng of biotin-labeled cRNA was hybridized onto an Illumina HumanHT-12 v4 Expression BeadChip (Illumina). Hybridized Beadchips were scanned with an Illumina iScan scanner. Raw intensity data was extracted from iScan scan image files using GenomeStudio software (Illumina), log₂ transformed, and normalized. Microarray raw data was analyzed by Dr. Ying Zhang at the University of Minnesota Supercomputing Institute using the R package: *lumi* and
limma according to the manufacturer’s protocol. Differential gene expression between CNOT1 siRNA and control siRNA transfected cells was assessed by a statistical linear model analysis using the bioconductor package limma, in which an empirical Bayes method was used to moderate the standard errors of the estimated log-fold changes of gene expression. The moderated t-statistic p values derived from the limma analysis were further adjusted for multiple testing by Benjamini and Hochberg’s method to control false discovery rate (FDR).

**Illumina RNA Sequencing**

Total RNA isolated from fresh frozen human CRC liver metastasis and normal colon samples was sequenced by the University of Minnesota BioMedical Genomics Center (BMGC) using an Illumina Highseq 2000 sequencer (Illumina). 3 µg of each total RNA sample was sent to BMGC in a 96 well plate on dry ice. Samples were quantified using fluorimetry (RiboGreen assay) and RIN determined. For the library creation, total RNA was enriched for polyadenylated RNA using oligo-dT purification, and the enriched polyA(+) RNA was converted to cDNA through reverse transcription. The cDNA was then fragmented, blunt-ended, and ligated to indexed (barcoded) adaptors. The fragmented cDNA library was size selected for 190-210 base-pair fractions, and size distribution was validated using capillary electrophoresis. The cDNA fragments were quantified using a fluorimetry (PicoGreen). The indexed libraries were then normalized, pooled, clustered on a flow cell, and loaded onto the Illumina Highseq 2000 instrument for paired-end sequencing. The resulting reads were then subjected to further analysis. Gene expression levels were calculated as fragments per kilobase of exon per million fragments mapped (FPKM).

**Global Gene Expression Analysis**

**Differential gene expression heatmap.** Differentially expressed genes showing fold change ≥ 2 and p value < 0.05 were used to generate a heat map by unsupervised hierarchical clustering. Extent of differential expression is shown as Z-value.

**Ingenuity Pathways Analysis (IPA).** To identify the gene networks and cellular
pathways associated with CNOT1 knockdown, the dataset consisting of 266 differentially expressed genes (fold change ≥ 2, p < 0.05, Table 4) identified by Illumina HumanHT-12 v4 Expression Beadchips analysis were used as input for network analysis using Ingenuity Pathway Analysis software (Ingenuity Systems, Redwood CA). For this type of analysis, differentially expressed genes, which interact with other molecules in the Ingenuity Knowledge Base are designated “Network Eligible” molecules. “Network eligible” molecules are combined into networks that maximize connections based on known biological attributes. Functional categories are assigned based on overlap with known function categories in Ingenuity Database. Networks are then scored based on the percent of dataset molecules they contain. The score is calculated using the Fishers Exact Test to determine the likelihood of the network containing a given number of Network Eligible molecules by chance. The score = -log (Fisher's Exact test result).

**Gene Set Enrichment Analysis (GSEA).** GSEA was performed using GSEA v.2.0, (Broad Institute, MIT [http://www.broad.mit.edu/gsea/](http://www.broad.mit.edu/gsea/)). To identify pathways altered in human liver CRC metastasis samples expressing high levels of CNOT1 a complete ranked dataset was generated in which genes were ranked according to ratio of expression in CNOT1 high expressing (CNOT1-Hi) to CNOT1 low expressing (CNOT1-Lo) samples. The ranked dataset was compared to genesets derived from Merlos-Suárez et al. characterizing gene expression in sub-populations of the intestinal crypt epithelium (Merlos-Suárez et al., 2011). The complete group of genesets tested consisted of: ISC Signature (EphB2), Proliferation Signature, Late TA Signature and Lgr5-defined Signature. This dataset was also compared to genesets of candidate CRC cancer genes identified by SB screens. These genesets were derived from reports in Starr et al. papers (Starr et al., 2009, Starr et al., 2011, Starr T.K., unpublished reports of a p53 mutant screen) as well as a report from March et al. (March et al., 2011). To identify pathways altered in CNOT1-depleted SW480 cells a complete ranked dataset was generated in which genes were ranked according to ratio of expression in CNOT1-depleted to control treated cells. This dataset was compared to genesets characterizing gene expression in adaptive vs. terminal UPR. The complete group consisted of Han et al. paper supplemental table 1 (representing both adaptive and terminal responses) and
supplemental table 2 (representing the terminal response) (Han et al., 2009), and So et al. paper supplementary table 1 (So et al., 2012). This table was split into two genesets, one consisting of adaptive response genes activated by sXBP1 and a separate geneset consisting of RIDD genes. This dataset was also compared to the MSigDB C5 CC collection. In all the comparisons, normalized gene expression data was ranked using the Signal2Noise metric, which represents a Z value. An enrichment score (ES) was calculated that reflects the degree to which a geneset is overrepresented at the extremes (top or bottom) of the entire ranked differentially expressed geneset using a running-sum statistic. The statistical significance (nominal p value) of the ES was estimated by using an empirical phenotype-based permutation test procedure. GSEA was performed against 1000 random gene set permutations. The ES for each gene set was then normalized to account for the size of the set, yielding a normalized enrichment score (NES). The false positives were controlled by calculating the false discovery rate (FDR) corresponding to each NES, which is the estimated probability that a set with a given NES represents a false positive finding. The proportion of false positives was controlled for by calculating the false discovery rate (FDR) corresponding to each NES.

**Assessment of Cell Function**

**Cell viability assay.** Cell viability was determined using a 3-(4,5-dimethylthiazol-2-yl)-2,5 diphenyltetrazolium bromide (MTT) assay (Cell Viability Kit 1, Roche Applied Sciences) in which absorbance at 595 nm is proportional to viable cell number. Cells were plated into 96 well plates at indicated cell densities in triplicate in DMEM medium. At each time point 100 µl fresh media was added with 10 µl MTT reagent. After 4 hour incubation, 100 µl of lysis buffer was added. After overnight incubation, absorbance at 595 nm was read using a plate reader. Cell viability was determined on days 2-6 after the second siRNA transfection or 1-4 days after plasmids transfection unless otherwise indicated.

**Apoptosis assay.** Apoptosis was determined using an Annexin V fluorescein isothiocyanate (AV-FITC) and propidium iodide (PI) double staining flow cytometry assay. AV-FITC binds to phosphatidylserine which translocates to the outer layer of the
plasma membrane lipid bilayer during the early stages of apoptosis. PI is a vital stain, which can only enter the cell when the membrane loses integrity during late apoptosis and cell death. Thus AV-FITC- PI- staining identifies live healthy cells; AV-FITC+ PI- identifies early apoptotic cells; and AV-FITC+ PI+ identifies late apoptotic and dying cells. For this assay 1 X 10^6 cells were washed twice with cold PBS and resuspended in 400 µl of 1 X Annexin V binding buffer (10 mM HEPES, pH 7.4, 140 mM NaOH, 2.5 mM CaCl₂). Suspension was divided into 4 x 100 µl aliquots to be stained with: no dye, 5 µl of Annexin V–FITC, 5 µl of PI, and 5 µl of Annexin V–FITC plus 5 µl of PI, respectively. Cells were stained for 15 minutes at room temperature in the dark. Then, flow cytometric analysis was conducted using a Becton-Dickinson FACScan cytofluorometry (Becton-Dickinson) and data analysis were performed with FlowJo software (Tree Star, Inc.).

**mRNA stability assay.** SW480 cells were plated at 7.5 X 10^5/dish in 6 cm dishes and transfected once with CNOT1 siRNA or with a non-targeting control siRNA. 48 hours after transfection, cells were treated with actinomycin D at 5 µg/ml final concentration and total RNA was extracted at indicated intervals from 0 to 120 minutes. TXNIP and CNOT1 mRNA levels were then measured by qRT-PCR. Values represent mean ± SD of three independent experiments. The best-fit lines, determined by least-squares methods, are shown as dotted lines.

**Analysis of unfolded protein response (UPR).** Tunicamycin is a pharmacological agent which blocks the first step in the synthesis of N-linked glycoproteins and thus causes accumulation of unprocessed proteins in the ER and induction of the UPR. In order to study factors affecting the UPR, SW480 cells were treated with tunicamycin under conditions described in results.

**Statistics**

All data were analyzed by comparing means with Student’s *t*-test or One-Way ANOVA method as indicated using SPSS (Version 19.0). Data is shown as mean ± SD and p < 0.05 is considered as statistically significant difference.
Bibliography


Andreas Boland, Ying Chen, Tobias Raisch, Stefanie Jonas, DuyguKuzuoğlu-Öztürk, Lara Wohlbold, Oliver Weichenrieder, Elisa Izaurralde. Structure and assembly of the NOT module of the human CCR4-NOT complex. nature structural & molecular biology advance online publication Oct.13 2013.


Clevers, H. Wnt/beta-Catenin signaling in development and disease. Cell 127, 469–480 (2006).


Duerkop BA, Vaishnava S, Hooper LV. Immune responses to the microbiota at the intestinal mucosal surface. *Immunity* 2009; 31: 368–376.


Körner CG, Worthington M, Muckenthaler M, Schneider S, Dehlin E, Wahle E. The deadenylating nuclease (DAN) is involved in poly(A) tail removal during the meiotic maturation of Xenopus oocytes. EMBO J. 17, 5427-5437 (1998).


Roche-Lestienne C, Soenen-Cornu V, Grardel-Duflos N, Laï JL, Philippe N, Facon T, Fenaux P, Preudhomme C. Several types of mutations of the Abl gene can be found in chronic myeloid
leukemia patients resistant to STI571, and they can pre-exist to the onset of treatment. *Blood.* 2002 Aug 1;100(3):1014-8.


Russell P, Benson JD, Denis CL. Characterization of mutations in NOT2 indicates that it plays an important role in maintaining the integrity of the CCR4-NOT complex. *J Mol Biol.* 2002 Sep 6;322(1):27-39.


Wu L, Fan J, Belasco JG. MicroRNAs direct rapid deadenylation of mRNA. *Proc Natl Acad Sci USA* 2006; 103:4034-9.


A Sleeping Beauty transposon-mediated screen identifies murine susceptibility genes for adenomatous polyposis coli (Apc)-dependent intestinal tumorigenesis

Timothy K. Starr\textsuperscript{a}, Patricia M. Scott\textsuperscript{b}, Benjamin M. Marsh\textsuperscript{b}, Lei Zhao\textsuperscript{b}, Bich L. N. Than\textsuperscript{b}, M. Gerard O’Sullivan\textsuperscript{a,c}, Aaron L. Sarver\textsuperscript{d}, Adam J. Dupuy\textsuperscript{e}, David A. Largaespada\textsuperscript{a}, and Robert T. Cormier\textsuperscript{b}

\textsuperscript{a}Department of Genetics, Cell Biology and Development, Center for Genome Engineering, Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455; \textsuperscript{b}Department of Biochemistry and Molecular Biology, University of Minnesota Medical School, Duluth, MN 55812; \textsuperscript{c}Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, St. Paul, MN 55108; \textsuperscript{d}Department of Biostatistics and Informatics, Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455; and \textsuperscript{e}Department of Anatomy and Cell Biology, University of Iowa, Iowa City, IA 52242


Abstract

It is proposed that a progressive series of mutations and epigenetic events leads to human colorectal cancer (CRC) and metastasis. Furthermore, data from resequencing of the coding regions of human CRC suggests that a relatively large number of mutations occur in individual human CRC, most at low frequency. The functional role of these low-frequency mutations in CRC, and specifically how they may cooperate with high-frequency mutations, is not well understood. One of the most common rate-limiting mutations in human CRC occurs in the adenomatous polyposis coli (APC) gene. To identify mutations that cooperate with mutant \textit{APC}, we performed a forward genetic screen in mice carrying a mutant allele of \textit{Apc} (\textit{Apc\textsubscript{Min}}) using Sleeping Beauty (SB) transposon-mediated mutagenesis. \textit{Apc\textsubscript{Min}} SB-mutagenized mice developed three times as many polyps as mice with the \textit{Apc\textsubscript{Min}} allele alone. Analysis of transposon common
insertion sites (CIS) identified the Apc locus as a major target of SB-induced mutagenesis, suggesting that SB insertions provide an efficient route to biallelic Apc inactivation. We also identified an additional 32 CIS genes/loci that may represent modifiers of the Apc\textsuperscript{Min} phenotype. Five CIS genes tested for their role in proliferation caused a significant change in cell viability when message levels were reduced in human CRC cells. These findings demonstrate the utility of using transposon mutagenesis to identify low-frequency and cooperating cancer genes; this approach will aid in the development of combinatorial therapies targeting this deadly disease.

Human colorectal cancers (CRC) generally can be divided into two classes based on whether they display chromosomal instability (CIN) or microsatellite instability (MSI). The majority of CRC (~80–90%) have a CIN phenotype; the remaining cases are characterized by MSI (Grady and Markowitz 2002). CRC displaying CIN frequently harbor allelic losses or mutations in adenomatous polyposis coli (APC), v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS), SMAD family member 4 (SMAD4), and tumor protein p53 (TP53), whereas MSI-type CRC usually have a mutation in one of six DNA mismatch repair genes (Takayama et al. 2007). In both CIN and MSI CRC complete functional loss of a gatekeeper tumor suppressor gene typically is the rate-limiting event in intestinal cell transformation. For CIN CRC, APC plays the key gatekeeping role, and its loss underlies the great majority of CIN CRC and >80% of all CRC. Although both classes of CRC are characterized by high-frequency mutations, such as those in APC, it is evident that many more low-frequency mutations are required for CRC development, and the majority of these low-frequency mutations are unknown (Wood et al., 2007).

To identify these low-frequency mutations, we performed a forward genetic screen in mice using the Sleeping Beauty (SB) DNA transposon as a mutagen in intestinal epithelial cells. To focus on mutations that contribute to the CIN phenotype, we conducted the screen in mice carrying the Apc\textsuperscript{Min} allele. Apc\textsuperscript{Min} mice harbor a T→A nonsense mutation in the Apc gene (Moser et al., 1990, Su et al., 1992) that results in a truncated protein product that is unable to bind β-catenin and promote its degradation,
thus leading to abnormal levels of β-catenin protein and up-regulation of β-catenin target genes such as cyclin D1 (Ccnld1) and myelocytomatosis oncogene (C-Myc). The Min mutation corresponds to a mutational hotspot in the human APC ortholog, and these mutations similarly result in dysregulation of the WNT/β-catenin signaling pathway. There is strong evidence that β-catenin dysregulation is a common transformative event in tumorigenesis in the Apc<sup>Min</sup> mouse and in both the inherited form of APC-deficient CRC (familial adenomatous polyposis, FAP) and in sporadic CRC (Fodde R 2002). Thus, the Apc<sup>Min</sup> mouse is an informative genetic model for APC-deficient intestinal cancer. Apc<sup>Min</sup> mice on the C57BL/6J background strain rarely survive beyond 120 d and can develop >100 tumors throughout the small and large intestine, with the phenotype dependent on diet, mouse strain, and other environmental factors (Dietrich et al., 1993, Cormier et al., 1997).

As in human CRC patients, loss of heterozygosity (LOH) leading to inactivation of both alleles of Apc is necessary for tumorigenesis to commence in Apc<sup>Min</sup> mice (Levy et al., 1994, Luongo et al., 1994). However, in contrast to LOH events in many human CRC, LOH in Apc<sup>Min</sup> tumors occurs predominantly by homologous somatic recombination (Haigis and Dove 2003). In this study we screened for mutations that cooperate with the Apc<sup>Min</sup> mutation by randomly mutating genes through selective activation of SB transposition in intestinal cells of Apc<sup>Min</sup> mice. The results of our screen support the importance of the loss of the second allele of Apc, because the great majority of tumors analyzed contained a transposon insertion in Apc, in particular in tumors in which there was maintenance of heterozygosity (MOH) for the Min allele. In addition to Apc, we identified 32 other genes and loci that probably facilitate the development of intestinal cancer in an Apc<sup>Min</sup> model. The function of these additional mutations could be to remove the requirement for Apc LOH, or they may function in some other manner. The majority of these genes have not been associated with CRC previously. To confirm that these genes play a causal role in tumor development, we used siRNA to knock down message levels of nine of the candidate genes in human colon cancer cell lines and demonstrated that five of these genes affected the growth rate of these cells.
Results

Design of a Forward Genetic Screen for CRC Genes. In a previous study we demonstrated that SB transposon-mediated mutagenesis in the intestinal tract of C57BL/6J Apc\(^{+/+}\) mice resulted in polyp formation (Starr et al., 2009). By mapping transposon insertions in DNA extracted from these tumors, we were able to identify 77 genetic loci which probably harbored genes that, when mutated, contributed to tumor development. Because APC loss is rate limiting in the development of most human CRC (Powell et al., 1992), we reasoned that SB mutagenesis in a mouse already harboring a mutation in Apc might generate more tumors with a shorter latency and reveal mutations that cooperate with Apc during tumor development.

To identify these genes, we performed a forward genetic screen using SB transposon-mediated mutagenesis in Apc\(^{Min}\) mice. The screen consisted of a cohort of Apc\(^{Min}\) SB transgenic test mice along with three groups of Apc\(^{Min}\) control mice. The Apc\(^{Min}\) SB test mice harbored three transgenes required for targeting SB mutagenesis to the gastrointestinal tract (Figure 35). The first transgene was a concatamer of oncogenic transposons (T2/Onc) that were resident on chromosome 1 (Collier et al., 2005). To enhance the mutagenic potential of the transposon, T2/Onc contains a strong viral promoter, splice acceptors in both orientations, and a bidirectional polyA signal. The second transgene was a conditionally expressed knockin SB11 transposase allele downstream of the Rosa26 promoter (Rosa26-LsL-SB11) (Zambrowics et al., 1997, Dupuy et al., 2009). Because of the presence of a floxed stop cassette, the transposase allele is not expressed unless Cre recombinase protein is present. The third transgene was Cre recombinase driven by the gastrointestinal tract-specific Villin promoter (Vil-Cre) (el Marjou et al., 2004). We have shown previously that these three transgenes effectively limit SB mutagenesis to the intestinal tract (Starr et al., 2009). All mice were heterozygous for the Apc\(^{Min}\) allele, and all the transgenes were fully congenic on the C57BL/6J genetic background. The first control group contained Rosa26-LsL-SB11 and either Vil-Cre or T2/Onc; the second control group contained T2/Onc and/or Vil-Cre but
not Rosa26-LsL-SB11; and the third control group harbored only the \( \text{Apc}^{\text{Min}} \) allele. Mice were killed when moribund or at 120 d.

**Intestinal Tumorigenesis Is Enhanced Significantly in \( \text{Apc}^{\text{Min}} \) SB Test Mice.** \( \text{Apc}^{\text{Min}} \) mice that harbored all three transgenes (Rosa26-LsL-SB11, T2/Onc, and Vil-Cre) developed an average of 360 polyps (test mice, Table 9). In contrast, mice carrying the \( \text{Apc}^{\text{Min}} \) allele alone developed an average of 112 polyps (control group 3, Table 9), a result that is consistent with the phenotype of \( \text{Apc}^{\text{Min}} \) mice in our colony (McAlpine *et al.*, 2006). Surprisingly, we also observed an enhanced rate of polyp development in control group 1 that carried the Rosa-26-LsL-SB11 allele but not the complete combination of alleles required for transposition. This control group developed an average of 182 polyps (control group 1, Table 9), a result that was unexpected based on previous screens. It is possible that the increased polyp number in these animals is caused by one or more modifiers linked to the Rosa26-LsL-SB11 transgene, because strain-specific modifiers are known to exist (Fodde and Smits 2001). Control animals carrying the \( \text{Apc}^{\text{Min}} \) allele, T2/Onc, and/or Vil-Cre, but not Rosa26-LsL-SB11 (control group 2, Table 9) developed the same number of polyps as the control mice carrying the \( \text{Apc}^{\text{Min}} \) allele alone. Although the Rosa26-LsL-SB11 allele alone contributes to polyp formation, the effect of active SB transposition was much greater, resulting in twice as many polyps in the test mice. In addition, the tumor burden was so extensive that \( \text{Apc}^{\text{Min}} \) SB test mice became moribund earlier than any of the three control groups (Table 9). Indeed, in a subset of \( \text{Apc}^{\text{Min}} \) SB test mice the tumor load was very severe, with some animals developing as many as 700 tumors.
Table 9. Polyp number and age of death for transgenic mice.

Although polyp number was greatly increased by SB mutagenesis, there was no evidence of local or systemic metastasis in experimental or control mice. We performed histopathologic analysis of tumors collected from 10 animals. These analyses identified numerous microadenomas and adenomas in the small intestine and a much smaller number of these lesions in the large intestines. No adenocarcinomas were identified, perhaps because of the short lifespan of $Apc^{Min}$ SB test mice. Immunohistochemistry for β-catenin was performed on 24 adenomas from seven animals. There was increased expression of β-catenin in all tumors compared with the adjacent normal mucosa epithelium (Figure 34).

Figure 34. A pedunculated adenoma stained with H&E (A) or immunostained for β-catenin (B and C). (B) There is increased staining for β-catenin (arrow) in the adenoma. (C) Higher-power magnification of a different section showing increased cytoplasmic and nuclear (arrows) staining for β-catenin in tumor cells compared with adjacent normal tissue seen in lower right and bottom of picture. (Scale bars: A and B, 500 µm; C, 50 µm.)
Analysis of Common Insertion Sites Identifies 30 Candidate Cancer Genes. To identify genes that contribute to tumor initiation and development in \( Apc^{\min} \) mice, we analyzed transposon insertions in 96 polyps, representing all regions of the intestines, from 12 mice to find common insertion sites (CIS). A CIS is defined by analyzing transposon insertions in many tumors and identifying genomic loci that contain transposon insertions at a higher rate than would be expected by chance (SI Materials and Methods). The presence of a CIS indicates that a transposon-mediated mutation in that locus probably has contributed to tumor development. By analyzing the genes within the CIS, one can identify candidate cancer genes. To map transposon insertions, we isolated DNA from the 96 tumors, digested the DNA with restriction enzymes, and performed ligation-mediated PCR (LM-PCR) to amplify transposon-genomic fragments specifically (Mueller and Wold 1989). Barcodes and fusion sequences were attached to the LM-PCR primers to enable pooling of the amplicons, which then were sequenced using the Roche GS FLX pyrosequencing machine. Six separate sequencing runs produced 347,993 sequence reads, 93% of which contained a barcode, the transposon sequence, and sufficient genomic sequence (>16 bp) for BLAST analysis. We were able to map more than half of these sequences (53%) unambiguously to the mouse genome. Of the 173,101 mapped sequences, 100,171 (67%) were redundant, leaving 72,930 nonredundant mapped insertions. Roughly half of the nonredundant insertions mapped to the same chromosome as the donor transposon concatamer (Chr 1), as expected because of the phenomenon of local hopping seen in other SB screens (Starr et al. 2009, Collier et al. 2005, Keng et al. 2009). To eliminate statistical bias in the dataset, these sequences were eliminated along with a smaller number of insertions that probably represent PCR artifacts (SI Materials and Methods). The remaining 30,088 insertions (Table 15) were analyzed to determine CIS. We used Monte Carlo simulations to find insertion rates in a given genomic window size that would not be expected to occur by chance (Starr et al., 2009). For example, based on a random assignment of 30,088 insertions to the mouse genome, one would not expect to find five or more insertions within a 12-kb window. Using these Monte Carlo defined parameters, we identified 37 CIS. Two of these CIS were removed from further analysis because all the tumors contributing to these two CIS
originated from a single mouse, indicating the tumors may be clonally related. Two more CIS were removed because they also were identified in a control dataset of tail-snip DNA from mice harboring unselected SB insertions and may represent hotspots for SB insertions (Starr et al., 2009). Because this control dataset was generated from tail snips, it is possible that other hotspots exist in other types of cells. After removal of these possible artifacts, 33 CIS remained (Table 10).

We assigned a candidate gene to each CIS if the majority of the insertions were in or near a single gene (Table 10). Four of the 33 CIS did not have an annotated gene within 40 kb and were not assigned a candidate gene. Another CIS contained two overlapping genes, SET domain-containing 5 (Setd5) and lipoma HMGIC fusion partner-like 4 (Lhfpl4), and all insertions in this CIS were in both genes. Notably, this CIS is located adjacent to the Rosa26 locus where the conditional SB11 knockin is located, and eight of nine insertions in this CIS are oriented with the internal promoter in the direction that would cause overexpression of the transgene. Rather than tagging an endogenous cancer gene, this CIS could represent selective pressure for increased mutagenesis via overexpression of SB11 transposase. Whether the transposon insertion caused a gain- or loss-of-function mutation sometimes can be predicted by analyzing the location and orientation of the insertions in all the tumors that comprise a single CIS. If all the tumors in a single CIS have transposon insertions in the same intron, and all the transposons are oriented in the direction of transcription, we predict the insertion causes a gain-of-function mutation. If the distribution of transposon insertions in all the tumors of a CIS is apparently random, and there is no bias in orientation, we predict a loss-of-function effect. Table 10 lists the predictions for the CIS. In total we identified 30 genes and four genomic loci with no annotated genes that probably contribute to intestinal tract cancer when mutated.
<table>
<thead>
<tr>
<th>Candidate gene</th>
<th>Chromosome</th>
<th>Start address*</th>
<th>End address*</th>
<th>Number of insertions†</th>
<th>Tumors with insertions‡</th>
<th>Predicted effect on gene§</th>
</tr>
</thead>
<tbody>
<tr>
<td>49.30422G04Rik</td>
<td>3</td>
<td>127241424</td>
<td>127431267</td>
<td>23</td>
<td>12</td>
<td>NP</td>
</tr>
<tr>
<td>AC11590.7</td>
<td>3</td>
<td>127697338</td>
<td>127730035</td>
<td>8</td>
<td>8</td>
<td>NP</td>
</tr>
<tr>
<td>AC131780.5</td>
<td>9</td>
<td>3001410</td>
<td>3030207</td>
<td>7</td>
<td>7</td>
<td>NP</td>
</tr>
<tr>
<td>Adams6</td>
<td>13</td>
<td>105093822</td>
<td>105104050</td>
<td>5</td>
<td>3</td>
<td>NP</td>
</tr>
<tr>
<td>Ap1er</td>
<td>3</td>
<td>127454370</td>
<td>127633377</td>
<td>17</td>
<td>10</td>
<td>NP</td>
</tr>
<tr>
<td>Apc</td>
<td>18</td>
<td>34324389</td>
<td>34514767</td>
<td>185</td>
<td>72</td>
<td>Loss</td>
</tr>
<tr>
<td>Atf2</td>
<td>2</td>
<td>73708689</td>
<td>73773260</td>
<td>9</td>
<td>9</td>
<td>NP</td>
</tr>
<tr>
<td>Atf2</td>
<td>17</td>
<td>80239837</td>
<td>80369888</td>
<td>10</td>
<td>7</td>
<td>Loss</td>
</tr>
<tr>
<td>Cnot1</td>
<td>8</td>
<td>98254541</td>
<td>98300340</td>
<td>8</td>
<td>8</td>
<td>Loss</td>
</tr>
<tr>
<td>Cske1a1</td>
<td>18</td>
<td>61726208</td>
<td>61750047</td>
<td>7</td>
<td>6</td>
<td>Loss</td>
</tr>
<tr>
<td>Elac1</td>
<td>18</td>
<td>73903365</td>
<td>73913215</td>
<td>5</td>
<td>3</td>
<td>Gain</td>
</tr>
<tr>
<td>Emcn</td>
<td>3</td>
<td>136918543</td>
<td>137092521</td>
<td>17</td>
<td>9</td>
<td>NP</td>
</tr>
<tr>
<td>Eso1</td>
<td>18</td>
<td>10578932</td>
<td>10777953</td>
<td>15</td>
<td>11</td>
<td>Loss</td>
</tr>
<tr>
<td>Fnbp11</td>
<td>3</td>
<td>122217901</td>
<td>122319064</td>
<td>11</td>
<td>11</td>
<td>Loss</td>
</tr>
<tr>
<td>Itgam</td>
<td>7</td>
<td>135180578</td>
<td>135240752</td>
<td>11</td>
<td>10</td>
<td>NP</td>
</tr>
<tr>
<td>Mys5b</td>
<td>18</td>
<td>74613029</td>
<td>74766289</td>
<td>18</td>
<td>17</td>
<td>Gain</td>
</tr>
<tr>
<td>No Gene 16</td>
<td>16</td>
<td>39019897</td>
<td>39031672</td>
<td>5</td>
<td>3</td>
<td>NP</td>
</tr>
<tr>
<td>No Gene 18</td>
<td>18</td>
<td>26169184</td>
<td>26282443</td>
<td>12</td>
<td>7</td>
<td>NP</td>
</tr>
<tr>
<td>No Gene 4</td>
<td>4</td>
<td>13170124</td>
<td>131238609</td>
<td>12</td>
<td>8</td>
<td>NP</td>
</tr>
<tr>
<td>No Gene Y</td>
<td>Y</td>
<td>2781406</td>
<td>2897989</td>
<td>16</td>
<td>15</td>
<td>NP</td>
</tr>
<tr>
<td>Nsd1</td>
<td>13</td>
<td>55352993</td>
<td>55372133</td>
<td>6</td>
<td>6</td>
<td>Gain</td>
</tr>
<tr>
<td>Pdcd5 dp</td>
<td>9</td>
<td>113560446</td>
<td>11372693</td>
<td>14</td>
<td>14</td>
<td>Loss</td>
</tr>
<tr>
<td>Pde4dp</td>
<td>3</td>
<td>97593369</td>
<td>97718572</td>
<td>14</td>
<td>11</td>
<td>NP</td>
</tr>
<tr>
<td>Pigl</td>
<td>11</td>
<td>6203495</td>
<td>62365763</td>
<td>16</td>
<td>16</td>
<td>NP</td>
</tr>
<tr>
<td>Setd5 or Lhpf4</td>
<td>6</td>
<td>113039133</td>
<td>113100176</td>
<td>9</td>
<td>9</td>
<td>NP</td>
</tr>
<tr>
<td>Sfi1</td>
<td>11</td>
<td>3004743</td>
<td>3179859</td>
<td>21</td>
<td>19</td>
<td>Loss</td>
</tr>
<tr>
<td>Smc24</td>
<td>18</td>
<td>53440600</td>
<td>53638364</td>
<td>19</td>
<td>12</td>
<td>Loss</td>
</tr>
<tr>
<td>Srflp1</td>
<td>18</td>
<td>52654288</td>
<td>52675270</td>
<td>6</td>
<td>5</td>
<td>NP</td>
</tr>
<tr>
<td>Stag1</td>
<td>9</td>
<td>100583003</td>
<td>100698850</td>
<td>12</td>
<td>12</td>
<td>Loss</td>
</tr>
<tr>
<td>Tmem122b</td>
<td>5</td>
<td>126077980</td>
<td>126152976</td>
<td>11</td>
<td>4</td>
<td>Loss</td>
</tr>
<tr>
<td>Wac</td>
<td>18</td>
<td>7855248</td>
<td>8046126</td>
<td>17</td>
<td>15</td>
<td>Loss</td>
</tr>
<tr>
<td>Zfp397</td>
<td>18</td>
<td>24110107</td>
<td>24157877</td>
<td>8</td>
<td>7</td>
<td>NP</td>
</tr>
<tr>
<td>Zfp609</td>
<td>9</td>
<td>65561432</td>
<td>65607064</td>
<td>8</td>
<td>8</td>
<td>Loss</td>
</tr>
</tbody>
</table>

*Genomic address based on National Center for Biotechnology Information Mouse genome Build 37.
†Number of nonredundant S8 transposon insertions within the locus.
‡Number of independent tumors with an insertion within the locus.
§Predicted effect is based on an analysis of the location and orientation of S8 transposon insertions in all tumors in a single CIS (see text for discussion). Gain, gain of function; Loss, loss of function; NP, no prediction.

