Title
Loss of stemness in the normal crypt, intestinal wounding, and colon cancer

Author
Habowski, Amber Nicole

Publication Date
2020

UNIVERSITY OF CALIFORNIA,
IRVINE

Loss of stemness in the normal crypt, intestinal wounding, and colon cancer

DISSERTATION

Submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
in Biomedical Sciences

by

Amber Nicole Habowski

Dissertation Committee:
Professor Marian L. Waterman, Chair
Professor Klemens J. Hertel
Professor Yongsheng Shi
Assistant Professor Selma Masri
Assistant Professor Kai Kessenbrock

2020

DEDICATION

This dissertation is dedicated to

My husband,
Tyler Habowski

And

My PI,
Marian Waterman

For their immense support, encouragement, and love which has been the foundation for my success.

"There is plenty of WOW left in the universe to discover!"

- Dr. Waterman

TABLE OF CONTENTS

Page

LIST OF FIGURES vi

LIST OF TABLES ix

ACKNOWEDLEGEMENTS x

CURRICULUM VITAE xiv

ABSTRACT OF THE DISSERTATION xxii

CHAPTER 1: Introduction

Introduction to the Intestine. 2

Signaling and the Transcriptome 3

Intestinal Homeostasis 5

Investigating Intestinal Stemness 9

Molecular Biology Experimental Approaches 9

Chapter Guide 13

Figures 14

References 17

CHAPTER 2: Transcriptomic and Proteomic Signatures of Stemness and Differentiation in the Colon Crypt

Abstract 22

Introduction 22

Results 25

Discussion 37

Methods 41

Figures 49

Supplementary Figures 62

Tables 115

References 119

CHAPTER 3: Regionally Distinct Responses to Irradiation Wounding of the Intestine

Abstract 126

Introduction 126

Results 129
Discussion and Future Work 136
Methods 144
Figures 149
Supplementary Figures 161
References 178

CHAPTER 4: α-Ketoglutarate Attenuates Wnt Signaling and Drives Differentiation in Colorectal Cancer

Abstract 183
Introduction 183
Results 185
Discussion 194
Methods 196
Figures 207
Tables 234
References 237

CHAPTER 5: Discussion

Normal Intestine 242
Wounded Intestine 245
Colon Cancer. 246
Intestinal Stemness 248
References 249

APPENDIX A: Isolation of Murine Large Intestinal Crypt Cell Populations with Flow Sorting

Abstract 253
Introduction 253
Reagents and Equipment 254
Procedure 256
Troubleshooting 263
Results 266
Figures 267

References 272

APPENDIX B: Tissue Biobanking and Human Colon Organoids

Introduction to Organoids 274
Status of Biobank. 275
Solutions 277
Conditioned Media 278
Organoid Protocols 281
Figures 286
Tables 291
References 297

APPENDIX C: GECO: Gene Expression Clustering Optimization App for Non-Linear Data Visualization of Patterns

Abstract 299
Background 300
Implementation 303
Results and Discussion 306
Conclusion. 308
Availability and Requirements 309
GECO README. 309
Figures 318
References 326

LIST OF FIGURES

Page

CHAPTER 1

Figure 1.1 Structure of the colonic crypt . . . 14

Figure 1.2 mRNA processing influences the transcriptome . . . 16

CHAPTER 2

Figure 2.1 A novel flow sorting protocol that purifies six murine colon crypt cell populations . . . 49

Figure 2.2 Characterization of intestinal stemness based on differential gene expression . . . 50

Figure 2.3 A burst of alternative mRNA processing activity during loss of stemness . . . 52

Figure 2.4 Alternative splicing and polyadenylation changes that occur during intestinal crypt loss of stemness . . . 54

Figure 2.5 mRNA processing contributes to lineage commitment . . . 56

Figure 2.6 Lineage commitment to secretory and absorptive lineages are influenced by signaling pathways . . . 57

Figure 2.7 Fate commitment and signaling characteristics of mature crypt cells . . . 59

Figure 2.8 Summary of transcriptome and proteome changes in colon crypt homeostasis . . . 61

Supp. Fig. 2.1 FACS gating strategies that define six colon crypt cell populations . . . 62

Supp. Fig. 2.2 FACS plots of colon crypts dissociated with or without TrypLE protease treatment . . . 64

Supp. Fig. 2.3 Validation of sorted stem cell population with Lgr5-EGFP-IRES-creERT2 mice . . . 65

Supp. Fig. 2.4 Expression of known intestinal crypt markers . . . 66

Supp. Fig. 2.5 SecPDG marker genes label both goblet and deep crypt secretory cells . . . 68

Supp. Fig. 2.6 Expression of known markers detects different subtypes of enteroendocrine cells . . . 71

Supp. Fig. 2.7 Markers enriched in tuft cells . . . 73

Supp. Fig. 2.8 Markers enriched in enteroendocrine cells . . . 76

Supp. Fig. 2.9 Markers enriched in enterocytes . . . 79

Supp. Fig. 2.10 Colonic tuft cells show a distinct potential for taste pathway activation via fatty acids . . . 82

Supp. Fig. 2.11 Mass Spec analysis detected proteins expressed in the crypt cell types . . . 84

Supp. Fig. 2.12 Comparison of mRNA to protein in each crypt cell type . . . 86

Supp. Fig. 2.13 Surface marker protein expression correlations with mRNA gene expression . . . 89

Supp. Fig. 2.14 Sorting procedure is works broadly in other mouse strains and gender 90
Supp. Fig. 2.15 Additional markers of intestinal stemness 91
Supp. Fig. 2.16 Apoptosis related gene expression in the intestinal crypt 93
Supp. Fig. 2.17 Regulation of the circadian clock in the crypt 95
Supp. Fig. 2.18 Expression of proliferation and cell cycle genes 96
Supp. Fig. 2.19 Alternative mRNA processing changes in the intestinal crypt 98
Supp. Fig. 2.20 Gene ontology analysis of alternatively processed mRNA in stem cells and progenitors 100
Supp. Fig. 2.21 Common mRNA processing events 102
Supp. Fig. 2.22 Expression and processing data for examples of alternatively processed genes 104
Supp. Fig. 2.23 Intestinal staining of alternatively polyadenylated genes in human intestinal tissues . . 106
Supp. Fig. 2.24 Expression and splicing of adhesion genes 107
Supp. Fig. 2.25 Prostaglandin and Leukotriene production and signaling in the intestinal crypt 109
Supp. Fig. 2.26 Unsupervised clustering of crypt signaling pathway components 111
Supp. Fig. 2.27 Additional influences on signaling 112
Supp. Fig. 2.28 Human intestinal expression of the EGFR 113
Supp. Fig. 2.29 Colonic expression of Aquaporins 114

CHAPTER 3

Figure 3.1 Assessment of intestinal damage and recovery 1-7 days following irradiation 149
Figure 3.2 Single cell RNA-sequencing analysis of the colonic epithelial crypt response to irradiation. 151
Figure 3.3 Identifying colonic epithelial crypt cell populations. 153
Figure 3.4 Irradiation-induced changes in colonic crypt cell populations 155
Figure 3.5 Proliferation increases four days post irradiation. 157
Figure 3.6 Mucin and Antimicrobial Peptide expression increases following irradiation 159
Supp. Fig. 3.1 Murine small intestine displays severe villi damage following irradiation. 161
Supp. Fig. 3.2 Murine colonic crypt displays regional differences in length and wounding severity from irradiation. 162
Supp. Fig. 3.3 Clustree analysis of single cell RNA-sequencing 164
Supp. Fig. 3.4 Sub-clustering of Cluster 0 reveals a secretory cell type enriched on day 4 post irradiation. 166

Supp. Fig. 3.5 Elevated mRNA expression of apoptosis pathway components in cells located at the base of crypts. 168
Supp. Fig. 3.6 Wnt signaling potential and downstream target expression are minimal in non-stem cells. 170
Supp. Fig. 3.7 β-catenin is broadly expressed in colonic crypt cells. 172
Supp. Fig. 3.8 Cma1 expression peaks on day 4 following irradiation. 174
Supp. Fig. 3.9 Retnlb1 expression increases during recovery from irradiation. 176

CHAPTER 4

Figure 4.1 The effect of glutamine starvation on *Apc*$^{Min/+}$ and wildtype organoids 207
Figure 4.2 Environmental glutamine restriction hyperactivates Wnt signaling and blocks cellular differentiation 209
Figure 4.3 Genetic alterations do not contribute to low-glutamine induced stemness 211
Figure 4.4 Glutamine restriction promotes self-renewal and niche independence in *Apc*$^{Min/+}$ organoids 213
Figure 4.5 The role of αKG in low-glutamine induced stemness 215
Figure 4.6 αKG supplementation rescues low-glutamine induced stemness and suppresses Wnt signaling 217
Figure 4.7 The effect of αKG and glutamine supplementation on intestinal differentiation 219
Figure 4.8 αKG promotes hypomethylation of histone and DNA in CRC cells 220
Figure 4.9 αKG supplementation leads to DNA hypomethylation of genes related to differentiation and Wnt inhibition 222
Figure 4.10 DM-αKG treatment inhibits initiation and growth of PDOs 224
Figure 4.11 αKG supplementation drives terminal differentiation and suppresses growth of patient-derived colon tumor organoids 226
Figure 4.12 αKG supplementation inhibits the growth of highly mutated CRC tumors in vivo 228
Figure 4.13 αKG supplementation is an effective therapeutic intervention in a mouse model of intestinal cancer 230
Figure 4.14 The effect of DM-αKG treatment in mice 232

LIST OF TABLES

Page

CHAPTER 2

Table 2.1 Genes enriched in stem and non-stem 115

Table 2.2 Genes enriched in progenitors and absorptive and secretory lineages 116

Table 2.3 Genes enriched in differentiated cells Tuft and Ent 117

Table 2.4 Genes enriched in differentiated cell type EECs 118

CHAPTER 4

Table 4.1 Genes with DNA hypomethylation and upregulated gene expression upon DM-αKG treatment 234

Table 4.2 Identified intestinal differentiation-associated genes with decreased DNA methylation upon DM-aKG treatment are shown 235

Table 4.3 Sequences of primers used in the study 236

ACKNOWLEDGEMENTS

Although this doctorate degree will be awarded to me, it truly does not belong to me alone, but rather the dozens of people who have propelled me along in this journey. My success was only possible with their support and I am forever indebted for their love, friendship, inspiration, and guidance.

While I grew up dreaming of being a professional soccer player, **my parents, John and Pam Givens**, realized early on my aptitude for learning and understood my current school would not enable me to reach my full potential. For six years they helped me attend an out of district school, applying every year for a waiver at the first possible moment to ensure acceptance, and personally driving me to and from school for many years as I wasn't allowed to take the bus. Although many parents end up playing the role of a less than glorified chauffeur, I am grateful mine were willing to do so. Be it soccer practice and games, the library, or school, I *always* arrived early and *always* had a ride waiting to get home. Continuing on in my education, I was extremely fortunate for the financial support of my parents during my undergraduate degree. Even with many scholarships, college is expensive and I am forever appreciative of my parents for helping to carry that financial burden. My parents have always showed me unending love and support, always encouraging me to follow my dreams. They have always believed I could do anything I put my mind to, and with that kind of support system how could I not succeed? *Thank you Momio and Dadio, I love you with all my heart.*

My parents' support enabled me to follow my dreams, but I only had those dreams because of the inspiration of many science mentors. Through **Ms. Sue Black's** IB Biology class at Inglemoor HS, I became enamored with science and in particular biology. The reason there are at least 4 students (that I know of) from our 30-person class that are pursuing Biology PhDs is because of the infectious veneration Ms. Black has for biology. While an undergraduate student at Seattle Pacific University I had an enormous amount of Biology and Chemistry professors who contributed to molding me into the scientist I am today. I am particularly indebted to all of the professors who gave me the opportunity to TA for them, **Dr. Jenny Tenlen** for being immensely caring and supportive, and **Dr. Derek Wood** for challenging me academically in Microbiology, Genetics, and Molecular Biology and allowing me to find my passion in Molecular Biology research. The end of my freshman year at SPU I was lucky enough to join **Dr. Rick Ridgway's** lab. He was my research advisor for the next three years but he also quickly became my 'SPU Dad'. We could talk endlessly about research, science, faith, and life. My favorite memories are the research trips to Blakely Island and because of this I will always have a love of Marine Biology. I am thankful for everything he has taught me – both on a personal and professional level – and one of my greatest achievements is just knowing he is proud of me. **Dr. Traci Czyzyk**, while at Mayo Clinic, accepted me into her lab for a summer research internship. My time working in her lab reinforced my interest in graduate school and gave me research skills that propelled me to future success. Undergrad can be a challenge but I had the support of **Dr. Leena Adamian**, who was the best research partner, roommate, co-worker, classmate, study buddy, maid-of-honor, and best friend I could ask for. Although we do not see each other as much as I would like these days, I am forever appreciative of her continued friendship.

Before even officially starting at UCI, **Dr. Mike Mulligan** was a phenomenal mentor and played a crucial role in selecting my rotations. I grew as a scientist and immediately felt welcomed to UCI during my first rotation with **Dr. Diane O'Dowd**. I am also appreciative of my time in **Dr. Chris Hughes'** lab as a rotation student and our continued collaboration. The Waterman lab is well known for our far reaching collaborations, and I would be remiss if I didn't mention and thank all of the people and labs I have had the privilege of working with these past 5 years. **Dr. Peter Donovan** has been a longstanding collaborator and always contributed to lively discussion during our joint lab meetings while lending his expertise in developmental biology. I will forever have the desire to greet him with a jolly 'Hello Peter' in my best British accent. **Dr. Eric Stanbridge** and I got off to a bit of a bumpy start when we first met and he thought I was a postdoc during the first week of my rotation (And thus thought I should have a project outlined more than I did!). We were able to move past that and since I have been extremely thankful for his scientific rigor, challenging questions during lab meetings, and continued support.

Upon joining the Waterman lab, I was involved in an RNA collaboration with **Dr. Yongsheng Shi** and **Dr. Klemens Hertel**. They both welcomed me into their labs, their offices, and have continued to provide guidance in our collaboration while also writing me letters of recommendation and serving on my thesis committee. I would like to thank the other members of my committee – **Dr. Marian Waterman**, **Dr. Kai Kessenbrock**, and **Dr. Selma Masri**. In the past year, the Waterman lab also developed an exciting collaboration with the Masri lab I have enjoyed being a part of and am also excited to see how the project continues. I am appreciative for Selma's fierce support of me and my career, her willingness to write me letters of recommendation, and offer her advice. I also have her to thank for bringing me into the world of science twitter.

Additional collaborations I have been a part of include with **Dr. Mei Kong's lab** and her previous student **Dr. Thai Tran**. The published worked from this collaboration is presented in Chapter 4. Through an Opportunity Award Grant through UCI's Center for Complex Biological Systems (CCBS) I was able to work with **Dr. Ali Mortazavi** and student **Gaby Balderrama**. In the past year, the Waterman lab started a collaboration with **Dr. Steve Wiley** and his team at Pacific Northwest National Lab. Our first proteomic work was just recently published (Chapter 2), but I look forward to exciting future work that will come from this collaboration.

I have been fortunate to have a wonderful relationship with many pathology experts who have not only provided human tissue, but their mouse expertise, and spent time staring at slides under a microscope with me. **Dr. Robert Edwards** is a fantastic pathologist and I am not sure how he finds the time to do everything he does, but I am always thankful when he sneaks some time for me in. The knowledge he has when he looks at slides under a microscope is unprecedented. **Dr. Delia Tifrea** is equally as busy, but somehow finds the time to help with anything she can, be it ordering mice, mouse surgeries, scanning slides, or trying to track down tissue and deliver it to us. I am extremely thankful and impressed with her endearing personality and intense work ethic. **Dr. Shu-Yuan Liao** (at St. Joseph Hospital) has been crucial in procuring human tissue allowing us to build our colon biobank and grow organoids.

Much of my research would not have been possible without the assistance and work of UCI core facilities. The Flow Core and **Dr. Jenn Atwood** were crucial in developing my intestinal flow sort procedure and helping to perform countless sorts. I will always fondly remember when we came up with the Harry Potter theme for the 'Sorting Room' and spent down time picking out decorations. Sequencing has been crucial to my thesis work and I would like to acknowledge the GHTF and **Dr. Melanie Oaks**. I have benefited immensely from being able to drop into Melanie's office anytime and get her expertise on all things sequencing related. Developing skills in bioinformatics has been a challenge but I would like to thank **Dr.**

Jenny Wu and **Dr. Harry Mangalam** for their assistance. In addition to HPC help, Harry has also been our lab resident 'techie' fixing computers, broken rotators, and even helping with succulent planters for Oncoslayer fundraisers. Outside of lab he has been my greatest UniHills neighbor coming to my literal rescue when needed and also feeding me delicious food (his 'grumpy granola' and bread are to die for). He always puts on the best firepit shows and I will miss the Watermangalam backyard get togethers immensely and the days when I was able to borrow their adorable doggo Cooper.

I have been privileged with the financial support I have received during my graduate education. I was funded by the UCI CRI T32 training grant (led by **Dr. David Fruman** and **Dr. Aimee Edinger**) and also received the NSF GRFP. During this past year I have received support from Graduate Division through a Dean's Dissertation Fellowship. I have received numerous travel grants from AGS (Associated Graduate Students) and SOM (School of Medicine), in addition to a Professional Development Scholarship from SOM for an international research experience. I have also received funds from UCI's Anti-Cancer Challenge research grant to support publications. I was honored to be selected as an Orange County ARCS fellow and have greatly enjoyed being a part of the wonderful ARCS community. I am thankful for the many donors who support graduate research through SOM at UCI and it was a privilege to be a Stanley Behrens Fellow and to meet Stanley personally.

Being a part of SOM and the community of students, faculty, and staff has been a highlight of graduate school. I was lucky to be able to work with **Leora Fellus** and **Dr. Klemens Hertel** on the SOM Graduate Council and am extremely grateful for their hard work, leadership, and of course their friendship. They prioritize SOM students and show every day how much they care for our wellbeing and success. I will also never be able to let go of the ongoing 'student versus faculty' rivalry that started with Dicki and I. In my time at UCI the score is faculty-2; students-1 and I am hopeful this will be corrected in future years.

In addition to broadly thanking those I have worked with in SOM, I would also like to thank the rest of the **Microbiology and Molecular Genetics (MMG) Department** faculty, students, and staff. Although our lab is not in MedSci, I have always felt immensely welcomed and a part of our department community. My fellow MMG grad students have been encouraging, helpful, and kind. I am sad our time physically together has been cut short by Covid and look forward to the time we can gather in-person again. The MMG department only functions thanks to our administration staff who work behind the scenes, so thank you to **Janet Horwitz**, **Shanti Iyer**, **Lesley Dowd**, and **Kimberly Smith-Lyons**. Your quick email responses, dealing with all things related to orders/grants/reimbursements, making sure we get paid among other duties truly are appreciated. I would also like to thank the **Cancer System Biology (CaSB)** U54 group who have provided wonderful feedback on my research and strengthened my understanding of scRNA-seq. This group is only able to be maintained by the hard work of administrator **Karen Martin** who I am very grateful for.

Along with my department, I have also had a home in Sprague Hall and within the **Cancer Research Institute (CRI)**. I have enjoyed and benefited from being a part of the CRI, connecting with other labs and faculty, sharing my work, and hearing about others' work. **Nita Driscoll** and **Krystina Jarema** have been my go-to people whenever a problem arises and I am thankful that they have not only put up with me but solved every problem. I am certain our building would crumple and all research would cease without their hard work to keep everything going (and we would also go hungry without Nita). For several years I have had the honor to work with them on both the UCI Anti-Cancer Challenge Team, Oncoslayers, and the Youth Science Fellowship Program (YSFP), both of which would not be successful without their involvement. Another key player in the success of both of these programs is **Jessica Flesher**. Not only has she been my co-leader for more events and programs than I can count, but she has also been a fantastic collaborator and a true friend. Jessica has helped me troubleshoot experiments, struggle through HPC challenges, and also been my infamous 'mouse dealer'. She has been my conference roomie, poster presenting buddy, and the first person I would go to when I needed to talk to someone. I was always thankful for the open chair next to her desk and even more thankful for her friendship.

From the moment I joined the Waterman lab I felt at home. Part of it was the big windows with a beautiful view and the clean and organized lab, but the biggest part of it were the people. **Dr. Stephanie Sprowl-Tanio** was the first to welcome me into the lab and I learned more from her about grad school and UCI during my first year than from anyone else. I am also appreciative of the mentorship and collaboration with **Dr. George Chen** and **Dr. Yung Lyou**. It has been a privilege to work with many other Waterman Lab members including: **Dr. Beibei Wu** and **Linzi Hosohama**; UCI undergrad students: **Samantha Renusch**, **Sonia Park**, and **Ghadi Ghanem**; my high school students: **Kyle Krueger** and **Ashley Nguyen**, and several other rotation students and interns. However, the person that makes the Waterman lab truly what it is, is **Dr. Marian Waterman**.

Everyone who knows Marian cannot help but admire her if not for her scientific accolades, then for her leadership and dedication, or for her down-to-earth and endlessly supportive personality. Every bit of that admiration is well placed. I have mentioned more than once to others that Marian is "my favorite person ever", "I won the PI lottery", or that "I want to be Marian when I grow up". These sentiments have continued to hold true throughout my graduate work. Although Marian is my 'boss', I am certain few people adore their boss as much as I do and at the end of the day I know she is also my dear friend and one of my biggest supporters and advocates for my future success. She has taught me more than anyone how to be a scientist and how to ask worthy questions. Outside of bench science she has taught me more lessons than I can count on how to be a phenomenal PI, a good leader, and a better person. She has written me more letters of recommendation than I am sure most PIs ever write (and sadly is not done yet!) and always encouraged me to attend conference/events or be a part of volunteer activities or other programs. There is life outside of lab, and Marian not only understands that but has been a counselor when I needed one or other times just a shoulder to cry on. She has also been transparent when she has faced difficult situations, using them as invaluable teaching moments. I have shared my saddest and hardest, but also happiest and funniest moments of graduate school with Marian and I am so thankful she welcomed me into her lab and into her life. *Thank you Marian for building me into the scientist and person I am today and for every opportunity you have given me.*

The very last, but certainly not least, person I need to thank is my husband, **Tyler Habowski**. Most people do not meet the love of their life when they are 12 years old, but I am so thankful I did. I have been beyond blessed to have him by my side through high school, undergrad, graduate school, and now as we move into the next chapter of our lives. His unfailing love and support have helped me through every rough patch, and every success we have celebrated together. It was his courage that helped me enter the world of science over a decade ago and I truly would not have become the person I am without him. True love is finding the person that helps you grow into the best person you can be and someone you want to share every up and every down with along the way. I could not be luckier to share my life with Tyler – *I love you sweetie*.

CIRRICULUM VITAE
AMBER NICOLE HABOWSKI

EDUCATION

University of California Irvine (Irvine, CA, USA) Fall 2014-Fall 2020 (expected)
Doctor of Philosophy, Biomedical Sciences – Molecular Biology, 3.99/4.0 GPA

Seattle Pacific University (Seattle, WA, USA) Fall 2010-Spring 2014
Bachelor of Science, Cellular and Molecular Biology
Bachelor of Science, Biochemistry, 3.92/4.0 GPA, *Summa Cum Laude*; *Alpha Kappa Sigma*

RESEARCH EXPERIENCE

Graduate Student Researcher (Advanced to Candidacy July 2017) 2015-present
University of California, Irvine. Advisor: Marian Waterman.
Department of Microbiology and Molecular Genetics, School of Medicine

- Focus on high through-put data and bioinformatics approaches to allow a genome wide understanding of transcriptomics and mRNA processing in colon cancer and normal crypt homeostasis.
- Use of human cancer cell lines, human colon organoids, and human and mouse primary tissue as model systems to understand the role of Wnt signaling in stem cell biology.

Visiting Scientist Jan 2017
Hubrecht Institute, Utrecht University, Netherlands. Advisor: Hans Clevers.
Training to grow normal and tumor human colon organoids.

Graduate Rotation Student Spring 2015
University of California, Irvine. Advisor: Chris Hughes.
Mechanistic role of transcription factor Slug in angiogenic sprouting involving Wnt and Notch signaling.

Graduate Rotation Student Winter 2015
University of California, Irvine. Advisor: Marian Waterman.
Alternate polyadenylation of long non-coding RNA *MALAT1* and its role in cancer.

Graduate Rotation Student Fall 2014
University of California, Irvine. Advisor: Diane O'Dowd.
Generation of CRISPR/Cas9 edited induced pluripotent stem cells as a neuronal model for Genetic Epilepsy.

Undergraduate Researcher 2013-2014
Seattle Pacific University, Seattle, WA. Advisor: Derek Wood.
Molecular investigation of functional genomics of proline biosynthesis in *Agrobacterium rhizogenes* A4.

Summer Undergraduate Research Fellow Summer 2013
Mayo Clinic, Scottsdale, AZ. Advisor: Traci Czyzyk.
Uncovering novel metabolic protein (MPZL3) mechanism relating to lipid production and cholesterol regulation *in vivo* and *in vitro.*

Undergraduate Researcher 2011-2014
Seattle Pacific University, Seattle, WA. Advisor: Richard Ridgway.
Immunofluorescence Localization of Dopamine in the Marine Macroalga, *Ulvaria obscura.*

PUBLICATIONS

Habowski, A.N., J.L. Flesher, J.M. Bates, C. Tsai, R. Zhao, K. Martin, A.K. Ganesan, R.A. Edwards, T. Shi, H.S. Wiley, Y. Shi, K.J. Hertel, and M.L. Waterman. (2020) Transcriptomic and proteomic signatures of stemness and differentiation in the colon crypt. *Communications Biology*, 3:453. DOI: 10.1038/s42003-020-01181-z.

Habowski, A.N, J.M. Bates, J.L. Flesher, R.A. Edwards, and M.L. Waterman. (2020) Isolation of murine large intestinal crypt cell populations with flow sorting. *Nature Protocol Exchange*. DOI: 10.21203/rs.3.pex-994/v1.

Habowski, A.N., T.J. Habowski, and M.L. Waterman. GECO: Gene expression clustering optimization app for non-linear data visualization of patterns. *In Revision.*

Habowski, A.N., A.A Almet, S. Wang, Q. Nie, and M.L. Waterman. Regionally distinct responses to irradiation wounding of the intestine. *In Preparation.*

Tran, T.Q., E.A. Hanse*, **A.N. Habowski***, H. Li, M.I. Gabra, Y. Yang, X.H. Lowman, A.M. Ooi, S.Y. Liao, R.A. Edwards, M.L. Waterman, and M. Kong. (2020) α-Ketoglutarate attenuates Wnt signaling and drives differentiation in colorectal cancer. *Nature Cancer*, 1:345-358. DOI: 10.1038/s43018-020-0035-5.

> Highlighted in News & Views: Dana Krauß & Eyal Gottlieb. (2020) Restraining colorectal cancer with αKG. *Nature Cancer*, 1:267-269. DOI: 10.1038/s43018-020-0044-4.

Lyou, Y., **A.N. Habowski**, G. Chen, M.L. Waterman. (2017) Inhibition of nuclear Wnt signalling: challenges of an elusive target for cancer therapy. *British Journal of Pharmacology*, 174(24):4589. DOI: 10.1111/bph.13963.

Sprowl-Tanio, S., **A.N. Habowski**, K.T. Pate, M.M. McQuade, K. Wang, R.A. Edwards, F. Grun, Y. Lyou, and M.L. Waterman. (2016) Lactate/Pyruvate transporter MCT-1 is a direct Wnt target that confers sensitivity to 3-bromopyruvate in colon cancer. *Cancer & Metabolism*, 4:20. DOI: 10.1186/s40170-016-0159-3.

*denotes equal contributions

EXTERNAL GRANTS/FUNDING

Total Award: $204,000 and 5 years of funding

National Science Foundation GRFP Fellowship ($138,000)	Aug 2016-2019
Trainee - Cancer Biology & Therapeutics (CBT) Training Grant (NIH T32) ($66,000)	2015-2016

AWARDS AND RECOGNITION

Total Award: $161,700 including $27,000 for direct research expenses

UCI MMG Harris Moyed Dissertation Award *(*$1,500*)*	2020
UCI Systems Biology Opportunity Award Research Grant ($10,000)	2019-2020
UCI Graduate Dean's Dissertation Fellowship Award ($12,000)	2019-2020

UCI Campus Nominee for NIH-NCI F99/K00 Transitional Fellowship Award	2018
Beckman Coulter ARCS Scholars (Orange County Chapter) ($10,000)	2017-2019
UCI School of Medicine Stanley Behrens Fellow ($20,000)	2017
UCI School of Medicine Professional Development Supplement ($2,000)	Nov 2016
UCI Ayala Fellowship – Top Graduate Recruit ($15,000)	2014-2015
SPU Trustees' Scholar Award for academic performance ($56,000)	2010-2014
FIRST Robotics Scholarship for leadership and excellence ($20,000)	2010-2014
Matthew Kelley Scholarship for involvement in medical mission work ($13,000)	2012-2014
Deitzmann Scholarship for academic excellence as a SPU Chemistry major ($3,000)	2013-2014
Causey Scholarship for academic performance as a SPU Biology major ($700)	2013-2014
SPU Cellular/Molecular Biology Student of the year	2014
Awarded standing on the Dean's List all quarters while in attendance at SPU	2010-2014
Who's Who Among Students in American Universities	2013
Paul Lepse Award for Excellence in Organic Chemistry (SPU)	2011-2012

ORAL SCIENTIFIC PRESENTATIONS

- *Loss of stemness and lineage commitment in the intestinal crypt* June 2020
  AACR Virtual Annual Meeting II
- *Loss of Stemness in Colon Crypts and Cancer* Feb 2019
  UCI MMG Departmental Seminar, Irvine, CA.
- *mRNA Splicing in the Colon.* March 2018
  UCI RNA Club Seminar and MMG Departmental Seminar, Irvine, CA.
- *mRNA splicing in the initiation and progression of Wnt-dependent colorectal cancer.* April 2017
  Cancer Biology and Therapeutics Retreat, Pasadena, CA.
- *mRNA splicing in the development and progression of Wnt-dependent colorectal cancer*. Nov 2016
  Research in Progress Talk, UCI Cancer Biology and Therapeutics Core, Irvine, CA.
- *The Role of MPZL3 in Hepatic Lipid Uptake and Synthesis.* Aug 2013
  Summer Undergraduate Research Fellowship Symposium, Mayo Clinic, Scottsdale, AZ.

SHORT TALKS/PUBLIC SPEAKING

- *Loss of Stemness in Colon Cancer* March 2019
  Annual ARCS Scholar Award Program, Irvine, CA.
- *The Why, Who, and How of Cancer Research.* Dec 2018, 2019
  IUSD 21st Century Career Conference, Irvine, CA.
- *The RNA Puzzle in Cancer*. **(Second Place Award)** May 2017
  NSF GRFP UCI Research Symposium Science Blitz, Irvine, CA.
- *Did you get your Colonoscopy?* Nov 2016
  Brews and Brains Guest Speaker, Costa Mesa, CA.
- *The RNA Puzzle.* **(Finalist Award)** Nov 2016
  UCI Graduate Student & Postdoctoral Elevator Pitch Competition, Irvine, CA.

POSTER PRESENTATIONS

- *GECO: Gene expression clustering optimization app for visualization of patterns* Sept 2020
  2020 CSBC NIH Annual Investigators Meeting (Virtual)
- *Transcriptomic and Proteomic Signatures of Stemness and Differentiation in the Colon* Aug 2020
  NCI Cancer Systems Biology Consortium – Junior Investigator Meeting (Virtual)
- *Loss of stemness and lineage commitment in the intestinal crypt*
  UCI “Grad Day” Symposium, Irvine, CA. Oct 2019
  Gordon Research Conference: Wnt Signaling, West Dover, VT. **(Best Poster Award)** Aug 2019
- *A global analysis of colon crypt transcriptome defined by deep sequencing of sorted cells*
  UCI Campus-Wide Symposium on Basic Cancer Research, Irvine, CA May 2019
  UCI Systems Biology (CCBS/MCSB/CaSB/CMCF) Annual Retreat, Universal City, CA March 2019
- *Wnt regulation of mRNA Splicing in Colon Crypts and Colon Cancer.*
  AACR Intestinal Stem Cells and Colon Cancer: Biology to Therapy, Washington D.C. Sept 2018
- *Dynamic Changes along the Colon Crypt: Understanding the Complex Regulation of Stem Cells and Differentiation.*
  UCI “Grad Day” Symposium, Irvine, CA. Oct 2018
  UCI RNA Symposium, Irvine, CA. April 2018
- *mRNA Splicing in the Initiation and Progression of Wnt-dependent Colorectal Cancer.*
  UCI “Grad Day” Symposium, Irvine, CA. Sept 2017
  Gordon Research Conference: Wnt Signaling, Stowe, VT. Aug 2017
  NSF GRFP UCI Research Symposium, Irvine, CA May 2017
  UCI Campus-Wide Symposium on Basic Cancer Research, Irvine, CA **(Best Poster Award)** May 2017
  Keystone Conference: mRNA Processing and Human Disease, Taos, NM. March 2017
- *The Role of MPZL3 in Hepatic Lipid Uptake and Synthesis.*
  Erickson Conference, Seattle Pacific University, Seattle, WA. May 2014
  Murdock Conference, Lewis and Clark College, Portland, OR. Nov 2013
- *Immunofluorescence Localization of Dopamine in the Marine Macroalga, Ulvaria obscura.*
  Erickson Conference, Seattle Pacific University, Seattle, WA. **(Best Poster Award)** May 2013
  Murdock Conference, Whitman College, Walla Walla, WA. Oct 2012
  Erickson Conference, Seattle Pacific University, Seattle, WA. **(Best Poster Award)** May 2012

TEACHING EXPERIENCE

Graduate Student TA. University of California, Irvine.

Systems Biology Bootcamp Fall 2018

- Week-long lab intensive bootcamp (25+ hours) for incoming PhD students to teach them a range of molecular, cellular, and developmental biology techniques and model systems.

Advanced Molecular Genetics Lab (M116L), Instructor of record: Pavan Kadandale. Spring 2017

- Instructor of 4-hour lab section once a week (20 students).
- Graded reading assignments (40), lab reports (80), and final exams (100+).
- Guest lectured during class seminar (100+ students).

Organic Chemistry Grader. Seattle Pacific University. 2012-2013
Graded quizzes and quiz corrections for organic chemistry. Instructor of record: Kevin Bartlett.

Biology Laboratory TA. Seattle Pacific University. 2011-2014

- Created and delivered lectures during lab, graded pre/post lab reports and lab notebooks, wrote and graded quizzes, and helped students troubleshoot experiments.
- Lower division Courses:
    - General Biology (Cell/Molecular) 2101 Fall 2011, Fall 2012
    - General Biology (Plants) 2103 Winter 2012, Spring 2012
    - Marine Biology Summer 2011, Summer 2012
    - Biology for Education Majors Winter 2013
- Upper Division Courses:
    - Freshwater Biology Spring 2013
    - Cell Biology (including Cell Culture Room Manager) Spring 2013, Spring 2014
    - Genetics Winter 2014

LEADERSHIP AND INVOLVEMENT

Co-Captain of UCI "Oncoslayers" Fundraising Team 2017-2019

- "Oncoslayers" are the official UCI Anti-Cancer Challenge team of the Cancer Research Institute and School of Medicine Graduate Studies.
- Lead and coordinate all fundraising efforts including bake sales, plant sale, t-shirt design and sales, raffles, auctions, and "Art of Research" competition.
- Volunteer and participate at the UCI Anti-Cancer Challenge Walk, Run, and Bike Event each year.

Program Coordinator for UCI CRI Youth Science Fellowship Program 2016-2019
Program Coordinator in Training for UCI CRI Youth Science Fellowship Program 2016

- Hosted by UCI's Cancer Research Institute, this program provides selected Orange County high school students a free 6-week summer internship with hands on lab training.
- Coordinate reviewing of applications and the selection of about 20 students out of the 200 that apply.
- Recruit faculty/graduate students who are willing to be mentors and assign students to labs.
- Host and coordinator of the high school students' orientation, weekly seminars, and final research symposium event.
- Advisor to program participants on students/mentor issues or lab problems.
- Mentored and oversaw all work by Waterman Lab high school students.

Appointed Member of School of Medicine Graduate Student Advisory Council 2016-2020

- Quarterly meetings to discuss the SOM graduate programs and plan future events.
- Coordinator of Events: (1) "Careers in Academia: Insights into the Academic Hiring Process at an R1 Institution", Feb 2017. (2) "Grad Day", Sept 2017. (3) "Bone Marrow Registration Drive (with DKMS)", Nov 2017. (4) The Perfect Specific Aims Page: Practical Grant Writing Advice for Biomedical Specific Aims, Day #1 – Faculty Panel, Day #2 – Writing Workshop, March 2018. (5)

“Grad Day”, Oct 2018. (6) Faculty and Student Soccer Match & BBQ, April 2019. (7) “Grad Day”, Oct 2019.

- Event volunteer including new student orientation, recruitment, and graduate social events.
- Lab host for current and prospective donors allowing them to hear about and see the types of research that occurs in the school of medicine and meet current graduate students.

Member on MMG Department Student Seminar Committee Coordinator 2016-2018

- Worked with department labs to select speakers to invite to department seminars. In charge of inviting the speakers and further correspondence to coordinate the visits.

Mentor for UCI’s Women and Minorities in STEM Mentorship Program 2016-2017

- Quarterly meetings with two UCI undergraduate students to mentor them on their career paths in STEM fields. Available when issues arise or they are in need of guidance.

AAAS Fellow for ELISS (Emerging Leaders in Science and Society) 2015-2017

- ELISS is a leadership development program that prepares graduate students to collaborate across boundaries for the benefit of society, focused on bridging between science, policy, and society.
- Competitively selected as one of three fellows at UCI, and 18 fellows across the nation (at 5 universities). 2016 cohort’s topic was: How to ensure safe, sustainable, and affordable drinking water in the United States.
- Met with national stakeholders and government officials on Capitol Hill in Washington DC and recruited national level advisors. Conducted over 50 informational interviews with water experts, recruited and worked with a team of 16 local advisors across academic, government, and industry sectors, collaborated with fellows across the country.
- Coordinated Events: (1) “Drinking Water in the US: Improving Well Water Safety” (National Forum), Washington DC, Dec 2016. (2) “Seeking Resilience: A Discussion on the Future of Southern California’s Water System” (Local Forum), Irvine, CA, Oct 2016. (3) Launch of UCI ELISS – “Winter Social”, UCI, Jan 2016.

Preliminary Exam Mentor to first year CMB Students 2015-2016

- Students in the Cellular and Molecular Biology (CMB) umbrella program at UCI must pass a preliminary exam at the end of their first year in order to join a lab and become a second-year student. This exam is a research presentation followed by a chalk talk and specific-aim-driven research proposal.
- Selected as a mentor based on excellent performance during previous year preliminary exam.
- Mentored and helped train 5 first year students in preparation for their exams. Worked with them through their practice talks, provided feedback, and gave example talks.

Member of UCI ‘STEM-ED’ Group 2014-2016

- Participated in monthly STEM Education (‘STEM-ED’) group meetings led by Dr. Pavan Kadandale and Dr. Brian Sato focused on primary research on education.
- Discussed effective ways of teaching, new education techniques, how to handle tough situations, how to get teaching experience, and pursuing a teaching career.

President of SPU Biology Club 2012-2014

- Hosted key events including the Bone Marrow Registration Drive, Blood Drives, Bowling with Professors (co-hosted with Chemistry Club), Science presentation/talks, video nights (planet earth and other documentaries). Personally acted as the liaison between Biology majors and department and also in charge of recruiting volunteers for events hosted by the department (research symposia).

Vice President of Meditrack 2012-2014

- Meditrack is a volunteer-based chemistry tutoring club at SPU.
- Coordinated weekly 2-hour tutor sessions, and personally tutored the Organic Chemistry sessions for 15-30 students. Explained practice problems and reinforced ideas they learned in class.

SELECT VOLUNTEER WORK

FIRST Robotics Volunteer (For Inspiration and Recognition of Science and Tech) (100 hours) 2016-present

- FRC Judge. Orange County Regional, Irvine, CA, March 2017-2019.
- FRC Defense Coordinator. Orange County Regional, Placentia, CA, April 2016.

Science Fair Judge for 6th-12th grade District and Regional Science Fairs (60 hours) 2014-present

- Judge Captain Orange County Science and Technology Fair (OCSEF), Anaheim, CA, March 2019.
- Judge at Astounding Inventions Fair, Irvine, CA, Jan 2018-2019.
- Judge at Orange County Science and Technology Fair (OCSEF), Costa Mesa, CA, March 2017-2018.
- Judge at Irvine Unified School District (IUSD) Science Fair, Irvine, CA, Feb 2015-2018
- Judge at Intel International Science and Engineering (ISEF) Fair Qualifying Competition, Irvine, CA, March 2015.
- Scientist at "Ask-a-Scientist Night" (IUSD), Irvine, CA, Oct 2014-2016, 2019.

Biomedical Research Representative and Counselor at UCI CampMed (100 hours) 2014-2015

- CampMed is an undergraduate led program that provides an opportunity to underserved local high school students to attend a free camp that inspires them to pursue medicine and sciences and higher education in general.
- Involved in activities prior to the camp such as bi-weekly mentorship at the high schools to help the students with homework, college applications, financial aid, and general life guidance.
- Attended weekly meetings with all counselors to help coordinate events. Fundraised throughout the year to raise enough money to provide this event free to the campers.
- Co-counselor to 12 high school students during the camp and also the biomedical research representative. Gave talks on what pursuing a graduate degree is like and the career options available in biomedical research.

STUDENTS MENTORED

- Sonia Park, University of California Irvine 2019-present
- Ghadi Ghanem, University of California Irvine, now UCLA medical school 2017-2019
- Samantha Renusch, University of California Irvine, now UCSD 2017-2019
- Ashley Nguyen, Villa Park High School, now at UCLA Summer 2018
- Kyle Krueger, Tesoro High School, now at UC Davis Summer 2017

ABSTRACT OF DISSERTATION

Loss of stemness in the normal crypt, intestinal wounding, and colon cancer

By

Amber Nicole Habowski

Doctor of Philosophy in Biomedical Sciences

University of California, Irvine, 2020

Professor Marian L. Waterman, Chair

The intestine performs essential functions to absorb water and nutrients and serve as an important barrier preventing pathogens from harming the body. The innermost layer of the intestinal mucosa is a single layer of epithelial cells that form invaginating crypt structures. Adult epithelial stem cells reside at the base of each of these structures. The stem cells are rapidly proliferating and give rise to daughter cells which differentiate into either the absorptive or secretory lineage. This stem cell activity is continuous and ensures that the intestine is an actively renewing tissue that can perform its normal homeostatic functions and can withstand injury by repairing itself rapidly. The precise mechanism(s) behind stem cell differentiation, wounding, and homeostasis are not very well understood. We used high throughput RNA sequencing to follow homeostatic and wounded processes. We revealed that during normal colonic stem cell differentiation, loss of stemness is accompanied by a burst of mRNA processing but minimal changes in mRNA expression level. Upon colonic crypt wounding, the rapidly proliferating stem cells readily undergo apoptosis. We determined that the remaining, surviving cells respond by quickly fortifying an epithelial barrier, secreting protective proteins, and ramping up proliferation just four days following wounding. Differentiated cells also display a high level of plasticity, and eventually new stem cells emerge via de-differentiation. Proper wound healing enables the intestine to return to homeostasis, but in some instances of genetic insult as in cancer initiation, there is no return to homeostasis. Cancer can also be

referred to as a wound that does not heal, and chronic wounding in the intestinal mucosa can even led to colorectal cancer. We demonstrated that shifts in glutamine metabolism can facilitate stemness and differentiation of colonic cancer cells and is relevant to cancer progression. Overall, this dissertation investigates stemness as it relates to the normal colon, wounding, and colon cancer and highlights the dynamic nature of the intestine transcriptome and proteome.

CHAPTER 1

Introduction

Unpublished work:
***Habowski, A.N.**,2020.*

INTRODUCTION TO THE INTESTINE

Intestine function

Once food is digested in the stomach it passes through the small intestine and then the large intestine (colon) where nutrient and water absorption take place. Intestinal absorption is a life-essential function that supplies important molecular building blocks and energy to the body. The inner most layer of the intestine is also immensely important since it acts as a barrier between the germ-riddled material that enters the body and infection-susceptible tissue inside the body. In addition to battling potentially harmful germs, the intestine is also impacted by environmental factors including being bombarded with unhealthy diets and even harmful chemicals or carcinogens that we consume [1]. At the same time, the cells in the intestine lining communicate with and enable growth of helpful, and necessary, flora that create the gut microbiome [2,3]. To maintain its essential function, the intestine has a phenomenal structure and out of necessity, is immensely resilient.

Intestine structure

The intestine is composed of three general layers: (1) muscle on the outside, (2) a submucosal layer made of stroma and other support cells, and (3) a layer of mucosa which forms the innermost barrier. The inner mucosal layer is comprised of an array of different epithelial cell types performing important functions and organized together in a structure called the crypt. Crypts are present in the small intestine and the colon, but the small intestine contains shorter crypts with very long extended villi, whereas no villi are observed in the colon. Given the average colon size, and an estimated 100 colon crypts per mm^2 tissue, the average human intestine has approximately 10 million colon crypts [4]. At the base of each crypt is a stem cell niche with a small persistent population of dividing stem cells that give rise to both stem cells and committed progenitor daughter cells [5] (Figure 1.1). These daughter cells terminally differentiate and migrate up the colon crypt before dying and sloughing off, being replaced by the next wave of cells such that every 5 days the mucosal lining is completely replaced – a process that makes the intestine one of the most proliferative, active tissues of the body [5,6]. The terminally differentiated cells found in the colon are enterocytes (absorptive cells), goblet cells (mucus secreting), enteroendocrine cells (peptide secreting), tuft cells (sensory and secretory), and deep crypt secretory cells (peptide secreting). All cell types are important for maintaining normal intestinal homeostasis.

SIGNALING AND THE TRANSCRIPTOME

Signaling pathways that regulate the intestine

The loss of stemness, the process of cell migration, and the differentiation of these cells into distinct lineages is closely regulated by important signal transduction pathways. The stem cell niche at the base of the crypt is nurtured by a Wnt signaling gradient that promotes proliferation and maintenance of stem cells [7] (Figure 1.1). Wnt ligands are secreted by supporting Paneth cells (small intestine only) and stroma-fibroblast cells, and they are closely juxtaposed to the stem cells since Wnt ligands can defuse distances of only a few cells [8]. Stem cells are triggered by the secreted Wnt ligands to activate canonical, beta-catenin dependent signaling. [9]. This results in activation of Wnt target genes and gene programs, and in the case of stem cells at the base of crypts, promotes self-renewal [10]. Wnt signaling does not act alone, but rather in concert with many other signaling pathways, such as the Notch signal transduction pathway.

Notch signaling is elevated at the base of the crypt and in transit amplifying daughter cells and it plays an important role in lineage specification via a process called lateral inhibition [8] (Figure 1.1b). The secretory lineage produces a large amount of Notch ligands which then activate Notch signaling in stem cells and uncommitted transit amplifying cells (which have Notch receptors) [11]. Activated Notch signaling pushes for an absorptive lineage since Notch target genes repress expression of a gene program that would otherwise promote a secretory cell fate [8]. This lateral inhibition allows for a homeostatic and finely regulated equilibrium between the number of secretory cells compared to absorptive cells. Other signaling pathways, such as BMP, promote differentiation and are elevated at the top of the crypt [8]. These signaling pathways, along with many others (Myc, Hippo, Hedgehog, Ephrin, Egf, and Kit) help to establish and maintain stemness and promote appropriate differentiation [11]. While we know a great deal about these signals and how they direct the rapid differentiation of colon epithelia, we still lack a full understanding of the earliest steps in this process – that is – how the newly produced daughter cells leave the stem cell niche, lose stemness and quickly commit to a lineage.

Regulation of the transcriptome

Activation of signal transduction pathways kicks off a cascade of events inside the cell which often result in alteration of gene expression and mRNA production. Changing gene expression by either activating or repressing genes alters the transcriptome of the cell and this can then lead to changes in the proteome which ultimately changes cell function or responds to environmental challenges. However, in addition to mRNA level changes, the transcriptome can be modified by altering the transcripts through mRNA processing [12–14]. As RNA transcripts are actively transcribed, a whole host of machinery is processing the RNA. This processing includes adding a protective 5' cap and a protective 3' polyadenylated tail, and splicing exons together via the removal of introns (Figure 1.2a). Each RNA transcript is not always processed in the same manner and these variations are known as alternative splicing and alternative polyadenylation (1.2b). Alternative mRNA processing can have a profound impact on the transcriptome as well as the proteome, and in fact 95% of protein-coding mRNAs are known to be alternatively processed [13–15].

Epigenetics is another factor that contributes to the regulation of gene expression. Epigenetics encompasses modifications (i.e. "marks") of genetic material that does not actually change the coding information itself. Epigenetic changes can influence how accessible a gene is for transcription and mRNA production [16]. This includes DNA methylation or histone modifications and can have immediate and/or long-lasting effects [16]. Many regulatory enzymes facilitate adding or removing epigenetics marks and their activity is influenced by the cellular environment [17]. The presence or absence of co-factors, such as specific metabolites, can lead to epigenetic changes as well [17]. Signal transduction pathways, mRNA processing, and epigenetics are important factors that contribute to how the transcriptome is maintained in normal intestinal homeostasis, how it changes during wounding, as well as the aberrant activity that occurs in cancer.

INTESTIONAL HOMEOSTASIS

Maintaining normal intestinal homeostasis

Multipotent intestinal stem cells that could transform into differentiated cell types were first reported by Cheng and Leblond in 1974 [18]. Further investigation of these cells confirmed that they are continuously cycling, that they give rise to all intestinal epithelial cell types, and that they persist for a lifetime – the basis for our understanding that these are adult stem cells [8]. Interestingly, stem cells at the base of the crypt undergo daily symmetric division and compete to remain the cycling stem cell [8,19]. Cells that end up leaving the stem cell niche at the base rapidly differentiate, and thus the position of the stem cell is indicative of whether it will remain a stem cell or not. This competition is described in the neutral drift model, which over time results in clonal crypts as one stem cell clonally tends to dominate [8,19]. This process is tightly regulated by neighboring stroma cells and other epithelial cells. The previously discussed signaling pathways play an important regulatory role and their disruption can be detrimental. For example, loss of Wnt signaling can lead to developmental defects and loss of stem cells in the intestine [20]. Healthy stem cells are crucial for crypt homeostasis and global intestine function.

A majority of our understanding of the intestine is based on research of the small intestine, and although it is similar to the colon, there are important differences. Most notable is the very high rate of colon cancer compared to only rare cases of small intestinal cancer in humans. Cancer has often been described as a ‘wound that never heals’ and thus it is logical to compare normal colon with wounded and cancer. Understanding how normal colon homeostasis is maintained at steady state and how recovery proceeds following wounding can inform our understanding of cancer progression. Aberrant activity of signaling pathways that are important for maintaining normal colon homeostasis can lead to cancerous growth. Importantly, we can better define colonic stemness and how it is appropriately maintained in the normal setting, and how this is altered in wounding and further skewed in colon tumors.

Disruption to homeostasis: Intestinal wounding

The important roles of the intestine, nutrient/water absorptive and barrier function, necessitates its resilience to wounding. When acute wounding occurs, the colon responds rapidly and quickly to promote recovery. When differentiated cells near the top of the crypt are damaged and lost, the stem cells at the base continue to function, giving rise to replacement differentiated cells and homeostasis eventually returns [21]. Paradoxically, targeted ablation of stem cells and wounding at the base of the crypt does not have long term effects and homeostasis, along with normal proliferative stem cells, are restored [21]. The initial hypothesis to explain rapid recovery from wounding when stem cells are lost is the presence of a quiescent, or ‘+4 cell’, able to act as a reserve stem cell [8,21,22]. Subsequent studies have not supported this hypothesis, but rather suggested that mature cells are able to de-differentiate and regain stemness [8,21,23–27]. This ability is shared by cells of both the secretory and absorptive lineage and suggests that cells in the intestinal crypt are incredibly ‘plastic’ and differentiated cells are able to retain some level of stemness [21,25,27].

Inflammatory bowel disorders (IBD; Crohn’s disease and ulcerative colitis) can result in chronic intestinal wounding. More than 3 million adults in the US have been diagnosed with IBD, and while about half are in remission, many face regular flair-ups with mild/moderate disease, and some are subject to severe, debilitating symptoms. The causes of IBD are still being investigated but they are attributed to many factors such as: genetic predisposition, environmental factors, the microbiome, and altered immune responses [21,28]. Severe inflammation can damage tissue leading to a loss of epithelial cells as well as degradation of other layers of the intestine [29]. A long term complication of chronic IBD is colon cancer, which further highlights the idea that cancer is a wound that never heals [28].

Loss of homeostasis: Colon cancer

In the US, colorectal cancer is the second leading cause of cancer-associated deaths following lung cancer. One in every 20 Americans will be diagnosed with colon cancer during their lifetime and more than 53,000 individuals are predicted to die from colon cancer in 2020. Although survival rates for those diagnosed in earlier stages are decent (Stage I at 92%, Stage II at 75%, Stage III at 70%), the outcome is very grim for those diagnosed in the late metastatic stage (Stage IV at 11%). Compliance with recommended colonoscopy screening and new, less invasive measures such as the FIT test, have improved early detection and increased survival rates. However, colon cancer trends now show occurrence in younger adults - a population that is not routinely screened (25-45 yrs). There is a clear need for continued research to better understand the molecular basis of colon cancer. A better understanding can lead to even further improved early detection and therapeutics, and treatment methods for patients with Stage IV disease.

Colon cancer arises when there is improperly controlled growth of intestinal cells. Mutations leading to constitutively active Wnt signaling (such as *APC*), result in hyper-proliferation of intestinal cells – often leading to cancer [20]. More than 80% of sporadic colon tumors contain an inactivating mutation of *APC*, suggesting that chronically overactive Wnt signaling is oncogenic [6]. Inherited mutations in *APC* results in syndromes such as Familial Adenomatous Polyposis (FAP), a condition in which hundreds of polyps appear in the colon of affected family members by ten years of age [6]. Without colectomy these patients have a nearly 100% chance of developing colon cancer [30]. Given that constitutively active Wnt signaling can drive colon cancer, it has been important to understand the mechanism of how this occurs. Wnt signaling leads to the activation and transcription of downstream target genes, some of which have been shown to play roles in cell proliferation, survival, metabolism, and migration (metastasis) [31,32]. However, inhibiting Wnt signaling as a cancer therapeutic has not only been very challenging, but might lead to adverse effects including more invasive tumors [33,34].

The role of stem cells in cancer has been long debated and has led to two opposing views of how tumors arise. The Stochastic Model argues there is not a specific identifiable tumor-initiating cell, but rather all cells in a tumor could have this ability and stochastic probability determines the initiator [35]. On the other hand, the Cancer Stem Cell Theory (developed from Hierarchy Theory), suggests there is a small population of cells that are tumor initiators [35]. Although the Hierarchy Theory, i.e. a cancer that arose from embryonic-like cells - was first proposed by Rudolph Virshow in 1855, it was not until the 1990s that the first experimental evidence for a stem-like cell in acute myeloid leukemia was reported [35,36]. Later a complete cellular hierarchy of cancer cells was identified, with a clear cancer initiator cell with stem-like properties[37]. Work in solid tumors has identified populations of cancer cells that have more tumor-initiating potential than other populations, but this has led to conflicting ideas of what actually defines a cancer stem cell [35,38]. Despite an evolving definition of cancer stem cells and tumor initiating cells, the underlying complexity is the heterogeneity of tumors and implications for development of new cancer therapies. Subtypes of tumor cells, such as cancer stem cells, are thought to be more resistant to therapy and thus an insufficient therapeutic treatment could result in the more aggressive/invasive cells remaining and poor patient outcome [38]. Recent work to classify colon cancer has identified a mesenchymal subtype, which has the worst relapse-free and overall survival compared to other subtypes [39]. This subtype (CMS4) has the most enriched cancer stem cell signature, the lowest Wnt signature, the least immune infiltration, and phenotypically, is the most invasive [39]. More work is needed to fully understand cancer stem cells (tumor initiators), how they are different and similar to other stem cells (embryonic and normal adult) and how to appropriately target their stemness therapeutically.

INVESTIGTATING INTESTINAL STEMNESS

In this dissertation intestinal stemness is investigated in three settings: 1) normal colonic crypts, 2) intestinal wounding , and 3) colorectal cancer. Although data from the small intestine will be presented, the focus will be on that of the colon since this is the most prominent location of human intestinal cancer. The overarching question is that of how stemness is regulated in the colon, including how it is lost and gained. In the normal colon (Chapter 2), we redefine the transcriptome and proteome of stem cells and raise questions of what facilitates a rapid loss of stemness and differentiation in homeostatic conditions. During intestinal wounding (Chapter 3), there is an immediate loss of stem cells, and we examine how other epithelial cells respond. We investigate the long-standing question of how crypt homeostasis returns and how stemness is regained following wounding. In Chapter 4, we explore the mechanism of the metabolite α-ketoglutarate that can alter stemness in colon cancer cells. The question of how a more differentiated versus a more stem-like colon cancer influences tumor growth and survival has important therapeutic implications. This dissertation and the data herein investigate these exciting research questions and greatly advances our understanding of the colon.

MOLECULAR BIOLOGY EXPERIMENTAL APPROACHES

Much of the advances in our understanding of the molecular world is facilitated by innovations in experimental model systems and technology. Three important tools highlighted throughout this dissertation are flow cytometry, organoids, and sequencing (+bioinformatics).

Flow cytometry

Flow cytometry is a technique used to measure characteristics, such as the shape/size/presence of biomarkers of individual cells in a suspension. Biomarkers (including cell surface proteins) are detected using cell permeable dyes or by incubating cells with antibodies that recognize an antigen of interest. These antibodies have a fluorophore attached that emits a reliable and detectable light signal of a specific wavelength when excited by a laser that excites the fluorophore. Enormous advances in flow cytometer technology since their development in the 1950s enables thousands of cells to be analyzed each second [40]. Multiple fluorophores can be detected at a time, allowing flow cytometers with multiple lasers to identify numerous markers on cell surfaces at one time in addition to cell shape/size.

Not long after the first flow cytometers were built did other scientists expand on the technology to create flow sorters (FACS = Fluorescence Activated Cell Sorting) [40]. This added an additional step to the flow cytometer that was the ability to not just measure and analyze cell characteristics, but to collect cells based on those characteristics. Following analysis (excitation with a laser and detection) the single cell suspension flows into microscopic liquid droplets containing ~one cell each. This droplet is then charged and passes through an electrostatic deflection system that diverts the droplet to the appropriate collection container. The flow analysis is also sensitive enough that cells can be sorted not just on a binary presence or absence of a marker, but high, medium, or low expression. Multiple markers can be layered to create a gating scheme to identify populations that have a specified set of markers. This allows a mixture containing heterogenous populations of cells to be separated by surface markers and sorted for downstream applications and analysis.

In Chapter 2, FACS was crucial for sorting colonic epithelial cell populations for bulk RNA-seq [11]. Methodology for the FACS protocol that was developed is presented in Appendix A [41].

Organoids

Culturing human cell lines has enabled phenomenal advances in our understanding of human biology and development of medicine. However, although these model systems are still frequently used today, there are limitations. Cells must be transformed in some manner to enable constant proliferation; cells from normal tissue cannot be maintained in this way as the Hayflick effect emerges to drive senescence and the loss of proliferative capacity. Many cancer cell lines were derived from cancer patients decades ago and now, in some biological respects, the cultured cells no longer accurately represent the initial cancer cells from the patient tumor. Additionally, the standard manner of culturing cell lines as a monolayer on hard plastic does not recapitulate the soft tissue mechanics of the human body. For these reasons, and more, improved *in vitro* systems that more accurately mimic the human body have been sought after.

In recognition of the limitations of 2D culture systems, researchers have been developing 3D culture systems since the 1970s [42,43]. The first organoids, or 'mini organs', were small intestine organoids initially created and published by the Clevers lab in 2009, following their identification of Lgr5+ intestinal stem cells in 2007 [44,45]. Sustained growth of normal untransformed tissue is enabled by identifying the adult stem cells and optimizing culture conditions to enable their growth [46]. Importantly, adult stem cells in organoid cultures not only maintain stemness but preserve the ability to form differentiated progeny and distinct lineages when given the appropriate signals. Tumor tissue may not need all of the same growth factors/conditions as its paired normal tissue but it can recapitulate human disease and drug responses [47–49]. Importantly organoids are grown in a thick artificial extracellular matrix that better approaches the tissue mechanics in the body, a matrix that enables 3D structures and growth in all directions and requires no adhesion to a plastic surface. Researchers have also been developing specific organoid conditions for all types of tissue and disease states [46,50]. Organoids can be rapidly grown directly from patient tissue and many researchers are currently using freshly derived patient organoids to test drug therapies for individual patients [48,51,52].

In Chapter 4, human and mouse organoids are used as a model system to investigate the effects of altered levels of metabolite α-ketoglutarate [53]*. Methodology for human organoid culturing and biobank development are presented in Appendix B.*

Sequencing

Sequencing describes the process of determining the composition of nucleic acid in a sample. Original methodologies relied on Sanger sequencing and were slow, tedious, and costly. The sequencing field has greatly advanced to where efficient high throughput sequencers now enable faster, more accurate and more efficient technologies. The Next Generation Sequencing revolution has resulted in the production of an enormous amount of sequencing data [54,55]. The sequencing data is in the form of millions of reads that must then be bioinformatically aligned to a reference genome and quantified in order to gain information about gene expression. Using additional pipelines, transcriptome information related to RNA processing, such as splicing and polyadenylation, can be investigated [56].

Two types of sequencing used to generate data in this dissertation are bulk RNA-seq and single cell RNA-seq (scRNA-seq). For bulk RNA-seq, thousands of cells are pooled together as one sample, and the transcriptome of these cells are merged as one signature. In contrast, scRNA-seq, looks at the transcriptome of many cells, but individually sequences each cell. One advantage of bulk RNA-seq is that it enables a greater depth of sequencing and transcript coverage than scRNA-seq. This means that lowly expressed genes can be detected in bulk RNA-seq and more information about isoforms and RNA processing can be gathered. In scRNA-seq lowly expressed genes may not be detected, and moderately expressed genes may only be detected in some cells. However, scRNA-seq is advantageous to investigate cell-to-cell heterogeneity, variations in the transcriptome that are not possible with bulk RNA-seq since cells are pooled. scRNA-seq also allows the identification of rare cell type transcriptomes, signatures that are lost in bulk RNA-seq because of their underrepresentation. Both techniques are valuable tools that enabled investigation of the transcriptome.

Bulk RNA-seq data is presented in Chapter 2 and 4, while scRNA-seq is shown in Chapter 3. Data generated via bioinformatics analyses facilitated discoveries in all chapters, most notably in Chapter 2 where *alternative mRNA processing analyses were performed. In Appendix C, an RNA-seq data analysis app we developed, GECO, is described.*

CHAPTER GUIDE

The remainder of this dissertation will take a deeper dive into intestinal stemness by first investigating normal intestinal homeostasis (Chapter 2) and then the activity of distorted intestines that arise during wounding (Chapter 3) and cancer (Chapter 4). Identification of transcriptomic and proteomic signatures of stemness and differentiation in Chapter 2 was enabled by the development of an intestinal flow sorting procedure (Appendix A). Human colonic organoids and our human tissue biobank (Appendix B) were a crucial model system in exploring the role of the metabolite α-ketoglutarate in promoting differentiation in colon cancer (Chapter 4). A common theme throughout this dissertation is the incorporation of RNA-sequencing and the use of bioinformatic analysis of big datasets to answer questions about intestinal cell transcriptomes. GECO, an app that was developed to observe patterns in bulk RNA-seq datasets, is highlighted in Appendix C. Overall, this dissertation contributes to our understanding of intestinal stemness and the role stemness plays in normal and disease states.

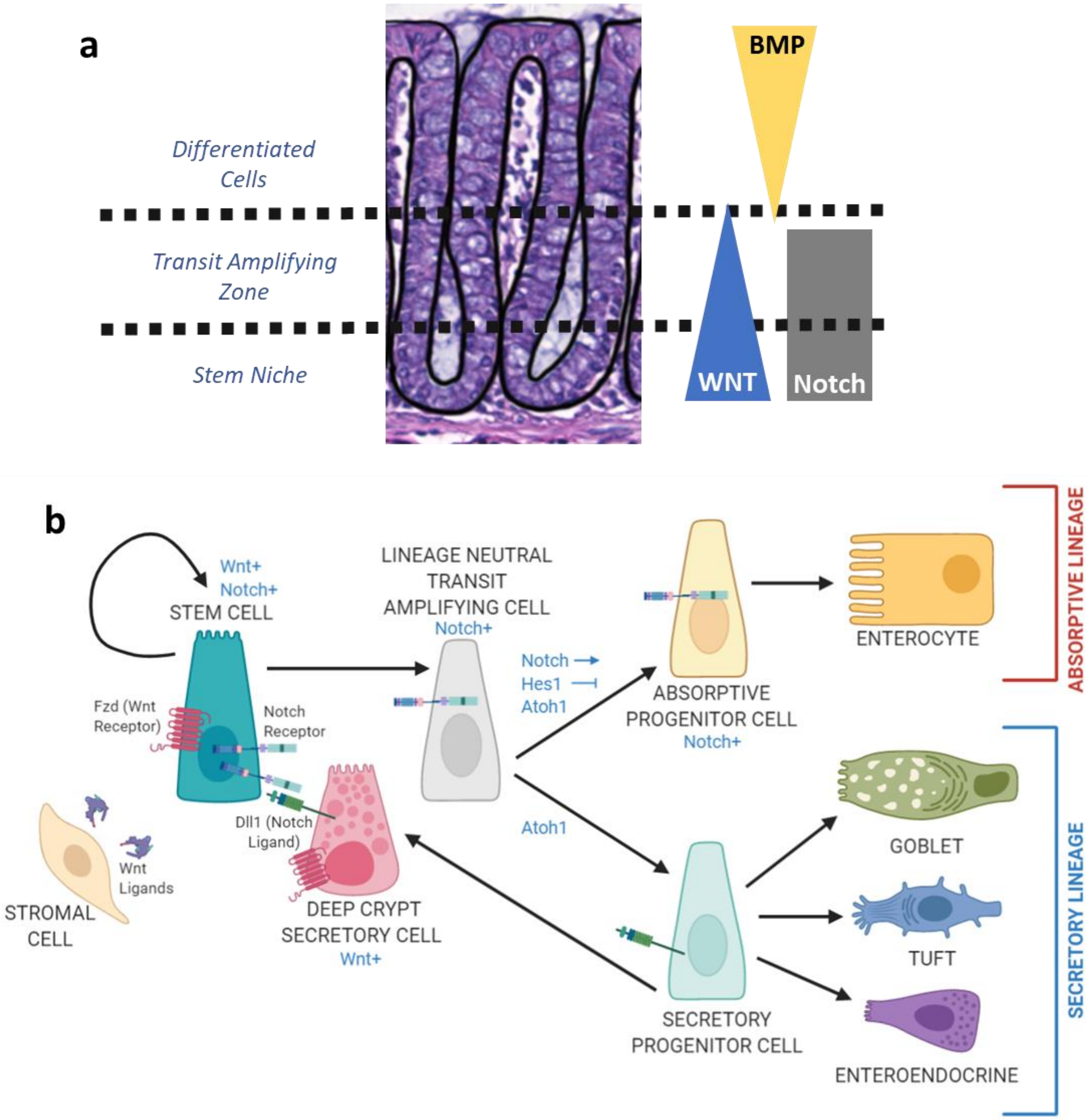

a
Differentiated Cells
Transit Amplifying Zone
Stem Niche
BMP
WNT
Notch
b
Wnt+
Notch+
STEM CELL
Fzd (Wnt Receptor)
Notch Receptor
Dll1 (Notch Ligand)
Wnt Ligands
STROMAL CELL
DEEP CRYPT SECRETORY CELL
Wnt+
LINEAGE NEUTRAL TRANSIT AMPLIFYING CELL
Notch+
Notch →
Hes1 ⊣
Atoh1
Atoh1
ABSORPTIVE PROGENITOR CELL
Notch+
ENTEROCYTE
ABSORPTIVE LINEAGE
SECRETORY PROGENITOR CELL
GOBLET
TUFT
ENTEROENDOCRINE
SECRETORY LINEAGE

Figure 1.1: ***Structure of the colonic crypt***

a H&E section of a mouse colonic crypt, with epithelial crypt cell outlined in black. Three distinct regions are identified and include (1) the stem cell niche at the base, (2) the transit amplifying zone containing proliferating daughter cells, and (3) differentiated cells that dominate the top of the crypt. Many signals regulate stemness, loss of stemness, and lineage commitment in the intestinal crypt. The levels of three important signals in the crypt are shown in the diagram: Wnt (promoting stem cell niche), Notch (lineage commitment), and Bmp (promotes differentiation). **b** Notch signaling plays an important role in lineage specification; receiving cells express Notch receptor (Stem and absorptive) and sending cells expression Notch ligand (secretory) and this facilitates lateral inhibition. If there are too few secretory cells, the level of Notch signaling will be lowered and more secretory cells will be produced enabling a homeostatic proportion of both lineages. Progenitor cells differentiate into mature differentiated cells of each lineage.

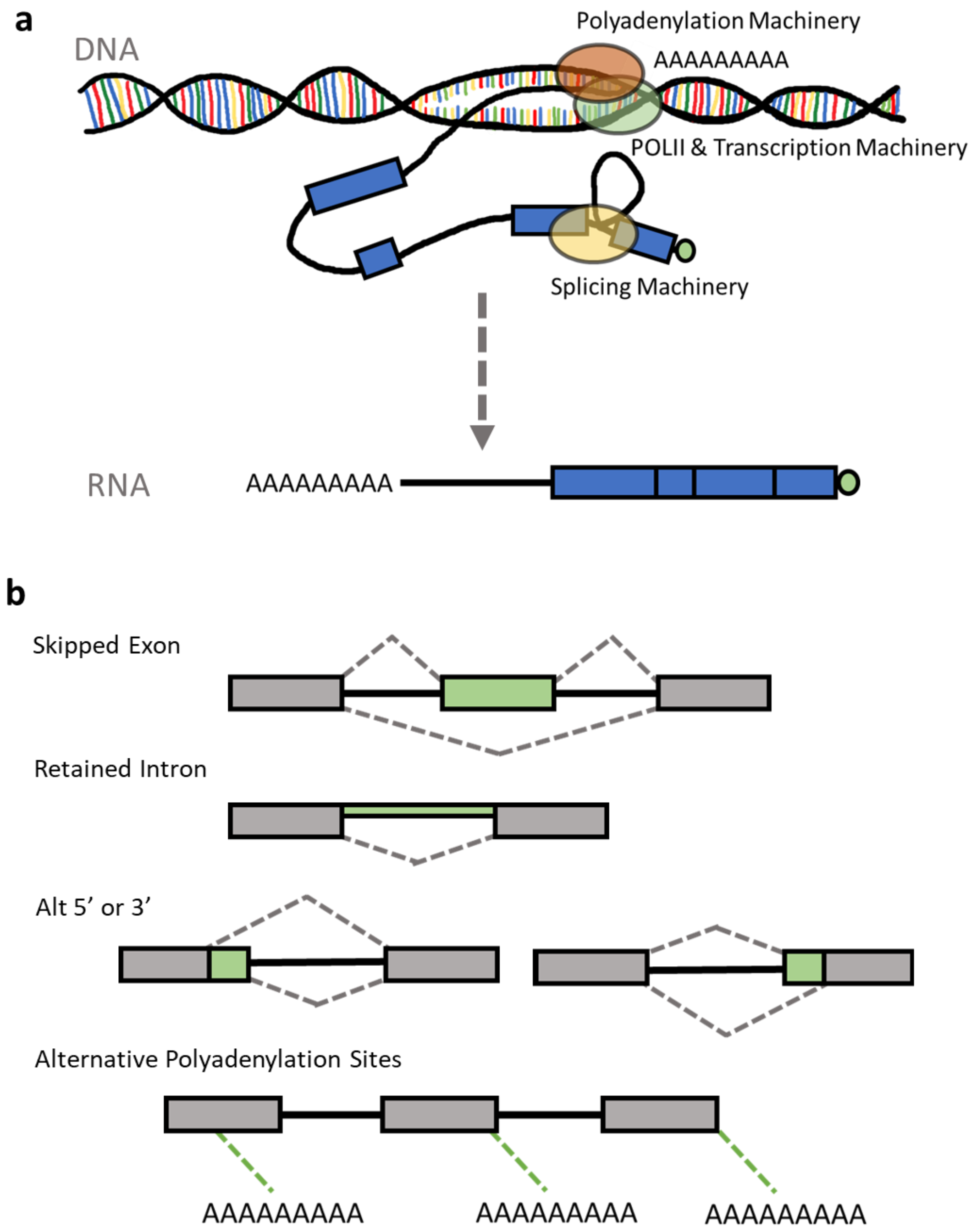


Figure 1.2: ***mRNA processing influences the transcriptome***

a As RNA is transcribed from DNA, mRNA processing machinery assembles and modifies the RNA. These modifications include a 5′ protective cap, splicing together of exonic regions, and truncating the messages by adding a polyadenylation tail. **b** Variations in the processing events are known as alternative processing. There are several different alternative splicing events (skipped exon, retained intron, alternate 5′ or 3′, or any combination of these events). Alternative polyadenylation can result in a polyA tail at several different points in the transcript.

REFERENCES

1. Conlon, M. A. & Bird, A. R. The impact of diet and lifestyle on gut microbiota and human health. *Nutrients* **7**, 17–44 (2015).
2. Velasquez-Manoff, M. Gut microbiome: The peacekeepers. *Nature* **518**, S3–S11 (2015).
3. Shreiner, A. B., Kao, J. Y. & Young, V. B. The gut microbiome in health and in disease. *Current Opinion in Gastroenterology* **31**, 69–75 (2015).
4. Nguyen, H. *et al.* Deficient Pms2, ERCC1, Ku86, CcOI in field defects during progression to colon cancer. *J. Vis. Exp.* **41**, (2010).
5. Humphries, A. & Wright, N. A. Colonic crypt organization and tumorigenesis. *Nat. Rev. Cancer* **8**, 415–424 (2008).
6. Kwong, L. N. & Dove, W. F. APC and its modifiers in colon cancer. *Adv. Exp. Med. Biol.* **656**, 85–106 (2009).
7. Krausova, M. & Korinek, V. Wnt signaling in adult intestinal stem cells and cancer. *Cellular Signalling* **26**, 570–579 (2014).
8. Beumer, J. & Clevers, H. Cell fate specification and differentiation in the adult mammalian intestine. *Nat. Rev. Mol. Cell Biol.* 1–15 (2020). doi:10.1038/s41580-020-0278-0
9. Farin, H. F. *et al.* Visualization of a short-range Wnt gradient in the intestinal stem-cell niche. *Nature* **530**, 340–343 (2016).
10. Reya, T. & Clevers, H. Wnt signalling in stem cells and cancer. *Nature* **434**, 843–850 (2005).
11. Habowski, A. N. *et al.* Transcriptomic and proteomic signatures of stemness and differentiation in the colon crypt. *Commun. Biol.* **3**, 1–17 (2020).
12. Licatalosi, D. D. & Darnell, R. B. RNA processing and its regulation: Global insights into biological networks. *Nature Reviews Genetics* **11**, 75–87 (2010).
13. Herzel, L., Ottoz, D. S. M., Alpert, T. & Neugebauer, K. M. Splicing and transcription touch base: Co-transcriptional spliceosome assembly and function. *Nature Reviews Molecular Cell Biology* **18**, 637–650 (2017).
14. Tian, B. & Manley, J. L. Alternative polyadenylation of mRNA precursors. *Nature Reviews Molecular Cell Biology* **18**, 18–30 (2016).
15. Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nature Reviews Molecular Cell Biology* **18**, 437–451 (2017).
16. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nature Reviews Genetics* **17**, 487–500 (2016).
17. Etchegaray, J. P. & Mostoslavsky, R. Interplay between Metabolism and Epigenetics: A Nuclear Adaptation to Environmental Changes. *Molecular Cell* **62**, 695–711 (2016).
18. Cheng, H. & Leblond, C. P. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine V. Unitarian theory of the origin of the four epithelial cell types. *Am. J. Anat.* **141**, 537–561 (1974).

19. Snippert, H. J. *et al.* Intestinal Crypt Homeostasis Results from Neutral Competition between Symmetrically Dividing Lgr5 Stem Cells. *Cell* **143**, 134–144 (2010).

20. Gregorieff, A. & Clevers, H. Wnt signaling in the intestinal epithelium: from endoderm to cancer. *Genes Dev.* **19**, 877–90 (2005).

21. Andersson-Rolf, A., Zilbauer, M., Koo, B. K. & Clevers, H. Stem cells in repair of gastrointestinal epithelia. *Physiology* **32**, 278–289 (2017).

22. Potten, C. S., Owen, G. & Booth, D. Intestinal stem cells protect their genome by selective segregation of template DNA strands. *J. Cell Sci.* **115**, 2381–2388 (2002).

23. Buczacki, S. J. A. *et al.* Intestinal label-retaining cells are secretory precursors expressing Lgr5. *Nature* **495**, 65–69 (2013).

24. Muñoz, J. *et al.* The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers. *EMBO J.* **31**, 3079–3091 (2012).

25. van Es, J. H. *et al.* Dll1+ secretory progenitor cells revert to stem cells upon crypt damage. *Nat. Cell Biol.* **14**, 1099–1104 (2012).

26. Tian, H. *et al.* A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. *Nature* **478**, 255–259 (2011).

27. Tetteh, P. W. *et al.* Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters In Brief. (2016). doi:10.1016/j.stem.2016.01.001

28. Rubin, D. C., Shaker, A. & Levin, M. S. Chronic intestinal inflammation: Inflammatory bowel disease and colitis-associated colon cancer. *Frontiers in Immunology* **3**, (2012).

29. Rieder, F., Brenmoehl, J., Leeb, S., Schölmerich, J. & Rogler, G. Wound healing and fibrosis in intestinal disease. *Gut* **56**, 130–139 (2007).

30. Armaghany, T., Wilson, J. D., Chu, Q. & Mills, G. Genetic alterations in colorectal cancer. *Gastrointest. Cancer Res.* **5**, 19–27 (2012).

31. Sprowl-Tanio, S. *et al.* Lactate/pyruvate transporter MCT-1 is a direct Wnt target that confers sensitivity to 3-bromopyruvate in colon cancer. *Cancer Metab.* **4**, 20 (2016).

32. Pate, K. T. *et al.* Wnt signaling directs a metabolic program of glycolysis and angiogenesis in colon cancer. *EMBO J.* **33**, 1454–1473 (2014).

33. Lyou, Y., Habowski, A. N., Chen, G. T. & Waterman, M. L. Inhibition of nuclear Wnt signalling: challenges of an elusive target for cancer therapy. *Br. J. Pharmacol.* **174**, 4589–4599 (2017).

34. Chen, G. *et al.* Disrupting ß-catenin dependent Wnt signaling activates an invasive gene program predictive of colon cancer progression. *Disrupting ß-catenin Depend. Wnt Signal. Act. an invasive gene Progr. Predict. colon cancer Progress.* 667030 (2019). doi:10.1101/667030

35. Croker, A. K. & Allan, A. L. Cancer stem cells: Implications for the progression and treatment of metastatic disease: Stem Cells Review Series. *J. Cell. Mol. Med.* **12**, 374–390 (2008).

36. Lapidot, T. *et al.* A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367**, 645–648 (1994).

37. Hope, K. J., Jin, L. & Dick, J. E. Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. *Nat. Immunol.* **5**, 738–743 (2004).

38. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105–111 (2001).

39. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).

40. Melamed, M. R. A brief history of flow cytometry and sorting. *Methods in Cell Biology* 3–17 (2001). doi:10.1016/s0091-679x(01)63005-x

41. A.N. Habowski, J.M. Bates, J.L. Flesher, R.A. Edwards, M. L. W. Isolation of murine large intestinal crypt cell populations with flow sorting. *Protoc. Exch.* (2020).

42. Simian, M. & Bissell, M. J. Organoids: A historical perspective of thinking in three dimensions. *J. Cell Biol.* **216**, 31–40 (2017).

43. Lancaster, M. A. & Knoblich, J. A. Organogenesis in a dish: Modeling development and disease using organoid technologies. *Science (80-.).* **345**, 1247125–1247125 (2014).

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

45. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003–1007 (2007).

46. Kaushik, G., Ponnusamy, M. P. & Batra, S. K. Concise Review: Current Status of Three-Dimensional Organoids as Preclinical Models. *Stem Cells* **36**, 1329–1340 (2018).

47. Sato, T. *et al.* Long-term Expansion of Epithelial Organoids From Human Colon, Adenoma, Adenocarcinoma, and Barrett's Epithelium. *Gastroenterology* **141**, 1762–1772 (2011).

48. Van De Wetering, M. *et al.* Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell* **161**, 933–945 (2015).

49. Drost, J. *et al.* Sequential cancer mutations in cultured human intestinal stem cells. *Nature* **521**, 43–47 (2015).

50. Clevers, H. Modeling Development and Disease with Organoids. *Cell* **165**, 1586–1597 (2016).

51. Fatehullah, A., Hui Tan, S. & Barker, N. Organoids as an in vitro model of human development and disease. *Nat. Cell Biol.* **18**, 246–254 (2016).

52. Driehuis, E., Kretzschmar, K. & Clevers, H. Establishment of patient-derived cancer organoids for drug-screening applications. *Nat. Protoc.* 1–30 (2020). doi:10.1038/s41596-020-0379-4

53. Tran, T. Q. *et al.* α-Ketoglutarate attenuates Wnt signaling and drives differentiation in colorectal cancer. *Nat. Cancer* **1**, 345–358 (2020).

54. Kchouk, M., Gibrat, J. F. & Elloumi, M. Generations of Sequencing Technologies: From First to Next Generation. *Biol. Med.* **09**, (2017).

55. Muir, P. *et al.* The real cost of sequencing: Scaling computation to keep pace with data generation. *Genome Biol.* **17**, 53 (2016).

56. de Klerk, E. & ’t Hoen, P. A. C. Alternative mRNA transcription, processing, and translation: Insights from RNA sequencing. *Trends in Genetics* **31**, 128–139 (2015).

CHAPTER 2

Transcriptomic and Proteomic Signatures of Stemness and Differentiation in the Colon Crypt

Text adapted from published work:

Habowski, A.N*., J.L. Flesher, J.M. Bates, C. Tsai, R. Zhao, K. Martin, A.K. Ganesan, R.A. Edwards, T. Shi, H.S. Wiley, Y. Shi, K.J. Hertel, and M.L. Waterman. (2020) Transcriptomic and proteomic signatures of stemness and differentiation in the colon crypt. Communications Biology, 3:453. DOI: 10.1038/s42003-020-01181-z.*

ABSTRACT

Intestinal stem cells are non-quiescent, dividing epithelial cells that rapidly differentiate into progenitor cells of the absorptive and secretory cell lineages. The kinetics of this process is rapid such that the epithelium is replaced weekly. To determine how the transcriptome and proteome keep pace with rapid differentiation, we developed a new cell sorting method to purify mouse colon epithelial cells. Here we show that alternative mRNA splicing and polyadenylation dominate changes in the transcriptome as stem cells differentiate into progenitors. In contrast, as progenitors differentiate into mature cell types, changes in mRNA levels dominate the transcriptome. RNA processing targets regulators of cell cycle, RNA, cell adhesion, SUMOylation, and Wnt and Notch signaling. Additionally, global proteome profiling detected >2,800 proteins and revealed RNA:protein patterns of abundance and correlation. Paired together, these data highlight new potentials for autocrine and feedback regulation and provide new insights into cell state transitions in the crypt.

INTRODUCTION

The intestinal crypt is a good model for studying how stem cells support a rapidly renewing tissue. Crypts are invaginating structures of single layer epithelium in which stem cells reside in a supportive niche at the base where they produce daughter cells (progenitors). Progenitors move up the crypt to differentiate and replace mature cells that are dying at the mucosal surface - a process with an average lifetime of only 4-5 days [1]. Constant replacement maintains homeostasis and proper absorption of water and nutrients, but the fast timescale of birth-to-death places great demand on both stem and daughter cells. Stem cells are by necessity non-quiescent and rapidly dividing, and progenitor cells exhibit rapid loss of stemness and commitment to differentiation. Multiple studies have shown how absorptive and secretory cell types can respond to wounding by de-differentiation and repopulation of the stem cell compartment [2–5].

Although this de-differentiation process occurs promptly, it is unknown if these reverse changes in cell state and gain of stemness occur on a similar rapid time scale as loss of stemness.

Quantitative imaging and lineage tracing tools have shown that newly produced progenitor cells lose stemness as they move from the stem cell niche into a compartment called the transit amplifying zone (TAZ) [6]. The progenitor's first round of cell division and commitment to either an absorptive (AbsPro) or secretory (SecPro) lineage happens nearly simultaneously with entrance into this zone. These changes occur within minutes-to-hours of each other, suggesting that loss of stemness and choice of cell lineage are connected and directed by processes that occur on this timescale.

Several signal transduction systems are important for the early changes in cell state. A decrease in Wnt signaling and activation of the Unfolded Protein Response (UPR) correlates with loss of stemness [7]. In addition, Notch signaling balances commitment to either the absorptive or secretory lineage through lateral inhibition signaling [4]. There has been a longstanding expectation that stem cells are defined by a unique transcriptome and that loss of stemness and lineage commitment are similarly defined by unique signatures. However, while signaling systems are capable of altering transcription, it is not known how much of the rapid changes in cell state are due to the turning ON/OFF of signal-targeted gene programs versus more immediate processes of co- and post-transcriptional processing, such as alternative mRNA splicing and alternative polyadenylation [8–10]. Each of these processes can quickly modify the nascent transcriptome and its attendant proteome by altering the coding sequences of mRNAs, the localization or interactions of mRNA and proteins, or by changing protein abundance through alterations in mRNA stability and/or protein translation rates [11–14].

To study how transcription and post-transcription processes contribute to stemness and differentiation, it is necessary to separate stem cells, daughter cells, and their differentiated progeny. Multiple cell sorting protocols have been optimized to isolate stem cells, but each lack resolution of these three cell types [15,16].

For example, the transgenic stem cell lineage marker Lgr5-EGFP enables purification of GFP-bright stem cells, but a mosaic expression pattern of the transgene in the intestine has made it difficult to confidently separate daughter cells from GFP-negative stem cells and differentiated cells [17,18]. Single cell RNA sequencing captures the diversity when analyzing mixed cell populations and has been useful for defining intestinal lineage trajectories and diversity of mature cells (for example enterocytes and enteroendocrine cells) [19–23]. However, its low sequencing depth misses moderate-to-lowly expressed transcripts and mRNA splicing and polyadenylation analyses are not yet reliable. Therefore, the transcriptome and proteome basis for loss of stemness and early commitment is unknown.

Here we developed a new cell sorting protocol for purification and comparative analysis of colon stem cells, their immediate daughters (SecPro, AbsPro), and their differentiated cell types, including tuft cells, enteroendocrine cells (EEC) and enterocytes (Ent). The protocol can be used with non-transgenic mice of any strain and importantly, when coupled to bulk RNA sequencing and mass spectrometry-based global proteome profiling, can provide a deeper analysis of cellular transcriptomes and proteomes. Using this protocol, we found that while the transcriptome and proteome of each cell type are generally correlated, deeper analyses of the bulk RNA-seq data reveal that loss of stemness and lineage commitment are accompanied by a greater change in mRNA splicing and polyadenylation than in gene expression, a pattern that largely resolves as progenitor cells mature. Sequencing analysis also enabled higher resolution of signal transduction systems (Wnt and Myc signaling), environmental sensing pathways, and patterns of lineage distinction, including prostaglandins and Fgf signaling pathways. These patterns were seen at both the RNA and protein level and are likely key to understanding the processes of homeostasis, namely: (i) loss of stemness, (ii) lineage commitment, and (iii) signaling connections between mature cell types. We relate how these findings are relevant to the earliest events that happen during loss of stemness and we highlight ways in which mature cells might de-differentiate to re-acquire the state of stemness.

RESULTS

Flow sorting purification of colon crypt cell populations

To create a high-resolution profile of colon crypt stem cells and their daughter cells, we developed a new flow sorting protocol using freshly dissected, wild-type C57BL/6N mouse colons and antibodies to validated intestinal cell surface markers including Cd44 (Figure 2.1a,b, Supplementary Figure 2.1). Upon discovery that Cd44 is highly sensitive to TrypLE, and other commonly used proteases [16] (Supplementary Figure 2.2a), we developed a dissociation protocol that uses only EDTA and mechanical force. This change resulted in a 10-fold increase in detectable Cd44 antigen surface expression and therefore higher resolution for cell sorting (Figure 2.1c, Supplementary Figure 2.2b). Using additional commonly used cell surface markers, six cryptal populations could be isolated. A previously validated intestinal stem cell signature of Cd44-high, Cd24-low, and cKit-negative was used to identify and isolate an abundant fraction of stem cells (Figure 2.1c). This cell population directly overlapped with Lgr5-EGFP+ cells from Lgr5-EGFP-IRES-creERT2 mice, confirming their stem cell identity [24] (Supplementary Figure 2.3a).

In addition to the stem cell population, five additional Epcam-positive populations were collected (Figure 2.1c) and replicate biological samples of the six populations were processed for bulk RNA-seq (Supplementary Data #1). These populations represent cellular subtypes from two distinct cryptal lineages (secretory and absorptive), each revealing a trajectory of differentiation from stem to mature populations (Figure 2.1d). Principal Component (PCA) analysis of mRNA and protein from the sorted cells showed that these cryptal populations were distinct and replicates tightly clustered (Figure 2.1e-f). Known markers for colon crypt cells were used to identify cell types (Supplementary Figure 2.4), which clearly demonstrated the presence of stem cells, two distinct populations of progenitor cells (absorptive and secretory), and three mature, differentiated populations (enterocytes, tuft cells and enteroendocrine (EEC) cells). Thus, our new protocol for crypt isolation and the greater range of Cd44 surface expression it preserves, enables a meaningful improvement in the resolution and sorting of stem cells from daughter cells and differentiated progeny (Figure 2.1c,d). Specifically, it is now possible to distinguish stem cells from Absorptive Progenitor cells (AbsPro; $Cd44^{Med}$) and from mature enterocytes (Ent; $Cd44^{Low/-}$). Secretory progenitors were identified as SecPDG as this population contains mostly secretory progenitors and deep crypt secretory cells, with a possible minor contribution of goblet cells, a cell type that is largely missing from our isolated cells (SecPDG, $Cd44^{Med}$, Figure 2.1c,d, Supplementary Figure 2.5). SecPDG markers were confirmed via immunohistochemical staining of human colon and small intestine (Supplementary Figure 2.5). Finally, our protocol's preservation of Cd44 expression, along with cKit expression, enabled resolution of two rare Epcam+/$Cd24^{high}$ populations identified as tuft cells and enteroendocrine cells, which are mature cell types from the secretory lineage (Figure 2.1c,d). EECs were predominantly enterochromaffin (Supplementary Figure 2.6). Tuft cells were comprised of both Tuft-1 and Tuft-2 subtypes (Supplementary Figure 2.7; [19]). For each of the isolated cell types we identified strongly associated biomarkers, including novel highly expressed proteins confirmed via proteomic analysis and immunohistochemistry (Supplementary Figure 2.7-2.9, Supplementary Data #2). In the case of tuft cells, we detected taste-directed signaling pathways that are distinct from tuft cells in the small intestine (Supplementary Figure 2.10) [25,26]

Pairing the FACS sorting protocol with new methods for global proteome analysis of small numbers of cells (<200 cells) enables us to compare the transcriptome and proteome for all six cell populations [27,28]. Despite the use of small cell numbers, particularly for the rare EEC and Tuft cell populations, we were able to quantify the expression of over 2,800 proteins and investigate RNA:protein correlation patterns (Supplementary Figure 2.11a, 2.12). General crypt markers, such as Epcam and Vil1 (Supplementary Figure 2.11b), were detected along with markers of mature cell types (Supplementary Figure 2.7g, 2.8g, 2.9g), and progenitor cell types (Supplementary Figure 2.11c,d). We also confirmed that mRNA expression levels of the surface protein markers used in the FACS protocol could accurately cluster cell types

(Supplementary Figure 2.13), confirming that at least for the sorting markers, the mRNA and protein expression are congruent. To determine whether our protocol is broadly useful we sorted colon epithelia from four additional commonly used mouse strains (Agouti, FVB, BALB/c, and NSG) and from female mice (Supplementary Figure 2.14). The sorting results were nearly the same, demonstrating that the procedure reliably distinguishes colon crypt cell types regardless of mouse strain or gender.

Redefining markers of crypt stem cells

The clear separation of stem cells from progenitors and mature cell types enabled us to define global patterns of gene expression from the early stages of loss of stemness and lineage commitment (comparing stem cells with AbsPro and SecPDG) to the final steps of differentiation (Ent, tuft, EEC; Figure 2.2a). We observed several notable trends in gene expression. First, differentiation is generally accompanied by an increase rather than a decrease in gene expression (Figure 2.2a). This is especially striking during loss of stemness and commitment to the absorptive and secretory lineages where there is a significant increase in the expression of 232 and 1,177 genes in the absorptive and secretory progenitors, respectively, in contrast to a decrease in 69 and 492 genes in those populations (Figure 2.2a; padj < 0.01 + minimum mean 50 counts). Fully committed, differentiated enterocytes, tuft cells, and EEC populations show 4.1, 2.7, and 4.2-fold differences in gene activation:suppression compared with stem cells, suggesting that differentiation in the colon crypt is defined more by gene activation rather than suppression of a distinct stem cell program. In addition, the transcriptomic stem cell signature is not shut off abruptly, but instead declines gradually (Figure 2.2b). Thus, stem cells are defined more by the absence of differentiated cell markers. This applies to well-known intestinal stem cell markers such as *Lgr5, Smoc2, Cd44, Cdca7, Notch1,* and *Rnf43* which show elevated expression in stem cells, but are well expressed in the other cell populations (Supplementary Figure 2.15a). *Lgr5* is a notable example as its levels decrease by 4-fold in AbsPro and SecPDG, but only 2-fold in the fully differentiated tuft cells demonstrating that *Lgr5* expression is not unique to the stem compartment (Supplementary Figure 2.3b). Indeed, we could demonstrate *Lgr5* expression in tuft cells at the protein level using flow cytometry of colon crypt epithelia from Lgr5-EGFP-IRES-creERT2 (Supplementary Figure 2.3c).

Although these data suggest that colon crypt stem cells have few specific markers, our analysis identified a set of 16 highly enriched mRNAs that distinguish stem cells from all other cell populations (Figure 2.2c; padj < 0.01 + minimum mean 50 counts). Some of these mRNAs are known stem cell markers (*Cd44, Rnf43, Notch1*) and Wnt signaling targets (*Axin2, Rnf43*), but newly identified markers are connected to epigenetics processes (*Lmnb1, Whsc1, Mybbp1a, Nap1l1, Prmt1, Aatf,* and *Arid5b*), regulation of the cell cycle (*Aatf, Cdk4, Trp53* – Supplementary Figure 2.16), and transcription regulators (*Mybbp1a, Arid5b, Zbtb38* – Supplementary Figure 2.17). Several markers were detected in the proteomics analysis as consistently elevated in stem cells (Figure 2.2d). We also identified several additional protein markers that gradually decrease in protein and mRNA expression as cells transition to the progenitor stage (and thus do not pass our stringent significance cut-off of differentially expressed between stem and progenitor) (Supplementary Figure 2.15b). RNA markers of proliferation (*Mki67, Pcna*, and *Mcms*) are highest in stem cells, but interestingly, their protein products are readily detectable in differentiated cells, thus highlighting inconsistencies between mRNA and protein biomarkers of proliferation (Supplementary Figure 2.18). When we limit the differential gene expression analysis to a comparison of stem and daughter cells, SecPDG and AbsPro, there are an additional 11 mRNAs that are stem cell-enriched (Supplementary Figure 2.15c), bringing the total number of genes that are most highly expressed in stem cells to 27. In contrast, the number of genes/proteins that increase as cells transition to the progenitor stage is larger. The top genes activated at this early step (e.g. *Fcgbp, Tff3, Ptprn2, Zg16,* etc.), are shown in Figure 2.2e. If the comparison is extended to all cell types, there are 107 genes that significantly increase in expression in all cell stages and all cell types compared with stem cells (Supplementary Figure 2.15d, example in Supplementary Figure 2.15e, Table 2.1). Gene ontology analysis (Enrichr and Panther) indicates these 107 “non-stem” genes demarcate the large intestine and are cytoplasmic and plasma membrane components (as opposed to factors in the nucleus), such as ion transporters that are involved in the function of mature epithelial cells in the mucosa (Figure 2.2f,g).

RNA processing remodels the intestinal crypt transcriptome

Given that the majority of gene expression changes as measured by mRNA levels are gradual and do not sharply distinguish stem cells from progenitor cell states, we investigated whether other transcriptomic signatures better delineate the rapid transitions of loss of stemness and early commitment. Since alternative pre-mRNA processing has been shown to be important in the differentiation of embryonic stem cells, we asked whether there are differences in alternative splicing and polyadenylation [29–38]. We used two computational pipelines, rMATS Turbo and MAJIQ, to analyze the RNA-seq data to identify significant changes in mRNA splicing patterns among the six cell populations (Figure 2.3a-c, Supplementary Figure 2.19a) [39,40]. With rMATS, we identified 3,659 changes in mRNA splicing amongst all possible comparisons, with the vast majority of these changes detected as skipped exon (SE) events (Figure 2.3b, Supplementary Figure 2.19b). The largest number of alternative mRNA splicing events were during the transition from stem to AbsPro (926 SE events; rMATS, FDR< 0.05), even though there are 3-fold fewer changes in gene expression (301 significant changes in mRNA levels, Figure 2.2a). The relative number of changes in splicing compared to the number of changes in gene expression (mRNA level) can be represented by a Splicing Abundance Ratio (SAR) (Figure 2.3c; [# of significant alternative splicing changes ÷ # of significant gene expression changes x 100]). This metric reveals the extent to which splicing changes dominate the changes in the transcriptome during the transition from stem to AbsPro and stem to SecPDG. During the latter stages of differentiation, however, the number of alternative splicing events is much less than the number of gene expression changes (Figure 2.3c; Supplementary Figure 2.19c).

We next used an alternative polyadenylation (APA) analysis platform, DaPars, to identify changes in APA and to determine the length of 3’ UTR regions in mRNAs [41]. Similar to the patterns of alternative splicing, the largest number of APA events were detected in the transition from stem to AbsPro, followed by stem to SecPDG (Figure 2.3d). However, unlike APA changes observed during embryonic stem cell differentiation [34,37], there is not a dominant, global trend towards lengthening or shortening of 3’ UTRs (Figure 2.3e). Similar to SAR, a Polyadenylation Abundance Ratio was used to quantify the number of changes in polyadenylation relative to the number of changes in gene expression (Figure 2.3f, Supplementary Figure 2.19d). This analysis revealed a pattern similar to mRNA splicing in that there are a greater number of APA events compared to gene expression changes as stem cells transition through loss of stemness and lineage choice and fewer changes during the final stages of differentiation into mature cell types. These data suggest that pre-mRNA processing, rather than gene expression changes, remodels the transcriptome and proteome during loss of stemness and/or lineage commitment.

RNA processing in the loss of intestinal stemness

RNA processing activities can be influenced by regulators and transcription rates and therefore linked to changes in mRNA levels. Alternatively, RNA processing can be a separate regulatory network that modifies the sequences of the existing transcriptome without altering mRNA abundance. We observe that the latter is the case for early stages of differentiation. Fewer than 5% and 20% respectively. of the AbsPro and SecPDG alternatively processed mRNAs showed significant changes in the level of mRNA (Figure 2.4a). This suggests that during loss of stemness, alternative mRNA processing and activation of gene transcription are distinct regulatory programs. Because changes in RNA processing are more common than alterations in gene transcription, the functional role of the processed mRNAs could reveal important details about crypt stem cell biology and loss of stemness. Thus, we identified mRNAs that displayed differential processing in both AbsPro and SecPDG. These changes included 332 genes with SE events in common, and 194 genes with common APA events (Figure 2.4b; the majority of these genes contain the same event). As these changes occur in both progenitor populations, they could potentially be some of the earliest changes in the stem cell transcriptome before lineage transition to an absorptive or secretory progenitor state. Gene ontology analysis of these commonly processed mRNAs shows the most dominant function is protein SUMOylation, such as SUMO enzymes and 'SUMO conjugation to E1' (Figure 2.4c, Supplementary Figure 2.20, 2.21b-c). Other enrichments include programs of mitosis, signaling (SMADs, mTOR, TP53, NMDA, ion channels), and glycosylation, with a number of these genes connected to Notch and Wnt signaling, which are pathways that direct stemness and differentiation in the intestine.

For example, split ends protein (Spen) has four RRM RNA binding domains and functions in splicing and transcription regulation, including suppression of Notch and activation of Wnt signaling [42–46]. Approximately 50% of *Spen* mRNA in stem cells is missing the 4th RRM domain, whereas in the secretory and absorptive progenitor populations, this domain is present in nearly 100% of the *Spen* mRNA (Figure 2.4d, Supplementary Figure 2.22a). Delta-catenin (*Ctnnd1*) has known functions in adhesion as well as Wnt and Notch signaling (Figure 2.4d, Supplementary Figure 2.22a). A C-terminal domain of Ctnnd-1 that binds the Notch1 regulator Numb and the GTPase activator Arhgap-1 is more often encoded in *Ctnnd1* mRNA in stem cells than in progenitor populations. Three examples of APA differences between stem cells and daughter cells (*Top2a* (DNA replication), *Wdhd1* (DNA replication), and *Cby1* (Wnt signaling regulator)) show significant increases in distal polyA choice and lengthening of the 3' UTR (Figure 2.4e, Supplementary Figure 2.22b). Interestingly, strong protein expression of Top2a and Wdhd1 is detected in the TAZ of crypts rather than at the base of the stem cell niche. (Supplementary Figure 2.23).

Previous work using variant-specific antibodies demonstrated that two isoforms of integrin α6 (Itga6) are present in the crypt with Itga6 isoform A (inclusion of exon 25) being more abundant in the base of the crypt, and isoform B (skipping of exon 25) being more abundant near the top of the crypt (Supplementary Figure 2.24a) [47]. Consistent with this, our analysis revealed that exon 25 has the highest inclusion in stem cells, and the lowest in SecPDG and Ent (Supplementary Figure 2.24b). Our global proteomics assays did not detect these isoforms, but it does reveal uniformly high Itga6 protein expression in all cell types along with expression of other adhesion proteins (Supplementary Figure 2.24c-d). Splicing of exon 25 alters the cytoplasmic domain of Itga6 (PDZ-binding domain) and has been linked to stem cell fate determination in several different model systems [48].

RNA processing in intestinal lineage commitment

Commitment of progenitor cells to an absorptive or secretory lineage is a nearly simultaneous event with loss of stemness [6] - an event influenced by signals (e.g. Notch, Wnt, UPR, etc) that activate expression of lineage-specific genes. Significantly, in addition to common splicing and APA changes in both lineages, our analysis detected 469 and 207 lineage-specific changes in alternative mRNA splicing (AbsPro and SecPDG, respectively; see Figure 2.4b). Similarly, 562 distinct changes in polyadenylation were detected in the AbsPro lineage and 431 changes in the SecPDG lineage (see Figure 2.4b). These lineage-specific patterns suggest an important role for splicing and APA in specifying cell fate and lineage choice, and again, the number of processing changes exceeded the number of changes in gene expression (SAR, PAR $\geq$ 100; Figure 2.3c,f). Functional analysis of the alternatively processed genes revealed that the predominant associated processes were chromatin binding and membrane trafficking (Figure 2.5a). In addition, there were enriched functions connected to signaling (Egfr, Wnt), as well as splicing and the cell cycle events (Figure 2.5a-c).

The mRNA encoding the translation regulator *Eif4a2*, a DEAD box RNA helicase involved in translation repression [49], is alternatively spliced in a lineage-specific manner. *Eif4a2* mRNA encodes a full-length protein isoform in the secretory populations (SecPDG, Tuft, EEC) whereas nearly half of the *Eif4a2* mRNA in the stem, AbsPro, and Ent populations encodes a truncated protein isoform (inclusion of exon 11a, Figure 2.5b, Supplementary Figure 2.22a). Total Eif4a2 protein levels in these populations are between 2 to 3-fold less abundant suggesting that this processing, which truncates the open reading frame of *Eif4a2,* could influence protein abundance.

Exon 2 of *Cbfa2t2* (also known as *Mtgr1*) is largely missing in AbsPro mRNA (34% inclusion) but mostly present in SecPro mRNA (71% inclusion) (Supplementary Figure 2.22a). This protein is a transcription regulatory co-factor that interacts with co-repressors (e.g. Prdm14, Ncor, Hdacs and Zbtb33 (Kaiso)), as well as transcription regulatory factors in the Notch (Rbpj) and Wnt (Lef/Tcf) signaling pathways [50,51]. While the functional consequence of this splicing event is not known, *Cbfa2t2* is known to be important in the secretory lineage since knockout of *Cbfa2t2* leads to a loss of secretory cell types as well as a surge in cell proliferation of remaining cell populations [52].

Although the functional consequences of many distal polyA choices are not known, two striking examples of changes in polyA choice in absorptive versus secretory lineage are shown in Figure 2.5c (Supplementary Figure 2.22b). *Rbm3* mRNA encodes an RNA binding protein that enhances Wnt signaling [53], stemness and mRNA stability, and *Ihh* mRNA encodes a Hedgehog signaling ligand that opposes Wnt signaling in intestinal crypts [54]. Polyadenylation of *Rbm3* shifts to a more distal site in AbsPro (Wnt suppressed) whereas polyadenylation of *Ihh* mRNA is shifted to a more distal site in SecPDG (Wnt enhanced). Alternate processing of these genes could potentially contribute to the skewing of Wnt and Notch activities in cells [55].

Gene expression changes in intestinal lineage commitment

Although there are minimal gene expression changes during the initial loss of stemness and transition to progenitor states, changes in mRNA levels become increasingly apparent as progenitor cells differentiate. Our analyses not only identified well established transcriptional signatures of loss of stemness (e.g. UPR) and lineage commitment steps (e.g. Notch), but also identified expression patterns suggesting additional autocrine/paracrine signaling that could impact lineage choice. For example, Notch signaling is known to direct lineage choice via lateral inhibition signaling in small intestinal crypts. Our RNA-seq data indicates that secretory lineage (SecPDG, tuft) cells express high mRNA levels for Notch ligands *Dll1* and *Dll4*, as well as a third ligand *Nov* (Figure 2.6a). Stem and AbsPro populations express the *Notch1* receptor as well as the Notch target gene *Hes1*, showing that Notch signaling is activated to the greatest extent in stem cells and AbsPro [4]. Also consistent with lineage commitment in the small intestine, the Hes1-repressed target gene *Atoh1*, and its downstream target *Spdef* are expressed at the highest levels in the secretory lineage [4,56]. These expression patterns show that the populations we have characterized in the colon align with the Notch-directed lateral inhibition feedback loop identified in the small intestine wherein Notch signaling by secretory cells to absorptive cells balances the proportions of the two mature cell types.

UPR directs cellular responses to ER stress such as growth arrest, apoptosis and/or survival, and can trigger loss of stemness as intestinal stem cells exit their niche [7]. While our analysis indicates activation of UPR in colon crypt progenitors, we observe that UPR signaling is lineage-skewed and most active in secretory populations (Figure 2.6b). Active UPR, as evidenced by target gene expression, protein expression of modulator Ndgr1, and increased splicing of the activated form of *Xbp1*, was observed mostly in SecPDG and tuft cells (Figure 2.6b) [57]. Furthermore, the UPR signal appears to direct survival rather than growth arrest. Specifically, while genes for three UPR sensors (*Atf6, Ern1, Eif2ak3*) were detected in the secretory lineage, the downstream target genes for two of them – Atf6, Ern1 (*Hspa5* and *Hsp90b1*) displayed the highest expression in this lineage. These targets promote ER expansion and survival from stress. Taken together, the increased expression of sensors and downstream targets in the secretory lineage suggests a sensitization to UPR stress that might play a role in lineage choice and/or stabilization. Interestingly, ER stress can slow migration, consistent with recent observations that secretory progenitors migrate up the crypt at a slower rate than absorptive cells [58].

Our analysis also discovered potentials for Fgf autocrine/paracrine signaling that could explain reported knockout phenotypes. Fgf has an important role in crypt homeostasis, although many Fgf ligands in adult mice come from the surrounding stroma [59,60]. Our transcriptional profiling indicated that only a few Fgf ligands are expressed by the epithelia (*Fgf1*, 9, 11, and 12), and predominantly by EECs. Fgf receptors, in contrast, are broadly expressed across the different cell types with *Fgfr3* detected at the highest level in the secretory lineages of SecPDG and tuft (Figure 2.6c). Fgf target gene *Etv5* is most highly expressed in secretory cell types, indicating that the pathway is most active in this lineage (10-fold enriched in SecPDG and tuft; Figure 2.6c).

Gene expression analysis also uncovered potential for lineage-specific autocrine/paracrine activities in prostaglandin signaling (Figure 2.6d, Supplementary Figure 2.25). Consistent with previous reports, we observed that tuft cells express key enzymes for prostaglandin and leukotriene synthesis, including *Ptgs1* (Cox-1), which converts arachidonic acid into prostaglandin H2 (Supplementary Figure 2.25) [61]. Enzymes that convert prostaglandin H2 to the more stable E2 form (*Ptges2* and *Ptges3*) and the prostaglandin transporter *Abcc4* are expressed in all cell types (Supplementary Figure 2.25d). *Ptger4*, a receptor for PGE2 is highly expressed in SecPDG (Figure 2.6d, Supplementary Figure 2.25d). Enterocytes express both an importer (*Slco2a*) for prostaglandins and an enzyme that degrades these molecules (*Hpgd*), suggesting that enterocytes might act as sinks for prostaglandin-mediated signals.

Transcription Regulators and Signaling in mature crypt cells

Wnt signaling and its broader network of cross-talking signaling systems (e.g. Myc, Hippo, Egf, Kit) have a well-established role in maintaining the intestinal stem cell niche and allowing for differentiation of progenitor cells upon exit from that niche. Wnt transcription factors *Tcf7* and *Tcf7l2* are the predominant family members in stem cells, but *Tcf7l1* and *Tcf7l2* are even more highly expressed in progenitor and mature populations along with negative regulators, such as the Tle repressors (Figure 2.7a, Supplementary Figure 2.26 a-d, 2.27a). Hippo mediator Yap1 is expressed ubiquitously but is highest in stem cells. Its binding partners (notably *Tead1* and *Tead3*) are broadly expressed while the direct negative regulator *Insm1* is elevated in EECs (Figure 2.7b, Supplementary Figure 2.26e) [62]. Interestingly, EECs express very high levels of bone morphogenic ligand (*Bmp2*, Figure 2.7b, Supplementary Figure 2.26f). Bmp2 and Yap1 function in a well-characterized signaling circuit in multiple systems [63,64], suggesting that EECs are likely to utilize the autocrine Bmp2-Yap1 signaling pathway. The transcription factor Myc is most highly expressed in stem cells, but its binding partner Max, which can heterodimerize with multiple different E-box factors, is broadly expressed with strong elevation in Ent (Figure 2.7c). Binding partners of Max are strongly expressed in the various populations including direct repressors (Mondo Family, Mnt, Mga, Mad repressors; Figure 2.7c). Expression of these transcriptional regulators and an array of negative regulators in most cryptal cell types implies that there is inherent capacity for gene regulation by their networks, suggesting that the absence of signal-activating ligands and the expression of direct inhibitors keeps these networks in a silent or quiescent state.

Kit and Egf signaling pathways are known to be critical for stem cell homeostasis [65,66], yet their expression patterns suggest that there is potential for additional crosstalk signaling with the rare tuft and EEC cell types. Previous work has suggested that Kit (cKit; Cd117), the receptor for kit ligand (*Kitl;* Stem cell factor) that directs cell survival pathways in stem cell niches, is specific for Paneth cells in the small intestine and DCS/goblet cells in the colon [65]. Although we observed highly expressed *Kit* mRNA in SecPDG and tuft populations (Figure 2.7d), our FACS protocol using Kit as a tuft cell sorting marker shows that at the protein level it is only detectable in tuft cells (Supplementary Figure 1; $\geq$ 5-fold cKit geometric mean and median in tuft compared to SecPDG). We also found that the ligand *Kitl* is most highly expressed in stem and AbsPro populations and to a lesser extent in the Ent population. Complementary expression patterns between the absorptive (Kit ligand) and secretory (Kit receptor) cell populations suggest that Kit could be an intra-cryptal signal from the absorptive lineage to tuft cells. This is mainly a soluble signal since the dominant spliced isoform of *Kitl* (inclusion of exon 6) is the secreted isoform (Figure 2.7d; inset). We observed a related pattern of Egf ligand expression in the colon, with the highest expression detected in SecPDG and tuft (Figure 2.7e). Other Egf-related ligands are most highly expressed in enterocytes. Egf receptor mRNA (*Egfr*), and its negative regulators (*Lrig1*, *Cbl*, and *Ptpn6*; Figure 2.7e,f, Supplementary

Figure 2.27b) are expressed broadly but receptor mRNA levels are highest in EEC and Ent. Interestingly, immunohistochemistry shows that in each crypt, Egfr is only evident in a few cells with morphologies indicative of EEC and tuft cells (Supplementary Figure 2.28; [67]). Other Egf receptor family members, *Erbb2*, and *Erbb3*, are highly expressed in all cell types including stem cells (Figure 2.7e).

Finally, the intestinal crypt is known for its impressive plasticity to rapidly regenerate stem cells at the base of wounded crypts. Multiple studies have shown that the epithelial cell populations, including Ent, tuft, EEC, and progenitor cells have the capacity to de-differentiate into stem cells and restore the niche [2–4]. Although the process of re-acquisition of stemness is not fully understood, our data indicate that colon epithelial populations continue to express mRNAs encoding stem cell regulators (Figure 2.7f), including Lgr5 in tuft cells and embryonic stem cell markers in EECs (Figure 2.7f, Supplementary Figure 2.3b,c). Importantly, EECs and tuft cells also express intestinal reserve stem cell markers, most notably *Msi1*, *Msi2*, and *Prox1*. Furthermore, at least five known RNA regulators that promote pluripotency in embryonic stem cells (*Ddx6, Rbfox2, Son, Srsf2*) are robustly expressed in all colon crypt populations. These expression patterns show that subsets of known stemness regulators are a broadly shared feature of all intestinal crypt cell types.

DISCUSSION

This study presents a high-resolution cell sorting protocol for mouse colon crypt epithelia, an advance that permitted deep RNA-seq and proteomics analyses of multiple cell types including progenitor cells for absorptive and secretory lineages (Figure 2.8). A key feature of our protocol was the elimination of protease treatments, which maximized biomarker sensitivity and cellular resolution and allowed us to clearly separate daughter-progenitor cells from parental stem cells. This advance enabled transcriptomics and proteomics profiling of the early changes occurring during loss-of-stemness and lineage commitment. Our analysis showed that before there are major changes in gene expression, changes in RNA processing

i) “re-configures” the stem cell transcriptome as stem cells lose stemness – altering splicing and polyadenylation patterns, ii) likely influences cell fate choice or stabilization of lineage transitions, and iii) that it does so through global changes in the regulatory networks that shape signal transduction and the proteome, including protein SUMOylation and epigenetic regulation. This suggests that the early stages of cellular differentiation involve a fundamental change in the activity, and/or stability of mRNA and their protein products rather than changes in mRNA levels. In addition to identifying altered RNA processing patterns, our analysis also identified new potentials for autocrine/paracrine signaling between different cell populations in the colon crypt.

Global analysis of gene expression in all six sorted cell populations enabled a more precise identification of stem cell markers (Figure 2.2c), revealing that some commonly used stem cell markers are not exclusive to stem cells (e.g. Lgr5, Smoc2, Cd44; Supplementary Figure 2.15a). We also identified protein markers such as Aquaporin 1, a transporter protein enriched in stem cells even though its mRNA is expressed in other cell types (Aqp-1; Supplementary Figure 2.29). As stem cells differentiate, biomarkers of tissue identity emerge in progenitors and remain expressed in mature cells despite their disparate phenotypes and functions. In other words, we found that stem cells are not so much defined by what they express, but by what they do not express. Most notably, they are distinguished by unique patterns of alternative mRNA splicing and polyadenylation, patterns that dominate transcriptome changes as stem cells begin to differentiate.

The precise point that stem cells lose stemness is not clearly defined but is thought to happen soon after stem cell division as the newly produced progenitor (daughter) cell leaves the stem cell niche and enters the TAZ [6]. The shared changes in RNA processing in the AbsPro and SecPDG populations might therefore represent events that occur during these earliest steps of transition. Indeed, changes in splicing and polyadenylation were detected in regulators of Wnt, Notch, and other known regulators of intestinal stem cells. For example, alternative RNA splicing of delta-catenin mRNA removes an exon that encodes a Numb binding domain in progenitor populations, and it removes an exon for RNA binding domain in Split ends (Spen) mRNA in stem cells (Figure 2.4d). Numb is a regulator of asymmetric cell divisions and a repressor of Notch signaling; Spen regulates early commitment choices of intestinal stem cells with activities that suppress Notch and activate Wnt [68].

SUMOylation is the most significant ontology category associated with commonly processed mRNA targets (Figure 2.4c, Supplementary Figure 2.20). SUMO proteins are ubiquitin-like proteins that are covalently attached to proteins but unlike ubiquitination, SUMOylation influences the activity and/or localization of proteins rather than triggering degradation. Alternative RNA processing of mRNA encoding SUMOylation regulators suggests that this network may function differently in stem cells versus progenitor cells, and perhaps contributes to the earliest changes in transition between states (Supplementary Figure 2.21b,c).

A comparison of gene expression patterns in the secretory and absorptive cell populations revealed new potential intra-cryptal signaling networks, most of which appear to target the secretory lineage. For example, we found that prostaglandin signaling is a potential feedback signal from tuft cells to secretory progenitors. Tuft cells can direct production of prostaglandin E2, while SecPDG progenitors express the PGE2 receptor *Ptger4* (Figure 2.6d), an expression pattern that could explain why knockout of *Ptger4* in the intestine leads to the loss of secretory cells [69]. Another class of prostaglandin receptor, Peroxisome proliferator-activated receptors alpha and gamma (*Ppara* and *Pparg*) are most highly expressed in the absorptive lineage, implying that prostaglandins have different roles in the two lineages (Supplement 2.25d) [70–72]. The overall expression pattern of prostaglandin genes suggests that tuft cells could provide prostaglandin precursors to all cryptal cell types for conversion and whole-cryptal production of PGE2 (*Ptges2, 3*), a potential form of “crowd-sourcing” of a signal known to be important for responding to wounding (Figure 2.6d, Supplementary Figure 2.25).

The activity of the Egf, Fgf, and Kit signaling systems are also enriched in the secretory lineage. Fgf receptors 2 and 3 (Figure 2.6c) are most highly expressed in secretory cells and the Fgf target gene *Etv5* is most highly expressed in SecPDG and tuft (Figure 2.6c). These patterns are consistent with knockout phenotypes in the intestine. For example, *Fgfr3*-knockout mice display enhanced proliferation in the TAZ [73], and knockout of *Fgfr2c* in zebrafish leads to a loss of goblet and EEC [74]. UPR is most active in the secretory lineage (Figure 2.6b), suggesting that like Egf, Fgf, and prostaglandins, this stress signal has a prominent role in commitment and differentiation along the secretory lineage and that the absorptive lineage can exert influences on these four signals. Enterocytes express prostaglandin uptake and degradation enzymes and therefore have potential to function as signal silencers, limiting the concentration and/or duration of signaling to the SecPDG population (Figure 2.6d).

Our study contributes to understanding how multiple cryptal cell types can respond to damage via rapid de-differentiation for crypt repair and stem cell replacement [2–5]. We found that the gene expression patterns of known pluripotency and multi-potency regulators are expressed broadly in the crypt. Thus, these gene loci are not silenced and inactive, but open and expressed, and in some mature cell types expressed at high levels (Figure 2.7c,f) [75]. Likewise, the loss of Wnt signaling during differentiation is not owing to a loss of expression of signaling pathway components. Although decreased Wnt signaling during differentiation is partly due to decreased Wnt ligand availability outside the stem cell niche, our data also show that Wnt inhibitors are expressed in mature populations (Supplementary Figure 2.26c) [76]. As Wnt signaling components continue to be expressed (Supplementary Figure 2.27a), the pathway could be re-activated if ligands become available and/or inhibitor action is overcome. In addition, proposed biomarkers of reserve/quiescent stem cells are expressed broadly in all cell types and strongly expressed in EEC and the secretory lineage. Perhaps most striking is the broad expression pattern of RNA regulators that play key roles in pluripotency by specifying patterns of APA and splicing in embryonic stem cells. Given that RNA processing patterns change markedly in the transition between stem cells and progenitor cells, and then also resolve as mature cells differentiate, the repair of the stem cell niche in wounded crypts might depend on reactivation of these processing changes in wounding and de-differentiation.

In summary, the early emergence of splicing and polyadenylation changes during stem cell differentiation is a novel and unexplored process in the intestinal crypt. This process might not be limited to the intestine but may be a general feature of somatic stem cell differentiation. There are known mRNA processing changes during loss of pluripotency in mouse embryonic stem cells [29–38], and changes in RNA processing have been identified in various disease states. However, very little is known regarding global changes in RNA processing during normal adult tissue homeostasis and especially during the earliest changes that occur during loss-of-stemness. The datasets and analyses presented here lay the groundwork for establishing an important role of mRNA processing as it relates to the rapid crypt dynamics and the de-differentiation potentials of mature intestinal cells.

METHODS

Mouse colon preparation

All mouse work was performed in accordance with NIH guidelines and was approved by the Institutional Animal Care and Use Committee (IACUC) of the University of California, Irvine, approval numbers AUP-17-053. Male C57BL/6N(NJ), obtained from the KOMP repository, mice aged 5-7 weeks were used unless otherwise noted (see Supplementary Figure 2.14). Other mice used include Lgr5-EGFP-IRES-creERT2 mice (B6.129P2-*Lgr5*$^{tm1(cre/ERT2)Cle}$/J, Stock Number 008875) [24], agouti mice (129S1/SvImJ, Stock Number 002448), and NSG mice (NOD.Cg-*Prkdc*^{scid} *Il2rg*$^{tm1Wj1/SzJ}$, Stock Number 005557) which were purchased from The Jackson Laboratory. FVB/NCrl mice (Strain Code 207) and BALB/cAnNCrl mice (Strain Code 028) were purchased from Charles River. A detailed step-by-step procedure is available through Nature Protocol Exchange [77]. In brief, mouse colons (cecum to rectum) were removed, flushed, and linearized. Tissue was dissociated at a slow rotation at 4 °C for 1 hr in a solution of 2 mM EDTA and 10 µM Rock inhibitor.

Aggressive shaking of the tissue solution, filtering (using 100 µm followed by 40 µm filters), and centrifugation (500 – 1,000 xg for 5-10 min at 4 °C depending on the step) were performed to isolate single cells. Data in the Supplementary Figure 2.2show crypt analysis, which included TrypLE (5 mL for 8 min; Life Technologies #12605010) dissociation step after the 100 µm filter step to ensure single cell suspension. A key feature of the sorting protocol is eliminating the use of any protease treatment (notably TrypLE) to preserve maximum cell surface levels of Cd44. The absence of protease action decreased cellular yield, but it also increased biomarker sensitivity and cellular resolution, effectively isolating daughter cells (secretory and absorptive progenitor populations) away from the parental stem cells.

Flow cytometry antibody preparation

Colon crypt single cell suspensions were DNAse treated for 5 min (Sigma-Aldrich #4716728001). Following a wash step, cells were incubated for 30 min in FACS buffer (PBS with 3% FBS + 10 µM Rock inhibitor (Y-27632 AdipoGen Life Sciences from Fisher #501146540)) with the following pre-conjugated validated flow antibodies: CD45-BV510 (1:200, Clone 30-F11; BD Biosciences #563891), CD31-BV510 (1:200, Clone MEC 13.3; BD Biosciences #563089), CD326-eFluor450 (1:100, Clone G8.8; eBioscience #48-5791-82), CD44-PerCP-Cy5.5 (1:100, Clone IM7; Thermo Fisher #A26013), CD24-PECy7 (1:200, Clone M1/69; eBioscience #25-0242-82), and CD117-APC-Cy7 (1:100, Clone 2B8; Thermo Fisher #A15423). Following wash steps, cells were resuspended in FACS buffer and Live/Dead Aqua (Thermo Fisher # L34957). An alternative CD45-APC (1:200, Clone 30-F11; BD Biosciences #561018) antibody was used in the Supplementary Figure 2.7f where specified.

Flow sorting

Cells were bulk sorted on a BD FACS Aria Fusion using a 100 µm nozzle (20 PSI) at a flow rate of 2.0 with a maximum threshold of 5,000 events/sec. The sample chamber and collection tubes were kept at 4 °C. Following exclusion of debris and singlet/doublet discrimination, cells were gated as demonstrated in Supplementary Figure 1. For RNA-seq, populations were sorted into TRIzol (Invitrogen # 15-596-018) for downstream RNA isolation. For global proteome profiling, populations were sorted into PCR tubes containing 50 µL of 100 mM ammonium bicarbonate. At least 100 cells were sorted for each sample and tubes were promptly spun down and frozen until further processing. FACS plots and analysis was done using BD FACSDiva software.

RNA preparation and RNA-seq

RNA was extracted from TRIzol samples using a Direct-zol RNA Micro-Prep kit (Zymo #11-330M) and associated guidelines. Sorted samples of each cell type were pooled as needed at the start of RNA preparations to ensure a minimum of 2,500 cells per sample. RNA sample quality and concentration was evaluated using an Agilent Bioanalyzer on an RNA high sensitivity pico chip. RNA samples were then pooled as needed to allow 1 ng library preps with Clontech Low Input Pico Kit (Takara #634940). Following confirmation of library quality by Agilent Bioanalyzer DNA high sensitivity chip, a total of 22 samples were sequenced (biological replicate numbers stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, EEC=2). Samples were multiplexed and sequencing was performed with 100 bp paired-end run on Illumina HiSeq 4000.

RNA-seq data analysis and visualization

Paired-end sequencing reads were trimmed of adapter sequences and analyzed for quality using Fastqc (version 0.11.7). Data was aligned to the mouse genome (UCSC mm10 from Illumina iGenome) using STAR (version 2.5.2a), converted to bam files and merged (samtools 1.3) and read counts were generated using HTSeq (version 0.6.1p1, with enthought_python version 7.3.2; option -s no). Differential gene expression analysis was done in RStudio (version 1.0.153) with R (version 3.6.1) using default setting of the DESeq2 pipeline for statistical analysis (version 1.16.1; with cooksCutoff = FALSE option) [78]. Gene expression significance was determined by DESeq2 Wald P-value test with a padj < 0.01 with a minimum mean of 50 normalized counts. Heatmaps and PCA plot were generated in RStudio (version 1.0.153) with R (version 3.6.1) using pheatmap (with default scale settings) and plotPCA, respectively, of r-log-transformed (regularized log) DESeq2 data. r-log-transformation is a robust way to transform the count data, used in differential gene expression analysis, to a log2 scale in a way which minimizes differences between samples and normalizes with respect to library size, it is also a standard function for downstream analysis such as clustering or linear discriminant analysis. Bar graphs of gene expression data were generated in GraphPad Prism (version 6.01) with normalized read counts (output of DESeq2) and error bars defining standard deviation. Supplementary Data #1 contains processed global mRNA gene expression data. Raw fastq files along with processed data (counts files) are available for download on GEO (GSE143915).

Splicing and polyadenylation analysis

Merged bam files were sorted and indexed (samtools 1.3) for downstream analysis. Alternative splicing was investigated using rMATS Turbo (rMATS.4.0.1) with STAR 2.5.2a, Samtool 1.3, and enthought_python 7.3.2 comparing two cell types at a time using UCSC mm10 gtf. MAJIQ (v1.1) was also run for alternative splicing with anaconda 3-2.0.1 and recommended mm10 ensembl gff3 reference with type=strand-specific followed by VIOLA for visualization. DaPars (v0.9.1) was used for alternative polyadenylation analysis with recommended mm10 UCSC reference files and python 2.7.15, bedtools 2.25.0, R 3.4.1, and the following settings (Num_least_in_group1=1, Num_least_in_group2=1, Coverage_cutoff=30, FDR_cutoff=0.05, PDUI_cutoff=0.15, Fold_change_cutoff=0.32). rMATS significance was defined in three different levels of significance: FDR < 0.05, FDR < 0.01, FDR < 0.01 with +/- 25% dpsi. Similarly, DaPars significance was defined in three different levels of significance: FDR < 0.05, FDR < 0.01, FDR < 0.01 with +/- 25% PDUI. Alternative processing gene lists are provided in Supplementary Data #3 and rMATS and DaPars output files for cell types compared to stem are available for download on GEO (GSE143915).

Gene ontology and enrichment analysis

Gene ontology and gene enrichment analysis of mRNA-seq data was performed on specified gene lists using Panther [79] (http://pantherdb.org/) and Enrichr [80,81] (https://amp.pharm.mssm.edu/Enrichr/).

Cell lysis and trypsin digestion for proteomic analysis

Prior to sample processing PCR tubes were centrifugated at 1000 xg for 10 min at 4 °C to keep the cells at the bottom of the tube to avoid potential cell loss. In all, 2 µL of 0.1% DDM in 25 mM ammonium bicarbonate was added to each PCR tube with gentle shaking. Intact cells were lysed using sonication 5 times at 1-min intervals over ice and then centrifuged for 3 min at 3000 xg. Samples were then incubated on a thermocycler for denaturation at 75 °C for 1 h. 1 µL and 2 µL of 10 ng/µL trypsin (Promega) in 25 mM ammonium bicarbonate was added to the PCR tubes at a total amount of 10 ng for <1000 cells and 20 ng for >1000 cells. Samples were digested for overnight (~16 h) at 37 °C with gentle sharking at ~500 xg. After digestion, 2 µL of 5% FA was added to the tube to stop enzyme reaction. The final sample volume was reduced down to ~20 µL using SpeedVac and the sample PCR tube was inserted into the liquid chromatagraphy vial for direct liquid chromatography-mass spectrometry (LC-MS) analysis. The processed samples were either analyzed directly or stored at -20 °C for later LC-MS analysis.

LC-MS/MS analysis

The cell subpopulation digests were analyzed using a commonly available Q Exactive Plus Orbitrap MS (Thermo Scientific, San Jose, CA). The standard LC system consisted of a PAL autosampler (CTC ANALYTICS AG, Zwingen, Switzerland), two Cheminert six-port injection valves (Valco Instruments, Houston, USA), a binary nanoUPLC pump (Dionex UltiMate NCP-3200RS, Thermo Scientific), and an HPLC sample loading pump (1200 Series, Agilent, Santa Clara, USA). Both SPE precolumn (150 µm i.d., 4 cm length) and LC column (50 µm i.d., 70 cm Self-Pack PicoFrit column, New Objective, Woburn, USA) were slurry-packed with 3 µm C18 packing material (300-Å pore size) (Phenomenex, Terrence, USA). Sample was fully injected into a 20 µL loop and loaded onto the SPE column using buffer A (0.1% formic acid in water) at a flow rate of 5 µL/min for 20 min. The concentrated sample was then separated at a flow rate of 150 nL/min and a 75 min gradient of 8-35% buffer B (0.1% formic acid in acetonitrile). The LC column was washed using 80% buffer B for 10 min and equilibrated using 2% buffer B for 20 min. Q Exactive Plus Orbitrap MS (Thermo

Scientific) was used to analyze the separated peptides. A 2.2 kV high voltage was applied at the ionization source to generate electrospray and ionize peptides. The ion transfer capillary was heated to 250 °C to desolvate droplets. The data dependent acquisition mode was employed to automatically trigger the precursor scan and the MS/MS scans. Precursors were scanned at a resolution of 35,000, an AGC target of 3×10^6, a maximum ion trap time of 100 ms. Top-10 precursors were isolated with an isolation window of 2, an AGC target of 2×10^5, a maximum ion injection time of 250 ms (for >300 cells, the AGC target of 2×10^5 and 100 ms ion injection time was used), and then fragmented by high energy collision with an energy level of 32%. A dynamic exclusion of 30 s was used to minimize repeated sequencing. MS/MS spectra were scanned at a resolution of 17,500.

Proteomics data analysis

The freely available open-source MaxQuant software was used for protein identification and quantification. The MS raw files were processed with MaxQuant (Version 1.5.1.11)[82,83] and MS/MS spectra were searched by Andromeda search engine against the against mouse UniProt database (fasta file dated April 12, 2017) (with the following parameters: tryptic peptides with 0-2 missed cleavage sites; 10 ppm of parent ion tolerance; 0.6 Da of fragment ion mass tolerance; variable modifications (methionine oxidation). Search results were processed with MaxQuant and filtered with a false discovery rate ≤1% at both protein and peptide levels. For label-free quantification, the match between runs (MBR) function was activated with a matching window of 0.4 min and the alignment window of 20 min. The quantitation results were extracted from MaxQuant outputs based on at least two valid values in one sample type by using Peruses (Version 1.5.8.3)[84]. Supplementary Data #2 contains processed global protein expression data.

Protein staining

All protein staining images are from the Human Protein Atlas and readily available at http://www.proteinatlas.org. Tissue in these images are from the intestine and labeled with the specific location including duodenum, small intestine, colon, or rectum [85,86].

Statistics and reproducibility

More than 200 mice were used to optimize and validate the flow sorting procedure and perform mRNA sequencing and proteomics. For proteomics three biological replicate were collected for each cell type, each biological replicate is treated as one sample during data analysis. These biological replicates are from independent mice and independent flow sorts. For mRNA-sequencing additional mice had to be used and pooled in order to isolate enough cells for sequencing, particularly for rarer cell types. In total we sequenced the following number of biological replicates (aka samples) per cell type stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, EEC=2. These biological replicates are from independent mice (sometimes sets of pooled mice) and independent flow sorts. Pooling of independent sorts was done as needed to ensure >2,500 cells for RNA preparation as described in the 'RNA Preparation and RNA-seq' method section. The number of independent mice for each of the biological replicates per cell type is as follows: stem=1,1,1; AbsPro=1,1,1; SecPDG=2,2,4,4; tuft=4,4,4,4,4; Ent=5,5,6,5,5; EEC=8,10.

No data exclusion were performed and no randomization or blinding methods were used in data analysis. Gene expression and mRNA processing bioinformatic packages (DESeq2, rMATS, and DaPars) were used for statistical analysis as specified in the appropriate methods section. GraphPad Prism (version 6.01) was used for additional analysis including: Figure 2.6b – Xbp1 splicing (unpaired two-sided t-test); standard deviation quantitation for mRNA expression graphs (Figures 2.6, 2.7 and throughout the Supplemental material; and Supplemental Figure 2.12 – linear best fit lines and R^2 values.

Data availability

Raw sequencing data (fastq) and processed data (counts files, rMATS, DaPars, and MAJIQ data) are available for download on GEO (GSE143915). The proteomics raw data sets and identified proteins groups lists generated from Maxqaunt have been deposited in Japan ProteOme STandard Repository [87] (jPOST; https://repository.jpostdb.org/). The accession number is JPST000853 for jPOST and PXD019351 for ProteomeXchange.

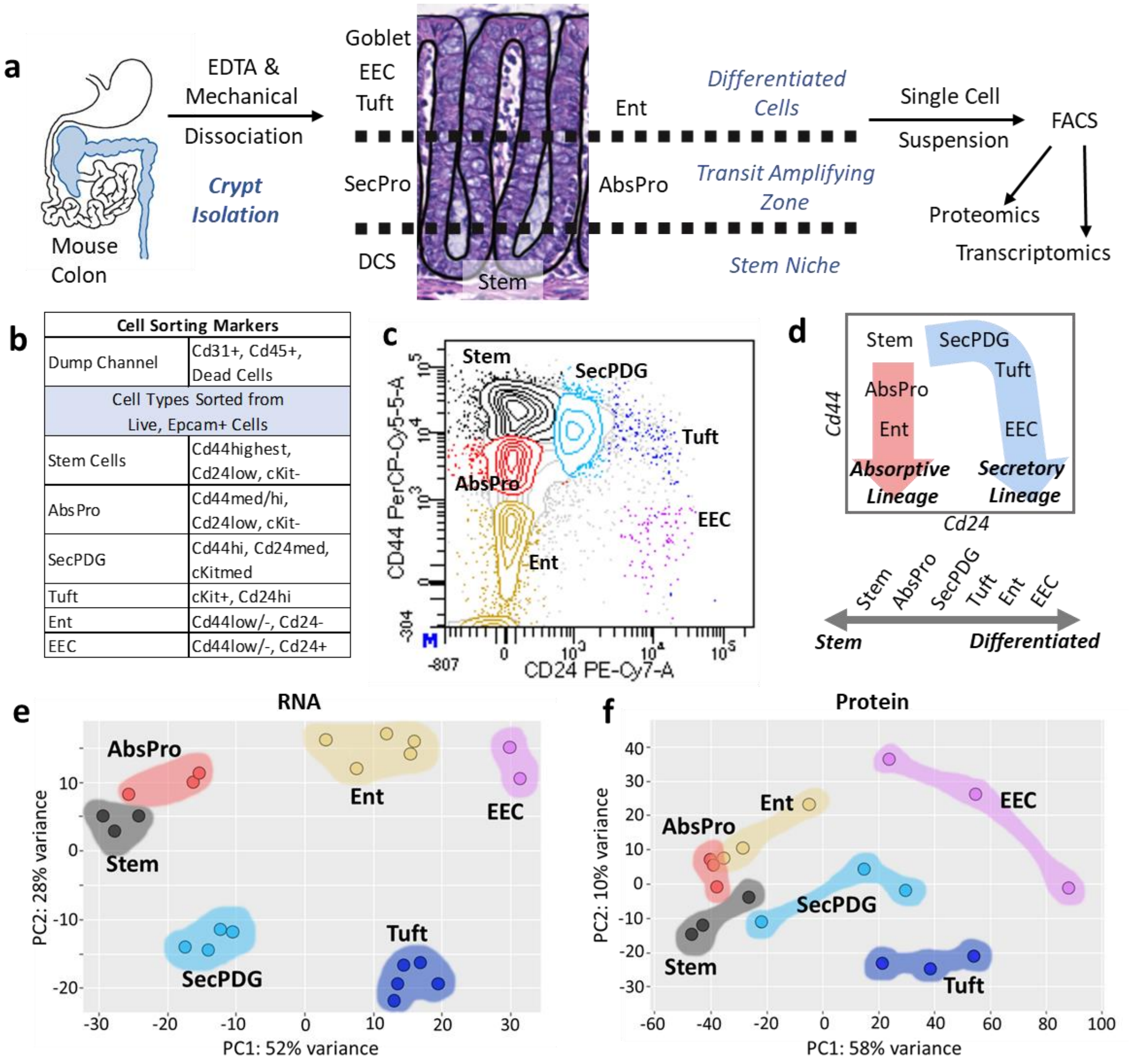


Figure 2.1: ***A novel flow sorting protocol that purifies six murine colon crypt cell populations.***

a Schema representing methods used for single cell isolation and, **b** sorting markers used for Flow Activated Cell Sorting (FACS). **c** FACS plot for membrane biomarkers Cd44 and Cd24 show six distinct populations including stem, absorptive progenitor (AbsPro), secretory progenitor/deep crypt secretory cells/goblet (SecPDG), tuft cells, enterocytes (Ent), and enteroendocrine (EEC). **d** Crypt cell populations diagramed in the FACS plot by lineage (Secretory and Absorptive) and on a scale from stem to differentiated. **e-f** Principle component analysis of bulk RNA-seq data (**e**, with top 100 genes) and proteomics data (**f**) from the six crypt cell populations. For protein biological replicates for each cell type n=3 samples, for RNA, biological replicate numbers are as follows: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2.

(Figure adapted from Fig. 1 of Habowski et al., 2020; Communications Biology)

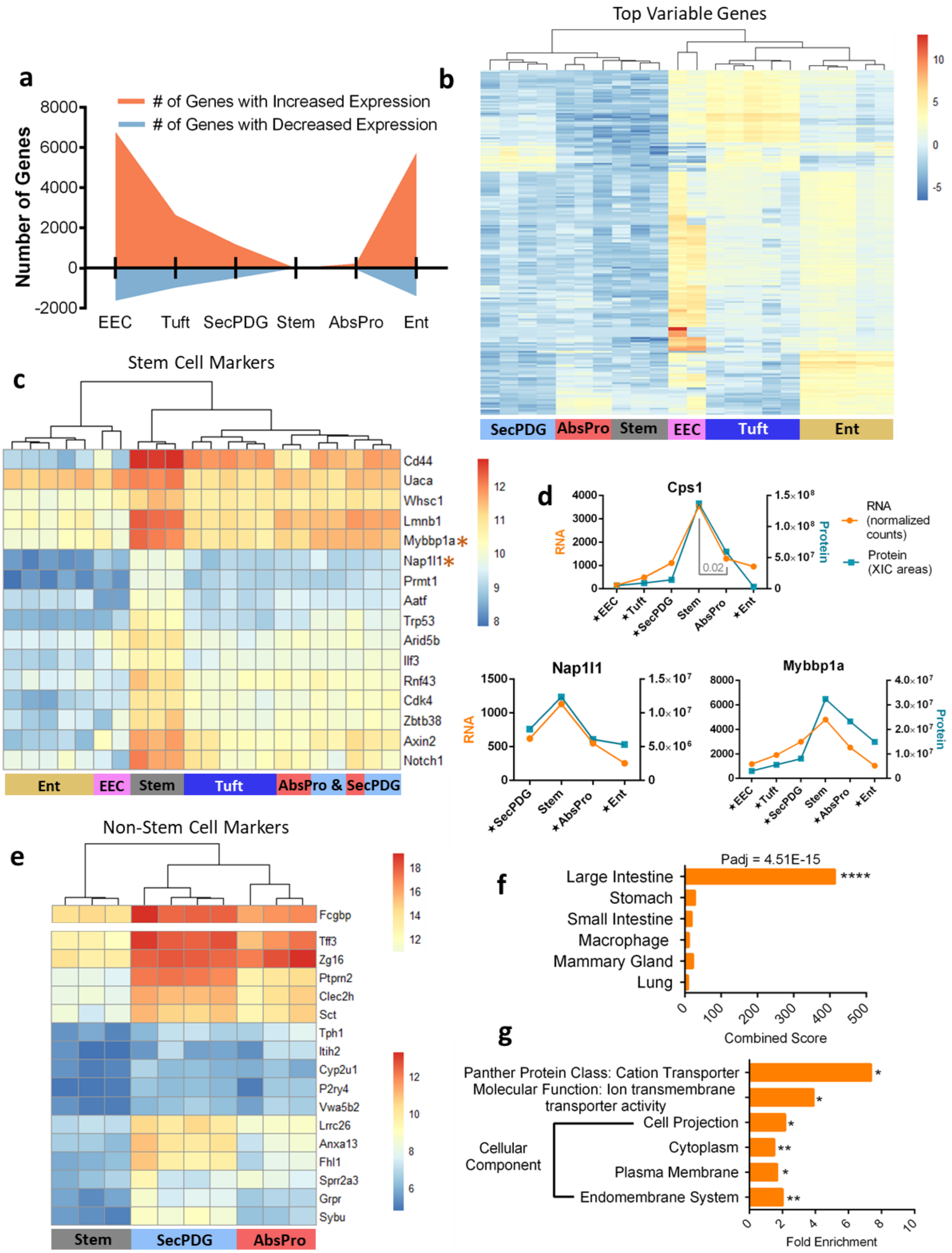
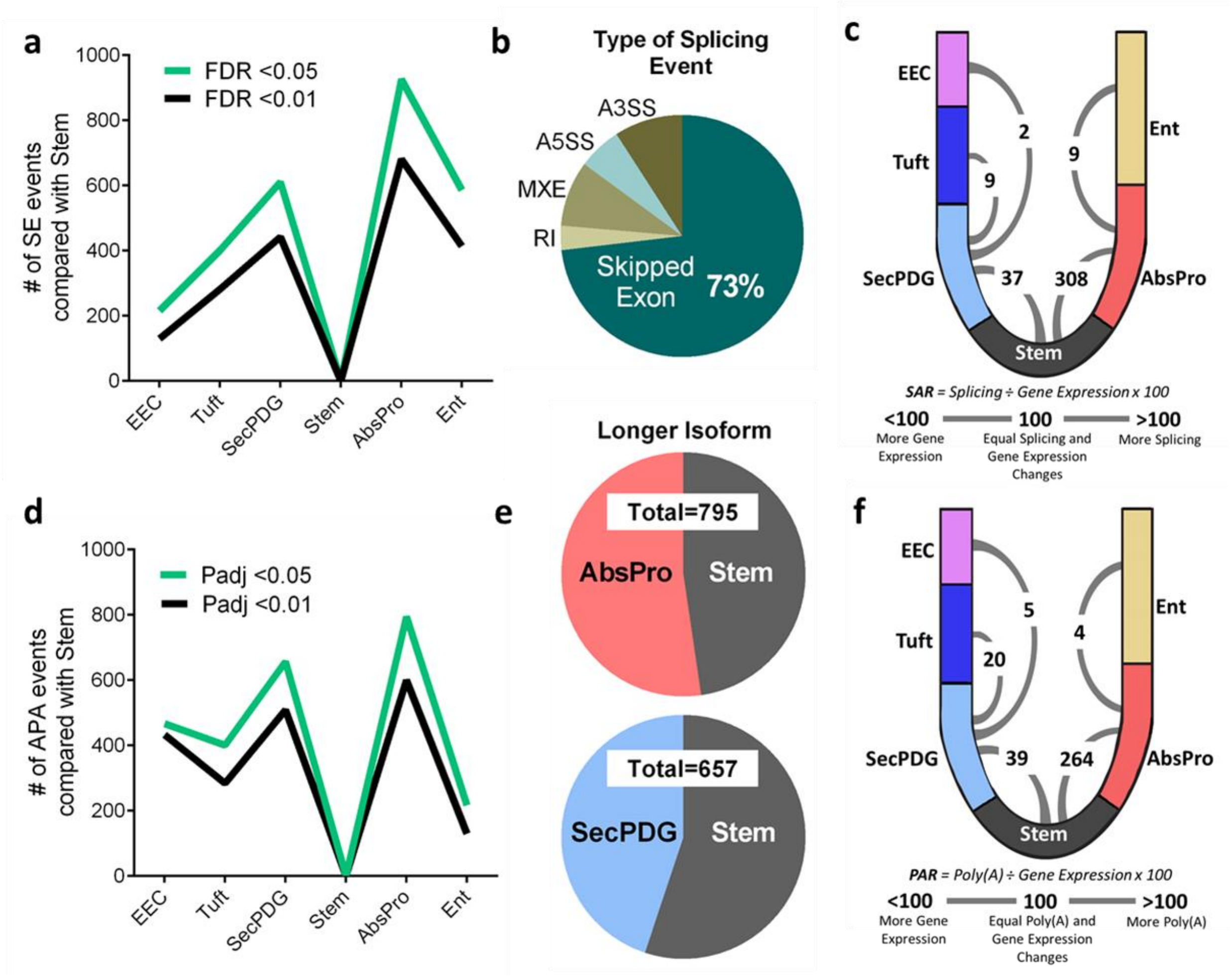

a
# of Genes with Increased Expression
# of Genes with Decreased Expression
Number of Genes
8000
6000
4000
2000
0
-2000
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
b
Top Variable Genes
10
5
0
-5
SecPDG
AbsPro
Stem
EEC
Tuft
Ent
c
Stem Cell Markers
Cd44
Uaca
Whsc1
Lmnb1
Mybbp1a*
Nap1l1*
Prmt1
Aatf
Trp53
Arid5b
Ilf3
Rnf43
Cdk4
Zbtb38
Axin2
Notch1
12
11
10
9
8
Ent
EEC
Stem
Tuft
AbsPro & SecPDG
d
Cps1
RNA
Protein
4000
3000
2000
1000
0
1.5×10⁸
1.0×10⁸
5.0×10⁷
0
0.02
⋆EEC
⋆Tuft
⋆SecPDG
Stem
AbsPro
⋆Ent
RNA (normalized counts)
Protein (XIC areas)
Nap1l1
1500
1000
500
0
1.5×10⁷
1.0×10⁷
5.0×10⁶
0
⋆SecPDG
Stem
⋆AbsPro
⋆Ent
Mybbp1a
8000
6000
4000
2000
0
4.0×10⁷
3.0×10⁷
2.0×10⁷
1.0×10⁷
0
⋆EEC
⋆Tuft
⋆SecPDG
Stem
⋆AbsPro
⋆Ent
e
Non-Stem Cell Markers
Fcgbp
Tff3
Zg16
Ptprn2
Clec2h
Sct
Tph1
Itih2
Cyp2u1
P2ry4
Vwa5b2
Lrrc26
Anxa13
Fhl1
Sprr2a3
Grpr
Sybu
18
16
14
12
12
10
8
6
Stem
SecPDG
AbsPro
f
Padj = 4.51E-15
Large Intestine ****
Stomach
Small Intestine
Macrophage
Mammary Gland
Lung
0
100
200
300
400
500
Combined Score
g
Panther Protein Class: Cation Transporter *
Molecular Function: Ion transmembrane transporter activity *
Cellular Component
Cell Projection *
Cytoplasm **
Plasma Membrane *
Endomembrane System **
0
2
4
6
8
10
Fold Enrichment

Figure 2.2: ***Characterization of intestinal stemness based on differential gene expression.***

a The number of genes that significantly change gene expression (mRNA level) between non-stem cells and stem cells; orange are the number of genes that increase expression and blue are the number of genes that decrease expression compared with stem (padj < 0.01 + minimum mean 50 counts). **b** Auto-scaled heatmap showing gene expression and unsupervised clustering of the top 200 most variably expressed genes. **c** Gene expression heatmap and unsupervised clustering of n=16 genes that are significantly enriched in stem cells compared to all non-stem cells (padj < 0.01 + minimum mean 50 counts). **d** Examples of stem enriched markers showing both mRNA expression paired with protein expression. Star annotation by cell type symbolizes significant differential mRNA expression compared to stem (padj < 0.01). **e** Unsupervised clustering of genes that significantly increase in expression from stem to both SecPDG and AbsPro (8-fold change cut-off, padj < 0.01 + minimum mean count 50 counts). Upper panel: The highly expressed *Fcgbp* gene is reported on a separate color scale. **f, g** Enrichr (Mouse Gene Atlas) and Panther (Cellular Component analysis, Molecular Function, and Panther Protein Class) gene ontology analysis of n=107 genes that are significantly higher in expression in all non-stem cell types compared to stem cells (gene list in Table 2.1). Significance is defined by: * < 0.05 FDR, ** < 0.01 FDR, *** < 0.005 FDR, **** < 0.001 FDR and analysis was performed with the following biological replicate numbers: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2.

(Figure adapted from Fig. 2 of Habowski et al., 2020; Communications Biology)



a
1000
800
600
400
200
0
FDR <0.05
FDR <0.01
# of SE events compared with Stem
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
b
Type of Splicing Event
A3SS
A5SS
MXE
RI
Skipped Exon
73%
c
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
2
9
9
37
308
SAR = Splicing ÷ Gene Expression x 100
<100
More Gene Expression
100
Equal Splicing and Gene Expression Changes
>100
More Splicing
d
1000
800
600
400
200
0
Padj <0.05
Padj <0.01
# of APA events compared with Stem
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
e
Longer Isoform
Total=795
AbsPro
Stem
Total=657
SecPDG
Stem
f
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
5
20
4
39
264
PAR = Poly(A) ÷ Gene Expression x 100
<100
More Gene Expression
100
Equal Poly(A) and Gene Expression Changes
>100
More Poly(A)

Figure 2.3: ***A burst of alternative mRNA processing activity during loss of stemness.***

a Alternative splicing analysis with rMATS determined the abundance of skipped exon events in non-stem cell types compared to stem cells. **b** Breakdown of average percentages of rMATS splicing changes (events) detected between stem and all non-stem cell types by type of event (SE = Skipped Exon, RI = Retained Intron, MXE = Mutually Exclusive Exon, A5SS = Alt 5 Splice Site, A3SS = Alt 3 Splice Site) showing predominance of SE (73%) (FDR < 0.05). **c** Crypt diagram illustrating cell types in the secretory lineage (SecPDG, tuft, EEC) versus absorptive lineage (AbsPro, Ent). A numeric SAR (Splicing Abundance Ratio = # of significant alternative splicing changes ÷ # of significant gene expression changes x 100) arc indicates the number of splicing changes relative to gene expression between stem, progenitors, and differentiated cells. **d** Alternative polyadenylation (APA) analysis with DaPars quantitated the number of APA changes (events) in non-stem cells compared with stem. **e** APA events characterized by which cell type has the longer 3'UTR isoform for each polyadenylated mRNA in stem versus AbsPro (top) and stem versus SecPDG (bottom) (padj < 0.05). **f** Crypt diagram illustrating PAR (Polyadenylation Abundance Ratio = # of significant alternative polyadenylation changes ÷ # of significant gene expression changes x 100) comparing polyadenylation changes to gene expression between stem, progenitors, and differentiated cells. Splicing and polyadenylation analysis was performed with the following biological replicate number of mRNA-seq samples: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2.

(Figure adapted from Fig. 3 of Habowski et al., 2020; Communications Biology)

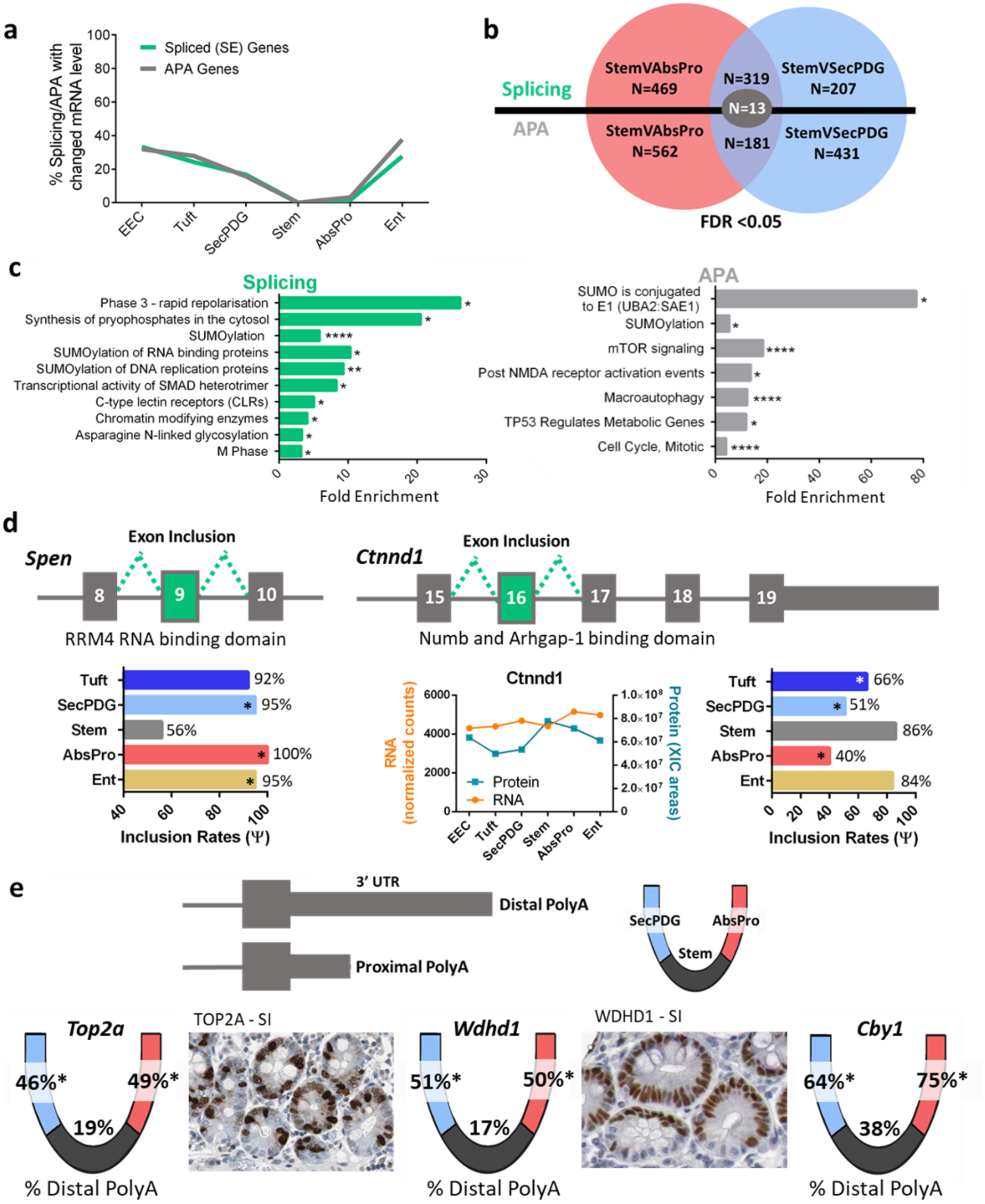

a
% Splicing/APA with changed mRNA level
100
80
60
40
20
0
Spliced (SE) Genes
APA Genes
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
b
Splicing
APA
StemVAbsPro N=469
N=319
StemVSecPDG N=207
N=13
StemVAbsPro N=562
N=181
StemVSecPDG N=431
FDR <0.05
c
Splicing
Phase 3 - rapid repolarisation *
Synthesis of pryophosphates in the cytosol *
SUMOylation ****
SUMOylation of RNA binding proteins *
SUMOylation of DNA replication proteins **
Transcriptional activity of SMAD heterotrimer *
C-type lectin receptors (CLRs) *
Chromatin modifying enzymes *
Asparagine N-linked glycosylation *
M Phase *
0
10
20
30
Fold Enrichment
APA
SUMO is conjugated to E1 (UBA2:SAE1) *
SUMOylation *
mTOR signaling ****
Post NMDA receptor activation events *
Macroautophagy ****
TP53 Regulates Metabolic Genes *
Cell Cycle, Mitotic ****
0
20
40
60
80
Fold Enrichment
d
Spen
Exon Inclusion
8
9
10
RRM4 RNA binding domain
Ctnnd1
Exon Inclusion
15
16
17
18
19
Numb and Arhgap-1 binding domain
Tuft 92%
SecPDG * 95%
Stem 56%
AbsPro * 100%
Ent * 95%
40
60
80
100
Inclusion Rates (Ψ)
Ctnnd1
RNA (normalized counts)
6000
4000
2000
0
Protein (XIC areas)
1.0×10^8
8.0×10^7
6.0×10^7
4.0×10^7
2.0×10^7
0
Protein
RNA
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
Tuft * 66%
SecPDG * 51%
Stem 86%
AbsPro * 40%
Ent 84%
0
20
40
60
80
100
Inclusion Rates (Ψ)
e
3' UTR
Distal PolyA
Proximal PolyA
SecPDG
AbsPro
Stem
Top2a
46%*
49%*
19%
% Distal PolyA
TOP2A - SI
Wdhd1
51%*
50%*
17%
% Distal PolyA
WDHD1 - SI
Cby1
64%*
75%*
38%
% Distal PolyA

Figure 2.4: ***Alternative splicing and polyadenylation changes that occur during intestinal crypt loss of stemness.***

a Percentage of alternatively spliced (FDR < 0.05) or polyadenylated genes (padj < 0.05) that also change gene expression (padj < 0.01) compared to stem. **b** Venn diagram overlap of APA and alternatively spliced genes between stem versus AbsPro and stem versus SecPDG (FDR < 0.05). n=13 genes were both APA and alternatively spliced differently in stem cells versus SecPDG and stem cells versus AbsPro (Supplementary Figure 2.21a). **c** Gene ontology (Reactome Pathway) analysis of the commonly spliced genes when comparing stem versus AbsPro and stem versus SecPDG (rMATS; n=332 (319+13) genes) and common APA genes (DaPars, n=194 (181+13) genes); FDR < 0.05. Sumoylation and cell cycle ontologies of alternatively processed genes are common to both splicing and APA changes (Supplementary Figure 2.21b,c). **d** Two examples of alternatively spliced genes, *Spen* and *Ctnnd1*, that are differentially processed in stem cells versus progenitor cells. The exon inclusion rate for each event is shown in the bar graph. Ctnnd1 protein was detected by MS and the abundance compared to mRNA is displayed. **e** Three examples of genes with significant changes in alternative polyadenylation choice: *Top2a*, *Wdhd1*, and *Cyb1*. Overlaid on the crypt-base diagram are the percentage of distal polyA usage for each of the genes in the three cryptal cell compartments: stem, AbsPro and SecPDG. Human protein atlas images show strong staining patterns of TOP2A and WDHD1 in the transit amplifying zone but a lack of staining in the stem cell niche despite the fact that *Top2a* mRNA levels are elevated in stem cells compared to progenitor cells and *Wdhd1* mRNA levels are the same among the cell types (Supplementary Figure 2.22). Additional immunohistochemistry images of human intestine are provided in Supplementary Figure 2.23. Significance is defined by: * < 0.05 FDR, ** < 0.01 FDR, *** < 0.005 FDR, **** < 0.001 FDR.

(Figure adapted from Fig. 4 of Habowski et al., 2020; Communications Biology)

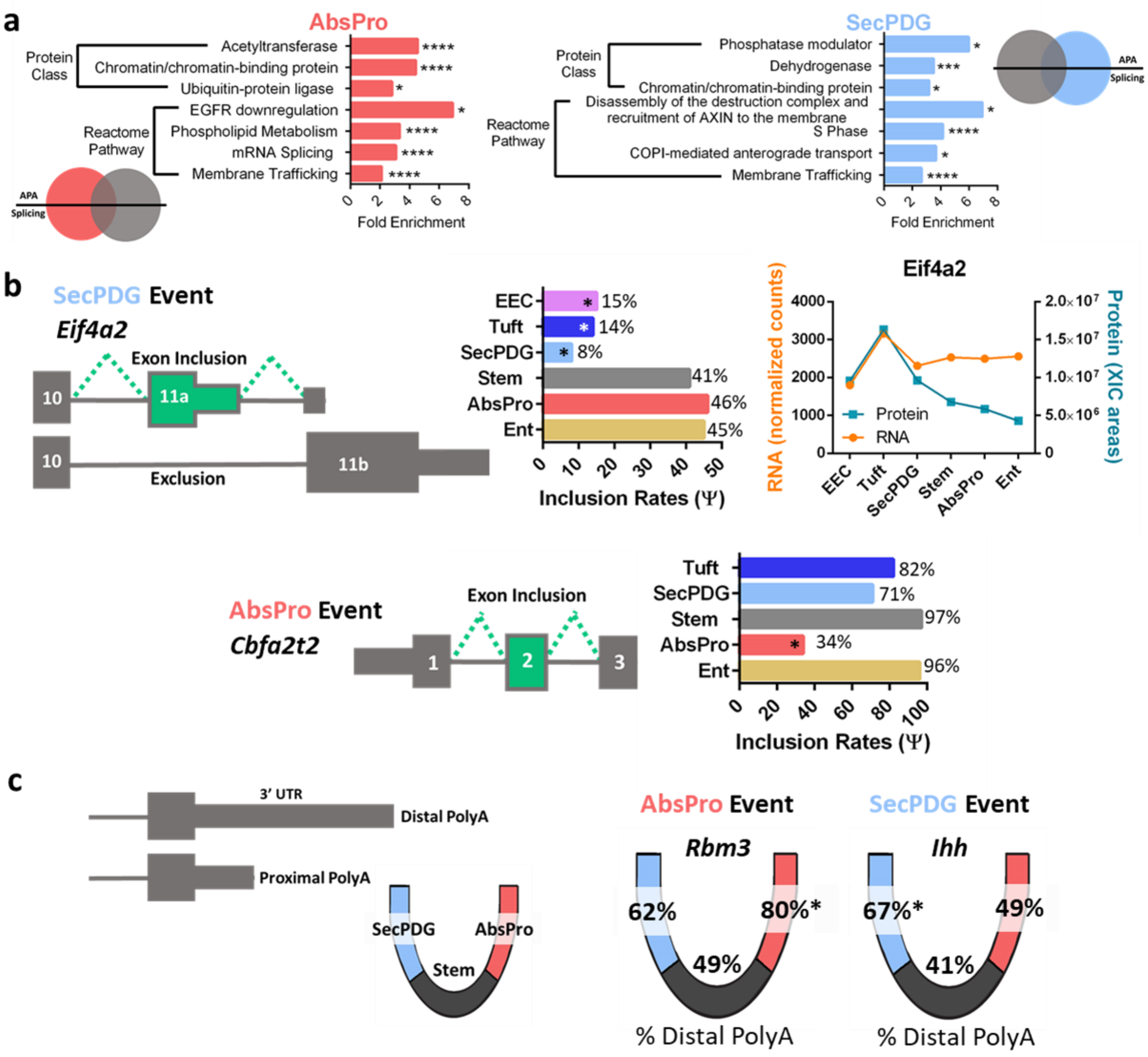


Figure 2.5: ***mRNA processing contributes to lineage commitment.***

a Gene ontology analysis (Panther) was performed on alternatively spliced (FDR < 0.05) and polyadenylated genes (padj < 0.05) with processing events specific to either AbsPro (red) or SecPDG (blue). Events are unique and not part of the overlap shown in Figure 2.4b. **b** Two examples of alternatively spliced genes: *Eif4a2*, specific to SecPDG, and *Cbfa2t2*, specific to AbsPro. The exon inclusion rate for each event is shown in the bar graph. Eif4a2 was detected by MS and the abundance compared to mRNA is displayed (Cbfa2t2 protein was not detected). **c** Two examples of alternatively polyadenylated genes: *Rbm3*, AbsPro specific, and *Ihh*, specific to SecPDG. The percentage of distal polyA usage for each of the events is overlaid on the crypt-base diagram. Significance is defined by: * < 0.05 FDR, ** < 0.01 FDR, *** < 0.005 FDR, **** < 0.001 FDR.