Table 10. List of 33 CIS.
Transposon Insertions Implicated in LOH of the Wild-Type Allele of *Apc*. The most commonly mutated gene in this study was *Apc* (in 72 of 96 tumors), reflecting the strong selective pressure for loss of the wild-type allele in *Apc<sup>Min</sup>* mice. Previous studies have demonstrated that loss of the *Apc<sup>+</sup>* allele is an early event that occurs in almost every adenoma in *Apc<sup>Min</sup>* mice (Levy *et al.*, 1994, Luongo *et al.*, 1994). In addition, inactivation of the wild-type *Apc* allele is caused predominantly by homologous somatic recombination events, leading to the replacement of the *Apc* wild-type allele with a second *Apc<sup>Min</sup>* allele. We reasoned that in our transgenic model LOH could be accomplished by an inactivating transposon insertion, as opposed to duplication of the *Min* allele. To test this hypothesis, we performed PCR on DNA from tumors to amplify the region surrounding the *Min* mutation (T2860A). By sequencing the PCR amplicon, LOH can be ascertained in *Apc<sup>Min</sup>* mice by measuring the ratio of the T:A trace peak heights at the location of the *Min* mutation. In heterozygous tissue the T:A ratio is between 0.8 and 1.2, which is considered MOH, but in tissue that has lost the wildtype allele the ratio drops below 0.5 (Figure 36). Ratios between 0.5-0.8 and >1.2 are considered uninformative, most likely caused by contamination from nontumor tissue. Of the 96 tumors tested, 47 gave informative results (Table 11). Of these 47 tumors, 32 had an identified transposon insertion in the *Apc* locus, and 15 did not. The majority (73%) of tumors lacking a transposon insertion in *Apc* had T:A ratios < 0.5, indicating LOH probably caused by loss of the entire allele. In support of our hypothesis, 53% of the tumors that had a transposon insertion in the *Apc* locus had T:A ratios between 0.8 and 1.2, indicating maintenance of the wild-type *Apc* locus at the site of the *Min* mutation. This result suggests that in these tumors the wild-type *Apc* allele is inactivated by the transposon more frequently than by duplication of the *Min* allele.

Set of CIS Identified in *Apc<sup>Min</sup>* Mice Differs Significantly from Those Found in *Apc* Wild-Type Mice. We compared the list of genes identified in this study with the 77 genes identified in the screen we performed on an *Apc* wild-type background (Starr *et al.*, 2009). Surprisingly, only four genes were identified in both studies: *Apc*, nuclear receptor binding SET domain protein 1 (*Nsd1*), *Sfi1* homolog, spindle assembly associated (*Sfi1*), and WW domain containing adaptor with coiled-coil (*Wac*). There are several reasons
that could explain why the overlap between the two studies was low. First, the total number of genes that could contribute to tumor formation may be large enough that the size of these two studies is not sufficient to saturate the candidate genes. Second, because we use a statistical method to identify cancer genes, it is likely that transposon insertions contributed to carcinogenesis in some of the tumors, but the insertions did not occur at a rate high enough to qualify as a CIS. In support of this hypothesis, 70% of the loci identified as CIS in this study (23 of 33) also had one or more insertions in the same locus in the previous study. Third, the overlap may be small because selection pressure for specific genetic mutations in cells that already have an Apc mutation is biased toward a different set of cancer genes than in cells with a different initial mutation. Fourth, because of technical limitations, our method of amplifying and sequencing transposon insertions does not identify all transposon insertions, so a portion of driver mutations will not be identified. For example, analysis of replicate sequencing runs indicates that 20–40% of the PCR amplicons in a given library are not sequenced in a given GS FLX sequencing run (SI Materials and Methods and Table 12).

Relevance to Human Disease. To determine the relevance of these findings to human cancers, we analyzed the regions of human orthology to the CIS loci and the orthologous human genes. Of the 30 candidate mouse genes associated with a CIS, 28 had human orthologs. We queried the literature for mutations and recurrent copy number changes in these genes in human CRC. Three of the genes, APC, NSD1, and phosphodiesterase 4D interacting protein (PDE4DIP), are considered bona fide cancer genes based on the cancer gene census maintained by the Wellcome Trust Sanger Institute (Futreal et al. 2004). Eight genes, APC, activating transcription factor-2 (ATF2), atlastin GTPase 2 (ATL2), casein kinase 1, alpha 1 (CSNK1A1), integrin, alpha M (ITGAM), programmed cell death 6-interacting protein (PDCD6IP), and WAC, have documented mutations in human cancers cataloged in the COSMIC database (Forbes et al., 2006).

We found strong concordance between the CIS mouse loci and orthologous regions in the human genome showing recurrent chromosomal losses and gains in human CRC (Vogelstein et al., 1989, Reid et al., 1996, Nakao et al., 2004, Habermann et al., 2007, Lassmann et al., 2007, Shih et al., 2001, Derks et al., 2008, Kurashina et al., 2008).
Of the 33 identified CIS, 31 can be mapped to an orthologous human locus. Of these 31 candidate cancer loci, 24 are found in regions that commonly are lost or gained in human CRC (Table 13), including several of the CIS that are orthologous to human chromosomal arms 18q, 17p, 5q, and 4q. Interestingly, one CIS that has no annotated genes nearby (CIS No gene 16) is in an orthologous region (3q21–24) that is associated with CRC based on genome-wide linkage analyses (Kemp et al., 2006, Neklason et al., 2008). To determine the significance of this overlap, we performed the analysis using randomly generated CIS lists and a single dataset of regions that are lost recurrently in human CRC (Nakao et al., 2004). Roughly 250 genomic regions in this dataset were lost in > 5% of the human samples tested, and 22 of the 31 CIS were located within these regions. In 10,000 simulations using equivalent-sized randomly generated CIS lists, we find an overlap of this magnitude < 0.3% of the time. These results suggest that our SB screen may be capable of pinpointing the affected genes in these regions.

**Candidate Genes Regulate Proliferation of Human CRC Cell Lines.** We tested nine genes [CCR4-NOT transcription complex, subunit 1 (CNOT1), PDE4DIP, PDCD6IP, ATF2, SFII, formin-binding protein 1-like (FNBP1L), myosin VB (MYO5B), sorting nexin 24 (SNX24), and stromal antigen 1 (STAG1)] for their effect on proliferation of the human CRC cell line SW480 by knocking down message levels using siRNA. We used the SW480 line because it has an APC gene-truncation mutation similar to the ApcMin mutation (Nishisho et al., 1991). Cells were transfected two times at 48-h intervals with siRNA targeting the human genes. Knockdown efficiency was at least 50% for all nine genes as measured by quantitative real-time PCR. Cell proliferation was measured using a tetrazolium-based colorimetric assay on days two and six after the second transfection. Depletion of five (CNOT1, PDE4DIP, PDCD6IP, ATF2, and SFII) of the nine genes tested resulted in a significant decrease in cell viability compared with a control siRNA of at least 33% at day six after transfection (Table 14).

**Discussion**

Using a transposon-based forward genetic screen in mice, we identified 33 genomic loci that probably cooperate with a germline mutation in the Apc gene to cause
intestinal tumorigenesis. The most frequently mutated locus was the Apc locus, a result that supports the hypothesis that there is strong selective pressure to lose the wild-type copy during tumor formation. SB insertions in Apc were found in 72 of 96 tumors (75%), suggesting that 75% of the tumors in Apc\textsuperscript{Min} SB test mice undergo LOH at the Apc locus via SB insertional mutagenesis. This hypothesis is supported by sequencing of the region spanning the 1-bp Min mutation, which indicated that the majority of tumors containing SB insertions at the Apc locus maintained heterozygosity at the location of the T:A Min mutation. In contrast, the majority of tumors lacking a transposon insertion in Apc showed loss of the wild-type sequence at the Min mutation site. These results suggest that the majority of tumors underwent biallelic loss of Apc activity through transposon insertion or somatic recombination. However, it also is possible that in some cases activity was lost through other mechanisms of Apc inactivation or through transposon insertion substituting for loss of the wildtype allele.

Although LOH at the Apc locus is the rate-limiting event in tumor initiation in the Apc\textsuperscript{Min} mouse and in familial and sporadic APC-deficient CRC, loss of APC probably is insufficient for the survival and growth of transformed cells into adenomas and, eventually, adenocarcinomas. SB-mutagenized animals showed increased polyp number but no evidence of adenocarcinoma or metastasis. Thus, it is likely that the CIS candidate genes identified in this SB screen contribute in a diverse fashion to initiation, establishment, and survival of adenomas. Moreover, depending on the complexity of the mechanism or pathway, the CIS candidates discovered in our screen, like the relatively large number of genes reported to be mutant in individual human CRC (Wood et al., 2007), might be expected to occur at a low frequency if mutations at any one of multiple genes in a complex pathway can contribute equally to tumorigenesis.

Aside from Apc, only a few of the remaining 28 known genes that we identified as CIS in our screen have been implicated directly in CRC development, although > 90% are located in genomic regions that are lost or gained in human CRC. To eliminate false positives, we removed CIS that also were identified in a control dataset of unselected transposon insertions mapped in tail snips. However, it is possible that other tissuespecific hotspots could result in false positives. Nevertheless, the known functions
of several of the CIS candidate genes make them plausible candidates for drivers of human CRC. For example, **CNOT1** is a member of the Ccr4-Not complex, which is implicated in mRNA decay and transcriptional repression. In human cells, CNOT1 has been reported to be a repressor of nuclear receptor-mediated transcription (Winkler et al., 2006). One target of CNOT1 repression appears to be estrogen receptor alpha (ERα), via interactions between CNOT1 and the ligand-binding domain of ERα. Inhibition of CNOT1 caused an increase in the expression of ERα target genes in breast cancer cells, and ERα has been shown to be a tumor suppressor gene in the intestinal tract (Cho et al., 2007); in particular, knockout of ERα in Apc^{Min} mice caused a significant increase in intestinal tumorigenesis (Cleveland et al., 2009).

Another gene identified in this study, **Pdcd6ip** (also known as ALG-2 interacting protein X, Alix) is involved in membrane trafficking and apoptosis (Odorizzi G 2006). **Pdcd6ip** produces a protein that binds to the protein product of **Pdcd6**, a proapoptotic gene involved in T-cell receptor—, Fas—, and glucocorticoid-induced cell death (Vito et al., 1996). **Pdcd6ip** also can block down-regulation of the EGF receptor (EGFR), thereby having a positive effect on growth factor signaling (Schmidt et al., 2004). These contradictory roles could explain why loss of **Pdcd6ip** in the mouse tumors promoted growth (via the loss of the proapoptotic function) but the loss of **Pdcd6ip** in SW480 cells caused decreased proliferation (via increased down-regulation of EGFR). Further functional studies are required to elucidate the role of this adaptor protein.

In summary, our approach identified 30 genes that probably modify tumorigenesis in the Apc^{Min} model of human CRC. Further functional analysis of these CIS candidate genes may provide insights into the etiology and treatment of human CRC, especially those cancers arising downstream of APC deficiency.

**Materials and Methods (Detailed protocols are given in SI Materials and Methods.)**

**Mice.** Mice containing Apc^{Min}, Rosa26-LsL-SB11, Villin-Cre, and T2/Onc were reared using Institutional Animal Care and Use Committee-approved protocols. All mice were
on an isogenic C57BL/6J background. Mice were monitored daily and killed and necropsied when moribund or after 120 d.

**Histopathology and Immunohistochemistry.** Formalin-fixed tissues were embedded in paraffin, and standard techniques were used to stain tissue sections with H&E. Standard immunohistochemistry techniques were used to detect β-catenin.

**Linker-Mediated PCR.** Linkers [described previously (Wu et al. 2003)] were ligated to NlaIII- (right-side) or BfaI- (left side) digested genomic DNA using T4 DNA ligase. A secondary digest (Xhol, right side; BamHI, left side) was performed to destroy concatamer-generated products. Primary and secondary PCR was performed using primers specific for linker and SB transposon sequences along with Fusion and barcode sequences. PCR amplicons were sequenced using the GS FLX (Roche).

**Sequence Analysis.** Sequences were analyzed for the presence of the barcode, inverted repeat/direct repeat (IR/DR) sequences required for transposition, and linker sequences. Genomic sequence was blasted against the mouse genome using BLASTN at 95% stringency and requiring a single match. Of the 324,898 sequences analyzed, 53% could be uniquely mapped to the mouse genome. Sequences were removed if they were redundant, on the donor concatamer resident chromosome (Chr 1), in the En2 gene (because the En2 sequence is present in the transposon), and when a single TA dinucleotide contained multiple insertions from several tumors from multiple mice (because of the possibility of a PCR artifact). The remaining 30,088 nonredundant sequences were used to identify CIS. A CIS was defined by Monte Carlo simulations using a random dataset of 30,088 insertions.

**Apc LOH Analysis.** To measure LOH for the \( Apc^{Min} \) mutation, DNA was isolated from individual polyps, and PCR was performed using primers that flank the mutation (sense primer: CCGAGTAAGCAGAGACAA; antisense primer: GGGAGGTATGAATGGCTGAT). The PCR product was purified using Qiagen 96 MinElute vacuum purification plates per the manufacturer’s protocol and was sequenced using the sense primer as the sequencing primer. Trace peak heights at the location of the mutation were measured for each tumor, and the ratio of the T peak to the A peak was calculated.
Comparisons with Human Data. Eight publicly available studies measuring DNA copy number in CRC compared with normal tissue were analyzed. Mutations in human tumors were examined using the Catalog of Somatic Mutations in Cancer database (Forbes et al. 2006), and cancer gene status was based on the Census of Human Cancer Genes maintained by the Wellcome Trust Sanger Institute (Futreal et al. 2004).

Knockdown of CIS Candidate Genes using siRNA in SW480 Cells. SW480 cells were obtained from ATCC (catalog no. CCL-228) and were cultured under recommended conditions. Transient siRNA transfection was used to deplete expression of Min CIS genes.

Cell Viability Assay. Viability of siRNA-treated SW480 cells was determined using a 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay (Cell Viability Kit 1; Roche Applied Sciences).

Supporting Information

SI Materials and Methods

Mice. \( Apc^{Min} \) mice (C57BL/6J-ApcMin /J), which harbor a T→A nonsense mutation in the \( Apc \) gene that results in a truncated protein product, were obtained from a breeding colony at the University of Minnesota Medical School Animal Services facility. Rosa26-LsL-SB11 mice (backcrossed to C57BL/6J) were a generous gift from Adam Dupuy (University of Iowa, Iowa City, IA). Villin-Cre mice [B6.D2-Tg(Vil-Cre)20Syr], strain 01XE7, were purchased from the National Cancer Institute Mouse Repository. T2/Onc mice (mixture of C57BL/6J and FVB) were described previously (Collier et al. 2005). Mice were necropsied, and both normal and tumor tissues were collected by snap-freezing in liquid nitrogen or overnight fixation in 10% buffered formalin followed by 70% ethanol. PCR primer sequences were as follows: T2/Onc forward: CGCTTCTCGCTTCTGTTCGC, T2/Onc reverse: CACCCCCAGCATTCTAGTT; Villin-Cre forward: CAAGCCTGGCTCGACGGCC, Villin-Cre reverse: CGCGAACATCTTCAGGTTCT; Rosa26-lox-stop-lox-SB11 and Rosa26-SB11 knock-in3- primer: wild-type forward: CTGTCTTGGAGGCAGGAA, wildtype reverse:
CCCCAGATGACTACCTATCCTCCC, knock-in reverse: CTAAAAGGCCTATCACAAAC. Apc\textsuperscript{Min} mice were genotyped as described previously (Dietrich \textit{et al.} 1993).

**Histopathology and Immunohistochemistry.** Histopathological analysis of tumors and adjoining normal tissue was performed on tissues that were fixed in 10% neutral buffered formalin, routinely processed into paraffin, sectioned at a thickness of 4µm, and stained with H&E. Multiple H&E sections were obtained from tumors from the colon and duodenum. All tissues were analyzed by an American College of Veterinary Pathologists-certified veterinary pathologist (M.G.O.) from the University of Minnesota Masonic Cancer Comparative Pathology Shared Resources facility and using the standardized nomenclature of the 2003 Consensus Report and Recommendations for pathology of mouse models of intestinal cancer (Boivin \textit{et al.} 2003). Immunohistochemistry was performed on 4-µm formalin-fixed, paraffin-embedded sections of small intestine which were deparaffinized and rehydrated, followed by antigen retrieval using 10 mM citrate buffer, pH 6.0, in a steamer. Staining for β-catenin was performed on a Dako Autostainer using a goat anti-human β-catenin polyclonal antibody (catalog no. sc-1496; Santa Cruz) as primary antibody (after blocking endogenous peroxidase and application of a protein block), with detection by a biotinylated donkey anti-goat antibody (Jackson ImmunoResearch Laboratories) and streptavidin-linked horseradish peroxidase (Dako) using diaminobenzidine (Dako) as the chromogen. Mayer’s hematoxylin (Dako) was used as the counterstain. Small intestinal adenomas from Apc\textsuperscript{Min/+} mice were used as a positive control tissue, and for negative control slides the primary antibody was substituted with Super Sensitive Goat Negative Serum (Biogenex).

**Linker-Mediated PCR.** Linkers used to sequence insertions weredescribed previously (Wu \textit{et al.} 2003). Genomic DNA was digested with NlaIII (for sequencing from the right side of T2/Onc) or BfaI (for sequencing from the left side of T2/Onc) and ligated to the linker using T4 DNA ligase. A secondary digestion was performed to destroy concatamer-generated products (XhoI for right-side cloning and BamHI for left-side cloning). These two enzymes do not cut the transposon distal to the NlaIII or BfaI sites.
but do cut the plasmid backbone present in the transgene concatamer, effectively destroying concatamer linker amplicons. Primary PCR was performed using primers that flank the inverted repeat/direct repeat (IR/DR) sequences and the linker. Primer sequences are listed below. Primary PCR products were diluted 1:75 and used in a secondary PCR with nested primers. Secondary PCR was performed using FusA+BC+Left2°Primer or FusA+BC+Right2°Primer and FusB+linker2°Primer primers (see below). FusA and FusB are sequences required for pyrosequencing using the 454 GSFlex machine (BC, barcode). PCR products were quantified using QuantIT picogreen assay (Invitrogen) and diluted to a concentration of 200,000 molecules/µL. All tumor samples were combined and diluted by the number of samples added, for a final concentration of 200,000 molecules/µL. Samples were sequenced using the Roche Genome Sequencer FLX using 454 pyrosequencing technology (Roche Applied Science) by the University of Minnesota Biomedical Genomics Center. Sequences (all sequences are listed in the 5′ → 3′ direction) used were

Left linker+: GTAATACGACTCATAAGGGCTCCGCTTTAAGGGAC
Left linker−: TAGTCCCTTAAGCGGAG
Right linker+: GTAATACGACTCATAAGGGCTCCGCTTTAAGGGACCATG
Right linker−: GTCCCTTAAGCGGAGCC
Linker 1°Primer: GTAATACGACTCATAAGGGC
Linker 2°Primer: AGGGCTCCGCTTTAAGGGAC

(Note, linker primers work for both the Left linker and the Right linker)

Left 1°Primer: CTGGAAATTTCCTTGTTTTTAAGGACAGTGCAAC
Left 2°Primer: GGACATCTCTTTGGCATGACACAAGTC
Right 1°Primer: GCTTGAGGCTACTCGAAATGGGTGAC
Right 2°Primer: CCACTGGAATGTGATGAAAGAAATAAGC

FusA+BC+Left2°Primer:
GCCTCCCTCGCCCATCAGAATGCGCATTTTAAGTGATGTATGAACCTTC
Fusion A: GCCTCCCTCGCCCATCAG

Example barcode (each tumor is different): AATGCCGCAT.
FusA+BC+Right2°Primer:  
GCCTCCCTCGGCCATCAGAATGCCGCATTAAGGTGTATGTAIIACTTC  
FusB+linker2°Primer: GCCTTGCCAGCCCGCTCAGAGGGCTCCGCTTAAGGGAC  
Fusion B: GCCTTGCCAGCCCGCTCAG

A full list of barcodes is available upon request.

**Processing of Sequence Files.** PCR amplicons were generated so that the 10-bp library-identifying barcode always appeared in the beginning of the sequence in the sense orientation. Sequence quality was outstanding even up to the first base. Using a custom Perl script, we scanned positions 1–12 of all reads for the presence of the library barcode, allowing 0 or 1 mismatch. We typically found perfect matches to a single barcode at positions 1–11, with matches at 2–12 occurring rarely. We did not find barcode sequences (0–1 mismatch) anywhere else in the read sequences. We successfully assigned 98% of all sequence reads to a library barcode. All barcodes differed by at least 2 bp, and no sequence read matched two or more barcodes in a given region. This process was carried out separately for each of the six 454 pyrosequencing regions used in these analyses. Following barcode identification, the data for each of the six runs were merged to allow uniform handling of the entire dataset. Note that we did not attempt to assemble reads into contigs for several reasons: (i) The read quality was outstanding, matching the consensus with >99.9% accuracy when assembly was performed; (ii) the contigs tended not to tile at all, because the reads all were primed at the same location and are of similar length, conferring little advantage to using contigs; and (iii) assembly might introduce chimeric artifacts, particularly when two relatively closely spaced insertion sequences appeared on opposite strands.

To identify and remove IR/DR and linker sequences from each read, we applied EMBOSS Vectorstrip (Rice *et al.* 2000) with custom-designed modifications for pipeline application and assessed the best mismatch parameters to use. We sequentially attempted to match both construct elements (IR/DR and linker) in sense and antisense orientations with four successively less stringent parameter sets:

i) 10% mismatch allowed, long-construct elements (17–32 bp)  
ii) 10% mismatch allowed, short-construct elements (<26 bp)
iii) 15% mismatch allowed, short-construct elements (<26 bp)
iv) 20% mismatch allowed, short-construct elements (<26 bp)

Note that the short-construct elements affect only the IR/DR sequence, because the linker element already is shorter than 26 bp. As we reduced the stringency, more construct elements could be detected, but the risk of finding spurious chance matches increased. To guard against the introduction of spurious matches, we used generic scripts to assign a label to each recognized construct that was encountered from the 5’end to the 3’ end. We assigned the following labels: Ideal: a construct with both IR/DR and linker elements in the same orientation
No-linker: a construct with an IR/DR but missing a linker
No-IR/DR: a construct missing the IR/DR
Bad: a construct with no recognizable elements or multiple IR/DRs
Unknown: anything else

For example, an ideal construct might look like 11-[+IR/DR]- 42-[+linker]-4, where, in this case, the 42-bp insertion is in the sense orientation (as indicated by the + signs for IR/DR and linker). The numbers between elements in this representation indicate the number of base pairs between elements. Insertions < 16 bp were considered empty or unmappable. Ideal and nearly ideal construct configuration counts were monitored as the stringency parameters were relaxed. We expected the number of ideal configurations detected to increase and the unknown count to remain steady until we hit a parameter set that was too lax. We started with the strictest set of parameters (stringency level 0) and collected all ideal configurations. Then we took all sequences with nonideal configurations through stringency levels 1–3 to see if we could move them into the ideal category. In our final merged summary sequence table, the best status (ideal > no-linker > no-IR/DR > bad) was reported for each sequence along with the data for the most stringent run level (0–3) where it achieved this label.

**Mapping Insertion Sequences.** To map sequences to the mouse genome [National Center for Biotechnology Information (NCBI) Build 37], we used BLASTN (DeCypher’s TeraBLASTN, Active Motif, http://timelogic.com), requiring query sequences to align within 1 bp of the start of right-IR/DR sequences or within 1 bp of the end of left-IR/DR
sequences (i.e., within 1 bp of the transposon insertion site with both types of reads). Additionally, the query was required to match with at least 95% identity. Lower thresholds of 90% and 85% identity were tested but failed to yield sufficiently higher percentages of newly mappable insertions to warrant lowering the matching stringency. Because we were most interested in the IR/DR position, we were careful to ensure that the query matched within 1 bp of the IR/DR insertion site, but we did not require the 3′ end of the query to match, in case cloning artifacts had altered that end of the sequence. If secondary genome hits were found that were at least 95% as long as the first match, their count was recorded, and the insertion location was considered ambiguous. However, if all secondary hits appeared within 5,000 bp of the primary hit on the same chromosome, we considered the insertion to be uniquely mappable to that locus. Of the 324,898 sequences analyzed by BLASTN, 173,101 (53%) could be uniquely mapped to the mouse genome. We removed redundant sequences that arose from the same tumor and mapped to the same TA dinucleotide insertion site in the genome. Of the 173,101 mapped sequences, 100,171 (67%) were redundant, leaving 72,930 nonredundant mapped insertions. Three more filtering steps were performed:

i) To avoid the bias of “local hopping,” all nonredundant insertions mapping to the chromosome containing the transposon donor concatamer (Chr 1) were removed.

ii) The T2/Onc transposon contains sequence from an intron and splice acceptor of the murine En2 gene. Any insertions mapping to this sequence were removed because they might represent a transposition event back into the concatamer and not an insertion into the genomic En2 locus.

iii) Three-primer PCR was used to validate a sampling of nonredundant mapped insertions. We were successful in validating mapped insertions except in one circumstance. When a single TA dinucleotide contained multiple insertions from several tumors from multiple mice, we could not validate the insertions with three-primer PCR in some cases. We hypothesize that some of these TA dinucleotides are neasequences that cause PCR artifacts and do not represent true transposon insertions. We chose to adopt a conservative approach to avoid these artifacts, so we eliminated all insertions in TA dinucleotides containing insertions from two or more tumors from two or more different
mice. These three filtering steps removed 42,842 (59%) insertions, leaving a total of 30,088 nonredundant mapped insertions. This set was used to determine CIS.