(Figure adapted from Fig. 5 of Habowski et al., 2020; Communications Biology)

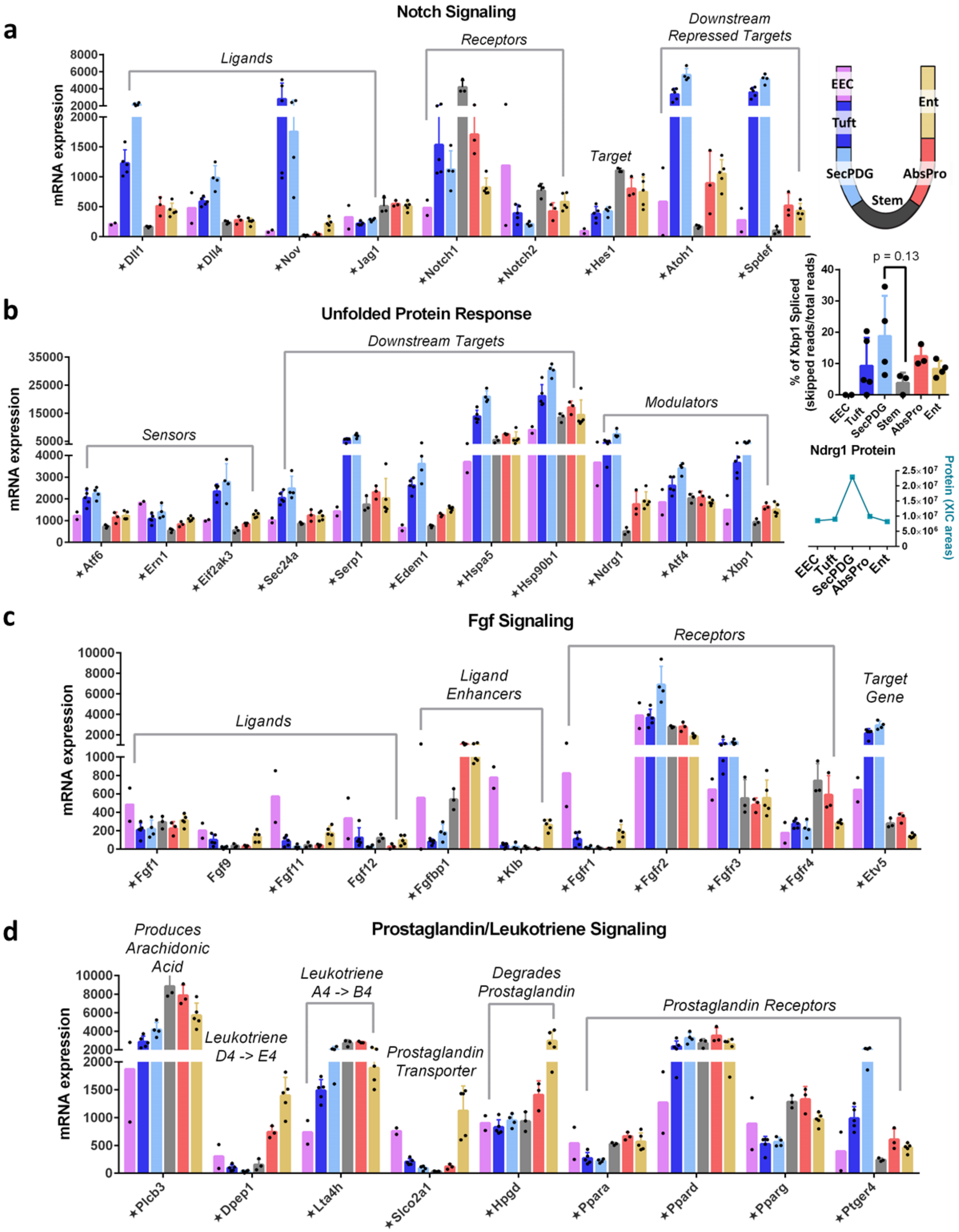

a
Notch Signaling
mRNA expression
Ligands
Receptors
Target
Downstream Repressed Targets
★Dll1
★Dll4
★Nov
★Jag1
★Notch1
★Notch2
★Hes1
★Atoh1
★Spdef
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
b
Unfolded Protein Response
Sensors
Downstream Targets
Modulators
★Atf6
★Ern1
★Eif2ak3
★Sec24a
★Serp1
★Edem1
★Hspa5
★Hsp90b1
★Ndrg1
★Atf4
★Xbp1
p = 0.13
% of Xbp1 Spliced (skipped reads/total reads)
Ndrg1 Protein
Protein (XIC areas)
c
Fgf Signaling
Ligands
Ligand Enhancers
Receptors
Target Gene
★Fgf1
Fgf9
★Fgf11
Fgf12
★Fgfbp1
★Klb
★Fgfr1
★Fgfr2
★Fgfr3
★Fgfr4
★Etv5
d
Prostaglandin/Leukotriene Signaling
Produces Arachidonic Acid
Leukotriene D4 -> E4
Leukotriene A4 -> B4
Prostaglandin Transporter
Degrades Prostaglandin
Prostaglandin Receptors
★Plcb3
★Dpep1
★Lta4h
★Slco2a1
★Hpgd
★Ppara
★Ppard
★Pparg
★Ptger4

Figure 2.6: ***Lineage commitment to secretory and absorptive lineages are influenced by signaling pathways.***

a mRNA expression of Notch ligands (elevated in SecPDG/tuft), receptors (elevated in stem), and downstream targets. **b** mRNA expression of Unfolded Protein Response (UPR) components including sensors, modulators, and downstream targets are elevated in SecPDG. Xbp1 activation, determined by a cytoplasmic splicing event, is elevated in SecPDG (inset - unpaired two-sided t-test). Modulator Ndrg1 was detected via MS and shows protein is elevated in SecPDG (inset graph) consistent with mRNA expression (Ndrg1 protein was not detected in Stem). **c** mRNA expression of Fgf signaling components including ligands (showing some EEC expression), ligand enhancer *Fgfbp1* expressed in AbsPro and Ent, receptors (well expressed in all cells, highest in SecPDG and tuft) and target gene (highly expressed in SecPDG and tuft). **d** Prostaglandin and Leukotriene precursors and final products are produced by tuft cells (Supplementary Figure 2.25), but absorptive lineage cells AbsPro and Ent, might contribute to the production (*Plcb3*, *Dpep1,* and *Lta4h*) and degradation (*Slco2a1* and *Hpgd*) of prostaglandin signals. Prostaglandin receptor *Ptger4* is enriched in SecPDG while alternate receptors *Pparg* and *Ppara* are enriched in the absorptive lineage. Star annotation by gene name symbolizes significant differential mRNA expression in at least one cell type compared to stem (padj < 0.01). mRNA expression values are normalized counts and error bars are standard deviation. mRNA differential expression analysis was performed with the following biological replicate numbers: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2.

(Figure adapted from Fig. 6 of Habowski et al., 2020; Communications Biology)

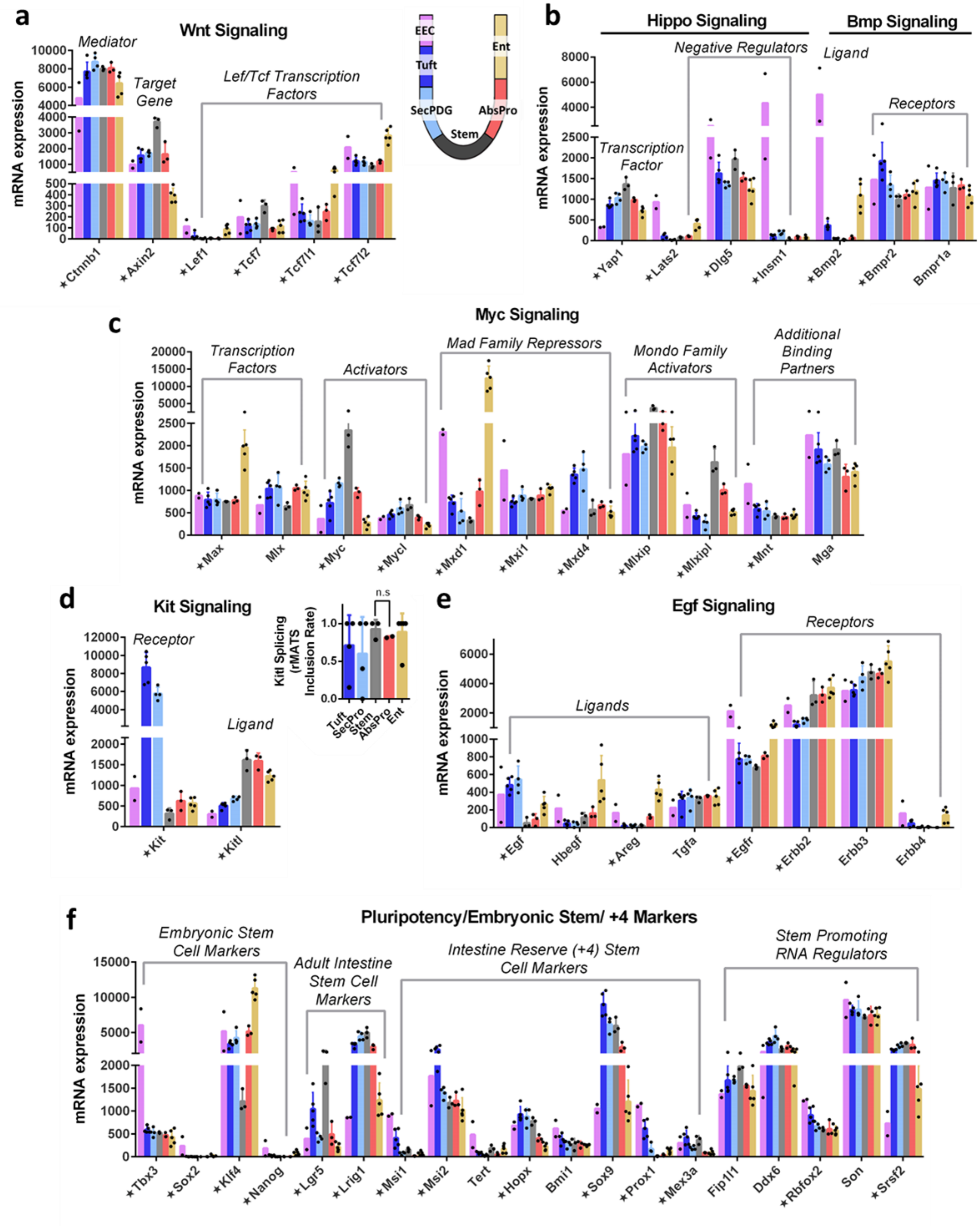

a
Wnt Signaling
Mediator
Target Gene
Lef/Tcf Transcription Factors
mRNA expression
10000
8000
6000
4000
4000
3000
2000
1000
500
400
300
200
100
0
★Ctnnb1
★Axin2
★Lef1
★Tcf7
★Tcf7l1
★Tcf7l2
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
b
Hippo Signaling
Bmp Signaling
Negative Regulators
Ligand
Receptors
Transcription Factor
8000
6000
4000
2500
2000
1500
1000
500
0
★Yap1
★Lats2
★Dlg5
★Insm1
★Bmp2
★Bmpr2
Bmpr1a
c
Myc Signaling
Transcription Factors
Activators
Mad Family Repressors
Mondo Family Activators
Additional Binding Partners
20000
15000
10000
5000
2500
2000
1500
1000
500
0
★Max
Mlx
★Myc
★Mycl
★Mxd1
★Mxi1
★Mxd4
★Mlxip
★Mlxipl
★Mnt
Mga
d
Kit Signaling
Receptor
Ligand
12000
10000
8000
6000
4000
2000
1500
1000
500
0
★Kit
★Kitl
Kitl Splicing (rMATS Inclusion Rate)
n.s
1.0
0.5
0.0
Tuft
SecPro
Stem
AbsPro
Ent
e
Egf Signaling
Receptors
Ligands
6000
4000
2000
1000
800
600
400
200
0
★Egf
Hbegf
★Areg
Tgfa
★Egfr
★Erbb2
Erbb3
Erbb4
f
Pluripotency/Embryonic Stem/ +4 Markers
Embryonic Stem Cell Markers
Adult Intestine Stem Cell Markers
Intestine Reserve (+4) Stem Cell Markers
Stem Promoting RNA Regulators
15000
10000
5000
2000
1500
1000
500
0
★Tbx3
★Sox2
★Klf4
★Nanog
★Lgr5
★Lrig1
★Msi1
★Msi2
Tert
★Hopx
Bmi1
★Sox9
★Prox1
★Mex3a
Fip1l1
Ddx6
★Rbfox2
Son
★Srsf2

Figure 2.7: ***Fate commitment and signaling characteristics of mature crypt cells.***

a mRNA expression of key Wnt signaling factors including Lef/Tcf transcription factors, β-catenin mediator (Supplementary Figure 2.27a), and downstream target gene *Axin2*. **b** mRNA expression of Hippo and Bmp signaling components. **c** mRNA expression of Myc signaling components including strong expression of repressive transcription factors. **d** mRNA expression of Kit signaling components including ligand (high in stem and absorptive), receptor (high in SecPDG and tuft). Splicing rates of *Kitl* in crypt cell types showed predominance of exon 6 inclusion which encodes a protease site for protease release and secretion of Kit ligand (inset). **e** mRNA expression of Egf signaling components including some epithelial ligand expression and receptors expressed in all cells. *Egfr* is elevated in EEC (Supplementary Figure 2.28), whereas the lowest levels of *Erbb2* is in SecPDG and tuft. **f** mRNA expression of stem promoting markers including classic adult intestine stem cell markers, embryonic stem cell markers, intestinal reserve (+4) stem cell markers, and RNA regulators showing enriched expression in some differentiated cell types. Star annotation by gene name symbolizes significant differential mRNA expression in at least one cell type compared to stem ($padj < 0.01$). mRNA expression values are normalized counts and error bars are standard deviation. mRNA differential expression analysis was performed with the following biological replicate numbers: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2.

(Figure adapted from Fig. 7 of Habowski et al., 2020; Communications Biology)

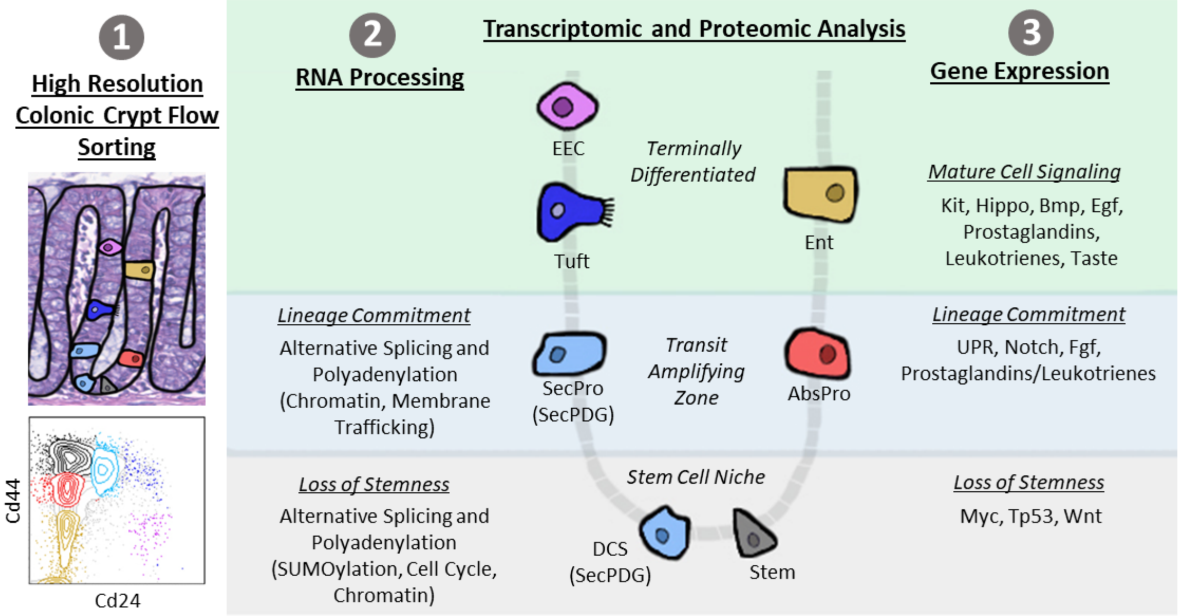
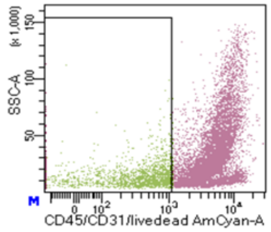


Figure 2.8: ***Summary of transcriptome and proteome changes in colon crypt homeostasis.***

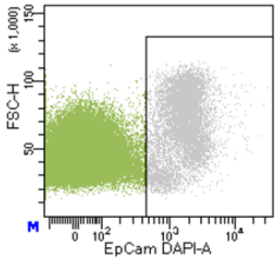
Our findings encompass three main themes: (1) Methodology – High resolution colonic crypt flow sorting to isolate stem cells, SecPDG (secretory progenitors/deep crypt secretory cells), AbsPro (absorptive progenitors), tuft cells, enterocytes (Ent), and enteroendocrine cells (EEC), (2) RNA processing (splicing and polyadenylation) influences the transcriptome most during loss of stemness and lineage commitment, and (3) gene expression changes influencing lineage commitment and mature cell signaling.

(Figure adapted from Fig. 8 of Habowski et al., 2020; Communications Biology)

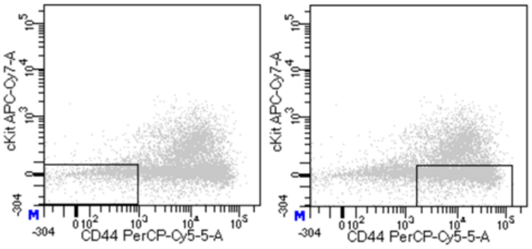
1. Select single, live cells based on forward/side scatter.
2. Dump Cd45+/Cd31+/Dead cells



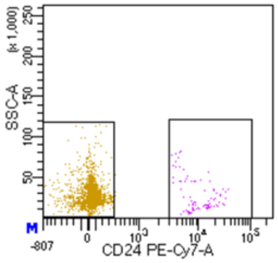
3. Select Epcam+



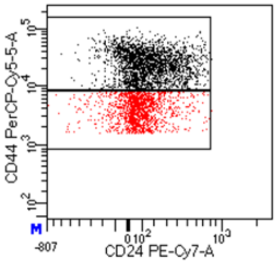
4. Gate two populations in Epcam+: (1) cKit-/Cd44low/- and (2) cKit-/Cd44+



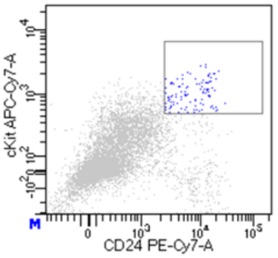
5. cKit-/Cd44low/- is gated on Cd24 = Cd24- (**ENT**) and Cd24+ (**EEC**)



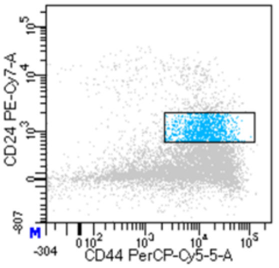
5. cKit-/Cd44+ is gated on Cd44/Cd24 = Cd44highest(**STEM**) and Cd44med/hi (**ABSPRO**)



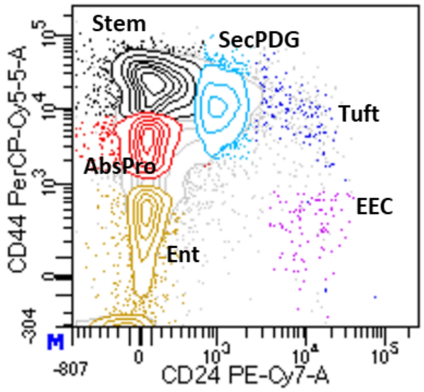
6. Epcam+ gated on cKit+/Cd24+ (**TUFT**)



7. Epcam+ gated on Cd44hi/Cd24med (**SECPDG**)



8. Resulting Populations



Relative Abundance of Cell Populations

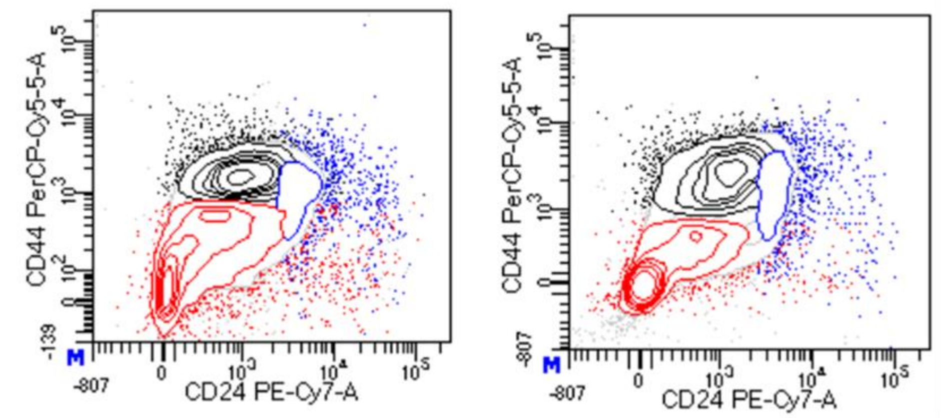
	Average % of Epcam+	Standard Deviation
Stem	22.1%	3.9%
AbsPro	15.6%	2.5%
SecPDG	9.0%	4.8%
Ent	19.7%	5.2%
Tuft	0.8%	0.5%
EEC	0.7%	0.4%

n = 10 independent sorts

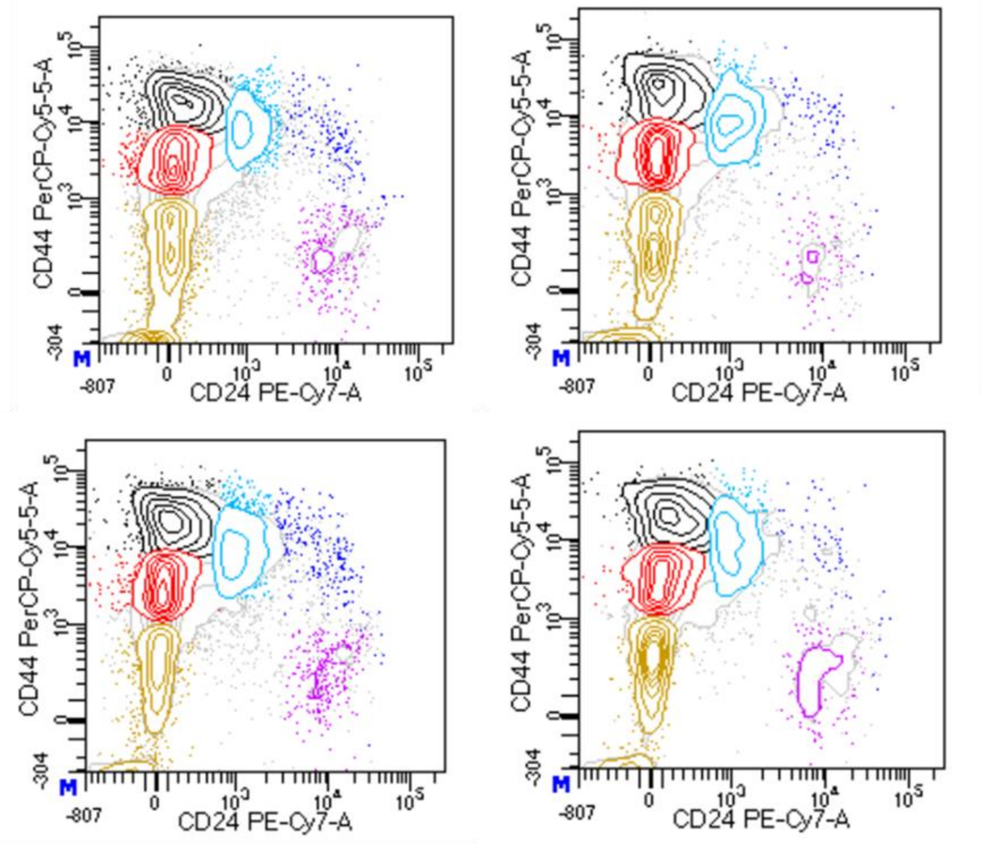
Supplementary Figure 2.1: ***FACS gating strategies that define six colon crypt cell populations.***

As a first step, standard gating is performed to select single, live cells based on forward and side scatter (Step 1). A dump channel then removes dead cells along with immune cells (Cd45+) and endothelial cells (Cd31+) (Step 2). Epcam+ cells (Step 3) are then gated using Cd44, Cd24, and cKit to isolate six distinct populations (Steps 4-7). The resulting populations are enterocytes (Ent), enteroendocrine (EEC; predominantly enterochromaffin cells), stem cells, absorptive progenitors (AbsPro), tuft, and SecPDG (a mixed population of secretory progenitors, deep crypt secretory cells, with a minor contribution from goblet cells) (Step 8). Relative average abundance and standard deviation of these populations in n=11 independent sorts are shown as a percentage of Epcam+ cells. Note: Some Epcam+ cells are not clearly gated into one of these cell populations and thus the sum of the six populations does not add to 100%.

(Figure adapted from Supplementary Fig. 1 of Habowski et al., 2020; Communications Biology)



a Dissociation with TrypLE

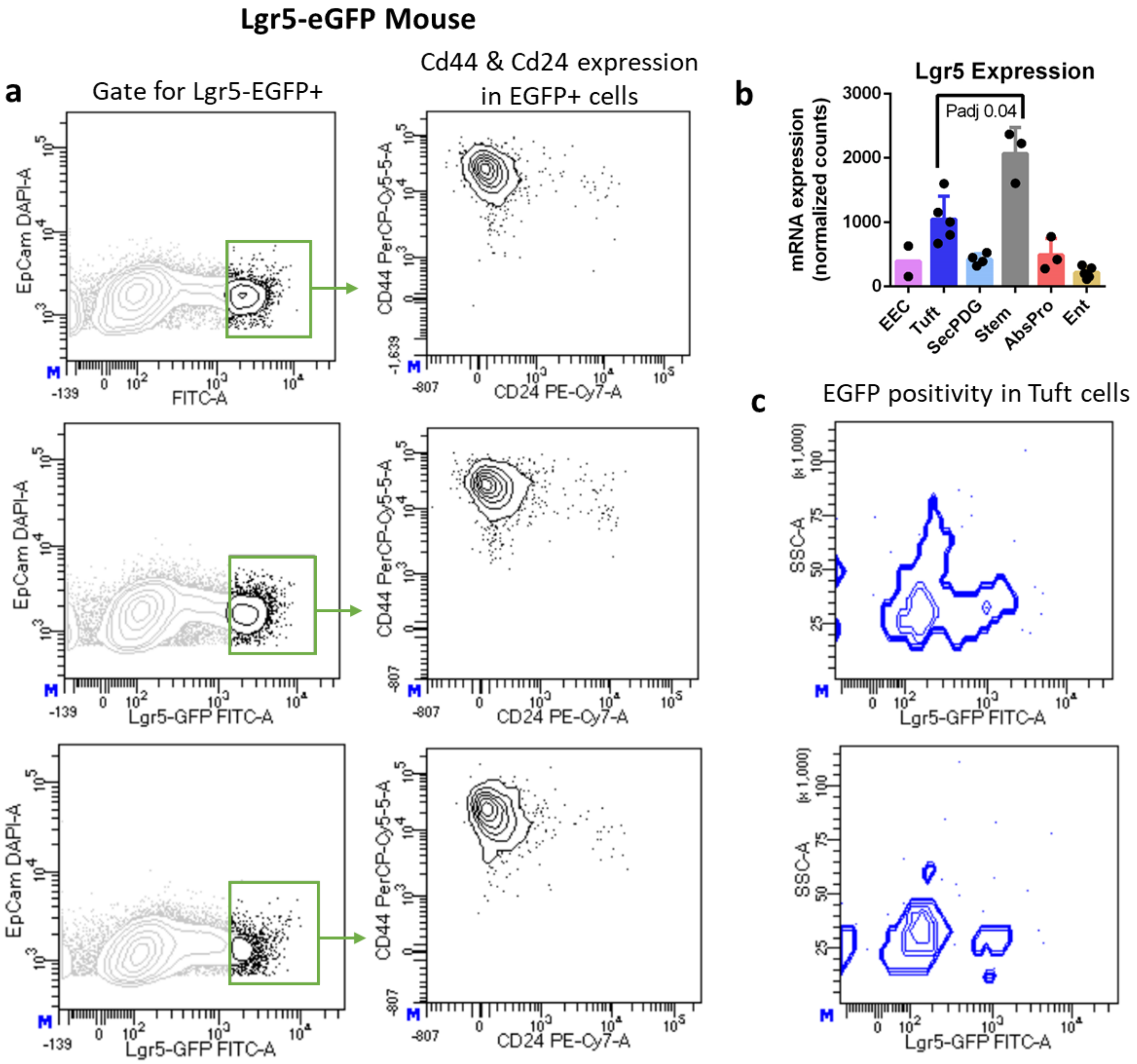


b Dissociation with **NO** TrypLE (EDTA only)

Supplementary Figure 2.2: ***FACS plots of colon crypts dissociated with or without TrypLE protease treatment.***

a When the TrypLE cocktail is used during intestine dissociation, FACS detects decreased Cd44 surface expression and the plots show a compressed population resolution compared to **b** no TrypLE. Each plot is from one mouse and is a representative image.

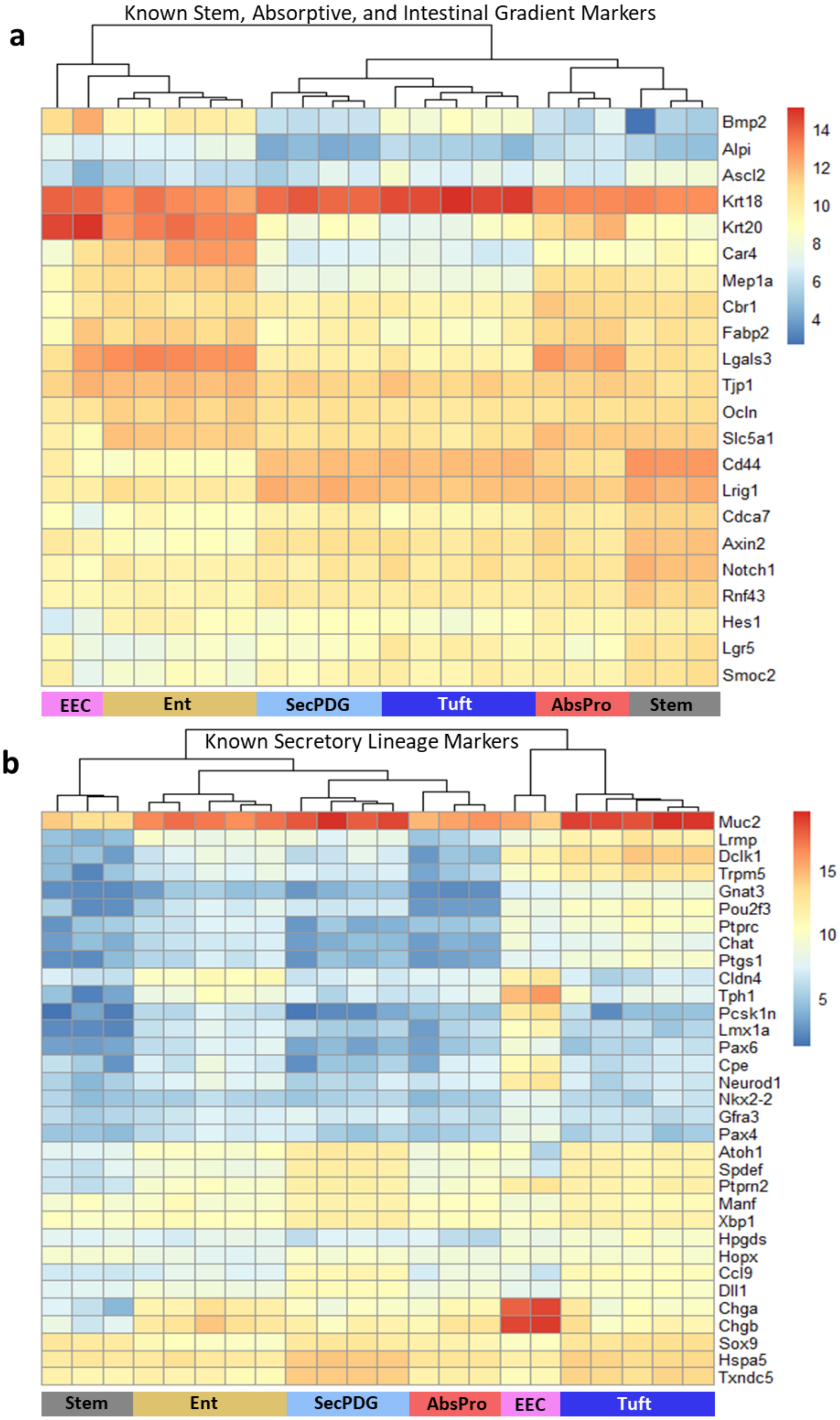
(Figure adapted from Supplementary Fig. 2 of Habowski et al., 2020; Communications Biology)



Supplementary Figure 2.3: ***Validation of sorted stem cell population with Lgr5-EGFP-IRES-creERT2 mice.***

a EGFP expression, used for labeling intestinal crypt stem cells via Lgr5-locus-directed transcription, is detected by FITC fluorescence and gated for +/high population. These cells are then displayed on a FACS plot with Cd44 and Cd24. The complete correlation between the GFP^{hi} stem cells and the Cd44^{hi}/Cd24^{lo} signature demonstrates a direct overlap of the two populations and confirms that Cd44^{hi}/Cd24^{lo} cells are colon crypt stem cells. n=3 independent mice/sorts. **b** mRNA expression of *Lgr5* shows the highest expression in the stem cell population, followed by only 2-fold lower in tuft cells and then 4-fold lower expression in AbsPro and SecPDG populations. mRNA differential expression analysis was performed with the following biological replicate numbers: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2. **c** Lgr5-EGFP (FITC) expression in tuft cells reveals positivity confirming that tuft cells express Lgr5 at detectable levels.

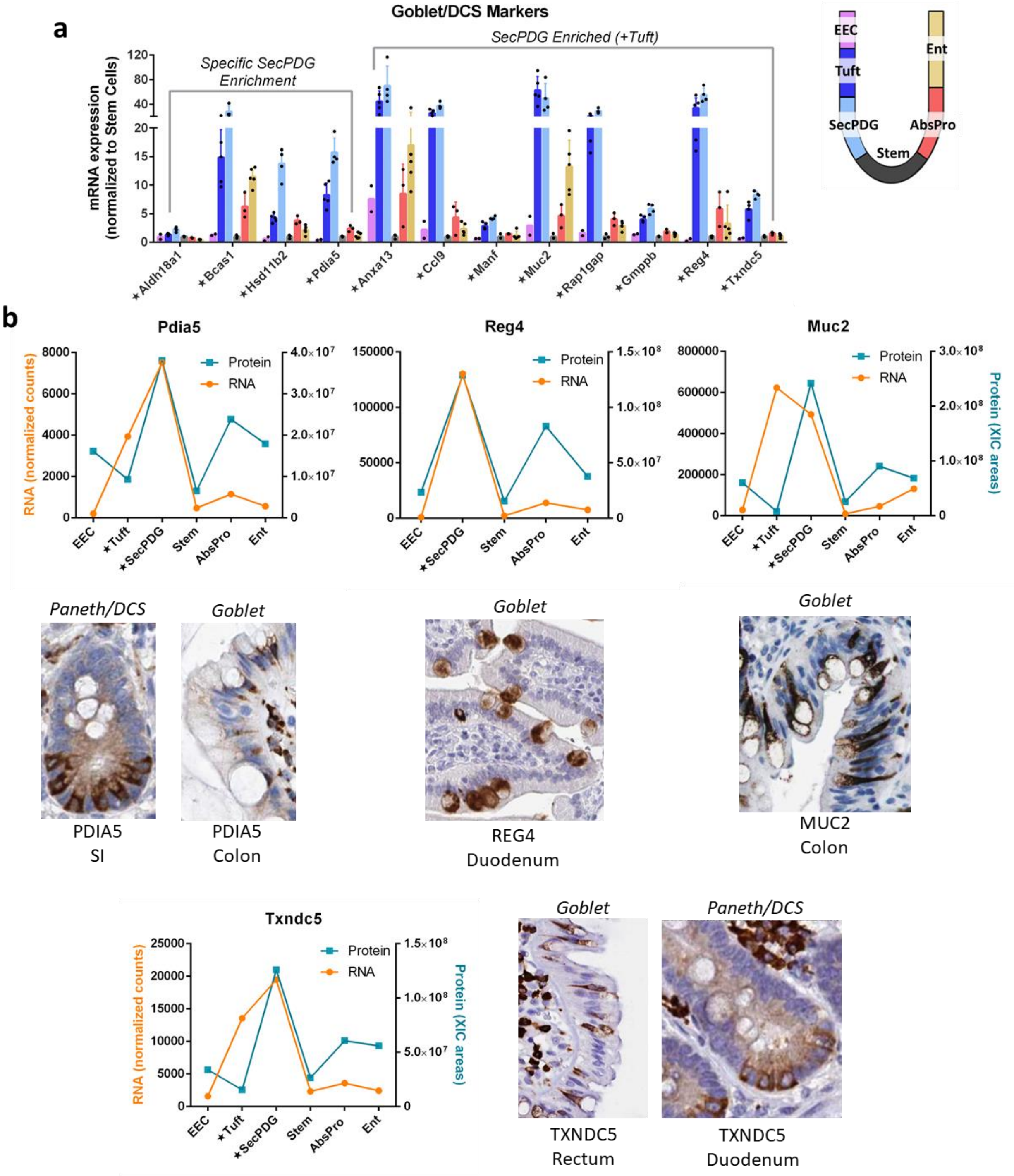
(Figure adapted from Supplementary Fig. 3 of Habowski et al., 2020; Communications Biology)


a
Known Stem, Absorptive, and Intestinal Gradient Markers
Bmp2
Alpi
Ascl2
Krt18
Krt20
Car4
Mep1a
Cbr1
Fabp2
Lgals3
Tjp1
Ocln
Slc5a1
Cd44
Lrig1
Cdca7
Axin2
Notch1
Rnf43
Hes1
Lgr5
Smoc2
14
12
10
8
6
4
EEC
Ent
SecPDG
Tuft
AbsPro
Stem
b
Known Secretory Lineage Markers
Muc2
Lrmp
Dclk1
Trpm5
Gnat3
Pou2f3
Ptprc
Chat
Ptgs1
Cldn4
Tph1
Pcsk1n
Lmx1a
Pax6
Cpe
Neurod1
Nkx2-2
Gfra3
Pax4
Atoh1
Spdef
Ptprn2
Manf
Xbp1
Hpgds
Hopx
Ccl9
Dll1
Chga
Chgb
Sox9
Hspa5
Txndc5
15
10
5
Stem
Ent
SecPDG
AbsPro
EEC
Tuft

Supplementary Figure 2.4: ***Expression of known intestinal crypt markers.***

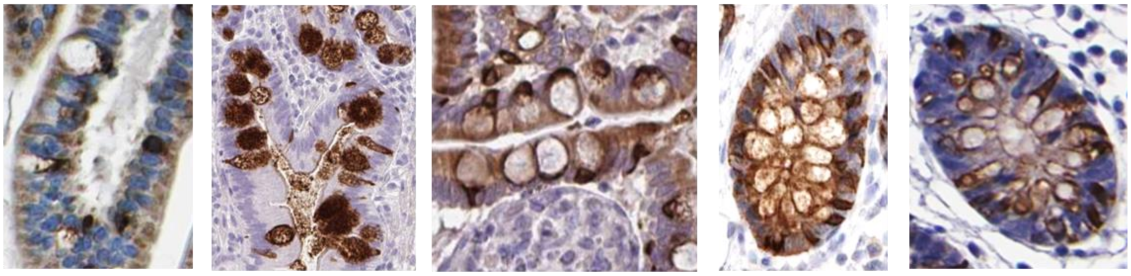
a Unsupervised clustering of known stem, absorptive lineage, and differentiation markers distinguishes cell types and confirms population identities. **b** Unsupervised clustering of known secretory lineage markers distinguishes secretory cell types, highlighting tuft markers (such as *Dclk1*) and high expression of *Chga/b* (EECs) and *Muc2*.

(Figure adapted from Supplementary Fig. 4 of Habowski et al., 2020; Communications Biology)



a
Goblet/DCS Markers
Specific SecPDG Enrichment
SecPDG Enriched (+Tuft)
mRNA expression (normalized to Stem Cells)
120
80
40
20
15
10
5
0
★Aldh18a1
★Bcas1
★Hsd11b2
★Pdia5
★Anxa13
★Ccl9
★Manf
★Muc2
★Rap1gap
★Gmppb
★Reg4
★Txndc5
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
b
Pdia5
RNA (normalized counts)
8000
6000
4000
2000
0
4.0×10⁷
3.0×10⁷
2.0×10⁷
1.0×10⁷
0
Protein
RNA
EEC
★Tuft
★SecPDG
Stem
AbsPro
Ent
Reg4
150000
100000
50000
0
1.5×10⁸
1.0×10⁸
5.0×10⁷
0
Protein
RNA
EEC
★SecPDG
Stem
AbsPro
Ent
Muc2
800000
600000
400000
200000
0
3.0×10⁸
2.0×10⁸
1.0×10⁸
0
Protein (XIC areas)
Protein
RNA
EEC
★Tuft
★SecPDG
Stem
AbsPro
Ent
Paneth/DCS
Goblet
PDIA5
SI
PDIA5
Colon
Goblet
REG4
Duodenum
Goblet
MUC2
Colon
Txndc5
RNA (normalized counts)
25000
20000
15000
10000
5000
0
1.5×10⁸
1.0×10⁸
5.0×10⁷
0
Protein (XIC areas)
Protein
RNA
EEC
★Tuft
★SecPDG
Stem
AbsPro
Ent
Goblet
Paneth/DCS
TXNDC5
Rectum
TXNDC5
Duodenum

c **SecPDG Markers enriched in Goblet Cells**



ALDH18A1
Colon

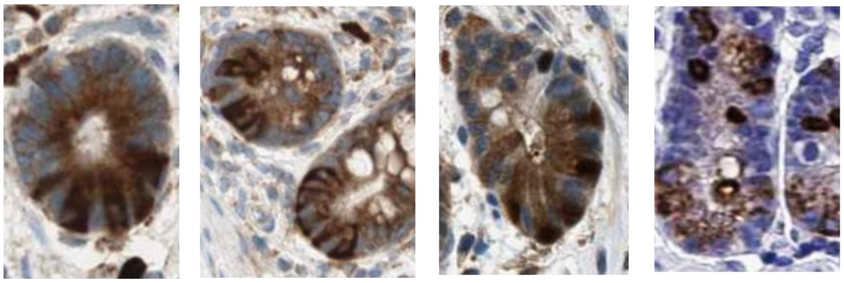
BCAS1
SI

HSD11B2
SI

ANXA13
Rectum

RAP1GAP
Colon

d **SecPDG Markers enriched Paneth Cells = Deep Crypt Secretory Cells**

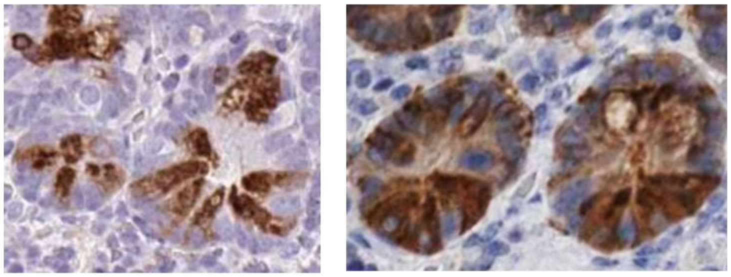


MANF
Colon

MANF
SI

GMPPB
Colon

RAP1GAP
Duodenum



BCAS1
SI

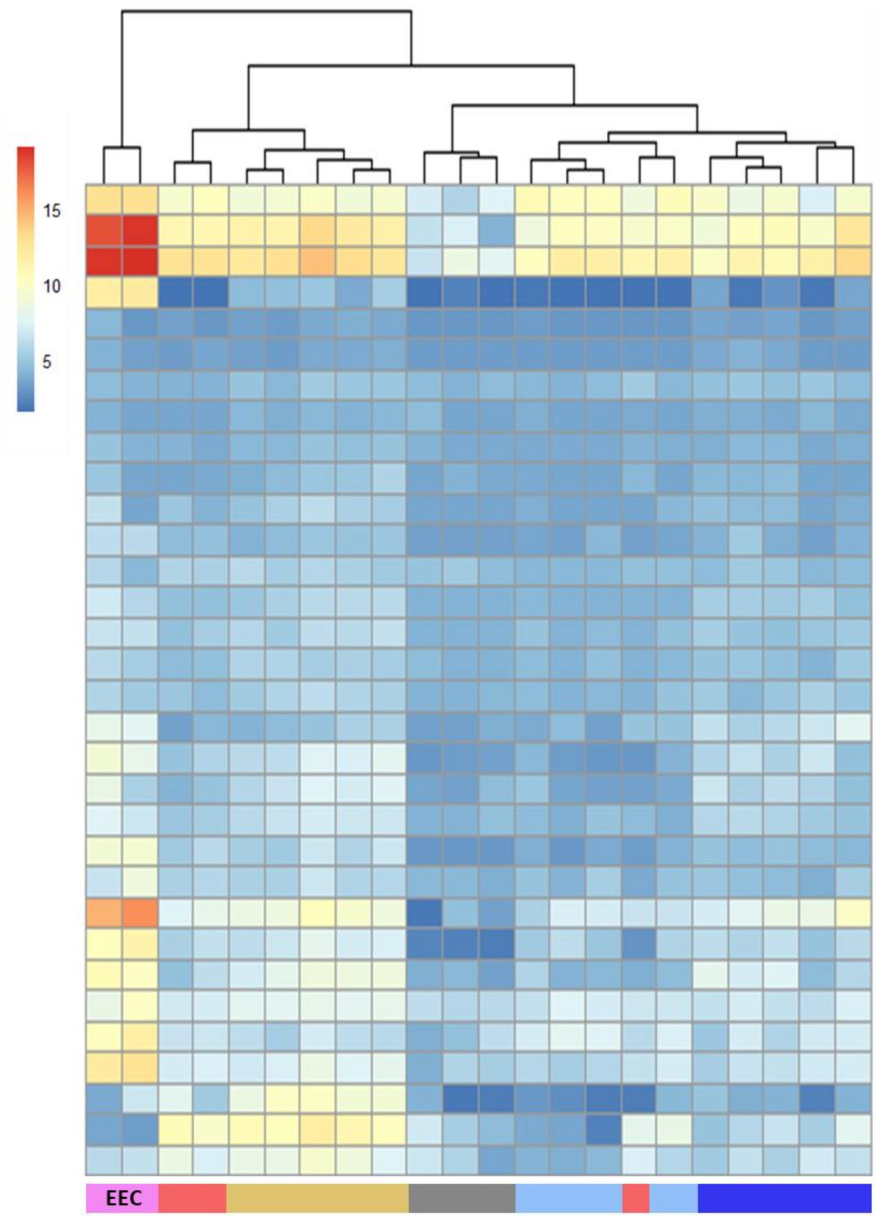
ANXA13
Duodenum

Supplementary Figure 2.5: ***SecPDG marker genes label both goblet and deep crypt secretory cells (DCS).***

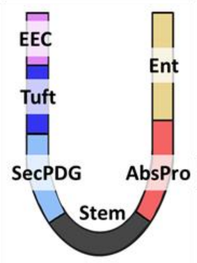
a mRNA expression of marker genes enriched selectively in SecPDG or in both SecPDG and tuft populations. List of all genes enriched in SecPDG is in Table 2.2. Star annotation by gene name symbolizes significant differential mRNA expression in at least one cell type compared to stem (padj < 0.01). mRNA differential expression analysis was performed with the following biological replicate numbers: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2. **b** Quantitative protein expression of marker genes Pdia5, Reg4, Muc2, and Txndc5 was detected with MS. Line graphs show the abundance of each protein compared to its mRNA expression across different cell types. Protein staining of these marker genes is evident in Paneth/DCS and Goblet cells. Star annotation by cell type symbolizes significant differential mRNA expression compared to stem (padj < 0.01). Staining of additional SecPDG markers in human tissues highlights **c** goblet cells, and **d** DCS/paneth cells (images from Human Protein Atlas). Some markers are expressed in both secretory progenitors, goblet, and DCS/paneth cell populations, whereas other markers are only expressed in specific subset populations. For example, REG4 and ALDH18A1 appear to be highly expressed in goblet cells along the crypt but not in DCS cells at the base of the crypt. MANF and GMPPB protein levels are high in DCS/paneth, but not goblet cells. PDIA5 antibody detects strong expression in paneth cells in the small intestine, but the same antibody does not clearly distinguish DCS in colon crypt (however it did mark colonic mature goblet cells).

(Figure adapted from Supplementary Fig. 5 of Habowski et al., 2020; Communications Biology)

EEC Subtype Markers



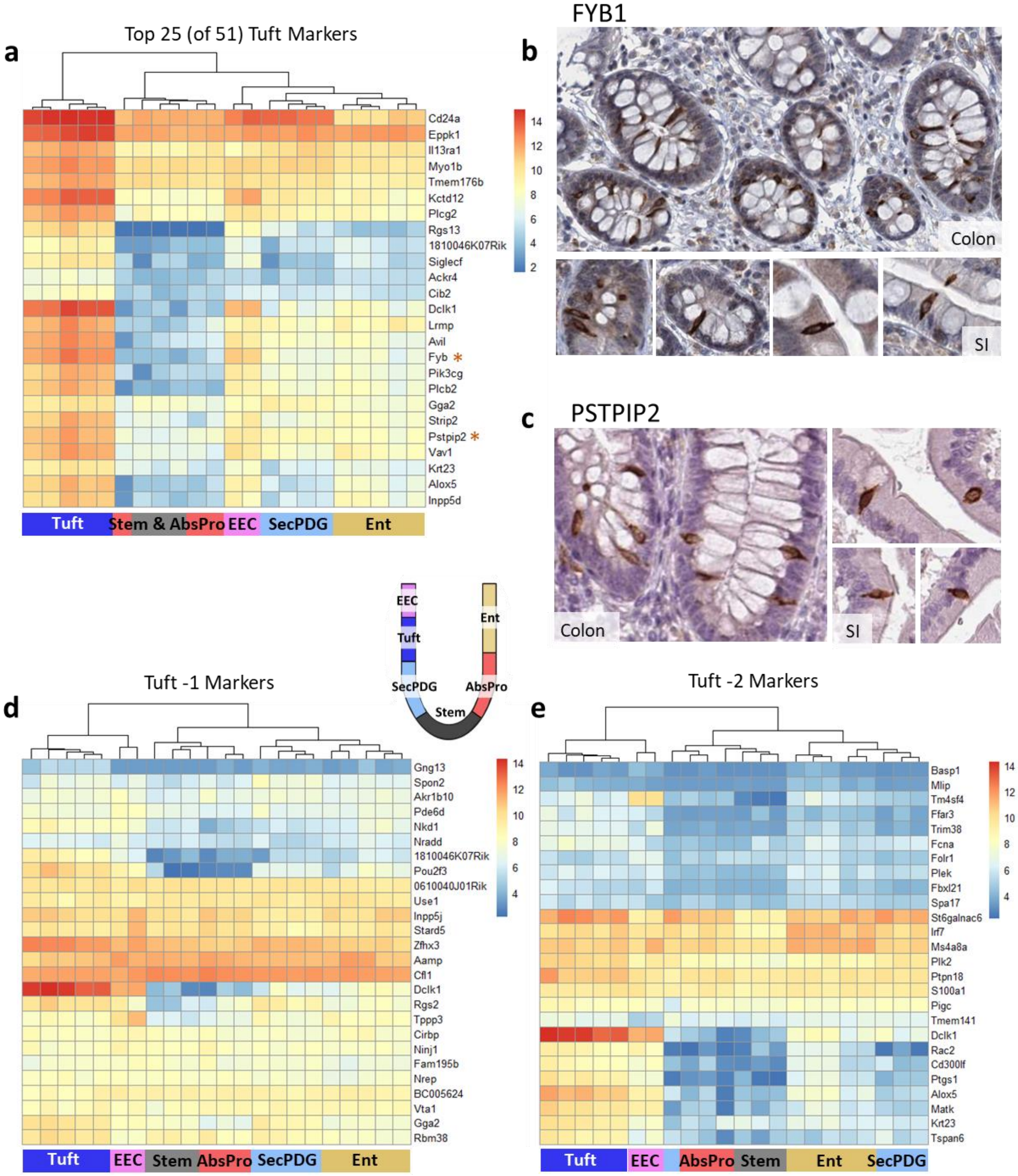
Gene	EEC Subtype
Sct	L, Ecm
Chga	Ecm
Chgb	Ecm
Iapp	
Agtr1a	L
Gal	
Ghrl	X
Ppy	L
Arx	X, I, L, K
Gip	K
Cck	L
Olfr558	Ecm
Insl5	L
Bdnf	
Isl1	N, L, K, D, I, X
Hhex	D
Scgn	
Sst	D
Pax6	L, K
Adgrd1	
Avpr1b	L
Olfr78	Ecm
Tac1	Ecm
Tph1	Ecm
Lmx1a	Progenitor
Piezo2	Ecm
Gch1	
Insm1	L
Neurod1	Progenitor
Nts	L
Gcg	L
Pyy	L



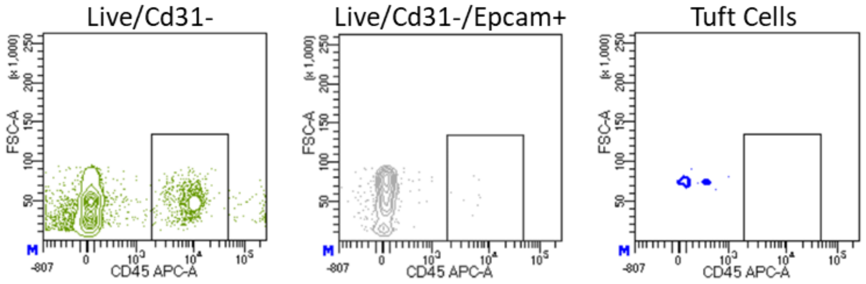
Supplementary Figure 2.6: ***Expression of known markers detects different subtypes of enteroendocrine cells (EEC).***

Unsupervised clustering of marker genes for EEC subtypes including L, X, I, K, N, D – cells, along with Ecm (enterochromaffin) and progenitors [88]. The sorted and sequenced EEC population presented here is predominantly Ecm, demarcated by the high expression of *Chga*, *Chgb*, and *Tph1*. Some L-cell markers, *Nts*, *Gcg*, and *Pyy*, are weakly expressed in Ent (and *Gcg* also in AbsPro). This suggests that rare L-cells might be pooled with Ent and additionally that Cd24 might be a distinguishing marker between different subtypes of EECs (with Ecm being Cd24 high).

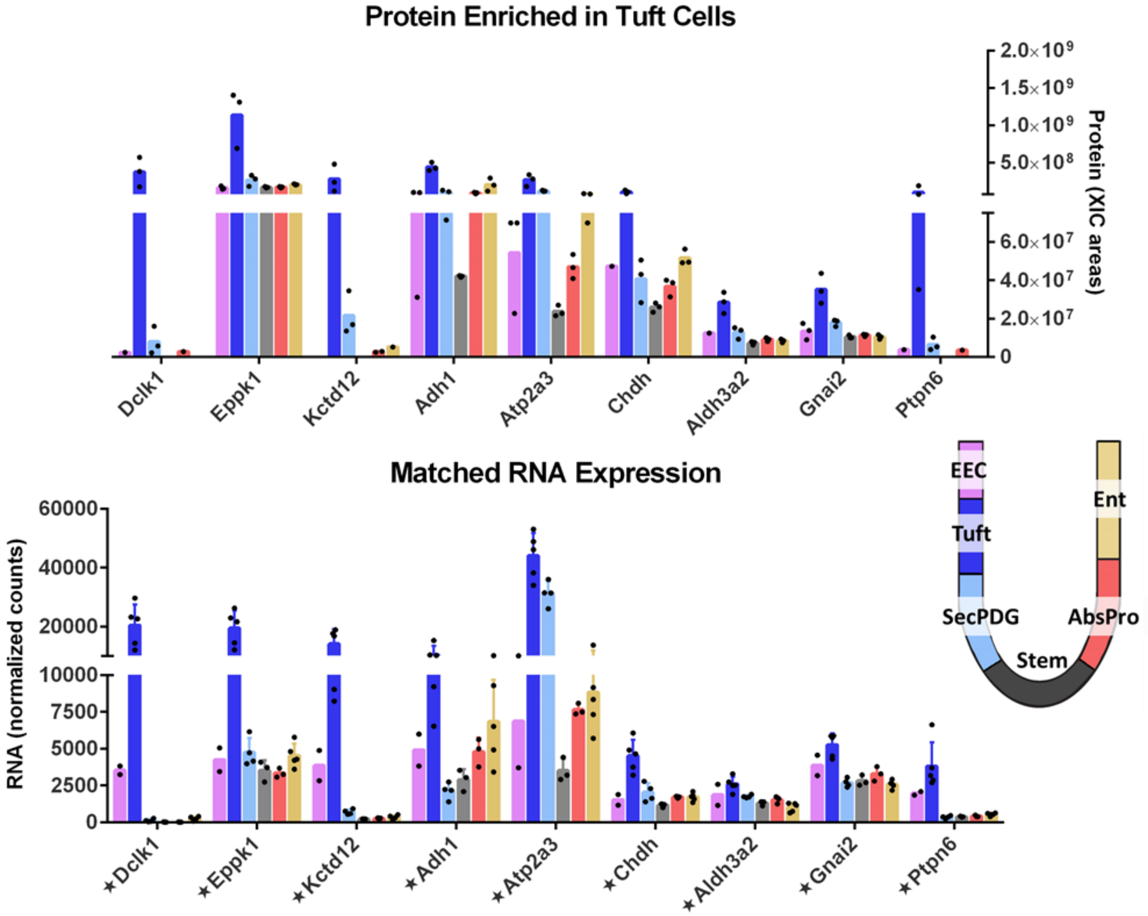
(Figure adapted from Supplementary Fig. 6 of Habowski et al., 2020; Communications Biology)


a
Top 25 (of 51) Tuft Markers
Cd24a
Eppk1
Il13ra1
Myo1b
Tmem176b
Kctd12
Plcg2
Rgs13
1810046K07Rik
Siglecf
Ackr4
Cib2
Dclk1
Lrmp
Avil
Fyb *
Pik3cg
Plcb2
Gga2
Strip2
Pstpip2 *
Vav1
Krt23
Alox5
Inpp5d
14
12
10
8
6
4
2
Tuft
Stem & AbsPro
EEC
SecPDG
Ent
b
FYB1
Colon
SI
c
PSTPIP2
Colon
SI
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
d
Tuft -1 Markers
Gng13
Spon2
Akr1b10
Pde6d
Nkd1
Nradd
1810046K07Rik
Pou2f3
0610040J01Rik
Use1
Inpp5j
Stard5
Zfhx3
Aamp
Cfl1
Dclk1
Rgs2
Tppp3
Cirbp
Ninj1
Fam195b
Nrep
BC005624
Vta1
Gga2
Rbm38
14
12
10
8
6
4
Tuft
EEC
Stem
AbsPro
SecPDG
Ent
e
Tuft -2 Markers
Basp1
Mlip
Tm4sf4
Ffar3
Trim38
Fcna
Folr1
Plek
Fbxl21
Spa17
St6galnac6
Irf7
Ms4a8a
Plk2
Ptpn18
S100a1
Pigc
Tmem141
Dclk1
Rac2
Cd300lf
Ptgs1
Alox5
Matk
Krt23
Tspan6
14
12
10
8
6
4
Tuft
EEC
AbsPro
Stem
Ent
SecPDG

f


Live/Cd31-
Live/Cd31-/Epcam+
Tuft Cells
FSC-A
(x 1,000)
250
200
150
100
50
CD45 APC-A
-807
0
10^3
10^4
10^5

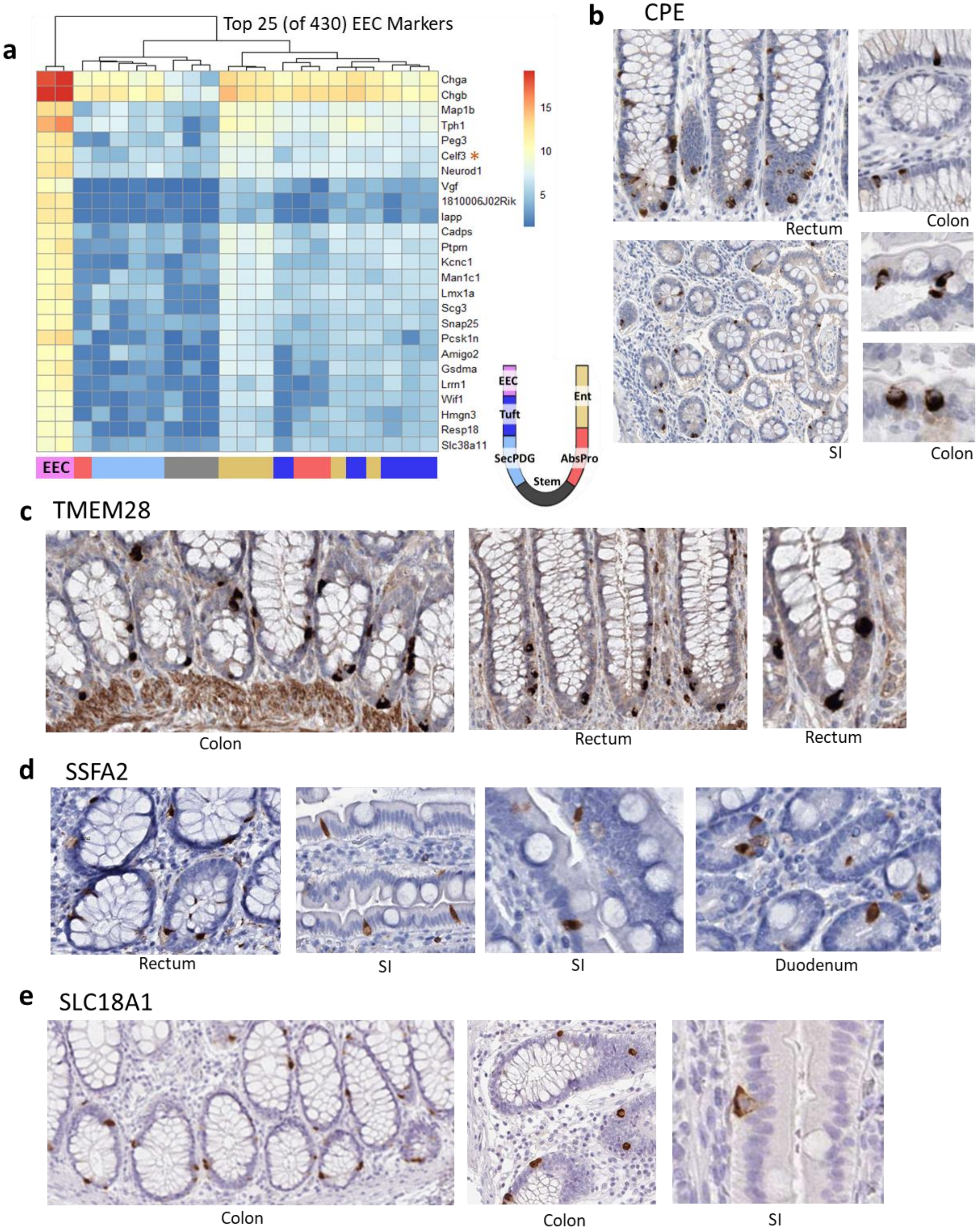
g


Protein Enriched in Tuft Cells
Protein (XIC areas)
2.0×10^9
1.5×10^9
1.0×10^9
5.0×10^8
6.0×10^7
4.0×10^7
2.0×10^7
0
Dclk1
Eppk1
Kctd12
Adh1
Atp2a3
Chdh
Aldh3a2
Gnai2
Ptpn6
Matched RNA Expression
RNA (normalized counts)
60000
40000
20000
10000
7500
5000
2500
0
★Dclk1
★Eppk1
★Kctd12
★Adh1
★Atp2a3
★Chdh
★Aldh3a2
★Gnai2
★Ptpn6
EEC
Tuft
SecPDG
Stem
AbsPro
Ent

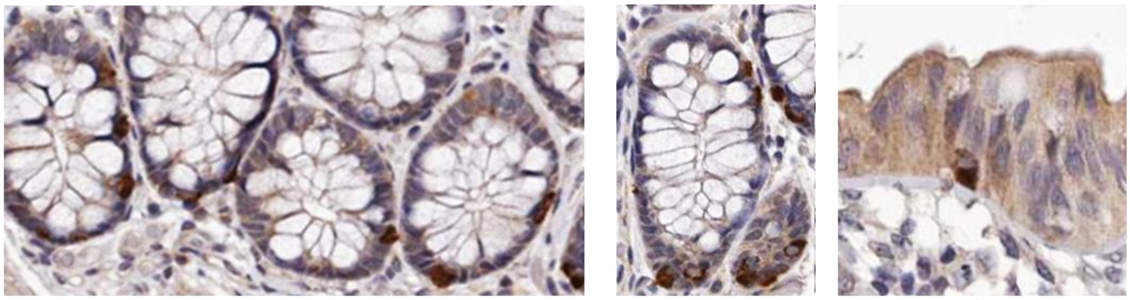
Supplementary Figure 2.7: ***Markers enriched in tuft cells.***

There were n=51 genes statistically enriched in mRNA expression in tuft cells compared to the five other isolated colon crypt cell types (padj < 0.01 + minimum mean 50 counts) (genes listed in Table 2.3). **a** Unsupervised clustering of the top 25 genes confirms their differential expression in tuft cells versus non-tuft cells. Protein staining is shown for genes with orange asterisk. Human protein atlas images confirm specificity with unique staining of tuft cells for the following markers **b** FYB1 and **c** PSTPIP2 staining shows positivity in cells with cells matching tuft morphologies. Haber *et al* identified two populations of Dclk1+ tuft cells in the small intestine, with the Tuft-1 populations denoted as Cd45- and the Tuft-2 population denoted as Cd45+ [19]. Unsupervised clustering of **d** Tuft-1 and **e** Tuft-2 marker genes as determined by Haber *et al* shows that despite Cd45+ cells being dumped in our sorting procedure (Cd45 is a marker for immune cells), we observe high expression of Tuft-2 markers (*Rac2, Cd300lf, Ptgs1, Alox5, Matk, Krt23,* and *Tspan6* for example) in our sorted Tuft population [89]. **f** Additionally, when we removed Cd45 from the dump channel we did not observe any Cd45+ tuft cells. Live/Cd31- cells show a Cd45+ population that disappears once Epcam+ cells are sorted. We conclude that colonic tuft cells are not expressing Cd45+ at detectable levels under homeostatic conditions and therefore it is not a marker for this population. **g** Using the MS proteomics data we identified proteins enriched in tuft cells which included known marker genes such at Dclk1. Data is displayed in two graphs: protein expression and mRNA expression of the same genes. Star annotation by gene name symbolizes significant differential mRNA expression in at least one cell type compared to stem (padj < 0.01) and error bars are standard deviation. For proteomics there are n=3 independent biological replicates for each cell type; for mRNA differential expression analysis the following biological replicate numbers were used: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2.

(Figure adapted from Supplementary Fig. 7 of Habowski et al., 2020; Communications Biology)

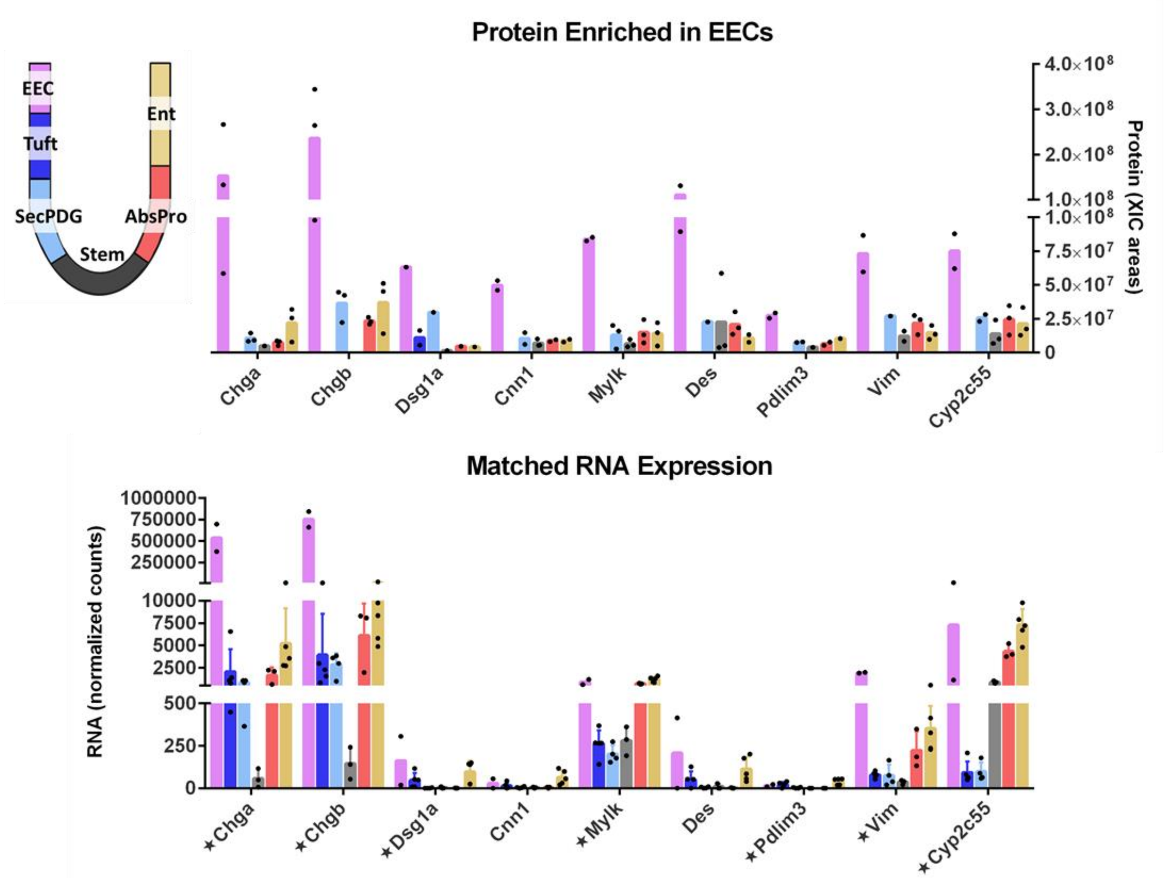

a
Top 25 (of 430) EEC Markers
Chga
Chgb
Map1b
Tph1
Peg3
Celf3 *
Neurod1
Vgf
1810006J02Rik
Iapp
Cadps
Ptprn
Kcnc1
Man1c1
Lmx1a
Scg3
Snap25
Pcsk1n
Amigo2
Gsdma
Lrrn1
Wif1
Hmgn3
Resp18
Slc38a11
15
10
5
EEC
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
b
CPE
Rectum
Colon
SI
Colon
c
TMEM28
Colon
Rectum
Rectum
d
SSFA2
Rectum
SI
SI
Duodenum
e
SLC18A1
Colon
Colon
SI

f CELF3



Colon Colon Colon

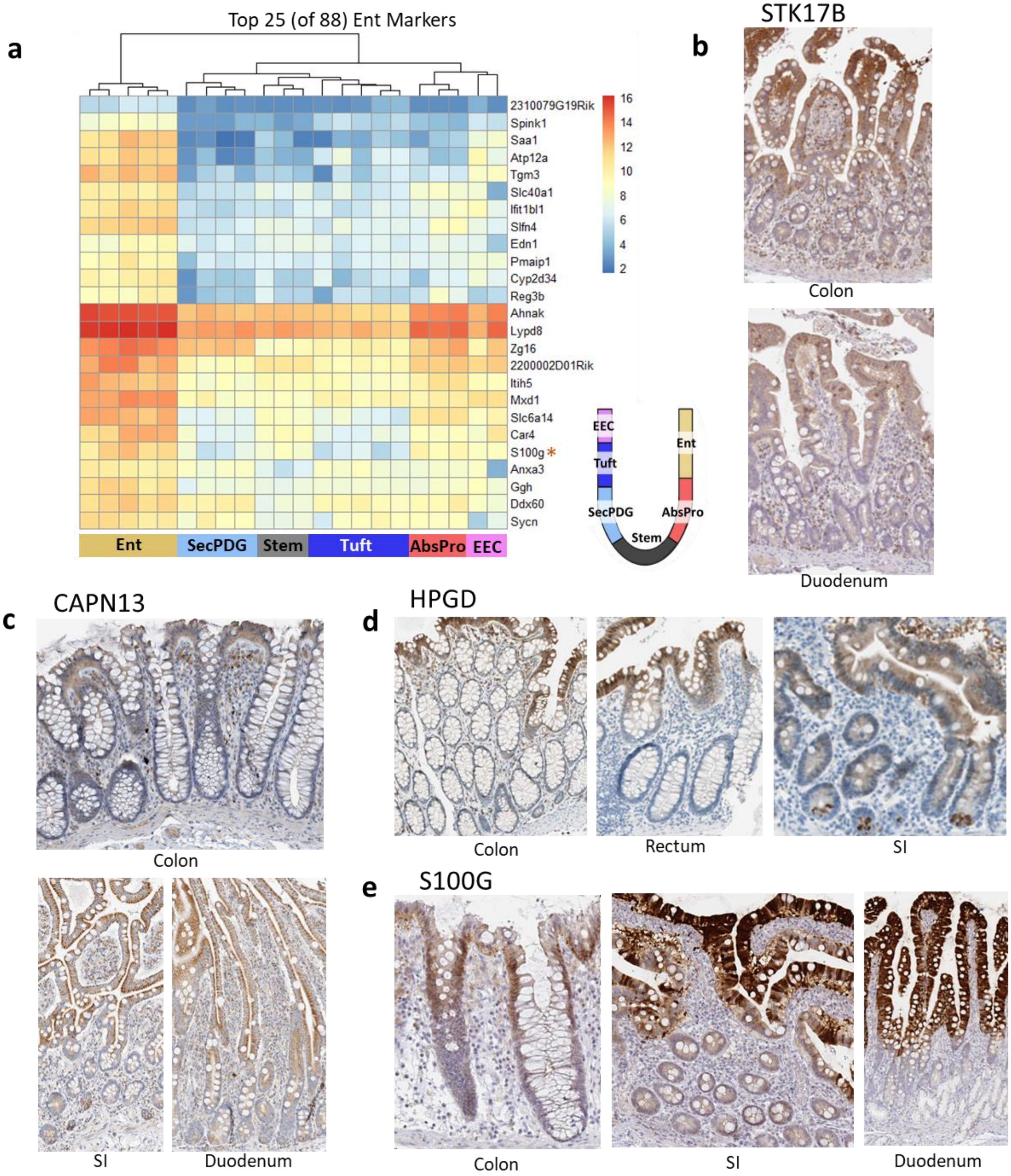
g



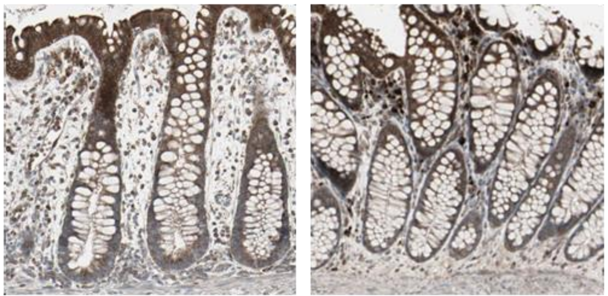
Supplementary Figure 2.8: ***Markers enriched in enteroendocrine cells.***

Our analysis identified n=430 genes as statistically enriched in EECs compared to the five other isolated colon crypt cell types (padj < 0.01 + minimum mean 50 counts) (genes listed in Table 2.4). **a** Unsupervised clustering of the top 25 genes differentially expressed between EEC and non-EEC. Protein staining in human intestinal tissues is shown for genes with an orange asterisk. Human protein atlas staining showing unique EEC expression of the following markers **b** CPE, **c** TMEM28, **d** SSFA2, **e** SLC18A1, and **f** CELF3. **g** Using the MS proteomics data we identified proteins enriched in EECs which included known marker genes such as Chga and Chgb. Data is displayed in two graphs: protein expression and mRNA expression of the same genes. Star annotation by gene name symbolizes significant differential mRNA expression in at least one cell type compared to stem (padj < 0.01) and error bars are standard deviation. For proteomics there are n=3 independent biological replicates for each cell type; for mRNA differential expression analysis the following biological replicate numbers were used: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2.

(Figure adapted from Supplementary Fig. 8 of Habowski et al., 2020; Communications Biology)

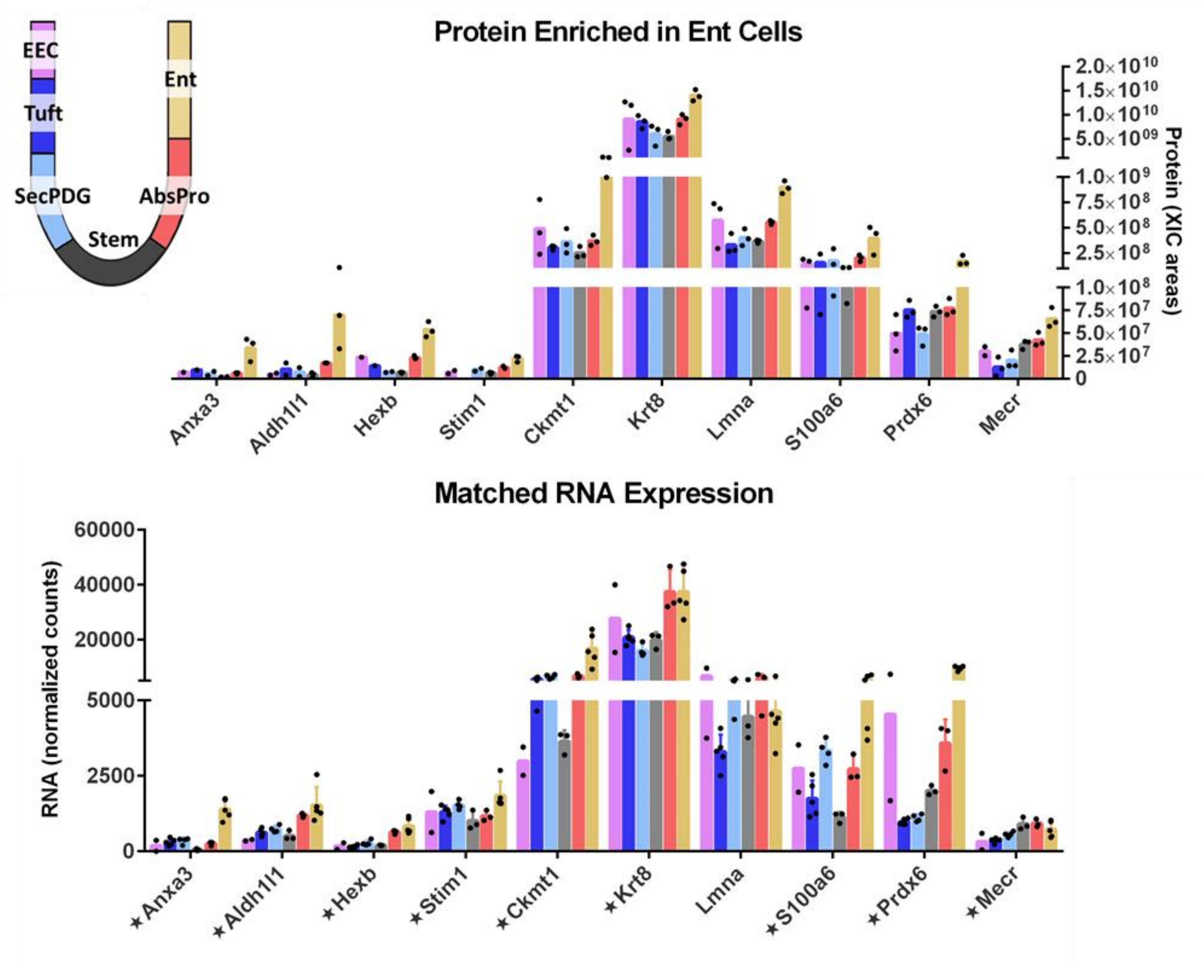

a
Top 25 (of 88) Ent Markers
2310079G19Rik
Spink1
Saa1
Atp12a
Tgm3
Slc40a1
Ifit1bl1
Slfn4
Edn1
Pmaip1
Cyp2d34
Reg3b
Ahnak
Lypd8
Zg16
2200002D01Rik
Itih5
Mxd1
Slc6a14
Car4
S100g *
Anxa3
Ggh
Ddx60
Sycn
16
14
12
10
8
6
4
2
Ent
SecPDG
Stem
Tuft
AbsPro
EEC
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
b
STK17B
Colon
Duodenum
c
CAPN13
Colon
SI
Duodenum
d
HPGD
Colon
Rectum
SI
e
S100G
Colon
SI
Duodenum

f RNASEL



Rectum Colon

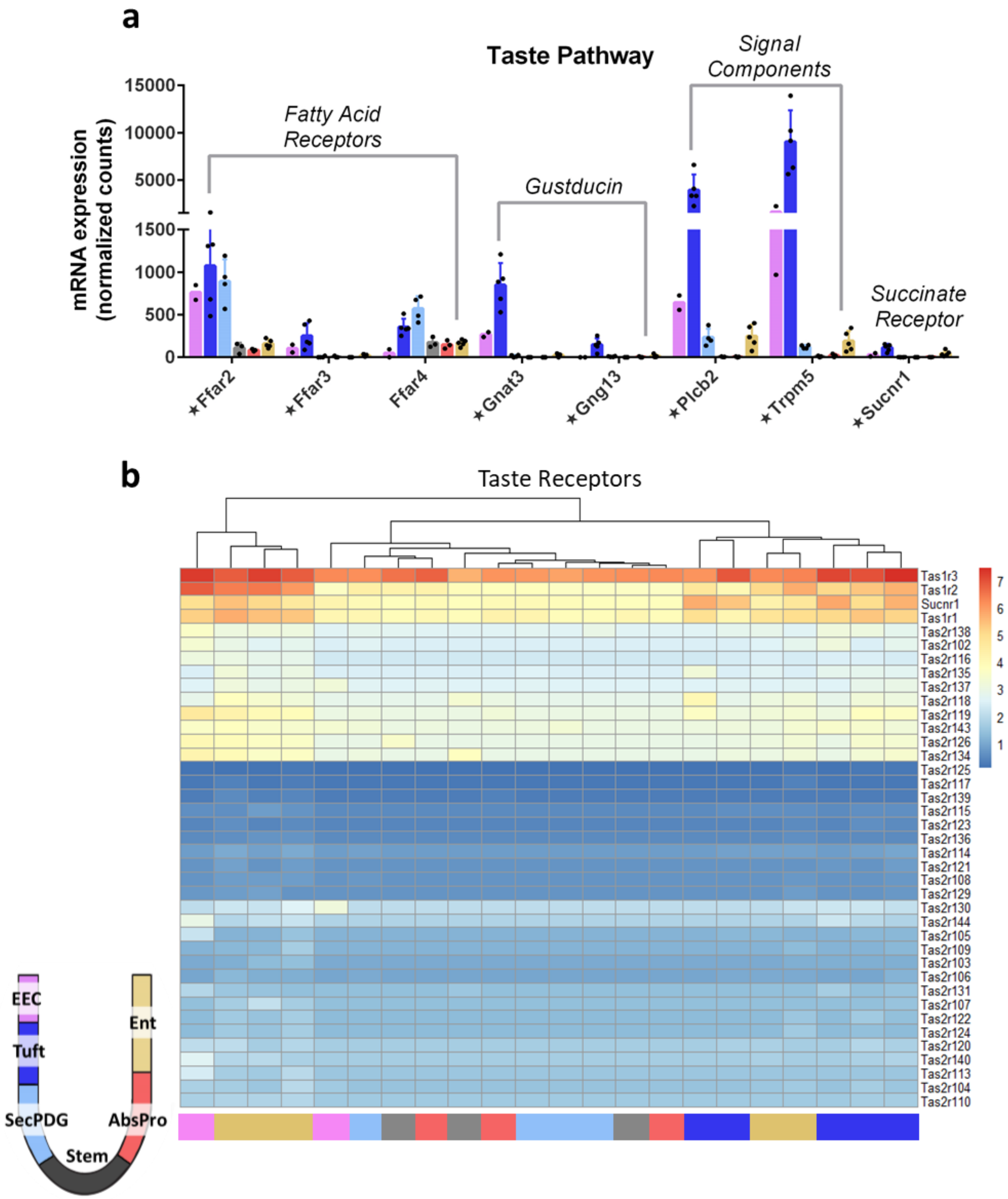
g



Supplementary Figure 2.9: ***Markers enriched in enterocytes.***

There were n=88 genes statistically enriched in the Ent population compared to the five other isolated colon crypt cell types (padj < 0.01 + minimum mean 50 counts) (genes listed in Table 2.3). **a** Unsupervised clustering of the top 25 genes differentially expressed between Ent and non-Ent. Protein staining is shown for genes with orange asterisk. Human protein atlas images show enriched protein staining in Ent at the top of the crypt for the following markers **b** STK17B, **c** CAPN13, **d** HPGD, **e** S100G and **f** RNASEL. **g** Using the MS proteomics data we identified proteins enriched in the Ent population. Data is displayed in two graphs: protein expression and mRNA expression of the same genes. Star annotation by gene name symbolizes significant differential mRNA expression in at least one cell type compared to stem (padj < 0.01) and error bars are standard deviation. For proteomics there are n=3 independent biological replicates for each cell type; for mRNA differential expression analysis the following biological replicate numbers were used: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2.

(Figure adapted from Supplementary Fig. 9 of Habowski et al., 2020; Communications Biology)


a
Taste Pathway
mRNA expression (normalized counts)
15000
10000
5000
1500
1000
500
0
Fatty Acid Receptors
Gustducin
Signal Components
Succinate Receptor
★Ffar2
★Ffar3
Ffar4
★Gnat3
★Gng13
★Plcb2
★Trpm5
★Sucnr1
b
Taste Receptors
Tas1r3
Tas1r2
Sucnr1
Tas1r1
Tas2r138
Tas2r102
Tas2r116
Tas2r135
Tas2r137
Tas2r118
Tas2r119
Tas2r143
Tas2r126
Tas2r134
Tas2r125
Tas2r117
Tas2r139
Tas2r115
Tas2r123
Tas2r136
Tas2r114
Tas2r121
Tas2r108
Tas2r129
Tas2r130
Tas2r144
Tas2r105
Tas2r109
Tas2r103
Tas2r106
Tas2r131
Tas2r107
Tas2r122
Tas2r124
Tas2r120
Tas2r140
Tas2r113
Tas2r104
Tas2r110
7
6
5
4
3
2
1
EEC
Tuft
SecPDG
Stem
AbsPro
Ent

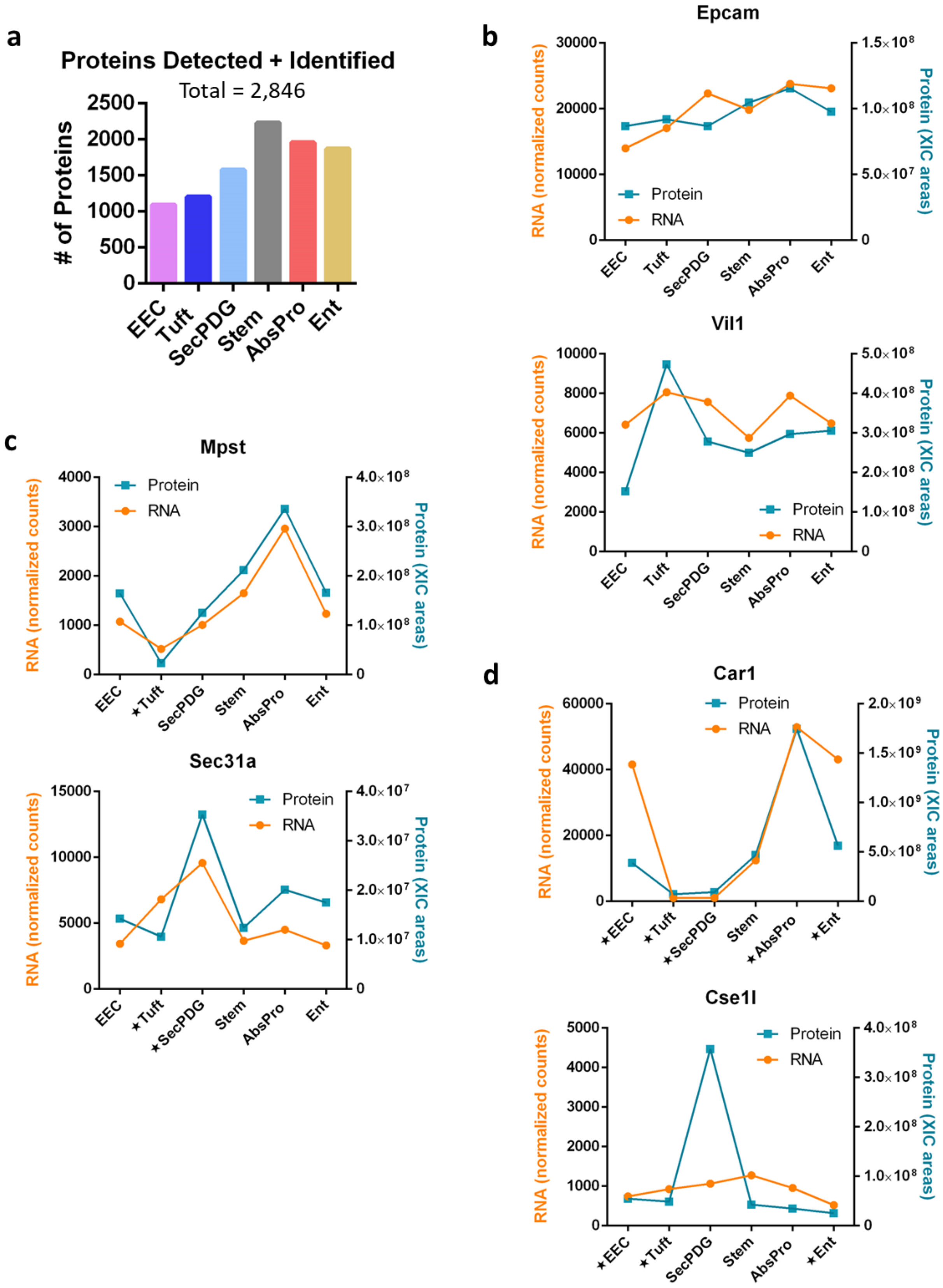
Supplementary Figure 2.10: ***Colonic tuft cells show a distinct potential for taste pathway activation via fatty acids.***

a mRNA expression of taste pathway signaling components elevated in tuft cells including three fatty acid receptors (*Ffar 2,3,4*), a primarily taste-specific G protein alpha-subunit (*Gnat3*; also known as Gustducin) and a related G protein subunit (*Gng13*) and cytoplasmic signaling components involved in taste signaling. Star annotation by gene name symbolizes significant differential mRNA expression in at least one cell type compared to stem (padj < 0.01) and error bars are standard deviation. For differential mRNA expression analysis, the following biological replicate numbers were used: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2. **b** Taste receptor family expression with minimal expression of Tas2 receptors and non-tuft specific expression of Tas1 receptors. See *Supplementary Discussion* section for additional information.

Discussion for Supplementary Figure 2.10:

Tuft cells in the small intestine function as chemo-sensory cells for “tasting” pathogens and coordinating a protective immune response. Several studies recently showed that small intestinal tuft cells have the ability to detect pathogens via the small molecule succinate due to their high expression of the succinate receptor (*Sucnr1*) and taste sensory signaling components gustducin (*Gnat3*) and *Trmp5* [26,90–92]. Whether tuft cells in the colon exhibit a similar chemo-sensory “taste” profile was not known. We detected high expression of taste signaling components for signal transduction in tuft cells, (*Gnat3*, *Gng13*, high expression of *Trmp5* and taste sensing signal facilitator *Plcb2*; Supplementary Figure 2.10a), however, there is little to no expression of Tas2 taste family receptors or the succinate receptor *Sucnr1,* (Supplementary Figure 2.10a,b) [90]. Tas1 taste receptors are expressed in tuft cells, but they are also expressed by all cell types and therefore not tuft cell-specific. Instead, we discovered tuft cell-specific expression of the free fatty acid receptors, *Ffar2*, *Ffar3,* and *Ffar4*, receptors that utilize the same downstream sensory signaling components as for succinate. These data indicate that colon tuft cells have a unique ability for metabolic sensing of fats, particularly short chain fatty acids.

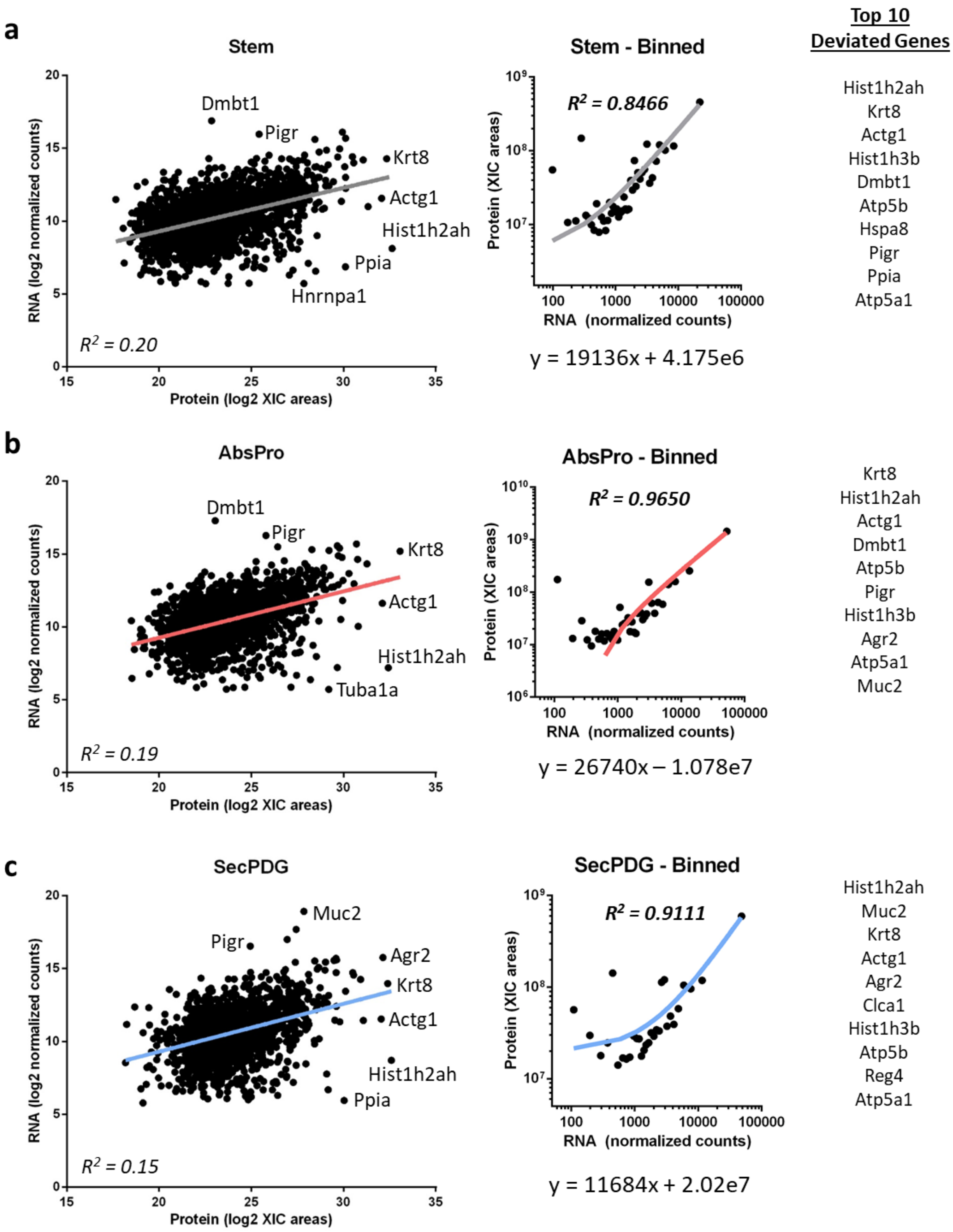
(Figure adapted from Supplementary Fig. 10 of Habowski et al., 2020; Communications Biology)


a
Proteins Detected + Identified
Total = 2,846
# of Proteins
2500
2000
1500
1000
500
0
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
b
Epcam
RNA (normalized counts)
30000
20000
10000
0
Protein (XIC areas)
1.5×10⁸
1.0×10⁸
5.0×10⁷
0
Protein
RNA
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
Vil1
RNA (normalized counts)
10000
8000
6000
4000
2000
0
Protein (XIC areas)
5.0×10⁸
4.0×10⁸
3.0×10⁸
2.0×10⁸
1.0×10⁸
0
Protein
RNA
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
c
Mpst
RNA (normalized counts)
4000
3000
2000
1000
0
Protein (XIC areas)
4.0×10⁸
3.0×10⁸
2.0×10⁸
1.0×10⁸
0
Protein
RNA
EEC
⋆Tuft
SecPDG
Stem
AbsPro
Ent
Sec31a
RNA (normalized counts)
15000
10000
5000
0
Protein (XIC areas)
4.0×10⁷
3.0×10⁷
2.0×10⁷
1.0×10⁷
0
Protein
RNA
EEC
⋆Tuft
⋆SecPDG
Stem
AbsPro
Ent
d
Car1
RNA (normalized counts)
60000
40000
20000
0
Protein (XIC areas)
2.0×10⁹
1.5×10⁹
1.0×10⁹
5.0×10⁸
0
Protein
RNA
⋆EEC
⋆Tuft
⋆SecPDG
Stem
⋆AbsPro
⋆Ent
Cse1l
RNA (normalized counts)
5000
4000
3000
2000
1000
0
Protein (XIC areas)
4.0×10⁸
3.0×10⁸
2.0×10⁸
1.0×10⁸
0
Protein
RNA
⋆EEC
⋆Tuft
SecPDG
Stem
AbsPro
⋆Ent

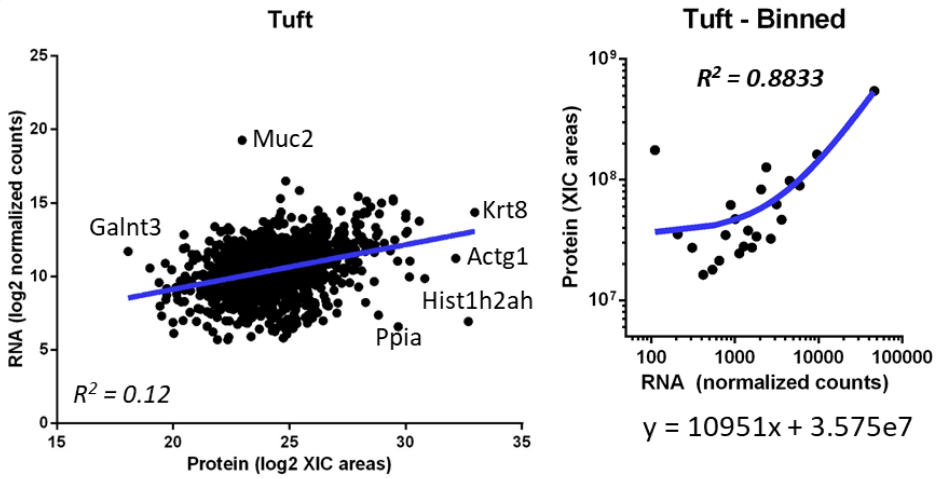
Supplementary Figure 2.11: ***Mass Spec analysis detected proteins expressed in the crypt cell types.***

a In total 2,846 proteins were detected and identified with more than 1,000 proteins in all cell types (including rare cell types such as tuft and EEC). **b** Protein from known genes expressed in the crypt, such as Epcam, were identified in all cell types and consistent with mRNA expression. Other genes such as Vil1, were also detected in all cell types at the protein level, but not as consistent with mRNA expression. **c** We identified genes that are markers of AbsPro (Mpst) and SecPDG (Sec31a) populations on both an mRNA and protein level. **d** We also observed some genes that were not consistent between mRNA and protein. for example, in AbsPro, Car1 protein is enriched whereas the mRNA is also enriched in Ent. Cse1l is only enriched in SecPDG on a protein level. Star annotation by gene name symbolizes significant differential mRNA expression in at least one cell type compared to stem (padj < 0.01). For differential mRNA expression analysis, the following independent biological replicate numbers were used: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2.

(Figure adapted from Supplementary Fig. 11 of Habowski et al., 2020; Communications Biology)


a
Stem
Dmbt1
Pigr
Krt8
Actg1
Hist1h2ah
Ppia
Hnrnpa1
$R^2 = 0.20$
RNA (log2 normalized counts)
Protein (log2 XIC areas)
0
5
10
15
20
15
20
25
30
35
Stem - Binned
$R^2 = 0.8466$
Protein (XIC areas)
RNA (normalized counts)
10^9
10^8
10^7
100
1000
10000
100000
y = 19136x + 4.175e6
Top 10 Deviated Genes
Hist1h2ah
Krt8
Actg1
Hist1h3b
Dmbt1
Atp5b
Hspa8
Pigr
Ppia
Atp5a1
b
AbsPro
Dmbt1
Pigr
Krt8
Actg1
Hist1h2ah
Tuba1a
$R^2 = 0.19$
RNA (log2 normalized counts)
Protein (log2 XIC areas)
AbsPro - Binned
$R^2 = 0.9650$
Protein (XIC areas)
RNA (normalized counts)
10^{10}
10^9
10^8
10^7
10^6
100
1000
10000
100000
y = 26740x – 1.078e7
Krt8
Hist1h2ah
Actg1
Dmbt1
Atp5b
Pigr
Hist1h3b
Agr2
Atp5a1
Muc2
c
SecPDG
Muc2
Pigr
Agr2
Krt8
Actg1
Hist1h2ah
Ppia
$R^2 = 0.15$
RNA (log2 normalized counts)
Protein (log2 XIC areas)
SecPDG - Binned
$R^2 = 0.9111$
Protein (XIC areas)
RNA (normalized counts)
10^9
10^8
10^7
100
1000
10000
100000
y = 11684x + 2.02e7
Hist1h2ah
Muc2
Krt8
Actg1
Agr2
Clca1
Hist1h3b
Atp5b
Reg4
Atp5a1

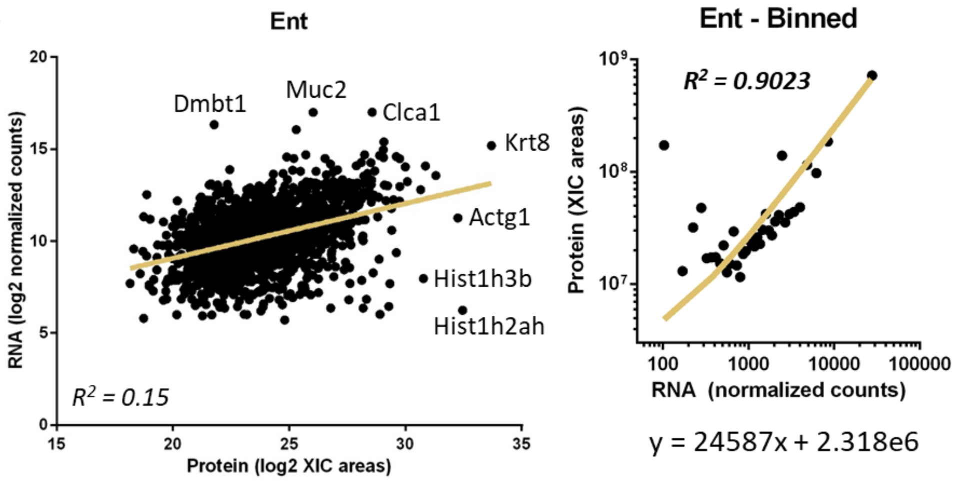
d



Top 10 Deviated Genes

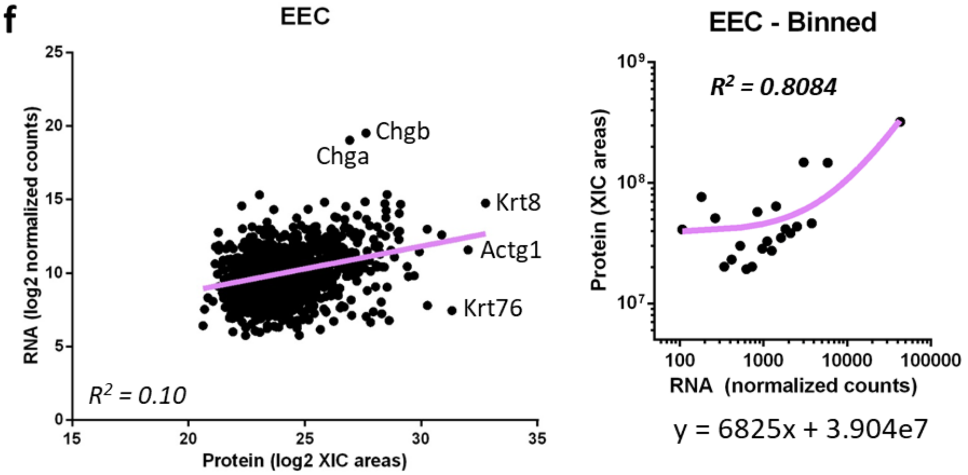
Krt8
Hist1h2ah
Muc2
Actg1
Hist1h3b
Atp5b
Tuba1a
Hist2h2ac
Clca1
Hspa8

e



Krt8
Hist1h2ah
Actg1
Muc2
Clca1
Atp5b
Dmbt1
Hist1h3b
Pigr
Atp5a1

f

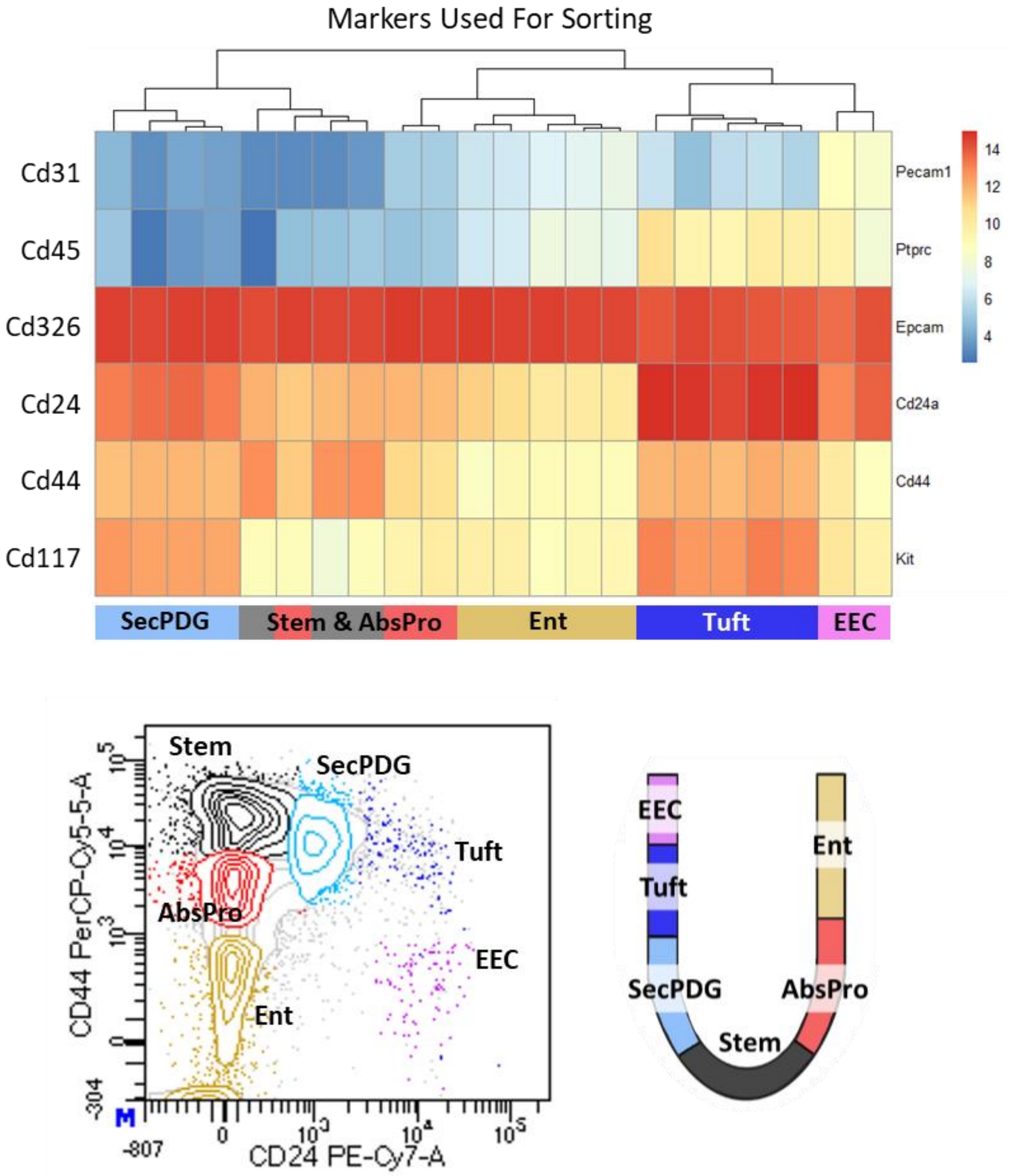


Krt8
Chgb
Actg1
Chga
Krt76
Atp5b
Hist1h3b
Atp5a1
Agr2
Hist2h2ac

Supplementary Figure 2.12: ***Comparison of mRNA to protein in each crypt cell type.***

Each gene that is detected on the protein level with MS and has mRNA expression >50 normalized counts is graphed for each cell type (scales are log2). A linear best fine line is drawn, however the R^2 values are poor. When the data is sorted based on mRNA expression and binned, with 50 genes in each bin, variation can be minimized and linear best fit lines fit the data well resulting in stronger R^2 values. Using the binned best fit line equation, we identified the top 10 deviating genes based on standard deviation of actual protein versus expected protein. The cell types include **a** Stem, **b** AbsPro, **c** SecPDG, **d** Tuft, **e** Ent, and **f** EEC. The strongest binned R^2 value was observed in AbsPro ($R^2 = 0.965$).

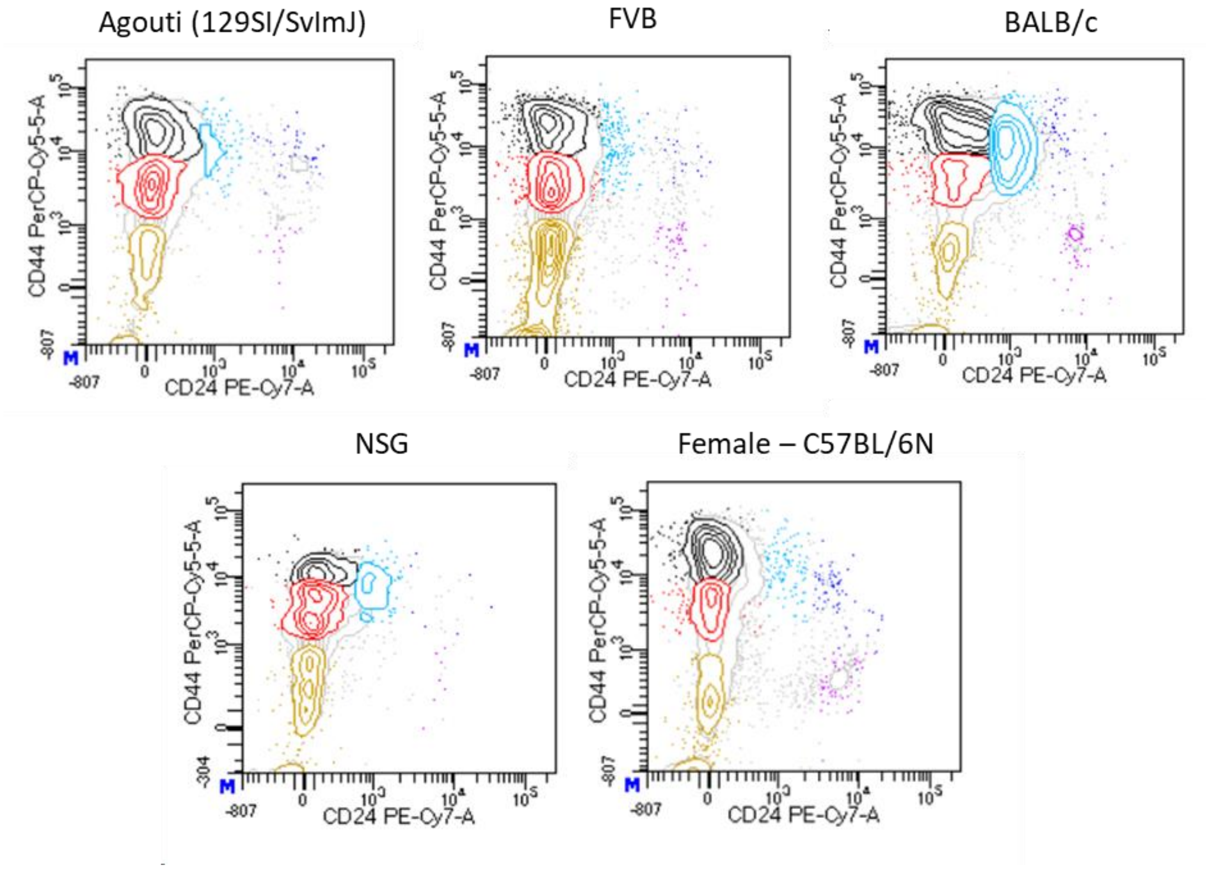
(Figure adapted from Supplementary Fig. 12 of Habowski et al., 2020; Communications Biology)



Supplementary Figure 2.13: ***Surface marker protein expression correlations with mRNA gene expression.***

Unsupervised clustering of mRNA gene expression of surface markers used for sorting. The dump channel includes Cd31 and Cd45 (to discard endothelial and immune system cells, respectively), whereas positive markers Cd326, Cd24, Cd44, and Cd117 are used to distinguish and purify their respective populations. Although no Cd45 protein was detected in tuft cells (Supplementary Figure 2.7f), a low, but clearly detectable level of mRNA (*Ptprc)* was detected and elevated in tuft cells compared to other isolated colon crypt cell types.

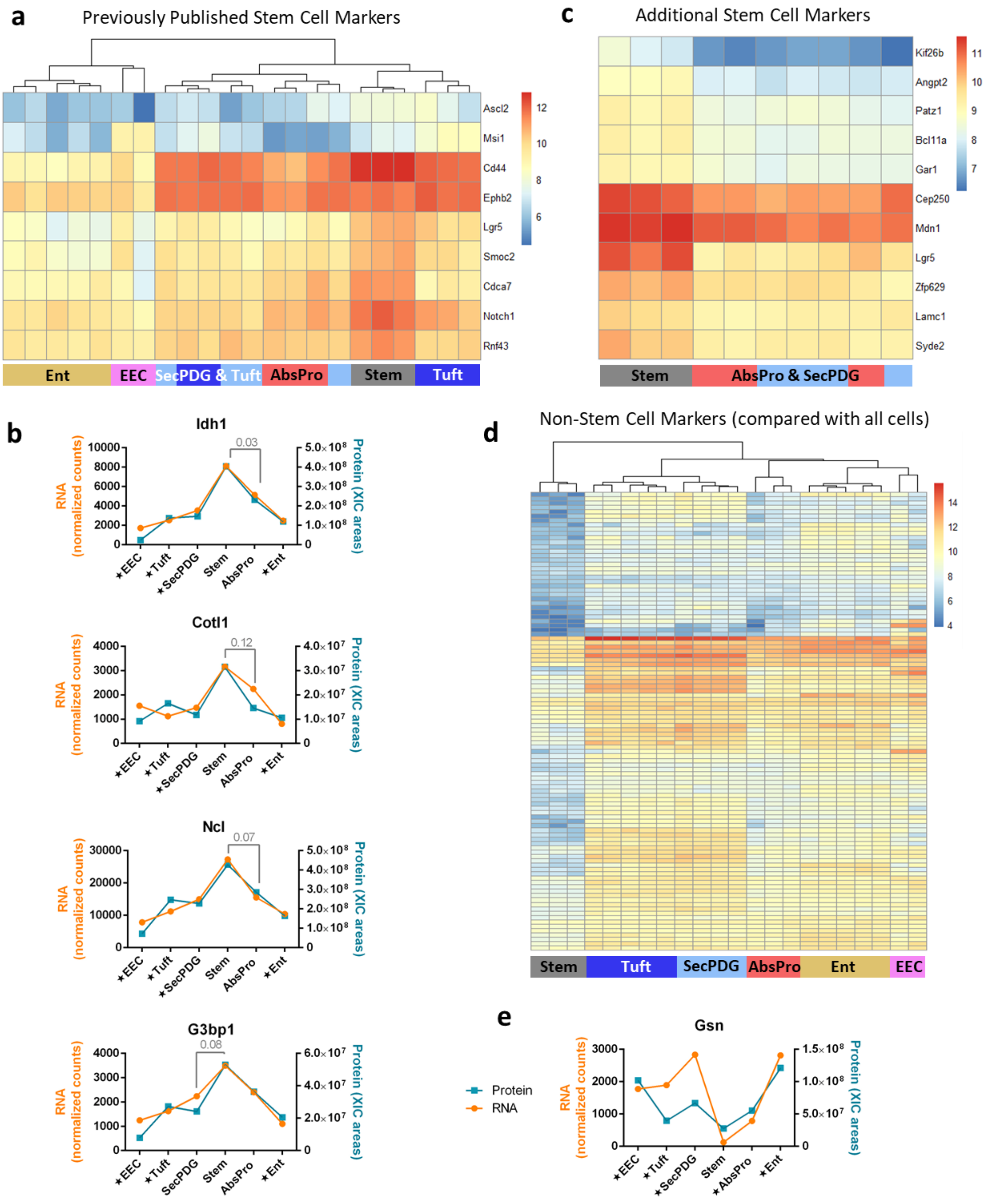
(Figure adapted from Supplementary Fig. 13 of Habowski et al., 2020; Communications Biology)



Supplementary Figure 2.14: ***Sorting procedure is works broadly in other mouse strains and gender.***

Sorting procedure is universal and resolves cell types in the colon of other mouse strains and gender including Agoutti, FVB, Balbc, NSG, and females. Each FACS plot is a representative image from one mouse, n=3 independent sorts.

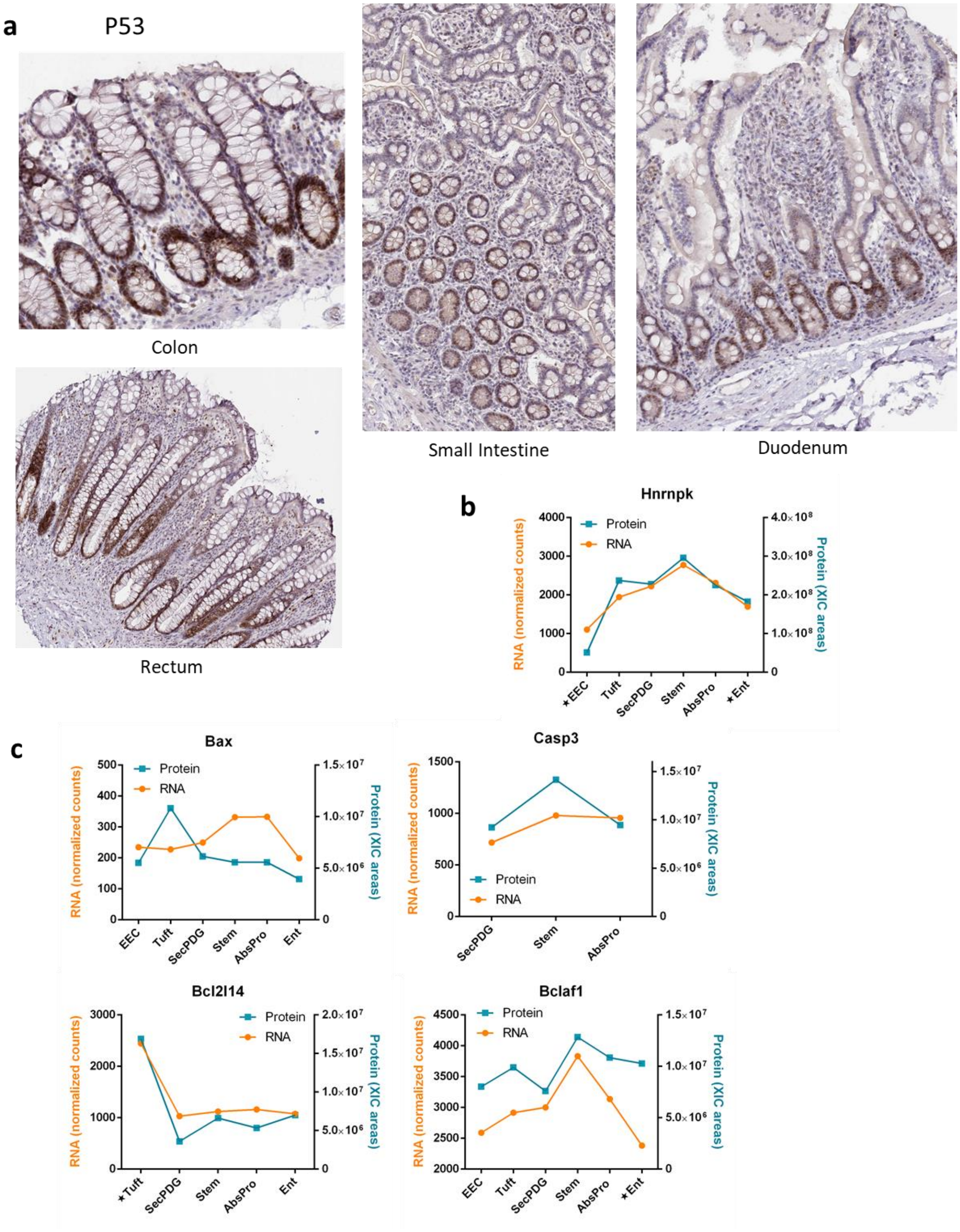
(Figure adapted from Supplementary Fig. 14 of Habowski et al., 2020; Communications Biology)


a
Previously Published Stem Cell Markers
Ascl2
Msi1
Cd44
Ephb2
Lgr5
Smoc2
Cdca7
Notch1
Rnf43
12
10
8
6
Ent
EEC
SecPDG & Tuft
AbsPro
Stem
Tuft
c
Additional Stem Cell Markers
Kif26b
Angpt2
Patz1
Bcl11a
Gar1
Cep250
Mdn1
Lgr5
Zfp629
Lamc1
Syde2
11
10
9
8
7
Stem
AbsPro & SecPDG
b
Idh1
0.03
RNA (normalized counts)
Protein (XIC areas)
10000
8000
6000
4000
2000
0
5.0×10⁸
4.0×10⁸
3.0×10⁸
2.0×10⁸
1.0×10⁸
0
⋆EEC
⋆Tuft
⋆SecPDG
Stem
AbsPro
⋆Ent
Cotl1
0.12
4000
3000
2000
1000
0
4.0×10⁷
3.0×10⁷
2.0×10⁷
1.0×10⁷
0
Ncl
0.07
30000
20000
10000
0
5.0×10⁸
4.0×10⁸
3.0×10⁸
2.0×10⁸
1.0×10⁸
0
G3bp1
0.08
4000
3000
2000
1000
0
6.0×10⁷
4.0×10⁷
2.0×10⁷
0
⋆EEC
⋆Tuft
SecPDG
Stem
AbsPro
⋆Ent
d
Non-Stem Cell Markers (compared with all cells)
14
12
10
8
6
4
Stem
Tuft
SecPDG
AbsPro
Ent
EEC
e
Gsn
Protein
RNA
3000
2000
1000
0
1.5×10⁸
1.0×10⁸
5.0×10⁷
0
⋆EEC
⋆Tuft
⋆SecPDG
Stem
⋆AbsPro
⋆Ent

Supplementary Figure 2.15: ***Additional markers of intestinal stemness.***

a Unsupervised clustering of known/classically defined intestinal stem cell markers. **b** mRNA and protein expression of markers elevated in stem, but not passing the stringent mRNA padj < 0.01 cut-off between stem and progenitors. These marker genes continue to decrease in mRNA and protein expression and are lowest in the most differentiated cell types. **c** When the analysis of stem versus non-stem in Figure 2.2c is performed *only* between stem versus AbsPro and SecPDG, n=11 additional stem markers are identified for a total of n=27 gene expression differences that distinguish stem cells from progenitors. **d** Unsupervised clustering of n=107 genes that were significantly elevated in expression in all non-stem cells (SecPDG, AbsPro, Ent, tuft, EEC) compared to stem cells (see Supplementary Data #3 for gene list). The highly abundant *Chga* and *Chgb* mRNAs were not included in the heatmap. Their expression is several logs greater than all the genes shown here. **e** Gsn is one of the n=107 differentiation genes that is elevated compared to stem and graph displays mRNA and protein expression. Star annotation by cell type symbolizes significant differential mRNA expression compared to stem (padj < 0.01). For differential mRNA expression analysis, the following independent biological replicate numbers were used: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2.

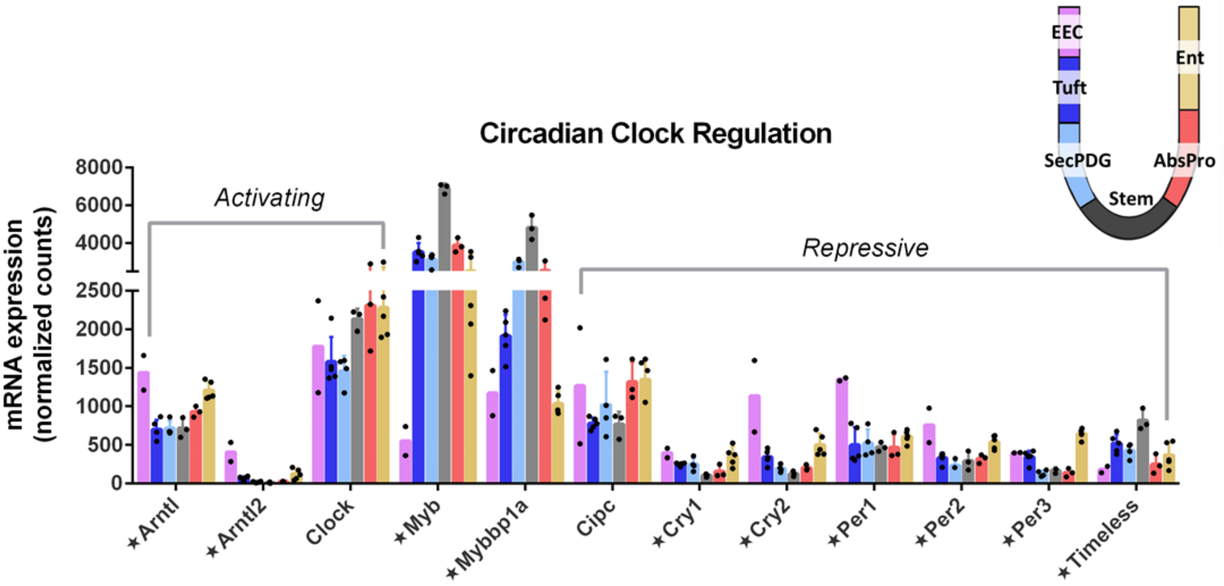
(Figure adapted from Supplementary Fig. 15 of Habowski et al., 2020; Communications Biology)


a
P53
Colon
Rectum
Small Intestine
Duodenum
b
Hnrnpk
Protein
RNA
RNA (normalized counts)
Protein (XIC areas)
4000
3000
2000
1000
0
4.0×10⁸
3.0×10⁸
2.0×10⁸
1.0×10⁸
0
*EEC
Tuft
SecPDG
Stem
AbsPro
*Ent
c
Bax
Protein
RNA
500
400
300
200
100
0
1.5×10⁷
1.0×10⁷
5.0×10⁶
0
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
Casp3
1500
1000
500
0
1.5×10⁷
1.0×10⁷
5.0×10⁶
0
SecPDG
Stem
AbsPro
Bcl2l14
3000
2000
1000
0
2.0×10⁷
1.5×10⁷
1.0×10⁷
5.0×10⁶
0
*Tuft
SecPDG
Stem
AbsPro
Ent
Bclaf1
4500
4000
3500
3000
2500
2000
1.5×10⁷
1.0×10⁷
5.0×10⁶
0
EEC
Tuft
SecPDG
Stem
AbsPro
*Ent

Supplementary Figure 2.16: ***Apoptosis related gene expression in the intestinal crypt.***

a Human protein atlas images showing TP53 protein levels are highest at the base of crypts in the colon, rectum, small intestine, and duodenum. **b** Protein and mRNA expression of Hnrnpk, a pre-mRNA binding protein. Hnrnpk acts as a transcriptional co-activator of p53 when SUMOylated [93]. Protein and mRNA expression for additional apoptosis related genes are shown in **c**. Bax promotes activation of Casp3 (protein only detected in stem and progenitors) which triggers apoptosis. Bcl2l14 also contributes to apoptosis and Bclaf1 is a death-promoting transcriptional repressor. Star annotation by cell type symbolizes significant differential mRNA expression compared to stem ($padj < 0.01$). For differential mRNA expression analysis, the following independent biological replicate numbers were used: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2.

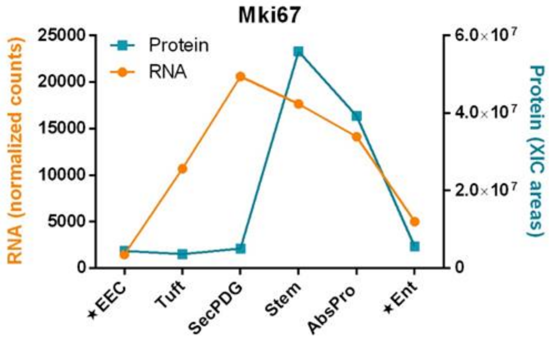
(Figure adapted from Supplementary Fig. 16 of Habowski et al., 2020; Communications Biology)

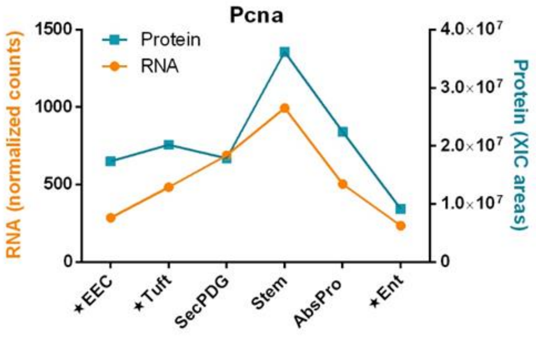


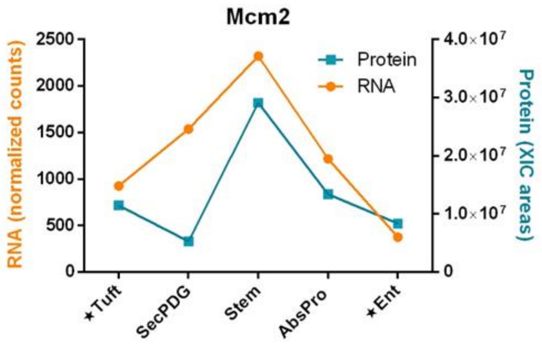
Supplementary Figure 2.17: ***Regulation of the circadian clock in the crypt.***

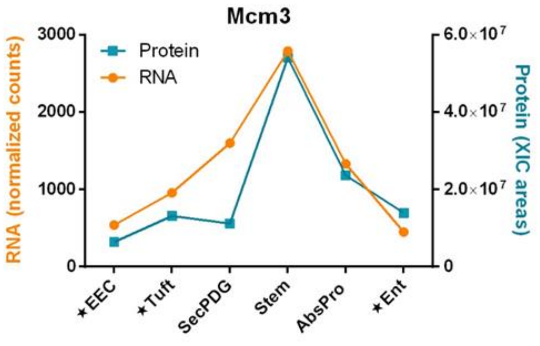
Arntl (Bmal1) and *Clock* are highly expressed in all cell types, including stem cells. In general, clock repressors are expressed at a lower level. Myb and Mybbp1a functions in the circadian clock are less well characterized, but expression is enriched in stem. Star annotation by gene name symbolizes significant differential mRNA expression in at least one cell type compared to stem (padj < 0.01) and error bars are standard deviation. For differential mRNA expression analysis, the following biological replicate numbers were used: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2.

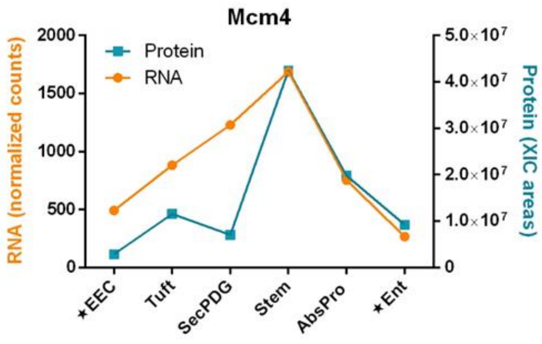
(Figure adapted from Supplementary Fig. 17 of Habowski et al., 2020; Communications Biology)

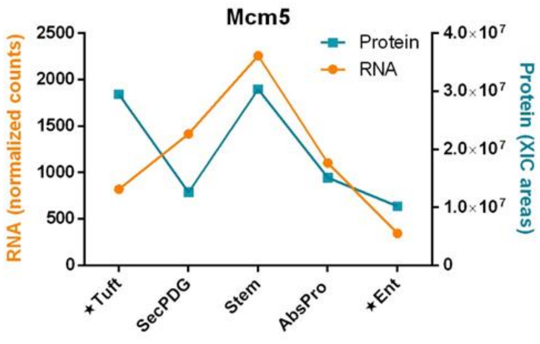

Mki67
Protein
RNA
RNA (normalized counts)
Protein (XIC areas)
25000
20000
15000
10000
5000
0
6.0×10⁷
4.0×10⁷
2.0×10⁷
0
★EEC
Tuft
SecPDG
Stem
AbsPro
★Ent

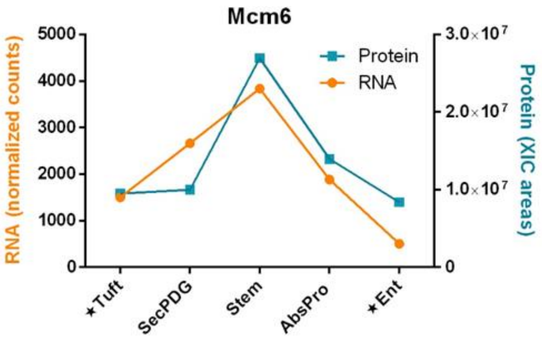

Pcna
Protein
RNA
RNA (normalized counts)
Protein (XIC areas)
1500
1000
500
0
4.0×10⁷
3.0×10⁷
2.0×10⁷
1.0×10⁷
0
★EEC
★Tuft
SecPDG
Stem
AbsPro
★Ent

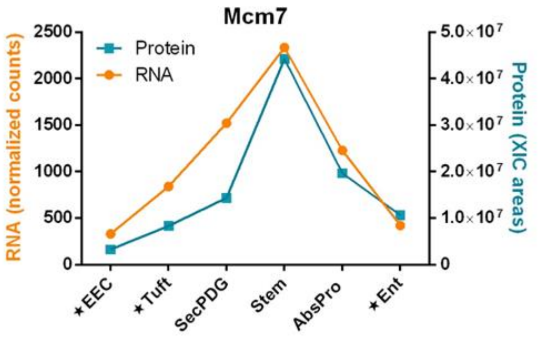

Mcm2
Protein
RNA
RNA (normalized counts)
Protein (XIC areas)
2500
2000
1500
1000
500
0
4.0×10⁷
3.0×10⁷
2.0×10⁷
1.0×10⁷
0
★Tuft
SecPDG
Stem
AbsPro
★Ent


Mcm3
Protein
RNA
RNA (normalized counts)
Protein (XIC areas)
3000
2000
1000
0
6.0×10⁷
4.0×10⁷
2.0×10⁷
0
★EEC
★Tuft
SecPDG
Stem
AbsPro
★Ent


Mcm4
Protein
RNA
RNA (normalized counts)
Protein (XIC areas)
2000
1500
1000
500
0
5.0×10⁷
4.0×10⁷
3.0×10⁷
2.0×10⁷
1.0×10⁷
0
★EEC
Tuft
SecPDG
Stem
AbsPro
★Ent


Mcm5
Protein
RNA
RNA (normalized counts)
Protein (XIC areas)
2500
2000
1500
1000
500
0
4.0×10⁷
3.0×10⁷
2.0×10⁷
1.0×10⁷
0
★Tuft
SecPDG
Stem
AbsPro
★Ent


Mcm6
Protein
RNA
RNA (normalized counts)
Protein (XIC areas)
5000
4000
3000
2000
1000
0
3.0×10⁷
2.0×10⁷
1.0×10⁷
0
★Tuft
SecPDG
Stem
AbsPro
★Ent

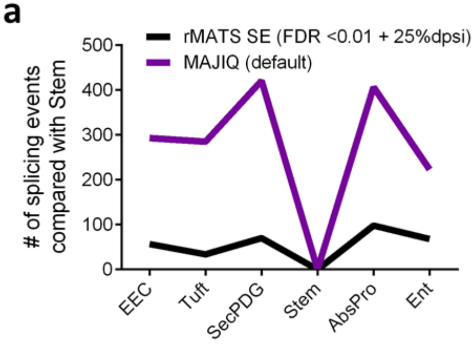

Mcm7
Protein
RNA
RNA (normalized counts)
Protein (XIC areas)
2500
2000
1500
1000
500
0
5.0×10⁷
4.0×10⁷
3.0×10⁷
2.0×10⁷
1.0×10⁷
0
★EEC
★Tuft
SecPDG
Stem
AbsPro
★Ent

Supplementary Figure 2.18: ***Expression of proliferation and cell cycle genes.***

Expression of Mki67 (mitotic chromosome stability), Pcna (DNA replication control), and Mcm2-7 (DNA replication licensing factors) are highest in both stem mRNA and protein. Star annotation by cell type symbolizes significant differential mRNA expression compared to stem ($padj < 0.01$). For differential mRNA expression analysis, the following biological replicate numbers were used: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2.

(Figure adapted from Supplementary Fig. 18 of Habowski et al., 2020; Communications Biology)

a



b

		Percent of each Splicing Event				
Cell Type (Compared with Stem)	Total Number of Splicing Events	SE	RI	MXE	A5SS	A3SS
AbsPro	1157	80.03	2.42	4.93	4.49	8.12
SecPDG	832	73.44	2.40	8.77	6.01	9.38
Tuft	542	73.99	1.85	11.25	6.09	6.83
Ent	791	74.08	3.03	9.23	5.56	8.09
EEC	337	64.09	6.53	9.50	6.82	13.06
	Average Percent	73.13	3.25	8.74	5.80	9.09

c

Comparison	Gene Expression (Padj <0.01)	mRNA Splicing - SE Events (FDR <0.05)	SAR = Splicing : Gene ratio x 100
Stem to AbsPro	301	926	308
AbsPro to Ent	5,201	489	9
Stem to Ent	7,138	586	8
Stem to SecPDG	1,669	611	37
SecPDG to Tuft	1,546	138	9
SecPDG to EEC	8,887	137	2
Stem to Tuft	3,615	401	11
Stem to EEC	8,417	216	3

d

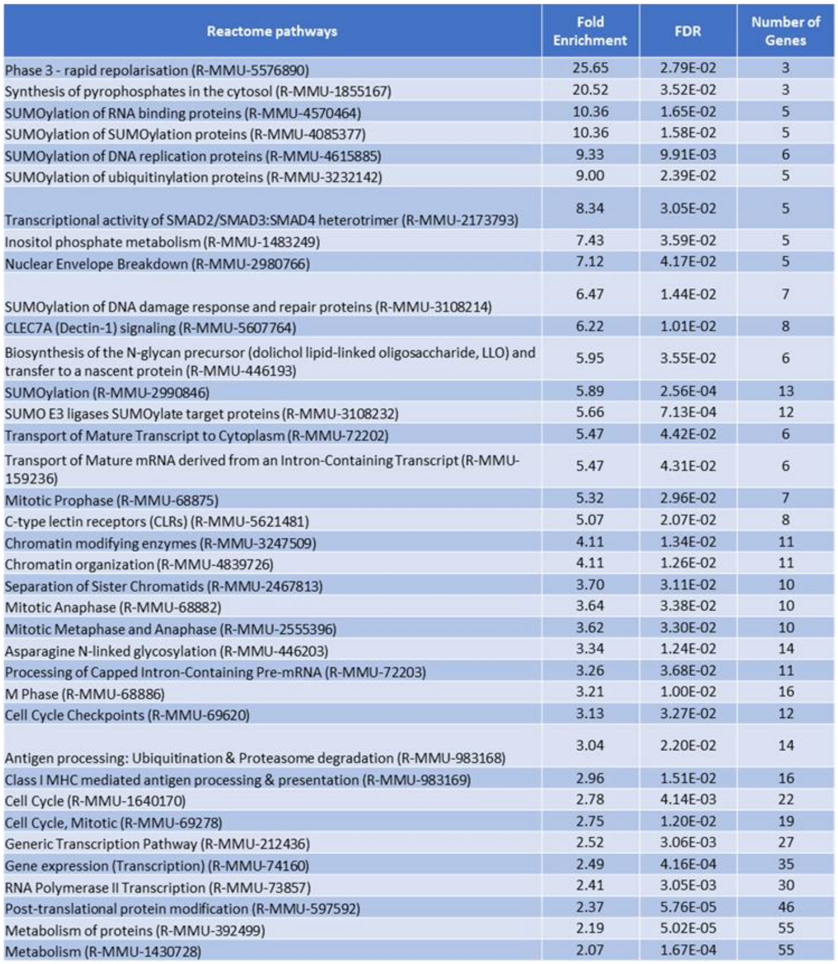
Comparison	Gene Expression (Padj <0.01)	mRNA Polyadenylation Events (Padj <0.05)	PAR = Poly(A) : Gene ratio x 100
Stem to AbsPro	301	795	264
AbsPro to Ent	5,201	185	4
Stem to Ent	7,138	217	3
Stem to SecPDG	1,669	657	39
SecPDG to Tuft	1,546	302	20
SecPDG to EEC	8,887	415	5
Stem to Tuft	3,615	401	11
Stem to EEC	8,417	467	6

Supplementary Figure 2.19: ***Alternative mRNA processing changes in the intestinal crypt.***

a Alternative splicing events in non-stem cells compared to stem cells as determined by an additional splicing pipeline, MAJIQ, or stringent filtering with rMATS. **b** Table showing the total number of splicing events detected by rMATS in each cell type compared to stem cells. There are a total of five different types of splicing events (SE = Skipped Exon, RI = Retained Intron, MXE = Mutually Exclusive Exon, A5SS = Alt 5 Splice Site, A3SS = Alt 3 Splice Site). The table provides the averages of each type of splicing event shown in the Figure 2.3b pie chart. SE events are the most abundant type of splicing event. **c** Splicing Abundance Ration (SAR) and **d** Polyadenylation Abundance Ratio (PAR) are metrics that report the relative abundance of specific RNA processing changes relative to changes in mRNA levels (i.e. gene expression). Statistical cut-offs for gene expression (padj < 0.01 + minimum mean 50 counts) and splicing (FDR < 0.05) or polyadenylation (padj < 0.01) values were used to generate the SAR and PAR values in Figure 2.3c and 2.3f, respectively. For differential mRNA expression and processing analyses, the following biological replicate numbers were used: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2.

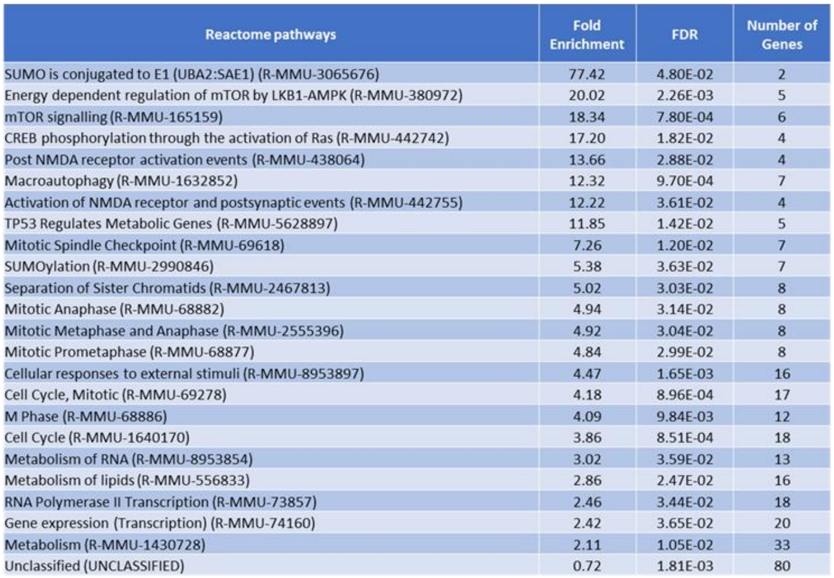
(Figure adapted from Supplementary Fig. 19 of Habowski et al., 2020; Communications Biology)

a **Splicing**



Reactome pathways	Fold Enrichment	FDR	Number of Genes
Phase 3 - rapid repolarisation (R-MMU-5576890)	25.65	2.79E-02	3
Synthesis of pyrophosphates in the cytosol (R-MMU-1855167)	20.52	3.52E-02	3
SUMOylation of RNA binding proteins (R-MMU-4570464)	10.36	1.65E-02	5
SUMOylation of SUMOylation proteins (R-MMU-4085377)	10.36	1.58E-02	5
SUMOylation of DNA replication proteins (R-MMU-4615885)	9.33	9.91E-03	6
SUMOylation of ubiquitinylation proteins (R-MMU-3232142)	9.00	2.39E-02	5
Transcriptional activity of SMAD2/SMAD3:SMAD4 heterotrimer (R-MMU-2173793)	8.34	3.05E-02	5
Inositol phosphate metabolism (R-MMU-1483249)	7.43	3.59E-02	5
Nuclear Envelope Breakdown (R-MMU-2980766)	7.12	4.17E-02	5
SUMOylation of DNA damage response and repair proteins (R-MMU-3108214)	6.47	1.44E-02	7
CLEC7A (Dectin-1) signaling (R-MMU-5607764)	6.22	1.01E-02	8
Biosynthesis of the N-glycan precursor (dolichol lipid-linked oligosaccharide, LLO) and transfer to a nascent protein (R-MMU-446193)	5.95	3.55E-02	6
SUMOylation (R-MMU-2990846)	5.89	2.56E-04	13
SUMO E3 ligases SUMOylate target proteins (R-MMU-3108232)	5.66	7.13E-04	12
Transport of Mature Transcript to Cytoplasm (R-MMU-72202)	5.47	4.42E-02	6
Transport of Mature mRNA derived from an Intron-Containing Transcript (R-MMU-159236)	5.47	4.31E-02	6
Mitotic Prophase (R-MMU-68875)	5.32	2.96E-02	7
C-type lectin receptors (CLRs) (R-MMU-5621481)	5.07	2.07E-02	8
Chromatin modifying enzymes (R-MMU-3247509)	4.11	1.34E-02	11
Chromatin organization (R-MMU-4839726)	4.11	1.26E-02	11
Separation of Sister Chromatids (R-MMU-2467813)	3.70	3.11E-02	10
Mitotic Anaphase (R-MMU-68882)	3.64	3.38E-02	10
Mitotic Metaphase and Anaphase (R-MMU-2555396)	3.62	3.30E-02	10
Asparagine N-linked glycosylation (R-MMU-446203)	3.34	1.24E-02	14
Processing of Capped Intron-Containing Pre-mRNA (R-MMU-72203)	3.26	3.68E-02	11
M Phase (R-MMU-68886)	3.21	1.00E-02	16
Cell Cycle Checkpoints (R-MMU-69620)	3.13	3.27E-02	12
Antigen processing: Ubiquitination & Proteasome degradation (R-MMU-983168)	3.04	2.20E-02	14
Class I MHC mediated antigen processing & presentation (R-MMU-983169)	2.96	1.51E-02	16
Cell Cycle (R-MMU-1640170)	2.78	4.14E-03	22
Cell Cycle, Mitotic (R-MMU-69278)	2.75	1.20E-02	19
Generic Transcription Pathway (R-MMU-212436)	2.52	3.06E-03	27
Gene expression (Transcription) (R-MMU-74160)	2.49	4.16E-04	35
RNA Polymerase II Transcription (R-MMU-73857)	2.41	3.05E-03	30
Post-translational protein modification (R-MMU-597592)	2.37	5.76E-05	46
Metabolism of proteins (R-MMU-392499)	2.19	5.02E-05	55
Metabolism (R-MMU-1430728)	2.07	1.67E-04	55

b **APA**



Reactome pathways	Fold Enrichment	FDR	Number of Genes
SUMO is conjugated to E1 (UBA2:SAE1) (R-MMU-3065676)	77.42	4.80E-02	2
Energy dependent regulation of mTOR by LKB1-AMPK (R-MMU-380972)	20.02	2.26E-03	5
mTOR signalling (R-MMU-165159)	18.34	7.80E-04	6
CREB phosphorylation through the activation of Ras (R-MMU-442742)	17.20	1.82E-02	4
Post NMDA receptor activation events (R-MMU-438064)	13.66	2.88E-02	4
Macroautophagy (R-MMU-1632852)	12.32	9.70E-04	7
Activation of NMDA receptor and postsynaptic events (R-MMU-442755)	12.22	3.61E-02	4
TP53 Regulates Metabolic Genes (R-MMU-5628897)	11.85	1.42E-02	5
Mitotic Spindle Checkpoint (R-MMU-69618)	7.26	1.20E-02	7
SUMOylation (R-MMU-2990846)	5.38	3.63E-02	7
Separation of Sister Chromatids (R-MMU-2467813)	5.02	3.03E-02	8
Mitotic Anaphase (R-MMU-68882)	4.94	3.14E-02	8
Mitotic Metaphase and Anaphase (R-MMU-2555396)	4.92	3.04E-02	8
Mitotic Prometaphase (R-MMU-68877)	4.84	2.99E-02	8
Cellular responses to external stimuli (R-MMU-8953897)	4.47	1.65E-03	16
Cell Cycle, Mitotic (R-MMU-69278)	4.18	8.96E-04	17
M Phase (R-MMU-68886)	4.09	9.84E-03	12
Cell Cycle (R-MMU-1640170)	3.86	8.51E-04	18
Metabolism of RNA (R-MMU-8953854)	3.02	3.59E-02	13
Metabolism of lipids (R-MMU-556833)	2.86	2.47E-02	16
RNA Polymerase II Transcription (R-MMU-73857)	2.46	3.44E-02	18
Gene expression (Transcription) (R-MMU-74160)	2.42	3.65E-02	20
Metabolism (R-MMU-1430728)	2.11	1.05E-02	33
Unclassified (UNCLASSIFIED)	0.72	1.81E-03	80

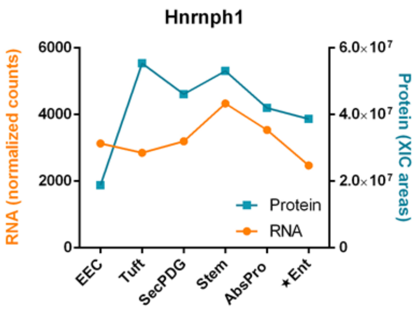
Supplementary Figure 2.20: ***Gene ontology analysis of alternatively processed mRNA in stem cells and progenitors.***

a Complete list of enriched gene ontologies for n=332 genes that are alternatively processed (spliced) in common between stem versus AbsPro and stem versus SecPDG. **b** Complete list of enriched gene ontologies for n=194 genes that are alternatively polyadenylated in common between stem versus AbsPro and stem versus SecPDG. Shortened list of ontology is presented in Figure 2.4c.

(Figure adapted from Supplementary Fig. 20 of Habowski et al., 2020; Communications Biology)

a

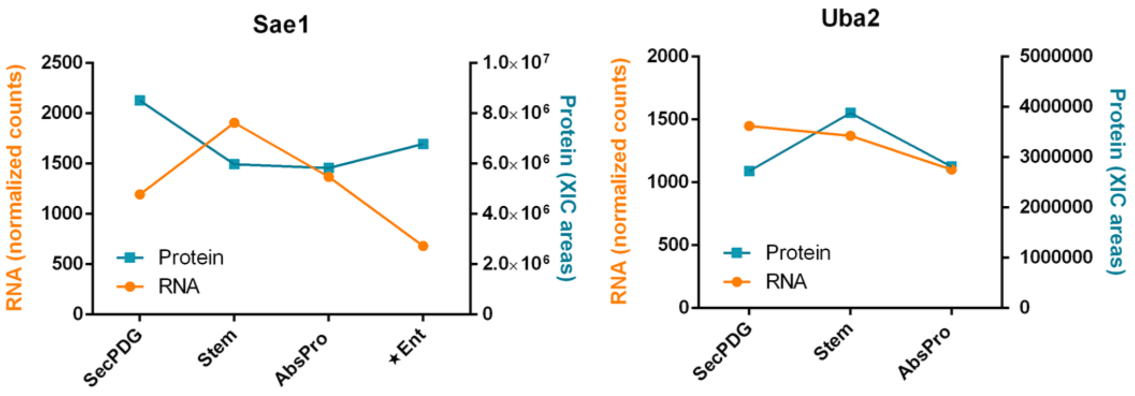
	In AbsPro + SecPDG			
Gene (n=13)	**SE event**	**terminal exon?**	**APA loci**	**Note:**
Metap1	chr3:138466223-138466367	Yes*	chr3:138458960-138460500	*Terminal exon for shorter isoform, distinct from APA loci
Hnrnph1	chr11:50385140-50385190, chr11:50381458-50381597	No	chr11:50385777-50386528	Second to last exon
Mknk2	chr10:80671927-80672015	No	chr10:80665318-80667332	
Muc4	chr16:32763910-32763999	No	chr16:32781826-32782390	
Ngly1	chr14:16260563-16260800	No	chr14:16310859-16311537	
Ranbp2	chr10:58453311-58453423	No	chr10:58493663-58494154	
Rexo4	chr2:26960223-26960417	No	chr2:26953564-26954596	
Rnps1	chr17:24418037-24418218	No	chr17:24425086-24425897	
Slc25a17	chr15:81327216-81327362	No	chr15:81318921-81319776	
Tlk1	chr2:70786876-70786995	No	chr2:70712539-70714165	
4833439L19Rik	chr13:54559197-54559324	No	chr13:54551290-54552817	
Akap9	chr5:4028373-4028536, chr5:4003560-4003782	No	chr5:4079659-4080204	
Cc2d1a	chr8:84146313-84146449	No	chr8:84132828-84133257	



b

Polyadenylation Events	**Distal PolyA Usage of Replicates**	**Average Distal PolyA Usage**	**Loci Coordinates**	**Padj (Compared with Stem)**	**Gene Expression Change (Compared with Stem)**
Sae1	Stem: 0.79,0.51,0.40	0.57	chr7:16327051-16327929		
	AbsPro: 0.21,0.18,0.80	0.40		7.66E-06	ns
	SecPDG: 0.00,0.52,0.41,0.52	0.36		1.63E-03	Padj = 0.011
Uba2	Stem: 0.42,0.75,0.46	0.53	chr7:34140697-34141466		
	AbsPro: 0.23,0.10,0.54	0.29		4.51E-08	ns
	SecPDG: 0.09,0.52,0.23,0.48	0.33		4.32E-05	ns
Splicing Events	**Inclusion Rates of Replicates**	**Average Inclusion Rate**	**Exon Coordinates**	**FDR (Compared to Stem)**	**Gene Expression Change (Compared with Stem)**
Sae1	Stem: 1.00, 1.00, 1.00	1.00	chr7: 16327830-16327929		
	AbsPro: 0.273, 1.00, 1.00	0.76		1.80E-02	ns
	SecPDG: 1.00, 1.00, 0.867	0.96		ns	Padj = 0.011

c



Supplementary Figure 2.21: ***Common mRNA processing events.***

a Table shows the n=13 genes (in Figure 2.4b Venn diagram) that are both alternatively spliced and polyadenylated in stem versus AbsPro and stem versus SecPDG. Coordinates for the events are given and show there is not an overlap in the processing event location in each gene confirming the splicing and polyadenylation events are distinct. Hnrnph1 was detected by MS, and the protein and mRNA levels are shown in the graph (Right). **b** Ontology analysis (Figure 2.4c) shows that SUMOylation processes impacted by both splicing and polyadenylation. The polyadenylation and splicing events of *Sae1* and *Uba2* are shown in the table. **c** mRNA and protein expression (when detected) for Sae1 and Uba2 are graphed. Star annotation by cell type symbolizes significant differential mRNA expression compared to stem (padj < 0.01). For differential mRNA expression and processing analyses, the following biological replicate numbers were used: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2.

(Figure adapted from Supplementary Fig. 21 of Habowski et al., 2020; Communications Biology)

a

Splicing Events	Inclusion Rates of Replicates	Average Inclusion Rate	Exon Coordinates	FDR (Compared to Stem)	Gene Expression Change (Compared with Stem)
Ctnnd1	Stem: 0.954,0.786,0.837	0.86	chr2: 84605183-84605246		
	AbsPro: 0.531,0.179,0.49	0.40		2.98E-08	ns
	SecPDG: 0.573,0.696,0.181,0.598	0.51		2.91E-05	ns
Spen	Stem: 0.145,0.976,0.57	0.56	chr4: 141485558-141485672		
	AbsPro: 1.0,1.0,1.0	1.00		4.32E-06	ns
	SecPDG: 1.0,0.916,0.952,0.94	0.95		7.45E-04	ns
Cbfa2t2	Stem: 1.0,1.0,0.9	0.97	chr2: 154500399-154500543		
	AbsPro: 0.449,0.0,0.57	0.34		8.56E-08	ns
	SecPDG: 0.874, 0.55	0.71		ns	ns
Eif4a2	Stem: 0.467,0.388,0.362	0.41	chr16: 23112350-23112457		
	AbsPro: 0.578,0.337,0.462	0.46		ns	ns
	SecPDG: 0.0,0.13,0.072,0.137	0.08		0.00E+00	ns

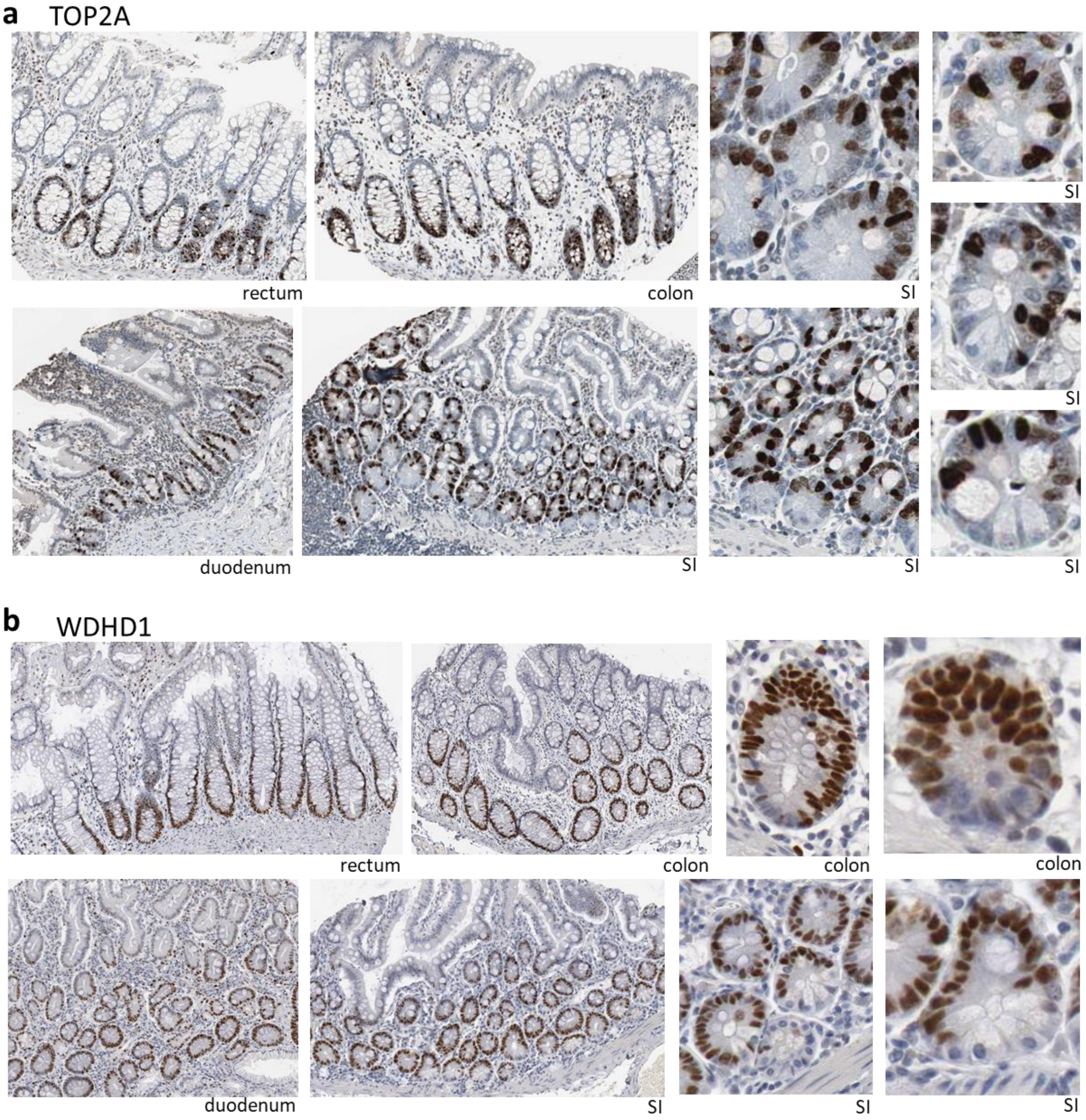
b

Polyadenylation Events	Distal PolyA Usage of Replicates	Average Distal PolyA Usage	Loci Coordinates	Padj (Compared with Stem)	Gene Expression Change (Compared with Stem)
Top2a	Stem: 0.1,0.12,0.36	0.19	chr11:98992947-98993413		
	AbsPro: 0.54,0.58,0.36	0.49		1.07E-05	ns. Padj= 0.06, log2fc = 1 (elevated in Stem)
	SecPDG: 0.18,0.5,0.8,0.35	0.46		1.88E-05	ns
Wdhd1	Stem: 0.19,0.21,0.11	0.17	chr14:47240944-47241812		
	AbsPro: 0.74,0.54,0.22	0.50		1.08E-03	ns
	SecPDG:1,0.16,0.38	0.51		9.70E-07	ns
Cby1	Stem: 0.16,0.55,0.42	0.38	chr15:79666988-79667660		
	AbsPro: 0.81,0.79,0.65	0.75		6.31E-09	ns
	SecPDG:1,0.5,0.59,0.48	0.64		4.95E-03	ns
Rbm3	Stem: 0.21,0.63,0.63	0.49	chrX:8142356-8142955		
	AbsPro: 0.7,0.93,0.78	0.80		9.83E-16	ns
	SecPDG: 0.31,0.67,0.92,0.59	0.62		ns	ns
Ihh	Stem: 0.37,0.41,0.45	0.41	chr1:74945315-74946747		
	AbsPro:0.66,0.35,0.47	0.49		ns	ns
	SecPDG: 0.45,0.62,0.8,0.7	0.67		9.66E-06	ns

Supplementary Figure 2.22: ***Expression and processing data for examples of alternatively processed genes.***

a Two splicing events that occur in both stem versus AbsPro and stem versus SecPDG (*Ctnnd1* and *Spen*), and two processing events that are unique to either stem versus Abspro (*Cbfa2t2*), or stem versus SecPDG (*Eif4a2*). Table lists inclusion rates for the skipped exon, the average rates of the inclusion, coordinates of skipped exon along with the splicing event significance. No significant changes in mRNA levels occurred for any of these genes when compared to the stem cell population. **b** Three alternative polyadenylation events that occur in both stem versus AbsPro and stem versus SecPDG (*Top2a, Wdhd1,* and *Cby1*) are listed. Two events that are unique to stem versus AbsPro (*Rbm3*) or are unique to stem versus SecPDG (*Ihh*) are also listed. Table includes rates of distal polyA usage in replicates, average distal polyA use, chromosome location of the polyA event within the gene locus, and polyA event significance. No significant gene expression (mRNA level) change occurred in any of these genes compared with stem, although *Top2a* expression was increased in stem. For differential mRNA expression and processing analyses, the following biological replicate numbers were used: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2.

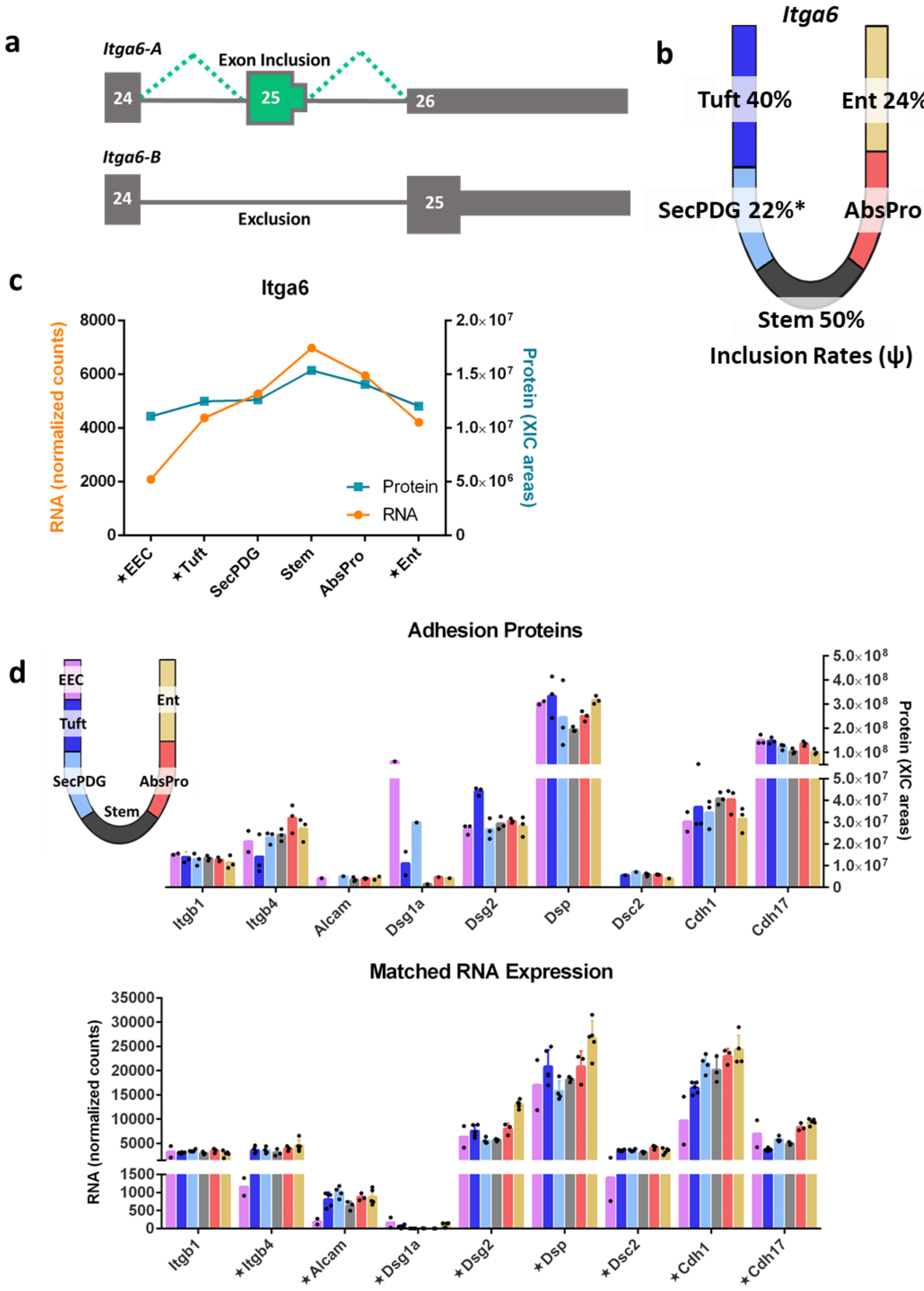
(Figure adapted from Supplementary Fig. 22 of Habowski et al., 2020; Communications Biology)



Supplementary Figure 2.23: ***Intestinal staining of alternatively polyadenylated genes in human intestinal tissues.***

Both **a** *Top2a* and **b** *Wdhd1* were defined as alternatively polyadenylated in stem versus AbsPro and stem versus SecPDG in mouse colon. Even though *Wdhd1* and *Top2a* (padj = 0.06) had modest increases in mRNA levels in stem, minimal staining was observed at the base of the stem cell niche. Both TOP2A and WDHD1 staining appear to be limited to the transit amplifying zone with no evidence of expression in differentiated cells towards the top of the crypt.