**Statistics Used to Identify CIS.** To identify CIS, nonredundant insertions were assigned to clusters if the local density of insertions in a given window size exceeded that which would be expected by chance. Window sizes were determined by exact Monte Carlo simulation (see below). Based on a dataset of 30,088 insertions, the significance thresholds obtained are:

- Five or more insertions within 12,000 bp
- Six or more insertions within 22,000 bp
- Seven or more insertions within 34,000 bp
- Eight or more insertions within 50,000 bp,
- Nine or more insertions within 65,000 bp
- Ten or more insertions within 82,000 bp
- Eleven or more insertions within 105,000 bp
- Twelve or more insertions within 124,000 bp
- Thirteen or more insertions within 150,000 bp
- Fourteen or more insertions within 175,000 bp
- Fifteen or more insertions within 200,000 bp

The assumption of standard Poisson statistics that potential insertion sites are randomly distributed throughout the genome is not strictly correct, because (i) TA dinucleotides are naturally clustered in genomes, and (ii) numerous unfinished regions in the mouse genome are “off-limits” because they are long tracts of Ns. For example, the initial telomeric region of every chromosome except Y is padded with 3 million consecutive Ns. Both these factors lead standard analytical approaches to underestimate the size and number of clusters that actually would be encountered by simply picking randomly chosen real TA sites. In other words, by ignoring the natural clustering of TA sites in the genome, the number of false-positive CIS that will be predicted is increased systematically. The magnitude of deviation gets larger as more and more insertion sites are scattered about the genome, as one would expect intuitively. Hence, we wrote a program to compute exactly the expected number of CIS of a given size in a specified
window across all the chromosomes that one would encounter by chance via Monte Carlo simulation. The observed number of unambiguous mappable nonredundant insertions was used for each chromosome separately as input. For example if chromosomes 1 and 2 had 2,100 and 1,420 insertions, respectively, then we randomly distributed 2,100 insertions among the real TA dinucleotide sites on mouse chromosome 1 and another 1,420 among the TA sites of chromosome 2. Once the total count of insertions was distributed randomly among the real TA sites across the whole genome, a tally of the number of CIS of size ≥3, ≥4, . . . , ≥15 was recorded within windows of 10,000 bp, 20,000 bp, . . . 150,000 bp. This process was repeated 100 times, and the average counts over those 100 iterations were computed. Four independent simulations of 100 iterations each were performed, yielding SE bars between simulations of < 1%, indicating sufficient convergence. The values obtained can be interpreted as expected values (E-values), because they indicate the expected number of CIS of a given number of insertions that would be observed within a given window size merely by chance. We chose a threshold with an E-value < 1. Thus, finding 11 CIS of ≥15 insertions within 200,000 bp, when not even a single CIS was expected, is highly significant. We compared the thresholds obtained by this method with the thresholds obtained using standard Poisson statistics with the assumption of random insertion in the genome. We found our method was uniformly more stringent and yielded fewer false positives than the standard Poisson statistics. A modified Poisson model that takes into account the local density of TA sites in the genome yielded excellent agreement with the Monte Carlo calculations.

**Annotation and Sequence Information Management.** We created two primary annotation files: one outlining details of each unique insertion, and one describing each CIS. These files provide information on the chromosomal mapping position of each insertion or CIS, redundancy information on each insertion, and characteristics of the nearest EnsEMBL gene that flanks the insertion. EnsEMBL mappings were identified by a custom Perl script that uses the published Application Programmer Interface (Stabenau *et al.* 2004). To facilitate the management of all sequence information, an MySQL relational database was constructed to store (i) genotypic and phenotypic information on all mice and tumors from which the insertion sequences were derived; (ii)
metainformation on the sequencing runs themselves; (iii) raw read sequences; (iv) construct element matching characteristics; (v) final processed insertion sequences; (vi) mapping information for each processed insert sequence to the mouse genome; (vii) clustering assignments of inserts into CIS; and (viii) annotation information on all mapped inserts and CIS. SQL queries were performed to facilitate the merging of distinct gastrointestinal tract tumor datasets and the annotation process.

**Analysis of Replicate Sequencing Runs to Determine Percentage of Library Capture.** To estimate the extent of undersampling of transposon insertions in our study, we analyzed the GS FLX sequencing replicates separately. Using a four-region plate, one sequencing run of the GS FLX machine can sequence four separate samples. In one of the sequencing runs we ran two aliquots of the left-side ligation-mediated (LM)-PCR pool in separate regions (1 and 4) and two aliquots of the right-side LMPCR in separate regions (2 and 3). We analyzed the sequence reads by custom Perl script to determine the extent of overlap (Table 11). For example, region 1 returned 68,371 total reads, of which 56,435 contained a perfect match to a barcode, the transposon-specific sequence, and a genomic TA dinucleotide along with at least 16 bases of genomic DNA. When duplicate sequences were combined, 20,654 unique transposon insertion reads remained. Region 4, which was a replicate, had 18,863 unique reads. The overlap between these two regions was 10,448 reads, a little more than 50%. Because this process is similar to a mark-recapture experiment, one can use the Lincoln–Peterson method (Seber GAF 2002) to estimate the number of amplicons in the total population (see below). Based on the overlap between the replicates sampled in regions 1 and 4, we estimate that there were 37,289 unique amplicons in the original pool. As a rough approximation, our protocol sampled about 78% of the amplicons present in the original left-side LM-PCR pool. If we apply the same analysis to the right-side LM-PCR pool, where the overlap was lower (~35%), we estimate that we sampled only 58% of the amplicons in the original right-side LM-PCR pool. In either case, it is apparent that by increasing the number of sequencing runs, or perhaps by using a different sequencing platform, we could find more transposon insertions and perhaps more candidate cancer genes.
Lincoln–Peterson Method to Estimate Total Population in a Mark-and-Recapture Experiment. A mark-and-recapture experiment can be used to estimate population size in an ecological setting where the researcher marks all animals captured during a first visit. The researcher then returns and makes a second capture, noting how many of the animals in the second capture were marked in the first capture. To use the Lincoln–Peterson method (Seber GAF 2002) in our study, we assume that the duplicate pools of amplicons contain equal numbers of the same amplicons. This assumption could be invalid, because the concentration in the amplicon pool is very small (~200,000 molecules/µL). Randomly removing a small volume could result in the two volumes containing different amplicons. Another caveat to using the Lincoln–Peterson method is that our use of unique reads instead of actual reads is not directly analogous to a mark-and-recapture experiment. Nevertheless, if we assume the aliquots are similar, then the number of reads in the first region (M) is analogous to the number of animals caught and marked in the first capture of a mark-and-recapture experiment. The number of reads in the second region (C) is analogous to the number of animals caught in the second capture, and the number that overlaps (R) is analogous to the number of animals marked in the first capture that are caught in the second capture. To estimate the total population (N), the Lincoln–Peterson method states that the proportion of marked individuals in the second capture to the number in the first capture (R/M) should equal the proportion of the number of animals in the second capture to the total population (C/N). Rearrangement of this equation gives $n = \frac{M \times C}{R}$. Using the number of unique reads in regions 1 and 4 (Table 12) equates to $(20,654 \times 18,863)/10,448 = 37,289$. The total number of unique reads we captured in both sequencing regions $(20,654 + 18,863 - 10,448 = 29,069)$ represents 78% of the estimated total population of 37,289. Using the same analysis of the unique reads in region 2 and 3 (Table 12) equates to an estimated total population of $(25,753 \times 25,804)/9,079 = 73,194$. In this case, we have sampled only $(25,753 + 25,804 - 9,079 = 42,478)$ 58% of the total population.

_Apc Loss of Heterozygosity Analysis_. To measure loss of heterozygosity (LOH) for the ApcMin mutation, DNA was isolated from individual polyps, and PCR was performed
using primers that flank the mutation (sense primer: CGGAGTAAGCAGAGACACAA; antisense primer: GGGAGGTATGAATGGCTGAT). The PCR product was purified using Qiagen 96 MinElute vacuum purification plates according to the manufacturer’s protocol and was sequenced using the sense primer as the sequencing primer. Trace peak heights at the location of the mutation were measured for each tumor. Wild-type DNA will only have a single peak, representing thymidine, whereas ApcMin nontumor DNA will have two peaks of equal height representing the wild-type thymidine base and the mutant adenine base. If significant LOH has occurred, the ratio of the height of the thymidine peak to the adenine peak will be < 0.5, whereas a ratio > 0.8 indicates maintenance of heterozygosity. Ratios between 0.5 and 0.8 and ratios > 1.2 are considered to be contaminated with nontumor tissue.

**Chromosomal Copy Number Analysis.** Eight studies measuring DNA copy number in CRC compared with normal tissue were analyzed (Vogelstein *et al.* 1989, Ried *et al.* 1996, Nakao *et al.* 2004, Habermann *et al.* 2007, Lassmann *et al.* 2007, Shih *et al.* 2001, Derks *et al.* 2008, Kurashina *et al.* 2008). Of the 33 CIS loci, 31 were mapped to the homologous human region using the Batch Coordinate Conversion (Lift-Over) utility from the University of California, Santa Cruz (UCSC) genome browser (Kent *et al.* 2002). The genome build appropriate to each study was analyzed. A CIS locus was determined to be lost or gained recurrently based on the individual study methodology. To determine the chances of this overlap occurring randomly, we ran 10,000 simulations using randomly generated CIS lists compared with the recurrently lost regions in Nakao *et al.* (Nakao *et al.* 2004). In this study they hybridized genomic DNA from 135 human CRC along with normal lymphocytic DNA to a BAC array (HumArray1.14, University of California San Francisco; UCSF) and used Spot/Sproc analysis software (UCSF) to determine log2 ratios. They published an Excel file with the log2 ratios for each BAC clone for each human sample. Chromosomal losses were based on a log2 ratio lower than −0.225. The genomic coordinates of the BAC clones were downloaded from the UCSF website ([http://cancer.ucsf.edu/ docs/cores/array/analysis/HA1.14_clonepos_May04.2006 0811.txt](http://cancer.ucsf.edu/docs/cores/array/analysis/HA1.14_clonepos_May04.2006_0811.txt)) and were mapped to the Human genome build HG19 using the Batch Coordinate Conversion utility from the UCSC genome browser.
(http://genome.ucsc.edu/cgi-bin/hgLiftOver). After BAC clones that did not have at least 65% of samples giving a reading and positions that did not map to the HG19 build were removed, log2 ratios for 2,075 BAC clones remained. We selected the subset of these clones that had log2 ratios lower than −0.225 in at least 5% of the samples. Genomic regions recurrently lost were defined based on the genome coordinates of consecutive BAC clones that had log2 ratios lower than −0.225. For example, if three adjacent clones all showed a loss, we assigned the region based on the start coordinate of the first clone and the end coordinate of the last clone. If the neighboring clones on both sides of a BAC were not decreased, the region was defined as the coordinates of the single BAC. Next, we wrote a Perl script that created 10,000 random sets of 31 CIS using the UCSC Golden Path coordinates. Each set of 31 CIS contained 31 randomly generated genomic regions that were the same length as the actual ApcMin dataset CIS. The Perl script then counted the overlap between the random CIS list and the Nakao recurrently lost regions, and the process was repeated 10,000 times. In 10,000 simulations, only three random CIS lists had an overlap of 22 regions or more.

**Knockdown of CIS Genes Using siRNA in SW480 Cells.** SW480 cells were purchased from American Type Culture Collection. Cells were maintained in DMEM supplemented with 10% FBS, 2 mmol/L glutamine, 100 U/mL penicillin, and 100 µg/mL streptomycin and were incubated in a humidified atmosphere of 95% air and 5% CO2 at 37 °C. Cells were switched to antibiotic-free medium before siRNA transfection. Cells at 70% confluence were transfected twice, 48 h apart, with siRNA oligonucleotides (siRNA oligos) targeting the human ortholog of the CIS candidate gene (CIS siRNA) or with a nontargeting control (control siRNA), at a final concentration of 50 nmol/L, using Lipofectamine 2000 transfection reagent. Oligos targeting CIS genes were obtained from Qiagen (Hs_ATF2_3, Hs_PDCD6IP_5 Hs_SFI1_7, and Hs_PDE4DIP_15) and from Dharmacon (CNOT1 On-TARGETplus SMARTPool). The control siRNA oligo (OnTARGETplus Nontargeting siRNA#3) was obtained from Dharmacon. Total RNA was harvested from CIS siRNA- and control siRNA-treated cells 2 or 3 d after the second siRNA transfection using the RNeasy kit (Qiagen). For each experimental sample, 1.5 µg RNA was converted to cDNA with random nonomer primers (Integrated DNA
Technologies) and recombinant Omniscript Reverse Transcriptase using the Omniscript RT kit (Qiagen) according to manufacturer’s instructions. CIS cDNA and 18S ribosomal cDNA (used as internal control) were amplified from total cDNA by PCR. PCR was carried out in 10 µL using 2 µL cDNA from a 10× dilution of the 20 µL RT reaction (equivalent to 15 ng reverse transcribed RNA) for CIS genes and a 100× dilution for 18S rRNA, 500 nM of each primer, and 5 µL LightCycler 480 SYBR Green 1 Master Mix (Roche Applied Science). PCR was performed on a LightCycler 480 System (Roche Diagnostics) in 96-well plates using the amplification protocol: one cycle of preincubation, 5 m at 95 °C; 45 cycles of amplification each consisting of denaturation at 95 °C for 5 s, annealing at 60 °C for 5 s, and elongation at 72 °C for 10 s; one cycle melting at 95 °C for 5 s, 65 °C for 1 m, heating to 97 °C; one cycle cooling at 40 °C for 30 s. Water was used as a template for negative control amplifications for each PCR run. All reactions were performed in duplicate. Standards were generated by reverse transcription of total RNA from untreated cells followed by PCR amplification to generate template DNA of the same sequence as the predicted CIS gene or 18S gene PCR product. Serial dilutions of template DNA were amplified in parallel with experimental samples and used to generate a standard curve for each gene. Data were analyzed using Roche LightCycler 480 software, and crossing points (CPs) were calculated using the absolute quantification-second derivative maximum method. The standard curve was used to determine efficiency of PCR amplification (E) for each gene. Relative mRNA levels represent the expression of the CIS gene in CIS siRNA-treated cells relative to expression in control siRNA-treated cells. CIS gene expression was normalized to 18S rRNA levels, and relative mRNA levels were calculated as described by Pfaffl (Pfaffl MW 2001): relative mRNA levels = ([Etarget] ΔCPtarget(control-treated))/[Eref] ΔCPref(control-treated)) × 100, where E is real-time PCR efficiency and CP is defined as the point at which fluorescence rises appreciably above background.

Primer sequences are as follows: activating transcription factor-2 (ATF2), 5′-TGACCGAAAGGATCATGAACTA-3′ and 5′-GCAGTCCTTTCTCAAGTTTCCA-3′; CCR4-NOT transcription complex, subunit 1 (CNOT1), 5′ CTTTCAACCCCAATCGACCC-3′ and 5′-AGGTTTCATCTTTACTCTGCTGGA-3′;
programmed cell death 6-interacting protein (PDCD6IP), 5′ AGGTGTCCCTGTCTGGCTGC-3′ and 5′ TTCATCATAGCGAGATGCCACTGT-3′; phosphodiesterase 4D-interacting protein (PDE4DIP), 5′-GAGAATCCAGGACAAGAAACAGCAT-3′ and 5′-GGATTCACTCCTCAGGAAGCTGG-3′; Sfi1 homolog, spindle assembly associated (SFII), 5′-AGCAGCAGGAGATGAGGAACAAG-3′ and 5′-CGAACAACCACGTAGATCAACCAG-3′.

**Cell Viability Assay.** Viability of siRNA-treated SW480 cells was determined using a 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay (Cell Viability Kit 1; Roche Applied Sciences) in which absorbance at 595 nm is proportional to viable cell number. One day after the second siRNA transfection, cells were replated into 96-well plates at 1,250 cells per well (ATF2, PDCD6IP, and PDE4IP) or 5,000 cells per well (CNOT1) in triplicate in regular growth medium. Cell viability was determined on days 2–6 after the second transfection. Day 6 A595 values for each treatment were normalized to day 2 values for the same treatment to correct for differences in plating. Relative cell viability represents the ratio of normalized absorbance at 595 nm of cells treated with CIS siRNA to absorbance of cells treated with control siRNA on day 6. Results shown for mRNA and cell viability are the mean ± SD of at least two experiments.

**Figure 35. Three alleles used to target SB mutagenesis to the intestinal tract.** (A) T2/Onc transposon. IR/DR, inverted repeat/direct repeat sequences required for transposition; MSCV 5′ LTR, murine stem cell virus 5′ long terminal repeat; SA pA, splice acceptor with a polyA signal. (B) Conditional SB allele. Rosa26 locus, endogenous Rosa26 locus [Gt(Rosa)26Sor]; LoxP-GFP/Stop-LoxP, GFP cDNA flanked by LoxP sites; SB11 transposase pA, SB11 transposase cDNA with polyA signal. (C) Cre recombinase cDNA driven by the Villin promoter.
Figure 36. PCR technique for detecting Apc LOH. Trace peak heights of the T→A Min mutation in (A) wild-type mice, (B) heterozygous Apc\textsuperscript{Min} mice, and (C) an adenoma with LOH. Arrows indicate Min mutation. Red peaks = T; green peaks = A.

<table>
<thead>
<tr>
<th>Table 11. LOH and MOH in Apc\textsuperscript{Min} tumors based on the ratio of T:A trace peaks.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transposon insertion in Apc</td>
</tr>
<tr>
<td>---------------------------------</td>
</tr>
<tr>
<td>Yes\textsuperscript{1}</td>
</tr>
<tr>
<td>No</td>
</tr>
</tbody>
</table>

MOH, maintenance of heterozygosity.
\textsuperscript{*}To qualify as an informative tumor, the peak height ratio T:A must be <0.5 or between 0.8-1.2.
\textsuperscript{1}Yes indicates the tumors had a mapped 58 transposon insertion in the Apc gene.

Table 12. Sequence read overlap between duplicate regions of a single GS FLX sequencing run.
Table 13. Human orthologous regions to the mouse CIS with recurrent chromosomal copy number changes based on published data.

<table>
<thead>
<tr>
<th>Mouse address*</th>
<th>Mouse gene</th>
<th>Mouse Entrez ID</th>
<th>Human address†</th>
<th>Human</th>
<th>Human Entrez ID</th>
<th>Human band</th>
<th>In region of loss/gain‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr:1:3001410-3030207</td>
<td>AC131780.5</td>
<td>114673?</td>
<td>Deleted in humans</td>
<td>No homology</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>chr:17:281406-2897989</td>
<td>No Gene Y</td>
<td>N/A</td>
<td>Deleted in humans</td>
<td>No homology</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>chr:12:27241424-127431287</td>
<td>4930422G0Rik</td>
<td>71043</td>
<td>chr:11:31228994-113576517</td>
<td>Cnorf21</td>
<td>55345</td>
<td>4q25</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:12:27697338-127730035</td>
<td>AC115907.7</td>
<td>100043382</td>
<td>chr:11:12526166-112649108</td>
<td>No Gene 3</td>
<td>N/A</td>
<td>4q25</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:13:105093822-105104050</td>
<td>Adamts6</td>
<td>108154</td>
<td>chr:6:67477713-67458225</td>
<td>ADAMTS6</td>
<td>11174</td>
<td>5q12.3</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:12:72453470-127633377</td>
<td>Apil1</td>
<td>211556</td>
<td>chr:11:13034449-112630726</td>
<td>PAPAR</td>
<td>55435</td>
<td>4q25</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:13:34324389-34514767</td>
<td>Apc</td>
<td>11789</td>
<td>chr:5:111979623-112236985</td>
<td>APC</td>
<td>324</td>
<td>5q22.2</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:17:82326977-8080688</td>
<td>Wilms</td>
<td>56298</td>
<td>chr:1:38205178-8613026</td>
<td>AT1L2</td>
<td>64225</td>
<td>2p22.2-22.1</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:18:6726206-67150047</td>
<td>Cnk1a1</td>
<td>38667</td>
<td>chr:1:46872355-148192809</td>
<td>CSNK1A1</td>
<td>1452</td>
<td>5q32</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:18:79303365-73913215</td>
<td>Elac1</td>
<td>114615</td>
<td>chr:18:48945579-48909488</td>
<td>ELAC1</td>
<td>55520</td>
<td>18q21.2</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:13:36919543-13705251</td>
<td>Emon</td>
<td>59308</td>
<td>chr:10:101232972-101544124</td>
<td>EMMCN</td>
<td>51795</td>
<td>4q24</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:18:10575893-10773935</td>
<td>Esco1</td>
<td>77805</td>
<td>chr:18:19121974-19393499</td>
<td>ESCO1</td>
<td>114299</td>
<td>18q12.2</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:12:22217801-122319064</td>
<td>Frip2</td>
<td>214439</td>
<td>chr:9:9161974-9409597</td>
<td>FNBP1L</td>
<td>54874</td>
<td>1p22.1</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:13:35180978-135240752</td>
<td>Itgam</td>
<td>16409</td>
<td>chr:16:31267174-31289751</td>
<td>ITGAM</td>
<td>3684</td>
<td>10p11.2</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:17:84713029-74976259</td>
<td>MyoD</td>
<td>1719</td>
<td>chr:18:4596944-47027268</td>
<td>MYOD1</td>
<td>4646</td>
<td>10q21.1</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:18:26181904-26243446</td>
<td>No Gene 18</td>
<td>N/A</td>
<td>chr:18:3553815-35494000</td>
<td>No Gene 18</td>
<td>N/A</td>
<td>18q12.2</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:13:1170124-13198099</td>
<td>No Gene 4</td>
<td>N/A</td>
<td>chr:9:29760033-29873048</td>
<td>No Gene 4</td>
<td>N/A</td>
<td>1p35.3</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:13:5532993-55372133</td>
<td>Ncl1</td>
<td>18193</td>
<td>chr:17:166657455-167671302</td>
<td>NCD1</td>
<td>63424</td>
<td>5q33.3</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:9:7953369-97718572</td>
<td>Pde4d</td>
<td>83679</td>
<td>chr:1:144494532-145112252</td>
<td>PDE4D</td>
<td>9659</td>
<td>1q21.1</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:11:6203495-62369763</td>
<td>Pigt</td>
<td>327942</td>
<td>chr:17:16049012-16299937</td>
<td>PIGL</td>
<td>9467</td>
<td>17p12.2</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:12:13039133-13130176</td>
<td>Sdt5Lhflp4</td>
<td>72895269788</td>
<td>chr:3:945056-9516832</td>
<td>SETD5</td>
<td>55209</td>
<td>2p23.3</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:11:3004743-3179859</td>
<td>Sfi1</td>
<td>78887</td>
<td>chr:2:213790517-32022116</td>
<td>SF1</td>
<td>8914</td>
<td>22q12.2</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:18:534060-53638864</td>
<td>Sro24</td>
<td>69226</td>
<td>chr:12:2221987-122440150</td>
<td>SRO24</td>
<td>28966</td>
<td>5q23.2</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:18:2654288-3267520</td>
<td>Srbp1</td>
<td>67222</td>
<td>chr:12:1270966-121397902</td>
<td>SRFBP1</td>
<td>15344</td>
<td>5q31.1</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:18:24110107-24151877</td>
<td>Zfp397</td>
<td>69256</td>
<td>chr:18:32186313-32817403</td>
<td>ZNF397</td>
<td>84307</td>
<td>18q12.2</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:6:5551342-65607064</td>
<td>Zfp609</td>
<td>214812</td>
<td>chr:15:64869232-64946883</td>
<td>ZNF609</td>
<td>23690</td>
<td>15q23.2</td>
<td>Yes</td>
</tr>
<tr>
<td>chr:2:3770869-37773640</td>
<td>Atf2</td>
<td>11909</td>
<td>chr:17:155979880-170769569</td>
<td>ATF2</td>
<td>1386</td>
<td>2p11.1</td>
<td>No</td>
</tr>
<tr>
<td>chr:9:8529451-9803000</td>
<td>Cnot1</td>
<td>234594</td>
<td>chr:16:65568279-3623838</td>
<td>CNOT1</td>
<td>23019</td>
<td>16q21</td>
<td>No</td>
</tr>
<tr>
<td>chr:16:3019987-29031672</td>
<td>No Gene 16</td>
<td>N/A</td>
<td>chr:13:212758030-19275409</td>
<td>No Gene 16</td>
<td>N/A</td>
<td>3q29</td>
<td>No</td>
</tr>
<tr>
<td>chr:11:31350446-113732933</td>
<td>Pde6c1</td>
<td>18571</td>
<td>chr:3:33684048-33911522</td>
<td>PDE6C1</td>
<td>10015</td>
<td>3p23.2</td>
<td>No</td>
</tr>
<tr>
<td>chr:10:6053600-10068950</td>
<td>Stag1</td>
<td>20842</td>
<td>chr:1:316220566-136384331</td>
<td>STAG1</td>
<td>10274</td>
<td>3p23.2</td>
<td>No</td>
</tr>
<tr>
<td>chr:12:6507980-126159706</td>
<td>Tmem132b</td>
<td>206151</td>
<td>chr:1:12215776600-125970181</td>
<td>TMEM132B</td>
<td>117495</td>
<td>12q24.3-24.32.3</td>
<td>No</td>
</tr>
<tr>
<td>chr:18:7855248-8046126</td>
<td>Wac</td>
<td>225131</td>
<td>chr:10:28820138-28913566</td>
<td>WAC</td>
<td>51322</td>
<td>10p12.1</td>
<td>No</td>
</tr>
</tbody>
</table>
*Based on mouse July 2007 Assembly (mm9) UCSC genome browser.
†Based on human GRCh37 Assembly (hg19) UCSC genome browser.
‡Region was identified in one of the following studies: Nakao et al., 2004; Derks et al., 2008; Lassmann et al., 2007; Habermann et al., 2007; Ried et al., 1996; Vogelstein et al., 1989; Shih et al., 2001.

Table 14. Knockdown of ApoMin CIS candidate genes affects viability of human colon cancer cells.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Relative mRNA levels of cells depleted for CIS gene*</th>
<th>Relative cell viability of cells depleted for CIS gene†</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATF2</td>
<td>17.2 ± 5.6%</td>
<td>61.0 ± 1.5%</td>
</tr>
<tr>
<td>CNOT1</td>
<td>13.7 ± 7.4%</td>
<td>18.5 ± 16%</td>
</tr>
<tr>
<td>PDCD6IP</td>
<td>10.5 ± 2.6%</td>
<td>64.2 ± 28%</td>
</tr>
<tr>
<td>PDE4DIP</td>
<td>27.9 ± 10%</td>
<td>55.0 ± 19%</td>
</tr>
<tr>
<td>SFI1</td>
<td>40.4 ± 13.2%</td>
<td>60.0 ± 14%</td>
</tr>
</tbody>
</table>

*Relative mRNA levels calculated using the ΔΔCt method (1) and represent levels of CIS mRNA in CIS siRNA-treated cells relative to levels in control, nontargeting siRNA-treated cells. 18S RNA was used as a housekeeping control. Results shown are the mean ± SD of at least two experiments.
†Relative cell viability represents the ratio of absorbance at 595 nm of cells treated with the indicated CIS gene siRNA to absorbance of cells treated with nontargeting control siRNA. Results shown are the mean ± SD of at least two experiments.

Table 15. Mapped transposon insertions in 96 tumors. Please see the weblink:

**Personal Contribution.** Lei Zhao performed MTT experiments analyzed data to determine the knockdown effect of the following genes on proliferation of human (SW480) CRC cells: CNOT1, PDCD6IP, PDE4DIP and SFI1.
APPENDIX 2

The role of KCNQ1 in mouse and human gastrointestinal cancers

BLN Than1,2, JACM Goos3, AL Sarver4, MG O’Sullivan5, A Rod1, TK Starr6,7, RJA Fijneman3, GA Meijer3, L Zhao1, Y Zhang8, DA Largaespada7, PM Scott1 and RT Cormier1

1Department of Biomedical Sciences, University of Minnesota Medical School, Duluth, MN, USA; 2Toxicology Graduate Program, University of Minnesota, Duluth, MN, USA; 3Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands; 4Department of Biostatistics and Informatics, Masonic Cancer Center, University of Minnesota, Minneapolis, MN, USA; 5College of Veterinary Medicine, University of Minnesota, St Paul, MN, USA; 6Department of Genetics, Cell Biology and Development, Center for Genome Engineering, Masonic Cancer Center, University of Minnesota, Minneapolis, MN, USA; 7Department of Obstetrics, Gynecology and Women’s Health, Masonic Cancer Center, University of Minnesota Medical School, Minneapolis, MN, USA and 8University of Minnesota Supercomputing Institute, Minneapolis, MN, USA.