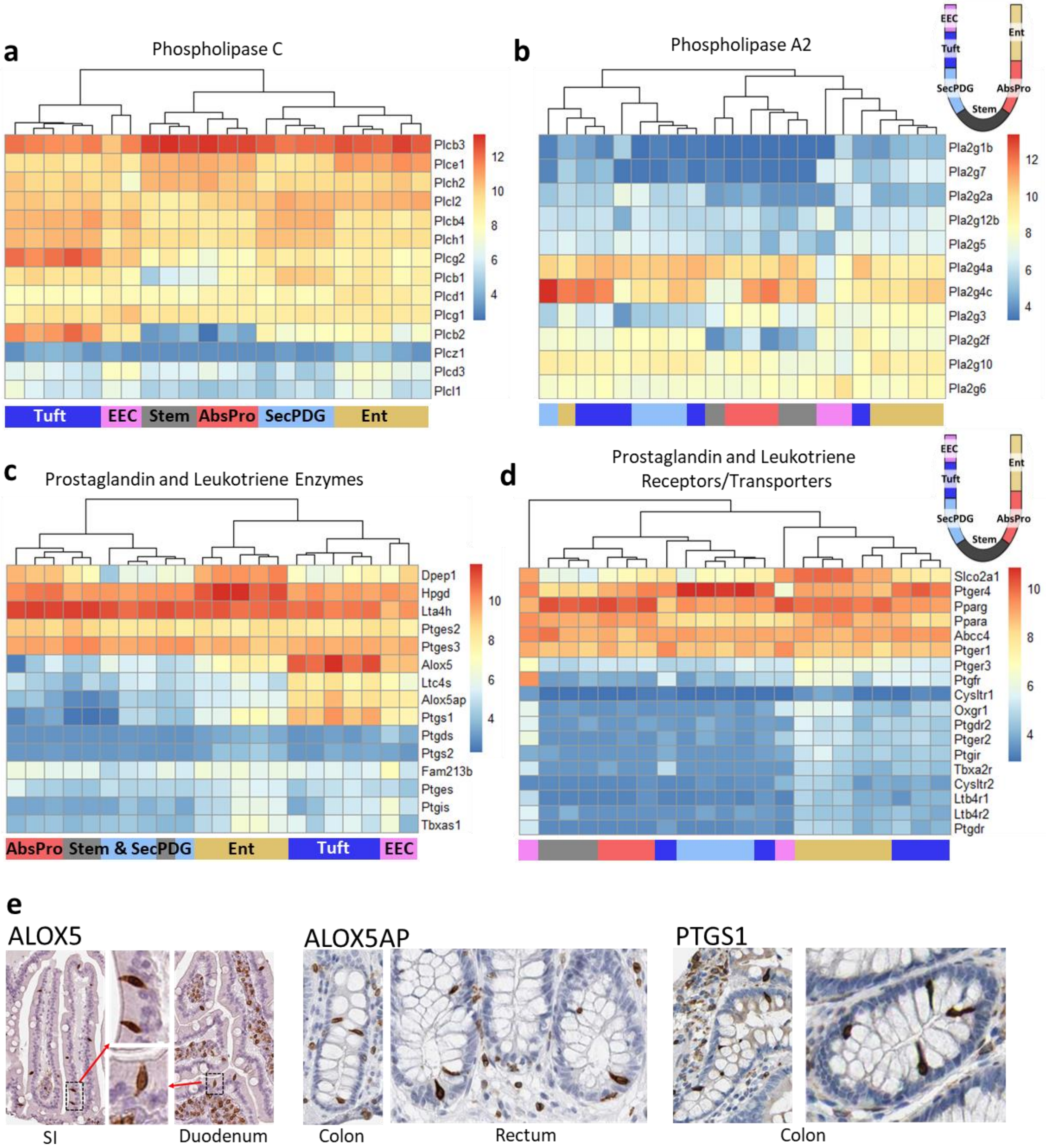
(Figure adapted from Supplementary Fig. 23 of Habowski et al., 2020; Communications Biology)


a
Itga6-A
Exon Inclusion
24
25
26
Itga6-B
24
25
Exclusion
b
Itga6
Tuft 40%
Ent 24%*
SecPDG 22%*
AbsPro 39%
Stem 50%
Inclusion Rates (ψ)
c
Itga6
RNA (normalized counts)
Protein (XIC areas)
8000
6000
4000
2000
0
2.0×10⁷
1.5×10⁷
1.0×10⁷
5.0×10⁶
0
Protein
RNA
★EEC
★Tuft
SecPDG
Stem
AbsPro
★Ent
d
Adhesion Proteins
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
Protein (XIC areas)
5.0×10⁸
4.0×10⁸
3.0×10⁸
2.0×10⁸
1.0×10⁸
5.0×10⁷
4.0×10⁷
3.0×10⁷
2.0×10⁷
1.0×10⁷
0
Itgb1
Itgb4
Alcam
Dsg1a
Dsg2
Dsp
Dsc2
Cdh1
Cdh17
Matched RNA Expression
RNA (normalized counts)
35000
30000
25000
20000
15000
10000
5000
1500
1000
500
0
Itgb1
★Itgb4
★Alcam
★Dsg1a
★Dsg2
★Dsp
★Dsc2
★Cdh1
★Cdh17

Supplementary Figure 2.24: ***Expression and splicing of adhesion genes.***

a Alternative splicing of exon 25 in *Itga6* produces two known, distinct isoforms. **b** Exon inclusion is at its highest level in stem cells compared to progenitors and differentiated cells. Inclusion rates were lowest in SecPDG and Ent (* marks significant decrease compared to stem; FDR < 0.05). **c** Itga6 protein was detected via MS in all cell types and compared to its cognate mRNA expression. Star annotation by cell type symbolizes significant differential mRNA expression compared to stem (padj < 0.01). **d** Additional adhesion proteins were detected and matched with mRNA expression. With the exception of Dsg1a and Dsg2, the expression level of each gene is similar across all cell types. There are however, striking differences in the ratio of mRNA:protein. For example: *Dsc2* mRNA is abundant, but there is very little protein; *Cdh1* mRNA is much higher than *Cdh17*, but there is more Cdh17 protein than Cdh1. Star annotation by gene name symbolizes significant differential mRNA expression in at least one cell type compared to stem (padj < 0.01) and error bars are standard deviation. For proteomics there are n=3 independent biological replicates for each cell type; for mRNA differential expression analysis the following biological replicate numbers were used: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2.

(Figure adapted from Supplementary Fig. 24 of Habowski et al., 2020; Communications Biology)


a
Phospholipase C
Plcb3
Plce1
Plch2
Plcl2
Plcb4
Plch1
Plcg2
Plcb1
Plcd1
Plcg1
Plcb2
Plcz1
Plcd3
Plcl1
12
10
8
6
4
Tuft
EEC
Stem
AbsPro
SecPDG
Ent
b
Phospholipase A2
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
Pla2g1b
Pla2g7
Pla2g2a
Pla2g12b
Pla2g5
Pla2g4a
Pla2g4c
Pla2g3
Pla2g2f
Pla2g10
Pla2g6
12
10
8
6
4
c
Prostaglandin and Leukotriene Enzymes
Dpep1
Hpgd
Lta4h
Ptges2
Ptges3
Alox5
Ltc4s
Alox5ap
Ptgs1
Ptgds
Ptgs2
Fam213b
Ptges
Ptgis
Tbxas1
10
8
6
4
AbsPro
Stem & SecPDG
Ent
Tuft
EEC
d
Prostaglandin and Leukotriene Receptors/Transporters
EEC
Tuft
SecPDG
Stem
AbsPro
Ent
Slco2a1
Ptger4
Pparg
Ppara
Abcc4
Ptger1
Ptger3
Ptgfr
Cysltr1
Oxgr1
Ptgdr2
Ptger2
Ptgir
Tbxa2r
Cysltr2
Ltb4r1
Ltb4r2
Ptgdr
10
8
6
4
e
ALOX5
SI
Duodenum
ALOX5AP
Colon
Rectum
PTGS1
Colon

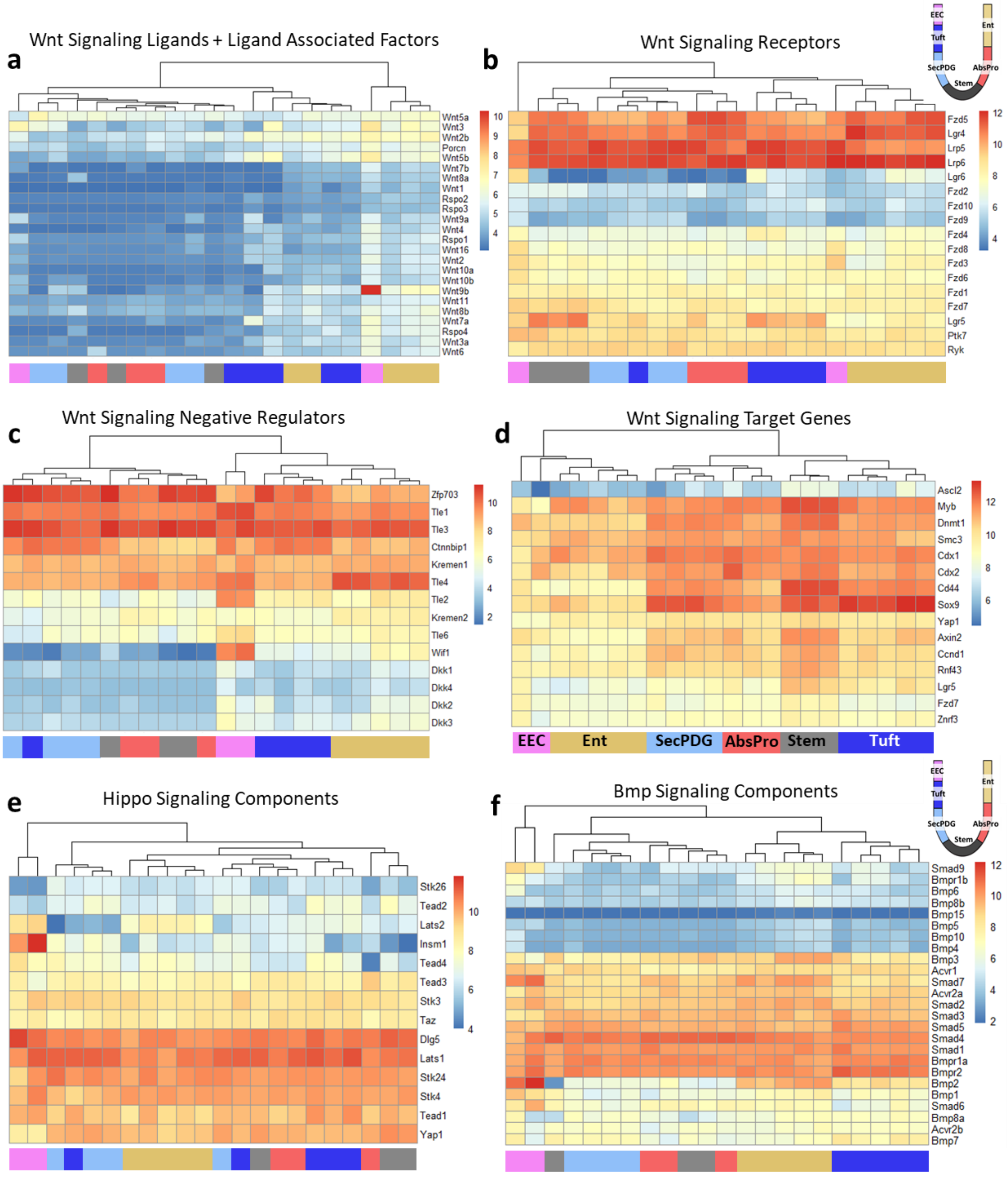
Supplementary Figure 2.25: ***Prostaglandin and Leukotriene production and signaling in the intestinal crypt.***

Diacylglycerol or phospholipids are converted to arachidonic acid by **a** Phospholipase C or **b** Phospholipase A2 enzymes. **c** Arachidonic acid is then provided to the leukotriene synthesis pathway (*Alox5, Alox5ap, Ltc4s*) or the prostaglandin synthesis pathway (*Ptgs1, Ptgs2*). Additional enzymes create different forms of prostaglandins and leukotrienes. **d** Prostaglandins and leukotrienes are often used to signal to non-epithelial cells, but there are some receptors and transporter expression in the crypt. **e** tuft cells are unique in the crypt epithelium in their ability to catalyze the initial conversion of arachidonic acid, which is confirmed by human protein atlas staining. See *Supplementary Discussion* section for additional information.

Discussion for Supplementary Figure 2.25:

Arachidonic acid is the molecular precursor for both prostaglandins and leukotrienes, each of which are potent inflammatory mediators for immune responses. Interestingly, all cell populations express high levels of *Plcb3*, an enzyme capable of producing arachidonic acid (Figure 2.6d), but only tuft cells express the machinery needed to convert arachidonic acid to leukotriene C4 (*Alox5*, *Alox5ap* (Flap), and *Ltc4s*; Supplementary Figure 2.18c). Further conversion of the leukotriene C4 precursor to D4 and E4 is possible in AbsPro and Ent which express high levels of *Dpep1*. This suggests that multiple cell populations cooperate with tuft cells to synthesize the full array of leukotriene mediators of inflammation. Unlike the prostaglandin pathway, the receptors for leukotrienes (Cysltr1 and Cysltr2) are minimally expressed in the epithelia underscoring how this signal is likely meant for cell populations in the stroma [94].

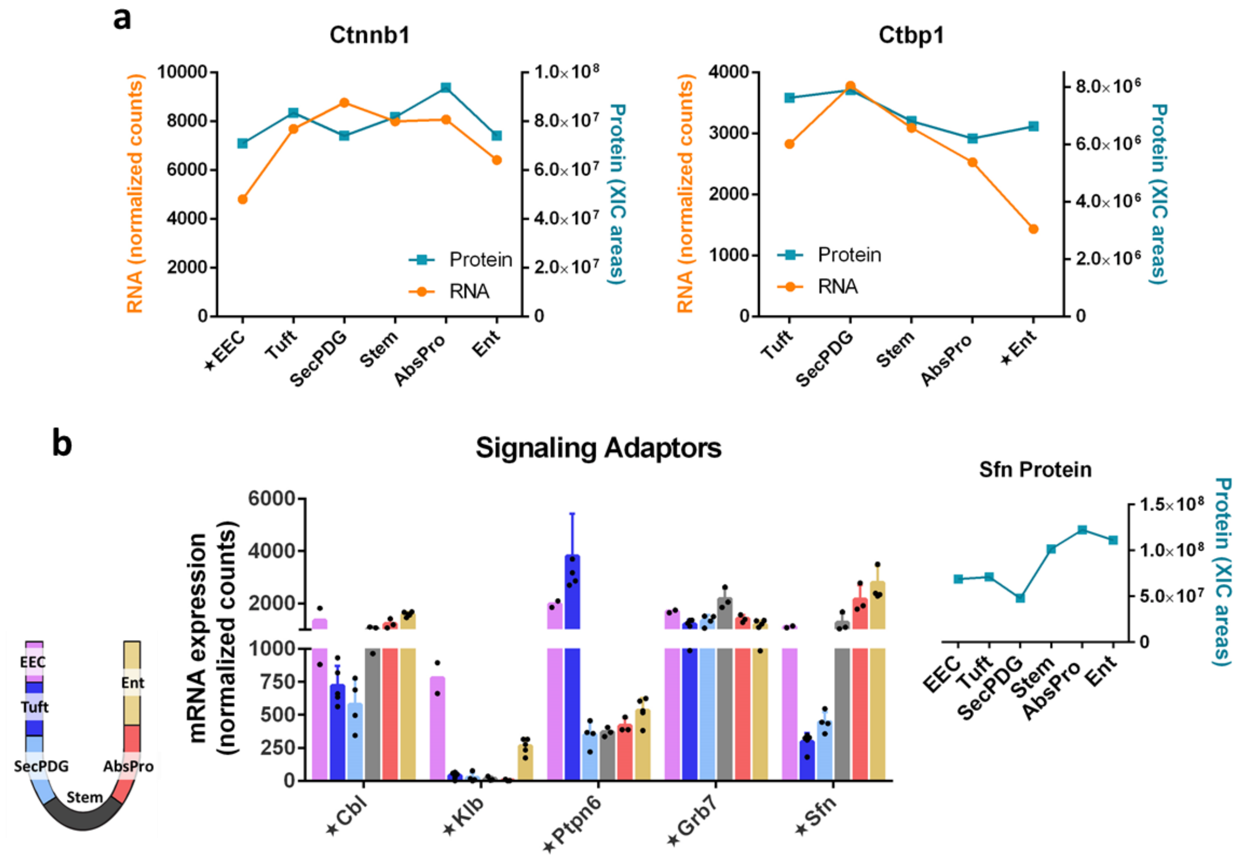
(Figure adapted from Supplementary Fig. 25 of Habowski et al., 2020; Communications Biology)



Supplementary Figure 2.26: ***Unsupervised clustering of crypt signaling pathway components.***

Wnt signaling **a** ligands and ligand associated factors, **b** receptors, **c** signaling inhibitors, and **d** Wnt signaling target genes. **e** Hippo signaling components. **f** Bmp signaling components.

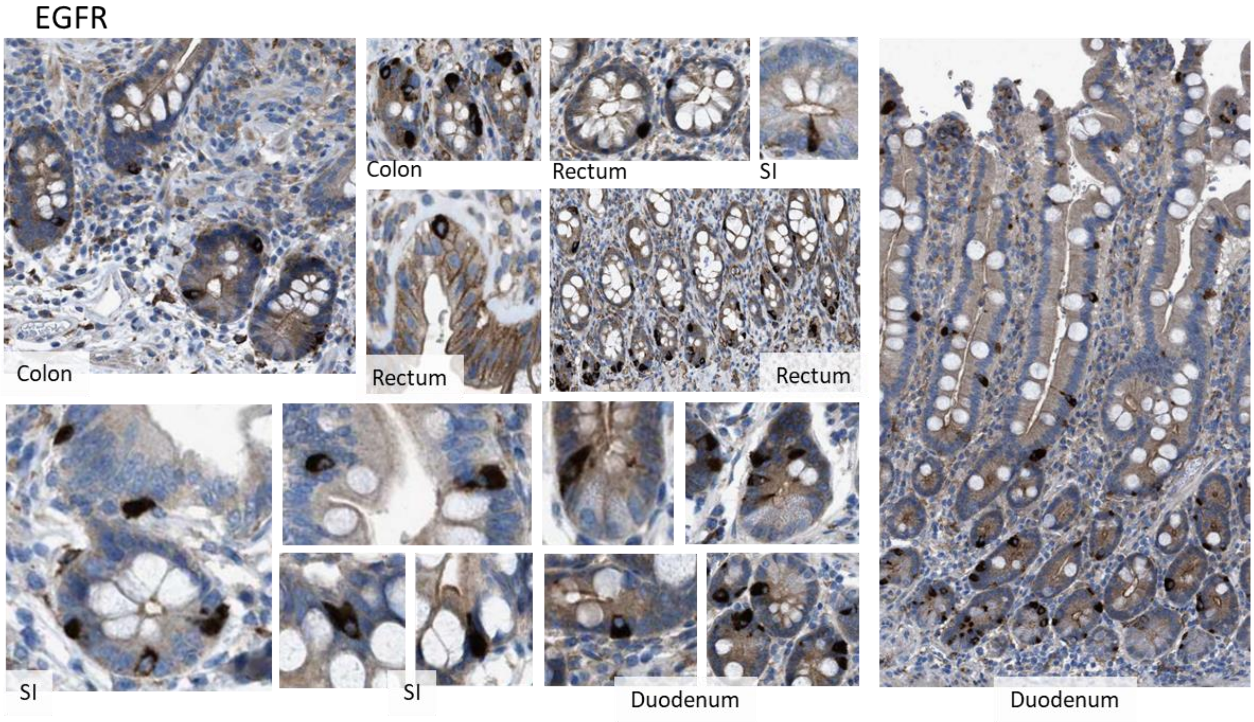
(Figure adapted from Supplementary Fig. 26 of Habowski et al., 2020; Communications Biology)



Supplementary Figure 2.27: ***Additional influences on signaling.***

a Although Wnt signaling is known to be elevated in the stem cell niche, the important mediator Ctnnb1 (β-catenin) was observed to be expressed in all cell types. Expression of co-repressor Ctbp1, which can also influence Wnt signaling, was also detected in all cell types. Star annotation by cell type symbolizes significant differential mRNA expression compared to stem (padj < 0.01). **b** Signaling adaptors can influence specific signaling pathways and we observed several that had unique expression. Cbl can influence Kit, Fgf, and Egf receptors, and Ptpn6 can also influence Kit and Egfr. We detected Sfn protein to be decreased in SecPDG (graph Right), consistent with mRNA expression. Star annotation by gene name symbolizes significant differential mRNA expression in at least one cell type compared to stem (padj < 0.01) and error bars are standard deviation. For mRNA differential expression analysis the following biological replicate numbers were used: stem=3, AbsPro=3, SecPDG=4, tuft=5, Ent=5, and EEC=2.

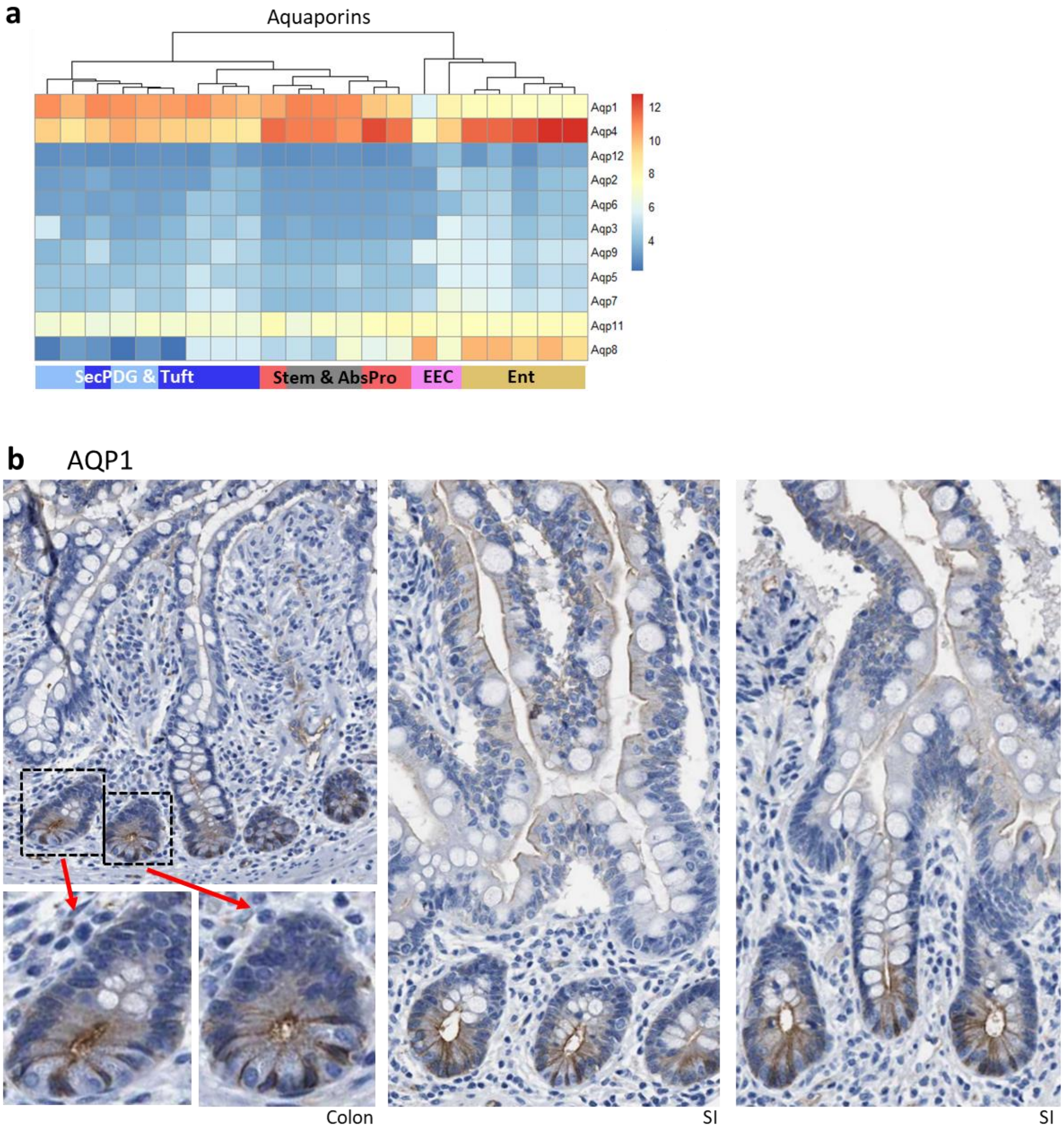
(Figure adapted from Supplementary Fig. 27 of Habowski et al., 2020; Communications Biology)



Supplementary Figure 2.28: ***Human intestinal expression of the EGFR.***

Human intestinal expression of the EGF receptor (EGFR; human protein atlas). EGFR expression appears to mark tuft and EEC-shaped cells in the rectum, colon, small intestine, and duodenum.

(Figure adapted from Supplementary Fig. 28 of Habowski et al., 2020; Communications Biology)



Supplementary Figure 2.29: ***Colonic expression of Aquaporins.***

a Heat map of unsupervised clustering showing expression of all aquaporin water channels. **b**. Note that AQP1 expression appears to be highly specific to stem cells at the crypt base in colon (including enlargement of crypt base), and small intestine.

(Figure adapted from Supplementary Fig. 29 of Habowski et al., 2020; Communications Biology)

Table 2.1: ***Genes enriched in stem and non-stem.***

Stem Enriched n=16 (Compared to all other cell types)	Stem Enriched n=27 (compared to SecPDG+AbsPro)	Non-Stem Enriched n=108 (what is elevated in non-Stem compared to Stem)		
Aatf	Aatf	Ano1	Gna14	Rassf6
Arid5b	Angpt2	Ano7	Gpt2	Rgs16
Axin2	Arid5b	Anxa13	Grpr	Rgs2
Cd44	Axin2	Arhgef37	Gsn	Rhoc
Cdk4	Bcl11a	Atp2a3	Guca2a	Rnf39
Ilf3	Cd44	Atp2c2	Hepacam2	S100a14
Lmnb1	Cdk4	Atp6v0d1	Hgfac	Scd1
Mybbp1a	Cep250	B3galt5	Hk1	Scnn1a
Nap1l1	Gar1	Best2	Hmgcr	Sct
Notch1	Ilf3	Cacna1h	Irf7	Sdcbp2
Prmt1	Kif26b	Camk2n1	Itih2	Selm
Rnf43	Lamc1	Capn13	Klf4	Sepp1
Trp53	Lgr5	Cbfa2t3	Lbh	Sh3pxd2a
Uaca	Lmnb1	Cbr2	Lrrc26	Sis
Whsc1	Mdn1	Cdhr5	Map1b	Slc17a9
Zbtb38	Mybbp1a	Chga	Mfhas1	Slc35b1
	Nap1l1	Chgb	Mfsd6	Slc41a2
	Notch1	Ckap4	Mia3	Slco2a1
	Patz1	Clec2h	Mical1	Sprr2a3
	Prmt1	Cmpk1	Mtmr6	St6gal1
	Rnf43	Cox7a1	Nab1	Stk38l
	Syde2	Cpeb2	Ndrg1	Susd6
	Trp53	Cpeb4	Ndrg1	Sybu
	Uaca	Creb3l2	Oas2	Synpo
	Whsc1	Cutal	Oas3	Syt7
	Zbtb38	Dap	Osbpl5	Sytl2
	Zfp629	Dock8	Parm1	Tff3
		Dst	Pck1	Tmc5
		Efnb2	Pde2a	Tmco3
		Fam102b	Plcb1	Tmem62
		Fam3b	Ppp1r3b	Tpsg1
		Fam3c	Prkaa2	Trp53inp2
		Fhl1	Ptprn2	Ttll7
		Fut2	Qsox1	Usp18
		Gde1	Ramp1	Ypel5
		Gm1123	Rasd2	Zbp1

(Table adapted from Supplementary Data 3 of Habowski et al., 2020; Communications Biology)

Table 2.2: ***Genes enriched in progenitors and absorptive and secretory lineages.***

AbsPro Enriched n=1	SecPDG Enriched n=23	Secretory Lineage (SecPDG + Tuft + EEC) n=50		Absorptive lineage (AbsPro + Ent) Enriched n=2
Trf	Aldh18a1	Agap1	Npdc1	Itih5
	Anpep	Amigo3	Ntn4	Slc6a14
	Bag3	Arfgef3	Pcsk1	
	Bcas1	Asah2	Pde2a	
	Cog1	Cbfa2t3	Plcb4	
	Dll1	Ccdc109b	Plet1	
	Ehd4	Cd164	Ppap2a	
	Fgfr2	Cd24a	Ptprn2	
	Hk2	Chn2	Rabep1	
	Hsd11b2	Cxxc5	Rasd2	
	Lbh	Dap	Rcan3	
	Lrig3	Dgkd	Rcn1	
	Lrrc26	Dpysl2	Rgs13	
	Ly6e	Eid1	Sh2d7	
	Myh9	Enpp4	Slc35g2	
	Pdia5	Etv5	Slc39a8	
	Rab15	Ffar2	Smpd1	
	Sec23ip	Foxa1	Stk38l	
	Surf4	Fxyd3	Suco	
	Tars	Gnai1	Syt7	
	Tmem109	Kctd12	Tbc1d16	
	Tmem97	Krt18	Tbc1d8	
	Tuba1c	Lrrc8d	Tox3	
		Maged1	Tpd52l1	
		Nbeal2	Tspan13	

(Table adapted from Supplementary Data 3 of Habowski et al., 2020; Communications Biology)

Table 2.3: ***Genes enriched in differentiated cells Tuft and Ent.***

Tuft Enriched n=51		**Ent Enriched n=88**		
1810046K07Rik	Pstpip2	43525	Ifit2	Slc6a14
Ackr4	Ptpn18	2200002D01Rik	Ifngr1	Slc6a8
Alox5	Ptprj	2310079G19Rik	Il22ra1	Slfn4
Anxa4	Rbm38	Acap2	Irf2	Specc1
Avil	Rgs13	Ahnak	Irf7	Spink1
Bcl2l14	Runx1	Anxa3	Itih5	Stk17b
Bub3	Sec14l1	Arih2	Kif13b	Stk25
Cables2	Siglecf	Atp12a	Klf4	Stom
Cd24a	St6galnac2	AW112010	Lamc2	Susd6
Chdh	Strip2	B630005N14Rik	Leng9	Sycn
Cib2	Suco	Bcar3	Lmo7	Tacc2
Dclk1	Svil	Birc3	Lrrfip1	Tgm3
Eppk1	Tmem176b	Capn13	Lypd8	Themis3
Fyb	Txndc16	Car4	Magi3	Tppp
Gga2	Vav1	Ccl25	Max	Trim12a
Il13ra1	Wnk2	Cd55	Mxd1	Trim30d
Inpp5d	Zfhx3	Cela1	Myo5c	Uqcrb
Jade1		Chmp2b	Naip1	Usp53
Jarid2		Ckmt1	Parp14	Xdh
Jmy		Clca4a	Pde7a	Zg16
Kctd12		Cxadr	Pfkp	
Kdm4a		Cyp2d34	Plac8	
Krt18		Ddx60	Pmaip1	
Krt23		Dsg2	Ppp1r3b	
Lrmp		Edn1	Prdx6	
Man2a1		Eif2s2	Reg3b	
Myo1b		Endod1	Rhbdl2	
Nav2		Eps8l3	Rnasel	
Nfe2l3		Fam129a	S100g	
Pgm2l1		Ggh	Saa1	
Pik3cg		Gm11538	Sat1	
Plcb2		Gyk	Sbds	
Plcg2		Hpgd	Slc40a1	
Plk2		Ifit1bl1	Slc5a3	

(Table adapted from Supplementary Data 3 of Habowski et al., 2020; Communications Biology)

Table 2.4: ***Genes enriched in differentiated cell type EECs.***

EEC Enriched n=430									
1190002N15Rik	Bmp2	Ddhd1	Fhl2	Hps1	Map7d1	Pde1c	Reps2	Snx30	Tph1
1700021K19Rik	Bsn	Dennd4a	Flnb	Hsd3b7	Mapk8ip1	Pde3b	Resp18	Sobp	Tppp3
1810006J02Rik	Btrc	Dennd5a	Fmn1	Hsf2	Mapk8ip3	Pdlim5	Ret	Spen	Trim3
2610035D17Rik	Cacna1a	Dhx40	Fmnl2	Iapp	Marcksl1	Peg13	Rfx2	Sqstm1	Trim39
5031426D15Rik	Cacna1h	Dlg4	Focad	Id2	Mast3	Peg3	Rgl2	Srgap1	Trim46
9530091C08Rik	Cacna2d1	Dlk1	Foxa1	Ids	Mast4	Pex5l	Rgs4	Srgap2	Trim62
Abca2	Cadm4	Dmpk	Foxj3	Ifi27	Mbd6	Phf2	Rimbp2	Ssfa2	Triobp
Acp2	Cadps	Dnaja4	Foxn3	Ift88	Mdm1	Phldb2	Rims2	Sstr1	Trpm2
Acta1	Calcoco1	Dnmt3a	Fry	Insm1	Meis3	Pitpnm2	Rnf145	Stk32a	Trrap
Acvrl1	Camsap1	Dock9	Fxyd5	Irf6	Mgll	Pkd1	Rnf217	Stx1a	Ttbk2
Adcy2	Camta1	Dpysl3	Fzd3	Itm2c	Miat	Pkib	Rnf34	Stxbp5	Ttc39b
Adgrf5	Cand2	Dusp16	Fzd8	Itpk1	Mical3	Pkn1	Rogdi	Stxbp5l	Ttll5
Adgrl1	Casp6	Dusp18	Gch1	Itpr1	Mras	Plcg1	Rora	Syn1	Tubb2a
Adprm	Ccdc112	Dusp4	Gdi1	Itpr3	Mreg	Plekhg2	Rufy3	Syn2	Tubb6
Aff4	Ccdc3	Dusp8	Glis3	Kcnc1	Mtcl1	Plxna2	Rundc3a	Syp	Tubgcp3
Ago4	Ccdc64	Ece1	Glp1r	Kcnh6	Mterf4	Plxna3	Rusc2	Syt11	Ube2o
Aim1	Cdip1	Egfr	Gltscr1	Kcnh8	Mtss1	Plxnb1	Scg3	Syt5	Ubr4
Akna	Cdkn1b	Eif4e3	Gm15800	Kctd17	Myo9a	Pnrc1	Scg5	Syt7	Ulk1
Alas2	Ceacam10	Ell	Gm3414	Kdm4c	Myt1	Ppfibp1	Sct	Tacc1	Unc13a
Ambp	Celf2	Elmo2	Gm7694	Kif12	Nav1	Ppl	Sdccag8	Taf1b	Unk
Amigo2	Celf3	Epb4.1l1	Gnao1	Kif1b	Nbea	Ppm1l	Setbp1	Taf3	Uqcc1
Angel1	Celsr2	Ets1	Gnaz	Kif21a	Ncor2	Ppp2r5b	Sez6l2	Taok2	Usp20
Ankrd24	Cep126	Evl	Gng4	Klf7	Neurod1	Prex1	Sh3bp4	Tbc1d17	Usp35
Anxa5	Chd2	Evpl	Gnl3l	Klhl2	Ngfrap1	Prkaca	Sh3kbp1	Tbc1d19	Vangl1
Anxa6	Chga	F2	Gnptg	Klhl7	Nin	Prkar1a	Shisa2	Tbx3	Vegfa
Apbb2	Chgb	Fam105a	Golim4	Kmt2d	Nlgn2	Prune	Sik3	Tcp11l2	Vgf
Aplp1	Chka	Fam117a	Gpr19	Krt7	Nmt2	Psd	Sis	Tenm4	Vim
Arhgef1	Ckb	Fam120aos	Gprasp1	Krt83	Nos1ap	Psen2	Slc12a7	Tfe3	Vps33b
Arhgef2	Cldn4	Fam122a	Grhl1	Lamc1	Npy1r	Ptch1	Slc16a10	Tgfb1	Wif1
Arhgef40	Cln3	Fam160b1	Grik5	Lcorl	Nt5e	Ptp4a3	Slc18a1	Tlcd2	Wipf1
Arl13b	Cnot4	Fam172a	Grina	Ldoc1l	Orai2	Ptpn23	Slc35d3	Tle1	Xrcc4
Arl3	Cobll1	Fam189b	Gripap1	Lmx1a	Osbpl5	Ptprg	Slc36a1	Tle2	Zbtb20
Arl8a	Cpb2	Fam193a	Gsdma	Lpin2	Ostm1	Ptprn	Slc38a11	Tm4sf4	Zc3h13
Atp6v1b2	Cpe	Fam193b	Gse1	Lrrn1	Otud7b	Ptpru	Slc39a6	Tmcc2	Zcchc11
Atp8a2	Cplx2	Fam219a	Gtf2f2	Lysmd2	Pacsin1	Qdpr	Slc43a2	Tmed8	Zdhhc8
Atxn10	Crebl2	Fam222b	Gtpbp6	Lyst	Palld	Rab3a	Smad6	Tmeff1	Zfand2a
Auts2	Ctsf	Fam43a	H1fx	Madd	Pam	Rab3c	Smarcal1	Tmem2	Zfhx2
Bace1	Ctso	Fbrs	Habp4	Magi2	Pcdhga10	Rab3gap2	Smcr8	Tmem28	Zfp184
Baiap3	Cwh43	Fbrsl1	Hdac6	Man1c1	Pcdhgb6	Ralgds	Smim13	Tmem44	Zfp335
Bcam	Cyb5r3	Fbxl16	Hivep2	Map1b	Pclo	Rap1b	Snap25	Tmem51	Zfp516
Bex2	Dapk1	Fbxo7	Hmces	Map3k10	Pcsk1	Rap1gap2	Sned1	Tmx4	Zfp518b
Bin1	Dclk3	Fev	Hmgn3	Map3k15	Pcsk1n	Rasal2	Sntb2	Tnfrsf11b	Zfp568
Bmf	Ddc	Fhad1	Hoxb4	Map3k5	Pde11a	Reep6	Snx11	Tnrc6c	Zfp608

(Table adapted from Supplementary Data 3 of Habowski et al., 2020; Communications Biology)

REFERENCES

1. Beumer, J. & Clevers, H. Regulation and plasticity of intestinal stem cells during homeostasis and regeneration. *Development* **143**, 3639–3649 (2016).

2. Buczacki, S. Fate plasticity in the intestine: The devil is in the detail. *World J. Gastroenterol.* **25**, 3116–3122 (2019).

3. Tetteh, P. W., Farin, H. F. & Clevers, H. Plasticity within stem cell hierarchies in mammalian epithelia. *Trends Cell Biol.* **25**, 100–108 (2015).

4. Gehart, H. & Clevers, H. Tales from the crypt: new insights into intestinal stem cells. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 19–34 (2019).

5. Murata, K. *et al.* Ascl2-Dependent Cell Dedifferentiation Drives Regeneration of Ablated Intestinal Stem Cells. *Cell Stem Cell* (2020). doi:10.1016/j.stem.2019.12.011

6. Tóth, B., Ben-Moshe, S., Gavish, A., Barkai, N. & Itzkovitz, S. Early commitment and robust differentiation in colonic crypts. *Mol. Syst. Biol.* **13**, 902 (2017).

7. Heijmans, J. *et al.* ER Stress Causes Rapid Loss of Intestinal Epithelial Stemness through Activation of the Unfolded Protein Response. *Cell Rep.* **3**, 1128–1139 (2013).

8. Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nature Reviews Genetics* **15**, 163–175 (2014).

9. Tian, B. & Manley, J. L. Alternative polyadenylation of mRNA precursors. *Nature Reviews Molecular Cell Biology* **18**, 18–30 (2016).

10. Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nature Reviews Molecular Cell Biology* **18**, 437–451 (2017).

11. Mayr, C. What are 3′ UTRs doing? *Cold Spring Harb. Perspect. Biol.* **11**, (2019).

12. Berkovits, B. D. & Mayr, C. Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. *Nature* **522**, 363–367 (2015).

13. Wang, Y. *et al.* Mechanism of alternative splicing and its regulation. *Biomed. Reports* **3**, 152–158 (2015).

14. Yang, X. *et al.* Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell* **164**, 805–817 (2016).

15. Nefzger, C. M. *et al.* A Versatile Strategy for Isolating a Highly Enriched Population of Intestinal Stem Cells. *Stem Cell Reports* **6**, 321–329 (2016).

16. Wang, F. *et al.* Isolation and Characterization of Intestinal Stem Cells Based on Surface Marker Combinations and Colony-Formation Assay. *Gastroenterology* **145**, 383-395.e21 (2013).

17. van der Flier, L. G. *et al.* Transcription Factor Achaete Scute-Like 2 Controls Intestinal Stem Cell Fate. *Cell* **136**, 903–912 (2009).

18. Schuijers, J., Van Der Flier, L. G., Van Es, J. & Clevers, H. Robust cre-mediated recombination in small intestinal stem cells utilizing the Olfm4 locus. *Stem Cell Reports* **3**, 234–241 (2014).

19. Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).

20. Yan, K. S. *et al.* Intestinal Enteroendocrine Lineage Cells Possess Homeostatic and Injury-Inducible Stem Cell Activity. *Cell Stem Cell* **21**, 78-90.e6 (2017).

21. Gehart, H., Van Es, J. H., Dekkers, J. F., Rios, A. & Clevers Correspondence, H. Identification of Enteroendocrine Regulators by Real-Time Single-Cell Differentiation Mapping Data Resources GSE113561 Gehart et al. *Cell* **176**, (2019).

22. Moor, A. E. *et al.* Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis. *Cell* **175**, 1156-1167.e15 (2018).

23. Herring, C. A. *et al.* Unsupervised Trajectory Analysis of Single-Cell RNA-Seq and Imaging Data Reveals Alternative Tuft Cell Origins in the Gut. *Cell Syst.* **6**, 37-51.e9 (2018).

24. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003–1007 (2007).

25. Luo, X. C. *et al.* Infection by the parasitic helminth Trichinella spiralis activates a Tas2r-mediated signaling pathway in intestinal tuft cells. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 5564–5569 (2019).

26. Howitt, M. R. *et al.* Tuft cells, taste-chemosensory cells, orchestrate parasite type 2 immunity in the gut. *Science* **351**, 1329–33 (2016).

27. Zhang, P. *et al.* Carrier-Assisted Single-Tube Processing Approach for Targeted Proteomics Analysis of Low Numbers of Mammalian Cells. *Anal. Chem.* **91**, 1441–1451 (2019).

28. Shi, T. *et al.* Facile carrier-assisted targeted mass spectrometric approach for proteomic analysis of low numbers of mammalian cells. *Commun. Biol.* **1**, 1–9 (2018).

29. Salomonis, N. *et al.* Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10514–9 (2010).

30. Yamazaki, T. *et al.* TCF3 alternative splicing controlled by hnRNP H/F regulates E-cadherin expression and hESC pluripotency. *Genes Dev.* **32**, 1161–1174 (2018).

31. Yeo, G. W. *et al.* Alternative Splicing Events Identified in Human Embryonic Stem Cells and Neural Progenitors. *PLoS Comput. Biol.* **3**, e196 (2007).

32. Salomonis, N. *et al.* Alternative Splicing in the Differentiation of Human Embryonic Stem Cells into Cardiac Precursors. *PLoS Comput. Biol.* **5**, e1000553 (2009).

33. Gabut, M. *et al.* An Alternative Splicing Switch Regulates Embryonic Stem Cell Pluripotency and Reprogramming. *Cell* **147**, 132–146 (2011).

34. Lackford, B. *et al.* Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO J.* **33**, 878–889 (2014).

35. Grassi, E. *et al.* Choice of alternative polyadenylation sites, mediated by the rna-binding protein Elavl3, plays a role in differentiation of inhibitory neuronal progenitors. *Front. Cell. Neurosci.* **12**, (2019).

36. de Morree, A. *et al.* Alternative polyadenylation of Pax3 controls muscle stem cell fate and muscle function. *Science (80-.).* **366**, 734–738 (2019).

37. Ji, Z., Lee, J. Y., Pan, Z., Jiang, B. & Tian, B. Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7028–7033 (2009).

38. Fiszbein, A. & Kornblihtt, A. R. Alternative splicing switches: Important players in cell differentiation. *BioEssays* **39**, 1600157 (2017).

39. Shen, S. *et al.* rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E5593-601 (2014).

40. Vaquero-Garcia, J. *et al.* A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife* **5**, e11752 (2016).

41. Xia, Z. *et al.* Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3′-UTR landscape across seven tumour types. *Nat. Commun.* **5**, 1–13 (2014).

42. Lin, H. V. *et al.* Splits ends is a tissue/promoter specific regulator of Wingless signaling. *Development* **130**, 3125–3135 (2003).

43. Monfort, A. & Wutz, A. Progress in understanding the molecular mechanism of xist RNA function through genetics. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, (2017).

44. Oswald, F. *et al.* SHARP is a novel component of the Notch/RBP-Jκ signalling pathway. *EMBO J.* **21**, 5417–5426 (2002).

45. Doroquez, D. B., Orr-Weaver, T. L. & Rebay, I. Split ends antagonizes the Notch and potentiates the EGFR signaling pathways during Drosophila eye development. *Mech. Dev.* **124**, 792–806 (2007).

46. Feng, Y. *et al.* Drosophila split ends homologue SHARP functions as a positive regulator of Wnt/β-catenin/T-cell factor signaling in neoplastic transformation. *Cancer Res.* **67**, 482–491 (2007).

47. Dydensborg, A. B. *et al.* Differential expression of the integrins α6Aβ4 and α6Bβ4 along the crypt–villus axis in the human small intestine. *Histochem. Cell Biol.* **131**, 531–536 (2009).

48. Zhou, Z. *et al.* α6-Integrin alternative splicing: Distinct cytoplasmic variants in stem cell fate specification and niche interaction. *Stem Cell Research and Therapy* **9**, 122 (2018).

49. Wilczynska, A. *et al.* eIF4A2 drives repression of translation at initiation by Ccr4-Not through purine-rich motifs in the 5′UTR. *Genome Biol.* **20**, 262 (2019).

50. Moore, A. C. *et al.* Myeloid Translocation Gene Family Members Associate with T-Cell Factors (TCFs) and Influence TCF-Dependent Transcription. *Mol. Cell. Biol.* **28**, 977–987 (2008).

51. Parang, B. *et al.* Myeloid translocation genes differentially regulate colorectal cancer programs. *Oncogene* **35**, 6341–6349 (2016).

52. Amann, J. M. *et al.* Mtgr1 Is a Transcriptional Corepressor That Is Required for Maintenance of the Secretory Cell Lineage in the Small Intestine. *Mol. Cell. Biol.* **25**, 9576–9585 (2005).

53. Venugopal, A. *et al.* RNA binding protein RBM3 increases β-catenin signaling to increase stem cell characteristics in colorectal cancer cells. *Mol. Carcinog.* **55**, 1503–1516 (2016).

54. Van Dop, W. A. *et al.* Loss of Indian hedgehog activates multiple aspects of a wound healing

response in the mouse intestine. *Gastroenterology* **139**, (2010).

55. Tian, H. *et al.* Opposing activities of Notch and Wnt signaling regulate intestinal stem cells and gut homeostasis. *Cell Rep.* **11**, 33–42 (2015).

56. Lo, Y.-H. *et al.* Transcriptional Regulation by ATOH1 and its Target SPDEF in the Intestine. *Cell. Mol. Gastroenterol. Hepatol.* **3**, 51–71 (2017).

57. Walter, P. & Ron, D. The unfolded protein response: From stress pathway to homeostatic regulation. *Science* **334**, 1081–1086 (2011).

58. Tawiah, A. *et al.* High MUC2 Mucin Biosynthesis in Goblet Cells Impedes Restitution and Wound Healing by Elevating Endoplasmic Reticulum Stress and Altered Production of Growth Factors. *Am. J. Pathol.* **188**, 2025–2041 (2018).

59. Pinchuk, I. V., Mifflin, R. C., Saada, J. I. & Powell, D. W. Intestinal mesenchymal cells. *Current Gastroenterology Reports* **12**, 310–318 (2010).

60. Danopoulos, S., Schlieve, C. R., Grikscheit, T. C. & Al Alam, D. Fibroblast Growth Factors in the Gastrointestinal Tract: Twists and Turns. *Dev. Dyn.* **246**, 344–352 (2017).

61. Bezençon, C. *et al.* Murine intestinal cells expressing Trpm5 are mostly brush cells and express markers of neuronal and inflammatory cells. *J. Comp. Neurol.* **509**, 514–525 (2008).

62. Guo, P. *et al.* Nerfin-1 represses transcriptional output of Hippo signaling in cell competition. *Elife* **8**, (2019).

63. Alarcón, C. *et al.* Nuclear CDKs Drive Smad Transcriptional Activation and Turnover in BMP and TGF-β Pathways. *Cell* **139**, 757–769 (2009).

64. Huang, Z. *et al.* YAP stabilizes SMAD1 and promotes BMP2-induced neocortical astrocytic differentiation. *Dev.* **143**, 2398–2409 (2016).

65. Rothenberg, M. E. *et al.* Identification of a cKit+ Colonic Crypt Base Secretory Cell That Supports Lgr5+ Stem Cells in Mice. *Gastroenterology* **142**, 1195-1205.e6 (2012).

66. Sato, T. *et al.* Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469**, 415–8 (2011).

67. McKinley, E. T. *et al.* Optimized multiplex immunofluorescence single-cell analysis reveals tuft cell heterogeneity. *JCI insight* **2**, (2017).

68. Andriatsilavo, M. *et al.* Spen limits intestinal stem cell self-renewal. *PLoS Genet.* **14**, (2018).

69. Matsumoto, Y. *et al.* Epithelial EP4 plays an essential role in maintaining homeostasis in colon. *Sci. Rep.* **9**, (2019).

70. Forman, B. M. *et al.* 15-Deoxy-Δ12,14-Prostaglandin J2 is a ligand for the adipocyte determination factor PPARγ. *Cell* **83**, 803–812 (1995).

71. Kliewer, S. A. *et al.* A prostaglandin J2 metabolite binds peroxisome proliferator-activated receptor γ and promotes adipocyte differentiation. *Cell* **83**, 813–819 (1995).

72. Lim, H. *et al.* Cyclo-oxygenase-2-derived prostacyclin mediates embryo implantation in the mouse via PPARδ. *Genes Dev.* **13**, 1561–1574 (1999).

73. Arnaud-Dabernat, S., Yadav, D. & Sarvetnick, N. FGFR3 contributes to intestinal crypt cell growth arrest. *J. Cell. Physiol.* **216**, 261–268 (2008).

74. Liu, D.-W., Tsai, S.-M., Lin, B.-F., Jiang, Y.-J. & Wang, W.-P. Fibroblast growth factor receptor 2c signaling is required for intestinal cell differentiation in zebrafish. *PLoS One* **8**, e58310 (2013).

75. Zhou, Z. Q. & Hurlin, P. J. The interplay between Mad and Myc in proliferation and differentiation. *Trends in Cell Biology* **11**, (2001).

76. Chodaparambil, J. V. *et al.* Molecular functions of the TLE tetramerization domain in Wnt target gene repression. *EMBO J.* **33**, 719–731 (2014).

77. A.N. Habowski, J.M. Bates, J.L. Flesher, R.A. Edwards, M. L. W. Isolation of murine large intestinal crypt cell populations with flow sorting. *Protoc. Exch.* (2020).

78. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

79. Thomas, P. D. *et al.* PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res.* **13**, 2129–2141 (2003).

80. Chen, E. Y. *et al.* Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, (2013).

81. Kuleshov, M. V *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, (2016).

82. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

83. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).

84. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **13**, 731–740 (2016).

85. Uhlen, M. *et al.* Tissue-based map of the human proteome. *Science (80-.).* **347**, 1260419–1260419 (2015).

86. Pontén, F., Jirström, K. & Uhlen, M. The Human Protein Atlas—a tool for pathology. *J. Pathol.* **216**, 387–393 (2008).

87. Okuda, S. *et al.* jPOSTrepo: an international standard data repository for proteomes. *Nucleic Acids Res.* **45**, (2016).

88. Billing, L. J. *et al.* Single cell transcriptomic profiling of large intestinal enteroendocrine cells in mice – Identification of selective stimuli for insulin-like peptide-5 and glucagon-like peptide-1 co-expressing cells. *Mol. Metab.* **29**, 158–169 (2019).

89. Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).

90. Nadjsombati, M. S. *et al.* Detection of Succinate by Intestinal Tuft Cells Triggers a Type 2 Innate Immune Circuit. *Immunity* **49**, 33–41 (2018).

91. Gerbe, F., Legraverend, C. & Jay, P. The intestinal epithelium tuft cells: specification and function. *Cell. Mol. Life Sci.* **69**, 2907–2917 (2012).

92. Schneider, C. *et al.* A Metabolite-Triggered Tuft Cell-ILC2 Circuit Drives Small Intestinal Remodeling. *Cell* **174**, 271-284.E14 (2018).

93. Lee, S. W. *et al.* SUMOylation of hnRNP-K is required for p53-mediated cell-cycle arrest in response to DNA damage. *EMBO J.* **31**, 4441–4452 (2012).

94. Savari, S., Vinnakota, K., Zhang, Y. & Sjölander, A. Cysteinyl leukotrienes and their receptors: bridging inflammation and colorectal cancer. *World J. Gastroenterol.* **20**, 968–77 (2014).

CHAPTER 3

Regionally Distinct Responses to Irradiation Wounding of the Intestine

Unpublished work:
***Habowski, A.N.**, 2020.*

ABSTRACT

The rapidly renewing epithelial cells that make up the inner layer of the intestine possess a stunning ability to recover following injury. Although many cell types have been hypothesized to enable wound repair and regeneration, few studies have taken an unbiased approach to look at the initiation of recovery following wounding. Here we present the results of single cell RNA-sequencing and histological analysis of murine intestines immediately following sub-lethal radiation. In the first few days we observe loss of radiosensitive stem cells and dynamic changes in the abundance of other cell types, as well as their transcriptomes, and proteomes. We determine that the dominate three activities during wounding and at the start of recovery are: 1) physical barrier maintenance, 2) secretion of protective peptides, and 3) increased proliferation to regain lost crypt cells. Based on this data we suggest that proliferative transit amplifying cells and immature absorptive cells are the first cells to repopulate the damaged stem cell niche.

INTRODUCTION

The intestinal mucosa is a rapidly renewing single layer of epithelial cells that perform essential absorptive and barrier functions. Crypts are invaginating structures of the epithelium that produce the multiple cell types that carry out these functions: i) proliferating stem cells at the base (marked by Lgr5 expression (Lgr5+)), transit amplifying cells that proliferate in a zone above the niche (Transit Amplifying Zone (TAZ)), and short-lived, mature, differentiated cells that perform secretory or absorptive activities. Cell fate choices of immature cells in the TAZ are maintained in homeostatic proportions. Upon wounding, the crypt structure and its homeostatic state are severely disrupted, yet the essential nature of the crypt necessitates that barrier function and nutrient absorption somehow continue. A remarkable feature of the intestine is that these functions are maintained while the crypt repairs and recovers. There is much that is not understood about how this happens - how the intestine responds to wounding and what the initial steps are that facilitate recovery.

Intestinal wounding can result from a variety of sources and different insults, some of which are due to external insults, some of which are internally generated. Micro-organisms such as bacteria and parasites can inflict intestinal damage, resulting in stem cell death, inflammation, and wound response [1]. Inflammatory bowel disease (IBD) is a disease of chronic inflammation of the intestine and surrounding stroma, a continuous process of mucosal wounding with serious, debilitating side effects. Although the exact causes of IBD is still not clear, researchers use chemicals such as the carcinogenic/neurotoxic agent Azoxymethane (AOM) and Dextran sodium sulfate (DSS), to damage murine intestinal mucosa to model IBD and colitis-associated cancer [2,3]. Radiation exposure can result in acute radiation syndrome which, depending on the dose, leads to multi-organ involvement and failure [4]. While the hematopoietic system is the most commonly known system for radiosensitivity, the gastrointestinal system, skin, and nervous system are also impacted, especially at high dosages [4]. High-dose radiation severely damages crypts and at high enough doses, causes such catastrophic levels of mucosal wounding that there is a complete loss of the mucosa which is a fatal condition[4]. The most common form of intestinal radiation exposure in humans occurs as a side effect of pelvic or gastric radiation for cancer treatment. While not as severe and usually temporary, this can still result in mucosal damage and sickness in the patient. However, these findings have highlighted the usefulness of radiation as a tool to study intestinal wounding and repair.

Christopher Potten, the grandfather of intestinal irradiation and apoptosis, was a meticulous scientist and one of the first to observe that intestinal stem cells (only theorized stem cells at this point in history) were extremely sensitive to irradiation [5]. Within a few hours of irradiation these cells can be seen undergoing apoptosis, although the response is variable throughout the gastrointestinal tract and highly dependent on the dose of irradiation [6,7]. 1 Gy of ionizing radiation triggers maximum apoptosis in the small intestine within 6 hours, whereas 8 Gy results in elevated and prolonged apoptosis for several days [6]. Within 6 hours of 12 Gy irradiation (the largest dose of whole-body irradiation mice reliably survive at least four days [8]) there is >50% decrease in proliferation and by two days crypt length has significantly shortened [9]. By day 4, proliferation activities emerge and increase in crypt lengthening is observed [9]. These data suggest that wounding and the initiation of recovery take place within four days of sub-lethal irradiation, although fully returning to homeostasis can take longer.

This work naturally led the field to encounter what has remained a longstanding, multi-decade conundrum: what cells replace the lost stem cells? Three hypotheses have been proposed: 1) A few residual stem cells that do not die, 2) quiescent stem cells, and 3) de-differentiation of mature epithelial cells. The idea that a few surviving stem cells replenish the stem cell pool has been discredited since crypts can recover and regain homeostasis after targeted and complete ablation of all Lgr5+ stem cells [10,11]. The quiescent/reserve stem cell hypothesis has been supported by several studies including initial work highlighting a label-retaining cell in the +4 position – a persistent cell that looked distinct from the classic Lgr5+ intestinal stem cell [12,13]. However, the greatest limitation of this work are that the marker genes (i.e. *Bmi1*, *Hopx*, *Tert*, *Lrig, Dclk1*, *Prox1*, *Mex3a*, *Sca1*) of the proposed quiescent stem cells are not unique to this cell type [1,13,14]. Instead they also label other epithelial cell populations including some rare populations (tuft and enteroendocrine), which can be found throughout the crypt including near the '+4 position' [13,15]. Although lineage tracing experiments using quiescent stem cell markers have been purported to show that these cells do regenerate crypts, due to the lack of specificity of the markers, this conclusion is tenuous at best. The lineage tracing data is equally congruent with the idea that de-differentiation of mature cell types can occur and that there is a general plasticity of intestinal epithelial cells and general de-differentiation potential [13,14,16]. On this front, several studies have provided evidence that mature, differentiated cells from both secretory and absorptive lineages can indeed de-differentiate and convert back into stem cells[13,17,18].

Collectively, the growing number of studies hypothesize that nearly any epithelial cell type can de-differentiate and replace stem cells, it is challenging to know which of these hypotheses are the most likely. It is possible that different types of wounding utilize different means of recovery, or that there is crypt-to-crypt variation and repair occurs through multiple de-differentiation events. However, one of the key unifying conclusions from all these studies is that the intestine is well equipped to respond to wounding events and rapidly return to homeostasis. In the work presented here we investigate the immediate response to wounding and the start of recovery using histology and a single cell RNA-seq time course of colon epithelial cells following irradiation. This enabled us to see broadly and in an unbiased manner, how all epithelial cell types respond in different manners in order to facilitate immediate survival as well as rapidly initiating recovery. This work provides further clarity on how intestinal plasticity is activated and how stemness is ultimately restored.

RESULTS

Colonic tissue and cellular response to irradiation

Following whole body sub-lethal irradiation, murine intestines continue to function, but are wounded and take time before returning to homeostasis. Gross anatomical analysis of the small intestine shows decreased intestinal length, loss of vasculature, and increased mucin production following irradiation (Figure 3.1a). The colon is also slightly truncated and the lumen contains loose stool (Figure 3.1a). Flow cytometry analysis of the colon (cecum to rectum) shows a 78% reduction in epithelial cells 1-day post irradiation (Figure 3.1b). By day 4 post irradiation there is a modest, but significant, increase in epithelial cells compared to day 1. By day 7, the abundance of epithelial cells has surpassed that of the wild-type colon with a 49% increase over baseline. This suggests that within the first week following irradiation there is a rapid wound response followed by repair and escalated proliferation. Importantly, with the first modest increase in epithelial cells already occurring at day 4 post irradiation, the important initial steps in wounding and repair have likely occurred in this time frame.

Analysis of H&E sections show a severe wounding response to irradiation in the small intestine (Supplementary Figure 3.1). In the colon, the cecum shows more severe signs of wounding than the rectum, consistent with previous reports (Figure 3.1c) [6]. Crypt base structures are disrupted and apoptotic cells are visible day 1 and 2. Although the tissue is still damaged on day 4, notably at the top of the crypt, the base of a substantial fraction of crypts have been repopulated with cells and there is semblance of normality suggesting some recovery has occurred (Figure 3.1c). By day 7, the base of all crypt has been repaired and is more densely packed with cells than in the WT. Measurements of crypt length show shorter crypts in the cecum of wild-type (WT) mice and nearly double the length in rectum (Figure 3.1d and Supplementary Figure 3.2b). By day 1 and 2 post irradiation, the cecum and mid colon crypt length is significantly shorter than WT; the rectum is unchanged. By day 4 post irradiation the mid-region of the colon is significantly longer than WT, and by day 7 an even larger middle region of the colon is longer than WT. In addition to changes in crypt length, wounding impacts the abundance and density of crypts. Decreased crypt density is most striking in the cecum day 2 post irradiation and persists in regions through day 4 (Figure 3.1e).

Single cell RNA-sequencing reveals an early transcriptomic response to wounding

Since some colonic recovery following irradiation appears to initiate prior to day 4, we sought to investigate the underlying transcriptomic changes that occurred at the start of wound healing. We isolated colonic crypt epithelial cells from control mice (n=4) and 1,2,3, and 4 days post whole-body X-ray irradiation (12 Gy) (n=3 mice for each irradiated condition) (Figure 3.2a). We then subjected these cells to 10X Chromium single cell RNA-sequencing. Clustree analysis revealed optimal resolution and appropriate clustering and highlighted the relationships between similar clusters (Figure 3.2b and Supplementary

Figure 3.3) [19]. A resolution of 1.3 resulted in 18 distinct clusters which are displayed via a UMAP plot (Figure 3.2c). We observed that a majority of clusters contained cells from WT and multiple days post-irradiation (Figure 3.2d). In instances where a cluster did not have cells contributed from specific days, we found this to be biologically relevant. For example, we observed a drop-off in the expression of stem cell markers on day 2 when looking at all cell types (Figure 3.2e) and the dwindling, then disappearance of the stem cell cluster (cluster 14) starting on day 2 (Figure 3.2d).

Although there are similarities between some of the clusters, we were able to identify distinct marker genes of even the smallest clusters (Figure 3.2f). These marker genes, in addition to known intestinal epithelial cell type markers (Figure 3.3a) were used to identify each cluster type (Figure 3.3b). For example, we identified cell types at the base of the crypt: stem cells and deep crypt secretory cells (DCS); cell types in the transit amplifying zone (TAZ); and differentiated cell types (Figure 3.3c). The TAZ consists of lineage neutral cells, as well as cells committed to either absorptive or secretory. Differentiated cells of absorptive (enterocytes) and secretory lineage (goblet and tuft) were identified.

Cell type abundance changes in response to wounding and apoptosis

As previously mentioned, we observe a loss of stem cells 2 days post irradiation (Figure 3.4a), but we also observed that the abundance of other cell type populations changed between day 1-4 as well. To our surprise, abundance of DCS cells decreased in a similar manner as stem cells, suggesting that the damage to the base of the crypt extends not just to stem cells but the neighboring support cells as well (Figure 3.4a). We also observed that the non-proliferative TA cell population decreased, while proliferative populations (ProlifTA and more absorptive committed AbsTA) initially decreased before substantial increasing by day 4. The size of the goblet cell population stayed steady, and the size of the SecPro population decreased transiently on day 1 but returned to WT levels and staying steady thereafter. A secretory-like population (Cluster 0) sharply increased on day 4 (Figure 3.4a and Supplementary Figure 3.4).

The tumor suppressor p53 can promote apoptosis or cell cycle arrest, and its activation upon irradiation and DNA damage could potentially trigger changes in populations that express it. Indeed, we observed *Trp53* mRNA expression to be elevated in stem cells, DCS, and proliferative TA and absorptive cells, even prior to irradiation (Figure 3.4b). Interestingly, the p53 target gene p21 (*Cdkn1a*) which promotes cell cycle arrest rather than apoptosis, was broadly expressed in all cell types (Supplementary Figure 3.5a). Other p53 target genes were more specifically enriched in stem and DCS cells, and interestingly, these populations also expressed components of the death receptor/Tumor Necrosis Factor (TNF) pathway (Supplementary Figure 3.5). Co-expression of two apoptosis-signaling pathways might more effectively facilitate the loss of stem cells and DCS following irradiation, while TAZ proliferative cells, lacking elevated co-expression, undergo cell cycle arrest, repair damage, and do not undergo apoptosis.

Enterocytes maintain an intact epithelial barrier immediately following irradiation

Relative to all other cryptal cell populations, enterocytes and less differentiated absorptive cells (Ent and Abs) were the only populations to dramatically increase within the first day following irradiation (Figure 3.4a). Enterocytes remained an abundant population until day 4 when there was a sharp decrease in number. This suggests that either mature enterocytes are more resistant to apoptosis and/or there is a rapid differentiation of immature cells into this cell fate. Although we did not observe any elevated response in Wnt signaling (i.e. expression of Wnt target genes) in cells following irradiation (Supplementary Figure 3.6), immunohistochemical staining of β-catenin (Ctnnb1) proved useful to follow changes in epithelial cells and structural features of the crypts. For example, in WT cecum, strong β-catenin membrane staining and some cytoplasmic staining was evident throughout the crypt (Figure 3.4c). On day 2, membrane staining highlighted the base of crypts to reveal a large proportion of large, bloated, and unhealthy cells, while also showing strong beta-catenin staining of a strikingly intact and adhesive epithelial barrier at the top of the crypt (Figure 3.4c). Based on the strong level of staining and a subcluster analysis of cells expressing the highest level of *Ctnnb1* mRNA, (Supplementary Figure 3.7a-e), this barrier layer is likely made of enterocytes. By day 4, there was a dramatic decrease in colonic enterocyte abundance (Figure 3.4a) and β-catenin staining in both the colon and small intestine revealed disruption of the epithelial barrier and distortion of the previously tightly packed adhesive cells (Figure 3.4c-d and Supplementary Figure 3.7f-g).

The changes in total enterocyte abundance was comprised of three enterocyte populations (clusters 2, 13, and 15) which responded in distinct manners to irradiation (Figure 3.4e). Ent-1 increased on day 1, then decreased, Ent-2 increased on day 2 and 3 (it is possible there is conversion of Ent-1 to Ent-2), and by day 4 only the Ent-3 remained. Two expression patterns that appeared unique to the enterocyte populations were strong increases in expression of the anti-oxidant factor *Sepp1* and the interferon pathway regulator *Irf7.* Both were elevated in expression on day 2 with persistent, strong expression through day 4, and may play a role in the survival of enterocytes (3.4f-g; see Discussion).

Proliferation increases four days after irradiation wounding

Although irradiation results in the loss of stem cells and therefore the loss of stem gene expression signatures, we observed continued expression of proliferation associated genes 1-3 days after irradiation (Figure 3.5a). By day 4 we observed dramatic increases in proliferation markers to levels that exceeded that in WT samples (Figure 3.5a). To evaluate which cell populations were becoming more proliferative, we utilized subclustering analysis of cells expressing proliferative markers (Pcna>0, Mki67>0, or Top2a>0). Subclustering revealed 6 populations with distinctive and lineage specific markers (Figure 3.5b-d). Cluster 4 showed signs of damage (high expression of mitochondrial markers) and was determined to be stem cells based on high *Lgr5* expression (Figure 3.5c-d). A small subset, Cluster 5, expressed the secretory progenitor marker *Spdef*, whereas cluster 2 and 3 expressed absorptive markers (Figure 3.5d). Two related clusters (0 and 1) did not express any lineage specific genes and the cluster marker genes were predominantly proliferation associated (Figure 3.5c-d). Interestingly, these lineage neutral proliferative cells along with absorptive proliferative cells (cluster 2,3) increased in abundance by day 4, in some cases by two-fold (Figure 3.5e). This increase was preceded by a drop in abundance on day 1, a drop that might indicate cell loss and/or also cell cycle stalling – conditions that are released prior to day 4, as previously discussed (Figure 3.5e). The proliferative secretory population (cluster 5, Fig. 3.5e) slowly declined and by day 4 was about half as abundant at WT (Figure 3.5e). These results suggest perhaps that uncommitted transit amplifying cells (cluster 0, 1) and absorptive committed progenitors (cluster 2, 3) might be responsible for the increased number of cells in crypts following wounding and they might also be the cell to eventually repopulate the base of the crypt and ultimately stem cells.

We were rather intrigued to find that cells positive for proliferation markers *Mki67* and *Top2a* were also *Pcna-positive* but not vice versa (Figure 3.5d). For example, *Krt20* positive absorptive cells were only *Pcna* positive, no *Mki67* or *Top2a* mRNA was detected (Figure 3.5d). We thus used *Pcna* antisera for immunohistochemical analysis of tissues as a more sensitive marker to identify all proliferative cells in crypts. IHC for Pcna showed prominent staining at the base of cecal crypts and in the TAZ of WT (Figure 3.5f). Although there was severe wounding day 1 and 2 post irradiation, there was still residual Pcna-positive cells in the TAZ and remnants of the crypt base (Figure 3.5f). By day 4, there was a dramatic increase in the number of Pcna positive cells, in some cases occupying the entire base of the crypt, the TAZ, and even further up the crypt (Figure 3.5f). This increase was diminished by day 7 where staining was once again limited to the base of the crypt (Figure 3.5f). In WT small intestine, Pcna is strongly expressed in the crypt base and clearly demarcates stem cells. By day 4 post irradiation, similar to the colon, the small intestine exhibits marked increases in Pcna positive proliferative cells (Figure 3.5g). However, these cells have not yet occupied the entire base of the crypt and instead appear to be migrating down to the base to replace the skeletal remains of old apoptotic cells (Figure 3.5g).

A novel secretory response to irradiation

Although secretory cells did not seem to respond to irradiation with dramatic changes in cell abundance relative to other cell types in the crypt, we observe changes in their expression of secreted proteins. Within a few days of irradiation, mucin production increased, with *Muc2* expression spiking on day 2 and *Muc3* and *Muc13* spiking on day 3 (Figure 3.6a). The increase in *Muc2* expression was not limited to the mature secretory goblet cells (cluster 3) or their SecPro progenitors (Cluster 10), but we observed broad increases of *Muc2* in nearly all cell clusters (Figure 3.6b). This suggests that increases in mucin gene expression are not limited to secretory cells following irradiation.

Another class of secreted protein that was altered in response to irradiation was several antimicrobial peptides (AMPs), in particular *Cma1*, *Ang4*, and *Retnlb*. These three AMPs increased in expression on day 3 and 4 post irradiation (Figure 3.6a). Gene expression of *Cma1* was not detectable in WT, modest by day 3, and strongly expressed by day 4 post irradiation (Figure 3.6a,c). As with mucin expression, this increase was observed not only in secretory cells (goblet, SecPro, and Cluster 0), but more broadly in non-secretory cells types (Figure 3.6c). IHC staining of Cma1 showed a gradual increase in Cma1 protein production, with the strongest staining evident on day 4 (Figure 3.6d and Supplementary Figure 3.8). Interestingly, day 4 Cma1 staining was confined to the crypt base, whereas day 7 staining highlighted sparse cells located further up the crypt consistent with the idea of a wave of expression/secretion originating at the base of the crypt and moving upward toward the mucosal surface. *Retnlb* and *Ang4* expression was detected in WT, and then following a decrease immediately after irradiation, strongly detected again in day 4 (Figure 3.6a,c). Similar to *Cma1*, the expression of *Retnlb* and *Ang4*, was not limited to secretory cells but broadly expressed in many cell types (Figure 3.6c). IHC staining of Retnlb revealed a regionally specific pattern -

cecal specific expression in WT, and no staining day 1, 2, or 7 (Figure 3.6d and Supplementary Figure 3.9). Staining of day 4 tissue detected Retnlb in the distal cecum and clearly evident signs of rapid production and secretion into the lumen of the crypt. Cma1 staining was restricted to the rectum, suggesting that despite the lack of severe signs of wounding via crypt histology, transcriptomic changes are clearly occurring – an indication that rectal cells are responding to irradiation (Supplementary Figure 3.8).

DISCUSSION AND FUTURE WORK

This chapter presents our study to investigate the immediate transcriptomic and histological changes that occur in the intestinal crypt following irradiation. The goal of the study was to understand how wounding occurs and how recovery is initiated. Using histological analysis, we identified regional differences in colonic wounding and recovery. The cecum bears the greatest damage, with rectal crypts showing minimal overt signs of wounding. Immediately following irradiation, signs of proliferation decrease dramatically, but then sharply return by 4 days after irradiation. The middle of the colon shows a strong "over" recovery in that the average crypt length is significantly longer than WT crypts by day 7 after wounding. An increase in colonic crypt length suggests that there is an overcompensation of cellular proliferation – a response documented by Christopher Potten a decade ago [9]. Single cell RNA-sequencing of epithelial cells every day for four days, confirmed there is a dramatic loss of stem cells. We also discovered that DCS cells are quickly lost. We suggest that the loss of these two populations is due to the concerted actions of the p53 and TNF death receptor pathways promoting apoptosis (Supplementary Figure 3.5). Proliferative non-stem cells do not strongly co-express components of these two pathways, and might therefore be subjected to cell cycle arrest, a condition that is released prior to day 4 of recovery. Although irradiation decreased the total number of epithelial cells, enterocytes seemed to be the least impacted with their relative abundance increasing immediately following irradiation. Secretory cells did not appear to drastically change in relative abundance following irradiation, but there were increases in secretory protein production in both secretory and non-secretory cell populations. We noted that members of the mucin and antimicrobial peptide families are expressed uniquely – proteins that likely provide additional protection of the wounded mucosa. Although there are likely many intricate responses of cells to irradiation we suggest that at least from the transcriptome perspective, the dominate three activities in the first few days following wounding are: 1) physical barrier maintenance, 2) secretion of protective peptides, and 3) increased proliferation of multiple non-stem cell populations to repair and restore damaged crypts.

Enterocytes and barrier function

Although we observed a global loss in epithelial cells following irradiation (Figure 3.1b), we also found that enterocytes become more abundant relative to other cell types in the first 3 days following wounding (Figure 3.4a). Previous studies have shown that irradiation damage leads to increased permeability of the intestine within the first few hours possibly due to damaged tight junctions and adhesion junctions (including β-catenin-dependent adhesions) [20,21]. However, our IHC staining of β-catenin showed strong staining of membrane connections between cells at the top of crypts on day 1 and day 2 after wounding (Figure 3.4c and Supplementary figure 3.7g). At least by histological appearance, these cells are not goblet and are therefore likely to be absorptive enterocytes (goblet cells are minimal in abundance). These enterocytes are strikingly abundant at the mucosal surface and they appear to form a strong barrier despite the fact that they are at the top of immensely damaged crypt bases (Figure 3.4c). On day 4, as enterocytes dramatically decrease in number, IHC staining highlighted what appear to be bloated, distended enterocytes forming a shapeless layer of cells that appear to be delaminating from the cells below (Figure 3.4c and Supplementary Figure 3.7g). By day 7, a new epithelial layer is formed with tight cell adhesion, healthy looking cells throughout the crypt, and secretory cells in relative positions that are reminiscent of WT tissue (Supplementary Figure 3.7g). Taken together, our histological and scRNA-seq analysis suggests that in response to irradiation, surviving enterocytes form a tight protective barrier as best they can despite still being wounded. This barrier lasts until day 4 such that by day 7 new cells have replaced it. This barrier is likely crucial to prevent bacteria or other pathogens from penetrating into the intestine which could result is septicemia [21].

The ability of enterocytes to evade the most severe, lethal damage that other intestinal cells succumb to likely relates to decreased *Trp53* levels (Figure 3.4b and Supplementary Figure 3.5a). That they are the more abundant surviving population prompted a search in the transcriptome for additional wound responses. *Sepp1*, a major selenium transporter protein, sharply increased in expression in enterocytes on day 2 (Figure 3.4f-g). Selenium is an essential trace element, necessary for forming selenocysteine in selenoproteins and it is normally absorbed via enterocytes in the intestine [22]. Importantly, selenoproteins, including Sepp1, are anti-oxidants and immune defense regulators that modulate intestinal inflammation and carcinogenesis [22,23]. Previous work has reported that Sepp1 is expressed in the intestine, predominantly in small intestinal villi, and that it has a directional secretion pattern and extracellular antioxidant enzyme activity [22]. Additionally, *in vitro* overexpression of SEPP1 in fibroblasts significantly decreases radiation induced toxicity and cell death by reducing the accumulation of radiation-induced late reactive oxygen species (ROS) [24]. This suggests that enterocyte expression of *Sepp1* may act in a similar protective manner and promote irradiation survival.

Selenbp1 is another selenium binding protein that increased in expression on day 1 (but then sharply decreased in expression by day 3 in both enterocytes and all cell types) (Figure 3.4f-g). Selenbp1 is a potential tumor suppressor as it is frequently decreased in expression in many tumor types [25]. When SELENBP1 was over-expressed in lung cancer cells there was reduced proliferation and increased apoptosis when cells were challenged with H_2O_2 [25]. Interestingly, SELENBP1 can transcriptionally modulate p21 (Cdkn1a) expression, independent of p53 to promote cell cycle arrest rather than apoptosis [26]. At this point it is therefore not clear what role *Selenbp1* is playing – a apoptosis promotion role (that Sepp1 might oppose), or a cell cycle arrest role (that Sepp1 contributes to with its ROS-scavenging activities). Overall, other than its ability to covalently bind selenium, there is relatively little known about how Selenbp1 contributes to the regulation of selenium and how it might interact with Sepp1 [25].

Irf7 is a key transcriptional regulator of type-I interferons that are expressed in response to pathogenic infections (predominantly viral) [27,28]. We were rather intrigued to see an enterocyte specific increase in *Irf7* expression starting on day 2 post irradiation, although no epithelial interferon expression was detected (stromal expression in likely) (Figure 3.4f-g). Irf7 is highly regulated via protein modification, so the observed increased in mRNA expression might indicate that enterocytes become heightened in their potential to respond to a viral infection. That Irf7 might contribute a pro-survival function is underscored by work in breast cancer which has highlighted that Irf7 provides a pro-tumor survival function via chemoresistance and cell dormancy [29]. More research is clearly needed to understand the role of increased *Irf7* expression in enterocytes.

Secreted peptides aid recovery following wounding

The colon contains two distinct layers of mucus, a smaller dense and tightly attached inner stratified layer with no bacteria (~50 µm thick in mice) and an outer, much thicker, looser layer that faces the lumen and contains commensal bacteria [30]. The dominant component of both layers is Muc2. In the absence of Muc2, bacteria can come into direct contact with epithelial cells and, as illustrated with *Muc2-/-* mice, this loss leads to disease (inflammation, diarrhea, colon prolapse, and increased risk of colon cancer) [30]. As a technical note, these layers of mucus collapse and shrink with FFPE tissue processing and thus are not visible in histology images [30]. Despite the severe intestinal wounding that we induced with irradiation, we did not observe any loss in *Muc2* expression, and instead observed an increase in *Muc2* mRNA levels by day 2. Muc2 mRNA levels returned to baseline by day 4 (Figure 3.6b). The day 2 increase is most prominent in goblet and SecPro cells but is also detected as broadly increased in most of the other cell populations

(Figure 3.6b). A general and broad increase in expression of *Muc2* is likely an important wound response to maintain the mucin layer and prevent encroachment of bacteria.

In addition to increases in mucin expression, we observed an increase in expression of three Antimicrobial Peptides (AMPs) starting on day 3 post irradiation (Figure 3.6a,c). AMPs are a widely used host defense mechanism of many organisms, including mammals. It is well known that small intestine paneth cells produce AMPs, notably a wide variety of defensins [31]. Interestingly our IHC staining revealed regional specific expression of Cma1 and Retnlb both of which were induced strongly on day 4 after wounding (Figure 3.6d and Supplementary Figure 3.8 and 3.9). Cma1 is a secreted protease, also known as Mast Cell Chymase since it is primarily expressed in mast cells . Although it has not been well studied as a gene expressed in epithelial cells, there is evidence for its role in host defense [32–34]. *Ang4* is part of the RNase subfamily and previously reported to be secreted into the lumen of the small intestine by paneth cells in response to bacterial infection [31,35]. Retnlb (RELMβ), which has previously been shown to be secreted from intestinal goblet cells, is a gastrointestinal cytokine, that recruits immune cells and is able to kill Gram-negative bacteria that invades too close to the epithelial layer [36–38]. The emergence of different patterns of AMP expression suggests that either 1) the intestine responds in a generic manner to wounding, be it irradiation or bacterial infection, and/or 2) there is a pre-emptive protective expression of AMPs to prevent an infection from occurring due to intestinal irradiation damage. Additionally, the disconnect between broad mRNA expression of AMPs and regional protein expression (Cma1 = rectum, Retnlb = cecum) suggests that translational or post-translational regulation of protein production governs regional expression. The mechanism of this regulation, as well as its functional implications would be interesting to pursue further.