Oncogene advance online publication 26 August 2013; doi: 10.1038/onc.2013.350

Abstract. Kcnq1, which encodes for the pore-forming a-subunit of a voltage-gated potassium channel, was identified as a gastrointestinal (GI) tract cancer susceptibility gene in multiple Sleeping Beauty DNA transposon-based forward genetic screens in mice. To confirm that Kcnq1 has a functional role in GI tract cancer, we created ApcMin mice that carried a targeted deletion mutation in Kcnq1. Results demonstrated that Kcnq1 is a tumor suppressor gene as Kcnq1 mutant mice developed significantly more intestinal tumors, especially in the proximal small intestine and colon, and some of these tumors progressed to become aggressive adenocarcinomas. Gross tissue abnormalities were also observed in the rectum, pancreas and stomach. Colon organoid formation was
significantly increased in organoids created from Kcnq1 mutant mice compared with wild-type littermate controls, suggesting a role for Kcnq1 in the regulation of the intestinal crypt stem cell compartment. To identify gene expression changes due to loss of Kcnq1, we carried out microarray studies in the colon and proximal small intestine. We identified altered genes involved in innate immune responses, goblet and Paneth cell function, ion channels, intestinal stem cells, epidermal growth factor receptor and other growth regulatory signaling pathways. We also found genes implicated in inflammation and in cellular detoxification. Pathway analysis using Ingenuity Pathway Analysis and Gene Set Enrichment Analysis confirmed the importance of these gene clusters and further identified significant overlap with genes regulated by MUC2 and CFTR, two important regulators of intestinal homeostasis. To investigate the role of KCNQ1 in human colorectal cancer (CRC), we measured protein levels of KCNQ1 by immunohistochemistry in tissue microarrays containing samples from CRC patients with liver metastases who had undergone hepatic resection. Results showed that low expression of KCNQ1 expression was significantly associated with poor overall survival.

INTRODUCTION

The KCNQ1 gene encodes for the pore-forming a-subunit of a voltage-gated potassium channel that enables a K⁺ current after electrical depolarization of the cell membrane. KCNQ1 is predominantly expressed in the myocardium, inner ear, stomach, intestine and pancreas, tissues in which KCNQ1 expression is critical for ion homeostasis (Peroz et al., 2008). Inherited mutations in KCNQ1 underlie several human disease syndromes. For example, mutations in KCNQ1 are associated with congenital long QT syndrome, a disorder caused by abnormal ventricular repolarization that increases the risk of sudden death from cardiac arrhythmias. Jervell and Lange–Nielson syndrome is a rare autosomal disorder characterized by deafness in addition to long QT syndrome. Recent reports indicate that Jervell and Lange–Nielson syndrome patients can also be susceptible to GI defects, including iron-deficiency anemia, gastric and duodenal hyperplasia, elevated gastrin levels and, more rarely, gastric adenocarcinoma (Winbo et al., 2012, Grahammer et al., 2001, Rice et al., 2011, Nikou et al., 2011). Variants in KCNQ1 are
also associated with enhanced risk for type 2 diabetes (Unoki et al., 2008, Yasuda et al., 2008).

Relatively little is known about the role of \textit{KCNQ1} in cancer, but a potential role in GI tract cancers was indicated by its identification as a high-frequency common insertion site locus in Sleeping Beauty DNA transposon-based forward mutagenesis genetic screens for intestinal (Starr et al., 2009, March et al., 2011) and pancreatic cancer (Mann et al., 2011, Perez-Mancera et al., 2012). In our forward genetic screen in \textit{Apc} wild-type mice, we found \textit{Kcnq1} mutations in 18 intestinal tumors (13%) (Starr et al., 2009) ranking it among the top 10 common insertion site genes. In this study, while transposon insertions in \textit{Kcnq1} were found in tumors from all regions of the intestine, approximately two-thirds (64%) were located in the duodenum and jejunum, with 26% in the ileum and 10% in the colon. Furthermore, transposon insertions in \textit{Kcnq1} were observed equally in both the forward and reverse strand orientation, consistent with a model whereby the transposon is acting to disrupt gene function. We performed a second forward genetic screen in \textit{Apc}\textsubscript{Min} mice and found \textit{Kcnq1} transposon insertions in 13 intestinal tumors (14%) (Starr et al., 2011). Another screen in \textit{Apc}\textsubscript{Min} mice found \textit{Kcnq1} transposon mutations in 120 intestinal tumors (27%), ranking it no. 15 out of 641 common insertion site genes (120 kb window) (March et al., 2011). We have also conducted a screen using p53R270H mutant mice and identified insertions in \textit{Kcnq1} in 15 intestinal tumors (23%) (TK Starr, personal communication). Six mutant mouse alleles of \textit{Kcnq1} have been created that partially model human disorders arising from inherited mutations in \textit{KCNQ1}, especially Jervell and Lange–Nielson syndrome (Lee et al., 2000, Casimiro et al., 2001, Casimiro et al., 2004, Takagi et al., 2007, Elso et al., 2004). These mutant alleles share similar but not identical phenotypes, including bilateral deafness from birth, gastric hyperplasia, gastric achlorhydria, elevated serum gastrin, headbobbing, bidirectional circling, ataxia, body tremor and hyperactivity with a consequent significant decrease in body weight. Conflicting results were reported regarding a cardiac phenotype in these models. In mouse small intestine, \textit{Kcnq1} expression is strongest in the duodenum and proximal jejunum and decreases towards the ileum (Demolombe et al., 2001, Dedek et al., 2001). In the large intestine, \textit{Kcnq1} expression is strongest in the distal colon
(Vallon et al., 2005, Warth et al., 2002). To further investigate the role of Kcnq1 in GI tract cancers, we measured GI cancer phenotypes following introgression of a targeted knockout (KO) allele of Kcnq1 (Casimiro et al., 2001) into the Apc<sup>Min</sup> model of intestinal tumorigenesis. We then investigated the role of KCNQ1 in progression of human colorectal cancer (CRC) by measuring KCNQ1 protein levels in tissue microarrays containing samples from more than 500 CRC patients with liver metastases. Results from both of these studies indicate that Kcnq1 is a tumor suppressor.

RESULTS

Loss of Kcnq1 enhances tumor multiplicity in Ape<sup>Min</sup> mice.

Haploinsufficiency for Kcnq1 significantly enhanced Ape<sup>Min</sup> intestinal tumor multiplicity overall, in both males and females, by ~40% (Table 16), with the phenotype strongest in the proximal small intestine and colon, where tumors increased 2-fold. Kcnq1<sup>−/−</sup> mice in our study manifested a range of previously reported phenotypes: rapid bidirectional circling, ataxia, tremor, head bobbing and they were smaller (~ 25%, body weight) and leaner than littermate Kcnq1 wild-type and heterozygous mice. Less than half of the expected number of Apc<sup>Min</sup> Kcnq1<sup>−/−</sup> mice survived until weaning (23 vs ~60). Survivor Ape<sup>Min</sup> Kcnq1<sup>−/−</sup> mice demonstrated a very strong increase in tumors in the proximal half of the small intestine (B3-fold), especially in the proximal quarter of the small intestine containing the duodenum and proximal jejunum (Table 17), the region exhibiting strongest expression of Kcnq1 in the intestine. Ape<sup>Min</sup> Kcnq1<sup>−/−</sup> mice developed B2-fold increase in colon tumors but no increase in the distal small intestine of males, and a decrease in females. No significant differences between Kcnq1 genotypes in tumor sizes were observed.
Table 16. Loss of Kcnq1 enhances tumor multiplicity in Apc<sup>Min</sup> mice.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Colon</th>
<th>Proximal SI</th>
<th>Distal SI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Males</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kcnq1&lt;sup&gt;+/+&lt;/sup&gt;</td>
<td>30</td>
<td>1.8 ± 1.7</td>
<td>11 ± 8</td>
<td>56 ± 41</td>
</tr>
<tr>
<td>Kcnq1&lt;sup&gt;+-/−&lt;/sup&gt;</td>
<td>34</td>
<td>3.8 ± 2.3*</td>
<td>18 ± 13*</td>
<td>76 ± 38*</td>
</tr>
<tr>
<td>Kcnq1&lt;sup&gt;−/−&lt;/sup&gt;</td>
<td>10</td>
<td>2.6 ± 2.8</td>
<td>36 ± 16*</td>
<td>52 ± 43</td>
</tr>
<tr>
<td><strong>Females</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kcnq1&lt;sup&gt;+/+&lt;/sup&gt;</td>
<td>26</td>
<td>1.2 ± 1.7</td>
<td>14 ± 9</td>
<td>58 ± 25*</td>
</tr>
<tr>
<td>Kcnq1&lt;sup&gt;+-/−&lt;/sup&gt;</td>
<td>46</td>
<td>2.5 ± 2.2*</td>
<td>20 ± 12</td>
<td>84 ± 37*</td>
</tr>
<tr>
<td>Kcnq1&lt;sup&gt;−/−&lt;/sup&gt;</td>
<td>13</td>
<td>2.1 ± 2.0</td>
<td>22 ± 11*</td>
<td>30 ± 16</td>
</tr>
</tbody>
</table>

Abbreviation: SI, small intestine. Apc<sup>Min</sup> Kcnq1<sup>+/−</sup>, Apc<sup>Min</sup> Kcnq1<sup>+/−</sup> and Apc<sup>Min</sup> Kcnq1<sup>−/−</sup> mice were killed at 120 days of age and tumors counted and measured. *P-values: <0.05; two-sided P-values for tumor counts were determined by use of the Wilcoxon rank-sum test comparing gender- and age-matched classes produced in the same genetic crosses.

Table 17. Apc<sup>Min</sup> Kcnq1<sup>−/−</sup> tumor phenotype is strongest in the proximal quarter of the small intestine.

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Tumors</th>
<th>Females</th>
<th>Tumors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kcnq1&lt;sup&gt;+/+&lt;/sup&gt;</td>
<td>4 ± 2</td>
<td>Kcnq1&lt;sup&gt;+/−&lt;/sup&gt;</td>
<td>5 ± 3</td>
<td></td>
</tr>
<tr>
<td>Kcnq1&lt;sup&gt;+-/−&lt;/sup&gt;</td>
<td>6 ± 3</td>
<td>Kcnq1&lt;sup&gt;−/−&lt;/sup&gt;</td>
<td>8 ± 4</td>
<td></td>
</tr>
<tr>
<td>Kcnq1&lt;sup&gt;−/−&lt;/sup&gt;</td>
<td>21 ± 11*</td>
<td>Kcnq1&lt;sup&gt;−/−&lt;/sup&gt;</td>
<td>15 ± 9*</td>
<td></td>
</tr>
</tbody>
</table>

Mice and tissues are the same as Table 1. *P-values: <0.05; two-sided P-values for tumor counts were determined by use of the Wilcoxon rank-sum test comparing gender- and age-matched classes produced in the same genetic crosses.

Loss of Kcnq1 promotes tumor progression

Tumors in Apc<sup>Min</sup> mice are invariably benign adenomas that usually vary in size between 0.3 and ~5 mm. Adenocarcinomas are extremely rare. In Apc<sup>Min</sup> mice carrying
either heterozygous or homozygous mutations in Kcnq1 development of adenocarcinomas occurred at a frequency of ~10–15%. These adenocarcinomas appeared as large lesions (some >20 mm), located in the proximal small intestine in a region approximate with the duodenal–jejunal flexure at the head of the pancreas. In almost all cases, there was buckling of the intestinal tube, in two cases there was a pronounced intussusception of the tissue and in two cases the tumors extended across the entire inner surface of the intestinal lumen. Six of these large tumors were analyzed by histopathology and all were found to be invasive adenocarcinomas (Figure 37). Figure 37a depicts a lesion of > 2 cm in length from the duodenum of an Apc\textsuperscript{Min} Kcnq1\textsuperscript{+/−} mouse, which when sectioned regionally revealed adenocarcinoma tissue in each section. No adenocarcinomas were detected in the remainder of the small intestine or in the large intestine, and no putative adenocarcinomas were detected in Apc\textsuperscript{Min} Kcnq1\textsuperscript{+/+} mice.

**Figure 37. Loss of Kcnq1 promotes tumor progression.** (a) Duodenum from Apc\textsuperscript{Min} Kcnq1\textsuperscript{+/−} mouse with stomach at right. Bar = 5 mm. Broad area of hyperplastic tissue with embedded tumors extends from pylorus–duodenum juncture past the ampulla to a region approximate with the duodenal jejunal flexure. Asterisk indicates area of normal duodenum–proximal jejunum. Box encloses a region of adenocarcinoma tissue of > 2 cm. Sections from this lesion were taken from various regions and all showed evidence of invasion. Two of these adenocarcinoma sections are
depicted in (b–f). Smaller arrow depicts normal large adenoma of ~5mm that is embedded in hyperplastic tissue. (b and c) Bars = 500 µm; invasion into muscularis mucosa, (d and e) bar = 250 µm; prominent scirrhous response to invading tubules that exhibit differentiated goblet cells (long arrow) and signet ring cells (short arrows) (f), bar = 50 µm.

**Loss of Kcnq1 is associated with other pathologies in the GI tract.**

Rectal prolapse was evident in ~10% of \(Apc^{\text{Min}}\) \(Kcnq1^{+/–}\) and \(Apc^{\text{Min}}\) \(Kcnq1^{-/-}\) mice and in several cases the rectal prolapses were associated with enlarged rectums characterized by diffused severe crypt hyperplasia (Figure 44). In addition, \(Kcnq1^{-/-}\) mice developed severe gastric hyperplasia that was characterized by mucosal hyperplasia, with parietal cells increased in size and crypt dilation with extensive cellular debris (Figure 45). Hyperplastic adenosmas in the pylorus were also present in \(Apc^{\text{Min}}\) \(Kcnq1\) mutant mice. Although \(Apc^{\text{Min}}\) \(Kcnq1^{+/–}\) mice showed an incidence of pyloric tumors of ~30%, the phenotype was much stronger in \(Apc^{\text{Min}}\) \(Kcnq1^{-/-}\) mice where the incidence of pyloric tumors was ~90% (21/23). Furthermore, in \(Apc^{\text{Min}}\) \(Kcnq1^{-/-}\) mice pyloric tumor multiplicity was much larger, with some mice developing as many as a dozen tumors. No pyloric tumors were observed in \(Apc^{\text{Min}}\) \(Kcnq1^{++}\) mice. Finally, roughly one-third of \(Apc^{\text{Min}}\) \(Kcnq1^{-/-}\) mice developed grossly and severely enlarged (~2- to 3-fold) pancreases, characterized by extra lobes. Histopathology of a matched pair of samples indicated that zymogen granules were less eosinophilic and less prominent, and that both exocrine cells and islets were smaller (Figure 46).

**Loss of Kcnq1 has a strong effect on gene expression**

We conducted cDNA microarray expression studies using tissue from distal colon and the proximal quarter of the small intestine of \(Apc^{+/+}\) \(Kcnq1^{+/+}\) and \(Apc^{+/+}\) \(Kcnq1^{-/-}\) mice. Overall, more than 400 genes showed a > 1.5-fold change in expression, and > 200 of these genes also had a P-value of < 0.05 (Tables 18 and 19). Expression of 20 genes was confirmed by reverse transcription–polymerase chain reaction (RT–PCR) (Figures 38 and 39).
Figure 38. qRT–PCR gene expression analysis of mouse colon. Samples were analyzed in triplicate and normalized to 18S ribosomal RNA. Data are presented as the mean fold change ± s.d. Each bar represents the mean and s.d. of multiple experiments that measured fold differences in the mRNA expression of whole colon tissue isolated from adult (~100 days), littermate and gendermatched pairs of Apc+/+ Kcnq1+/+ and Kcnq1−/− mice. RNA was isolated from 1 cm sections from the same region of distal colon. At least two matched pairs of mRNAs were tested for each gene with most genes tested in at least three matched pairs of mRNAs. To be included in this figure, genes showed a mean fold difference of at least 1.5. In all cases, the direction in changes in gene expression confirmed microarray data. *P<0.05.
Figure 39. qRT–PCR gene expression analysis of mouse proximal small intestine. Each gene sample was run in triplicate and gene expression was normalized to the expression of 18S. Data are presented as the mean fold change ± s.d. Each bar represents the mean and s.d. of multiple experiments that measured fold differences in the mRNA expression in proximal small intestine tissue isolated from adult (~100 days), littermate and gendermatched pairs of Apc<sup>+/−</sup> Kcnq1<sup>+/−</sup> and Kcnq1<sup>−/−</sup> mice. mRNAs were isolated from 1 cm sections of proximal small intestine from the same region for all mice. At least two matched pairs of mRNAs were tested for each gene with most genes tested in at least three matched pairs of mRNAs. To be included in this figure, genes showed a mean fold difference of at least 1.5. In all cases, the direction of changes in gene expression confirmed microarray data. *P<0.05.

Target genes were clustered into several functional groups: innate immune response/goblet and Paneth cell function (Ang4, Retnlb, Clps, Pnliprp2, Clca3, Muc2, Reg3a, Reg3b, Reg3g, Car1), ion channels (Trpv6, Slc12a8, Slc30a10, Aqp7, Aqp1, Aqp8, Aqp4, Clca3, Clca6), mucins (Muc2, Muc13, Muc4, Muc3), growth regulatory signaling pathways (Areg, Egr1, Fos, Ccdn1), cancer cell migration (Mmp9, Mmp15, S100a14, Mt1), apoptosis (Casp14, Casp2, Bcl6) inflammation, mediated by mechanisms such as detoxification and stress responses (Cyp2c55, Gstk1, Sirt1, Sirt3, Aldh1a1, Aldh1b1, Aldh2, Mt1, Gstm2), the adaptive immune response and intestinal stem cell-
related genes (*Clca4, Aqp4, Aldh1a1, Olfm4*), consistent with a role for *Kcnq1* in the intestinal stem cell compartment (Munoz et al. 2012). Interestingly, while loss of *Kcnq1* showed a strong effect on colon secretory cell genes, there was no change in colon tissues in the secretory cell populations or cell proliferation (Figure 47).

**Kcnq1, Cftr and Muc2 share common genetic pathways in the GI tract**

Ingenuity Pathway Analysis (IPA) identified significant networks and functional groups altered in *Kcnq1* KO mice (see Figures 48A and B, and Table 20 (colon); and Figure 49 and Table 21 (proximal small intestine). In particular, IPA analysis identified Cftr as a top upstream regulator of *Kcnq1* KO dysregulated genes, indicating that genes differentially expressed in the *Kcnq1* KO mice overlap with targets of Cftr (overlap P-value of 5.81E-13, activation Z-score -2.722). Furthermore, the Cftr network is predicted to be downregulated in *Kcnq1* KO small intestine, suggesting that *Kcnq1* may be necessary for some aspects of Cftr function.

To identify common pathways and genes disrupted by *Kcnq1* and *Cftr* deficiencies, we used Gene Set Enrichment Analysis (GSEA). The entire ranked expression data set from the *Kcnq1* KO proximal small intestine array was compared with gene sets compiled from a microarray analysis of gene expression in the small intestine of *Cftr* KO mice (Norkina et al. 2004). *Cftr* gene sets consisted of all genes up- or downregulated by twofold in the *Cftr* KO mouse. GSEA revealed significant correlation, with overlap of gene identity and direction of regulation (Figure 40 and Table 20; normalized enrichment score (NES) for Cftr KO upregulated genes 1.97, P < 0.001; NES for Cftr KO downregulated genes -1.92, P < 0.001). These shared targets included a large subset of genes involved in lipid metabolism and, in particular, downregulation of genes involved in lipid oxidation. In support of the significance of this overlap, IPA analyses of *Kcnq1* KO and *Cftr* KO mice each identified lipid metabolism as a major functional category, with oxidation of lipids as the subcategory showing the most significant down regulation (*Kcnq1*, P-value 9.78E-06, activation Z-score, -2.26; *Cftr*, P-value 9.37E-09, Z-score, -3.62). Common targets included upregulation of *Idi1*, involved in cholesterol synthesis, and downregulation of *Hsd174b* and *Acaa1b*, both involved in
fatty acid beta oxidation. Inhibition of fatty acid oxidation is associated with a switch to lipogenesis and this switch is characteristic of metabolic remodeling that commonly occurs during oncogenesis (DeBerardinis et al. 2008). Also, fatty acid metabolism has recently been shown to have a key role in stem cell maintenance with lipogenesis associated with more highly proliferating stem cells (Knobloch et al. 2013). Thus, altered lipid metabolism may contribute to the development of CRC in Kcnq1 and Cftr KO mice. A second group of genes dysregulated in both Kcnq1 KO mice and Cftr KO mice is involved in inflammation. Because the intestinal epithelium is constantly exposed to flora and foreign substances, it must maintain protective and repair mechanisms. Disruption of these protective or homeostatic processes can result in inflammatory changes, which in turn are associated with inflammatory bowel disease and increased risk of CRC (Duerkop et al. 2009). Genes involved in inflammation are dysregulated in both Kcnq1 KO mice and Cftr KO mice. These include upregulation of factors normally produced by intestinal epithelial cells to maintain immune homeostasis, including Il7 (Shalapour et al. 2010), Reg3g30 and Pglyrp1 (Vaishnava et al. 2011, Dziarski et al. 2010). In addition, a number of protective genes are downregulated including those involved in detoxification, Gsdt1, Cyp3a11, Cyp2b10, and others involved in protective signaling, such as Sema7a, which promotes Il10 signaling (Kang et al. 2012), and Ace2, which protects against inflammation-induced epithelial damage (Hashimoto et al. 2012). Further evidence of overlap between KCNQ1 and CFTR was provided by analyzing RNA Seq data from 20 CRC liver metastases samples for concordance between expression levels of each gene. Using Pearson’s correlation analysis of FPKM (fragments per kilobase pair of exon model per million fragments mapped) values in the 20 samples, it was found that KCNQ1 and CFTR were significantly coexpressed, despite up to a 10-fold sample-to-sample FPKM variability for each gene (see Figures 40A and B). Finally, both KCNQ1 and CFTR are regulated downstream of the serine threonine kinase gene SGK1, which is activated by factors including insulin, via the PI3K pathway, under conditions of osmotic stress (Lang et al. 2013). SGK1 levels in turn appear to be influenced by the activity of both KCNQ1 and CFTR, as Sgk1 expression was upregulated by more than twofold in Kcnq1 KO colon (Table 20) and was also upregulated by > 1.6-fold in Cftr KO colon.
(BLN Than, personal communication). Of further interest, \textit{Sgk1} expression levels were downregulated by 1.5-fold in the proximal small intestine of \textit{Kcnq1} KO mice and downregulated by 5.5-fold in the small intestine of \textit{Cftr} KO mice (Norkina \textit{et al.} 2004).

**Figure 40. Results of GSEA showing enrichment scores (ES) of \textit{Cftr} and \textit{Muc2} gene sets with respect to the ranked \textit{Kcnq1} expression data set.** Shown are enrichment plots for gene sets consisting of (a) genes upregulated in the small intestine of \textit{Cftr} KO mice (Norkina \textit{et al.} 2004); (b) genes downregulated in the small intestine of \textit{Cftr} KO mice (Norkina \textit{et al.} 2004); and (c) genes downregulated in the small intestine of \textit{Muc2} KO mice (Yang \textit{et al.} 2008); (d) NES and nominal P-value (nom. P-value) are shown for each comparison.

CFTR export of $\text{Cl}^-$ and $\text{HCO}_3^-$ maintains extracellular H2O homeostasis, which in turn is necessary for hydration and function of the protective lumenal mucin layer (De Lisle \textit{et al.}, 2011). Another gene directly involved in maintenance of the mucin layer is \textit{MUC2}, which encodes MUCIN2, one of the major glycoproteins making up this layer. As with \textit{Kcnq1} and \textit{Cftr}, \textit{Muc2} deficiency is implicated in CRC (Yang \textit{et al.}, 2008). To test if common pathways were disrupted by loss of \textit{Kcnq1} and \textit{Muc2}, we compared gene expression changes in the \textit{Kcnq1} KO mouse with those reported for \textit{Muc2} KO mice.
(Yang et al., 2008). GSEA identified significant overlap in gene expression, with genes downregulated in Muc2 KO mice enriched in genes downregulated in Kcnq1 KO mice (Figure 40 and Table 22, NES -1.80, P < 0.001). This group included a number of genes involved in detoxification, including cytochrome oxidase P450 enzymes and several glutathione S-transferases, Gsta2, Gstm2, Gstm4, Gstm6 and Gstt1. Moreover, in a separate study (Burger-van Paassen et al., 2012), deficiency for Muc2 in the mouse intestine resulted in significant increases in the innate immune responders Reg3g and Reg3b, matching our findings in Kcnq1 KO mice. Finally, we found that Areg, an epidermal growth factor receptor pathway gene that was upregulated in both the colon and proximal small intestine of Kcnq1 KO mice, was also upregulated by >10-fold in the colon of Muc2 KO mice (RTC and RJAF, unpublished microarray results) and in the colon of Cftr KO mice (BLNT, manuscript in preparation) (see Table 23).

**Loss of Kcnq1 promotes colon organoid development.**

To investigate a role for Kcnq1 in the intestinal stem cell compartment, we created intestinal organoids from the colons of a group of Apc<sup>+/+</sup> Kcnq1<sup>+/+</sup> and Kcnq1<sup>-/-</sup> age- and gendermatched littermate mice. At days 2–5, organoid growth was monitored and organoid number counted on day 5. Loss of Kcnq1 increased the number of colon organoids by greater than threefold, confirming a potential regulatory effect of Kcnq1 on the intestinal stem cell compartment (Munoz et al., 2012) (Figure 41).
Figure 41. Colon organoids. Colon organoids were created from 500 crypt bottoms and plated in triplicate in 24-well plates. Crypt bottoms were isolated on separate days from 5 of each, age-, gender- and littermate-matched *Apc*<sup>+/+</sup> *Kcnq1<sup>+/+</sup> and *Apc*<sup>+/+</sup> *Kcnq1<sup>−/−</sup> mice (one matched pair per day). Organoids were examined daily and counted at 5 days post plating. *P < 0.05.

KCNQ1 expression is associated with good prognosis in late-stage human CRC.

On the basis of our finding that deficiency for *Kcnq1* drove the progression of intestinal tumors in mice, we tested the hypothesis that low levels of KCNQ1 in human CRC would correlate with a worse prognosis. To this end, we evaluated KCNQ1 protein expression in tissue microarrays containing samples from more than 500 stage IV CRC patients with liver metastases who had undergone hepatic resection. Immunohistochemical staining for KCNQ1 expression was evaluated for 311 patients. Membranes of neoplastic CRC liver metastasis (CRCLM) epithelium were scored for KCNQ1 staining intensity (Figure 42). We found that high KCNQ1 expression was related to improved overall survival (OS) (hazard rate ratio (HRR) 0.64; 95% confidence interval (CI): 0.45–0.92; P = 0.02). Median OS for patients with high KCNQ1 expression
was 60 months, whereas patients with low KCNQ1 levels had a median OS of 37 months, indicating a significant 23-month difference in survival (Figure 43).

Figure 42. Representative human samples showing the expression pattern of KCNQ1 in the epithelium of CRCLM. Staining intensity was evaluated as (a) high or (b) low. Scale bars are 50 µm.
Next, we performed a multivariate analysis to evaluate whether the prognostic value of KCNQ1 protein expression was independent of established prognostic clinicopathological variables, that is, primary tumor-to-liver metastasis interval > 12 months, number of liver metastases > 1, maximal tumor diameter > 5.0 cm, lymph node positivity at the time of diagnosis of the primary tumor and serum carcinoembryonic level > 200 ng/ml. Upon stepwise backward Cox regression analysis with OS as the dependent variable, KCNQ1 expression (HRR, 0.49; 95%: 0.32–0.77; P = 0.002), lymph node positivity at the time of diagnosis of the primary tumor (HRR, 1.64; 95% CI: 1.08–2.48; P = 0.02) and maximal tumor diameter > 5.0 cm (HRR, 1.44; 95% CI: 0.96–2.16; P = 0.08) were retained as prognostic variables. Note that the association between maximal tumor diameter and survival was not significant; however, the variable was retained in the model because we only excluded variables when P > 0.1.

**DISCUSSION**

We first identified *Kcnq1* as a potential CRC driver gene in our forward genetic screens for GI tract cancer genes. Here, we have followed up with studies using a transgenic mouse model and analysis of human CRC tissues that support a tumor suppressor role for *Kcnq1*. We demonstrated the tumor suppressor function of *Kcnq1* in Apc\(^{Min}\) mice, where both hetero- and homozygous inactivation of *Kcnq1* resulted in significantly more adenomas and, importantly, progression to adenocarcinoma. This was especially evident in the proximal small intestine. Furthermore, Kcnq1 deficiency caused rectal adenomatous hyperplasia, pyloric tumorigenesis, gastric hyperplasia and pancreatic abnormalities. Notably, in human CRC that had metastasized to the liver, low KCNQ1 protein expression correlated with significantly lower OS, by almost 2 years, compared to patients with high KCNQ1 expression. These findings suggest that KCNQ1 expression could be a biomarker to help determine eligibility for hepatic resection in CRC patients with liver metastases. In addition, newly developed small molecules that enhance K\(^+\) ion
channel activity (Xiong et al., 2008, Wulff et al., 2009) could potentially be used to treat CRC patients with low levels of KCNQ1.

By what mechanism does loss of KCNQ1 influence cancer processes in the GI tract? A promising model is via KCNQ1’s functional interactions with CFTR and MUC2. In the healthy intestinal epithelium, KCNQ1 is physiologically linked to the CFTR ion channel, with basolateral export of $K^+$ ions by KCNQ1 providing the electrochemical driving force for apical export of $Cl^-$ ions by CFTR (Preston et al., 2010). Cftr was identified as a CRC driver gene in two Sleeping Beauty GI tract forward genetic screens (Starr et al., 2009, March et al., 2011), and individuals with cystic fibrosis, caused by inactivating mutations in CFTR, are at increased risk for GI cancers including CRC (Maisonneuve et al., 2003). To test if common pathways and genes are disrupted by loss of Kcnq1 and Cftr activity, we used GSEA to compare gene sets consisting of genes reported to be up- or downregulated in the small intestine of Cftr KO mice to the entire ranked Kcnq1 expression data set and found significant correlation (Figure 40). Genes upregulated in common include several involved in immune responses and inflammation, whereas genes downregulated in common include a subset involved in fatty acid metabolism, in particular lipid oxidation. Further evidence of overlap between KCNQ1 and CFTR was the finding of concordance between KCNQ1 and CFTR expression in RNA Seq analysis of 20 CRC liver metastases (Figures 50A and B).

As with Kcnq1 and Cftr, Muc2 deficiency is implicated in CRC (Yang et al., 2008). To test if common pathways were disrupted by loss of Kcnq1 and Muc2, we compared gene expression changes in the Kcnq1 KO mice with those reported for Muc2 KO mice. GSEA identified significant overlap in gene expression with genes downregulated in Muc2 KO mice enriched in genes downregulated in Kcnq1 KO mice (Figure 40). This group included a number of genes involved in detoxification including cytochrome oxidase P450 enzymes and several glutathione S transferases, and genes involved inflammation.

These results point to common pathways (inflammation, lipid metabolism, detoxification and stress responses) disrupted by Kcnq1, Cftr and Muc2 deficiencies. Therefore, we can hypothesize that KCNQ1, like MUC2 and CFTR, may act through one
or more of these pathways in preventing cancer in the GI tract. Moreover, precisely defining mechanisms of action for KCNQ1 can lead to its use as a prognostic predictor for CRC patients and potential therapeutic target.

**MATERIALS AND METHODS**

**Mice**

C57BL/6J mice and C57BL/6J-\(Apc^{Min}\) mice were obtained from the Jackson Laboratory (Bar Harbor, ME, USA). C57BL/6-\(Kcnq1\) knockout mice were obtained from Dr Karl Pfeifer (NIH, Bethesda, MD, USA) (Casimiro et al., 2001). Details of mouse husbandry are as described previously (Fijneman et al., 2012). The genotype of the \(Apc\) and \(Kcnq1\) loci were determined by PCR assays as described previously (Casimiro et al., 2001, Fijneman et al., 2012).

**Tumor analysis**

Scoring of GI tumor tissues was performed as described previously (Fijneman et al., 2012). Two-sided P-values for tumor counts were determined by use of the Wilcoxon rank-sum test comparing gender- and age-matched classes produced in the same genetic crosses.

**Histopathology**

Histopathological analysis of tumors and adjoining normal tissue was performed on formalin-fixed paraffin-embedded tissues by an ACVPcertifie veterinary pathologist (MGO0S) from the University of Minnesota Masonic Cancer Comparative Pathology Shared Resources facility using protocols as described previously (Fijneman et al., 2012).

**Organoid culture**

Gender- and age-matched littermate C57Bl/6J \(Kcnq1^{+/+}\) and \(Kcnq1^{-/-}\) mice were killed between 8 and 12 weeks of age. Colons were removed, cut open and washed in cold phosphate-buffered saline. Colon organoids were then cultured using the protocol of Sato et al. (Sato et al., 2013) following the plating of 500 crypt bottoms per well in triplicate per sample. Two-sample t-test was used to determine statistical significance.

**RNA processing for Illumina bead arrays**
Mouse intestinal tissues were removed, opened longitudinally and rinsed in phosphate-buffered saline. Colon tissues were processed in RNALater (Qiagen, Valencia, CA, USA) as per the manufacturer’s protocol. The distal quarter of the proximal quarter of the small intestine was flash frozen in liquid nitrogen. The tissue was then transferred from liquid nitrogen to precooled RNALaterIce (Ambion, Foster City, CA, USA) at -80°C for 30 min, and later transferred to -20°C for at least 48h. Immediately before RNA isolation, tissue in either RNALater or RNALaterIce was placed in 2–4 ml of RLT (guanidine thiocyanate buffer)/14.3 M b-mercaptoethanol buffer. The tissue was then homogenized using an IKA Ultra-Turrax T25 digital homogenizer (Fisher Scientific, Waltham, MA, USA). RNA isolation was then conducted using an RNeasy Mini Kit with an additional DNase Digestion step (Qiagen). RNA sample concentrations were measured using a Nanodrop-1000 Spectrophotometer (Thermo Scientific, Rochester, NY, USA). RNA was stored at -80°C.

**Illumina bead microarray and data analysis**

RNA labeling, microarray hybridization and scanning were performed at the University of Minnesota BioMedical Genomics Center using Illumina MouseWG-6 v.2.0 Expression BeadChips (Illumina, San Diego, CA, USA), according to the manufacturer’s instructions. Differential expression of genes was quantified by the moderated t-statistic. Data were analyzed using GeneData Expressionist Software (GeneData Inc., San Francisco, CA, USA); genes that showed statistically significant differences in expression between groups were determined using the two-group t-test and analysis of variance. Gene expression data has been submitted to the Gene Expression Omnibus, and accession number is pending.

**qRT–PCR**

Quantitative RT–PCR was performed as described previously (Starr et al., 2011). Significant differences in expression between groups were determined using the two-group t-test.

**List of primers used:**

All primers were designed using Primer-BLAST (National Center for Biotechnology Information) and obtained from Integrated DNA Technologies.
Mouse primers

Ang4: forward, 5’-AGCACACAGCTAGACTCGTCCC-3’; reverse, 5’-ACCAGACCCAGCAGACACAGACC-3’; Areg: forward, 5’-GCTTTAGGCTCAGGCGCCATTA-3’; reverse, 5’-CGCTTTAGGCTTGGGAAACCTCTC-3’; Cd55: forward, 5’-GGGGCTATCCGTGGGCG-3’; reverse, 5’-TGCCCAAGATGGGCTGGCA-3’; Clca3: forward, 5’-CCACACAAACGAGAAGGC-3’; reverse, 5’-TGCTTTAGGCTTGGGAAACCTCTC-3’; Ang4

Areg: forward, 5’-GGTCTTAGGCTCAGGCCATTA-3’; reverse, 5’-CGCTTTAGGCTTGGGAAACCTCTC-3’; Cd55: forward, 5’-GGGGCTATCCGTGGGCG-3’; reverse, 5’-TGCCCAAGATGGGCTGGCA-3’; Clca3: forward, 5’-CCACACAAACGAGAAGGC-3’; reverse, 5’-TGCTTTAGGCTTGGGAAACCTCTC-3’; Ang4

Cd55: forward, 5’-GGGGCTATCCGTGGGCG-3’; reverse, 5’-TGCCCAAGATGGGCTGGCA-3’; Clca3: forward, 5’-CCACACAAACGAGAAGGC-3’; reverse, 5’-TGCTTTAGGCTTGGGAAACCTCTC-3’; Ang4

Clca3: forward, 5’-GGGGCTATCCGTGGGCG-3’; reverse, 5’-TGCCCAAGATGGGCTGGCA-3’; Clca3: forward, 5’-CCACACAAACGAGAAGGC-3’; reverse, 5’-TGCTTTAGGCTTGGGAAACCTCTC-3’; Ang4

Gstk1: forward, 5’-TGGATGCGTGTATGGTCTCG-3’; reverse, 5’-CAGAAAGTGTTGGCGTTCG-3’; Egr1: forward, 5’-CCTCAAGGGGAGCCGAGCG-3’; reverse, 5’-AACCGAGTGGCTTTGGCTGGG-3’; Fos: forward, 5’-CGGGTTTCAACGCGACTA-3’; reverse, 5’-TTGGCACTAGAGACGGACAGA-3’; Id1: forward, 5’-CTCAGCACCCTGAACCGCGA-3’; reverse, 5’-CATCTGGGTCCCTACACCAAGGC-3’; Muc2: forward, 5’-AACGATGCCTACACCAAGGTC-3’; reverse, 5’-ACTGAACCTGATGCGTCCCTACACCAAGGTC-3’; Muc3: forward, 5’-ACTGAACCTGATGCGTCCCTACACCAAGGTC-3’; reverse, 5’-GCCT CCATTCTCAGTTCGCGCC-3’; Muc13: forward, 5’-TTTGGCTACAGCGGGATGAA-3’; reverse, 5’-TCTTTGACCTCGAGACGAC-3’; Mt1: forward, 5’-AAAC CCTTTGCGCCCGGACT-3’; reverse, 5’-AGCAGAGCGAGTGGGCGGCC-3’; Pim3: forward, 5’-GCTCATCGACTTCCGCTGGGGGTCC-3’; reverse, 5’-AGTGGCAGACCGCCCGTGATA-3’; and Retnlb: forward, 5’-AAGCCTACACTGTTTGCTTCCTCCTCCTCCT3’; reverse, 5’-GCTTCCTTGTACCTGTTGCCAC-3’.

Pathway analysis

GSEA. GSEA v.2.0, http://www.broad.mit.edu/gsea/ (Subramanian et al., 2005, Mootha et al., 2003). Gene sets analyzed were obtained from published reports and the National Center for Biotechnology Information Gene Expression Omnibus database. GSEA compared the entire ranked Kcnq1 proximal small intestine KO expression data set to Cfr KO25 (twofold change in expression, t ≤ 0.05) and Muc2 KO36 (1.5-fold change in the duodenum at 3 or 6 months of age).
IPA. Genes showing a 1.5-fold change in expression in whole colon (N = 89) and proximal small intestine (N = 345) microarrays were analyzed using IPA (IPA Ingenuity Systems).

Tissue microarrays

Patient study population. Patients who underwent liver resection with curative intent, sometimes with addition of RFA, in one of the seven Dutch hospitals affiliated with the DeCoDe PET group were identified by crossreferencing surgery and pathology databases using Dutch MeSH terms for ‘colon’, ‘rectum’, ‘carcinoma’, ‘adenocarcinoma’, ‘colorectal neoplasms’, ‘liver’, ‘neoplasm metastasis’ and ‘(hemi)hepatectomy’. Clinicopathological data from 507 patients that were operated on between 1990 and 2010 were extracted from these databases (see Supplementary Table 22; a manuscript describing the patient population in more detail is in preparation). Formalin-fixed paraffin-embedded tissue specimens were collected from one CRCLM sample and an adjacent control liver sample. Only specimens of patients with histologically confirmed CRCLM were included in the study, whereas tissue samples of patients with multiple primary tumors were excluded. Collection, storage and use of clinicopathological data and tissue specimens were performed in compliance with the ‘Code for Proper Secondary Use of Human Tissue in The Netherlands’, and approved in protocol 2011-03 of the Department of Pathology (Stichting FMWV Rotterdam 2011).

Tissue microarrays. A total of 21 tissue microarrays (TMAs) were generated as described previously (Simon et al., 2004, Belt et al., 2011). In brief, three tissue core biopsies of 0.6mm in diameter were punched from morphologically representative areas of all formalin-fixed paraffin-embedded donor blocks, and transferred into TMA recipient paraffin blocks using the 3DHISTECH TMA Master (v.1.14, 3DHISTECH Ltd, Budapest, Hungary).

Immunohistochemistry. Four-micrometer sections of TMAs were mounted on glass slides, deparaffinized by xylene and rehydrated with a decreasing alcohol series. Staining for KCNQ1 was performed upon antigen retrieval by microwave heating in citric acid (10mM, pH 6.0) and endogenous peroxidase neutralization in 0.3% hydrogen peroxide in methanol for 25 min. The primary rabbit polyclonal antibody directed against human
KCNQ1 (sc-20816; Santa Cruz Biotechnology Inc., Santa Cruz, CA, USA) was incubated overnight at a 1:200 dilution at 4°C, followed by incubation with anti-rabbit secondary antibodies for 30 min at room temperature (Envision Plus; Dako, Heverlee, Belgium). Secondary antibodies were visualized by liquid diaminobenzidine substrate chromogen system. Slides were counterstained with Mayer’s hematoxylin. Staining of formalin-fixed paraffin-embedded colon tissue was used as a positive control, and incubation without primary antibody as a negative control.

**Evaluation of KCNQ1 protein expression.** Immunohistochemical stainings were digitally captured using the Mirax slide scanner system equipped with a x20 objective with a numerical aperture of 0.75 (Carl Zeiss BV, Sliedrecht, The Netherlands) and a Sony DFW-X710 Fire Wire 1/3 in-type progressive SCAN IT CCD (pixel size 4.65 x 4.65 µm²). Actual scan resolution at x 20 was 0.23 µm. Computer monitors were calibrated using Spyder2PRO software (v.1.0–16; Pantone Colorvision, Regensdorf, Switzerland). TMA core biopsies were scored for intensity of KCNQ1 protein expression on membranes of neoplastic epithelial cells (categories negative, weak, moderate, strong) using dedicated TMA scoring software (v.1.14.25.1; 3DHISTECH Ltd). For facilitating scoring, a chart with visual analogue scales of staining patterns was used. Intensity scores were dichotomized into low and high KCNQ1 intensities using receiver operating characteristic curve analysis (Zlobec et al., 2007).

**Statistical analysis of TMAs.** OS was defined as the time in months after surgery until death in a follow-up period of 10 years. Patients were excluded from analysis if OS < 2 months (n = 72), if OS or survival status were unknown (n = 36) or if tissue cores could not be evaluated for technical reasons (n = 88). The relation between KCNQ1 expression and survival was studied using Kaplan–Meier curves and tested using log-rank statistics. HRR was calculated using Cox regression. Multivariate analysis was performed by inclusion of KCNQ1 protein expression in a proportional hazards (Cox regression) analysis together with established prognostic clinicopathological variables combined in the clinical risk score as defined by Fong et al. (Fong et al., 1999) primary tumor-to-liver metastasis interval > 12 months, number of liver metastases > 1, maximal tumor diameter > 5.0 cm, lymph node positivity at the time of diagnosis of the primary tumor, serum
carcinoembryonic level > 200 ng/ml. Stepwise backward regression was used to exclude variables from the model when \( P > 0.1 \). All statistical tests were two-sided and executed using IBM SPSS Statistics 20.0 software (SPSS Inc., Chicago, IL, USA). P-values < 0.05 were considered significant.

**Supplemental Information:**
**Figure 44. Rectal Adenomatous Hyperplasia. Panel A.** Gross image of rectum. Lines represent locations of cross sections taken for histopathological analysis. Panels B, C & D, three cross sections of the same tissue showing diffuse crypt hyperplasia. The incidence of abnormal rectums was 10 out of 102 $\text{Apc}^{\text{Min}} \text{Kcnq1}^{+/−}$ (79) and $\text{Apc}^{\text{Min}} \text{Kcnq1}^{−/−}$ (23) mice and 0 out of 58 for $\text{Apc}^{\text{Min}} \	ext{Kcnq1}^{+/+}$ mice, a difference that is highly significant ($p = 0.009$, Fisher’s Exact Test).
Figure 45. Stomach of Apc\textsuperscript{Min} Kcnq1\textsuperscript{-/-} mice. Panels A & B, 10X; Panel C, 20X. Severe gastric hyperplasia that was characterized by mucosal hyperplasia, with parietal cells increased in size and crypt dilatation with debris present. Hyperplastic adenomas developed in the pylorus (not shown). The incidence of pyloric tumors was 21 out of 23 Apc\textsuperscript{Min} Kcnq1\textsuperscript{-/-} mice and 0 out of 58 Apc\textsuperscript{Min} Kcnq1\textsuperscript{+/+} mice, a difference that is highly significant (p < 2.2e-16, Fisher’s Exact Test).

Figure 46. Pancreas. 20X. Kcnq1 Wildtype, left panels; Apc\textsuperscript{Min} Kcnq1\textsuperscript{-/-} right panels. Zymogen granules were less eosinophilic and less prominent. Exocrine cells seemed to be smaller and islets appeared to be less prominent (arrows). The incidence of enlarged pancreases was 8 out of 23 Apc\textsuperscript{Min} Kcnq1\textsuperscript{-/-} mice and 0 out of 58 in Apc\textsuperscript{Min} Kcnq1\textsuperscript{+/+} mice, a difference that was high significant (p = 1.52e-05, Fisher’s Exact Test). The maximum diameter of a total of 56 pancreatic islets was measured, 28 from Apc\textsuperscript{Min} Kcnq1\textsuperscript{-/-} mice and 28 from Apc\textsuperscript{Min} Kcnq1\textsuperscript{+/+} mice. The mean diameter of islets from Apc\textsuperscript{Min} Kcnq1\textsuperscript{-/-} mice was 100 microns and the mean diameter of islets from Apc\textsuperscript{Min} Kcnq1\textsuperscript{+/+} mice was 118 microns. While the difference in means was not significant, when the 56 islet diameters were ranked in size the top 7 islets were all from the Kcnq1\textsuperscript{+/+} group (246, 242, 212, 210, 199, 184, and 181 microns), while the largest islet from the Kcnq1\textsuperscript{-/-} group was 173 microns.
Figure 47. Immunohistochemistry for cellular markers in mouse colon. Age and gender matched *Kcnq1*+/+ and *Kcnq1*−/− adult littermate mice were sacrificed and FFPE tissues processed as previously described (Fijneman et al., 2012). Multiple sections were examined by immunostaining for the goblet cell markers Mucin2 and alcian blue (for detection of acidic mucopolysaccharides), and the proliferation marker Ki67. No significant differences were detected between genotypes. Representative colon section is shown above. Mucin2: Panel A (*Kcnq1*+/+), Panel B (*Kcnq1*−/−); Alcian Blue: Panel C (*Kcnq1*+/+), Panel D (*Kcnq1*−/−); Ki67: Panel E (*Kcnq1*+/+), Panel F (*Kcnq1*−/−). Methods. 4 µm formalin-fixed, paraffin-embedded sections of small intestine were deparaffinized and rehydrated, followed by antigen retrieval using 10mM Citrate buffer pH 6.0 in a steamer. A rabbit polyclonal antibody was used for Mucin 2, Catalog number: SC15334, (Santa Cruz Biotechnology). A rabbit monoclonal antibody was used for Ki67, Catalog number: CRM325, clone SP6, (Biocare Medical). Detection was achieved using a rabbit EnVision™+ Kit (catalog K4011, Dako) with DAB as chromogen. Alcian blue staining was performed using a standard protocol.
Figure 48. Colon Microarray - Top Functional Networks Identified by IPA

A. Cancer, Cellular Function and Maintenance, Molecular Transport

B. Immunology & Inflammatory Disease

Figure 49. Proximal Small Intestine Microarray - Top Gene Networks Identified by IPA.
Figure 50. Concordance between expression of KCNQ1 and CFTR in colorectal cancer liver metastases. Twenty stage IV CRC liver metastases were analyzed by RNA seq, followed by Pearson’s Correlation analysis to determine concordance between the expression levels of KCNQ1 and CFTR. Panel A. Using all twenty samples. Panel B. removal of two top and bottom outliers, N = 18.
RNA Seq analysis. Twenty fresh frozen stage IV CRC LM samples were obtained from the University of Minnesota BioNet Tissue Procurement facility. The samples were from 9 males and 11 females, average age 64 years. RNA was isolated using RNA LaterICE (Invitrogen) and Qiagen RNAEasy kits, and quantified using a Nanodrop-1000 Spectrophotometer (Thermo Scientific), and quantity and quality were confirmed using The extracted RNA-samples were sent to the University of Minnesota Genomics Center (UMGC) for sequencing. All raw reads that passed the CASAVA 1.8 P/F filter were directly deposited at the Minnesota Supercomputing Institute, where the data were
analyzed. Sequence quality was assessed via fastqc (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Since the raw sequence reads were of sufficient quality, no data trimming or filtering was necessary. Hence, the raw paired-end reads were mapped to human genome (hg19 assembly) using Tophat 2.0.6 with the iGenomes reference (UCSC human hg19 annotation, downloaded from http://cufflinks.cbcb.umd.edu/igenomes.html March 9, 2012). Final mapping percentages of the 20 samples range from 92% to 94%. The mapped samples were then analyzed via Cuffdiff (default parameters using the iGenomes UCSC hg19 GTF file previously mentioned) to quantify the expression level of each known gene in units of FPKM (Fragment mapped per kilobase of exons per million mapped reads). Additional statistical test of Pearson's correlation from R/Bioconductor was applied to selected gene pairs to check for expression concordance.

Table 18. List of top known genes 1.5 fold (A) up-regulated and (B) down-regulated in Kcnq1−/− mouse colons, represented as fold change vs control Kcnq1+/+. The asterisk in the Fold Change column indicates a P value of 0.05 or less.

Table 19. List of top known genes 1.5 fold (A) up-regulated and (B) down-regulated in Kcnq1−/− mouse small intestines, represented as fold change vs control Kcnq1+/+. The asterisk in the Fold Change column indicates a P value of 0.05 or less.

Table 20. IPA Analysis of Colon Microarray.

Table 21. IPA Analysis of Proximal Small Intestine Microarray.

Table 22. Core genes enriched in Kcnq1 KO.

Table 23. Expression of Areg.

Table 18-23 can be found in the supplemental tables. Please see the weblink:
http://www.nature.com/onc/journal/vaop/ncurrent/full/onc2013350a.html
CONFLICT OF INTEREST
The authors declare no conflict of interest.

ACKNOWLEDGEMENTS
We thank Dr Karl Pfeifer at the NIH for generously providing Kcnq1 mutant mice and reviewing the manuscript. Research was supported by a grant to RC and DL (NCI R01 CA134759-01A1) and to JG and RF (Center for Translational Molecular Medicine, DeCoDe project grant 03O-101) and to TKS (NCI R00 4R00CA151672-02).

Personal Contribution. Lei Zhao isolated RNA samples for Illumina RNA Sequencing and analyzed data.
APPENDIX 3

The roles of CNOT1 and CNOT2 in regulating mRNA levels of $p27^{kip}$ and CDX2

Introduction

Sleeping Beauty (SB) transposon-based screens have identified Cnot1 and Cnot2, two subunits of the CNOT complex, as candidate CRC genes (Starr et al., 2009). (See introduction for discussion of CNOT1). The TCGA project identified 48 somatic mutations in CNOT2 in different types of cancers and 3 of them were discovered in CRC. In addition, 47 somatic mutations in CNOT2 were documented in the COSMIC database and 76% of them were missense mutations. The presence of somatic DNA mutations in CNOT2 provided evidence for a correlation between genetic alterations in CNOT2 and CRC. The biological importance of CNOT2 has been reported in several model animals. It is critical for yeast survival (Russell et al., 2002). Apoptosis may be occurring in CNOT2-depleted C. elegans and D. melanogaster during early embryonic stages (Frolov et al., 1998; Sonnichsen et al., 2005). In Hela cells CNOT2 depletion caused apoptotic cell death in a caspase-dependent manner (Ito et al., 2011). As for the molecular function of CNOT2, accumulating results suggest that it plays a role in regulation of deadenylation as a component of the CNOT deadenylase complex and/or in transcriptional regulation (Zwartjes et al., 2004). CNOT2 plays an important role in maintaining the integrity of the CNOT complex, supported by the fact that CNOT2 mutations affect the structural integrity of the complex in yeast (Russell et al., 2002). Mammalian CNOT2 depletion also destabilized the CNOT complex and impaired its deadenylase activity. Ito et al showed that the disruption of the CNOT complex induced by CNOT2 depletion is accompanied by a loss of deadenylase activities (Ito et al., 2011). Furthermore, knockdown of CNOT2 inhibits mRNA decay in Drosophila S2 cells (Temme et al. 2004). These results suggest that the structural integrity of the complex is apparently important
for full deadenylase activity. However, the exact mechanism by which CNOT2 contributes to the regulation of the deadenylase activity of the CNOT complex remains obscure. CNOT2 also acts as a transcriptional repressor of Estrogen Receptor (ER)-mediated gene expression through direct repression of pol II transcription. The major repressive function of CNOT2 is localized in the Not-Box repression domain in the C-terminus of CNOT2 (Zwartjes et al., 2004). Depletion of deadenylase subunits of the CNOT complex was reported to affect cell viability (Ito et al., 2011; Morita et al., 2007), suggesting that the deadenylase activity of the CNOT complex is important for regulation of cell viability. Thus attenuated deadenylation activity of the CNOT complex may contribute to the decreased cell viability caused by CNOT2 depletion.

Cyclin-dependent kinase inhibitor 1B (CDKN1B, p27^Kip1^) is a well-known cell cycle inhibitor which prevents the activation of cyclin E-CDK2 or cyclin D-CDK4 complexes, and thus arrest cell cycle progression at G1. Cnot6L was shown to contribute to the deadenylation and degradation of p27^Kip1^ mRNA through a 3'UTR ARE-mediated mechanism (Morita et al., 2007). The gut-specific homeotic transcription factor caudal type homeobox 2 (CDX2) is a crucial regulator of intestinal development and homeostasis, which exhibits a tumor suppressor function in CRC (Saandi et al., 2013). Cnot1 and Cnot2 silencing was shown to lead to markedly increased expression of Cdx2 in mouse embryonic stem cells (Zheng et al., 2012). These data suggest that p27^Kip1^ and CDX2 are potential target genes of CNOT1- and CNOT2-mediated deadenylation or transcriptional regulation.

In our study, to investigate the role of CNOT1 and CNOT2 in CRC, we measured cell viability following depletion of CNOT1 and CNOT2 in SW480 CRC cells. We then studied the role of CNOT1- and CNOT2-mediated deadenylation in regulating p27^Kip1^ and CDX2 mRNA levels and the contribution of these two genes to CNOT1 and CNOT2 depletion caused cell viability reduction.

**Results**

**Depletion of the CNOT Complex Subunits Results in Reduced Cell Viability in SW480 Cells**
As shown in Figure 24 D, depletion of CNOT2 resulted in decreased cell viability in SW480 CRC cells, suggesting that CNOT2 may function as an oncogene. In addition, cell viability assay revealed that depletion of CNOT6L, CNOT7 and CNOT8 caused significant decrease in cell viability (Figure 24 D), suggesting that CNOT1 and CNOT2’s effects on cell viability may be mediated by deadenylation activity of these deadenylase subunits.

*p27<sup>kip</sup>* and *CDX2* mRNA Levels Were Upregulated in SW480 Cells Depleted for *CNOT1* and *CNOT8*

The CNOT deadenylases were reported to mediate deadenylation and subsequent degradation of cell cycle inhibitor *p27<sup>kip</sup>* and the intestinal differentiation factor *CDX2*. Consistent with published data, we found that *p27<sup>kip</sup>* (Figure 51A) and *CDX2* (Figure 51B) mRNA levels were increased by 1.5 and 2.5-fold, respectively, in *CNOT1*-depleted but not changed in *CNOT2*-depleted cells. To determine which deadenylase subunit shares a common phenotype with CNOT1 we examined the effect of their depletion on levels of *p27<sup>kip</sup>* (Figure 51A) and *CDX2* (Figure 51B) mRNA levels. Lack of CNOT8 caused a 1.5- and 2-fold increase in *p27<sup>kip</sup>* and *CDX2* mRNA levels, respectively. But depletion of other deadenylase subunits did not cause an obvious change in *p27<sup>kip</sup>* mRNA levels. Interestingly, depletion of *CNOT6* and 6L both caused decreased *CDX2* mRNA levels possibly through indirect mechanisms. These results suggest that CNOT1 and 8 may works in complexes to regulate the mRNA levels of *p27<sup>kip</sup>* and *CDX2* through deadenylation.
Figure 51. $p27^{kip}$ and CDX2 mRNA expression levels in SW480 cells depleted for CNOT subunits. SW480 cells were depleted for CNOT1, 2, 6, 6L, 7 and 8, respectively as described before. Total RNA was extracted at 48 hours after the second siRNA transfection. mRNA expression levels of $p27^{kip}$ (A) and CDX2 (B) were measured by qRT-PCR. Values represent the mean ± SD of at least two independent experiments except for CNOT6 (one experiment). Statistical significance was determined by a t-test. * = p <0.05.
Effects of Depletion for RBPs on Cell Viability and $p27^{kip}$ and CDX2 mRNA Levels

To test if 3’UTR RNA binding proteins (RBPs) might mediate the phenotype shared by CNOT1 and CNOT8 we examined the effect of depletion of several RBPs on cell viability and $p27^{kip}$ and CDX2 mRNA levels, including ZFP36, PUM1, PUM2 and CELF1. They were knocked down in SW480 cells by transfection of corresponding siRNA oligos, which was confirmed by qRT-PCR (Figure 52A). Cell viability was reduced in cells depleted for ZFP36, PUM1 and CELF1, but not in cells depleted for PUM2 (Figure 52A). mRNA levels of $p27^{kip}$ were increased by 1.35 fold in ZFP36-depleted cells but not in other cells (Figure 52B). CDX2 mRNA levels were increased by 1.5 fold in ZFP36 and CELF1-depleted cells but not in other cells (Figure 52C). Together, these results suggest that ZFP36 and CELF1 may work with CNOT1 and 8 in complexes to regulate cell viability and mRNA levels of $p27^{kip}$ and CDX2 through deadenylation.