Loss of stemness and proliferation

One explanation for the sensitivity of stem cells to irradiation is that they express higher levels of p53 mRNA (*Trp53*) [15,39–41]. We observed elevated *Trp53* mRNA expression levels in stem cells, as well as DCS, and proliferative TA and absorptive cells, even prior to irradiation (Figure 3.4b). There are two key functions of p53 following cellular stress and DNA damage: cell cycle arrest, enabling time for DNA repair prior to replication, and triggering apoptosis when the damage is too severe [42]. p21 (*Cdkn1a*) is a canonical p53 target gene that has been previously suggested to promote cell cycle arrest, rather than apoptosis; it has also been reported to carry out other diverse cellular activities [43]. We observed broad *Cdkn1a* expression in all cell types (albeit lowered in stem cells (Supplementary Figure 3.5a)), raising the question of whether this gene is promoting cell death or cell stalling/survival. Others have suggested that the level of p53 can be indicative of whether cell cycle arrest or apoptosis is the outcome, with higher levels of p53 promoting apoptosis and lower levels promoting survival [42]. Consistent with this, we observe that stem and DCS, the cell populations that are rapidly lost upon wounding, have the highest levels of p53 target gene expression (Supplementary Figure 3.5a) [44,45]. Additionally, stem and DCS cells express Tumor Necrosis Factor (TNF) pathway ligands, death receptors, and signaling components needed to trigger p53-independent caspase cascades and apoptosis (Supplementary Figure 3.5b) [46,47]. Our data suggests that expression of p53 with TNF signaling components may be indicative of cell death sensitivities to damage. Interestingly, previous work found that there was an identical overlap between apoptotic cells and p53 expression in the small intestine, but a weaker correlation in the colon [41]. Instead, at the base of colon crypts there was a larger percent of cells undergoing apoptosis than were p53 positive, suggesting other p53-independent apoptotic mechanism may be also involved in radiation response in the colon [41].

One of the unanswered questions in intestinal wounding and repair is how stemness is regained. Our transcriptomic data showed a significant loss of stem cell gene expression by two days post irradiation, a loss that persisted through day 4 (Figure 3.2e). However, our histological data also showed that by day 4, a significant number of cecum crypt bases were populated with highly proliferative cells, and by day 7 all crypt bases were tightly packed with compact Pcna+ epithelial cells (Figure 3.1c, 3.5f-g). This would indicate that although stem cell signatures of expression are not yet detectable at day 4, highly proliferative cells with histological similarities to stem cells have repopulated the niche (Supplementary Figure 3.6). Examining of all cluster types on day 4, one cluster - cluster 11 (absorptive) - stands out for having the greatest Wnt signaling potential due to higher levels of expression of Wnt signaling components (*Ctnnb1*, *Lrp6*, *Gsk3b*, *Apc*, and *Tcf7l1* (Supplementary Figure 3.6f)). Even so, no cluster has detectable levels of canonical Wnt target gene expression, suggesting that the reactivation of Wnt signaling takes longer than four days. We hypothesize that lineage neutral proliferative TA cells (ProlifTA) and immature absorptive cells (AbsTA) rapidly expand to repopulate the crypt base by day 4. In turn, the surrounding fibroblast cells in the stroma, which are known to express and secrete Wnt ligands [48,49] transform these cells into stem cells through establishing appropriate morphogen gradients.

Ongoing work

This chapter has presented exciting preliminary data; however, work is still ongoing to analyze the data, including using it to mathematically model crypt proliferation and dynamics. Also, additional single cell RNA-sequencing analyses are ongoing. Mathematical modeling is currently being performed to establish a 3D dynamic crypt homeostasis between stem cell, TAZ, and differentiated cells [50]. The crypt model can be ‘wounded’ to lose stem cells and then simulate recovery. This type of modeling allows us to investigate how factors such as crypt length, patterns of proliferation, and the abundance of TAZ cells relative to lineage abundances influence the recovery. We also hope to model the striking overcompensation in crypt length we observe on day 7 since that suggests features of positive and negative feedback signaling for homeostasis. Additional single cell RNA-sequencing work is focused on time and pseudo-time analyses to understand changes in lineage trajectory and predict how cell populations are transitioning [51,52]. The goal is to define a clear trajectory of cell differentiation in WT (also built on what is previously published) and then determine how this trajectory is altered following irradiation. We also aim to identify which of the proliferative day 4 cell population are most similar to stem cells and determine where those cells might be derived from.

Although crypt epithelial cells perform important intestinal functions and have been the focus on this chapter, other surrounding non-epithelial cells also perform crucial activities. Colonic epithelial cells are supported by a surrounding stroma that contains fibroblasts and immune cells, as well as a vasculature built of endothelial cells, and muscle and fat layers. The impact of irradiation, and wounding in general, on these populations has not been as well studied and there is much to do in this regard. For example, we observed gross anatomical differences in vasculature in the small intestine 4 days post irradiation (Figure 3.1a). Consistently, apoptosis of microvascular endothelial cells was previously published to occur following irradiation and contributes to the severity of the wound [53,54]. Endothelial apoptosis was inhibited by FGF, which then decreased crypt damage and organ failure/death [53,54]. Wnt ligands from intestinal stromal macrophages have been shown to be essential for long term survival and small intestine regeneration following irradiation [55]. Bone marrow stroma cell transplants promoted mouse survival following otherwise lethal doses of abdominal irradiation [56] and extracellular vesicles derived from mesenchymal stromal cells also decreased irradiation wounding in the small intestine [57]. These data highlight the interconnectedness of the intestine, and that all cell types play a role in maintaining homeostasis and wound repair. Future work should further investigate the wounded microenvironment because it plays a role not only in normal wound repair, but also when there is a lack of wound repair such as in chronic wounding events or in the development of cancer. Ultimately, defining the normal wounding and repair processes that occur in the colon will help to develop better strategies to treat chronic wounding such as IBD or even cancer, where the repair process is hijacked or entirely absent (see Discussion chapter 5).

METHODS

Mouse irradiation

All animal procedures were performed under the approval of the Institutional Animal Care and Use Committee at the University of California, Irvine. C57BL/6J male mice from Jackson Lab aged 7-10 weeks were subjected to 12 Gy whole body X-ray irradiation. Mice were harvested on post-irradiation days #1, 2, 3, 4, and 7 along with unirradiated control mice.

Tissue processing for histology

Mouse colons (cecum to rectum) and small intestine were dissected and processed for histology following previously published guidelines [58]. Briefly, intestines were flushed with cold PBS, linearized, and partially fixed with Bouin's fixative and then swiss rolled. Tissue was then fixed in formalin overnight and embedded in paraffin. 5 µm formalin-fixed paraffin-embedded (FFPE) sections were cut onto SuperFrost Plus slides.

Immunohistochemistry

Slides were deparaffinized and rehydrated. For H&E, slides were stained by hematoxylin and eosin, dehydrated, and mounted using Permount Mounting Medium (Fisher SP15-100). For antibody staining, antigen retrieval was performed in a pressure cooker using the optimized buffer for each antibody (specified below) for five minutes at pressure. Slides were blocked in 3% H_2O_2, goat serum, and then avidin and biotin blocking reagents (Vector Labs SP-2001). Sections were incubated in primary antibodies: anti-β-catenin (Cell Signaling 8480P; 1:100 4 °C overnight incubation; 10 mM pH 6 sodium citrate retrieval buffer), anti-Pcna (Abcam ab18197; 1:500 RT for 30 min; 10 mM pH 6 sodium citrate retrieval buffer), anti-Cma1 (Mast Cell Chymase; Abcam ab233103; 1:100 RT for 30 min; 10 mM pH 6 sodium citrate retrieval buffer), and anti-Retnlb (Relmβ; Abcam ab11429; 1:500 RT for 30 min; 10 mM pH 8 tris EDTA retrieval buffer). A biotinylated secondary antibody was used (goat-anti-rabbit; 1:200; Vector Labs BA-1000-1.5) for visualization using a peroxidase-conjugated avidin-based Vectastain protocol (ABC Elite – Vector Labs PK-4001; DAB Quanto – Fisher TA060QHDX). Slides were then counterstained with hematoxylin and mounted using Permount Mounting Medium. Whole slide scans were performed with a Roche Ventana DP 200 slide scanner and then analyzed using QuPath (https://qupath.github.io/) [59] and ImageJ [60]. Crypt length was measured from the lowest crypt base cell to the furthest cell at the top. Only measurements of crypts that were sectioned with the full length visible were measured. Each condition (WT or day post irradiation) had two biological replicate mice. H&E samples were assessed from each replicate and two sets of counts were taken of each replicate starting in the center (rectum) and counting in opposite directions through the layers of the swiss roll towards the cecum. In each of the 6-7 layers, 5 nearby crypts were counted, for a minimum of 120 crypt lengths measured per condition. Line graphs of crypt length were generated in GraphPad Prism (version 6.01) and two-tailed unpaired t-tests were performed and p-values are reported as * < 0.05, ** <0.01, *** < 0.001.

Tissue processing for FACS and sequencing

Irradiated and control mouse colons (caecum to rectum) were processed as previously published for isolation of crypt cell populations with flow sorting [15,61] with several modifications and additional steps to enable Multi-seq multiplexing (see methods section "Multi-seq cell hashing") [62] and 10x single cell sequencing. Tissue was dissociated in 2 mM EDTA and 10 µM Rock inhibitor and then aggressively shaken and filtered through 100 and then 40 µm filters. FACS buffer (PBS with 3% FBS + 10 µM Rock inhibitor) was not yet used as serum can interfere with Multi-seq, instead PBS with a Rock inhibitor supplement was used. The cell suspension was DNAse treated and centrifuged. Multi-seq reagents were then added (5x recommended volumes to account for the large number of cells). First the Anchor:Barcode mix and then the Co-Anchor mix with 5 min incubation after each. FACS buffer was then added and each sample centrifuged and processed following the previously published protocol including use of flow antibodies:

CD45-BV510 (1:200, Clone 30-F11; BD Biosciences #563891), CD31-BV510 (1:200, Clone MEC 13.3; BD Biosciences #563089), CD326-eFluor450 (1:100, Clone G8.8; eBioscience #48-5791-82), CD44-PerCP-Cy5.5 (1:100, Clone IM7; Thermo Fisher #A26013), CD24-PECy7 (1:200, Clone M1/69; eBioscience #25-0242-82), and CD117-APC-Cy7 (1:100, Clone 2B8; Thermo Fisher #A15423) and Live/Dead Aqua (Thermo Fisher # L34957) [15,61].

FACS for sequencing

Colonic crypt cells post-irradiation days 1, 2, 3, and 4 were collected for sequencing. Cells were sorted on a BD FACS Aria Fusion using a 100 µm nozzle (20 PSI) at a flow rate of 2.0 with a maximum threshold of 5000 events/sec. Live Epcam + cells were sorted directly into 10x-sequencing RT buffer. Each sample contained 5,000 – 10,000 Epcam + cells which came from 4 mice multiplexed together (with Multi-seq). The four mice consisted of one control mouse and three independent irradiated mice. In total, 4 control mice were sequenced, and 12 irradiated mice (n=3 for each post-irradiation day). FACS plots and analyses of colonic crypt populations were performed using BD FACSDiva software. Abundance of live epithelial cells was determined by normalizing Epcam+ cells from each sample to the number of single cells (as determine via flow with forward/side scatter) compared to the control for each day.

Multi-seq cell hashing

Samples were subjected to the Multi-seq method of cell hashing [62] and Multi-seq reagents (Anchor LMO and Co-Anchor LMO) and protocols received from Zev Gartner Lab (UCSF) were followed (protocol version dated December 21, 2018). Sample specific barcodes (BC1-16) and Multi-seq primer, and RPI Index primer, Universal I5, and TruSeq RPI-1 (Table) were ordered from Integrated DNA Technologies, Inc. Library preparation instructions provided by the Gartner lab were followed. Following sequencing, samples were de-multiplexed following instructions provided at https://github.com/chris-mcginnis-ucsf/MULTI-seq using R (version 1.2.1335 with R version 3.6.1) package deMULTIplex version 1.0.2. In some cases, thresholds were optimized manually and a semi-supervised negative cell reclassification was performed to rescue some cells. Other packages used for de-multiplexing include: KernSmooth (v 2.23-15), reshape2 (v 1.4.3), Rtsne (v 0.15), stringdist (v 0.9.5.5), shortRead (v 1.42.0), ggplot2 (v 3.2.1), ggpubr (v 0.2.5), Seurat (v 3.1.4), dplyr (v 0.8.3), and umap (v 0.2.5.0).

Single cell RNA sequencing

FACS-sorted Epcam+ cells prepared at a concentration of approximately 1,000 cell/µL were used for 10x single cell RNA-sequencing. Library generation was performed following the 10x Chromium Single Cell 3' Reagents Kit V3 manufacturers guidelines with added Multi-seq preparation steps. cDNA and library preparation were analyzed using Qubit dsDNA high sensitivity for quantitation. 10x libraries were sequenced on the Illumina HiSeq 4000 platform to achieve an average of >30,000 reads per cell. Multi-seq barcodes were sequenced on the Illumina NovaSeq S4 to achieve >10,000 reads per cell.

Single cell sequencing analysis

Fastq files were aligned using 10x Genomics Cell Ranger (v 3.0.2) and the Cell Ranger accompanying reference mm10 genome (v 1.2.0). Further analysis was performed in RStudio version 1.2.1335 with R version 3.6.1. First, samples were de-multiplexed following the Multi-seq pipeline (details above). A total of 6,561 cells were able to be confidently de-multiplexed (control = 2,453; day1 = 1,029; day2 = 937; day3 = 1,247; day4 = 895). Quality control parameters were used to filter cells with >200 genes with a mitochondrial percentage under 15% which resulted in 2,830 cells selected for further analysis. Day1-4 and WT cells were merged and cell clustering performed using the Seurat R package (v 3.2.2) [63], including column normalization and log-transformation. Principal component analysis (PCA) was performed, and based on JackStrawPlots and Clustree analysis[19] dimensions 1:20 and resolutions of 0.2 and 1.3 were used to identify 7 and 18 clusters, respectively. These clusters were then displayed using UMAP. Cluster cell identities were identified using FindAllMarker gene lists and known intestinal epithelial cell population markers. Line graphs of population abundances were generated in GraphPad Prism (version 6.01). Additional subsetting and subclustering was performed and both JackStrawPlots and Clustree were used to determine appropriate clustering resolutions/dimensions. Subclustering of *Ctnnb1*-high expressing cells was performed with a filter of ‘Ctnnb1>2’, dimension 1:10 and resolution of 0.4 which resulted in 497 cells. Cluster0 (409 cells) were identified and subclustered for further investigation with dimension 1:5 and resolution 0.4. A subset of proliferative cells (763 cells) was identified as ‘Pcna>0 | Mki67>0 | Top2a>0’ and subclustered with a dimension of 1:17 and resolution 0.5.

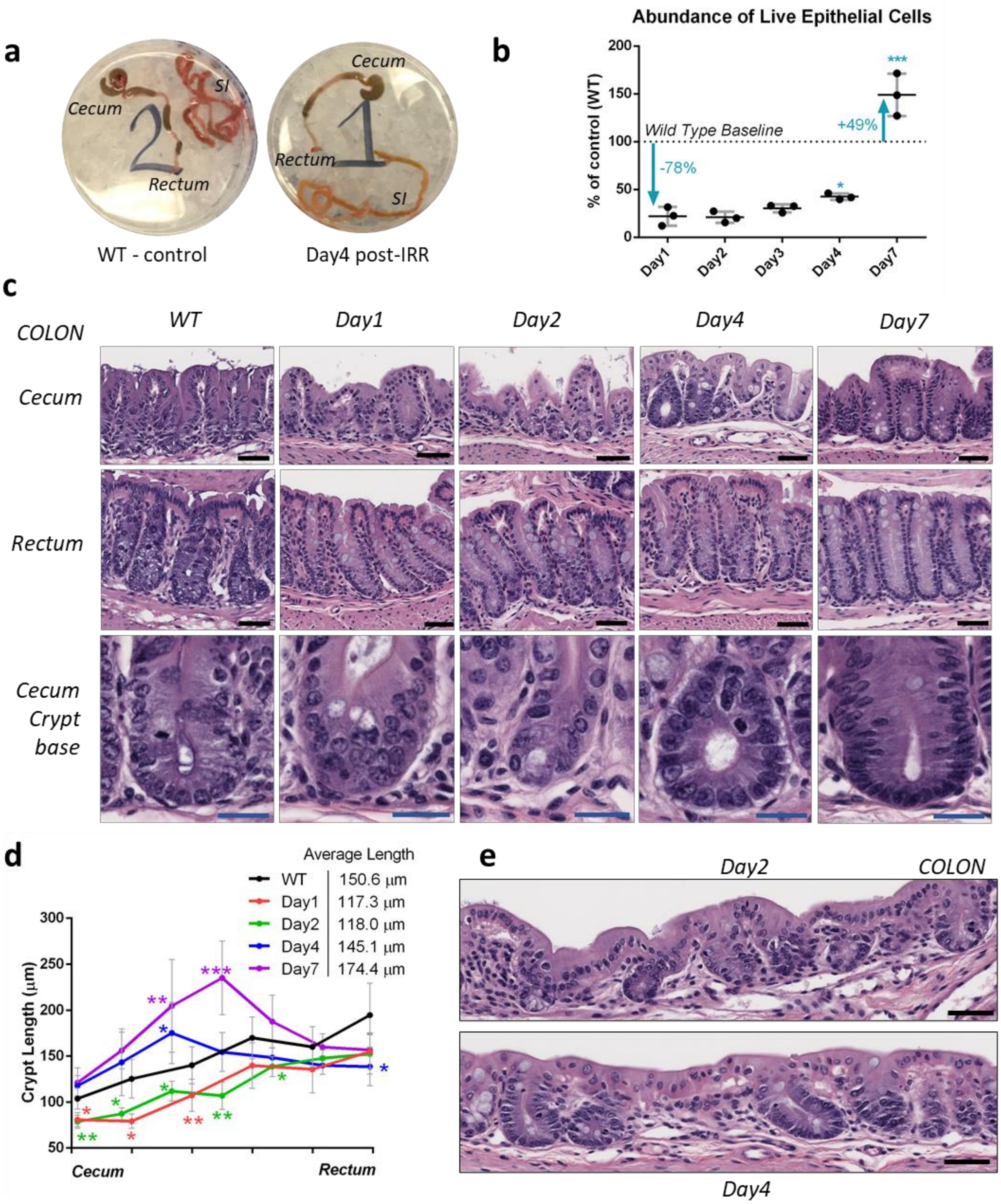

a
Cecum
SI
Rectum
2
WT - control
Cecum
Rectum
SI
1
Day4 post-IRR
b
Abundance of Live Epithelial Cells
% of control (WT)
200
150
100
50
0
Wild Type Baseline
-78%
+49%
*
***
Day1
Day2
Day3
Day4
Day7
c
COLON
WT
Day1
Day2
Day4
Day7
Cecum
Rectum
Cecum Crypt base
d
Average Length
WT 150.6 µm
Day1 117.3 µm
Day2 118.0 µm
Day4 145.1 µm
Day7 174.4 µm
Crypt Length (µm)
300
250
200
150
100
50
Cecum
Rectum
e
Day2
COLON
Day4

Figure 3.1: ***Assessment of intestinal damage and recovery 1-7 days following irradiation.***

a Freshly isolated mouse colon (cecum to rectum) and small intestine (SI) from a wild type (WT)/control mouse compared with Day4 post whole body 12 Gy irradiation. Irradiated intestine is shorter in length, shows decreased vascularization and increased mucus production (particularly evident in the SI), and contains loose stool (most evident in colon). **b** FACS quantitation of live colonic epithelial cells following irradiation. Three independent biological replicate mice were used for each time point, compared to a control mouse. Two-tailed unpaired T-tests were performed comparing Day2,3,4, and 7 with Day1 and p-values are reported as * < 0.05, ** <0.01, *** < 0.001. **c** H&E staining was performed on intestine swiss rolls of WT and irradiated mice. Colons were processed as swiss rolls of the cecum through to the rectum (Supplementary Figure 3.1a). Black scale bar represents 50 µm and blue scale bar represents 25 µm. **d** Crypt lengths were measured in the layers of WT and irradiated mouse colon swiss rolls (Supplementary Figure 3.1b). Two-tailed unpaired T-tests were performed and p-values are reported as * < 0.05, ** <0.01, *** < 0.001. **e** In addition to quantified changes in crypt length, there are also changes in the frequency and abundance of crypts. These regions which are located near the cecum have crypts that are less tightly packed with large batches of stroma as a result of irradiation. Black scale bar represents 50 µm.

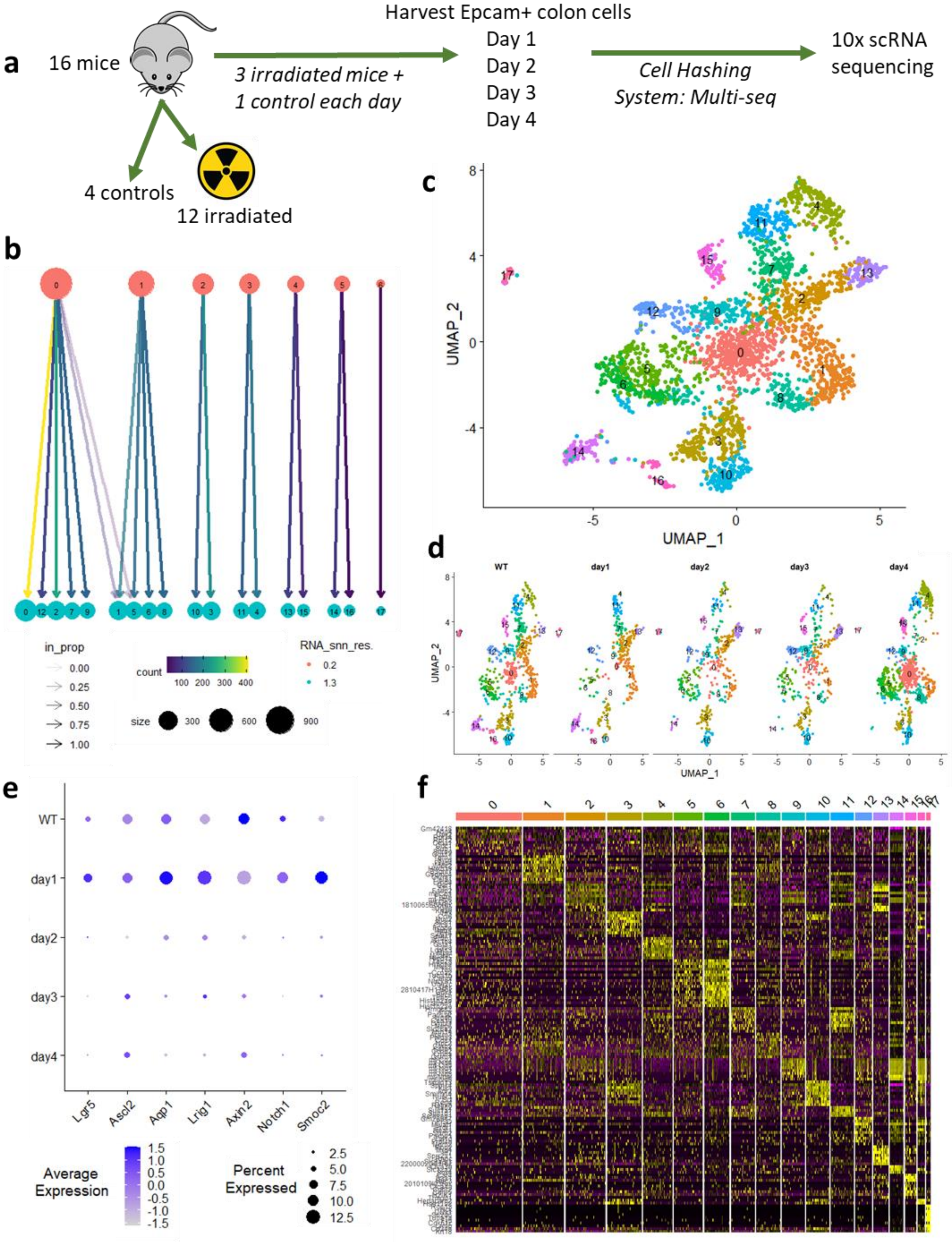
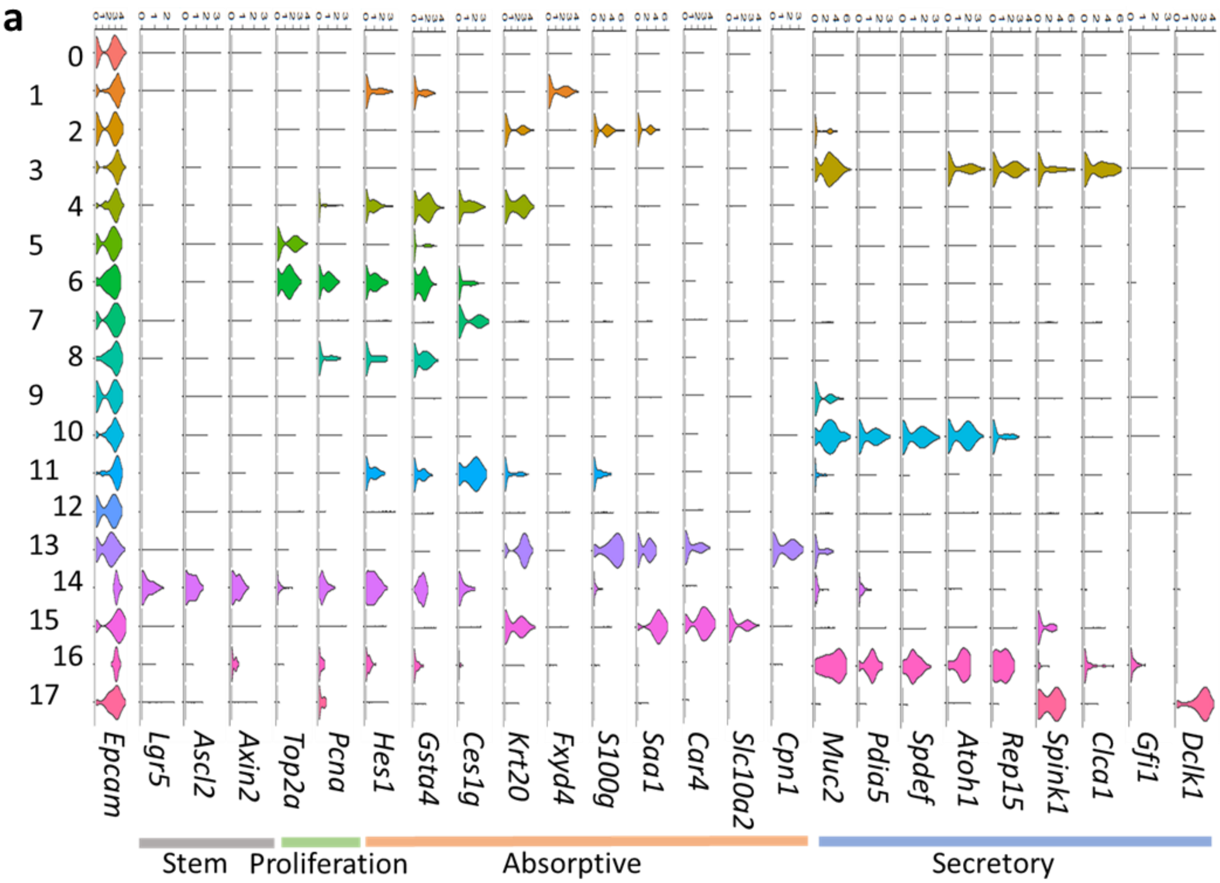

a
16 mice
4 controls
12 irradiated
3 irradiated mice +
1 control each day
Harvest Epcam+ colon cells
Day 1
Day 2
Day 3
Day 4
Cell Hashing
System: Multi-seq
10x scRNA
sequencing
b
in_prop
0.00
0.25
0.50
0.75
1.00
count
100 200 300 400
size
300
600
900
RNA_snn_res.
0.2
1.3
c
UMAP_2
UMAP_1
d
WT
day1
day2
day3
day4
UMAP_2
UMAP_1
e
WT
day1
day2
day3
day4
Lgr5
Ascl2
Aqp1
Lrig1
Axin2
Notch1
Smoc2
Average Expression
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
Percent Expressed
2.5
5.0
7.5
10.0
12.5
f
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

Figure 3.2: ***Single cell RNA-sequencing analysis of the colonic epithelial crypt response to irradiation.***

a Schema describing the single cell RNA-sequencing experiment performed on 16 mice (controls and whole body 12 Gy irradiation). Samples were multiplexed using the cell hashing system Multi-seq (see Methods section) [62]. **b** Clustree analysis was used to determine appropriate resolution for clustering [64]. A minimal resolution (0.2) and a high resolution (1.3) were used. This analysis also highlighted related clusters (for example proliferation, absorptive, or secretory). **c** UMAP display of all cells (WT + irradiated) with 1.3 resolution resulting in 18 clusters. **d** UMAP display showing that most clusters are generally composed of cells from WT and all days post irradiation. The disappearance of clusters over time or expansion of cells in a cluster is biologically relevant (i.e loss of stem cells cluster following wounding). **e** Dotplot display of all clusters shows a swift drop in stem cell marker expression by day2 post irradiation. **f** Heatmap of top 10 markers genes for each.



b

Cluster Number	Cell Type
0	Mixed
1	Abs-1
2	Enterocyte-1
3	Goblet
4	Abs TA-1
5	Proliferative TA-1
6	Proliferative TA-2
7	TA
8	Abs TA-2
9	Mixed
10	SecPro
11	Abs-2
12	Mixed
13	Enterocyte-2
14	Stem
15	Enterocyte-3
16	Deep Crypt Secretory
17	Tuft

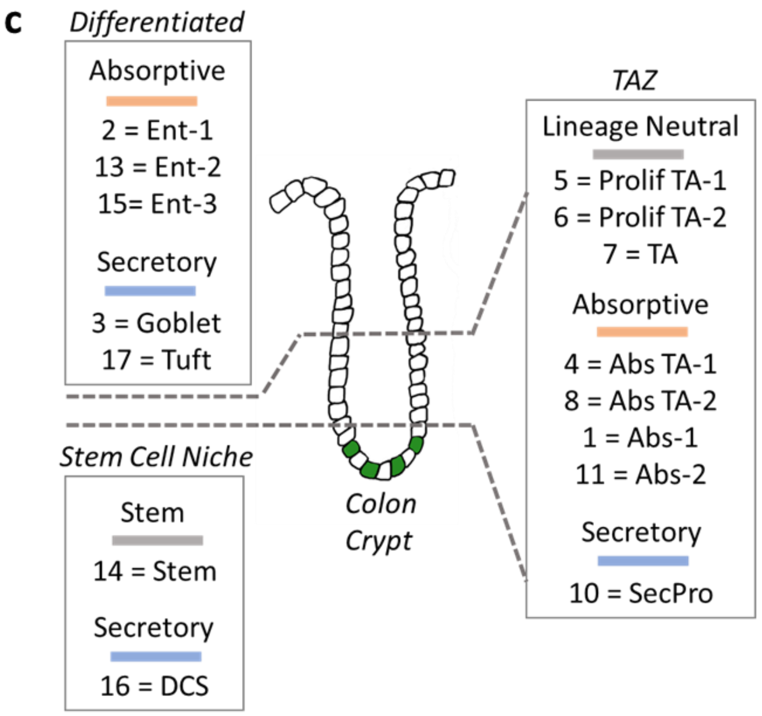


Figure 3.3: ***Identification of colonic epithelial crypt populations.***

a Violin plots showing mRNA expression of known intestine crypt marker genes (Stem, proliferative, absorptive, and secretory) for 18 clusters. **b** Assignment of cell type to clusters based on marker gene expression. A few clusters labeled as "mixed" do not have clear identities. Cluster 0 is further investigated in Supplementary Figure 3.4. Cluster 9 and 12 contain high levels of mitochondrial reads as well as long-noncoding RNA and as such, appear to be dying/apoptotic cells. Cluster 0, 9, and 12 exhibit expression signatures indicative of a secretory cell lineage. **c** Schematics of intestinal cell types in their predicted relative crypt position at either the bottom of the crypt in the stem cell niche (pattern of stem cells highlighted in green), in the transit amplifying zone (TAZ) or labeled as a differentiated cell. Although enterocytes are generally found above the TAZ, tuft cells are frequently found in the TAZ and throughout the crypt. Although a few enteroendocrine markers expressing cells were identified (ie. *Chga* and *Chgb positive*) these cells did not segregate into a unique cluster. Identified cell types include deep crypt secretory cells (DCS), transit amplifying (TA), proliferating (Prolif), absorptive (Abs), secretory progenitors (SecPro), and enterocytes (Ent).

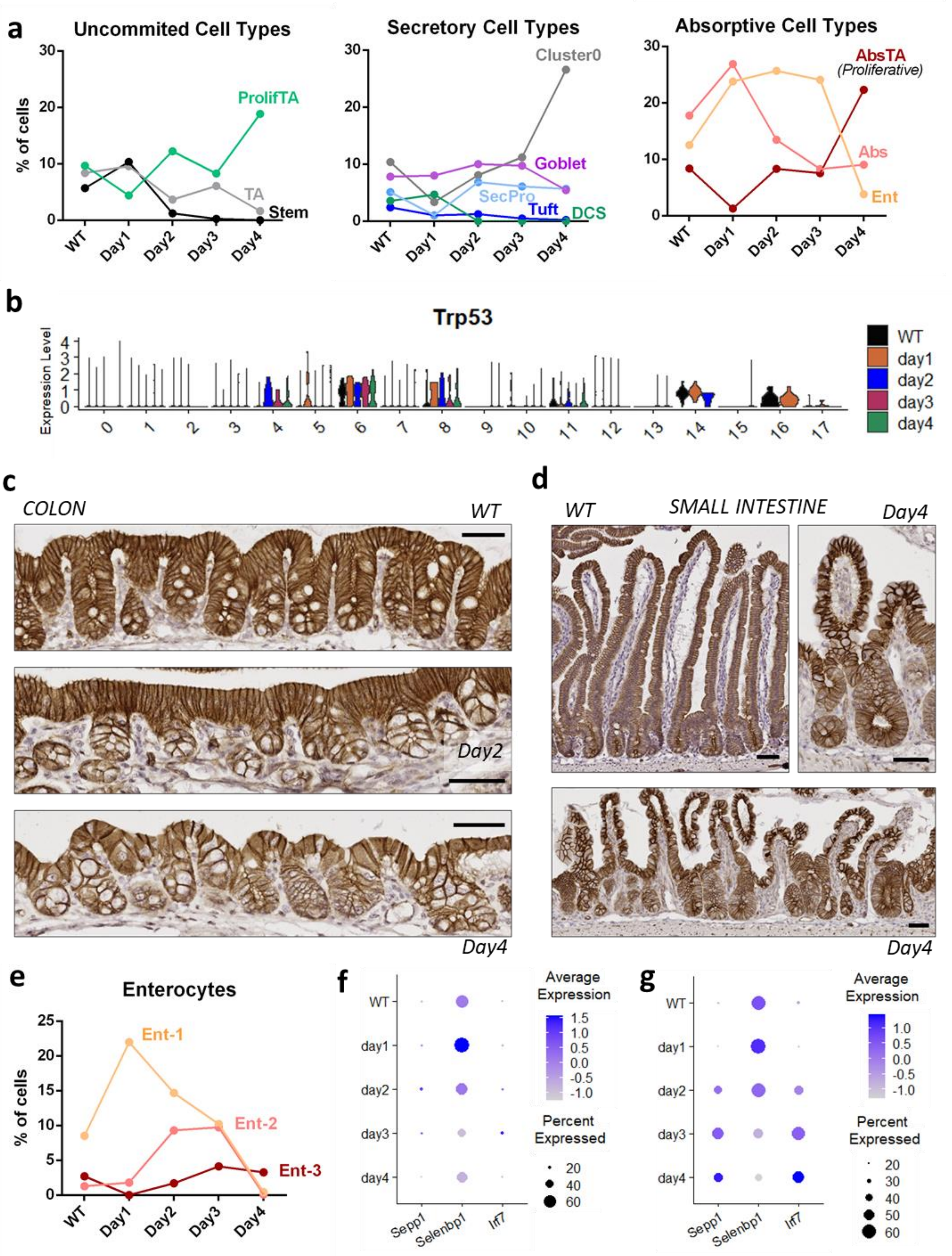

a
Uncommited Cell Types
% of cells
30
20
10
0
ProlifTA
TA
Stem
WT
Day1
Day2
Day3
Day4
Secretory Cell Types
Cluster0
Goblet
SecPro
Tuft
DCS
Absorptive Cell Types
AbsTA
(Proliferative)
Abs
Ent
b
Trp53
Expression Level
4
3
2
1
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
WT
day1
day2
day3
day4
c
COLON
WT
Day2
Day4
d
WT
SMALL INTESTINE
Day4
Day4
e
Enterocytes
% of cells
25
20
15
10
5
0
Ent-1
Ent-2
Ent-3
WT
Day1
Day2
Day3
Day4
f
WT
day1
day2
day3
day4
Sepp1
Selenbp1
Ifi7
Average Expression
1.5
1.0
0.5
0.0
-0.5
-1.0
Percent Expressed
20
40
60
g
WT
day1
day2
day3
day4
Sepp1
Selenbp1
Ifi7
Average Expression
1.0
0.5
0.0
-0.5
-1.0
Percent Expressed
20
30
40
50
60

Figure 3.4: Irradiation-induced changes in *colonic crypt cell populations.*

a Changes in cell type abundances post irradiation. Abundance percent is determined by the number of cells in a cluster on a day as a fraction of the total cells for that day. The cell types are pooled from the following clusters: ProlifTA = 5,6; TA = 7; Stem = 14; Cluster0 = 0; Goblet = 3; SecPro = 10; Tuft = 17; DCS = 16; AbsTA = 4,8; Abs= 1,11; and Ent = 2,13,15. Cluster0 is marked as a secretory cell type based on further Subclustering (Supplmentary Figure 3.9). **b** *Trp53* split violin plot highlighting cluster and day post irradiation expression. **c** IHC staining of Ctnnb1 in cecum. Black scale bar represents 50 µm. **d** IHC staining of Ctnnb1 in the small intestine. Black scale bar represents 50 µm. **e** Abundance of three enterocyte populations following irradiation. Dotplot displaying expression of *Sepp1*, *Selenbp1*, and *Irf7* in **f** all cell types following irradiation and **g** only enterocytes.

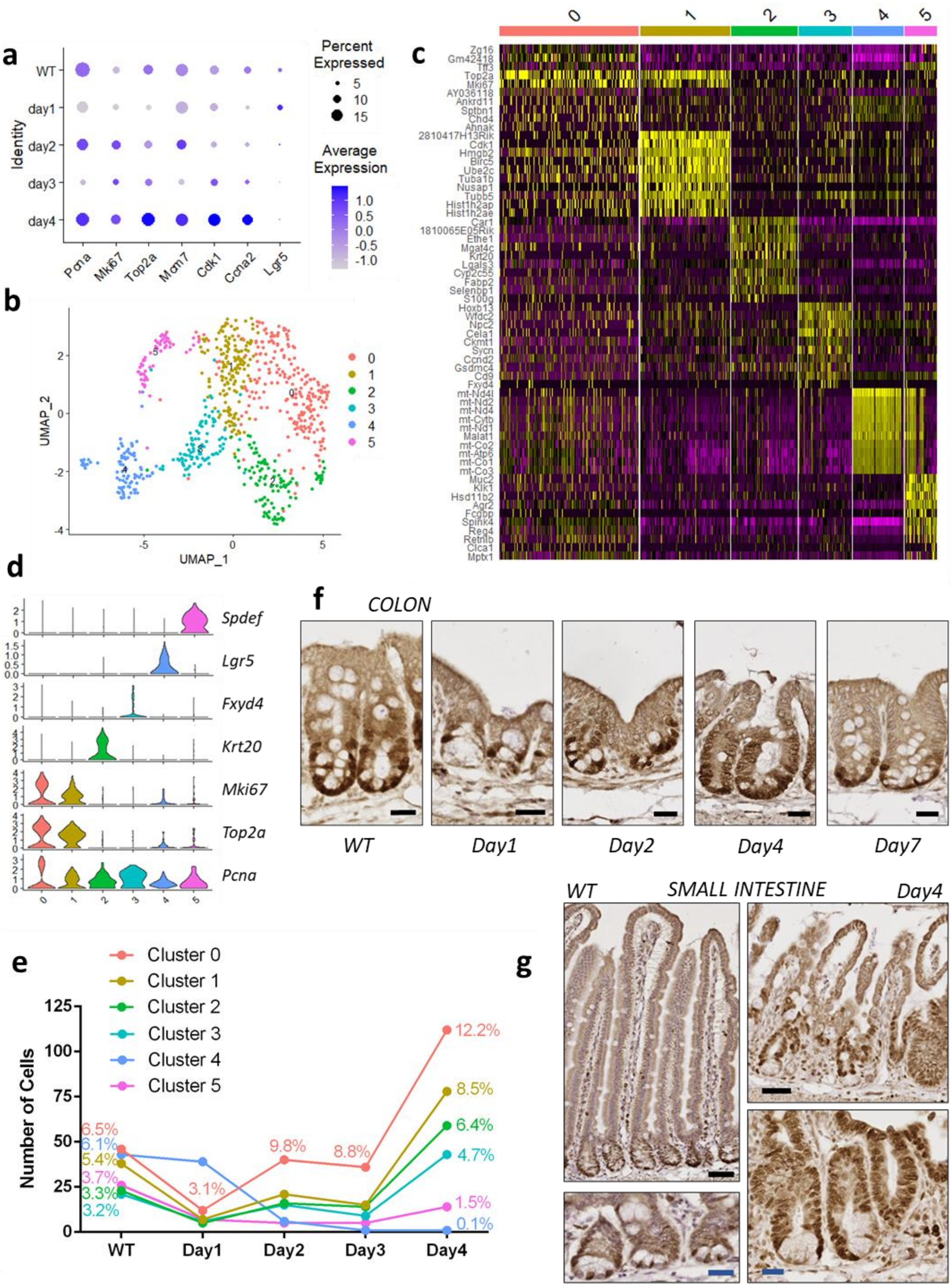

a
Percent Expressed
5
10
15
Average Expression
1.0
0.5
0.0
-0.5
-1.0
Identity
WT
day1
day2
day3
day4
Pcna
Mki67
Top2a
Mcm7
Cdk1
Ccna2
Lgr5
b
UMAP_2
UMAP_1
0
1
2
3
4
5
c
0
1
2
3
4
5
Zg16
Gm42418
Tff3
Top2a
Mki67
AY036118
Ankrd11
Sptbn1
Chd4
Ahnak
2810417H13Rik
Cdk1
Hmgb2
Birc5
Ube2c
Tuba1b
Nusap1
Tubb5
Hist1h2ap
Hist1h2ae
Car1
1810065E05Rik
Ethe1
Mgat4c
Krt20
Lgals3
Cyp2c55
Fabp2
Selenbp1
S100g
Hoxb13
Wfdc2
Npc2
Cela1
Ckmt1
Sycn
Ccnd2
Gsdmc4
Cd9
Fxyd4
mt-Nd4l
mt-Nd2
mt-Nd4
mt-Cytb
mt-Nd1
Malat1
mt-Co2
mt-Atp6
mt-Co1
mt-Co3
Muc2
Klk1
Hsd11b2
Agr2
Fcgbp
Spink4
Reg4
Retnlb
Clca1
Mptx1
d
Spdef
Lgr5
Fxyd4
Krt20
Mki67
Top2a
Pcna
e
Cluster 0
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Number of Cells
0
25
50
75
100
125
WT
Day1
Day2
Day3
Day4
6.5%
6.1%
5.4%
3.7%
3.3%
3.2%
3.1%
9.8%
8.8%
12.2%
8.5%
6.4%
4.7%
1.5%
0.1%
f
COLON
WT
Day1
Day2
Day4
Day7
g
WT
SMALL INTESTINE
Day4

Figure 3.5: ***Proliferation increases four days post irradiation.***

a Dotplot of all cells displaying increased expression of markers of proliferation on day4 post irradiation. Compare with loss of stemness (*Lgr5*). **b** UMAP displaying subclustering of proliferative cells (Pcna, Mki67, or Top2a >0) with 6 distinct populations. **c** Heatmap of cluster marker genes and **d** violin plot of proliferation and lineage markers. **e** Change in the number of cells in each cluster following irradiation, with percentages of total cells from each day overlaid. **f** IHC staining of Pcna in the cecum. Black scale bar represents 25 µm. **g** IHC staining of Pcna in the small intestine. Black scale bar represents 50 µm, blue scale represents 25 µm.

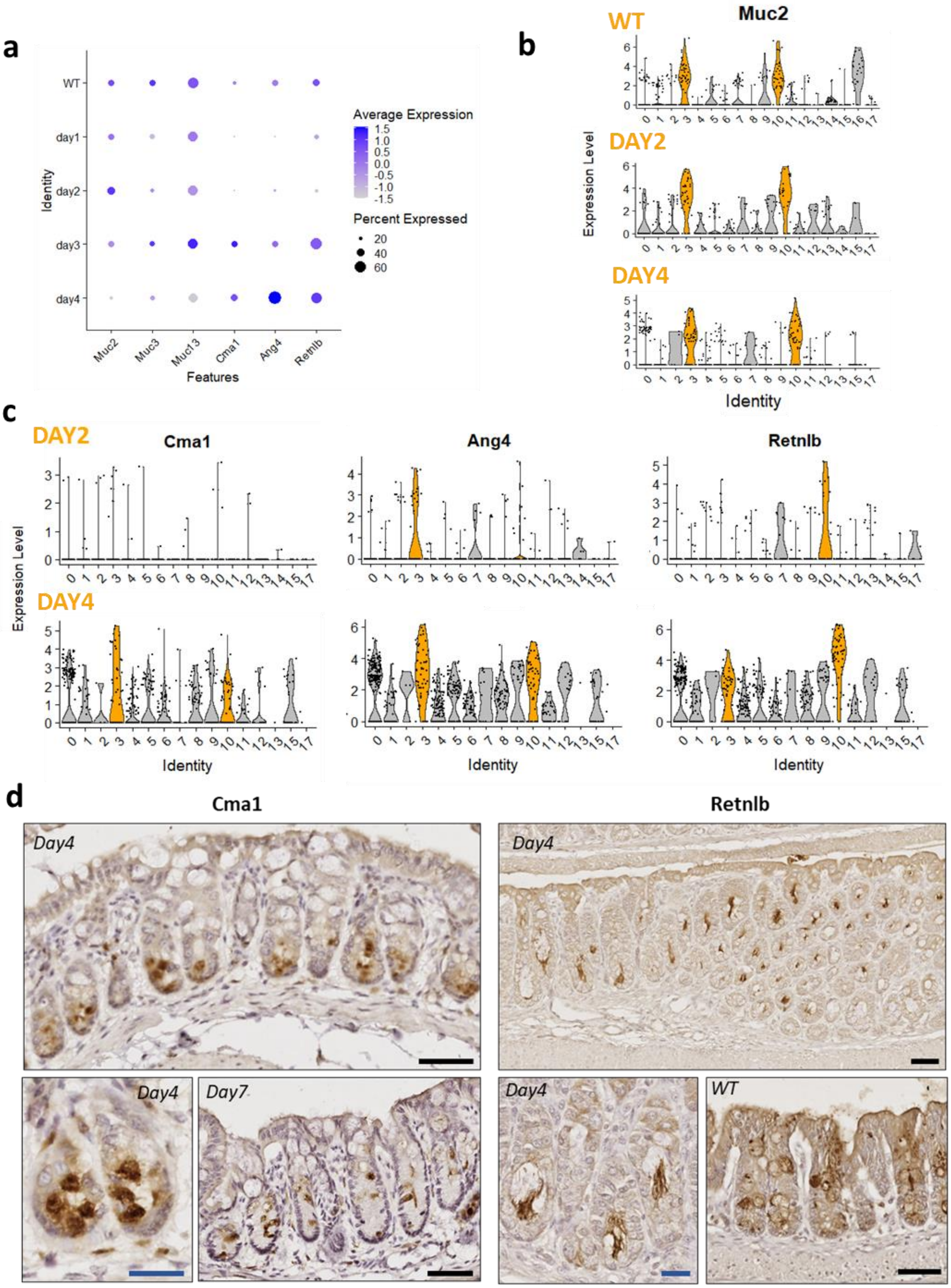
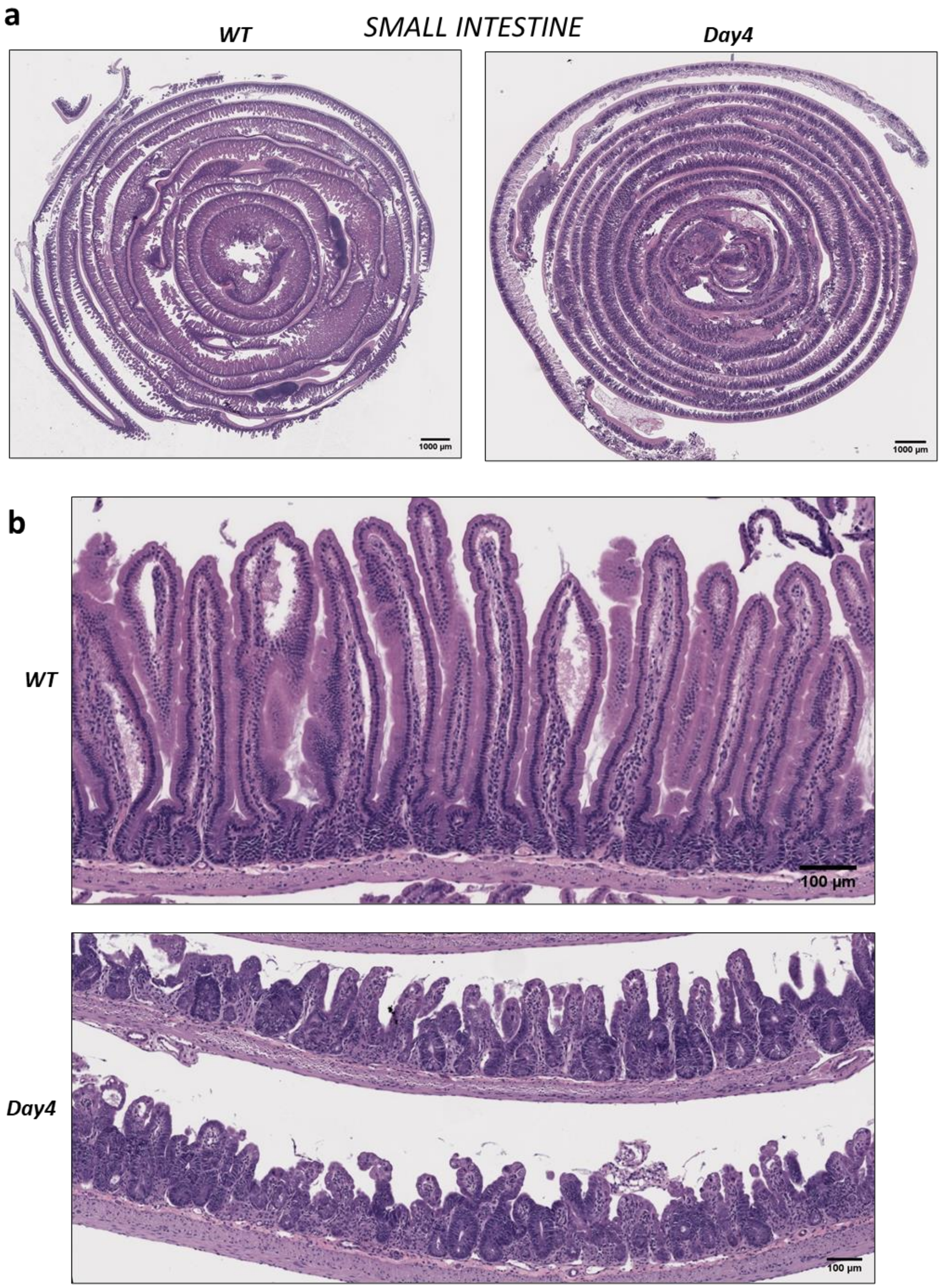

a
Identity
WT
day1
day2
day3
day4
Muc2
Muc3
Muc13
Cma1
Ang4
Retnlb
Features
Average Expression
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
Percent Expressed
20
40
60
b
Muc2
WT
DAY2
DAY4
Expression Level
Identity
c
DAY2
DAY4
Cma1
Ang4
Retnlb
Expression Level
Identity
d
Cma1
Retnlb
Day4
Day4
Day7
Day4
Day4
WT

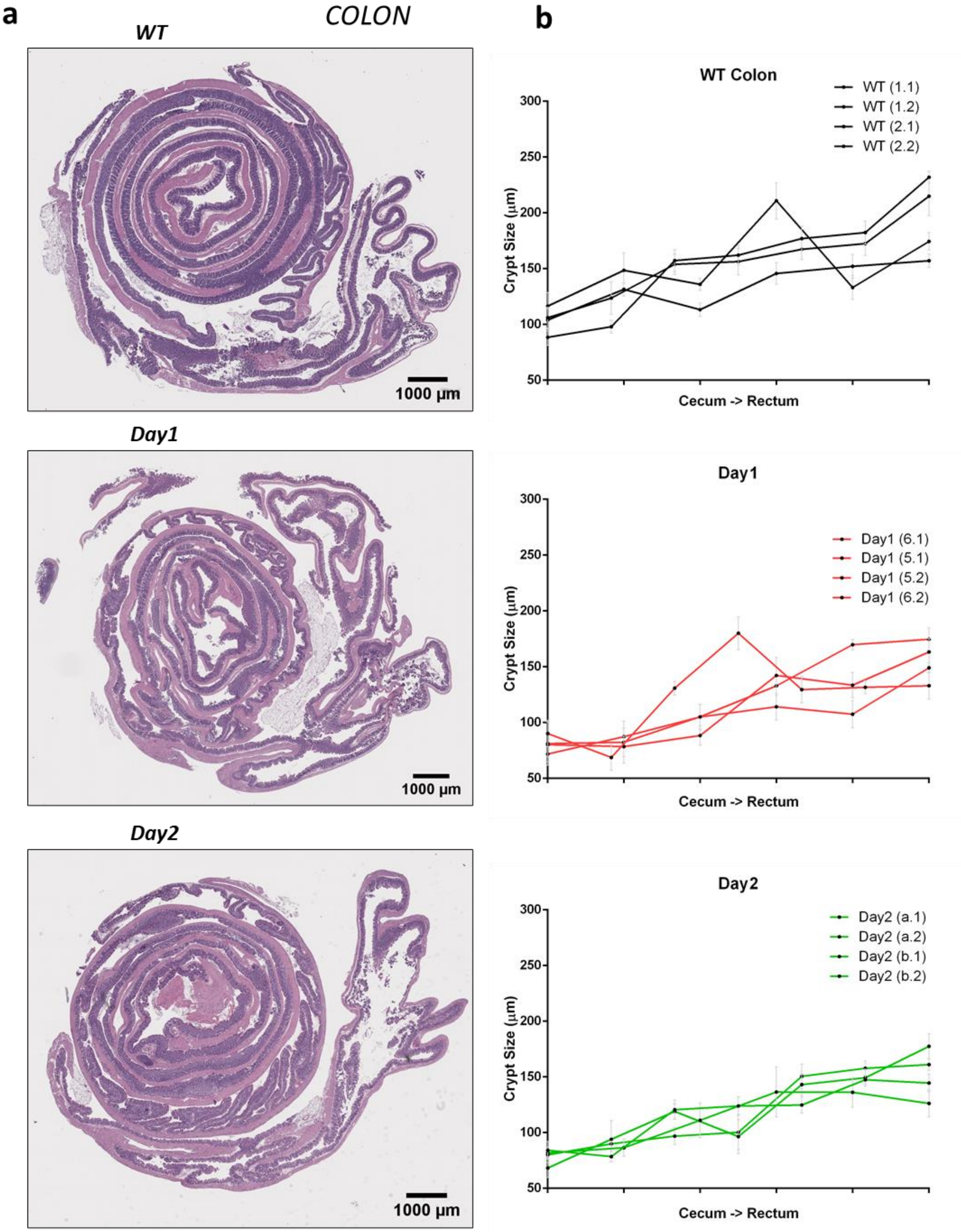
Figure 3.6: ***Mucin and Antimicrobial Peptide expression increases following irradiation.***

a Dot plot of all cells showing mRNA expression of genes that produce secreted proteins which are altered following irradiation. **b** Violin plots of *Muc2* expression in all clusters in WT, day2, and day4. Goblet and SecPro clusters are highlighted in orange. **C** Violin plots of antimicrobial peptide genes, *Cma1*, *Ang4*, and *Retnlb*, with decreased expression on day2 post irradiation and increased expression on day4. Goblet and SecPro clusters are highlighted in orange. **d** IHC staining of Cma1 and Retnlb in the colon. Black scale bars represent 50 µm, blue scale bars represent 25 µm.

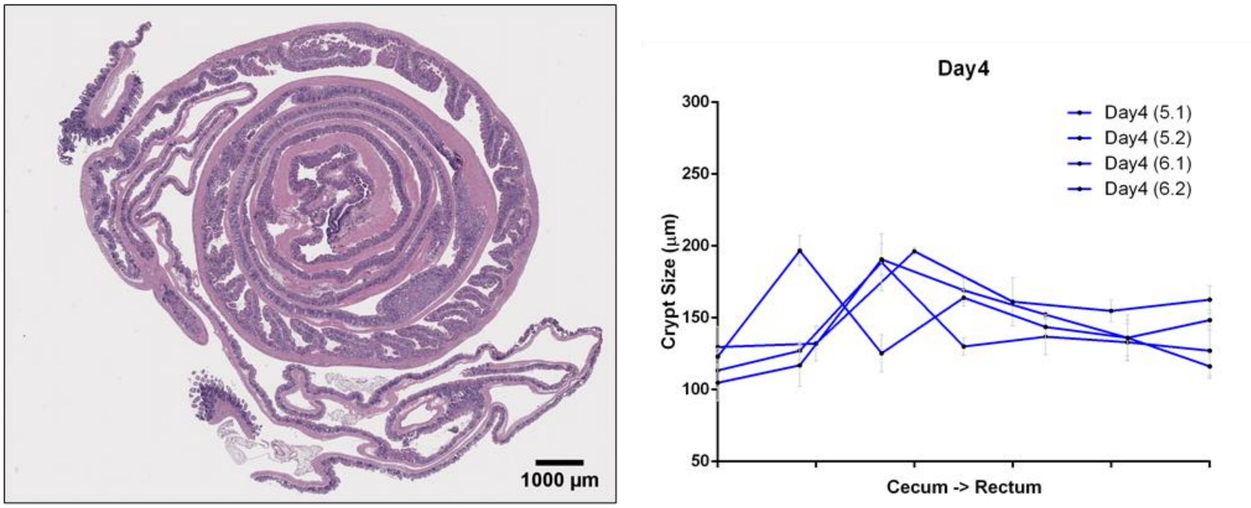


Supplementary Figure 3.1: ***Murine small intestine displays severe villi damage following irradiation.***

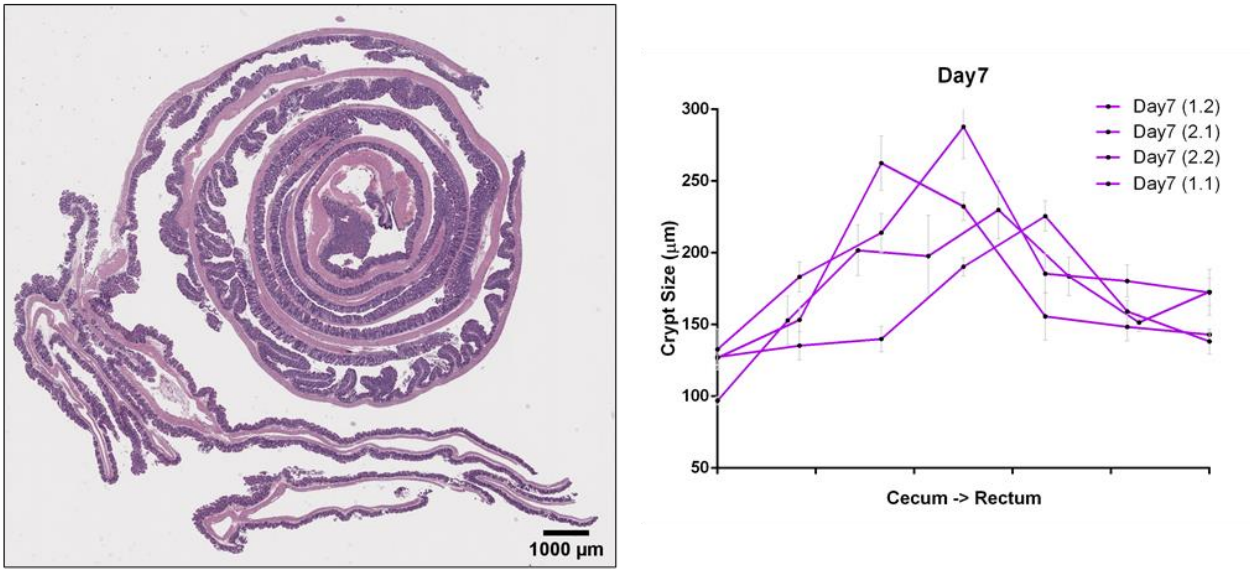
a Small intestine swiss roll of a wild type versus day4 post irradiation. **b** Close up views show the small intestine crypt and large villi which are severely damaged and truncated day4 post irradiation.


a
COLON
WT
1000 µm
Day1
1000 µm
Day2
1000 µm
b
WT Colon
WT (1.1)
WT (1.2)
WT (2.1)
WT (2.2)
Crypt Size (µm)
300
250
200
150
100
50
Cecum -> Rectum
Day1
Day1 (6.1)
Day1 (5.1)
Day1 (5.2)
Day1 (6.2)
Crypt Size (µm)
300
250
200
150
100
50
Cecum -> Rectum
Day2
Day2 (a.1)
Day2 (a.2)
Day2 (b.1)
Day2 (b.2)
Crypt Size (µm)
300
250
200
150
100
50
Cecum -> Rectum

Day4



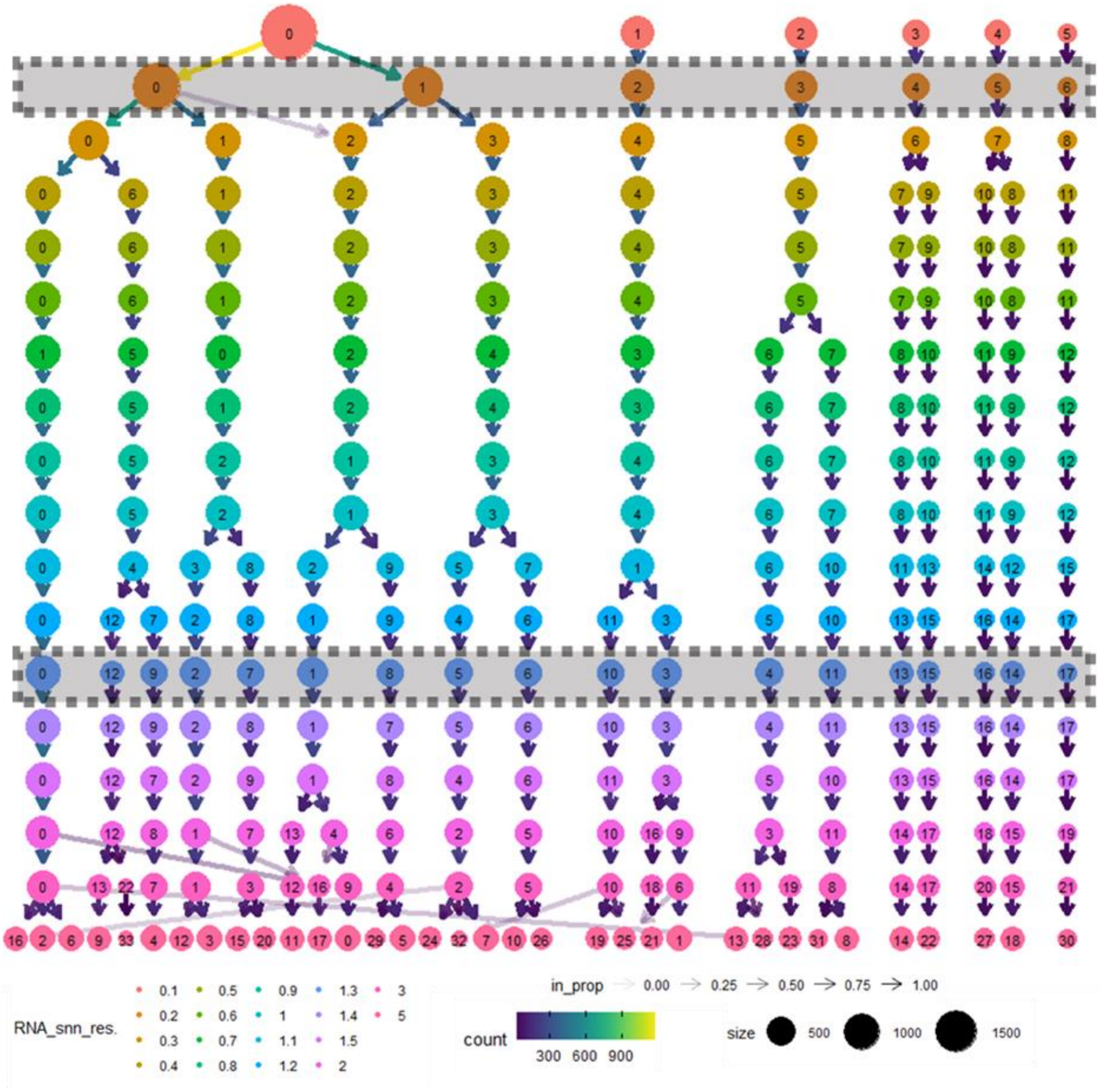
Day7



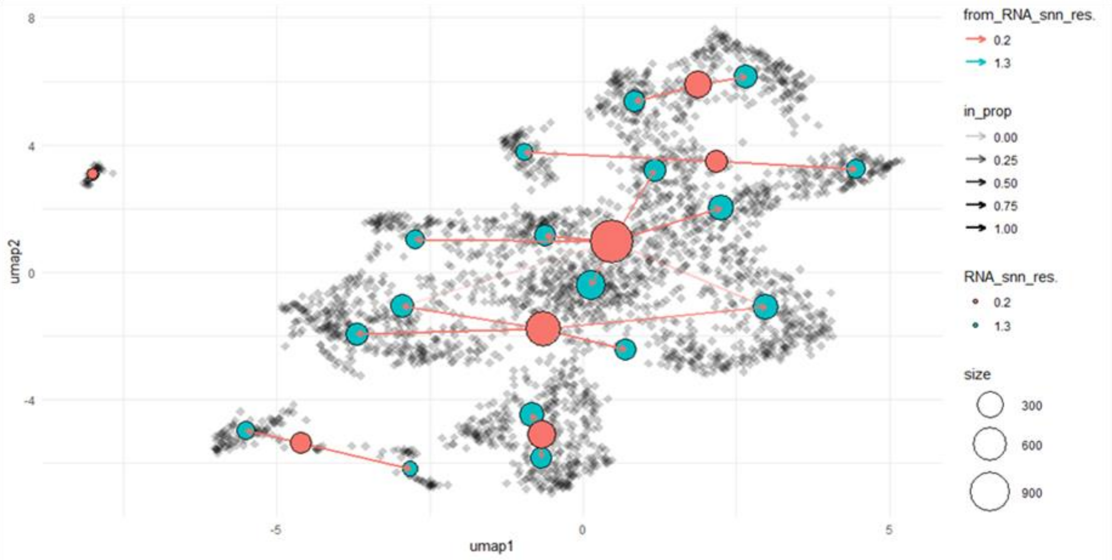
Supplementary Figure 3.2: Murine c*olonic crypt displays regional differences in length and wounding severity from irradiation.*

a Murine colonic swiss rolls with H&E staining. Rectum is in the center and the cecum is on the outer layer. **b** Crypt length measurements taken of layers in the colon swiss rolls. Two independent biological replicate mice were used for each condition. One H&E section from each mouse was measured twice, with 5 crypts measured in each position.

a

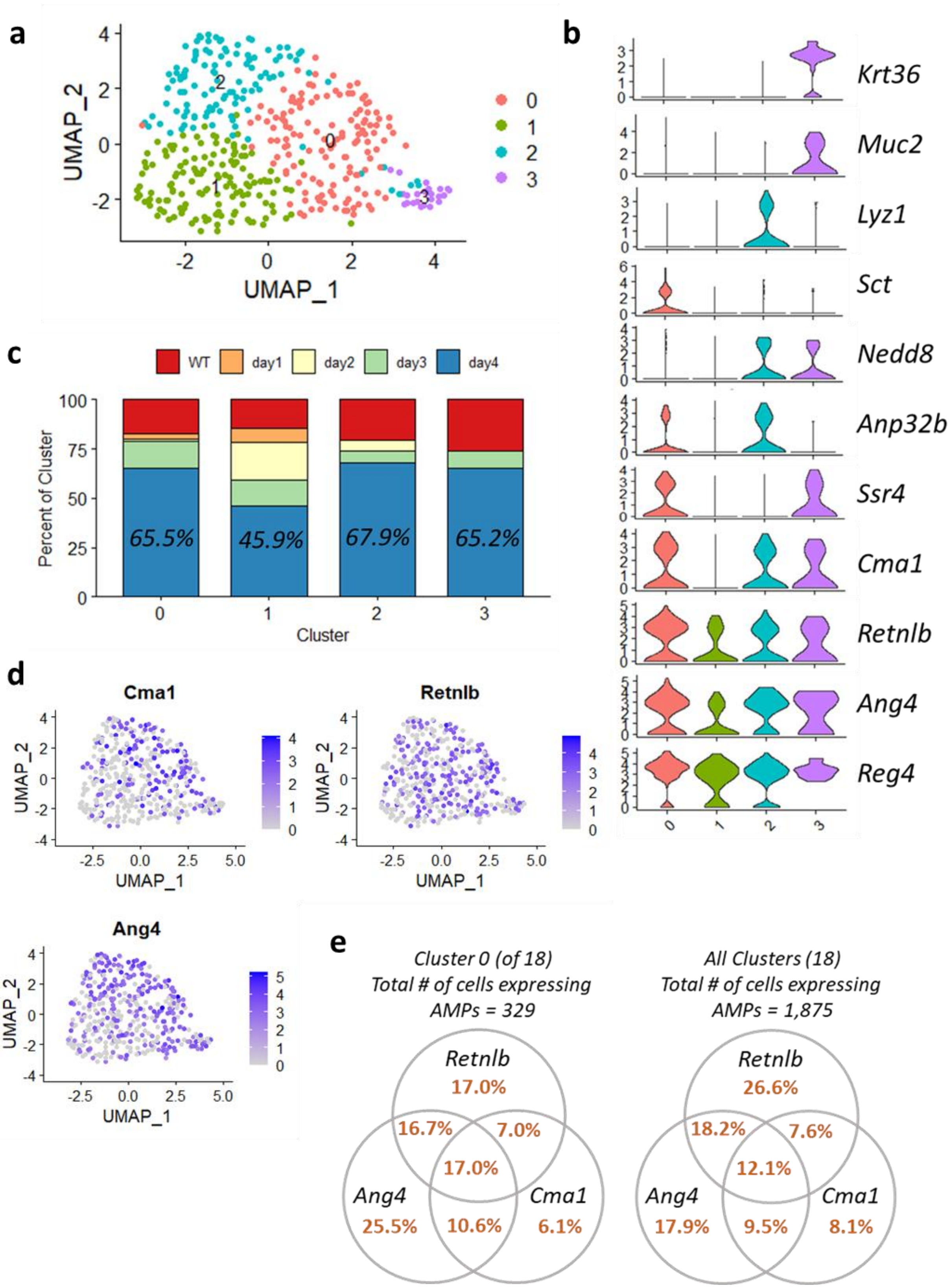

RNA_snn_res.
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
1.4
1.5
2
3
5
in_prop
0.00
0.25
0.50
0.75
1.00
count
300
600
900
size
500
1000
1500

b


umap2
8
4
0
-4
umap1
-5
0
5
from_RNA_snn_res.
0.2
1.3
in_prop
0.00
0.25
0.50
0.75
1.00
RNA_snn_res.
0.2
1.3
size
300
600
900

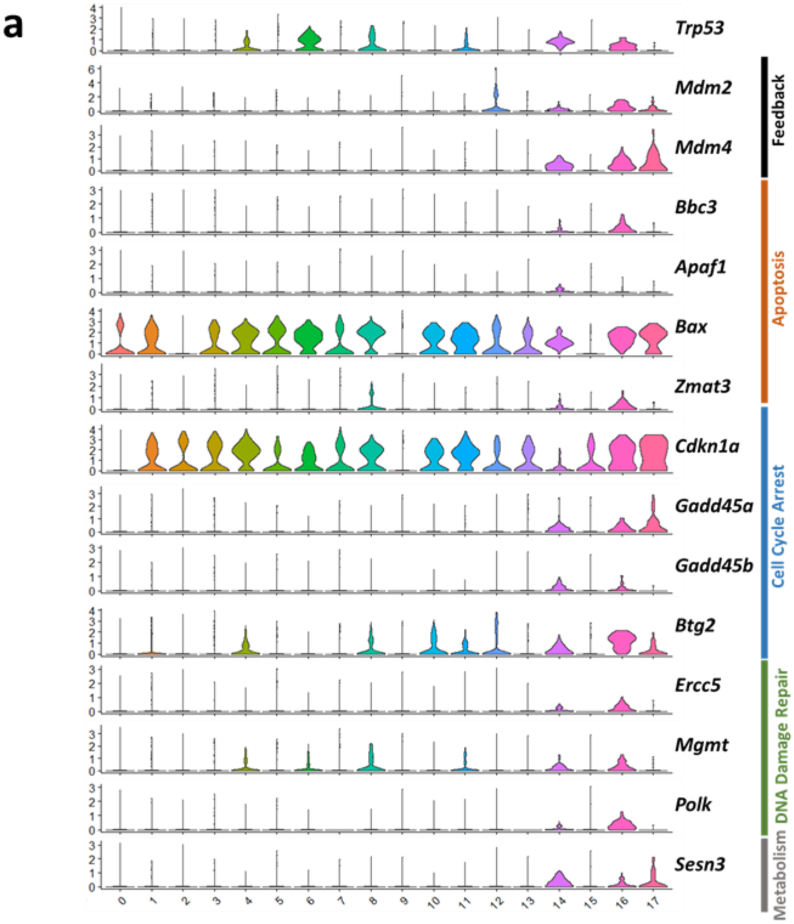
Supplementary Figure 3.3: ***Clustree analysis of single cell RNA-sequencing.***

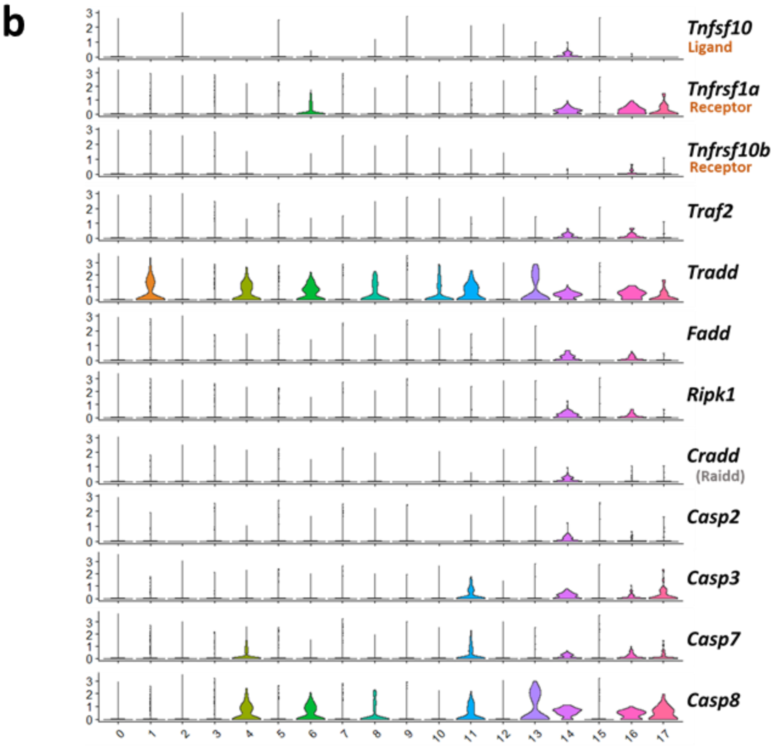
Clustree analysis was used to determine an optimal clustering level by testing the relationship between clusters formed with variable resolutions. **a** The “Clustree” tree shown was created using cells from all days and resolution 0.1 to 2 with each horizontal row resulting from the specified resolution. There is minimal interchange between clusters and many clusters that stay consistent at multiple resolutions. Highlighted in grey are the resolution and cluster numbers chosen for further analysis: 0.2 (7 clusters) and 1.3 (18 clusters). **b** The location of clusters identified at resolutions 0.2 and 1.3 are overlayed on a UMAP.


a
UMAP_2
UMAP_1
0
1
2
3
b
Krt36
Muc2
Lyz1
Sct
Nedd8
Anp32b
Ssr4
Cma1
Retnlb
Ang4
Reg4
c
WT
day1
day2
day3
day4
Percent of Cluster
Cluster
65.5%
45.9%
67.9%
65.2%
d
Cma1
Retnlb
Ang4
UMAP_2
UMAP_1
e
Cluster 0 (of 18)
Total # of cells expressing
AMPs = 329
Retnlb
17.0%
16.7%
7.0%
17.0%
Ang4
25.5%
10.6%
Cma1
6.1%
All Clusters (18)
Total # of cells expressing
AMPs = 1,875
Retnlb
26.6%
18.2%
7.6%
12.1%
Ang4
17.9%
9.5%
Cma1
8.1%

Supplementary Figure 3.4: ***Sub-clustering of Cluster 0 reveals a secretory cell type enriched on day 4 post irradiation.***

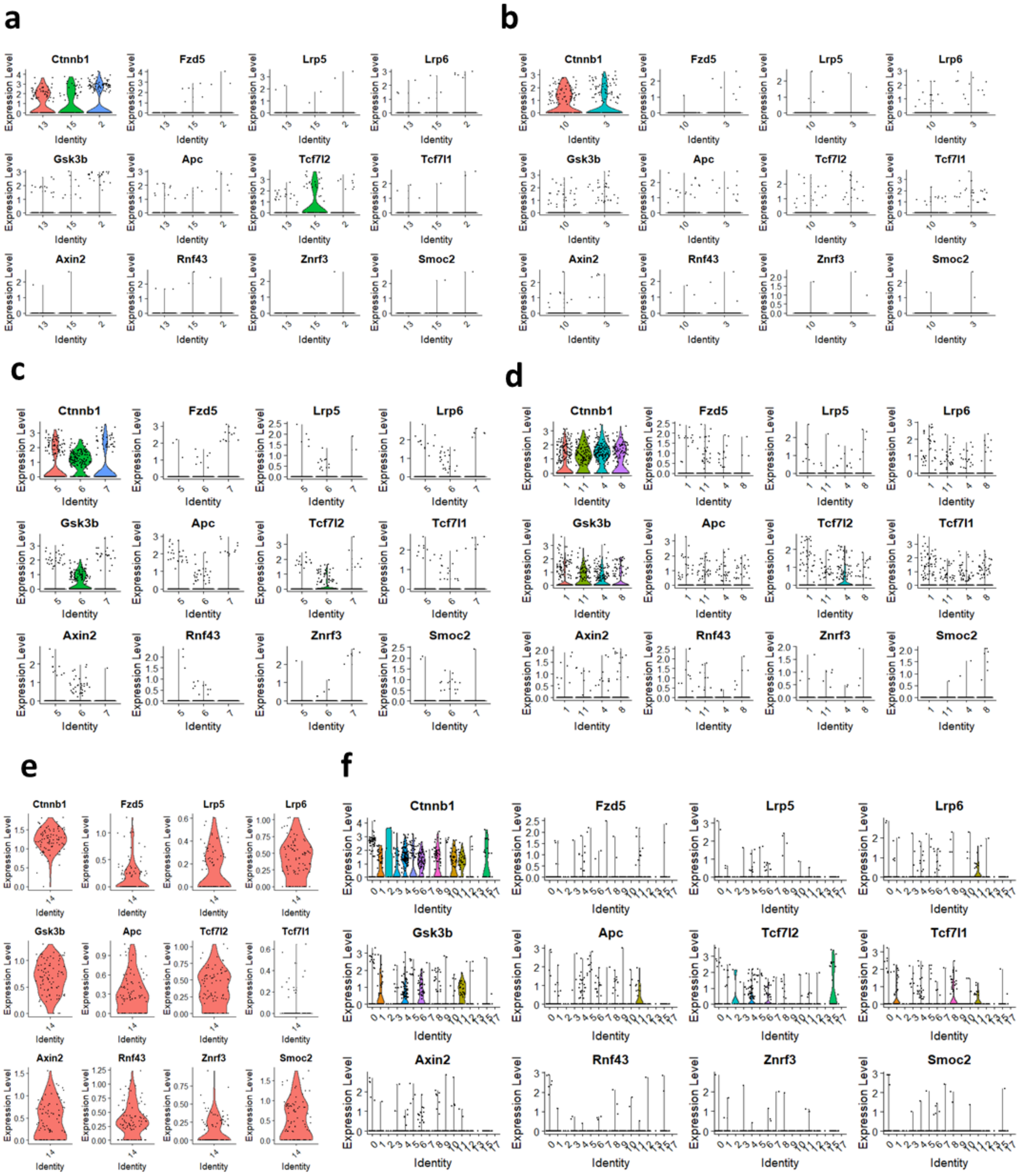
Cluster 0 was identified as a mixed population since it lacks strong specific marker gene expression compared to other clusters. In order to investigate further we performed a subcluster analysis to identify 4 subclusters as displayed in **a** the UMAP plot. **b** Violin plot display of the subcluster marker genes. Several of these subclusters are secretory genes often associated with the secretory fate – although *Muc2*, generally high in all secretory cells, was only expressed in the smallest subcluster 3. All of these subclusters expressed more than one antimicrobial peptide. **c** Graph of the relative abundance of cells from either WT or each day post irradiation shows the subclusters are predominantly made of cells from day4. **d** Feature plot showing antimicrobial peptides broadly expressed in all subclusters. **e** Venn diagram displaying the percent of cells expressing one or more antimicrobial peptide in Cluster 0 versus in all original 18 clusters. Cluster 0 represents 17.5% of all cells that express at least one antimicrobial peptide.


a
Trp53
Mdm2
Mdm4
Bbc3
Apaf1
Bax
Zmat3
Cdkn1a
Gadd45a
Gadd45b
Btg2
Ercc5
Mgmt
Polk
Sesn3
Feedback
Apoptosis
Cell Cycle Arrest
DNA Damage Repair
Metabolism
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17


b
Tnfsf10
Ligand
Tnfrsf1a
Receptor
Tnfrsf10b
Receptor
Traf2
Tradd
Fadd
Ripk1
Cradd
(Raidd)
Casp2
Casp3
Casp7
Casp8
0 1 2 3 4 5 6 7 8 10 11 12 13 14 15 16 17