A.

<table>
<thead>
<tr>
<th>Gene</th>
<th>mRNA levels</th>
<th>Cell viability</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZFP36</td>
<td>74.07±17.04%</td>
<td>57.14±9.21%</td>
</tr>
<tr>
<td>PUM1</td>
<td>60.03%</td>
<td>42.90%</td>
</tr>
<tr>
<td>PUM2</td>
<td>15.76%</td>
<td>92.67%</td>
</tr>
<tr>
<td>CELF1</td>
<td>46.16%</td>
<td>63.77%</td>
</tr>
</tbody>
</table>

B.
C.

Figure 52. Cell viability and $p27^{kip}$ and CDX2 mRNA expression levels in SW480 cells depleted for RBPs. SW480 cells were depleted for ZFP36, PUM1, PUM2 and CELF1 by siRNA transfection as described before. (A) On day 1 after the second transfection SW480 cells were plated at 2,500 cells per well in triplicate in 96 well plates. Cell viability was measured by MTT assay on day 6 after the second transfection. Relative cell viability represents = absorbance at 595 nm of cells treated with indicated siRNA/cells treated with non-targeting control. Results represent one experiment except for ZFP36 (3 experiments). (B and C) Total RNA was extracted 48 hours after the second transfection and $p27^{kip}$ (B) and CDX2 (C) mRNA expression levels were determined by qRT-PCR. Values represent one experiment except for ZFP36 (mean ± SD of 3 experiments). Significance of difference was calculated by a t-test. * = p <0.05.

Depletion of CDX2 Did Not Result in Changes in Cell Viability in SW480 Cells

To initially investigate whether the decreased cell viability caused by knockdown of CNOT1, CNOT8, ZFP36 and CELF1 was due to increased levels of CDX2, we depleted CDX2 by transfection of siRNA oligos targeting CDX2 in SW480 cells and measured cell viability using the MTT assay. Knockdown of CDX2 was confirmed by qRT-PCR (Figure 53A). No obvious difference in cell viability was observed in CDX2 siRNA transfected cells comparing to control cells (Figure 53B). These results suggest that upregulated CDX2 may not be the cause for the reduction in cell viability. However, this is the data from only one experiment which was performed in highly malignant SW480 cells. Therefore, more assays need to be performed, including overexpression of
CDX2 and knocking down CDX2 in normal colon epithelial cell lines, before we can draw any conclusions.

A. 

Figure 53. Depletion of CDX2 did not result in changes in cell viability in SW480 cells. SW480 cells were depleted for CDX2 by transfection of siRNA oligo targeting CDX2 at 50 nM final concentration twice 48 hours apart. Knockdown of CDX2 was confirmed by qRT-PCR (A) and cell viability (B) was measured by MTT assay. Data represents result of only one experiment. Each data point for MTT assay represents the mean ± SD of three replicates in one MTT assay. Significance of differences in cell viability on day 6 was determined by one-way ANOVA.

CNOT1 Does Not Regulate p27kip and CDX2 mRNA Levels Through 3’UTR-Mediated Deadenylation

To investigate whether CNOT1 regulates p27kip and CDX2 mRNA levels through 3’UTR-mediated deadenylation we performed a 3’UTR luciferase reporter assay in SW480 cells using a Dual-Glo reporter assay system (Promega). The 3’UTR of p27kip and CDX2 was cloned downstream of a sequence encoding renilla luciferase in a psi-Check2 vector where activity of renilla luciferase is indicative of 3’UTR mediated deadenylation. This vector also constitutively encodes the firefly luciferase to normalize cell number and transfection efficiency. Results revealed that transfection of CNOT1 siRNA did not cause a change in luciferase activity of p27kip (Figure. 54A) and CDX2 (Figure 54B) 3’UTR reporter plasmids compared with cells transfected with control
siRNA. This result suggested that CNOT1 might regulate mRNA levels of these two genes through mechanisms other than deadenylation. However, the data represent the result of only one experiment. More replicates and more assays need to be done before we can draw any conclusions.

Figure 54. CNOT1 does not regulate $p27^{kip}$ and $CDX2$ mRNA levels through 3’UTR-mediated deadenylation. A luciferase reporter assay was performed in SW480 cells to investigate whether CNOT1 regulates $p27^{kip}$ and $CDX2$ level through 3’UTR-mediated deadenylation. SW480 cells were co-transfected with 50 ng of luciferase reporter plasmid containing either $p27^{kip}$ or $CDX2$ 3’UTR together with 50 nM CNOT1 siRNA or control siRNA. Firefly and renilla luciferase activity was measured at 48 hours post-transfection using the Dual-Glo reporter assay system (Promega). Luciferase activity was analyzed by the ratio of renilla/firefly illuminance to normalize cell number and transfection efficiency. The luciferase activity of $p27^{kip}$ (A) and $CDX2$ (B) reporter was not obviously changed in CNOT1 siRNA transfected cells comparing to control cells. The data represent the result of only one experiment in which each treatment was performed in duplicates.

Discussion

Consistent with published studies, knockdown of $CNOT1$ and $CNOT2$ in SW480 CRC cells highly reduced cell viability, suggesting that $CNOT1$ and 2 may function as an oncogene in CRC and that the integrity of the CNOT complex is important in maintaining cell viability. It has been shown that the attenuated deadenylase activity of the CNOT complex could cause reduced cell viability (Ito et al., 2012; Morita et al., 2007).
Therefore, the suppression of deadenylation activity caused by \textit{CNOT1} and \textit{CNOT2} depletion might be responsible for the decreased cell viability. To test this hypothesis we depleted deadenylase subunits in SW480 cells and measured cell viability. We found that cell viabilities in \textit{CNOT6L}, 7 and 8-depleted cells were highly reduced, suggesting that decreased cell viability induced by \textit{CNOT1} and 2 depletion might be due to suppressed deadenylation.

\p27^{kip} and CDX2 are two well-known cell cycle inhibitors and are potential target genes of CNOT1 and 2 mediated deadenylation. We investigated the mRNA levels of these two genes in both \textit{CNOT1}- and \textit{CNOT2}-depleted SW480 cells. The results showed that mRNA levels of these two genes were upregulated in \textit{CNOT1}-depleted cells but had no obvious change in \textit{CNOT2}-depleted cells, suggesting that CNOT2 may not participate in regulation of these two genes. These results also revealed that CNOT1 and CNOT2 might have different roles in deadenylation mediated by the CNOT complex. Next, to initially study whether deadenylation contribute the regulation of \p27^{kip} and CDX2 mRNA levels by CNOT1 we measured the mRNA levels of these two genes in SW480 cells depleted for deadenylase subunits. We found that lack of \textit{CNOT8} caused an increase in \p27^{kip} and CDX2 mRNA levels. But depletion of other deadenylase subunits did not result in obvious changes in \p27^{kip} mRNA levels. Interestingly, depletion of \textit{CNOT6} and 6L both caused decreased CDX2 mRNA levels, suggesting an indirect effect. Together, these results suggest that CNOT1 and 8 may participate in a complex that regulates mRNA levels of \p27^{kip} and CDX2 by deadenylation. These findings also revealed that deadenylase subunits have specificity in targets or that their distribution and abundance might be different in cell lines.

RNA binding proteins (RBPs) promote deadenylation of target RNA transcripts by binding to AREs in their 3'UTRs. They recruit the deadenylases to the target mRNAs through direct binding to CNOT1 that in turn provides a platform to recruit the deadenylases and thereby trigger decay of target mRNAs. \textit{SB}-based screens have identified several RBPs as potential CRC cancer genes, including \textit{Pum1}, \textit{Pum2}, \textit{Elavl1} and \textit{Celf1}. To test if these RBPs might mediate the phenotype shared by CNOT1 and
CNOT8 we examined the effect of depletion of several RBPs on cell viability and $p27^{kip}$ and $CDX2$ mRNA levels. Cell viability was reduced in cells depleted for ZFP36, PUM1 and CELF1 but not in cells depleted for PUM2. mRNA levels of $p27^{kip}$ were increased only in ZFP36-depleted cells. $CDX2$ mRNA levels were increased by 1.5 fold in ZFP36, and CELF1-depleted cells. Together, these results suggest that ZFP36 and CELF1 may work with CNOT1 and 8 in complexes to regulate cell viability and mRNA levels of $p27^{kip}$ or $CDX2$ through deadenylation.

Two important questions are whether CNOT1 regulates $p27^{kip}$ and $CDX2$ mRNA levels through deadenylation and whether upregulated levels of these two genes were responsible for the reduced cell viability in CNOT1 insufficient cells. To answer the first question we performed a 3’UTR luciferase reporter assay. The results revealed that lack of CNOT1 did not result in a change in luciferase activity of $p27^{kip}$ and $CDX2$ 3’UTR reporter plasmids. This result suggested that CNOT1 might regulate the mRNA levels of these two genes through mechanisms other than deadenylation. It is also possible that the assay was not able to detect the changes in deadenylation possibly due to the reporter plasmids, so we may need to construct new reporter plasmids, such as a plasmid containing concatemers of targeting sites of known microRNAs that destabilizes these two genes. To answer the second question we depleted $CDX2$ in SW480 cells. But no obvious change in cell viability was observed. These results suggest that upregulated $CDX2$ may not be the cause for reduction in cell viability caused by CNOT1 depletion. However, this experiment was carried out in highly malignant SW480 cells. Therefore we need to knock down $CDX2$ in a normal colon epithelial cell line and perform an overexpression study of $CDX2$, before we can draw any conclusions.

In this study we failed to provide evidence that CNOT1 and CNOT2 could regulate $p27^{kip}$ and $CDX2$ mRNA levels through deadenylation and that upregulated $CDX2$ levels could be responsible for decreased cell viability following CNOT1 and CNOT2 depletion. However, there were several obvious limitations in this study. Some of the experiments were performed only once, thus the results need to be confirmed by more replicates. Knockdown of $CDX2$ was performed in a highly malignant CRC cell line, possibly making the acceleration in cell proliferation impossible to be observed.
Thus a less malignant cell line and overexpression studies are necessary. The 3’UTR reporter assay may not work for our purpose, so a new strategy to design reporter plasmids may be needed. It is also possible that upregulation of p27kip and CDX2 do not contribute to decreased cell viability caused by CNOT1 and CNOT2 depletion. Upregulation of other genes or a combination of other genes may be the reason.

**Material and methods**

siRNA transfection, MTT assay, RNA isolation, cDNA synthesis, qRT-PCR were performed as described in previous sections.

**Luciferase Reporter Plasmids Construction and Luciferase Assay**

Clones containing complete 3’UTR of p27kip (clone ID: 3458141) and CDX2 (clone ID: 4867859) were obtained from Openbiosystem. PCR primers containing restriction sites for XhoI and NotI were designed to amplify the complete 3’UTR of p27kip and CDX2. p27kip: forward, 5’- CAACTCGAGATGACCCACCGGGGTCTGCAG-3’; reverse, 5’-CAAGCGGCCCATCTGGAAGCTCATTTATCT-3’; CDX2: forward, 5’-CAACTCGAGATAACAGCTCGAATTAAGAAT-3’; reverse, 5’-CAAGCGGCCCATCTGGAAGCTCATTTATCT-3’.

PCR was carried out using a Phusion High-Fidelity DNA Polymerase (New England Biolabs) with the following reaction mix: 8 µl of 5 x Phusion HF buffer, 0.4 µl Phusion DNA Polymerase, 200 µM of each dNTP, 1 µl of each primer (final concentration 0.5 µM), 100 ng of DNA template, and corresponding amount of H2O to make a 40 µl volume. Cycling conditions were 95°C for 5 minutes, then 35 cycles of 95°C for 30 seconds, 60°C for 30 seconds, 72°C for 30 seconds, and finally 72°C for 10 minutes. PCR products was digested with XhoI and NotI and inserted between the corresponding restriction sites on a psiCHECK-2 vector (Promega) to generate 3’UTR reporter plasmids. psiCHECK-2 vector contains both the synthetic firefly luciferase (Fluc) gene and the synthetic renilla luciferase (hRluc) gene, each with its own promoter and poly(A)-addition sites. The correctness of all constructs was confirmed by sequencing.

To perform luciferase reporter assays, SW480 cells were plated at 2 x 10^4/well on 96 well plates and co-transfected with 50 ng of 3’UTR luciferase reporter plasmids for
p27kip or CDX2 together with 50 nM CNOT1 siRNA or a control siRNA. Firefly and renilla luciferase activities were measured at 48 hours post-transfection using a Dual-Glo reporter assay system (Promega) according to manufacturer's instructions. Briefly, 96 well plates was changed with 75 µl DMEM medium and then added with 75 µl of Dual-Glo luciferase reagent containing the firefly luciferase substrate. Incubate plates for 15 minutes to allow for cell lysis to occur, and then measure the firefly luminescence in a luminometer (Biotek). Add 75 µl of Dual-Glo Stop & Glo reagent that contains the inhibitor of firefly luciferase and substrate for renilla luciferase. Incubate for 15 minutes, and then measure renilla luminescence in the same plate order as the firefly luminescence. Total renilla luciferase activity was normalized to firefly luciferase activity in order to correct for differences in transfection efficiency.
APPENDIX 4

Construction of CNOT1 mutant plasmids

As discussed in the introduction, multiple somatic mutations in CNOT1 have been identified in various types of human cancers, in particular CRC that has the highest CNOT1 mutation rate among cancers (~8%). Furthermore, several recurrent mutations also have been documented, for example, R1733H and R1640Q. These results suggest that genetic mutations in CNOT1 may contribute to the development of CRC. To further study the potential contribution of CNOT1 somatic mutations to CRC we have began to create constructs that contain the documented mutations in CNOT1 to study their effects on CRC in cell culture as well as in mouse models. To start with we selected nine somatic mutations in CRC identified by the TCGA project (Table 24). Among these mutations, seven have been documented more than once in the COSMIC database and all of them are predicted to have a medium level impact on CNOT1 function. The distribution of these mutations in CNOT1 protein is illustrated (Figure 55).

<table>
<thead>
<tr>
<th>AA change</th>
<th>Codon</th>
<th>Mutation</th>
<th>COSMIC</th>
<th>FIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>R369C</td>
<td>CGT</td>
<td>TGT</td>
<td>4</td>
<td>Medium</td>
</tr>
<tr>
<td>L484R</td>
<td>CTG</td>
<td>CGG</td>
<td>2</td>
<td>Medium</td>
</tr>
<tr>
<td>R499C</td>
<td>CGC</td>
<td>TGC</td>
<td>2</td>
<td>Medium</td>
</tr>
<tr>
<td>R499H</td>
<td>CGC</td>
<td>CA C</td>
<td>2</td>
<td>Medium</td>
</tr>
<tr>
<td>R591C</td>
<td>CGT</td>
<td>TGT</td>
<td>6</td>
<td>Medium</td>
</tr>
<tr>
<td>L1232F</td>
<td>TTG</td>
<td>TTC</td>
<td>1</td>
<td>Medium</td>
</tr>
<tr>
<td>R1394C</td>
<td>CGT</td>
<td>TGT</td>
<td>4</td>
<td>Medium</td>
</tr>
<tr>
<td>A1396T</td>
<td>GCA</td>
<td>ACA</td>
<td>1</td>
<td>Medium</td>
</tr>
<tr>
<td>R1446H</td>
<td>CGT</td>
<td>CAT</td>
<td>2</td>
<td>Medium</td>
</tr>
</tbody>
</table>

Table 24. The 9 selected mutations in CNOT1. COSMIC indicates the number of times this mutation has been documented in the COSMIC database. FIS (Functional Impact Score): the predicted functional impact for missense mutations calculated by Mutation Assessor.
Figure 55. The distribution of selected mutations in the CNOT1 protein. Each dot represents a mutation.

Clones encoding CNOT1S (NM_206999.2) ORF (BC040523) in a pBluescriptR vector were obtained from Open Biosystem (Clone Id: 5266600). The CNOT1S ORF was inserted between the Xho1 and BamH1 sites from 5’ to 3’ (Figure 57). These nine somatic mutations were introduced into the ORF of CNOT1S via using a Transformer Site-Directed Mutagenesis Kit (Clontech). The general procedure of this method is illustrated in Figure 56. The mutagenic primer introduces the desired mutation and the selection primer mutates a unique restriction site in the plasmid for selection of mutated plasmids. These two primers are simultaneously annealed to one strand of a denatured double-stranded template and both mutations are incorporated in the same newly synthesized strand after elongation by T4 DNA polymerase. The mixture of original and mutated DNA is then digested with a restriction enzyme that cuts at the original restriction site and is used to transform E. coli BMH 71-18 mutS, which is mismatch repair deficient, to propagate the mutated plasmid. The uncut, mutated DNA has much higher transformation efficiency than linear DNA with no mutations, so the product of the first round of transformation will be mutated plasmids with the two mutations on one of the two strands. Then the DNA will be digested with the same restrictive enzyme again and used to transform a regular E. coli strain. The final product will be a mixture of original plasmids and mutated plasmids with the two mutations on both strands. Two rounds of DNA digestion and transformation ensure that a very high frequency of transformants carry the mutated plasmid, which nearly always contains both mutations.
The nine CNOT1 mutants were constructed following the protocol below.

A. Design selection and mutagenic primers (Table 25). The selection primer deletes the single Sca1 site on pBluescriptR-CNOT1s plasmid. And the mutagenic primers bear the nine desired point mutations in CNOT1.

<table>
<thead>
<tr>
<th>Primer type</th>
<th>Change</th>
<th>Original</th>
<th>Mutated</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection primer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#1</td>
<td>R369C</td>
<td>CGT</td>
<td>TGT</td>
<td>5’p-GTGACTGGTGAAAAGTC3’</td>
</tr>
<tr>
<td>#2</td>
<td>L484R</td>
<td>CTG</td>
<td>CGG</td>
<td>5’p-CCAGACAGTGGAACAGCAAGTC-3’</td>
</tr>
<tr>
<td>#3</td>
<td>R499C</td>
<td>CGC</td>
<td>TGC</td>
<td>5’p-GGCATACCTGTGCAATGTATCTC-3’</td>
</tr>
<tr>
<td>#4</td>
<td>R499H</td>
<td>CGC</td>
<td>CAC</td>
<td>5’p-GGCAACCTGGGTCAATTCAC-3’</td>
</tr>
<tr>
<td>#5</td>
<td>R591C</td>
<td>CGT</td>
<td>TGT</td>
<td>5’p-GCAGCTGCCTCGTTGGAATAC-3’</td>
</tr>
<tr>
<td>#6</td>
<td>L1232F</td>
<td>TTT</td>
<td>TTC</td>
<td>5’p-GGCAGGACAGGATCCCTTCTTGAGTGC-3’</td>
</tr>
<tr>
<td>#7</td>
<td>R1394C</td>
<td>CGT</td>
<td>TGT</td>
<td>5’p-CAGTGTGTCGTGACAGCTGG-3’</td>
</tr>
<tr>
<td>#8</td>
<td>A1396T</td>
<td>GCA</td>
<td>ACA</td>
<td>5’p-GTGCTGCAAGTATGAACCGGCTGCC-3’</td>
</tr>
<tr>
<td>#9</td>
<td>R1446H</td>
<td>CGT</td>
<td>CAT</td>
<td>5’p-GCATGATAGCAGTGACGCTGG-3’</td>
</tr>
</tbody>
</table>

Table 25. The sequence of selection and mutagenic primers.
Figure 57. The map of pBluescriptR-CNOT1 plasmid and the location of mutagenic and selection primers on the plasmid.

B. Denaturation of plasmids and annealing of primers to the DNA template.

Set up the primer/plasmid reaction system in a 0.5 ml microcentrifuge tube, as follows:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10X Annealing Buffer</td>
<td>2.0 μl</td>
</tr>
<tr>
<td>pBluescriptR-CNOT1</td>
<td>2.0 μl</td>
</tr>
<tr>
<td>(0.05 μg/μl)</td>
<td></td>
</tr>
<tr>
<td>selection primer</td>
<td>2.0 μl</td>
</tr>
<tr>
<td>(0.05 μg/μl)</td>
<td></td>
</tr>
<tr>
<td>mutagenic primer</td>
<td>2.0 μl</td>
</tr>
<tr>
<td>(0.05 μg/μl)</td>
<td></td>
</tr>
<tr>
<td>H2O</td>
<td>12.0 μl</td>
</tr>
<tr>
<td>Total</td>
<td>20 μl</td>
</tr>
</tbody>
</table>

Incubate at 100°C for 3 minutes.

Chill immediately in an ice and water bath (0°C) for 5 minutes.

C. Synthesis of the mutant DNA strands.

To the primer/plasmid annealing reaction from last step add:
Incubate at 37°C for 2 hours.

Heat the reaction at 70°C for 5 minutes to inactivate the enzymes.

D. Primary selection for mutant plasmids by restriction digestion. This step will selectively linearize the parental DNA to increase the percentage of mutant plasmids present in the pool of the first transformation.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixed plasmids from step B</td>
<td>30 µl</td>
</tr>
<tr>
<td>10X NEB buffer 4</td>
<td>3.5 µl</td>
</tr>
<tr>
<td>Sca1-HF</td>
<td>1.5 µl</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>35 µl</td>
</tr>
</tbody>
</table>

Incubate at 37°C for 2 hours.

Heat the reaction at 70°C for 5 minutes to inactivate the enzymes.

E. First Transformation.

Transform the BMH 71-18 mutS cells with digest reaction from step D to propagate the heterozygous mutant plasmids.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMH 71-18 mutS cells</td>
<td>100 µl</td>
</tr>
<tr>
<td>digested plasmids from step D</td>
<td>10 µl</td>
</tr>
</tbody>
</table>

Put on ice 30 minutes, incubate at 42°C for 1 minute, transfer to ice for 5 minutes.

Add 1 ml of S.O.C (with no antibiotic) to each tube. Incubate at 37°C for 60 minutes with shaking at 220 rpm to let cells recover.

Add 4 ml of LB medium containing ampicillin at 100 µg/ml to each tube.

Incubate the culture at 37°C overnight with shaking at 220 rpm.

F. Isolation of the mixed plasmid pool using a Qiagen Miniprep Kit.

Nanodrop the harvested plasmids and adjust the concentration to ≈ 100 ng/µl.

G. Selection of the mutant plasmids.
Incubate at 37°C for 2 hours.

Add an additional 0.5 µl Sca1-HF, and continue incubation at 37°C for another 1 hour.

Heat the reaction at 70°C for 5 minutes to inactivate the enzymes.

**H. Final Transformation**

Transform the NEB5α high efficiency cell with digest reaction from step G to propagate the final mutant plasmids.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digested plasmids from step G</td>
<td>5 µl</td>
</tr>
<tr>
<td>NEB5α high efficiency cell</td>
<td>50 µl</td>
</tr>
</tbody>
</table>

Put on ice 30 minutes, incubate at 42°C for 1 minute, transfer to ice for 5 minutes.

Add 450 µl S.O.C medium (with no antibiotic) to each tube. Incubate at 37°C for 60 minutes with shaking at 220 rpm to let cell recover.

Use 50 µl, 100 µl and 350 µl cell suspension to streak three LB plates containing 50 µg/ml ampicillin for each mutant.

Incubate plate at 37°C overnight.

**I. Propagate cells and harvest final plasmids.**

Pick 3 clones from each plate and grow cells in 2.5 ml LB containing 100 µg/ml ampicillin at 37°C overnight with shaking at 220 rpm.

Isolate final plasmids from 4 clones for each mutant using a Qiagen Miniprep Kit (Qiagen).

**J. Confirm the mutations by Sanger sequencing.**

The sequencing primers for confirming introduced mutations are listed in table 26.
Table 26. The sequencing primers used for validation of mutations.

<table>
<thead>
<tr>
<th>Mutant</th>
<th>AA change</th>
<th>Sequencing primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>R369C</td>
<td>CAAGGGTTTTGGAATGATG</td>
</tr>
<tr>
<td>#2</td>
<td>L484R</td>
<td>CCCTGTCATCGTTGCCACT</td>
</tr>
<tr>
<td>#3</td>
<td>R499C</td>
<td>CCCTGTCATCGTTGCCACT</td>
</tr>
<tr>
<td>#4</td>
<td>R499H</td>
<td>CCCTGTCATCGTTGCCACT</td>
</tr>
<tr>
<td>#5</td>
<td>R591C</td>
<td>TCGAATACTTGGATGGGCCC</td>
</tr>
<tr>
<td>#6</td>
<td>L1232F</td>
<td>TGTTAAAGGACACAAAGATTGC</td>
</tr>
<tr>
<td>#7</td>
<td>R1394C</td>
<td>CAGTACAGCTACCACGACATCA</td>
</tr>
<tr>
<td>#8</td>
<td>A1396T</td>
<td>CAGTACAGCTACCACGACATCA</td>
</tr>
<tr>
<td>#9</td>
<td>R1446H</td>
<td>CAGTACAGCTACCACGACATCA</td>
</tr>
</tbody>
</table>

Set up the reaction in a 0.5 ml eppendorf tube for Sanger sequencing as follows:

- CNOT1 mutant: 800 ng
- Sequencing primer (2 μM): 1.6 μl
- Add H20 to: 12 μl

Send the tubes for sequencing to the University of Minnesota BioMedical Genomic Center. The sequencing results revealed that all nine mutations except for #6 were successfully introduced into the CNOT1 ORF in the pBluescriptR vector (Figure 58).

![Figure 58. Validation of introduced mutations by Sanger sequencing.](image)

The mutated nucleic
acids were marked out in gray.

**K. Construction of plasmids expressing CNOT1 mutants.**

The pBluescriptR vector is non-expressing. Therefore, to ectopically express the CNOT1 mutants in cells we transferred the mutated CNOT1 ORF into a pcDNA3 vector (Life Technologies) that can be expressed in mammalian cells. The eight CNOT1 mutant constructs were digested with ApaI and NotI and inserted between the corresponding restriction sites on a pcDNA3 vector to generate pcDNA3-CNOT1Δ. The final plasmids were propagated by transforming NEB5α cells and confirmed by restrictive digestion and agarose gel electrophoresis.

**Discussion**

Multiple somatic mutations in CNOT1 have been documented in different types of cancers, in particular CRC. To characterize these mutations in CRC ectopic expression of CNOT1 bearing these mutations in cell lines represents a direct means to test the function of these mutations *in vitro*. The CNOT1 mutations in CRC span the whole CNOT1 protein and no obvious cluster was discovered. Due to a lack of understanding of CNOT1 structure, it is difficult to predict which mutations may have more influence than others. Therefore, we selected 9 mutations spanning the short CNOT1 isoform to make CNOT1 mutants. There are several methods in introducing a site-specific mutation into a construct. We employed Clontech’s Transformer Site-Directed Mutagenesis method to introduce the 9 mutations into the CNOT1 ORF. Finally, we successfully constructed 8 CNOT1 mutants, confirmed by Sanger sequencing. These results suggested that this method is an effective way to introduce a mutation into a plasmid construct of 8kb in length. We note that the short CNOT1 isoform lacks the amino acids after 1478 aa including part of the deadenylase binding domain and the NOT1 domain on the C-terminal region. Since long CNOT1 isoforms are the complete and predominant isoforms and may have different functions, we now need to make the mutated long CNOT1 isoforms bearing the same mutations in ways described in material and methods section. We also need to introduce the mutations discovered in the C-terminal region of CNOT1 into the construct for long CNOT1 isoforms. In the future, we will express these CNOT1
mutants in CRC and normal colon epithelial cells to study their effects on oncogenic phenotypes, assembly of the CNOT complex and deadenylation of target mRNA transcripts. If necessary, we can introduce another mutation to these CNOT1 mutants to silence the sites targeted by siRNA oligos used to deplete endogenous CNOT1. Furthermore, we may carry out in vivo studies by producing transgenic mice bearing CNOT1 mutations that show strong phenotypes in cell line based studies. These studies will be helpful in determining the role of CNOT1 in CRC and in understanding the structure and functions of CNOT1.
APPENDIX 5

Identification of fusion mRNA transcripts in human colorectal cancer and normal colon epithelial tissues using RNA seq

Introduction

A gene fusion is formed by combining a part of one gene with a segment of another. Until recently, most fusion genes were discovered in neoplasms having hematopoietic and mesenchymal origins (Li et al., 2008), such as leukemia, lymphoma and sarcoma. Recurrent and characteristic gene fusions are a common feature of these neoplasms (Li et al., 2008) that together represent only 10% of all human cancers, such as BCR-ABL1 in chronic myelogenous leukemia (CML) (Nowell et al., 2007), NPM1-ALK in lymphomas (Falini et al., 1999) and EWS-FLI-1 in Ewing's sarcoma (May et al., 1993). Recurrent gene fusions are widely used for disease classification, diagnosis, and therapeutic targets. For example, treatment of CML with Gleevec, a tyrosine kinase inhibitor specifically targeting the BCR-ABL chimeric protein, has been proved to increase the survival rate of CML patients (Deininger et al., 2005).