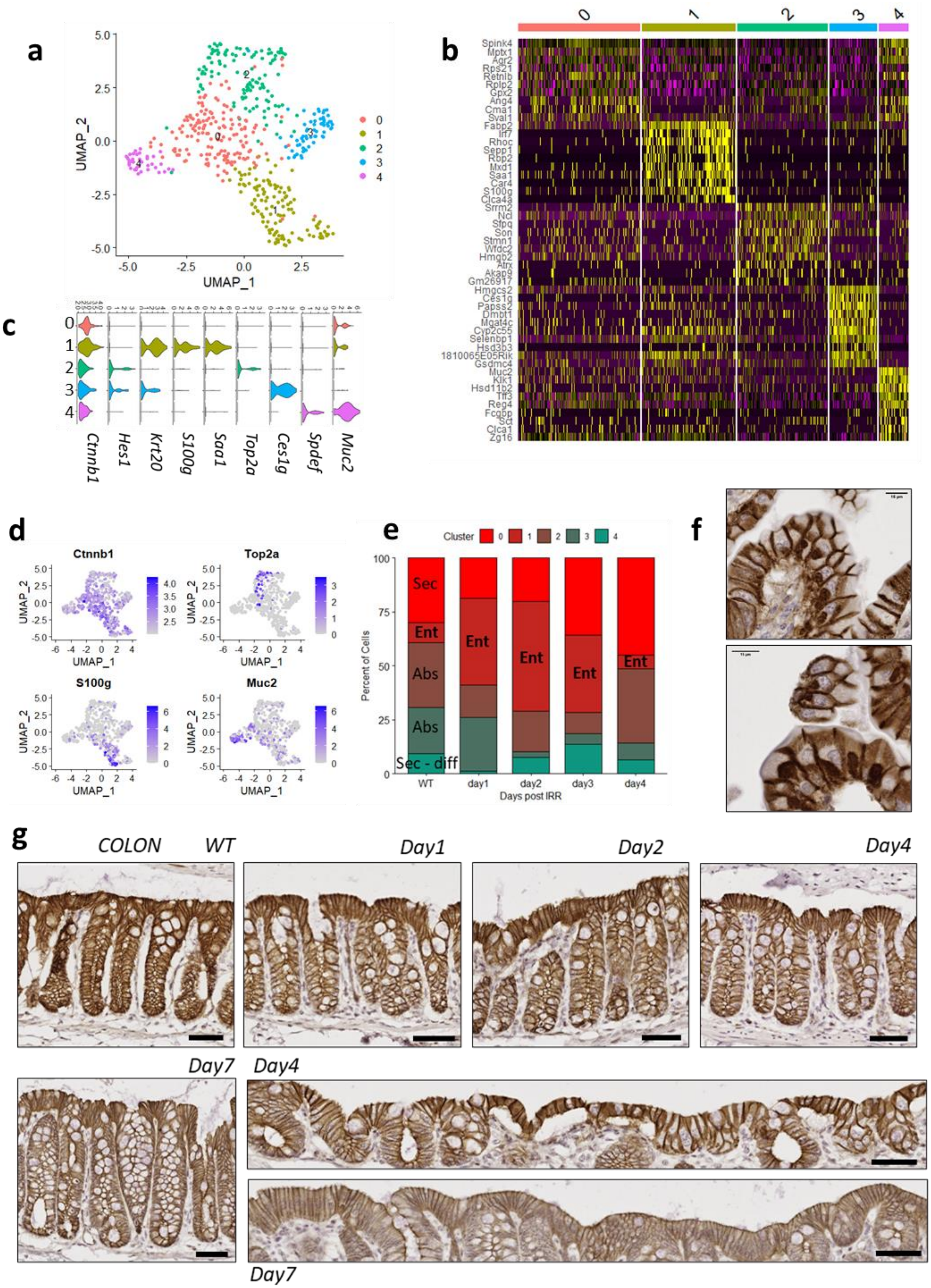
Supplementary Figure 3.5: ***Elevated mRNA expression of apoptosis pathway components in cells located at the base of crypts.***

Violin plots displaying expression of apoptosis related genes in all cells by cluster. **a** P53 and P53 target genes. **b** Death receptor/Tumor Necrosis Factor (TNF) signaling pathway components that can activate an apoptotic promoting caspase cascade.


a
b
c
d
e
f
Ctnnb1
Fzd5
Lrp5
Lrp6
Gsk3b
Apc
Tcf7l2
Tcf7l1
Axin2
Rnf43
Znrf3
Smoc2
Expression Level
Identity
13
15
2
10
3
5
6
7
1
11
4
8
14
0 1 2 3 4 5 6 7 8 9 10 11 12 13 15 17

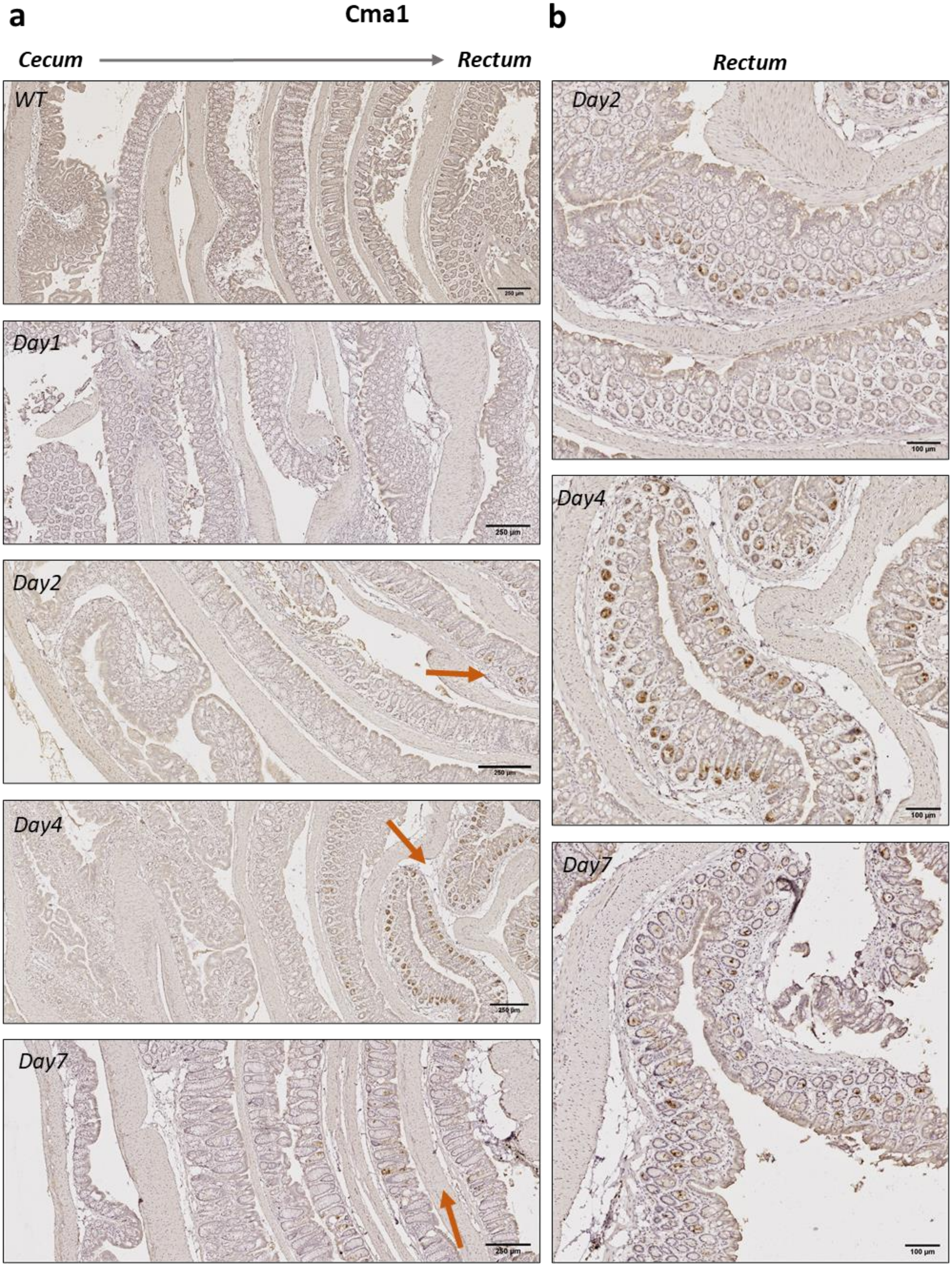
Supplementary Figure 3.6: ***Wnt signaling potential and downstream target expression are minimal in non-stem cells.***

Elevated Ctnnb1 expression, a Wnt signaling mediator, is not co-expressed with other Wnt signaling components or elevated target genes (receptors = *Fzd5*, *Lrp5, Lrp6*; Mediators = *Gskf3b, Apc*; Transcription Factors = *Tcf7l2, Tcf7l1*; and Target genes = *Axin2, Rnf43, Znrf3, Smoc2*) in cell populations other than stem cells. This suggests that there is low-to-minimal potential for response to Wnt ligands and active Wnt signaling. Violin plots display all cells (WT and all days post irradiation) from specified cluster numbers for **a** enterocytes, **b** goblet and SecPro (secretory), **c** TA cells, **d** Absorptive cells (less differentiated than enterocytes), and **e** stem cells. **f** Day 4 post irradiation cells only in all clusters.


a
b
c
d
e
f
g
UMAP_1
UMAP_2
0
1
2
3
4
Spink4
Mptx1
Agr2
Rps21
Retnlb
Rplp2
Gpx2
Ang4
Cma1
Sval1
Fabp2
Iff7
Rhoc
Sepp1
Rbp2
Mxd1
Saa1
Car4
S100g
Clca4a
Srrm2
Ncl
Sfpq
Son
Stmn1
Wfdc2
Hmgb2
Atrx
Akap9
Gm26917
Hmgcs2
Ces1g
Papss2
Dmbt1
Mgat4c
Cyp2c55
Selenbp1
Hsd3b3
1810065E05Rik
Gsdmc4
Muc2
Klk1
Hsd11b2
Tff3
Reg4
Fcgbp
Sct
Clca1
Zg16
Ctnnb1
Hes1
Krt20
S100g
Saa1
Top2a
Ces1g
Spdef
Muc2
Ctnnb1
Top2a
S100g
Muc2
Cluster
Percent of Cells
Sec
Ent
Abs
Abs
Sec - diff
Ent
Ent
Ent
Ent
WT
day1
day2
day3
day4
Days post IRR
COLON
WT
Day1
Day2
Day4
Day7
Day4
Day7

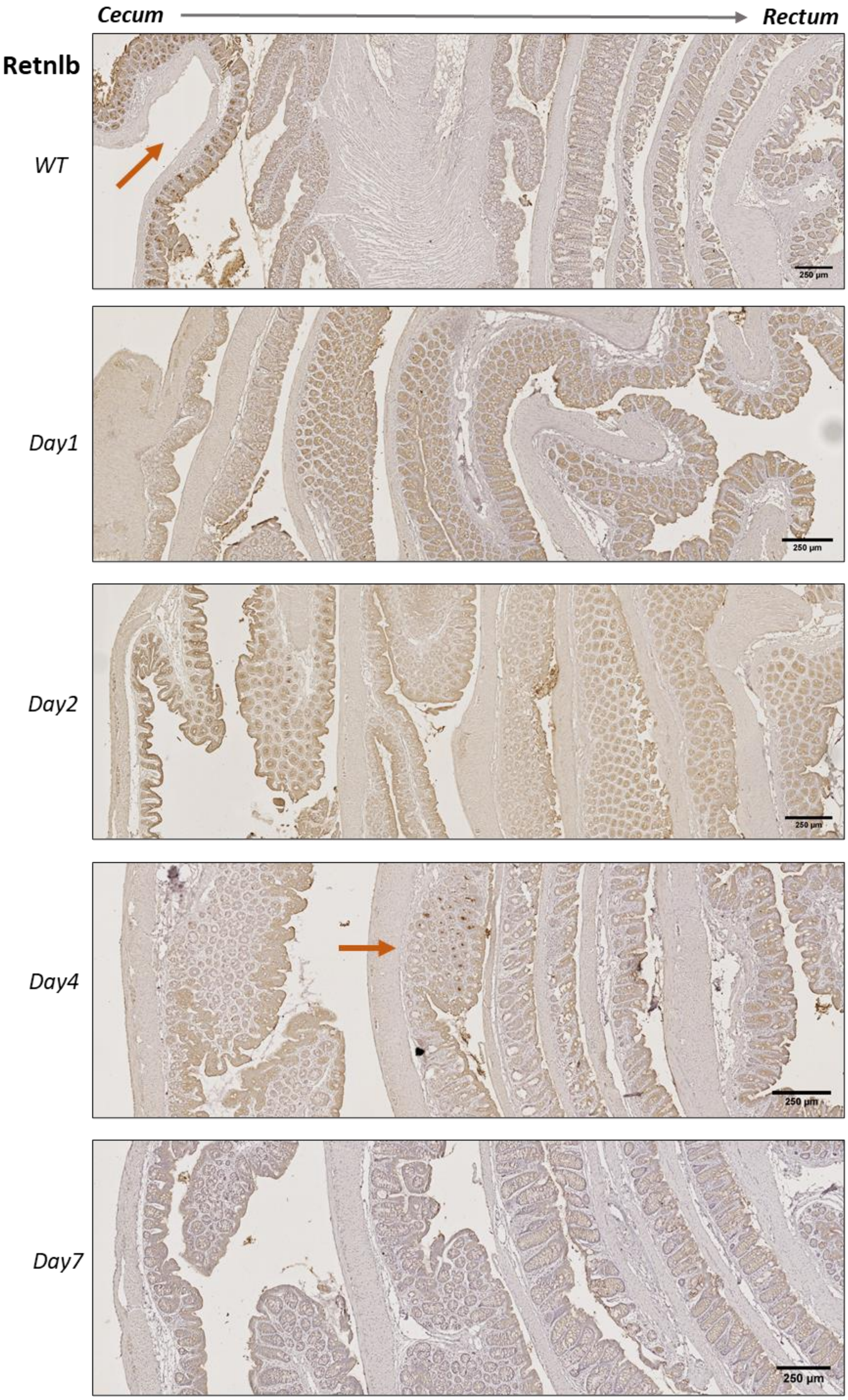
Supplementary Figure 3.7: ***β-catenin is broadly expressed in colonic crypt cells.***

Ctnnb1 expression was detected broadly in many cell clusters and its overall expression level was not strongly altered as a result of irradiation. The highest expressing *Ctnnb1* (>2) cells were subclustered. **a** UMAP showing 5 distinct population clusters of *Ctnnb1*-high expressing cells. The identities and broad lineages of these clusters were identified using **b** heatmap of cluster marker genes, and **c** violin plots of known markers. There are clear distinctions between secretory (cluster 0,4) and absorptive (cluster 1,2,3) clusters, with cluster 2 marked as proliferative, cluster 4 expressing differentiated secretory signatures, and cluster 1 consisting of differentiated absorptive cells = enterocytes. Feature plot in **d** confirms expression of key markers as well as *Ctnnb1* expression. **e** Abundance of clusters on each day post irradiation highlights relative expansion of enterocytes (Ent) on days1-3, declining on day4. **f** IHC staining on Ctnnb1 in small intestine day4 post irradiation shows what appears to be multivesicular bodies in enterocytes. **g** IHC staining of Ctnnb1 in the rectum shows strong staining throughout the crypt evident in WT and all days post irradiation. Cecal staining highlights the difference in enterocytes on the top of the crypt in day 4 compared with day7 post irradiation. Black scale bar represents 50 µm.


a
Cma1
b
Cecum
Rectum
Rectum
WT
Day1
Day2
Day4
Day7
Day2
Day4
Day7
250 µm
100 µm

Supplementary Figure 3.8: ***Cma1 expression peaks on day 4 following irradiation.***

a IHC staining of the anti-microbial Cma1 protein in colonic swiss rolls shows expression limited to the rectum in mice starting 2 days post irradiation, strongest at 4 days, and dwindling at 7 days. Staining is evident at the crypt base in day 2 and 4, but further up the crypt on day 7. No specific staining was detected in WT or 1-day post irradiation samples. Staining of two independent biological replicate mice was analyzed. **b** Enlarged images showing positive rectal regions.


Cecum
Rectum
Retnlb
WT
250 µm
Day1
250 µm
Day2
250 µm
Day4
250 µm
Day7
250 µm

Supplementary Figure 3.9: ***Retnlb1 expression increases during recovery from irradiation.***

IHC staining of Retnlb (Relmβ) in colonic swiss rolls shows expression limited to the cecum in wild type mice. No specific staining was detected 1, 2, or 7 days post irradiation. A limited region of the distal cecum was positive for staining on day 4. Staining of two independent biological replicate mice was analyzed.

REFERENCES

1. Nusse, Y. M. *et al.* Parasitic helminths induce fetal-like reversion in the intestinal stem cell niche. *Nature* **559**, 109–113 (2018).
2. Perše, M. & Cerar, A. Dextran Sodium Sulphate Colitis Mouse Model: Traps and Tricks. *J. Biomed. Biotechnol.* **2012**, 1–13 (2012).
3. De Robertis, M. *et al.* The AOM/DSS murine model for the study of colon carcinogenesis: From pathways to diagnosis and therapy studies. *J. Carcinog.* **10**, (2011).
4. Dörr, H. & Meineke, V. Acute radiation syndrome caused by accidental radiation exposure - therapeutic principles. *BMC Medicine* **9**, 126 (2011).
5. POTTEN, C. S. Extreme sensitivity of some intestinal crypt cells to X and γ irradiation. *Nature* **269**, 518–521 (1977).
6. Potten, C. & Grant, H. The relationship between ionizing radiation-induced apoptosis and stem cells in the small and large intestine. *Br. J. Cancer* **78**, 993–1003 (1998).
7. Cai, W. B., Roberts, S. A., Bowley, E., Hendry, J. H. & Potten, C. S. Differential survival of murine small and large intestinal crypts following ionizing radiation. *Int. J. Radiat. Biol.* **71**, 145–155 (1997).
8. Hornsey, S. & Vatistas, S. Some characteristics of the survival curve of crypt cells of the small intestine of the mouse deduced after whole body X irradiationE SMALL. *Br. J. Radiol.* **36**, 795–800 (1963).
9. Gândara, R. M. C., Mahida, Y. R. & Potten, C. S. Regional Differences in Stem and Transit Cell Proliferation and Apoptosis in the Terminal Ileum and Colon of Mice After 12 Gy Radiation Oncology. *Radiat. Oncol. Biol.* **82**, e521–e528 (2012).
10. Andersson-Rolf, A., Zilbauer, M., Koo, B. K. & Clevers, H. Stem cells in repair of gastrointestinal epithelia. *Physiology* **32**, 278–289 (2017).
11. Tian, H. *et al.* A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. *Nature* **478**, 255–259 (2011).
12. Potten, C. S., Owen, G. & Booth, D. Intestinal stem cells protect their genome by selective segregation of template DNA strands. *J. Cell Sci.* **115**, 2381–2388 (2002).
13. Bankaitis, E. D., Ha, A., Kuo, C. J. & Magness, S. T. Reserve Stem Cells in Intestinal Homeostasis and Injury. *Gastroenterology* **155**, 1348–1361 (2018).
14. Yan, K. S. *et al.* The intestinal stem cell markers Bmi1 and Lgr5 identify two functionally distinct populations. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 466–71 (2012).
15. Habowski, A. N. *et al.* Transcriptomic and proteomic signatures of stemness and differentiation in the colon crypt. *Commun. Biol.* **3**, 1–17 (2020).
16. Ayyaz, A. *et al.* Single-cell transcriptomes of the regenerating intestine reveal a revival stem cell. *Nature* **569**, 121–125 (2019).
17. Tetteh, P. W. *et al.* Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters In Brief. (2016). doi:10.1016/j.stem.2016.01.001

18. van Es, J. H. *et al.* Dll1+ secretory progenitor cells revert to stem cells upon crypt damage. *Nat. Cell Biol.* **14**, 1099–1104 (2012).

19. Zappia, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience* **7**, (2018).

20. Shukla, P. K. *et al.* Rapid disruption of intestinal epithelial tight junction and barrier dysfunction by ionizing radiation in mouse colon in vivo: protection by N-acetyl-L-cysteine. *Am. J. Physiol. Liver Physiol.* **310**, G705–G715 (2016).

21. Nejdfors, P., Ekelund, M., Weström, B. R., Willén, R. & Jeppsson, B. Intestinal permeability in humans is increased after radiation therapy. *Dis. Colon Rectum* **43**, 1582–1587 (2000).

22. Speckmann, B., Bidmon, H. J., Borchardt, A., Sies, H. & Steinbrenner, H. Intestinal selenoprotein P in epithelial cells and in plasma cells. *Arch. Biochem. Biophys.* **541**, 30–36 (2014).

23. Barrett, C. W., Short, S. P. & Williams, C. S. Selenoproteins and oxidative stress-induced inflammatory tumorigenesis in the gut. *Cellular and Molecular Life Sciences* **74**, 607–616 (2017).

24. Eckers, J. C., Kalen, A. L., Xiao, W., Sarsour, E. H. & Goswami, P. C. Selenoprotein P inhibits radiation-induced late reactive oxygen species accumulation and normal cell injury. *Int. J. Radiat. Oncol. Biol. Phys.* **87**, 619–625 (2013).

25. Elhodaky, M. & Diamond, A. M. Selenium-binding protein 1 in human health and disease. *International Journal of Molecular Sciences* **19**, (2018).

26. Wang, Y. *et al.* Selenium-binding protein 1 transcriptionally activates p21 expression via p53-independent mechanism and its frequent reduction associates with poor prognosis in bladder cancer. *J. Transl. Med.* **18**, 17 (2020).

27. Ning, S., Pagano, J. S. & Barber, G. N. IRF7: Activation, regulation, modification and function. *Genes and Immunity* **12**, 399–414 (2011).

28. Honda, K. *et al.* IRF-7 is the master regulator of type-I interferon-dependent immune responses. *Nature* **434**, 772–777 (2005).

29. Lan, Q. *et al.* Type I interferon/IRF7 axis instigates chemotherapy-induced immunological dormancy in breast cancer. *Oncogene* **38**, 2814–2829 (2019).

30. Johansson, M. E. V, Larsson, J. M. H. & Hansson, G. C. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc. Natl. Acad. Sci. U. S. A.* **108 Suppl 1**, 4659–65 (2011).

31. Bevins, C. L. & Salzman, N. H. Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. *Nature Reviews Microbiology* **9**, 356–368 (2011).

32. Caughey, G. H. Mast cell tryptases and chymases in inflammation and host defense. *Immunological Reviews* **217**, 141–154 (2007).

33. Gendrin, C. *et al.* Mast cell chymase decreases the severity of group B Streptococcus infections. *J. Allergy Clin. Immunol.* **142**, 120-129.e6 (2018).

34. Johnzon, C.-F., Rönnberg, E. & Pejler, G. The Role of Mast Cells in Bacterial Infection. *Am. J. Pathol.* **186**, 4–14 (2016).

35. Hooper, L. V., Stappenbeck, T. S., Hong, C. V. & Gordon, J. I. Angiogenins: A new class of microbicidal proteins involved in innate immunity. *Nat. Immunol.* **4**, 269–273 (2003).

36. Propheter, D. C., Chara, A. L., Harris, T. A., Ruhn, K. A. & Hooper, L. V. Resistin-like molecule β is a bactericidal protein that promotes spatial segregation of the microbiota and the colonic epithelium. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 11027–11033 (2017).

37. Bergstrom, K. S. B. *et al.* Goblet Cell Derived RELM-β Recruits CD4+ T Cells during Infectious Colitis to Promote Protective Intestinal Epithelial Cell Proliferation. *PLoS Pathog.* **11**, (2015).

38. Hogan, S. P. *et al.* Resistin-like molecule β regulates innate colonic function: Barrier integrity and inflammation susceptibility. *J. Allergy Clin. Immunol.* **118**, 257–268 (2006).

39. Xue, Y. *et al.* Bortezomib stabilizes and activates p53 in proliferative compartments of both normal and tumor tissues in vivo. *Cancer Res.* **79**, 3595–3607 (2019).

40. Bach, S. P., Renehan, A. G. & Potten, C. S. Stem cells: The intestinal stem cell as a paradigm. *Carcinogenesis* **21**, 469–476 (2000).

41. Merritt, A. J. *et al. The Role of p53 in Spontaneous and Radiation-induced Apoptosis in the Gastrointestinal Tract of Normal and p53-deficient Mice1*. *Cancer Research* **54**, (1994).

42. Wilson, J. W., Mark Pritchard, D., Hickman, J. A. & Potten, C. S. Radiation-Induced p53 and p21 WAF-1/CIP1 Expression in the Murine Intestinal Epithelium Apoptosis and Cell Cycle Arrest. *Am. J. Pathol.* **153**, 899–909 (1998).

43. Kreis, N. N., Louwen, F. & Yuan, J. The multifaceted p21 (Cip1/Waf1/CDKN1A) in cell differentiation, migration and cancer therapy. *Cancers (Basel).* **11**, (2019).

44. Aubrey, B. J., Kelly, G. L., Janic, A., Herold, M. J. & Strasser, A. How does p53 induce apoptosis and how does this relate to p53-mediated tumour suppression? *Cell Death and Differentiation* **25**, 104–113 (2018).

45. Fischer, M. Census and evaluation of p53 target genes. *Oncogene* **36**, 3943–3956 (2017).

46. Sheikh, M. S. *et al.* p53-dependent and -independent Regulation of the Death Receptor KILLER/DR5 Gene Expression in Response to Genotoxic Stress and Tumor Necrosis Factor α. *Cancer Res.* **58**, (1998).

47. Wajant, H., Pfizenmaier, K. & Scheurich, P. Tumor necrosis factor signaling. *Cell Death and Differentiation* **10**, 45–65 (2003).

48. Greicius, G. *et al.* PDGFRα + pericryptal stromal cells are the critical source of Wnts and RSPO3 for murine intestinal stem cells in vivo. *Proc. Natl. Acad. Sci. U. S. A.* 201713510 (2018). doi:10.1073/pnas.1713510115

49. Kabiri, Z. *et al.* Stroma provides an intestinal stem cell niche in the absence of epithelial Wnts. *Dev.* **141**, 2206–2215 (2014).

50. Dunn, S. J., Osborne, J. M., Appleton, P. L. & Näthke, I. Combined changes in Wnt signaling response and contact inhibition induce altered proliferation in radiation-treated intestinal crypts. *Mol. Biol. Cell* **27**, 1863–1874 (2016).

51. Wang, S., Karikomi, M., Maclean, A. L. & Nie, Q. Cell lineage and communication network

inference via optimization for single-cell transcriptomics. *Nucleic Acids Res.* **47**, 66 (2019).

52. Wagner, D. E. *et al.* Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science (80-.).* **360**, 981–987 (2018).

53. Maj, J. G. *et al. Microvascular Function Regulates Intestinal Crypt Response to Radiation*. *Cancer Research* **63**, (2003).

54. Paris, F. *et al.* Endothelial apoptosis as the primary lesion initiating intestinal radiation damage in mice. *Science (80-.).* **293**, 293–297 (2001).

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

56. Saha, S. *et al.* Bone Marrow Stromal Cell Transplantation Mitigates Radiation-Induced Gastrointestinal Syndrome in Mice. *PLoS One* **6**, e24072 (2011).

57. Accarie, A. *et al.* Extracellular vesicles derived from mesenchymal stromal cells mitigate intestinal toxicity in a mouse model of acute radiation syndrome. *Stem Cell Res. Ther.* **11**, 1–15 (2020).

58. Bialkowska, A. B., Ghaleb, A. M., Nandan, M. O. & Yang, V. W. Improved Swiss-rolling Technique for Intestinal Tissue Preparation for Immunohistochemical and Immunofluorescent Analyses Video Link. *J. Vis. Exp* 541613791–54161 (2016). doi:10.3791/54161

59. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **7**, 1–7 (2017).

60. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671–675 (2012).

61. A.N. Habowski, J.M. Bates, J.L. Flesher, R.A. Edwards, M. L. W. Isolation of murine large intestinal crypt cell populations with flow sorting. *Protoc. Exch.* (2020).

62. McGinnis, C. S. *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* 1 (2019). doi:10.1038/s41592-019-0433-8

63. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).

64. Zappia, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience* **7**, (2018).

CHAPTER 4

α-Ketoglutarate attenuates Wnt signaling and drives differentiation in colorectal cancer

Text adapted from published work:

Tran, T.Q., E.A. Hanse, **A.N. Habowski***, H. Li, M.I. Gabra, Y. Yang, X.H. Lowman, A.M. Ooi, S.Y. Liao, R.A. Edwards, M.L. Waterman, and M. Kong. (2020) α-Ketoglutarate attenuates Wnt signaling and drives differentiation in colorectal cancer. Nature Cancer, 1:345-358. DOI: 10.1038/s43018-020-0035-5.*

**denotes equal contribution*

ABSTRACT

Genetic-driven deregulation of the Wnt pathway is crucial but not sufficient for colorectal cancer (CRC) tumorigenesis. Here, we show that environmental glutamine restriction further augments Wnt signaling in APC mutant intestinal organoids to promote stemness and leads to adenocarcinoma formation *in vivo* via decreasing intracellular alpha-ketoglutarate (αKG) levels. αKG supplementation is sufficient to rescue low-glutamine induced stemness and Wnt hyperactivation. Mechanistically, we found that αKG promotes hypomethylation of DNA and histone H3K4me3, leading to an upregulation of differentiation-associated genes and downregulation of Wnt target genes, respectively. Using CRC patient-derived organoids and several *in vivo* CRC tumor models, we show that αKG supplementation suppresses Wnt signaling and promotes cellular differentiation, thereby significantly restricting tumor growth and extending survival. Together, our results reveal how metabolic microenvironment impacts Wnt signaling and identify αKG as a potent antineoplastic metabolite for potential differentiation therapy for CRC patients.

INTRODUCTION

Colorectal cancer (CRC) remains the second leading cause of cancer-associated deaths with more than a million people in the United States alone living with the disease [1]. Hyperactivation of the Wnt pathway due to adenomatous polyposis coli (APC) mutations occur in 80% of human CRC and is a crucial initiating step in carcinogenesis that disrupts cellular differentiation and promotes rapid proliferation [2,3]. However, genetic-driven Wnt deregulation only drives the formation of benign polyps and is insufficient to promote carcinoma [4,5]. While the accumulation of mutations in other oncogenes and tumor suppressor genes has been implicated in CRC progression, emerging evidences suggest that non-genetic factors, such as the microenvironment, can contribute to optimal Wnt activation and support oncogenic transformation in CRC [6–9].

Colon cancer cells are subjected to diverse metabolic fluctuations in gut microenvironment, yet little is known about the role of metabolism in regulating Wnt signaling and CRC tumorigenesis. It is well established that overactive Wnt signaling through the regulation of metabolic enzymes and transcription factors can directly reprogram metabolic pathways to support rapid proliferation. For example, Lef/Tct/Ctnnb1 transcription complexes, through the upregulation of pyruvate dehydrogenase kinase 1 (Pdk1) and the lactate transporter (Mct1) can divert glycolytic glucose toward the production of lactate which stimulates angiogenesis at tumor sites [10,11]. Consequently, Wnt-mediated aerobic glycolysis can render cancer cells more dependent on glutamine metabolism to meet an elevated bioenergetic demand and support the biosynthesis of macromolecules. A recent study demonstrates that human CRC tumors exhibit an increased uptake and catabolism of glutamine compared to healthy tissues consistent with this idea [12]. Furthermore, around 30% of circulating glutamine is consumed in the gut, suggesting intestinal cells utilize exogenous glutamine as a preferred fuel source to support intestinal regeneration and integrity [13]. Therefore, the increased glutamine dependency in CRC tumors, in addition to an already high glutamine demand by normal intestinal cells, may exhaust the local nutrient supply leading to intra-tumoral glutamine starvation. Glutamine catabolism in proliferative cells in part supports the biosynthesis of alpha ketoglutarate (αKG) to replenish TCA cycle intermediates [14]. Importantly, glutamine and αKG levels have been shown to play an important role in metabolic homeostasis, reactive oxygen species (ROS) control and epigenetic regulation in both normal and cancer cells [15–17]. While the Wnt pathway's effect on metabolism is well established, it remains largely unknown how changing metabolism, such as glutamine levels, modulate oncogenic signaling pathways and cancer progression.

In this study, we measure how changing glutamine and αKG levels affects oncogenic Wnt signaling and cancer cell differentiation using primary intestinal organoids and CRC mouse models. We demonstrate that while low glutamine concentrations drive Wnt hyperactivation to enhance intestinal stemness and tumorigenicity, the supplementation of a glutamine-derived metabolite, αKG, is sufficient to suppress

Wnt signaling and promotes terminal differentiation via epigenetic reprogramming. Collectively, we provide compelling evidence for the crucial role of glutamine and αKG homeostasis in regulating oncogenic Wnt signaling and CRC progression.

RESULTS

Environmental glutamine restriction hyperactivates Wnt signaling and blocks cellular differentiation

Accumulating evidence suggests colon cancer cells rapidly consume glutamine to support cell survival and proliferation [18]. In agreement with this evidence, we found a significant increase in glutamine uptake in organoids with hyperactive Wnt signaling (Figure 4.1a). To determine potential changes in glutamine levels in CRC *in vivo*, we measured glutamine concentrations in mouse intestinal tumors. We observed that glutamine concentrations were significantly lower in intestinal tumors from heterozygous *APC* mutant (*Apc*$^{Min/+}$) mice compared to healthy tissues (Figure 4.2a). This phenomenon is also observed in human tumors and may be due to increased glutamine utilization in CRC tumor [12,19]. To examine how this environmental glutamine restriction influences CRC progression, we generated *Apc*$^{Min/+}$ small intestinal organoids that recapitulate the genetic background of tumor-initiating cells. We found that low glutamine transformed *Apc*$^{Min/+}$ organoids with well-defined crypts into cystic organoids, which are less differentiated, in a dose and time dependent manner (Figure 4.2b and Figure 4.1b) [20]. We also observed a modest increase in the cystic phenotype of *Apc*$^{Min/+}$ organoids treated with a glutaminase inhibitor (Figure 4.1c). In contrast, metabolic stress had minimal effects on cellular differentiation and the expression of Wnt target genes from wild-type organoids (Figure 4.1d-g). Further examination revealed that glutamine-starved *Apc*$^{Min/+}$ organoids exhibit increased expression of the stem cell marker Lgr5 along with active β-catenin suggesting an increase in stemness (Figure 4.2c). Moreover, expression of the differentiation marker Krt20 and activity of the alkaline phosphatase enzyme Alpi were reduced, indicating decreased cellular differentiation in response to glutamine starvation (Figure 4.2d).

Using whole exome sequencing, we did not find genetic alterations in CRC driver pathways in glutamine-starved organoids, consistent with previous findings that cultured organoids are genetically and phenotypically stable (Figure 4.3a) [4]. Thus, it appears that genetic instability and other mutations do not contribute to low-glutamine induced stemness. Moreover, we detected full-length wildtype Apc protein in glutamine-starved *Apc*$^{Min/+}$ organoids (Figure 4.3b-c). Consistently, glutamine withdrawal did not reduce the percentage of cells expressing wildtype Apc based on immunofluorescence staining for the C-terminus of Apc, which specifically recognizes wildtype Apc protein (Figure 4.3d). Together, these data suggest glutamine deprivation induces stemness in heterozygous *Apc*$^{Min/+}$ cells instead of providing a selective advantage for cells that lost the wild-type Apc allele. To further confirm that the effect of low glutamine on stemness is independent of Apc mutation status, we examined the effect of low glutamine on organoids with Apc knockdown, Kras activation and p53 deletion (AKP) in which Apc expression can be controlled by a doxycycline inducible Apc shRNA. Consistent with the finding in *Apc*$^{Min/+}$ organoids, glutamine deprivation drastically delayed intestinal differentiation and maintained stemness, regardless of Apc status as observed via organoid morphology and gene expression (Figure 4.3e-g).

To better understand how low glutamine alters cellular differentiation, we performed whole transcriptome sequencing in *Apc*$^{Min/+}$ organoids. We observed a global reprogramming of gene expression in intestinal organoids in response to glutamine restriction (Figure 4.3h). Pathway analyses revealed the top upregulated pathways in glutamine-starved organoids are associated with CRC progression or metastasis including *Wnt*, *Mapk*, *Rho*, *Vegf*, *Il1* and *Il8* (Figure 4.2e) [21,22]. Interestingly, glutamine-starved organoids displayed upregulation of CRC-associated genes and hyperactivation of Wnt signaling as determined by GSEA analysis and the expression of the Wnt target gene *Axin2* (Figure 4.2f-g). In addition, glutamine restriction in SW620 colon cancer cells harboring an APC mutation induced the expression of the Wnt target genes *AXIN2* and *LGR5* in addition to Wnt ligands in a dose-dependent manner (Figure 4.3i). Furthermore, blockade of Wnt signaling with iCRT3, a small molecule inhibitor of the Wnt pathway, partially reversed cellular differentiation in glutamine-starved *Apc*$^{Min/+}$ organoids indicated by the ability to re-form crypts (Figure 4.2h). Similarly, other Wnt inhibitors, IWP2 and XAV939, also blocked the formation of the stem-like cystic organoids upon glutamine deprivation (Figure 4.2i). Taken together, these results demonstrate that a glutamine restricted environment promotes Wnt signaling and stemness in *Apc*$^{Min/+}$ organoids.

Glutamine restriction promotes enhanced self-renewal and niche independence in Apc$^{Min/+}$ organoids

It has been shown that APC mutation alone is not sufficient to promote cancer [4]. Thus, we asked whether metabolic inputs contribute to the oncogenic development in CRC. To test this, we examined the self-renew ability of organoids from single cells cultured under low glutamine conditions (Figure 4.4a) [23]. We found that glutamine-starved *Apc*$^{Min/+}$ organoids exhibited enhanced self-renewal based on the secondary organoids they produced along with increased proliferative capacity based on total cell numbers (Figure 4.4b), while no significant effect was found in Apc wildtype organoids (Figure 4.1f-g). Furthermore, glutamine restriction allowed these organoids to survive and grow independent of "stem cell niche" factors such as R-Spondin, Egf and Noggin even though there was a large reduction in secondary organoid formation in growth-factor free medium compared to fully supplemented medium (Figure 4.4c). In contrast, none of the control organoids survived in the medium lacking these factors. Together, these results suggest that exposure to low glutamine promotes stemness and niche independence in precancerous *Apc*$^{Min/+}$organoids.

To test whether pre-exposure to glutamine starvation promotes the tumorigenicity of *Apc*$^{Min/+}$ organoids *in vivo*, we injected control organoids and glutamine-starved organoids into the flanks of immunodeficient mice and let them develop (Figure 4.4d). While paired control *Apc*$^{Min/+}$ organoids failed to engraft, glutamine-starved organoids formed visible lesions with features of tubular adenoma in the subcutaneous environment where intestinal niche factors are not found (Figure 4.4e). Remarkably, the tumors, while limited in size, eventually developed into poorly differentiated adenocarcinomas based on histological analysis (Figure 4.4f). Thus, our results suggest that glutamine limitation enhances the stemness and tumorigenicity of *Apc* mutant cells.

αKG supplementation rescues low glutamine-induced stemness and suppresses Wnt signaling to restore cellular differentiation

In cancer and highly proliferative cells, glutamine metabolism contributes to several biological processes, including the TCA cycle, ROS regulation and epigenetic modifications (Figure 4.5a)[18]. Metabolomic analysis revealed that glutamine deprivation in *Apc*$^{Min/+}$ organoids leads to decreased levels of glutathione and many TCA intermediate metabolites including αKG (Figure 4.5b). Consistently, we found that αKG levels dramatically decreased in the tumors collected from *Apc*$^{Min/+}$ mice (Figure 4.5c). To further determine the mechanism by which glutamine metabolism affects organoid differentiation, we supplemented low glutamine medium with N-acetyl cysteine (NAC), an antioxidant; cell permeable succinate (DM-succinate), a TCA intermediate; and cell-permeable αKG (DM-αKG), a TCA intermediate with epigenetic modification potential (Figure 4.5d–f). We found both NAC and DM-succinate exerted limited ability to restore cellular differentiation, whereas DM-αKG alone was sufficient to inhibit the stem-like cystic morphology and partially rescued crypt formation in glutamine-starved Apc$^{Min/+}$ organoids (Figure 4.6a-b). In addition, αKG supplementation was sufficient to reverse low glutamine's effect on intestinal stemness and expression of Wnt target genes (Figure 4.5g-i). We found that αKG alone induced expression of the differentiation marker *Krt20* in Apc$^{Min/+}$ organoids, suggesting cellular differentiation can be induced in response to the metabolite's addition (Figure 4.7a). To further understand the anti-tumorigenic effect of glutamine and αKG, we compared the effect of glutamine supplementation on organoids with or without pre-exposure to glutamine starvation. We found that increased glutamine levels did not significantly affect the growth and differentiation processes of organoids that had not previously experienced low glutamine conditions. Interestingly, re-supplementation of glutamine or αKG suppressed growth and promoted cellular differentiation of organoids that have adapted to the chronic low glutamine conditions (Figure 4.7b).

We next compared the transcriptional change induced in organoids under low glutamine and DM-αKG treatment. We found that αKG supplementation antagonized Wnt signaling and promoted the expression of differentiation-related genes, while glutamine deprivation displayed an opposite effect on gene expression (Figure 4.6c & Figure 4.7c). GSEA analyses revealed that glutamine starvation and DM-αKG treatment have opposing influences on intestinal stemness and differentiation based on their respective expression profiles (Figure 4.6d). We next tested whether low glutamine and αKG supplementation regulate Wnt signaling by examining β-catenin and Lgr5 levels in organoids. Glutamine starvation increased the levels and nuclear localization of β-catenin and elevated Lgr5 expression while DM-αKG treatment displayed an opposite effect (Figure 4.6e-f). Together, these findings suggest that glutamine restriction favors Wnt signaling and stemness whereas supplementation of αKG is able to reverse this as indicated by decreased Wnt signaling and increased intestinal differentiation.

αKG supplementation promotes DNA and histone hypomethylation at genes related to intestinal differentiation and Wnt signaling

αKG is an essential co-factor for various chromatin modifying enzymes including the Ten-eleven translocation (TET) methylcytosine dioxygenase enzymes which demethylate DNA and Jumonji C (JmjC)-domain-containing histone demethylases [17]. Aberrant DNA methylation, especially hypermethylation at the promoters of tumor suppressor genes and Wnt antagonist genes, has been shown to contribute to CRC oncogenesis [15,24,25]. We found that inhibition of DNA methylation with Decitabine, a DNA hypomethylating agent, blocked *Axin2* induction in glutamine-starved organoids, indicating potential crosstalk between epigenetics and Wnt signaling mediated by intracellular αKG levels (Figure 4.8a). Consistently, low glutamine promoted DNA hypermethylation which was reversed by αKG addition as shown by 5-methyl cytosine dot blot analysis (Figure 4.8b-c). Next, we performed reduced representation bisulfite sequencing (RRBS) to analyze the genome-wide methylation profile in organoids. We found that DM-αKG treatment alone resulted in minimal DNA hypermethylation but drastic DNA hypomethylation, indicating a hyperactivity of the Tet enzymes (Figure 4.9a). A volcano plot demonstrates the most differentially methylated genes are hypomethylated while a much smaller subset of genes are hypermethylated in response to DM-αKG treatment (Figure 4.9b). Moreover, gene ontology analysis showed that αKG supplementation promotes DNA hypomethylation in genes involved in cellular differentiation, immune response and metabolism (Figure 4.9c). Strikingly, global transcriptome remodeling took place with a larger percentage of genes upregulated upon DM-αKG treatment (Figure 4.9d). The integration of the DNA methylation sequencing and the transcriptome profile in a Starburst plot identified 293 genes with DNA hypomethylation and increased gene expression upon DM-αKG treatment (Figure 4.9e & Table 4.1). Among these are differentiation-associated genes identified by overlapping with an intestinal differentiation signature and genes often hypermethylated in CRC including *Ndrg4* and *Stox2* [26]. Moreover, we found genes with tumor suppressive functions including *Bbc3* and *Bax* also affected by DM-αKG treatment (Figure 4.9f). Consistently, methylation EPIC microarray analyses revealed that αKG supplementation resulted in drastic DNA hypomethylation in human CRC cells, especially at regions of genes associated with intestinal differentiation (Figure 4.10d & Table 4.2).

We also found that αKG reshaped epigenetic marks on genes related to Wnt signaling. For example, αKG treatment induced DNA-demethylation and upregulation of *Dkk3* and *Dkk4*, Wnt antagonists that block Wnt ligand-receptor interaction, and *Fat1* which interferes with the nuclear localization and transcriptional activity of β-catenin (Figure 4.9f) [27]. Consistently, *Dkk4*, which has been shown to suppress Wnt signaling in APC mutant CRC cells, was upregulated in organoids upon αKG or DNA methylation inhibitor treatment (Figure 4.8e) [28]. Similarly, we confirmed αKG promoted global DNA demethylation, particularly on the upstream region of *DKK4* in CRC cells (Figure 4.8f). We found knocking down *Tet1* blocked *Dkk4* induction and partially inhibited the effect of αKG on the expression of stemness marker *Lg*r5 (Figure 4.8g-h), supporting that αKG-mediated DNA demethylation is in part through Tet enzymes.

In addition to DNA methylation, previous studies suggest that histone methylation is also regulated by intracellular αKG levels [15,29]. Consistently, we found that αKG supplementation drives drastic hypomethylation of histone marks in CRC cells, especially H3K4 trimethylation (H3K4me3) which is associated with active transcription (Figure 4.8i). Chromatin immunoprecipitation assays reveal that αKG supplementation led to the loss of H3K4me3 on Wnt target genes, consistent with previous studies showing that H3K4 plays a crucial role in the regulation of Wnt target genes in CRC cells (Figure 4.8j) [30,31]. In contrast, low glutamine conditions led to increased H3K4me3 levels, specifically at the loci of Wnt target genes (Figure 4.8k-l). Taken together, these results suggest that αKG promotes DNA hypomethylation to induce the expression of differentiation-related genes and H3K4 hypomethylation to suppress the expression of Wnt target genes.

αKG supplementation drives terminal differentiation and suppresses growth of patient-derived colon tumor organoids

As cellular dedifferentiation drives CRC pathogenesis and therapeutic resistance, reinforcement of terminal differentiation is a promising therapeutic approach that has gained success in other cancers that arise similarly from differentiation dysfunction [7,32,33]. Since αKG supplementation has resulted in remarkable responses in mouse organoids in terms of promoting cellular differentiation, we generated a panel of patient-derived organoids (PDOs) to test the therapeutic potential of αKG supplementation in human CRC cells (Extended data Figure 4.10a-b). We found that DM-αKG treatment inhibited the initiation and growth of a panel of PDOs (Figure 4.11a). In addition, DM-αKG treatment significantly reduced organoid size and blocked cystic morphology in PDOs (Figure 4.11b). Also, αKG-treated PDOs highly expressed intestinal differentiation markers including Krt20 and Muc2 compared to the control organoids (Figure 4.11c). Rather than inducing cytotoxicity, treatment with αKG arrested the cancer organoids in a terminally differentiated state with a limited proliferative potential that persisted even after the metabolite was washed out (Figure 4.11d & Figure 4.10c-d). Consistent with the response in mouse intestinal organoids, we found that Wnt signaling was suppressed and differentiation-related genes were partly induced upon αKG treatment in PDOs (Figure 4.11e). Drug response profiling in PDOs has been shown to accurately predict clinical responses in patients with gastrointestinal cancer [34]. Therefore, αKG supplementation with the ability to suppress Wnt signaling and induce cellular differentiation in PDOs represent a potential therapeutic opportunity for the treatment of colon cancer.

αKG supplementation inhibits the growth of highly mutated CRC xenograft tumors in vivo

In addition to Wnt dysregulation, CRC tumors commonly acquire mutations in other oncogenic pathways, including KRAS (43 % human CRC) and P53 (54 % human CRC) that contribute to cancer progression and drug resistance [35]. While the effect of αKG supplementation in *Apc*$^{Min/+}$ organoids is dramatic, the accumulation of these genetic mutations could render cancer cells insensitive to the metabolite. Thus, we tested the therapeutic potential for αKG supplementation in tumors with an activating Kras mutation, Apc loss and p53 disruption, as a model for human CRC. In poorly differentiated AKP organoids, we found that DM-αKG treatment restricted organoid growth and induced cellular differentiation as indicated by the ability of organoids to re-form crypts without restoring Apc expression (Figure 4.12a-b). Importantly, DM-αKG treatment either by IP injection or in the drinking water suppressed the growth of mouse subcutaneous xenograft tumors generated from AKP organoids (Figure 4.12c). Histological analysis revealed that DM-αKG treated tumors were predominantly occupied by stromal cells while control AKP organoids grew out as invasive carcinomas (Figure 4.12d). To further validate these results, we tested the effect of αKG on another xenograft tumor model generated from SW620 human colon cancer cells harboring mutations in the *APC, KRAS, TP53* and *SMAD4* genes. We found that DM-αKG treatment increased intra-tumoral αKG levels and limited tumor growth (Figure 4.12 e-f). Consistently, αKG treatment *in vivo* suppressed the expression of Wnt target genes in SW620 xenograft tumors (Figure 4.12g). Together, these findings suggest that αKG supplementation is effective for the treatment of CRC tumors bearing other oncogenic mutations in addition to *APC*.

αKG suppresses Wnt signaling and inhibits tumor initiation in Apc$^{Min/+}$ mice

Combating tumor initiation and preventing recurrence is a clinical challenge in CRC [36]. Thus, we evaluated the effects of αKG treatment in intestinal neoplasia using the *Apc*$^{Min/+}$ mouse, a model that develops spontaneous intestinal tumors due to Wnt deregulation [37]. DM-αKG treatment via IP injection raised αKG levels in the intestine without significantly affecting the body weight or intestinal homeostasis in healthy wildtype mice, suggesting DM-αKG dosage used here is well-tolerated (Figure 4.13a & Figure 4.14a). Strikingly, DM-αKG treatment in *Apc*$^{Min/+}$ mice partially protected against tumor-associated weight loss and significantly reduced tumor numbers (Figure 4.13b–d). To gain better insight into how αKG modulates tumor initiation, we performed whole transcriptome sequencing of the intestinal tissue of mice treated with DM-αKG. *Apc*$^{Min/+}$ mice displayed extensive hyperactivation of genes compared to the wild-type mice. Interestingly, DM-αKG treatment in the *Apc*$^{Min/+}$ mice alone suppressed a significant majority of these genes in Cluster 8 (Figure 4.13e). Gene analysis of Cluster 8 revealed gene networks involved in Wnt signaling, angiogenesis, and other oncogenic pathways suggesting that αKG treatment may suppress many cancer-associated genes to inhibit tumor initiation and progression in the *Apc*$^{Min/+}$ mice (Figure 4.13f). The suppression of Wnt signaling and stemness induced upon DM-αKG treatment was confirmed via GSEA analysis, IHC staining of β-catenin, and gene expression of stem cell markers (Figure 4.13g–i). To further address whether αKG-reduced oncogenic gene expression occurs in both adenomas and non-adenomatous crypts, we performed IHC staining for Cyclin D1, in which the expression is reduced upon DM-αKG treatment based on the RNA sequencing profile. We found that DM-αKG treatment *in vivo* led to reduced Cyclin D1 expression in both tumor regions and non-adenomatous crypts (Figure 4.14b). In addition, αKG treatment suppressed the expression of Wnt target genes and partially rescued the expression of differentiation-related genes (Figure 4.14c). In order to achieve a more direct and continuous delivery of αKG, we tested the therapeutic potential of DM-αKG supplementation in drinking water. We found that supplementation of DM-αKG in the drinking water had no observed effect on the general health of the mice, including body weight, liver function and kidney function (Figure 4.14d-e). Strikingly, we found that over 90% of $Apc^{Min/+}$ control mice developed rectal bleeding at day 50, an indication of intestinal tumors, while only 23% of the mice with αKG supplementation in the drinking water had rectal bleeding (Figure 4.13j). Importantly, Kaplan-Meier survival curves revealed a significant extension in the survival of $Apc^{Min/+}$ mice with DM-αKG supplementation (Figure 4.13k). Together, our data demonstrate that αKG represses Wnt signaling restricting tumor initiation and improving survival.

DISCUSSION

Despite enormous progress in understanding the molecular carcinogenesis of colon cancer, the crosstalk between environmental factors and colon cancer development including Wnt signaling hyperactivation remains unclear. Here, we demonstrate the glutamine-αKG axis contributes to Wnt signaling and cellular differentiation in colon cancer. These findings shed light on the role of the metabolic environment in tumor progression and provide a potential therapeutic opportunity using metabolites for CRC treatment.

We found glutamine concentration is dramatically depleted in mouse intestinal tumors compared to healthy tissues, consistent with previous studies using metabolomic analysis comparing glutamine levels in colon patient samples with healthy tissues [19]. Additionally, intra-tumoral glutamine deficiency is also observed in other solid tumors including hepatomas, melanomas, pancreatic carcinomas and sarcomas [38,39]. In the case of CRC, Wnt-driven metabolic reprogramming may contribute to rapid glutamine utilization, which might eventually deplete the local supply leading to a period of glutamine deprivation observed in tumors [10,40]. Glutamine is a vital nutrient that supports the survival of cancer cells, making glutamine metabolism an attractive target for cancer therapy [18]. While complete withdrawal of glutamine may inhibit cancer cell growth, our studies suggest that tumor cells residing in chronic low glutamine conditions may become well adapted to this stressful metabolic environment [41–44]. While our finding raises a potential paradoxical role for glutamine in tumor growth, it is important to consider that tumor development is a highly dynamic process. For example, oxygen is required for cancer cell proliferation, yet, it has been well established that hypoxia induces cancer stemness and drug resistance [45,46]. Interestingly, hyperbaric oxygen treatment can exert some antitumor effects on xenograft tumors [47–49]. Similarly, our studies suggest that acute increases in glutamine/αKG levels could be detrimental to tumor cells that have been exposed to low glutamine and adapted to these conditions.

Our data indicate that chronic glutamine restriction further activates Wnt signaling to promote cancer dedifferentiation in cells with *Apc* mutation. This finding is consistent with emerging evidence suggesting that other environmental factors contribute to the optimal activation of Wnt signaling in addition to genetic alterations [9,50–54]. Nearly 50% of people will develop intestinal polyps at some point in their lifetime, yet only a small fraction of these benign lesions eventually develop into invasive tumors [55]. It has been shown that multiple genetic mutations including APC, KRAS, P53 or SMAD are implicated in this process, yet whether non-genetic factors also contribute to the oncogenic transformation of CRC is less understood [56,57]. We found glutamine restriction endows *Apc* mutant intestinal organoids with the ability to grow independent of intestinal niche factors and form adenocarcinoma tumors *in vivo* via epigenetic reprogramming. Moreover, our data suggest that the effect of low glutamine on Wnt and intestinal stemness are more profound in cells with predisposed genetic alterations compared to those from wild type cells. Thus, an altered metabolic microenvironment such as glutamine deficiency, through potentiating Wnt signaling and dedifferentiation, may ultimately contribute to tumor progression rather than initiation in CRC. We also found that low glutamine augments Wnt signaling and stemness through the depletion of intracellular αKG. Importantly, increasing intracellular αKG levels antagonized oncogenic Wnt signaling and facilitated terminal differentiation via epigenetic reprogramming in intestinal organoids and PDOs. It was previously shown that αKG induces de-differentiation and stemness by installing a hypomethylation state in embryonic stem cells [15]. However, other studies also have reported the effect of αKG in driving differentiation in other cell types including iPS cells and melanoma cells [16,29]. Thus, the metabolic regulation of cellular fate may also depend on the tissue of origin.

While the molecular pathogenesis of colorectal cancer is well characterized, clinical efforts to inhibit the Wnt pathway remain unsuccessful [58,59]. As such, treatment regimens for advanced CRC still depend heavily on chemotherapies. Many Wnt signaling inhibitors exert detrimental effects on normal intestinal homeostasis and other tissues where physiological Wnt plays a crucial role in somatic stem cell maintenance [32]. Thus, the ability to safely modulate Wnt signaling to restore terminal differentiation in colon cancer cells presents a promising therapeutic approach for colon cancer. Indeed, cancer therapies that promote terminal cellular differentiation yield impressive clinical outcomes. For example, the all-trans retinoic acid in combination with chemotherapy drives terminal differentiation resulting in a cure rate of more than 80% in patients with promyelocytic leukemia [33]. Similarly, our results indicate that αKG could be used as a therapeutic agent to suppress Wnt signaling and drive CRC differentiation. This approach could represent a less aggressive therapy for the treatment of colon cancer. In conclusion, our work provides a potential therapeutic direction to harness the potency of αKG in driving intestinal differentiation to improve the clinical outcomes of patients with CRC.

METHODS

Mouse intestinal crypt isolation and organoid culture

Intestinal crypts from 6–8 week-old *Apc*$^{Min/+}$ mice and *Apc*^{WT} mice (Jackson Laboratory) were isolated from tumor-free small intestinal tissues based on a previous study [60]. No intestinal tumors were detected in the young *Apc*$^{Min/+}$ mice that were used for organoid generation. The isolated crypts were cultured in IntestiCult™ organoid media containing EGF, Noggin and R-Spondin (Stemcell, 06005) supplemented with penicillin/streptomycin and Primocin and mixed at 1:1 ratio with Growth Factor Reduced Matrigel (Corning, 356230). 50 µL of the organoid mix was plated into a pre-warmed 48-well plate followed by 10-minute incubation at 37°C to polymerize the matrigel. 200 µL of room temperature intestinal organoids medium was added on the side of the well to cover the Matrigel. For organoid maintenance, the medium was changed every other day, and the organoids were passaged 1:5 every week. To passage the organoids, the culture medium was removed and the matrigel was dissociated in EDTA-PBS and incubated briefly in EDTA-PBS at room temperature for 10 minutes on a rocking platform. For subcloning assays, organoids were dissociated into single cells using TrypLE (ThermoFisher, 12605036). For glutamine starvation experiments, DMEM/F-12 medium- No glutamine (Gibco, 21331020) was used as a basal media and L-glutamine (Corning, 25–005) was added back to the desired concentration. Small intestinal shApc/Kras/p53$^{-/-}$ organoids were cultured with 500 ng/ml doxycycline, and kindly provided by the Lukas Dow laboratory (Weill Cornell Medicine).

Patient-derived colon isolation and organoid culture

Colonic human tissue was collected following surgical resection with informed consent and IRB approval at St. Joseph Hospital Orange (Orange, CA). All patients were diagnosed with colorectal cancer and diagnosis was confirmed by a pathologist. Detailed diagnosis of participants were included in Figure 4.10a and the tissues were obtained from the following donors: T1 from a 74 year-old male donor, T9 from a 48 year-old female donor, T13 from a 69 year-old female donor, T23 from a 83 year old male donor, and T27 from a 72 year-old male donor. Establishment and culture of organoids was performed based on a previous study [61]. Tumor Organoids were grown in basement membrane extract (Cultrex PathClear BME Type 2) with medium consisting of: 50 % advanced DMEM/F12 (supplemented with penicillin/streptomycin, 10mM HEPES, Glutamax, and Primocin), 20% R-Spondin conditioned medium (from Cultrex Rspo1-expressing cells, Trevigen), 10% Noggin conditioned medium (from HEK293 cells stably transfected with pcDNA3 NEO mouse Noggin insert, kindly provided by Dr. Hans Clevers Laboratory), 20% Expansion medium 5x (final concentrations in medium of the following 1x B27, 1.25 mM n-Acetyl Cysteine, 10 mM Nicotinamide, 50 ng/ml human EGF, 10 nM Gastrin, 500 nM A83–01, 3 μM SB202190, 10 nM Prostaglandin E2). The VENTANA MMR IHC Panel was used as a qualitative immunohistochemistry (IHC) test for light microscopic assessment of mismatch repair (MMR) proteins (MLH1, PMS2, MSH2, and MSH6) and BRAF V600E proteins in formalin-fixed, paraffin-embedded colorectal cancer (CRC) tissue sections. The OptiView DAB IHC Detection Kit was used with MLH1, MSH2, MSH6 and BRAF V600E, and the OptiView DAB IHC Detection Kit with OptiView Amplification Kit was used for PMS2 detection. Methodology for BRAF, KRAS, NRAS mutation analysis was as follows: DNA was isolated from cells or microdissection-enriched FFPE tissue, mutation evaluation by high-sensitivity Sanger sequencing bidirectional method from Neogenomics (Aliso Viejo, CA).

Cell culture reagents and siRNA transfection

SW620 cells (ATCC) were cultured in Dulbecco's Modified Eagle Medium (DMEM, Corning) supplemented with 10% fetal bovine serum (FBS; Gemini Bio-Products) with penicillin/ streptomycin (Gemini Bio-Products) at 37 °C with 5 % CO2. For glutamine starvation experiments, DMEM-No glutamine (Corning) and 10 % dialyzed FBS (Gemini Bio-Products) were used to make glutamine-free media. For cell proliferation assays, cells were cultured in a 96-well plate for treatment. Relative cell number was determined by CellTiter-Glo assay (Promega) according to the manufacturer's protocol. For siRNA transfection, ON-TARGET plus Human TET1 siRNA (Dharmacon, L-014635-03-0005) or control siRNA (Dharmacon) was used in the presence of RNAi Max lipofectamine reagent (Invitrogen). DM-αKG treatment was performed two days after transfection. Reagents: Dimethyl α-ketoglutarate (Sigma, 349631), N-Acetyl-L-cysteine/NAC (Sigma, A7250), Dimethyl-succinate (Sigma), iCRT3 (Millipore, 219332), IWP2 (Stemgent 04–0034), XAV939 (Sigma, X3004), CB-839 (Selleckchem, 7655).

Metabolite, ALT, and AST measurement assay

For glutamine measurement from tissues or tumors, 20–40 mg of frozen/fresh tissue was homogenized in ice-cold 70% ethanol by Precellys 24 homogenizer, and the supernatant was collected and dried using a SpeedVac Vacuum Concentrator. The pellet was collected and suspended in water (1 µL water per mg of fresh tissue). The concentration of glutamine in the solution was then determined by the EnzyChrom Glutamine Assay Kit (BioAssay Systems) according to the manufacturer's protocol. For αKG measurement, Alpha Ketoglutarate Assay Kit (Abcam) was used according to the manufacturer's protocol for colorimetric assay without the deproteinization step. Intracellular succinate was measured using Succinate assay kit (Abcam) and ROS levels were detected using Cellular ROS Assay Red Kit (Abcam, 186027). For ALT and AST measurement in mouse serum, EnzyChrom™ Alanine Transaminase Assay Kit and EnzyChromTM Aspartate Transaminase Assay Kit from Bioassays were used according to the manufacturer's protocol.

Metabolomics

Metabolite extraction was performed as previous described [62]. Briefly, organoids after treatment were washed with NaCl buffer and 1 ml of 80% methanol/water (HPLC grade) was added to the cell pellet on dry ice. The pellet was then transferred to −80°C freezer for 15 min to further inactivate enzymes. The metabolite extract was dried by speed vacuum. The samples were prepared and analyzed by Liquid chromatography–mass spectrometry (LC-MS) at Duke University (Durham, NC) and described previously [63].

Immunofluorescence and Immunohistochemistry

For C-terminus APC staining, organoids were digested into single cells and cultured in glass chamber with organoid medium supplemented with 10% FBS overnight. Organoids after treatment were fixed with 4% formaldehyde for 10 minutes at room temperature followed by blocking with 1% BSA at 4°C overnight. The organoids were stained with primary antibodies against KRT20 (Cell Signaling, 13063), Non-phospho

(Active) β-catenin (Cell Signaling, 19807), Muc2 (Abcam, 11197), C-terminus APC, clone 28.9 (Millipore, 3786) or DAPI (ThermoFisher, D1306) at 4°C overnight. Secondary antibodies, goat anti-rabbit Alexa Fluor 488 (Invitrogen, 11037) and goat anti-mouse Alexa Fluor 594 (Invitrogen, A11029), were purchased from Millipore-Sigma. Alkaline Phosphatase activity was determined by Red Alkaline Phosphatase Substrate Kit (Vector, SK5100). Images were captured with 20x magnification by using a Zeiss LSM 700 Confocal Microscope and the ZEN Blue image acquisition software. IHC staining were performed on sections prepared from formalin fixed and paraffin-embedded tissue by San Diego Pathology Group (San Diego, CA) and UCI experimental tissue resource pathology core.

RNA extraction, quantitative real-time PCR

Total RNA was extracted and purified using Trizol (Invitrogen) or RNeasy kit (Qiagen) according to the manufacturer's protocol. qScript cDNA synthesis kit (Quanta Biosciences) was used to make cDNA. Quantitative real-time PCR were performed using SYBR-Green PCR master mix (Quanta Biosciences) and a BioRad real-time PCR machine. Relative gene expression was normalized to rRNA ribosomal 18S or Actin. The primers used in the study are listed in Table 4.3.

RNA and DNA sequencing

RNA sequencing: Kapa RNA mRNA HyperPrep kit (Kapa Biosystems, Cat KR1352) was used for RNA sequencing library preparation. 100 ng of total RNA from each sample was used for polyA RNA enrichment using magnetic oligo-dT beads. cBot cluster generation system (Illumina) with HiSeq SR Cluster V4 Kit was used to prepare library templates for sequencing. Sequencing run was performed using Illumina HiSeq 2500 with HiSeq SBS V4 Kits. Real time analysis 2.2.38 (RTA) software was used to process the image analysis and base calling. RNA-seq sequences are aligned to mouse genome (mm10) using HISAT2 and RNA-seq expression level were measured as RPKM using Partek Genomic Suite software (v6.6). Different expressed genes were detected using Partek GS's RNA-seq pipeline. The significantly different expressed genes were detected using FDR adjusted p-value <0.05 and +/−1.5-fold change as cutoff.

GSEA analysis was based on the following gene sets: Wnt signaling: KEGG_WNT_SIGNALING_PATHWAY (M19428), Intestinal stem cell signature and intestinal differentiation signature from a previous study [64].

RRBS-seq sequencing: The RRBS-seq sequencing libraries were prepared with City of Hope IGC modified protocol. Briefly, 250 ng of genomic DNA was digested with the methylation insensitive restriction enzyme MspI (NEB) at 37°C overnight, and followed by incubation at 80 °C for 20 minutes to inactivate MspI. In the same reaction tube, Klenow Fragment (3'→5' exo-minus; NEB) and dNTP was added and incubated at 37°C for 40 minutes to repair MspI digested DNA and add dAMP to the 3' end. Illumina methylated adapters and T4 DNA ligase (Promega) were added in the same tube for an overnight ligation. 1.6xAmpureXp beads purification was used for the cleanup after the ligation. EZ DNA Methylation-Gold kit (Zymo Resarch, Cat. D5005) was used for the DNA bisulfite conversion. Bisulfite converted DNA was amplified with a 12 cycle of PCR with Pfu Turbo Cx Hotstart DNA Polymerase (Agilent). The final PCR products were run on 6% TBE acrylamide gels and DNA fragments with size from 160–500 bp (insert size 40–380 bp) were excised and eluted in EB buffer. RRBS-seq library templates were prepared for sequencing with Illumina HiSeq PE Cluster V4 Kit, sequencing runs were performed in the paired end mode of 101cycle on Illumina HiSeq 2500 with HiSeq SBS V4 Kits. Real-time analysis (RTA) 2.2.38 software was used to process the image analysis and base calling. RRBS-seq sequences were aligned to mouse genome using novocraft's Novoalign (V3.08.02). The bisulphite reference sequence was generated based on mm10 genome sequence using Novocraft's novoindex with bisulphite mode. The alignment results were used to detect CpG Methylation site using Novocraft's novomyethyl tool. Each sample type has three biological replications and the results of each biology replication were merged to generate the total methylation sites list. Most of the CpG methylation sites are close to gene region (+/−20 Kbps) and over 80% of the methylation sites are located close to gene regions. Different methylation pattern between samples are detected using ANOVA with FDR adjusted p-value <0.05 and +/−1.5-fold change on methylation percentage. To identify the methylation status at the gene level, the average of methylation difference of significant (FDR adjusted p-value) differential methylation sites was calculated.

DNA methylation profiling microarray: DNA was extracted from SW620 cells after DM-αKG treatment by DNeasy Blood & Tissue Kit. DNA methylation profiling was performed using Infinium MethylationEPIC BeadChip Kit (Illumina) according to manufacturer's protocol at The UCSD IGM Genomics Center.

Exome sequencing: 250 ng genomic DNA was fragmented using Covaris S220 (Covaris, Woburn, MA) with the 200 bp peak setting. The fragmented DNA was end-repaired and ligated to Illumina adaptor oligonucleotides with Kapa Hyper Prep Kit (KAPA Biosystems, Wilmington, MA; Cat.KK8504). Ligation products were purified and amplified with a 7 cycle of PCR. The enriched PCR products were subject to the exome capture procedure using the SureSelecXT mouse All Exon kit (Agilent, Technologies, Santa Clara, CA; Cat 5190–4641) according to manufacturer's protocols. The captured products were further amplified with an 8-cycle of PCR. Exome-seq library templates were prepared for sequencing with Illumina HiSeq PE Cluster V4 Kit as described in RRBS sequencing. Whole exome capturing and sequencing on target regions designed by Agilent SureSelect Mouse All Exon V1 were performed. The circular binary segmentation (CBS) algorithm was used to identify abnormal copy number change between samples. Read depth of coverage in each exon region was calculated, and log2-based coverage difference was further adjusted by subtracting the mean of log2-based coverage difference between y. The outcome was smoothed and segmented to identify potential copy number change between the samples. Among the potential copy number change, only the segments that span across at least 2 markers and are at least log2 (1.5) distance away from the mean copy number level was retained.

Immunoblotting

Following treatment, organoids and cells were harvested in ice cold PBS, lysed in RIPA buffer containing protease and phosphatase inhibitor (Thermo Scientific) and followed by brief sonication. Equal amount of protein, as measured by BCA protein assay, were loaded into precast NuPAGE Bis-Tris gels (Life Technologies) followed by transfer onto nitrocellulose or PVDF membrane. After blocking with 5% milk-PBS, membranes were incubated with primary antibody overnight at 4°C with shaking followed by horseradish-peroxidase-conjugated secondary antibodies for 1 hour at room temperature. The signal was visualized by Western Lightning Plus-ECL (PerkinElmer). The antibodies used are LGR5 (Abcam, 75732), Non-phospho/ Active β-Catenin (Cell Signaling, 19807), Histone H3 (Cell Signaling, 4499), Apc (Millipore, MABC202), H3K4me3 (Cell Signaling, 9751), H3K27me3 (Cell Signaling, 9733) H3K9me3 (Cell Signaling, 13969), H3K36me3 (Cell Signaling, 4909).

For Dot blot assay, DNeasy Blood & Tissue Kit (Qiagen) was used to isolate genomic DNA. After incubation at 95 °C for 5 minutes, equal amount of DNA was loaded onto a positively charged nylon membrane (Amersham Hybond-N+; GE Healthcare). The membrane was UV cross-linked and blocked in 5 % milk-PBS followed by incubation with anti-5 methyl-cytosine (Millipore, MABE146) overnight at 4 °C. After the secondary antibody incubation for 1 hour at room temperature, the signal was visualized using Western Lightning Plus-ECL (Perkin-Elmer).

MeDIP and ChIP

For methylated DNA immunoprecipitation (MeDIP), genomic DNA following DM-αKG treatment was isolated using DNeasy Blood & Tissue Kit (Qiagen) and sonicated to yield 300–1000 bp DNA fragments. The sonicated DNA was immunoprecipitated with a monoclonal antibody against 5-methylcytosine (Millipore, MABE146) for 3 hours or IgG at 4 °C with shaking. DNA complexes was pulled down using

Protein A Agarose and purified with PCR purification kit. The DNA was then analyzed by RT-qPCR with the MeDIP-hDKK4 primers listed in Table 4.3.

The chromatin immunoprecipitation assay (ChIP) was performed using a ChIP assay kit (Millipore) according to the manufacturer's guideline. Cells after treatment were cross-linked with 1% formaldehyde followed by sonication to yield 200–1000bp DNA fragments. 1 µg of H3K4me3 antibody or IgG was used for each overnight immunoprecipitation with rotation at 4°C. The DNA was then analyzed by RT-qPCR with the primer listed in Table 4.3.

Animal studies

All animal procedures and experiments were performed in compliance with ethical guidelines approved by the Institutional Animal Care and Use Committee at the University of California, Irvine, under the protocol number AUP-17–222. Sample size was generally chosen based on preliminary data indicating the variance within each group and the differences between groups. No statistical method was used to predetermine sample size.

Xenotransplantation of organoids in mice: APC mutant organoids cultured in low glutamine (0.3 mM) condition and paired control organoids with similar passage were cultured for one week. Organoids were expanded in organoid culture DMEM/F12 medium containing (Recombinant Murine EGF 50 ng/ml (Peprotech), Recombinant murine Wnt-3A 100 ng/ml (Millipore GF-160), Recombinant Murine Noggin 50 ng/ml (R&D system), 20% R-Spondin conditioned medium (from Rspo1-expressing cells, Trevigen), penicillin, and streptomycin. Organoids were dissociated into single cell with TrypLE (Thermofisher), and 1 million cells, suspended in cold Matrigel and DMEM/F12 (1:1 ratio), were injected subcutaneously into the flanks of 10-week-old male NSG (NOD scid gamma) mice bred at UCI animal facility (n=10 mice). Control organoids were injected into the left flank and low glutamine organoids were injected into the right flank. The established tumors were collected and fixed for subsequent histological analysis. To generate xenograft tumors with AKP organoids, 7-week-old athymic nude male mice (NCr-Foxn1nu, Taconic Laboratories, Rensselaer, NY) were injected subcutaneously into both flanks with 0.5 × 10^6 cells of AKP organoids in DMEM/F12 media and Matrigel (1:1 ratio), n=12 mice. The mice were treated with doxycycline in drinking water at 0.5 mg/ml with sucrose to turn on *Apc* shRNA. After the tumors were established (around 100 mm^3), the mice were randomized into groups treated with PBS control, 600 mg/kg DM-αKG by intraperitoneal injection, or 25 mg/ml DM-αKG in drinking water. SW620 xenograft tumors were generated with 1 × 10^6 cells in 8-week-old athymic nude male mice, n=12 mice. After tumor engraftment, the mice were randomized into groups treated with control PBS or with 600 mg/kg DM-αKG by intraperitoneal injection four times per week. SW620 xenograft tumors were generated with 1 × 10^6 cells in 8-week-old athymic nude male mice. After tumor engraftment, the mice were randomized into groups treated with PBS control or DM-αKG by intraperitoneal injection.

DM-αKG treatment in *Apc*$^{Min/+}$ mice: 5–8 week-old C57BL/6J-ApcMin/J mice were obtained from Jackson laboratory (MIN-002020). *Apc*$^{Min/+}$ mice were bred by crossing the Male *Apc*$^{Min/+}$ mice with female wild-type C57BL/6J mice. APC mutant genotyping was determined by Transnetyx Inc. 8–9 week *Apc*$^{Min/+}$ mice were treated with 400 mg/kg DM-αKG by intraperitoneal injection (n=16 male mice) or in drinking water supplemented with 15 mg/ml of DM-αKG (n=17 male mice and 19 female mice) (Sigma-349631 or TCI Chemical-K0013). For DM-αKG in drinking water experiment, DM-αKG supplemented water was replaced every other day for 8–9 week *Apc*$^{Min/+}$ mice. Survival endpoint was determined by established parameters including 25 % of body weight loss, pale extremities, hunching, poor body condition/dehydration, lethargy and/or irreversible rectal relapse. For 18F-FDG PET imaging of *Apc*$^{Min/+}$, mice were fasted overnight and 100µCi FDG was i.v. injected and PET images were taken 1-hour post injection.

Statistical analysis and reproducibility

Graphic representation and statistical analysis such as unpaired Student's t-test was calculated by Graphpad Prism 7 or Excel and used to determine the statistical significance of differences between means (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Results are shown as averages; error bars represent standard deviation (SD) or standard error of the mean (SEM) if indicated. p-value for mice survival study was determined by Log-rank (Mantel-Cox) test in Graphpad. No statistical method was used to predetermine sample size. For the animal study, mice were randomly assigned to the groups before treatments and the investigators were blinded to allocation of mice into different treatment groups. For the survival study, two female *APC*$^{Min/+}$ mice that developed large mammary tumors were euthanized and were excluded from the survival study based on pre-established exclusion criteria. The investigators were not blinded to outcome assessment. For tumor analysis, at least three or more independent tumors from each group were used for the analysis. Experimental reports in the study were reliably reproduced in at least two independent experiments or by multiple biologically independent replicates.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

DNA and RNA sequencing data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE140270. Super Series GSE140270 is composed of the following SubSeries: GSE140263, GSE140264, GSE140265, GSE140266, GSE140267, GSE140269.

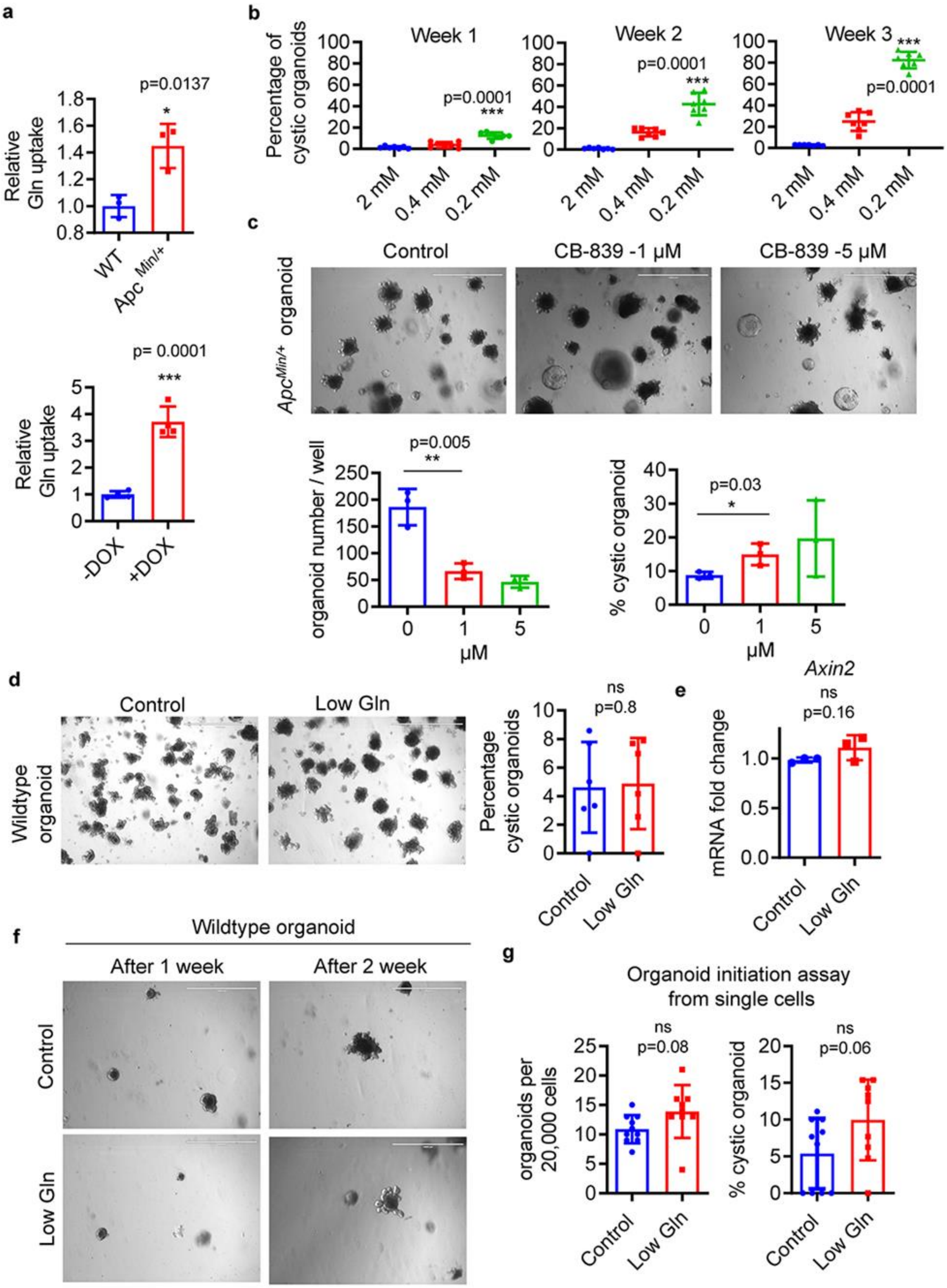

a
Relative Gln uptake
p=0.0137
*
1.8
1.6
1.4
1.2
1.0
0.8
WT
Apc Min/+
Relative Gln uptake
p= 0.0001
***
5
4
3
2
1
0
-DOX
+DOX
b
Percentage of cystic organoids
Week 1
p=0.0001
***
Week 2
p=0.0001
***
Week 3
***
p=0.0001
100
80
60
40
20
0
2 mM
0.4 mM
0.2 mM
c
Control
CB-839 -1 μM
CB-839 -5 μM
Apc^Min/+ organoid
organoid number / well
p=0.005
**
250
200
150
100
50
0
0
1
5
μM
% cystic organoid
p=0.03
*
40
30
20
10
0
0
1
5
μM
d
Control
Low Gln
Wildtype organoid
Percentage cystic organoids
ns
p=0.8
10
8
6
4
2
0
Control
Low Gln
e
Axin2
mRNA fold change
ns
p=0.16
1.0
0.5
0.0
Control
Low Gln
f
Wildtype organoid
After 1 week
After 2 week
Control
Low Gln
g
Organoid initiation assay from single cells
organoids per 20,000 cells
ns
p=0.08
25
20
15
10
5
0
Control
Low Gln
% cystic organoid
ns
p=0.06
20
15
10
5
0
Control
Low Gln

Figure 4.1: ***The effect of glutamine starvation on Apc$^{Min/+}$ and wildtype organoids.***

a Relative glutamine uptake in *Apc*$^{Min/+}$ organoids and wildtype organoids (n=3 biologically independent samples) and relative glutamine uptake in AKP organoids upon doxycycline addition (n=4 biologically independent samples). **b** Percentage of cystic organoid in *Apc*$^{Min/+}$ organoids upon glutamine deprivation (0.2 mM and 0.4 mM) overtime (n=7 biologically independent cultures). **c** Percentage of cystic organoids and organoid number of *Apc*$^{Min/+}$ organoids treated with CB-839 for 1 week (n=3 biologically independent cultures). **d** Percentage of wildtype organoids with cystic morphology after 4 passages in low glutamine conditon (n=6 biologically independent cultures). **e** qPCR analysis of *Axin2* in wildtype organoids cultured in control or low glutamine medium for 1 week. Data from n=3 independent experiments with a line marking the mean value. **f**, **g** Control and glutamine-starved wildtype organoids were dissociated into single cells, and equal number of organoid-derived cells were cultured in organoid medium with 3 mM or 0.3 mM glutamine (low gln). Secondary organoid formation and percentage of cystic organoids are shown (n=9 biologically independent cultures). Data in **a**-**e** and **g** represent means +/− SD, p values were determined by two-tailed unpaired Student's t-test. Scale bars represent 1000 µm (**c**, **d**, **f**).