However, detection of cancer-specific, recurrent gene fusions was rare in common types of epithelial carcinomas, which account for 80% of cancer-related deaths. This apparent discrepancy is mainly due to the inherent complexity of the solid cancer genome and technical limitations rather than fundamental genetic differences between hematological malignancies and solid cancers. With the application of new technological approaches, in particular the next-generation sequencing (NGS) techniques, and software for analyzing sequencing data, increasing numbers of gene fusions were being identified in solid tumors, such as TMPRSS2-ERG in prostate cancers (Taylor et al., 2010, Ishkanian et al., 2009), EML4-ALK in non-small cell lung cancer (NSCLC) (Soda et al., 2007) and VAPB-IKZF3 in breast cancer cell lines (Edgren et al., 2011). These results
have changed our views regarding the presence and frequency of fusion genes in solid tumors, sparking a revitalization of the quest to identify novel gene fusions in epithelial malignancies.

The Mechanisms for Formation of Fusion Gene Transcripts

A fusion gene transcript is a chimeric RNA joining the transcripts of two or more different genes. The fusion transcripts can be produced by different mechanisms.

Chromosomal rearrangement is a most common cause of gene fusions, which can cause translocations, deletions and inversions. The best-known example of a gene fusion caused by chromosomal translocation is the BCR-ABL fusion gene in CML (Wong and Witte, 2004).

In addition to chromosomal rearrangements, fusion transcripts can also be produced due to aberrant RNA splicing at the transcript level, including transcription-induced chimeras (TICs) and trans-splicing (Figure 59).

A TIC is a specialized cis-splicing event that results from read-through transcription of two adjacent, independent genes that share the same orientation. In this process, RNA polymerase II keeps transcribing over multiple polyadenylation sites, where transcripts are normally cleaved and processed. Most of these fusion transcripts are generated from tandem genes that are physically located within ~50 kb of each other and the median distance is ~8.5 kb, indicating the spatial proximity might be a prerequisite for forming TICs (Akiva et al., 2006). TICs were previously reported in isolated cases and thought to be a rare event in mammals. Recent studies, however, incorporating systematic in silico analyses of ESTs and cDNAs in the NCBI databases, found that as much as 5% of the tandem gene pairs in the human genome is comprised of chimeric sequences, indicating TICs could be quite frequent. (Parra et al., 2006, Unneberg et al., 2007). Despite the prevalence, the regulation and function of most TIC-induced fusion transcripts remains largely uncharacterized (Wang et al., 2007, Akiva et al., 2006). Some of these chimeric transcripts are thought to be produced as ‘noise’ in the transcription process, while some could encode chimeric proteins or have a logical function inferred.
For example, the RBM6-RBM5 TIC expression levels were associated with elevated levels of RBM6 and RBM5 mRNA, and with increased tumor size in breast tumor tissue (Wang et al., 2007). In addition, two tissue-specific TICs fusing tyrosine hydroxylase (TH) and insulin (INS) genes were reported to encode two TH protein isoforms whose functionality was somewhat diminished with respect to the canonical TH, providing an additional aspect in the control of TH/insulin gene expression. (Hernández-Sánchez et al., 2006, Mansilla et al., 2005).

Trans-splicing can generate chimeric mRNAs between pre-mRNAs originating from two different genes. It is a special form of RNA processing where exons from two different primary RNA transcripts are joined end to end and ligated (Figure 59). This process was found to follow similar mechanisms as cis-splicing of RNA, which is also catalyzed by the spliceosome (Faustino et al., 2003). Juxtaposition of the loci encoding the RNAs that participate in trans-splicing seems to be essential for some pre-mRNA substrates but not for some other ones (Kristi et al., 2008). Some resultant fusion transcripts may have functional effects. An example is the cyclin D1 (CCND1)-tumor-associated calcium signal transducer 2 (TROP2) fusion transcript detected in a variety of human cancers (Guerra et al., 2008). This RNA transcript did not encode a chimeric protein, but rather led to elevated levels of CCND1 in that the 3’ UTR in the CCND1 mRNA was deleted due to trans-splicing with TROP2 pre-RNA. Removal of the 3’UTR stabilizes the chimeric RNA transcript relative to normal CCND1 mRNA. In addition, the chimeric RNA was reported to be oncogenic as it could transform primary cells in culture and induce aggressive tumor formation in cooperation with activated RAS in both cultured cells and transgenic mice. These findings have turned the view of trans-splicing in higher eukaryotes from a bizarre phenomenon to a biological event that is attaining stronger recognition (Mansfield et al., 2004, Garcia-Blanco et al., 2003).
The Biological Effects of Gene Fusions

Previous studies suggested that the consequence of gene fusions could be either generation of a truncated or chimeric protein with a different function or an alteration of gene expression (Tomlins et al., 2005). In most cases, gene fusions occur within introns, giving rise to the expression of truncated or chimeric proteins with a new or altered activity (Mitelman et al., 2007), such as BCR-ABL chimeric protein which has constitutive tyrosine kinase activity (Wong and Witte, 2004). Some of the resultant chimeric proteins are abnormal, tumor-specific products that may provide tumor cells with a growth and/or survival advantage (Li et al., 2008, Mitelman et al., 2007). Besides encoding chimeric proteins, formation of gene fusions can affect gene expression (Lin et al., 2012). In some cases, gene fusion occurs between the coding region of a gene and the transcriptionally active promoter or enhancer region of another gene, which may lead to upregulated expression of the former gene. Examples include TMPRSS2-ERG and TMPRSS2-ETV1 gene fusions where expression of ERG or ETV1 genes are dramatically up-regulated due to their repositioning after the TMPRSS2 transcription start site which is actively transcribed in prostate cancer (Tomlins et al., 2005). Another type of gene fusion involves recombination between the coding region of one gene with the non-coding regulatory region, such as 3’UTR, of the second gene, which may alter the post-transcriptional regulation of the former gene. One example is the chimera between CCND1 mRNA with TROP2 pre-RNA, which led to the deletion of 3’ UTR of CCND1 mRNA transcript. The CCND1-TROP2 fusion transcript stabilized CCND1 mRNA levels and consequently resulted in elevated CCND1 levels, promoting cells to pass the G1 phase of the cell cycle (Guerra et al., 2008).

Next Generation Sequencing and RNA Sequencing
Next generation sequencing (NGS) technology is a fundamentally novel approach to DNA sequencing, which can sequence the fragments of a DNA template in a massively parallel fashion, producing thousands or millions of sequences concurrently. This advance enables rapid sequencing of large stretches of DNA base pairs spanning entire genomes (Hall et al., 2007, Church et al., 2006). Whole-genome DNA sequencing can be used to identify potential fusion-gene-creating rearrangements. However, only a fraction of gene fusions predicted based on DNA sequencing is expected to generate an expressed fusion mRNA, making this approach tedious to discover activated, oncogenic fusion gene events. In contrast, RNA sequencing (RNA-seq) using next-generation sequencing instruments directly identifies only those fusion genes that are expressed, providing an efficient tool to identify candidate oncogenic fusions (Edgren et al., 2011). RNA-seq also facilitates comprehensive characterization of cellular transcriptomes.

In a standard RNA-seq procedure, RNA sample (total or polyA+ enriched) is converted to cDNA through reverse transcription, which is then broken into a library of small fragments. Oligonucleotide adapters are then attached to a cDNA library on one or both ends. Finally the adaptor-ligated cDNA library is sequenced on a high-throughput platform either single-ended (single-end sequencing) or pair-ended (pair-end sequencing) (Martin et al., 2011). This process usually generates millions of short (30-400 bp) reads which can be analyzed by a variety of bioinformatics tools. The methods for inferring the sequencing data are generally classified as two types. In one approach the reads are aligned and mapped to a reference genome either of the organism itself (whose transcriptome is being studied) or of a closely related species. The other approach, de novo transcriptome assembly, uses bioinformatics tools to assemble the transcriptome directly from short sequence reads without referring to a reference genome (Wang et al., 2009). Generally speaking, the results of RNA-Seq have shown high levels of reproducibility, for both technical and biological replicates (Cloonan et al., 2008). The general procedure of RNA sequencing is summarized in Figure 60. Paired-end RNA-seq, where 36 to 100 bp are sequenced from both ends of 200 to 500 bp long cDNA molecules, is especially suitable for identification of chimeric mRNA transcripts. The employment of the paired-end RNA-seq, combined with effective analytic pipelines
enable the discovery of fusion RNA transcripts in breast cancer (Edgren et al., 2011, Asmann et al., 2012), lung cancer (Lin et al., 2009), colon cancer (Chu et al., 2011), and other types of solid tumors with high confidence.

Figure 60. The procedure of RNA-seq. (Solar home http://cmb.molgen.mpg.de)

Fusion Genes in CRC

The first recurrent fusion gene in CRC was reported by Lin et al in 2009 (Lin et al., 2009). They performed an exon array for the global search of gene rearrangements in a large collection of breast cancer, CRC and non-small cell lung cancer (NSCLC) samples and identified the echinoderm microtubule associated protein like 4 (EML4)-anaplastic lymphoma receptor tyrosine kinase (ALK) fusion in multiple samples. An RT-PCR assay was then used to screen a collection of tumor samples for the presence of the EML4-ALK fusion transcript, leading to detection of the EML4-ALK fusion in 2.4% of breast cancer (5 of 209), 2.4% of CRC (2 of 83), and 11.3% of NSCLC (12 of 106), which was confirmed by subsequent sequencing of PCR products. The presence of EML-ALK rearrangement was further verified by identifying genomic fusion points in tumor
samples using long-range genomic PCR. A FISH assay was used to confirm the EML4-ALK translocation in tumor samples and tumor-derived cell lines positive for the fusion transcript. The functional role of the EML4-ALK fusion in cell growth and proliferation was assessed by measuring cell viability following siRNA-mediated silencing of EML4 and ALK. Cell growth inhibition was not observed in control cell lines lacking the EML4-ALK fusion. On the contrary, significant growth inhibition was observed in lung and breast cancer cell lines but not in CRC cell lines harboring the EML4-ALK fusion, failing to show the functionality of this fusion in CRC (Lin et al., 2009).

Next generation sequencing provides the capacity for doing a deeper analysis of the entire genomes of CRC tumors. In the first study using NGS to discover gene fusions in CRC, Bass et al performed whole genome sequencing of 9 CRCs and paired non-neoplastic tissues (Bass et al., 2011). They found 11 chromosomal rearrangements that could give rise to in-frame fusion transcripts, including a fusion of vesicle transport through interaction with t-SNAREs 1A (VTI1A) and transcription factor 7-like 2 (TCF7L2). By screening cDNA from a panel of 97 primary CRCs, they found that the in-frame VTI1A-TCF7L2 fusion was recurrently expressed in 3 of 97 cases. The VTI1A-TCF7L2 chimeric transcript fuses the first three exons of VTI1A to the fourth exon of TCF7L2. VTI1A encodes a v-SNARE protein mediating fusion of intracellular vesicles within the Golgi complex 21 (Kreykenbohm et al., 2002). TCF7L2 encodes a transcriptional factor belonging to the TCF/LEF family 4 (Clevers et al., 2006, Waterman et al., 2004). TCF7L2 is known to have a significant role in CRC in collaboration with beta-catenin. To test the functional importance of the VTI1A-TCF7L2 fusion gene in CRC, they knocked down the fusion gene through RNAi targeting the sequence spanning the fusion, which caused a dramatic reduction in the anchorage-independent growth of NCI-H508 cells, suggesting the fusion gene has a substantial role in cancer cell survival. This paper provides an excellent example of employment of NGS technology in identification of important genomic rearrangements and fusion events occurring in CRC.

Using RNA-seq, Wu et al profiled the whole transcriptomes of CRC tissue, adjacent non-tumor tissue and distant normal tissue from a stage III CRC patient (Wu et
al., 2012). They identified a cancer-specific gene fusion between the prostaglandin F2 receptor inhibitor (PTGFRN) and NOTCH2 (NOTCH2) using two algorithms, deFuse and TopHat-Fusion, which was confirmed by RT-PCR and Sanger sequencing. NOTCH2 is a homolog of NOTCH1 and plays a role in a variety of developmental processes by controlling cell fate decisions. NOTCH2 expression has been shown to be a prognostic predictor and is related to tumor differentiation status in CRC (Chu et al., 2011). In the PTGFRN-NOTCH2 gene fusion the first intron of PTGFRN is fused with the 3′ junction of the 17th exon of NOTCH2, thus the fusion gene is under the control of the promoter of NOTCH2. The predicted protein encodes the first 917 aa of NOTCH2, lacking some essential domains in NOTCH2, such as NOTCH domain, and appearing to act like a loss-of-function mutation. This study employed RNA-Seq to successfully identify the tumor-specific PTGFRN-NOTCH2 fusion gene in CRC. However, this study only included one tumor sample and did not show the functional significance of the PTGFRN-NOTCH2 gene fusion. Therefore, a study using RNA-seq in larger group of CRC samples is necessary to identify recurrent gene fusions in CRC.

In another large scale study of fusion genes in CRC, Seshagiri et al (Seshagiri et al., 2012) applied exome sequencing and RNA-seq to 70 pairs of primary human CRC and normal tissues to characterize their exomes, transcriptomes and copy-number alterations. 36 chromosomal rearrangements that result in gene fusions were identified using RNA-seq. Among these gene fusions, recurrent gene fusions involving R-spondin family members R-spondin 2 homolog (RSPO2) (3%; 2 out of 68) and R-spondin 3 homolog (RSPO3) (8%; 5 out of 68) were discovered. R-spondins are secreted proteins that can potentiate canonical WNT signaling (Yoon et al., 2012). The recurrent RSPO2 fusion involves eukaryotic translation initiation factor 3, subunit E (EIF3E) exon 1 and RSPO2 exon 2, which is expected to produce a functional RSPO2 protein driven by the EIF3E promoter. Two different RSPO3 fusion variants were identified where either exon 1 or exon 7 of protein tyrosine phosphatase, receptor type, K (PTPRK) (5′) was fused to exon 2 of RSPO3 (3′). The PTPRK (e1)–RSPO3 (e2) transcript found in four samples is an in-frame fusion that preserves the entire coding sequence of RSPO3 and replaces its secretion signal sequence with that of PTPRK. The PTPRK (e7)-RSPO3 (e2) fusion,
detected in one sample, is also an in-frame fusion that potentially encodes a 70 kD protein consisting of the first 387 amino acids of PTPRK, including its secretion signal sequence, and the RSPO3 amino acids 34-272 lacking its native signal peptide. In studies of the biological activity of the R-spondin fusions, they showed that the RSPO fusion proteins were capable of potentiating WNT signaling in both 293T and HT-29 cells using a WNT-responsive luciferase reporter assay.

Furthermore, TCGA has comprehensively examined in genomic alterations in CRC, providing the most integrated view of CRC cancer genes to date (Cancer Genome Atlas Network, 2012). To identify new chromosomal translocations, they performed low-pass, paired-end, whole-genome sequencing on 97 tumors with matched normal samples. Despite the low genome coverage, they detected 250 candidate interchromosomal translocation events (0–10 per tumor). Among these events, 212 had one or both breakpoints in an intergenic region, whereas the remaining 38 juxtaposed coding regions of two genes in putative fusion events, of which 18 were predicted to code for in-frame events. They also found three separate cases (3%) in which the first two exons of the neuron navigator 2 (NAV2) gene are joined with the 3′ coding portion of transcription factor 7 like 1 (TCF7L1) through an interstitial deletion. TCF7L1 encodes TCF3, a member of the TCF/LEF class of transcription factors that heterodimerize with nuclear β-catenin to enable β-catenin-mediated transcriptional regulation. However, the predicted NAV2–TCF7L1 fusion protein lacks the TCF3 β-catenin-binding domain, suggesting a loss of function truncation. They also observed 21 cases of translocation involving tetratricopeptide repeat domain 28 (TTC28) that was identified as a target of P53 and an inhibitor of tumor cell growth (Brady et al., 2011). In all cases the fusions are predicted to lead to inactivation of TTC28.

These studies provided examples for employment of next-generation sequencing and RNA-seq in identification of recurrent and functional gene fusions in CRC as well as other solid tumors. Identification of these genomic rearrangements and fusion events will be beneficial in further understanding the mechanisms of CRC formation and developing novel therapeutic strategies targeting these gene fusions or their product.
All the published studies in gene fusions in CRC were performed using primary CRC samples. To study the fusion transcripts that are associated with advanced stage CRC and potentially involved in cancer progression we studied the fusion transcripts in 20 liver metastasis of stage IV CRC and 5 normal colon samples using RNA-seq and analyzed the sequencing data via deFuse software.

Results

In the 20 CRC samples, we discovered 414 fusion transcripts involving 737 genes, among which 289 fusion transcripts were unique, 42 were in-frame fusions, and 137 occurred between adjacent genes, suggesting that these fusion transcripts were transcriptional read-throughs, also known as TICs, rather than actual genome rearrangements. In addition, the number and characteristics of fusion transcripts vary from sample to sample. On average, each CRC sample had 33 fusion transcripts (range 15-51), 0.7 in-frame fusions (range 0-3), 18 TICs (range 9-31), and 5 fusion transcripts (range 1-12) which had a predicted probability $\geq 0.85$, suggesting the presence of these fusions are highly possible (Figure 61, table 27, 35 and 36, Sue Rathe, unpublished data). These results suggest that fusion transcripts and TICs are prevalent in CRC samples. However, the in-frame fusions that potentially encode chimeric proteins are rare.

In the 5 normal colon samples, 54 fusion events were discovered among which 37 were unique fusion transcripts, none was an in-frame fusion, and 24 were caused by TIC. On average, each normal colon sample had 10.8 fusion transcripts (range 9-14), no in-frame fusion, 7 TICs (range 6-8), and 1 fusion transcripts (0-2) having a predicted probability $\geq 0.85$ (Figure 62, table 28, 37 and 38). These results suggest that fusion transcripts and TICs also exist in normal colon samples, while the numbers are lower than in CRCs.

We then compared the lists of fusion transcripts discovered in both CRC and normal samples to identify the fusions occurring in both types of samples. We found that 20 unique transcripts were discovered in both CRC and normal colon samples, among which 17 were the same fusion transcripts and 3 were formed by same partners but with
different breakpoints. None of these fusions were in-frame and 15 of them were TICs (Figure 63, Table 39).

Follow-up filtering using a stringent threshold for in-frame fusions resulted in a workable set of 12 fusion candidates that we are testing in \textit{in vitro} assays. The criteria for selection of candidate fusion transcripts are: probability $\geq 0.85$, in coding regions, and in-frame. All the 12 fusion transcripts were validated by RT-PCR followed by DNA sequencing except four which were validated by RT-PCR only. Interestingly, 6 of these fusion transcripts were also discovered in normal colon samples by RT-PCR, although none of them were predicted by the RNA seq data. In addition, 7 of the candidate fusion transcripts were TICs and 3 of them were found to be cancer-specific (Table 29).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure61.png}
\caption{The map of fusion transcripts in 20 CRC liver metastasis. The thickness of the edges corresponds to the times that the same fusion was detected.}
\end{figure}

\begin{table}[h]
\begin{tabular}{cc}
\hline
\# fusions & 414 \\
\# unique fusion transcripts & 289 \\
\hline
\end{tabular}
\end{table}
(B)

<table>
<thead>
<tr>
<th></th>
<th>Per sample</th>
<th>Range</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td># fusions</td>
<td></td>
<td>15--51</td>
<td>33</td>
</tr>
<tr>
<td># in-frame</td>
<td></td>
<td>0--3</td>
<td>0.7</td>
</tr>
<tr>
<td># TIC</td>
<td></td>
<td>9--31</td>
<td>18</td>
</tr>
<tr>
<td># prob ≥ 0.85</td>
<td></td>
<td>1--12</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 27. Summary of fusion transcripts discovered in 20 CRC samples. (A) Characteristics of fusion transcripts. (B) Fusion transcripts discovered in per sample. # in-frame: number of the fusion transcripts remaining an open reading frame; # TICs: number of fusion transcripts predicted to be TICs; # probability ≥ 0.85: number of fusions having predicted probability ≥ 0.85.

Figure 62. The map of fusion transcripts in 5 normal colon samples. The thickness of the edges corresponds to the times that the same fusion was detected.

(A)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td># fusions</td>
<td>54</td>
</tr>
<tr>
<td># unique fusion transcripts</td>
<td>37</td>
</tr>
<tr>
<td># in-frame</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 28. Summary of fusion transcripts discovered in 5 normal colon samples. (A) Characteristics of fusion transcripts. (B) Fusion transcripts discovered in per sample. # in-frame: number of the fusion transcripts remaining an open reading frame; # TICs: number of fusion transcripts predicted to be TICs; # probability ≥ 0.85: number of fusions having predicted probability ≥ 0.85.

<table>
<thead>
<tr>
<th>Gene1</th>
<th>Gene2</th>
<th>#Sample</th>
<th>Probability</th>
<th>TIC</th>
<th>PCR</th>
<th>Sequencing</th>
<th>Cancer specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>KISS1</td>
<td>GOLT1A</td>
<td>13</td>
<td>0.978</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>OSGIN1</td>
<td>NECAB2</td>
<td>2</td>
<td>0.971</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>CDH1</td>
<td>CDH3</td>
<td>1</td>
<td>0.965</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>IGSF5</td>
<td>PCP4</td>
<td>1</td>
<td>0.959</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>ADCK4</td>
<td>NUMBL</td>
<td>3</td>
<td>0.957</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>GFER</td>
<td>ZNRF2</td>
<td>1</td>
<td>0.941</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>UBXN11</td>
<td>AIM1L</td>
<td>1</td>
<td>0.929</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>EHMT2</td>
<td>MIF</td>
<td>1</td>
<td>0.914</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>RGN</td>
<td>RP2</td>
<td>1</td>
<td>0.882</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>WFDC3</td>
<td>TNNC2</td>
<td>4</td>
<td>0.870</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>
Table 29. Lists of the fusions that have been validated by PCR and Sanger Sequencing.

<table>
<thead>
<tr>
<th>NUP155</th>
<th>WDR70</th>
<th>1</th>
<th>0.854</th>
<th>N</th>
<th>Y</th>
<th>Y</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>KDM2A</td>
<td>SYT12</td>
<td>1</td>
<td>0.854</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

Discussion

Given that common gene fusions are rare in CRC (Bass et al., 2011), identifying case-specific gene fusions can help to understand the complexity of the molecular basis of CRC development. In addition, all the published studies of gene fusions in CRC were performed using primary CRC samples. Therefore, we used 20 liver metastasis of stage IV CRC to study the fusion transcripts in advanced stage CRC. Since only 1 of the 5 normal colon samples matches one of the CRC samples, we are not able to compare the fusions in paired samples to identify fusion transcripts potentially involved in CRC progression. However, the fusion transcripts discovered in CRC samples provide some candidate fusion transcripts for functional characterization. The RNA-seq dataset was analyzed for presence of fusion transcripts by using deFuse which is a bioinformatics tool for discovering fusion transcripts from RNA seq data. In the 20 CRC samples, we discovered 414 fusion events and 289 unique fusion transcripts. The majority of gene fusions were only found in a single patient, however, several were highly recurrent. The number of fusion genes varies from sample to sample (15 to 51 fusions per sample). Only 42 fusion transcripts retained an open reading frame, suggesting that majority of fusion transcripts do not encode chimeric proteins and they may function through other mechanisms. About half of the fusion transcripts (137/289) occurred between adjacent genes, suggesting that these fusion transcripts were transcriptional read-throughs or TICs rather than actual genome rearrangements. This result suggested that TICs are an important mechanism in forming fusion transcripts in CRC. Together, these results indicated that fusion transcripts and TICs are prevalent in CRC samples, while the in-frame fusions that potentially encode chimeric proteins are rare. In the 5 normal colon samples, 54 fusion events were discovered and 37 were unique fusion transcripts. On average, each sample had 10.8 fusion transcripts that was a lower number than CRC samples. Among the 37 unique fusion transcripts 24 were caused by TICs, suggesting
that TICs are also widely involved in formation of fusion transcripts in normal colon tissues. In consistent with the low occurrence of in-frame fusions in CRC samples, none of these fusion transcripts was in-frame. These results suggest that fusion transcripts and TICs also exist in normal colon samples, while the numbers were lower than in CRCs.

12 fusion candidates that meet the criteria of probability ≥ 0.85, in coding regions, and in-frame were validated by RT-PCR followed by DNA sequencing, suggesting the predicted probability ≥ 0.85 is a good predictor for the presence of a fusion transcript. 6 of these fusion transcripts were also discovered in normal colon samples by RT-PCR, although none of them were predicted by the RNA-seq data. This result suggests that RNA-seq may not be able to pick all the fusion transcripts in a specific sample. 7 of these fusion transcripts were TICs consistent with RNA-seq data that about 50% fusions were formed by TICs, suggesting that TICs are an important mechanism in formation of fusion transcripts. We also found that 3 of these TICs were cancer-specific, which were reported for the first time.

In ongoing studies, we will functionally characterize some of the fusion transcripts, especially cancer-specific TICs, to reveal the influence of fusion transcripts in tumor formation and to identify potential targets for novel therapies.

**Material and methods**

**Tumor and Normal Colon Samples**

The liver metastasis of 20 stage IV colon cancer patients and 5 normal colon samples were obtained from the Biological Materials Procurement Network (BioNet) at the University of Minnesota. All tumor and normal tissues were snap frozen in liquid nitrogen at the time of surgical resection, and delivered to us on dry ice, and transferred into liquid nitrogen immediately upon they were received. The clinico-pathological profile of the 20 CRC (Table 30) and 5 normal colon samples (Table 31) are listed below.

**Table 30. Summery of clinico-pathological profile of the 20 CRC patients**
<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>11</td>
<td>55</td>
</tr>
<tr>
<td>Female</td>
<td>9</td>
<td>45</td>
</tr>
<tr>
<td>Histology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td>19</td>
<td>95</td>
</tr>
<tr>
<td>Adenocarcinoid</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Differentiation*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td>Moderate</td>
<td>6</td>
<td>75</td>
</tr>
<tr>
<td>Poor</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>20</td>
<td>1</td>
</tr>
</tbody>
</table>

* Differentiation was not available for 12 cases.
Two patients received chemotherapy, but drugs were unknown. AD = Adenocarcinoma.

Table 31. Summary of clinico-pathological profile of the 5 normal colon samples

<table>
<thead>
<tr>
<th>Variable</th>
<th>Median</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, yr</td>
<td>68</td>
<td>(49-77)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>3</td>
<td>60</td>
</tr>
<tr>
<td>Female</td>
<td>2</td>
<td>40</td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left colon</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Right colon</td>
<td>4</td>
<td>80</td>
</tr>
<tr>
<td>Rectum</td>
<td>1</td>
<td>20</td>
</tr>
</tbody>
</table>

RNA Isolation

Total RNA was isolated from 20 fresh frozen CRC samples and 5 fresh frozen normal colon tissues as described in chapter 4. RNA integrity was assessed by capillary electrophoresis and the RNA Integrity Number (RIN) for each sample is listed as below (Table 32).

Table 32. RIN of the total 25 RNA samples
### Experimental Validation of Fusion Transcripts

We used RT-PCR followed by DNA sequencing to validate the presence of fusion genes transcripts in RNA samples.

**cDNA synthesis** was performed as described before.

**PCR.** All standard PCR was carried out with HotStarTaq DNA Polymerase using a HotStarTaq Master Mix Kit (Qiagen) with the following reaction mix: 25 µl of 2 X HotStarTaq Master Mix (2.5 units HotStarTaq DNA Polymerase, 1 x PCR Buffer, 200 µM of each dNTP), 10 µl of each primer (final concentration 0.2 µM), and 5 µl of cDNA template in a 50 µl volume. Cycling conditions were 95°C for 5 minutes, then 35 cycles.
of 95°C for 30 seconds, 60°C for 30 seconds, 72°C for 1 minute per kb of target, and finally 72°C for 10 minutes. The primers used for PCR are listed (Table 33).

<table>
<thead>
<tr>
<th>Fusion</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>KISS1-GOLT1A</td>
<td>GCAACTTAAACCAAGGGTCAGC</td>
<td>GCAGTAGCTGCCAAGAAACC</td>
</tr>
<tr>
<td>OSGIN1-NECAB2</td>
<td>CAAGAAGGTTCTGGGGGCAATT</td>
<td>TCCACATGGTTGTTGTTCGA</td>
</tr>
<tr>
<td>CDH1-CDH3</td>
<td>GTCTCCTCTTGGCTCTTGCC</td>
<td>CTTTCCTGGACTGTCTCGCC</td>
</tr>
<tr>
<td>IGSF5-PCP4</td>
<td>AAAGACCACAGACACCCGCTT</td>
<td>ACTGAGACTGAATGGCCACC</td>
</tr>
<tr>
<td>ADCK4-NUMBL</td>
<td>TACATCGAGGGTGTTGAAGGC</td>
<td>GTAACCTTTGTTGTTTGGCC</td>
</tr>
<tr>
<td>GFER-ZNRF2</td>
<td>GTCCTGCTCACTCTCGCA</td>
<td>CCAGCCGCTATCCACTTTTGG</td>
</tr>
<tr>
<td>UBXN11-AIM1L</td>
<td>ACTGGAGACGATGTGGTGGTC</td>
<td>TTTCTGTCTCCACATGGTCC</td>
</tr>
<tr>
<td>EHMT2-MIF</td>
<td>CCAGAGTTTGGCTATGAGGCT</td>
<td>GCTTGGCTGTTAGGAGGCGGT</td>
</tr>
<tr>
<td>RP2-RGN</td>
<td>AATATCCCGGGGTCAGAGACA</td>
<td>TTGTGGCTTACCGGACAAATCC</td>
</tr>
<tr>
<td>WFDC3-TNNC2</td>
<td>CTCCGCGATCTCCGACAGGA</td>
<td>TCATCAGTTACACAGCGACCAA</td>
</tr>
<tr>
<td>NUP155-WDR70</td>
<td>TCCCTGTTTCCCTTCCTGCTACA</td>
<td>TTTTTGCCCTTGATCGCTCC</td>
</tr>
<tr>
<td>KDM2A-SYT12</td>
<td>CAGCCCTGGAGTGTTTCTT</td>
<td>GACTTCCGGAGGGTTCCCA</td>
</tr>
</tbody>
</table>

**Table 33. List of primers used for validation of candidate fusion transcripts.**

**Gel extraction.** PCR products were subjected to electrophorisis using a 1% low melting agarose (LMT) gel with 0.05% ethidium bromide. The desired bands on gel were excised from the gel using a sterile scalpel and the PCR products were then extracted from the gel slice using the QIAquick Gel extraction kit (Qiagen) following the manufacturer’s instructions. The extracted DNA was quantified using Nanodrop.

**Sanger sequencing.** The reaction mix comprises 50 ng DNA extracted from LMT gel, 1.6 µl of 2 µM sequencing primer stock and H2O to make a final volume of 12 µl. The reactions were sent to the University of Minnesota BioMedical Genomics Center for sequencing by using an ABI 3730 DNA sequencer (Applied Biosystems). Results were screened by eye using the Applied Biosystems software packages Sequence Scanner version 1.0 (Applied Biosystems). Sequencing results were compared with the RNA seq data to confirm the fusion transcripts. The sequencing primers are listed in Table 34.
Table 34. List of sequencing primers for validation of candidate fusion transcripts.

Illumina RNA-Sequencing was performed as described in chapter 4.

deFuse. Fusion transcript discovery was performed using deFuse (http://compbio.bccrc.ca). Briefly, deFuse identifies fusion transcripts by clustering discordantly aligning paired-end reads as potential evidence of reads that span a fusion breakpoint. Clusters of discordantly aligning reads are then used to inform a targeted search for reads split by fusion breakpoints. The results produced by deFuse are further filtered to reduce the number of false positives. After these series of filters are applied, a candidate list of fusions is carried forward for further analysis (McPherson et al., 2011).

Supplemental information

Table 35. Summary of fusion transcripts discovered in each CRC sample

<table>
<thead>
<tr>
<th>Sample</th>
<th># Fusions</th>
<th># In Frame</th>
<th># Adjacent</th>
<th># prob≥0.85</th>
</tr>
</thead>
<tbody>
<tr>
<td>T980429-5</td>
<td>51</td>
<td>3</td>
<td>31</td>
<td>7</td>
</tr>
<tr>
<td>T020556-4</td>
<td>49</td>
<td>2</td>
<td>25</td>
<td>11</td>
</tr>
<tr>
<td>T980321-1</td>
<td>49</td>
<td>8</td>
<td>26</td>
<td>8</td>
</tr>
<tr>
<td>T020201-4</td>
<td>48</td>
<td>8</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>T010314</td>
<td>45</td>
<td>5</td>
<td>29</td>
<td>12</td>
</tr>
<tr>
<td>T001683</td>
<td>42</td>
<td>3</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>T001090</td>
<td>38</td>
<td>2</td>
<td>26</td>
<td>10</td>
</tr>
<tr>
<td>T970192-5</td>
<td>35</td>
<td>0</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>T990896-2</td>
<td>33</td>
<td>2</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>T021380</td>
<td>32</td>
<td>2</td>
<td>22</td>
<td>12</td>
</tr>
<tr>
<td>T010004</td>
<td>31</td>
<td>2</td>
<td>19</td>
<td>2</td>
</tr>
<tr>
<td>T970854-6</td>
<td>29</td>
<td>3</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>T990569-2</td>
<td>26</td>
<td>3</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>Gene1</td>
<td>Gene2</td>
<td>#Sample (20 total)</td>
<td>In Frame</td>
<td>Adjacent</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------</td>
<td>-------------------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>RP11-150O12.3</td>
<td>RP11-150O12.5</td>
<td>14</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>GOLT1A</td>
<td>KISS1</td>
<td>13</td>
<td>0.23</td>
<td>Y</td>
</tr>
<tr>
<td>MT1P1</td>
<td>RP11-435O5.2</td>
<td>13</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>ANO1</td>
<td>RP11-626H12.1</td>
<td>12</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>NME4</td>
<td>Z97634.5</td>
<td>11</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>AC036222.1</td>
<td>HOXB9</td>
<td>11</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>AC022098.2</td>
<td>PALM3</td>
<td>11</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>AL360181.1</td>
<td>RP11-108K14.4</td>
<td>10</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>BAG6</td>
<td>C6orf47</td>
<td>10</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>AC017074.2</td>
<td>ACR</td>
<td>10</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>PAQR6</td>
<td>SMG5</td>
<td>9</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>PPP1R1B</td>
<td>STAR3D9</td>
<td>8</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>GSK3B</td>
<td>NR1I2</td>
<td>8</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>C19orf63</td>
<td>JOSD2</td>
<td>8</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>C2orf49</td>
<td>FHL2</td>
<td>8</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>AC007405.7</td>
<td>Y RNA.703</td>
<td>7</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>AC036222.1</td>
<td>HOXB7</td>
<td>7</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>AC011383.1</td>
<td>SYNPO</td>
<td>7</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>C5orf32</td>
<td>CTB-131B5.5</td>
<td>7</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>MICAL2</td>
<td>MICALCL</td>
<td>6</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>HOTAIR1M1</td>
<td>HOXA3</td>
<td>6</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>AP003774.1</td>
<td>CCDC88B</td>
<td>6</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>MRPS36P2</td>
<td>RP11-11N5.1</td>
<td>6</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>AC036222.1</td>
<td>CTD-2377D24.6</td>
<td>5</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>TNNC2</td>
<td>WDFC3</td>
<td>4</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>AKR1A1</td>
<td>PRDX1</td>
<td>4</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>CDC42EP2</td>
<td>POLA2</td>
<td>4</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>MORN1</td>
<td>RP4-740C4.4</td>
<td>4</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>AC018717.1</td>
<td>AC019330.1</td>
<td>4</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>AC090425.1</td>
<td>NDUFB5</td>
<td>4</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>CALM3</td>
<td>PTGIR</td>
<td>4</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>HADH</td>
<td>RP11-286E11.1</td>
<td>4</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>RP11-11N5.1</td>
<td>RP11-217E13.1</td>
<td>4</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Table 36. Summary of fusion transcripts discovered in the 20 CRC samples
<table>
<thead>
<tr>
<th>MIR633</th>
<th>TANC2</th>
<th>4</th>
<th>N</th>
<th>Y</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP11-191N8.2</td>
<td>RP11-185M8.1</td>
<td>3</td>
<td>N</td>
<td>Y</td>
<td>0.83</td>
</tr>
<tr>
<td>CLDN10</td>
<td>DZIP1</td>
<td>3</td>
<td>N</td>
<td>Y</td>
<td>0.67</td>
</tr>
<tr>
<td>CKAP5</td>
<td>F2</td>
<td>3</td>
<td>N</td>
<td>Y</td>
<td>0.6</td>
</tr>
<tr>
<td>ADCK4</td>
<td>NUMBL</td>
<td>3</td>
<td>Y</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>AL049836.1</td>
<td>DIO3</td>
<td>3</td>
<td>N</td>
<td>N</td>
<td>0.4</td>
</tr>
<tr>
<td>AC100791.3</td>
<td>TBC1D16</td>
<td>3</td>
<td>N</td>
<td>N</td>
<td>0.4</td>
</tr>
<tr>
<td>ARSD</td>
<td>GYG2P1</td>
<td>3</td>
<td>N</td>
<td>N</td>
<td>0.25</td>
</tr>
<tr>
<td>CARD10</td>
<td>MFNG</td>
<td>3</td>
<td>N</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>RP1-10C16.1</td>
<td>RP11-375F2.2</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0.8</td>
</tr>
<tr>
<td>NCBP2</td>
<td>PIGZ</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0.75</td>
</tr>
<tr>
<td>ATAD5</td>
<td>EIF5A</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0.75</td>
</tr>
<tr>
<td>AFM</td>
<td>RP11-622A1.2</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0.71</td>
</tr>
<tr>
<td>ACOT12</td>
<td>ZCCHC9</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0.67</td>
</tr>
<tr>
<td>PFKFB3</td>
<td>Y_RNA.738</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0.67</td>
</tr>
<tr>
<td>AL137145.2</td>
<td>PFKFB3</td>
<td>2</td>
<td>0.33_Y</td>
<td>N</td>
<td>0.67</td>
</tr>
<tr>
<td>METTL11A</td>
<td>RP11-492E3.2</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>MAP6D1</td>
<td>PARL</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>AC087699.1</td>
<td>CUX1</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>AC079080.2</td>
<td>RP11-162O12.2</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>RP11-255B23.3</td>
<td>RP11-255B23.4</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>AL139815.1</td>
<td>MSRB2</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>DUS3L</td>
<td>FUT6</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>PITPNA</td>
<td>YWHAE</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>ACE2</td>
<td>BMX</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0.4</td>
</tr>
<tr>
<td>AC007405.6</td>
<td>Y_RNA.703</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0.33</td>
</tr>
<tr>
<td>TTL12</td>
<td>TTLL12</td>
<td>2</td>
<td>Y</td>
<td>N</td>
<td>0.33</td>
</tr>
<tr>
<td>NECAB2</td>
<td>OSGIN1</td>
<td>2</td>
<td>0.33_Y</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>TNFRSF9</td>
<td>UBE4B</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>TBC1D5</td>
<td>ZMAT3</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.8</td>
</tr>
<tr>
<td>CDH1</td>
<td>CDH3</td>
<td>1</td>
<td>0.40_Y</td>
<td>Y</td>
<td>0.8</td>
</tr>
<tr>
<td>NFS1</td>
<td>RBM39</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.75</td>
</tr>
<tr>
<td>CEACAM7</td>
<td>RPS19</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.75</td>
</tr>
<tr>
<td>IGSF5</td>
<td>PCP4</td>
<td>1</td>
<td>0.50_Y</td>
<td>Y</td>
<td>0.75</td>
</tr>
<tr>
<td>SCN2A</td>
<td>SCN3A</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.75</td>
</tr>
<tr>
<td>C17orf99</td>
<td>SYNGR2</td>
<td>1</td>
<td>0.25_Y</td>
<td>Y</td>
<td>0.75</td>
</tr>
<tr>
<td>CIITA</td>
<td>DEX1</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.67</td>
</tr>
<tr>
<td>TBC1D5</td>
<td>Y_RNA.191</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.67</td>
</tr>
<tr>
<td>OAZ2</td>
<td>ZNF609</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.67</td>
</tr>
<tr>
<td>ATXN10</td>
<td>RP4-695O20_B.10</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.67</td>
</tr>
<tr>
<td>RP11-142A23.1</td>
<td>RP11-656G20.1</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.67</td>
</tr>
<tr>
<td>SLC25A33</td>
<td>SPSB1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.67</td>
</tr>
<tr>
<td>DCBLD2</td>
<td>MINA</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>HSP90AB1</td>
<td>SLC29A1</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>AMPD2</td>
<td>STXBP3</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>AQP5</td>
<td>AQP6</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>CAPRIN2</td>
<td>FAR2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>SERINC2</td>
<td>WASF2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>Gene</td>
<td>Gene</td>
<td>Value</td>
<td>Value</td>
<td>Value</td>
<td>Value</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>FAM19A1</td>
<td>PTPRG</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>TBL3</td>
<td>ZNRF2</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>KIF16B</td>
<td>RP3-410C9.1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>MND1</td>
<td>TRIM2</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>BLM</td>
<td>RCDC1</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>SYT8</td>
<td>TNNI2</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>RP5-1024N4.1</td>
<td>SKINTL</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>L2HGDH</td>
<td>MAP4K5</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>EIF2C2</td>
<td>WDR67</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>ACAP2</td>
<td>VPS8</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>BCL2L13</td>
<td>RP11-14C22.6</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>C1orf21</td>
<td>RNF2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>RGN</td>
<td>RP2</td>
<td>1</td>
<td>0.50</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>PLCE1</td>
<td>SLC5A11</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>RP11-285C1.2</td>
<td>TRIM2</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>BCLAF1</td>
<td>ENC1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>SMN1</td>
<td>TIMM10</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>RP4-610C12.1</td>
<td>SMUG1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>AC093768.1</td>
<td>SLC29A1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>GLG1</td>
<td>STAC2</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>MYOM1</td>
<td>NEDD4L</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>NPR2</td>
<td>RP11-112J3.16</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>AC005682.6</td>
<td>MAP4</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>CEP63</td>
<td>RP11-680B3.2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>ATP6V1H</td>
<td>PLEKHA2</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>AL591415.1</td>
<td>RP11-543D5.1</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>DTD1</td>
<td>RP11-379J5.2</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>CBX1</td>
<td>RP11-390F4.6</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>ATP5G1P7</td>
<td>TSPAN15</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>AP001610.9</td>
<td>AP006748.1</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>GGT1</td>
<td>IGSF3</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>TPD52</td>
<td>U6.413</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>AC091849.1</td>
<td>CTD-2245E15.3</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>CACNB1</td>
<td>RPL19</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>LRPPRC</td>
<td>SPDYA</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>BCL7C</td>
<td>MGRN1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>RP11-1L12.3</td>
<td>SLC5A12</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>CDC45</td>
<td>PARVB</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>AF064860.7</td>
<td>IGSF5</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>CTB-63M22.1</td>
<td>PLEKHM3</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>SPAG17</td>
<td>WARS2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>BDP1</td>
<td>NAIP</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>AC011897.2</td>
<td>NBAS</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>CTNND2</td>
<td>SRFBP1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>C4orf3</td>
<td>FABP2</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>AC019181.2</td>
<td>SLC38A11</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0.5</td>
</tr>
<tr>
<td>GLS</td>
<td>STAT4</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>CDH1</td>
<td>TMCO7</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>NCAPH2</td>
<td>NPTXR</td>
<td>1</td>
<td>0.10</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Gene</td>
<td>Gene</td>
<td>Value</td>
<td>Value</td>
<td>Value</td>
<td>Value</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>C17orf37</td>
<td>GRB7</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.17</td>
</tr>
<tr>
<td>PLEKHO1</td>
<td>VPS45</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>RP11-439A17.7</td>
<td>SRGAP2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>CDC37</td>
<td>KISS1R</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>AIM1L</td>
<td>UBXN11</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>TEX9</td>
<td>USP3</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>HOXC5</td>
<td>HOXC9</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>ATP6V1B1</td>
<td>VAX2</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>RP1-154K9.2</td>
<td>U6.507</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>SMAP1</td>
<td>TPBG</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>RP11-322M19.1</td>
<td>VT11A</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>NUP155</td>
<td>WDR70</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>THBP1</td>
<td>ZNF554</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>KDM2A</td>
<td>SYT12</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>ARFGEF1</td>
<td>KCNB2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>5S_rRNA,78</td>
<td>PMM2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>C1orf94</td>
<td>KIAA1522</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>GFER</td>
<td>ZNRF2</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>AC004066.3</td>
<td>PPA2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>C20orf26</td>
<td>NINL</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>CXorf31</td>
<td>PHF16</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>CORO7</td>
<td>HMOX2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>RNF38</td>
<td>RP11-84P7.2</td>
<td>4</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>AC005532.5</td>
<td>C1GALT1</td>
<td>3</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>ATF1</td>
<td>CSRNP2</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>FGB</td>
<td>RP5-857K21.4</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>RP5-857K21.6</td>
<td>RPS18</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>ATIC</td>
<td>FN1</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>FABP1</td>
<td>SMAD2</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>PDDPF</td>
<td>PTK6</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>AC006008.1</td>
<td>ACTR3B</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>KRT18</td>
<td>LOH12CR1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>BCL7C</td>
<td>PLCE1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>SKAP1</td>
<td>Y RNA.493</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>METTL4</td>
<td>SMCHD1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>AC091172.1</td>
<td>WNT3</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>FBXL6</td>
<td>ZNF252</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>ABLIM3</td>
<td>CTC-529P8.1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>GSDMD</td>
<td>RP11-661A12.5</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>RP11-113G13.2</td>
<td>USO1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>C1orf94</td>
<td>CSMO2</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>AC091320.2</td>
<td>RP11-328J2.1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>EDARADD</td>
<td>ERP44</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>C9orf163</td>
<td>SEC16A</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>CR381670.1</td>
<td>PHF20</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>NBAS</td>
<td>RP11-339A11.1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>PTGR1</td>
<td>RNASE10</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>RGS20</td>
<td>RP11-419C23.1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>GDAP1L1</td>
<td>HNF4A</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>Gene</td>
<td>Location</td>
<td>NUM</td>
<td>Upper</td>
<td>Lower</td>
<td>Score</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------</td>
<td>-----</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>LIMA1</td>
<td></td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>MIR1268</td>
<td>TMEM87A</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>EFHA1</td>
<td>ZDHHC20</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>CPSF1</td>
<td>WWOX</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>RP11-504G3.1</td>
<td>UEVLD</td>
<td>20</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>C19orf21</td>
<td>PALM</td>
<td>17</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>ANKDD1A</td>
<td>SPG21</td>
<td>16</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>PLCB3</td>
<td>RP11-783K16.13</td>
<td>16</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>AP000963.2</td>
<td>CXADR</td>
<td>13</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>EDF1</td>
<td>PHPT1</td>
<td>12</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>CCDC162</td>
<td>RP11-425D10.1</td>
<td>11</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>C4orf48</td>
<td>WHSC2</td>
<td>11</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>AC016735.1</td>
<td>AC016735.2</td>
<td>10</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>LRRC25</td>
<td>SSBP4</td>
<td>8</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>CTA-398F10.2</td>
<td>TSPAN15</td>
<td>7</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>GRIN2D</td>
<td>KDEL1</td>
<td>6</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>PDE4DIP</td>
<td>RP11-289I10.2</td>
<td>6</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>BLCAP</td>
<td>RPL7AP14</td>
<td>6</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>RP11-395P17.3</td>
<td>SLC25A25</td>
<td>5</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>FAM83H</td>
<td>RP11-429I17.6</td>
<td>5</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>AF064858.11</td>
<td>AF064858.8</td>
<td>5</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>PLCH2</td>
<td>RP4-740C4.4</td>
<td>5</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>CEACAM5</td>
<td>CEACAM7</td>
<td>5</td>
<td>0.12</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>CFR11</td>
<td>CTTNB2P2</td>
<td>4</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>ACTR8</td>
<td>IL17RB</td>
<td>4</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>KIAA1274</td>
<td>PRF1</td>
<td>4</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>NLRP6</td>
<td>RP11-326C3.2</td>
<td>4</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>DPFA2P1</td>
<td>UTY</td>
<td>4</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>MT-ATP6</td>
<td>MT-ND4</td>
<td>3</td>
<td>0.33</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>GBP4</td>
<td>GBP7</td>
<td>3</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>CTC-537E7.1</td>
<td>CTC-537E7.2</td>
<td>3</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>CTAGE5</td>
<td>MIA2</td>
<td>3</td>
<td>Y</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>GRPR</td>
<td>RP11-431J24.2</td>
<td>3</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>ITIH1</td>
<td>NEK4</td>
<td>3</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>RP5-857K21.8</td>
<td>TPT1</td>
<td>3</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>ATP5F1</td>
<td>C1orf162</td>
<td>3</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>PDE4DIP</td>
<td>RP11-277L2.2</td>
<td>3</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>AHSG</td>
<td>CT45A6</td>
<td>3</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>AC012363.4</td>
<td>EPB41L5</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>CATSPERG</td>
<td>KCNK6</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>CTD-2165H16.1</td>
<td>SMAD2</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>DNAH2</td>
<td>KDM6B</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>CCDC102A</td>
<td>GPR114</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>CCNO</td>
<td>DHX29</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>AC083849.1</td>
<td>CNTNAP2</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>MT-ATP6</td>
<td>MT-CYB</td>
<td>2</td>
<td>0.50</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>DMBX1</td>
<td>RP5-110J22.3</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>AC091814.3</td>
<td>DENND3</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>MT-ATP6</td>
<td>MT-ND4</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
</tbody>
</table>

232
<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene</th>
<th>Value</th>
<th>Value</th>
<th>Value</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDPSP8</td>
<td>NEK2</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>AC022201.4</td>
<td>BRD7P6</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>KNDC1</td>
<td>VENTX</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>CTD-3035K6.1</td>
<td>RP5-857K21.6</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>FSIP1</td>
<td>THBS1</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>NCRNA00106</td>
<td>SLC25A6</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>MALT1</td>
<td>U6.282</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>NCOA3</td>
<td>SULF2</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>EPCAM</td>
<td>RP5-857K21.4</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>AP001187.1</td>
<td>AP001187.9</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>FPR3</td>
<td>ZNF577</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>RP5-857K21.4</td>
<td>RPL8</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>MT-ND4</td>
<td>MT-ND6</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>AC006000.4</td>
<td>AC01288.2</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>LYPLAL1</td>
<td>RP11-135J2.3</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>MT-CYB</td>
<td>MT-ND2</td>
<td>2</td>
<td>0.40_Y</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>AVI.9</td>
<td>DPY19L1</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>SERPINF2</td>
<td>WDR81</td>
<td>2</td>
<td>0.33_Y</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>HPGD</td>
<td>RP11-440114.2</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>AAGAB</td>
<td>GRINL1A</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>AP2B1</td>
<td>PIPOX</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>RP11-434D9.2</td>
<td>RP11-44F21.4</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>AC013564.1</td>
<td>SMAD6</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>J01415.25</td>
<td>MT-ND6</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>CTD-215J2.1</td>
<td>WDR70</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>KDM6A</td>
<td>RP11-185O17.3</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>MT-ATP6</td>
<td>MT-ND2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>ATP8D4</td>
<td>NIT1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>AAK1</td>
<td>AC092431.3</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>ACTR3B</td>
<td>RP11-728K20.1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>KCTD8</td>
<td>STXBP3</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>HMGN3</td>
<td>Y RNA.293</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>ERBB2</td>
<td>WDR59</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>ATP8A2</td>
<td>GTF2F2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>BTBD8</td>
<td>KIAA1107</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>RP11-834C11.5</td>
<td>SMUG1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>ABCC1</td>
<td>ARID1B</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>ARL13A</td>
<td>TRMT2B</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>5S rRNA.160</td>
<td>PEX5L</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>TOP1</td>
<td>U2.20</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>KDM5B</td>
<td>MACF1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>BGN</td>
<td>SLC25A6</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>CACNB1</td>
<td>RYR2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>GATA1</td>
<td>U6.681</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>HPCAL1</td>
<td>ODC1</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>CEP70</td>
<td>MRPS22</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>GPR35</td>
<td>PLA2G6</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>RP4-779E11.3</td>
<td>SMOX</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>LCTL</td>
<td>SMAD6</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
</tbody>
</table>

233
Table 37. Summary of fusion transcripts discovered in each normal sample

<table>
<thead>
<tr>
<th>Sample</th>
<th># Fusions</th>
<th># In Frame</th>
<th># Adjacent</th>
<th># prob ≥ 0.85</th>
</tr>
</thead>
<tbody>
<tr>
<td>T011145-9</td>
<td>14</td>
<td>0</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>T020447-3</td>
<td>10</td>
<td>0</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>T020613-1</td>
<td>9</td>
<td>0</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>T020627-1</td>
<td>12</td>
<td>0</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>T022437-2</td>
<td>9</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 38. Summary of fusion transcripts discovered in 5 normal samples

<table>
<thead>
<tr>
<th>Gene1</th>
<th>Gene2</th>
<th>Sample (5 total)</th>
<th>In Frame</th>
<th>Adjacent</th>
<th>Probability ≥ 0.85</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP001187.1</td>
<td>AP001187.9</td>
<td>3</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>SYNPO</td>
<td>AC011383.1</td>
<td>4</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>FABP2</td>
<td>C4orf3</td>
<td>3</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>AP001610.9</td>
<td>AP006748.1</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0.5</td>
</tr>
<tr>
<td>BAG6</td>
<td>C6orf47</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>EPS15</td>
<td>CALR4P</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>GIMAP2</td>
<td>ATPB1</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>HOXB7</td>
<td>AC036222.1</td>
<td>3</td>
<td>N</td>
<td>Y</td>
<td>0.33</td>
</tr>
<tr>
<td>KDM6B</td>
<td>DNAH2</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>RP11-286E11.1</td>
<td>HADH</td>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>RP11-504G3.1</td>
<td>UEVLD</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>AC036222.1</td>
<td>HOXB9</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>Gene1</td>
<td>Gene2</td>
<td># In CRC</td>
<td># In normal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>----------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACR</td>
<td>AC017074.2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>ANKDD1A</td>
<td>SPG21</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>AP000963.2</td>
<td>CXADR</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>BCDIN3D</td>
<td>RP11-202D20.1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>CD200</td>
<td>RP11-231E6.1</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>CORO1C</td>
<td>Y_RNA.117</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>CTD-2165H16.1</td>
<td>SMAD2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>EDF1</td>
<td>PHPT1</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>EPCAM</td>
<td>PCBD2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>GADD45GIP1</td>
<td>DAND5</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>GSDMD</td>
<td>RP11-661A12.5</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>MEP1A</td>
<td>MEP1AP4</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>MICAL2</td>
<td>MICALCL</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>MT-ATP6</td>
<td>MT-ND6</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>MTCH2</td>
<td>MED15</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>MYH11</td>
<td>AC092024.2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>NME4</td>
<td>Z97634.5</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>RP11-230B22.1</td>
<td>ATG4C</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>RP11-271E2.1</td>
<td>FGD5</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>RP11-783K16.13</td>
<td>PLCB3</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>RP5-857K21.4</td>
<td>RP11-433O3.1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>SGMS2</td>
<td>HADH</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>SLC25A25</td>
<td>RP11-395P17.3</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>SLC25A45</td>
<td>FRMD8</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>UTY</td>
<td>DPPA2P1</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
</tbody>
</table>

# Samples: number of samples this gene fusions has been discovered in
In-frame: if the fusion transcript remains an open reading frame
Adjacent: if the two genes are located in adjacent
Probability ≥ 0.85: percentage of this fusions having predicted probability ≥ 0.85

Table 39. The overlapping fusion transcripts in CRC and normal samples
<table>
<thead>
<tr>
<th>DPPA2P1</th>
<th>UTY</th>
<th>4</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDF1</td>
<td>PHPT1</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>GSDMD</td>
<td>RP11-661A12.5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>HADH</td>
<td>RP11-286E11.1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>MICAL2</td>
<td>MICALCL</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>MT-ATP6</td>
<td>MT-ND6</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>NME4</td>
<td>Z97634.5</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>PLCB3</td>
<td>RP11-783K16.13</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>RP11-395P17.3</td>
<td>SLC25A25</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>RP11-504G3.1</td>
<td>UEVLD</td>
<td>20</td>
<td>2</td>
</tr>
</tbody>
</table>

# In CRC  number of CRC samples this fusion has been discovered in
# In normal  number of normal samples this fusion has been discovered in