(Figure adapted from Extended Data Fig. 1 of Tran, Hanse, Habowski et al., 2020)

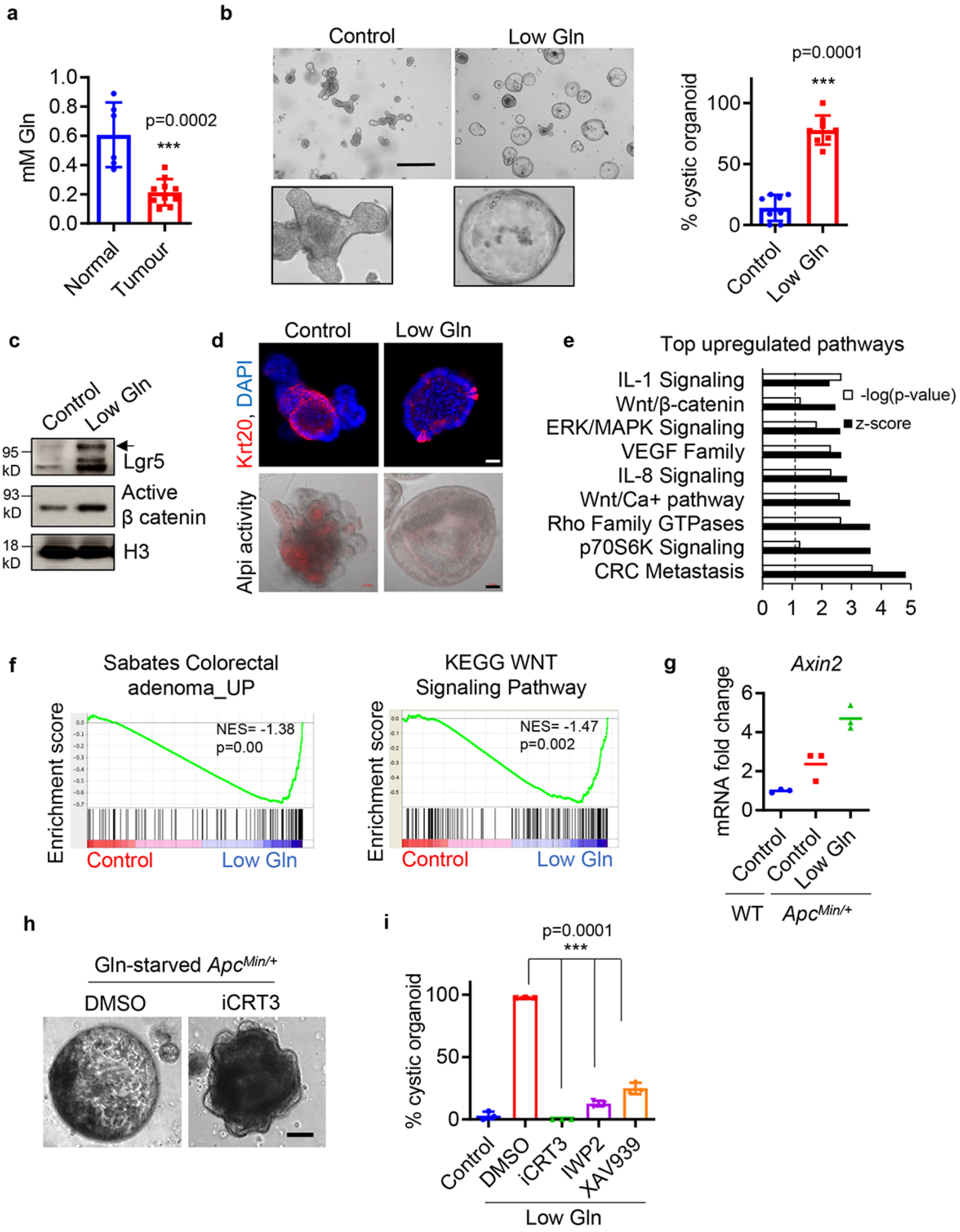

a
mM Gln
1.0
0.8
0.6
0.4
0.2
0.0
p=0.0002
***
Normal
Tumour
b
Control
Low Gln
% cystic organoid
p=0.0001
***
100
50
0
Control
Low Gln
c
Control
Low Gln
95
kD
Lgr5
93
kD
Active
β catenin
18
kD
H3
d
Control
Low Gln
Krt20, DAPI
Alpi activity
e
Top upregulated pathways
IL-1 Signaling
Wnt/β-catenin
ERK/MAPK Signaling
VEGF Family
IL-8 Signaling
Wnt/Ca+ pathway
Rho Family GTPases
p70S6K Signaling
CRC Metastasis
☐ -log(p-value)
■ z-score
0 1 2 3 4 5
f
Sabates Colorectal adenoma_UP
Enrichment score
NES= -1.38
p=0.00
Control
Low Gln
KEGG WNT Signaling Pathway
Enrichment score
NES= -1.47
p=0.002
Control
Low Gln
g
Axin2
mRNA fold change
6
4
2
0
Control
Control
Low Gln
WT
ApcMin/+
h
Gln-starved ApcMin/+
DMSO
iCRT3
i
p=0.0001
***
% cystic organoid
100
50
0
Control
DMSO
iCRT3
IWP2
XAV939
Low Gln

Figure 4.2: ***Environmental glutamine restriction hyperactivates Wnt signaling and blocks cellular differentiation*.**

a Glutamine levels in intestinal tumors from *Apc*$^{Min/+}$ mice (n=10 mice) and normal intestinal tissues of wildtype mice (n=6 mice). Data represent means ± SD, p value was determined by two-tailed unpaired Student's t-test. **b** Representative brightfield images and percentage of cystic morphology in *Apc*$^{Min/+}$ small intestine organoids cultured in control (3mM) or low glutamine (0.3 mM) medium after 4–6 passages (n=8 biologically independent cell cultures). Data represent means ± SD, p value was determined by two-tailed unpaired Student's t-test. **c** Immunoblots for stem cell marker Lgr5 and active Ctnnb1, the results are representative of two independent experiments. **d** Immunofluorescent images for intestinal differentiation markers Krt20 and Alpi enzyme activity in *Apc*$^{Min/+}$ organoids cultured in control (3mM) or low glutamine (0.3 mM) medium. The results are representative of three biologically independent samples. **e** IPA analysis of top upregulated pathways and **f** GSEA analysis between control organoids versus glutamine-starved (1 week) organoids from RNA sequencing (n=3 biologically independent samples). Dotted line in e indicates threshold of significance (p = 0.05) and p values were determined by a Right-Tailed Fisher's Exact Test. **g** qPCR analysis of *Axin2* in wildtype organoids and *Apc*$^{Min/+}$ organoids cultured in control and low glutamine (0.3 mM) medium. n= 3 technical replicates with a line marking the mean value. The experiment was repeated twice independently with similar results. **h** Representative brightfield images of glutamine-starved organoids treated with 5 µM iCRT3 for 1 week from three biologically independent samples. **i** Percentage of cystic organoids of control and glutamine-starved organoids treated with 10 µM of indicated Wnt inhibitors for 4 days (n=3 biologically independent samples). Data represent means ± SD and p values were determined by two-tailed unpaired Student's t-test. Scale bars, 400 µm **b**, 50 µm **d**, 100 µm **h**.

(Figure adapted from Fig. 1 of Tran, Hanse, Habowski et al., 2020)

a **Exome sequencing:** Control vs Low Gln

Chr	Start	End	Ref	Alt	Func.refGene	Gene.refGene	ExonicFunc.refGene	Other info
chr2	44844614	44844614	T	A	exonic	*Zeb2*	nonsynonymous SNV	heterozygous
chr5	24824820	24824820	T	G	exonic	*Kmt2c*	nonsynonymous SNV	heterozygous
chr17	37450924	37450924	T	G	exonic	*Olfr100, Olfr102*	nonsynonymous SNV	heterozygous

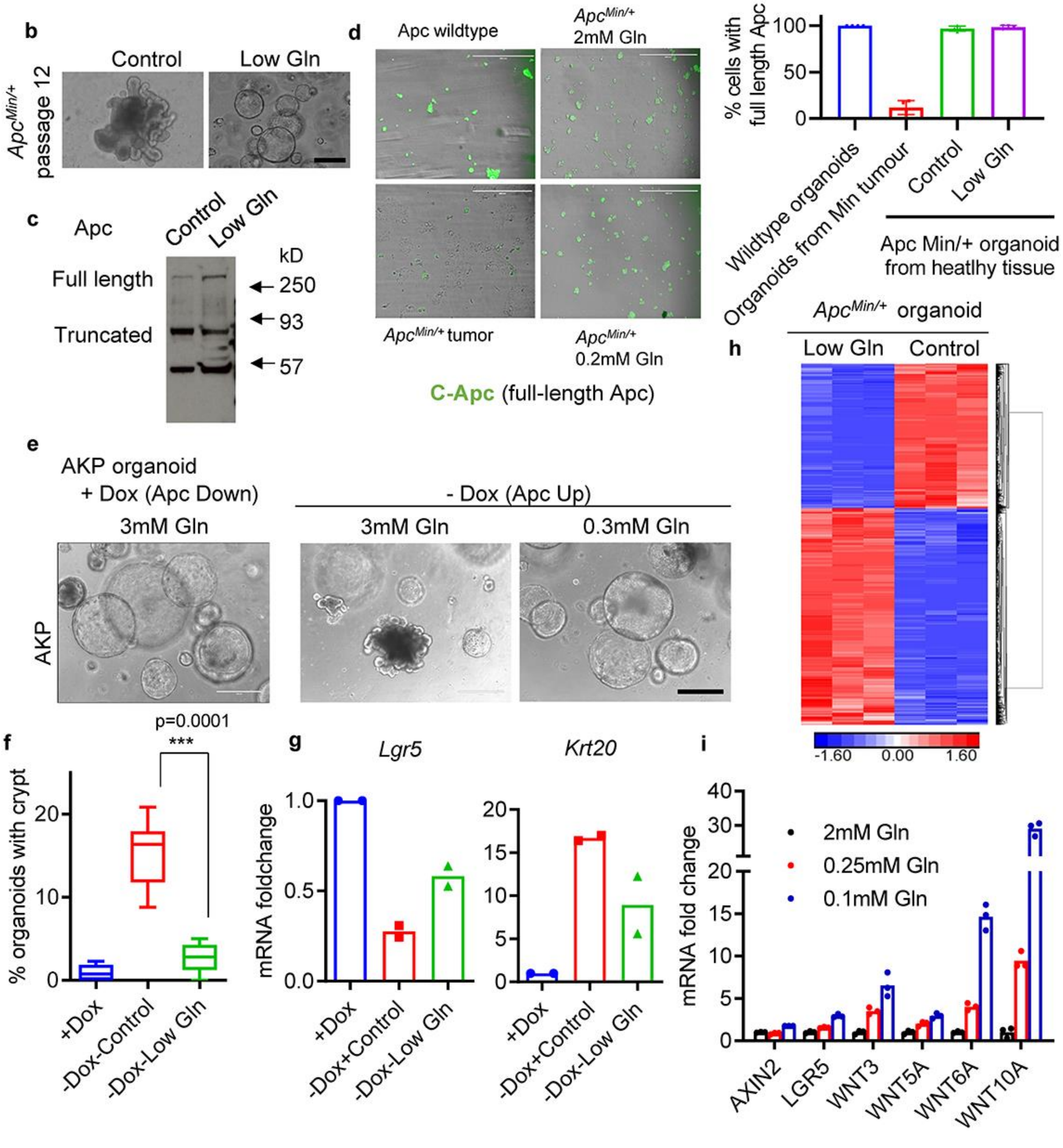


Figure 4.3: ***Genetic alterations do not contribute to low-glutamine induced stemness.***

a Identified genetic alterations in glutamine-starved *Apc*$^{Min/+}$ organoids compared to paired control organoids as determined by exome sequencing. **b** Brightfield images of control and glutamine-starved *Apc*$^{Min/+}$organoids after 12 passages. Results are representative of three biologically independent cultures. **c** Immunoblotting for full length and truncated Apc protein in control and glutamine starved *Apc*$^{Min/+}$ organoids after 8 passages. Results are representative of three independent experiments. **d** Representative images and percentage of cells with full-length Apc protein based on immunofluorescent staining with C-terminus Apc antibody in wildtype organoids, tumor organoids derived from adenomas derived in *Apc*$^{Min/+}$ mice, and *Apc*$^{Min/+}$organoids from healthy tissues in control medium and upon glutamine deprivation (n=4 biologically independent cultures), data represent means +/− SD. **e**, **f** Representative images and percentage of shApc /Kras^{G12D}/p53$^{fl/fl}$ (AKP) organoid with crypts cultured in control or low glutamine medium for 10 days (n=5 biologically independent cultures). **g** qPCR analysis of *Krt20* and *Lgr5* in a similar experiment described in **e** after 3 days of glutamine deprivation. Data from n=2 independent experiments with a line marking the mean value. **h** Hierarchical clustering of significant differentiated gene expression of *Apc*$^{Min/+}$ organoids cultured in control or low-glutamine medium (n=3 biologically independent samples). **i** qPCR analysis of the indicated genes in SW620 colon cancer cells (n=3 technical replicates and data represent means) cultured in medium with the indicated glutamine concentration for 3 days. A single experiment is shown that is representative of two independent experiments with similar results. Scale bars represent 200 µm **b** and 400 µm **d**, **e**.

(Figure adapted from Extended Data Fig. 2 of Tran, Hanse, Habowski et al., 2020)

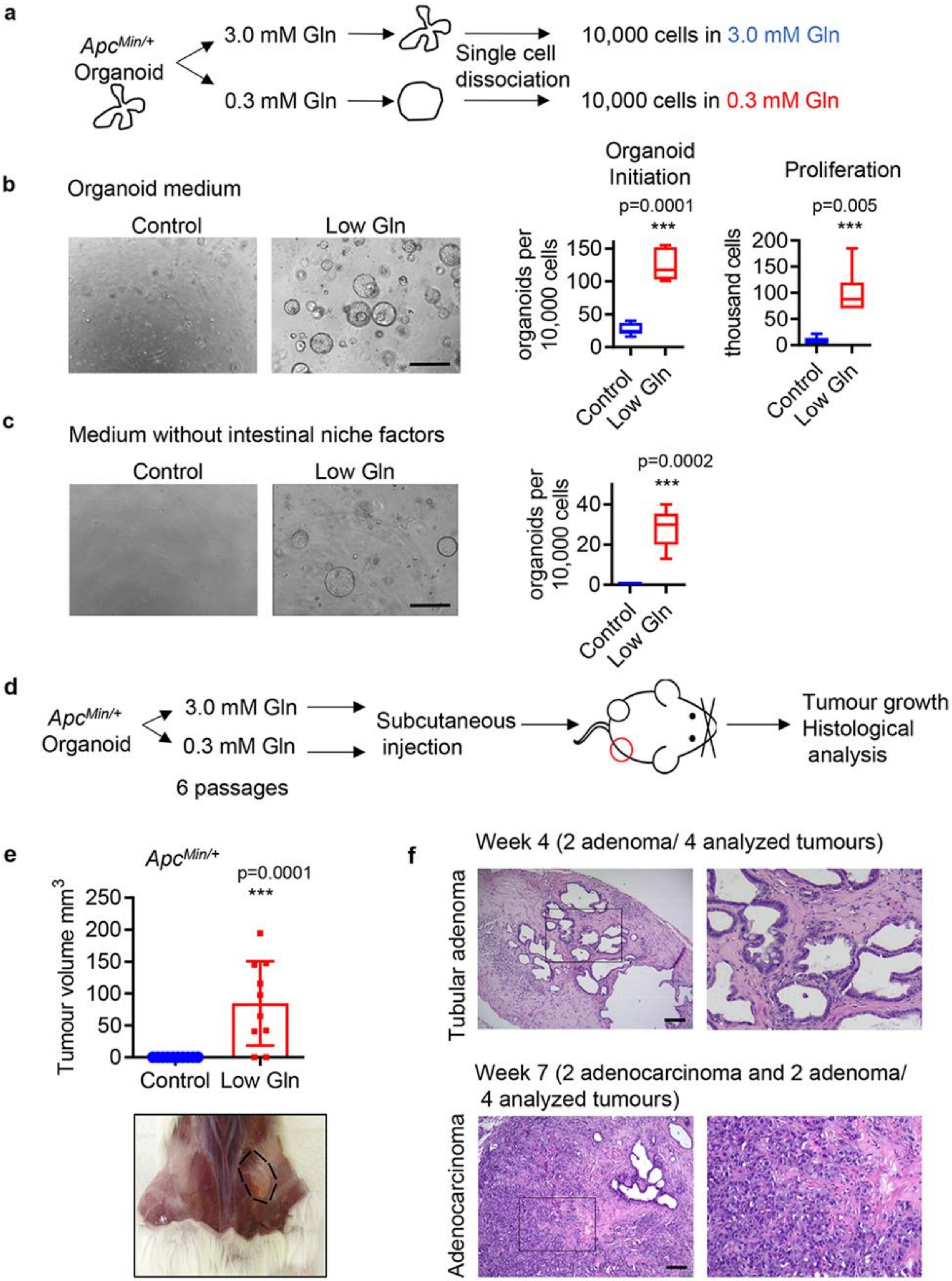

a
Apc^Min/+
Organoid
3.0 mM Gln
0.3 mM Gln
Single cell dissociation
10,000 cells in 3.0 mM Gln
10,000 cells in 0.3 mM Gln
b
Organoid medium
Control
Low Gln
Organoid Initiation
p=0.0001
***
organoids per 10,000 cells
0
50
100
150
Control
Low Gln
Proliferation
p=0.005
***
thousand cells
0
50
100
150
200
Control
Low Gln
c
Medium without intestinal niche factors
Control
Low Gln
p=0.0002
***
organoids per 10,000 cells
0
20
40
Control
Low Gln
d
Apc^Min/+
Organoid
3.0 mM Gln
0.3 mM Gln
6 passages
Subcutaneous injection
Tumour growth
Histological analysis
e
Apc^Min/+
p=0.0001
***
Tumour volume mm^3
0
50
100
150
200
250
Control
Low Gln
f
Week 4 (2 adenoma/ 4 analyzed tumours)
Tubular adenoma
Week 7 (2 adenocarcinoma and 2 adenoma/ 4 analyzed tumours)
Adenocarcinoma

Figure 4.4: ***Glutamine restriction promotes self-renewal and niche independence in Apc$^{Min/+}$ organoids.***

a Schematic of experimental design of organoid initiation assay from single cells. **b** Representative brightfield images of secondary organoid formation and cell proliferation after 1 week are shown (n= 8 biologically independent cultures for organoid initiation; n= 6 biologically independent cultures for proliferation). **c** Representative brightfield images of secondary organoid formation after 1 week in medium without R-Spondin, Egf and Noggin are shown (n= 5 biologically independent cultures). Box plots in **b**, **c** show the maximum, third quartile, median, first quartile and minimum values, and the p values were determined by two-tailed unpaired Student's t-test. **d** Schematic of experimental procedure to establish subcutaneous xenograft tumors with *Apc*$^{Min/+}$ organoids. **e** Tumor volume of subcutaneous xenografts generated with control organoids or glutamine-starved organoids harvested 2 weeks after injection (n=10 mice). Data represent means ± SD, p value was determined by two-tailed unpaired Student's t-test. **f** Representative H&E staining of n= 4 tumors generated from glutamine-starved organoids. Scale bars, 400 µm **b**, **c**, and 40 µm **f**.

(Figure adapted from Fig. 2 of Tran, Hanse, Habowski et al., 2020)

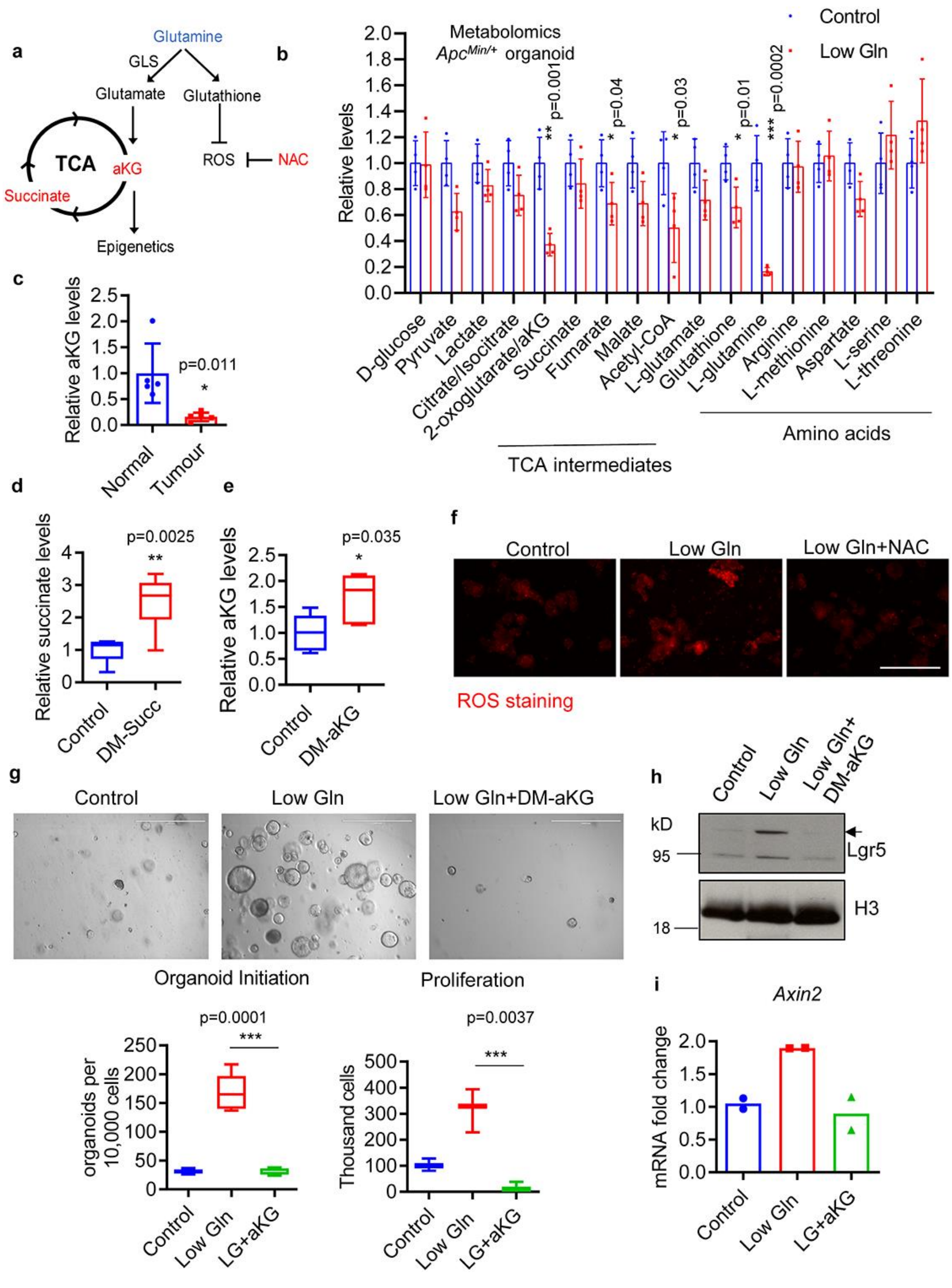

a
Glutamine
GLS
Glutamate
Glutathione
TCA
aKG
Succinate
ROS
NAC
Epigenetics
b
Metabolomics
$Apc^{Min/+}$ organoid
Control
Low Gln
Relative levels
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
p=0.001
**
p=0.04
*
p=0.03
*
p=0.01
*
p=0.0002
***
D-glucose
Pyruvate
Lactate
Citrate/Isocitrate
2-oxoglutarate/aKG
Succinate
Fumarate
Malate
Acetyl-CoA
L-glutamate
Glutathione
L-glutamine
Arginine
L-methionine
Aspartate
L-serine
L-threonine
TCA intermediates
Amino acids
c
Relative aKG levels
2.5
2.0
1.5
1.0
0.5
0.0
p=0.011
*
Normal
Tumour
d
Relative succinate levels
4
3
2
1
0
p=0.0025
**
Control
DM-Succ
e
Relative aKG levels
2.5
2.0
1.5
1.0
0.5
0.0
p=0.035
*
Control
DM-aKG
f
Control
Low Gln
Low Gln+NAC
ROS staining
g
Control
Low Gln
Low Gln+DM-aKG
Organoid Initiation
p=0.0001
***
organoids per 10,000 cells
250
200
150
100
50
0
Control
Low Gln
LG+aKG
Proliferation
p=0.0037
***
Thousand cells
500
400
300
200
100
0
Control
Low Gln
LG+aKG
h
Control
Low Gln
Low Gln+ DM-aKG
kD
95
18
Lgr5
H3
i
Axin2
mRNA fold change
2.0
1.5
1.0
0.5
0.0
Control
Low Gln
LG+aKG

Figure 4.5: ***The role of αKG in low-glutamine induced stemness.***

a Schematic diagram of glutamine metabolism. **b** Relative metabolite levels as measured by LC-MS in *Apc*$^{Min/+}$ organoids cultured in control and low glutamine medium (n=4 biologically independent samples). **c** Relative αKG levels in intestinal tumors from *Apc*$^{Min/+}$ mice and normal intestinal tissues of wildtype mice (n=5 mice per group). Data in b, c represent means +/− SD, and p values were determined by two-tailed unpaired Student's t-test. **d** and **e** Relative intracellular αKG and succinate levels in *Apc*$^{Min/+}$ organoids upon DM-αKG (n=5 biologically independent samples) or DM-succinate supplementation (n=6 biologically independent samples). **f** Immunofluorescent staining for ROS in *Apc*$^{Min/+}$ organoids under low glutamine or low glutamine medium supplemented with NAC. Results are representative from three biologically independent samples. **g** Control organoids, glutamine-starved organoids treated with or without 3.5 mM DM-αKG were dissociated into single cells. An equal number of organoid-derived cells were cultured, and secondary organoid formation (n=6 biologically independent cultures) and cell proliferation (n=3 biologically independent cultures) were measured after 1 week and are shown. p values were determined by two-tailed unpaired Student's t-test. **h** Immunoblotting for Lgr5 and **i** qPCR analysis for *Axin2* expression in *Apc*$^{Min/+}$ organoid cultured in control and low glutamine medium with or without DM-αKG (n= 2 independent experiments with a line marking the mean value). Box plots in **d**, **e**, **g** show the maximum, third quartile, median, first quartile and minimum values, and the p values were determined by two-tailed unpaired Student's t-test. Scale bar represents 400 µm **f** and 1000 µm **g**.

(Figure adapted from Extended Data Fig. 3 of Tran, Hanse, Habowski et al., 2020)

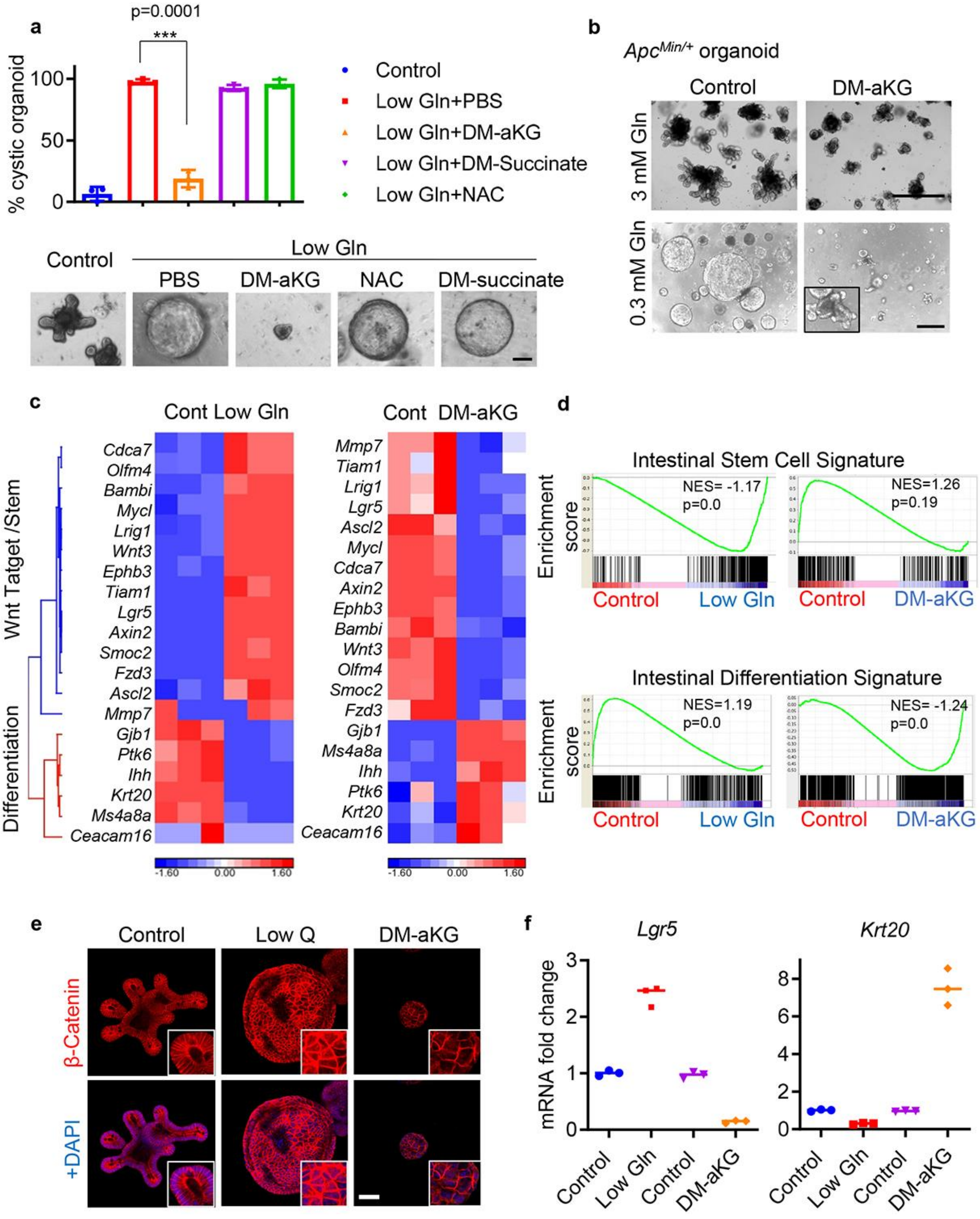
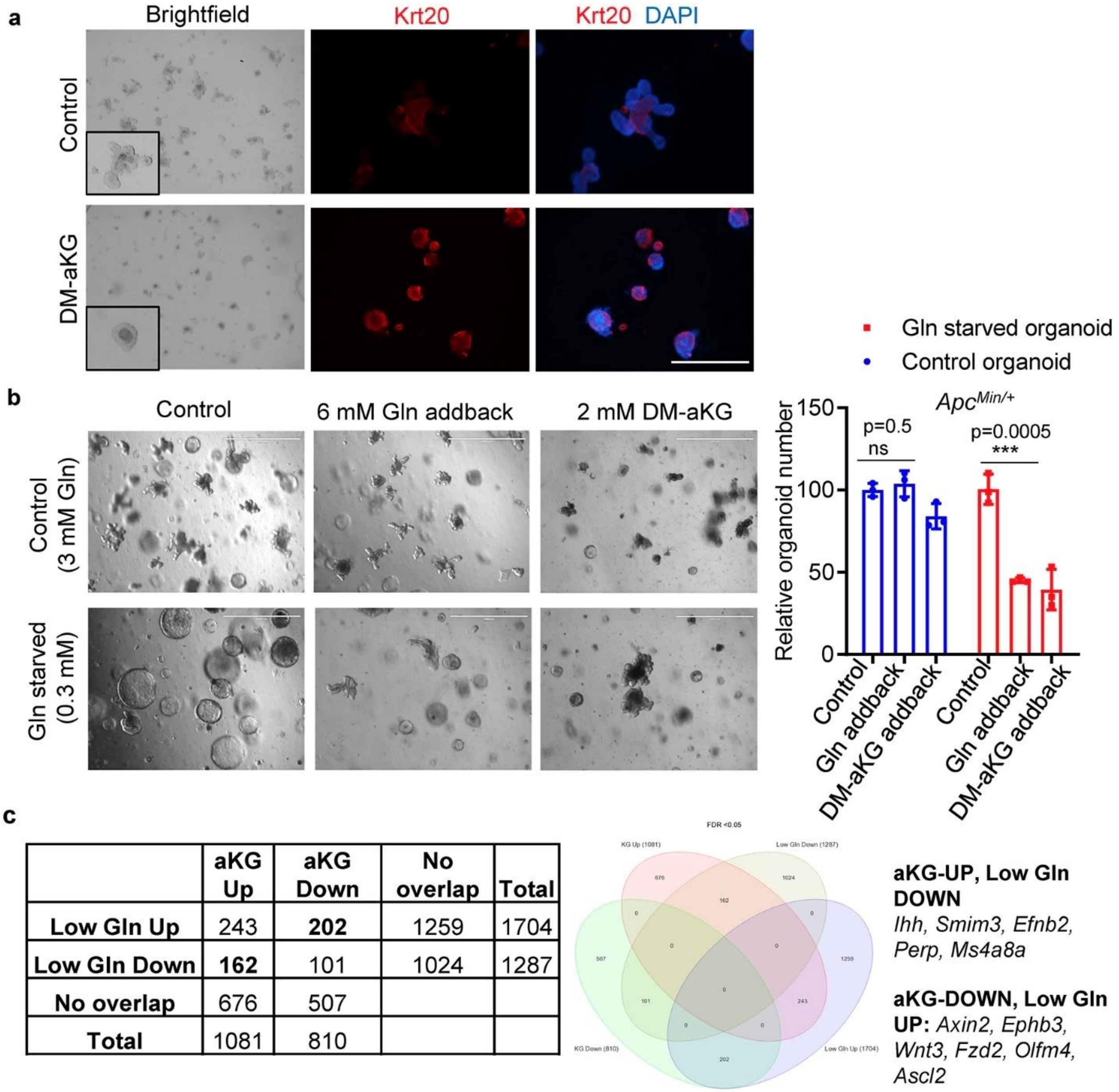

a
p=0.0001
***
% cystic organoid
100
50
0
Control
Low Gln+PBS
Low Gln+DM-aKG
Low Gln+DM-Succinate
Low Gln+NAC
Control
Low Gln
PBS
DM-aKG
NAC
DM-succinate
b
ApcMin/+ organoid
Control
DM-aKG
3 mM Gln
0.3 mM Gln
c
Cont Low Gln
Cont DM-aKG
Wnt Tatget /Stem
Differentiation
Cdca7
Olfm4
Bambi
Mycl
Lrig1
Wnt3
Ephb3
Tiam1
Lgr5
Axin2
Smoc2
Fzd3
Ascl2
Mmp7
Gjb1
Ptk6
Ihh
Krt20
Ms4a8a
Ceacam16
Mmp7
Tiam1
Lrig1
Lgr5
Ascl2
Mycl
Cdca7
Axin2
Ephb3
Bambi
Wnt3
Olfm4
Smoc2
Fzd3
Gjb1
Ms4a8a
Ihh
Ptk6
Krt20
Ceacam16
-1.60
0.00
1.60
-1.60
0.00
1.60
d
Intestinal Stem Cell Signature
Enrichment score
NES= -1.17
p=0.0
NES=1.26
p=0.19
Control
Low Gln
Control
DM-aKG
Intestinal Differentiation Signature
Enrichment score
NES=1.19
p=0.0
NES= -1.24
p=0.0
Control
Low Gln
Control
DM-aKG
e
Control
Low Q
DM-aKG
β-Catenin
+DAPI
f
Lgr5
Krt20
mRNA fold change
3
2
1
0
8
6
4
2
0
Control
Low Gln
Control
DM-aKG
Control
Low Gln
Control
DM-aKG

Figure 4.6: *αKG supplementation rescues low-glutamine induced stemness and suppresses Wnt signaling*.

a Percentage of cystic organoid morphology and representative brightfield images of control organoids and glutamine-starved *Apc*$^{Min/+}$ organoids supplemented with PBS control, 3 mM DM-αKG, 3 mM DM-succinate, or 5 mM NAC for 3 days (n=3 biologically independent cell cultures). Data represent means ± SD, p value was determined by two-tailed unpaired Student's t-test. **b** Brightfield images of control organoids and glutamine-starved *Apc*$^{Min/+}$ organoids treated with 3.5 mM DM-αKG for 3 days. Results are representative of 3 biologically independent samples. **c** Heat map of gene expression profile from RNA sequencing data and **d** GSEA analysis of *Apc*$^{Min/+}$ organoids cultured in control (3 mM glutamine), low glutamine (0.3 mM glutamine) or with 3.5 mM DM-αKG medium (n=3 biologically independent RNA samples). The nominal p values in **d** are the statistical significance of the enrichment score analyzed by GSEA. **e** Representative immunofluorescence for active Ctnnb1 from 3 biologically independent samples. **f** qPCR analysis for *Lgr5* and *Krt20* expression in *Apc*$^{Min/+}$ organoids cultured in low glutamine medium or treated with 3.5 mM DM-αKG. n=3 technical replicates with a line marking the mean value. The experiment was repeated three times independently with similar results. Scale bars, 100 µm **a**, 400 µm **b**, and 50 µm **e**.

(Figure adapted from Fig. 3 of Tran, Hanse, Habowski et al., 2020)



	aKG Up	aKG Down	No overlap	Total
Low Gln Up	243	**202**	1259	1704
Low Gln Down	**162**	101	1024	1287
No overlap	676	507		
Total	1081	810		

Figure 4.7: ***The effect of αKG and glutamine supplementation on intestinal differentiation.***
a Representative brightfield images and immunofluorescent staining of the differentiation marker Krt20 in *Apc*$^{Min/+}$ organoids treated with 3 mM DM-αKG for 3 days. Results are representative of three independent experiments. **b** Representative images and relative organoid number of control *Apc*$^{Min/+}$ organoid or glutamine-starved *Apc*$^{Min/+}$ organoids upon 2 mM DM-αKG treatment or 6mM glutamine addition for 1 week (n=3 biologically independent cultures). Data represent means +/− SD, and the p values were determined by two-tailed unpaired Student's t-test. **c** Overlapping gene expression profile of *Apc*$^{Min/+}$ organoids cultured in low glutamine medium or treated with αKG reveals opposing regulation on Wnt target genes and intestinal differentiation related genes. Scale bars represent 1000 µm (Brightfield), 200 µm (Immunofluorescence).

(Figure adapted from Extended Data Fig. 4 of Tran, Hanse, Habowski et al., 2020)

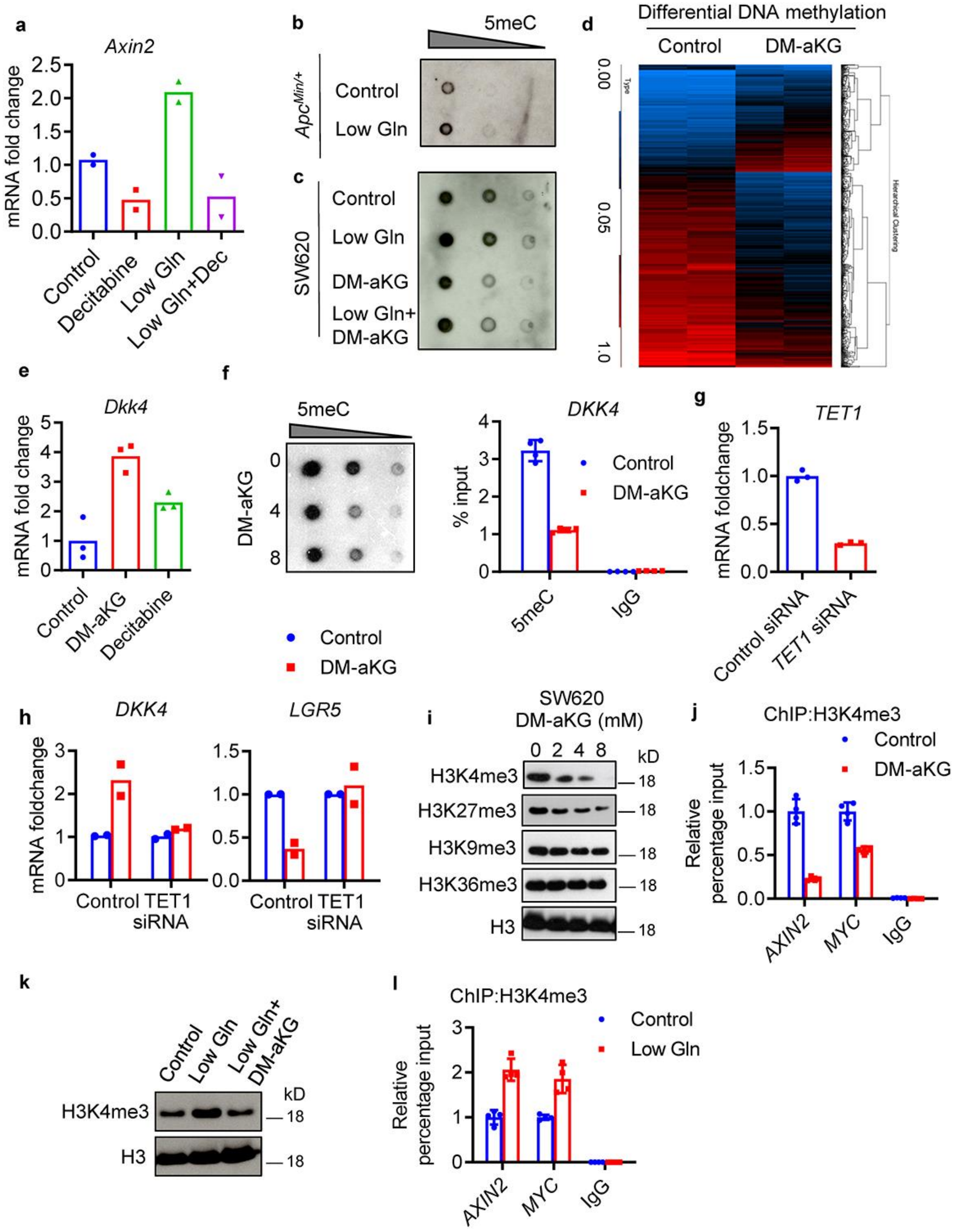

a
Axin2
mRNA fold change
2.5
2.0
1.5
1.0
0.5
0.0
Control
Decitabine
Low Gln
Low Gln+Dec
b
5meC
$Apc^{Min/+}$
Control
Low Gln
c
SW620
Control
Low Gln
DM-aKG
Low Gln+
DM-aKG
d
Differential DNA methylation
Control
DM-aKG
0.00
0.05
1.0
Type
Hierarchical Clustering
e
Dkk4
mRNA fold change
5
4
3
2
1
0
Control
DM-aKG
Decitabine
f
5meC
DM-aKG
0
4
8
Control
DM-aKG
DKK4
% input
4
3
2
1
0
Control
DM-aKG
5meC
IgG
g
TET1
mRNA foldchange
1.5
1.0
0.5
0.0
Control siRNA
TET1 siRNA
h
DKK4
LGR5
mRNA foldchange
3
2
1
0
1.5
1.0
0.5
0.0
Control
TET1 siRNA
Control
TET1 siRNA
i
SW620
DM-aKG (mM)
0
2
4
8
kD
H3K4me3
H3K27me3
H3K9me3
H3K36me3
H3
18
18
18
18
18
j
ChIP:H3K4me3
Control
DM-aKG
Relative percentage input
1.5
1.0
0.5
0.0
AXIN2
MYC
IgG
k
Control
Low Gln
Low Gln+ DM-aKG
kD
H3K4me3
H3
18
18
l
ChIP:H3K4me3
Control
Low Gln
Relative percentage input
3
2
1
0
AXIN2
MYC
IgG

Figure 4.8: ***αKG promotes hypomethylation of histone and DNA in CRC cells.***

a qPCR analysis of *Axin2* in control and glutamine-starved organoids treated with 1 µM decitabine for 3 days. Data from n=2 independent experiments with a line marking the mean value. (b, c) Dot blot analysis of 5meC levels in *Apc*$^{Min/+}$ organoids in control and low glutamine medium and SW620 cells in control, low glutamine medium or low glutamine medium supplemented with 8 mM DM-αKG. Results are representative from two independent experiments. **d** Heatmap of the differential methylated regions (different methylated ratio >+/−20%) in SW620 cells upon 8 mM DM-αKG treatment for 3 days. Beta value of the methylation ratio are shown (top) (n= 2 biologically independent samples). **e** qPCR analysis of *Dkk4* in *Apc*$^{Min/+}$ organoids treated with 3.5 mM DM-αKG or 1 µM decitabine (n=3 technical replicates). A repeat experiment showed similar results. **f** Dot blot analysis of 5meC levels in SW620 cells treated with DM-αKG (left), MeDIP experiment with 5meC antibody for *DKK4* promoter in SW620 cells upon 8 mM DM-αKG treatment (right). Data show means ± SD of n=4 technical replicates. Results are representative of two independent experiments. **g** qPCR analysis of TET1 expression in SW620 cells transfected with control siRNA or TET1 siRNA (data show means of n=3 technical replicates). **h** qPCR analysis of *DKK4* and *LGR5* expression in control SW620 cells or *TET1* siRNA knockdown cells following DM-αKG treatment. Data from n=2 independent experiments with a line marking the mean value. **i** Representative immunoblot of histone methylation in SW620 cells treated with DM-αKG from two independent experiments. **j** ChIP analysis of H3K4 levels on promoter regions of *AXIN2* and *MYC* in SW620 cells in response to 8 mM DM-αKG treatment for 3 days (n=4 technical replicates). **k** Representative immunoblot of H3K4me3 in SW620 cells in control, low-glutamine medium or low glutamine medium supplemented with 8 mM DM-αKG from two independent experiments. **l** ChIP analysis of H3K4me3 levels on promoter regions of *AXIN2* and *MYC* in SW620 cells in response to glutamine starvation after 1 week (n=4 technical replicates). Results in **j** and **l** represent means ± SD and are representative of two independent experiments.

(Figure adapted from Extended Data Fig. 5 of Tran, Hanse, Habowski et al., 2020)

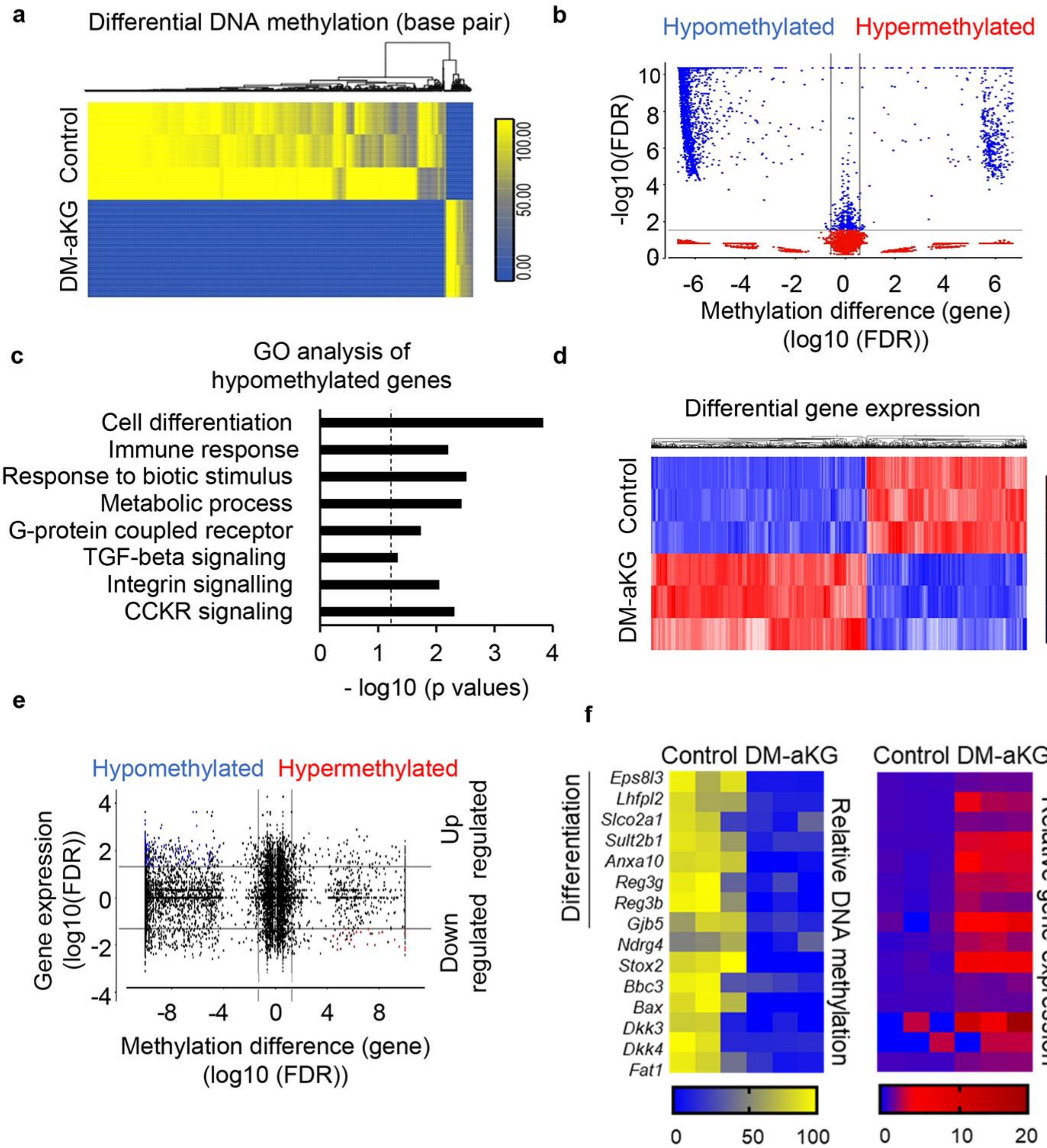

a
Differential DNA methylation (base pair)
Control
DM-aKG
100.00
50.00
0.00
b
Hypomethylated
Hypermethylated
-log10(FDR)
10
8
6
4
2
0
-6
-4
-2
0
2
4
6
Methylation difference (gene)
(log10 (FDR))
c
GO analysis of
hypomethylated genes
Cell differentiation
Immune response
Response to biotic stimulus
Metabolic process
G-protein coupled receptor
TGF-beta signaling
Integrin signalling
CCKR signaling
0
1
2
3
4
- log10 (p values)
d
Differential gene expression
Control
DM-aKG
1.60
0.00
-1.60
e
Hypomethylated
Hypermethylated
Gene expression
(log10(FDR))
4
2
0
-2
-4
-8
-4
0
4
8
Up
regulated
Down
regulated
Methylation difference (gene)
(log10 (FDR))
f
Control
DM-aKG
Differentiation
Eps8l3
Lhfpl2
Slco2a1
Sult2b1
Anxa10
Reg3g
Reg3b
Gjb5
Ndrg4
Stox2
Bbc3
Bax
Dkk3
Dkk4
Fat1
Relative DNA methylation
0
50
100
Control
DM-aKG
Relative gene expression
0
10
20

Figure 4.9: ***αKG supplementation leads to DNA hypomethylation of genes related to differentiation and Wnt inhibition.***

a Heatmap showing base pairs with differential DNA methylation based on RRBS sequencing (n=3 biologically independent samples). **b** Volcano plot showing genes with affected DNA methylation between the control and DM-αKG (3.5mM) treated *Apc*$^{Min/+}$ organoids, DNA methylation difference is plotted on the x-axis and p values are plotted on the y-axis (n=3 biologically independent samples). Different methylation pattern between samples are detected using ANOVA with FDR adjusted p-value <0.05 and +/−1.5-fold change on methylation percentage. The average of methylation difference of significant (FDR adjusted p-value) differential methylation sites was calculated. **c** Panther gene ontology analysis of hypomethylated genes in DM-αKG treated organoids (n=3 biologically independent samples). Dotted line indicates threshold of significance (p = 0.05) and p values were calculated by GO enrichment analysis software. **d** Hierarchical clustering of differential gene expression in control and DM-αKG treated *Apc*$^{Min/+}$ organoids (n=3 biologically independent samples). **e** Starburst plot for comparison between DNA methylation and gene expression. The black line represents the cutoff of FDR adjusted p values < 0.05. Blue dots represent genes with DNA hypomethylation and upregulated expression, red dots represent genes with DNA hypermethylation and downregulated expression between control and DM-αKG treated *Apc*$^{Min/+}$ organoids (n=3 biologically independent samples). Different methylation pattern between samples are detected using ANOVA with FDR adjusted p-value <0.05. **f** Heatmap showing selected genes with decreased DNA methylation at upstream regions and upregulated expression in *Apc*$^{Min/+}$ organoids upon DM-αKG treatment (n=3 biologically independent samples).

(Figure adapted from Fig. 4 of Tran, Hanse, Habowski et al., 2020)

a

ID	Location	Type	MMR Status	Other Mutation
T1	Cecum	Tubulovillous adenoma	Proficient	
T9	Colon	adenocarcinoma	Proficient	
T13	Colon	adenocarcinoma	Proficient	
T23	Colon	adenocarcinoma with liver metastasis	Proficient	KRAS (c.38G>A) (p.G13D)
T27	Cecum	adenocarcinoma	Proficient	

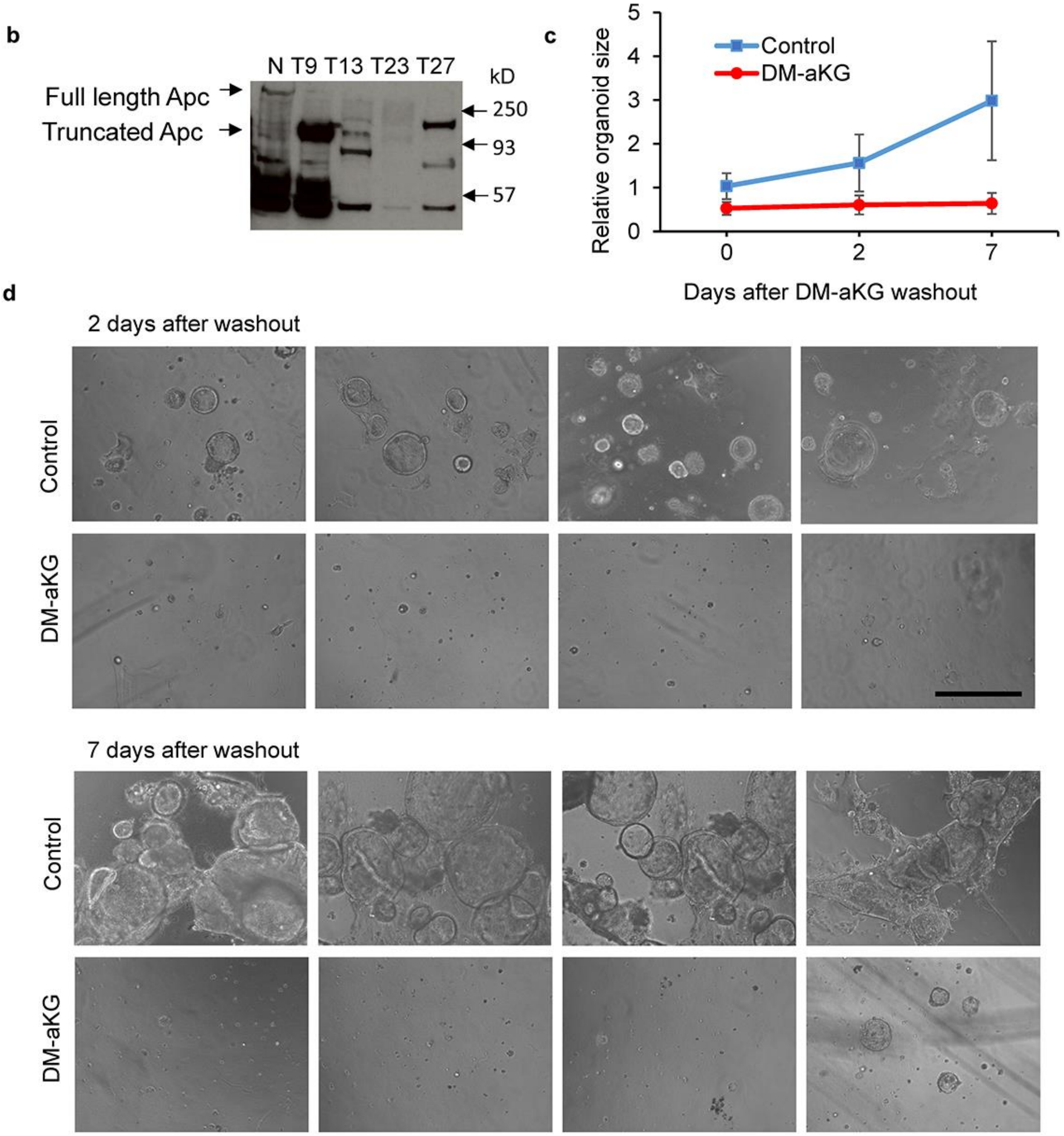


Figure 4.10: ***DM-αKG treatment inhibits initiation and growth of PDOs*.**

a Clinical information on PDOs used in the study. **b** Immunoblot probed for Apc protein in different PDOs. **c** Relative organoid size (n=50 organoids) and **d** representative images of four biologically independent cultures of T23 PDO treated with 6 mM DM-αKG for 7 days, followed by metabolite wash-out and subsequent culture for 7 days. Data in **c** represent means ± SD. Scale bar in **d** represents 400 µm.

(Figure adapted from Extended Data Fig. 6 of Tran, Hanse, Habowski et al., 2020)

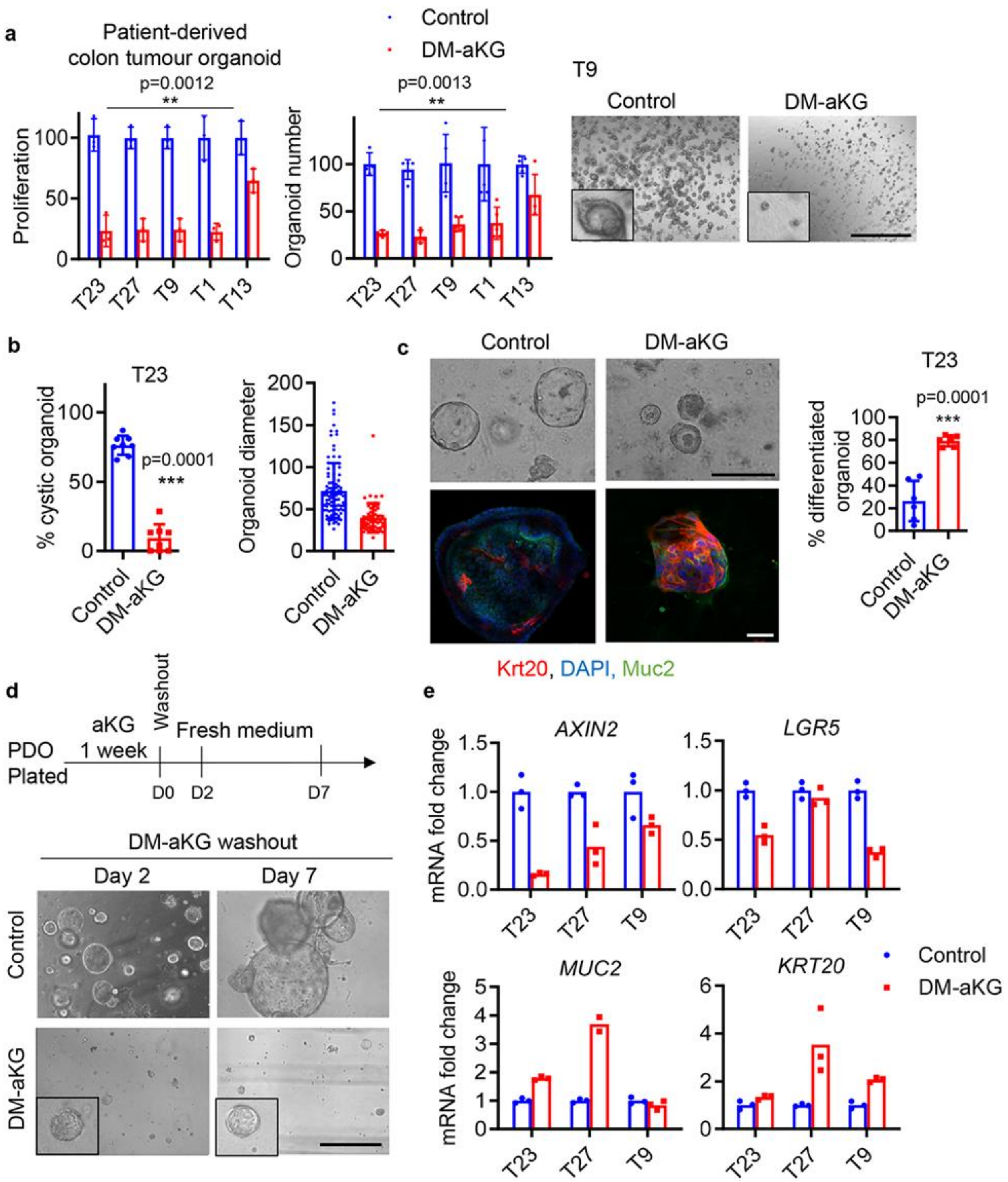

a
Patient-derived
colon tumour organoid
Control
DM-aKG
p=0.0012
**
p=0.0013
**
Proliferation
Organoid number
0
50
100
T23
T27
T9
T1
T13
T9
Control
DM-aKG
b
T23
% cystic organoid
p=0.0001
***
Control
DM-aKG
Organoid diameter
0
50
100
150
200
c
Control
DM-aKG
T23
p=0.0001
***
% differentiated organoid
0
20
40
60
80
100
Krt20, DAPI, Muc2
d
aKG
1 week
Washout
Fresh medium
PDO
Plated
D0
D2
D7
DM-aKG washout
Day 2
Day 7
Control
DM-aKG
e
AXIN2
LGR5
MUC2
KRT20
mRNA fold change
0.0
0.5
1.0
1.5
0
1
2
3
4
5
6
Control
DM-aKG

Figure 4.11: ***αKG supplementation drives terminal differentiation and suppresses growth of patient-derived colon tumor organoids.***

a Titer-Glo proliferation assay and relative organoid number of a panel of PDOs treated with 6 mM DM-αKG for 7 days (n=3 biologically independent samples). Representative brightfield images of T9 PDOs after DM-αKG treatment are shown. Data represent means ± SD, p values were determined by two-tailed paired Student's t-test between control and treated group. **b** Percentage of cystic organoids (n= 8 biologically independent cultures) and relative organoid size of T23 PDOs (n=95 organoids in control group; n=62 organoids in DM-αKG treated group) treated with DM-αKG as determined by ImageJ. Data represent means ± SD, p values were determined by two-tailed unpaired Student's t-test. **c** Brightfield images and immunofluorescent staining for the indicated differentiation markers of T23 PDOs following 6 mM DM-αKG treatment for 7 days, and the percentage of differentiated organoids as determined by Krt20 staining was shown (n=6 biologically independent cultures). Data represent means ± SD, p values was determined by two-tailed unpaired Student's t-test. **d** Representative images of T23 PDOs treated with DM-αKG followed by wash-out. Results are representative of three independent cell cultures. **e** qPCR analysis of the indicated genes in different PDOs treated with 6mM DM-αKG for 7 days, n=3 technical replicates. Scale bars, 1000 µm **a**, 200 µm and 50 µm **c**, and 400 µm **d**.

(Figure adapted from Fig. 5 of Tran, Hanse, Habowski et al., 2020)

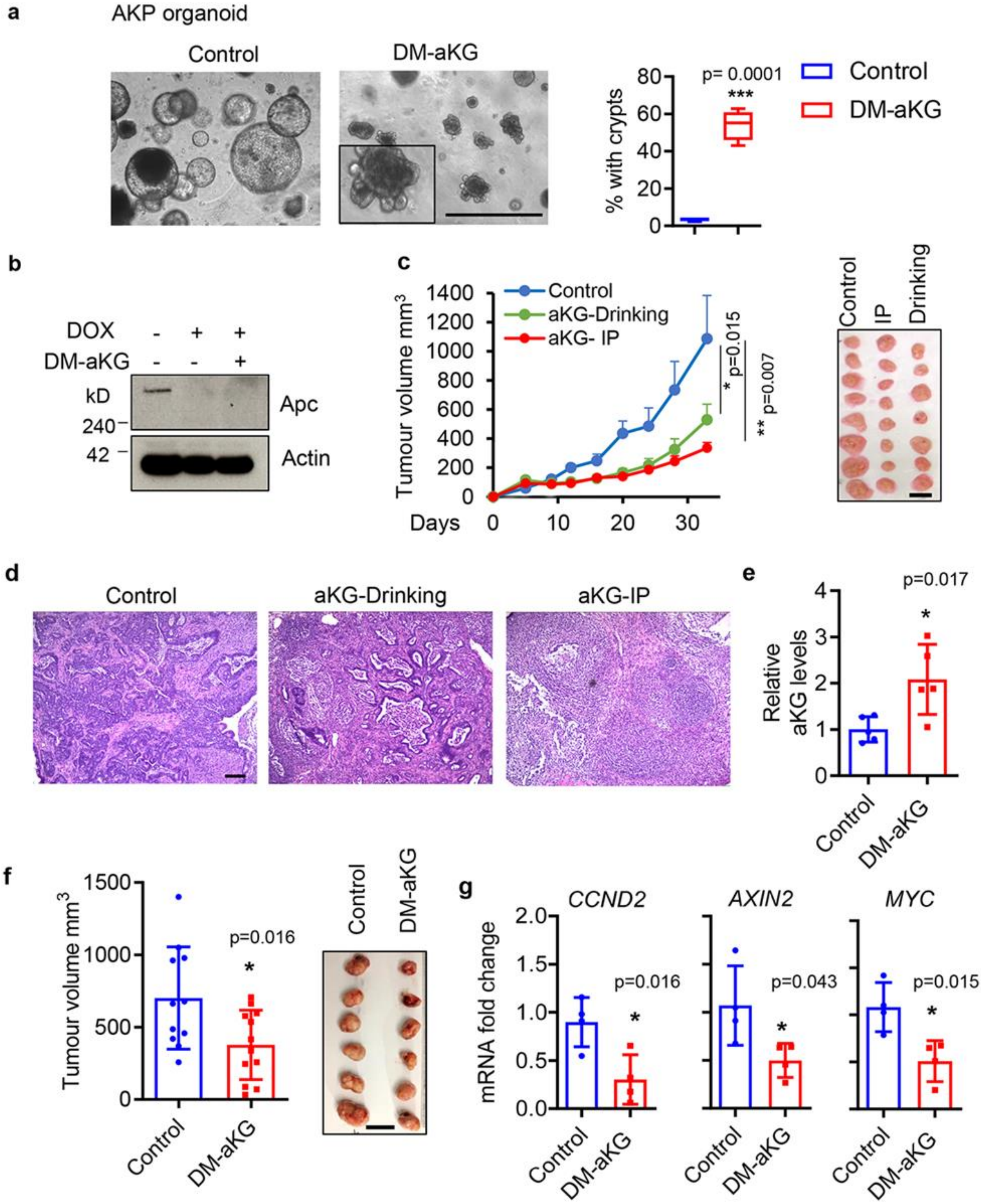

a
AKP organoid
Control
DM-aKG
% with crypts
80
60
40
20
0
p= 0.0001
***
Control
DM-aKG
b
DOX
- + +
DM-aKG
- - +
kD
240
42
Apc
Actin
c
Tumour volume mm^3
1400
1200
1000
800
600
400
200
0
Days
0
10
20
30
Control
aKG-Drinking
aKG- IP
* p=0.015
** p=0.007
Control
IP
Drinking
d
Control
aKG-Drinking
aKG-IP
e
Relative aKG levels
4
3
2
1
0
p=0.017
*
Control
DM-aKG
f
Tumour volume mm^3
1500
1000
500
0
p=0.016
*
Control
DM-aKG
Control
DM-aKG
g
mRNA fold change
2.0
1.5
1.0
0.5
0.0
CCND2
p=0.016
*
AXIN2
p=0.043
*
MYC
p=0.015
*
Control
DM-aKG
Control
DM-aKG
Control
DM-aKG

Figure 4.12: ***αKG supplementation inhibits the growth of highly mutated CRC tumors in vivo.***

a Brightfield images of shApc /Kras^{G12D}/p53$^{fl/fl}$ (AKP) small intestine organoids treated with 3 mM DM-αKG for 3 days, the percentage of organoids with crypts are shown (n=4 biologically independent cultures). Box plots show the maximum, third quartile, median, first quartile and minimum values, and the p value was determined by two-tailed unpaired Student's t-test. **b** Immunoblot of Apc in AKP organoids upon DM-αKG treatment. The results are representative of two independent experiments. **c** Tumor volume of xenograft tumors established with AKP organoids treated with 600 mg/kg DM-αKG via IP injection daily or 25 mg/ml in drinking water (n=8 tumors per group). Data represent means + SEM, p values at day 20 are shown and were determined by two-tailed unpaired Student's t-test. **d** H&E staining of control AKP tumor and DM-αKG treated tumors. Results are representative of 3 tumors per group. **e** Relative αKG levels in xenograft tumors established with SW620 colon cancer cells treated with 400 mg/kg DM-αKG by IP injection daily (n=5 tumors per group). **f** Tumor volume measured after DM-αKG treatment for 23 days (n=11 tumors per group). **g** qPCR analysis of Wnt target genes in SW620 xenograft tumors treated with DM-αKG or vehicle control (n=4 tumors per group). Data in **e-g** represent means ± SD, p values were determined by two-tailed unpaired Student's t-test. Scale bars, 1000 µm **a**, 20 mm **c,f**, and 100 µm **d**.

(Figure adapted from Fig. 6 of Tran, Hanse, Habowski et al., 2020)

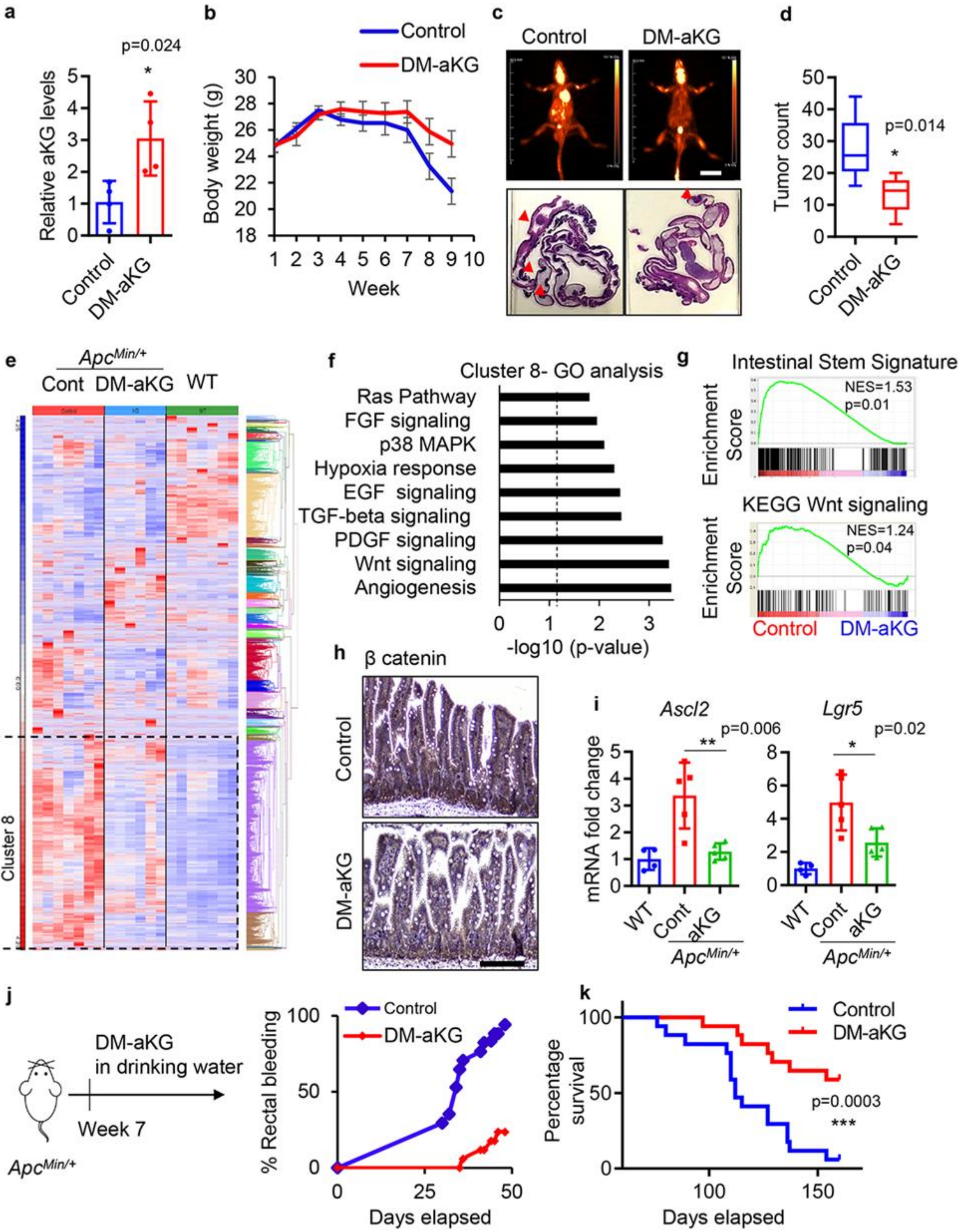

a
p=0.024
*
Relative aKG levels
0
1
2
3
4
5
Control
DM-aKG
b
Control
DM-aKG
Body weight (g)
18
20
22
24
26
28
30
Week
1 2 3 4 5 6 7 8 9 10
c
Control
DM-aKG
d
Tumor count
0
10
20
30
40
50
p=0.014
*
Control
DM-aKG
e
$Apc^{Min/+}$
Cont
DM-aKG
WT
Cluster 8
f
Cluster 8- GO analysis
Ras Pathway
FGF signaling
p38 MAPK
Hypoxia response
EGF signaling
TGF-beta signaling
PDGF signaling
Wnt signaling
Angiogenesis
0
1
2
3
-log10 (p-value)
g
Intestinal Stem Signature
NES=1.53
p=0.01
Enrichment Score
KEGG Wnt signaling
NES=1.24
p=0.04
Enrichment Score
Control
DM-aKG
h
β catenin
Control
DM-aKG
i
Ascl2
**
p=0.006
mRNA fold change
0
1
2
3
4
5
WT
Cont
aKG
$Apc^{Min/+}$
Lgr5
*
p=0.02
0
2
4
6
8
WT
Cont
aKG
$Apc^{Min/+}$
j
DM-aKG
in drinking water
Week 7
$Apc^{Min/+}$
Control
DM-aKG
% Rectal bleeding
0
50
100
0
25
50
Days elapsed
k
Control
DM-aKG
Percentage survival
0
50
100
100
150
Days elapsed
p=0.0003
***

Figure 4.13: ***αKG supplementation is an effective therapeutic intervention in a mouse model of intestinal cancer.***

a Relative αKG levels in intestinal tissues of *Apc*$^{Min/+}$ mice treated with 400 mg/kg DM-αKG via IP injection or vehicle control (n=4 mice per group). Data represent means ± SD, p value was determined by two-tailed unpaired Student's t-test. **b** Body weight of *Apc*$^{Min/+}$ mice treated with DM-αKG or vehicle control (n=8 mice per group), data represent means ± SEM. **c** PET scan images at day 70 after treatment (top panels) and H&E images of intestinal tissues of *Apc*$^{Min/+}$ mice treated with DM-αKG (bottom panels). The results are representative of 3 mice per group. **d** Number of visible intestinal tumors (n=6 mice per group) in *Apc*$^{Min/+}$ mice treated with DM-αKG or vehicle control. Box plots show the maximum, third quartile, median, first quartile and minimum values, and the p value was determined by two-tailed unpaired Student's t-test. **e** Hierarchical clustering of gene expression from RNA sequencing data in tumor-free intestinal tissues of wildtype B6 mice, *Apc*$^{Min/+}$ mice and *Apc*$^{Min/+}$ mice treated with DM-αKG (n= 7 mice in control group, 6 mice in DM-αKG treated group, 7 mice in wildtype group). **f**, **g** Panther gene ontology enrichment analysis of cluster 8 and GSEA analysis for the indicated gene signatures between control *Apc*$^{Min/+}$ mice and DM-αKG treated mice (n= 7 mice in control group and 6 mice in DM-αKG treated group). Dotted line in **f** indicate threshold of significance (p = 0.05) and p values were calculated by GO enrichment analysis software. The nominal p values in **g** are the statistical significance of the enrichment score analyzed by GSEA software. **h** IHC staining for Ctnnb1 of *Apc*$^{Min/+}$ mice treated with vehicle control or DM-αKG. The results shown are representative of 3 mice per group. **i** qPCR analysis of stem/Wnt target genes (n=5 tumors for *Apc*$^{Min/+}$ mice and n= 4 tumors for wildtype mice). Data represent means ± SD, p values were determined by two-tailed unpaired Student's t-test. **j** Percentage of rectal bleeding, an indication of intestinal tumors, and **k** percentage survival of *Apc*$^{Min/+}$ mice supplemented with 15 mg/ml DM-αKG in the drinking water (n=17 mice per group). The p value in **k** was determined by Log-rank (Mantel-Cox) test. Scale bars, 20 mm **c** and 100 µm **h**.

(Figure adapted from Fig. 7 of Tran, Hanse, Habowski et al., 2020)

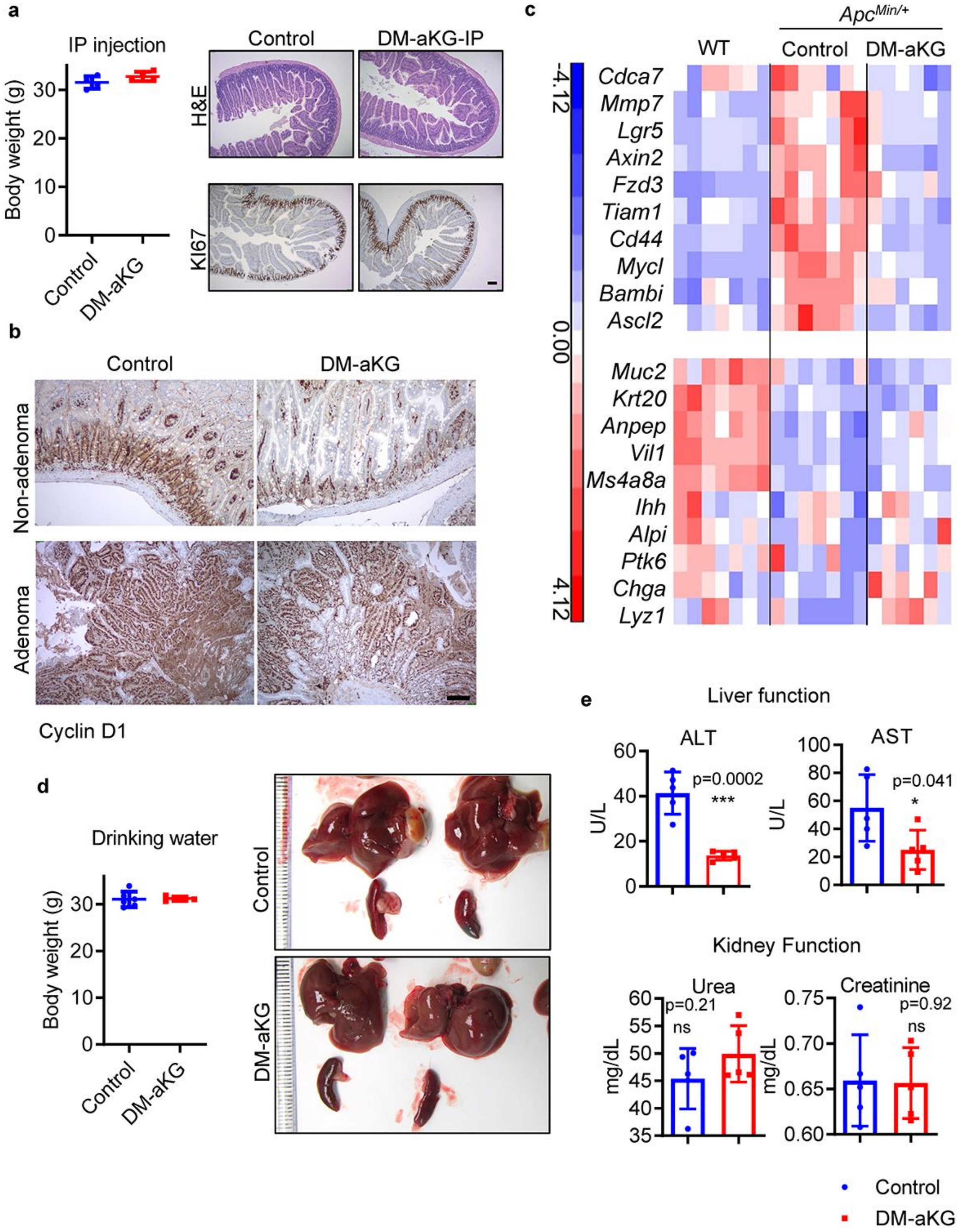

a
IP injection
Body weight (g)
0
10
20
30
Control
DM-aKG
Control
DM-aKG-IP
H&E
KI67
b
Control
DM-aKG
Non-adenoma
Adenoma
Cyclin D1
c
$Apc^{Min/+}$
WT
Control
DM-aKG
-4.12
0.00
4.12
Cdca7
Mmp7
Lgr5
Axin2
Fzd3
Tiam1
Cd44
Mycl
Bambi
Ascl2
Muc2
Krt20
Anpep
Vil1
Ms4a8a
Ihh
Alpi
Ptk6
Chga
Lyz1
d
Drinking water
Body weight (g)
0
10
20
30
Control
DM-aKG
Control
DM-aKG
e
Liver function
ALT
U/L
0
20
40
60
p=0.0002
***
AST
U/L
0
20
40
60
80
100
p=0.041
*
Kidney Function
Urea
mg/dL
35
40
45
50
55
60
p=0.21
ns
Creatinine
mg/dL
0.60
0.65
0.70
0.75
p=0.92
ns
● Control
■ DM-aKG

Figure 4.14: ***The effect of DM-αKG treatment in mice*.**

a Body weight and histological analysis of wildtype mice treated with 400 mg/kg DM-αKG via IP injection for more than 2 months (n=4 mice per group). **b** Representative IHC staining for Cyclin D1 in intestinal tissues collected from *Apc*$^{Min/+}$ mice treated with DM-αKG from three mice per group. **c** Gene expression analysis from RNA sequencing performed on the intestinal tissues of wildtype mice (n=7 mice), *Apc*$^{Min/+}$ mice (n=7 mice), and *Apc*$^{Min/+}$ mice treated with DM-αKG (n=6 mice). **d** Body weight changes and images of liver and spleen from *Apc*$^{Min/+}$ mice treated with DM-αKG (n=5 mice per group). **e** Liver and kidney function of wildtype mice treated with 15 mg/ml DM-αKG supplemented in drinking water for more than 4 months (n=5 mice per group). Data shown in **a**, **d**, **e** are means +/− SD. The p values in **e** were determined by two-tailed unpaired Student's t-test.

(Figure adapted from Extended Data Fig. 7 of Tran, Hanse, Habowski et al., 2020)

Table 4.1: ***Genes with DNA hypomethylation and upregulated gene expression upon DM-αKG treatment.***

1700057G04Rik	Bdnf	Dcun1d3	Gm12602	Klc3	Nfic	Rassf6	Sult2b1
2210408F21Rik	Bicd1	Ddah1	Gm13056	Klhl5	Nt5c1a	Rbpms	Svop
4933404O12Rik	Blvrb	Ddah2	Gm1943	Lats2	Nudt17	Reg3b	Sytl1
6030419C18Rik	Bmp8b	Ddx60	Gm19510	Ldlrap1	Oaf	Reg3g	Tacc2
8430408G22Rik	Btbd10	Dnah10	Gm44805	Lhfpl2	Oma1	Relb	Tacstd2
Aars	Btc	Dock6	Gm7367	Lmna	Omp	Rfx5	Tbc1d2
Abcc4	Camsap2	Drc1	Gm9530	Lonrf3	Osgin1	Rgl3	Tdrd12
Abhd2	Capg	Dscaml1	Gmip	Lrrc28	Palld	Rhoc	Tead4
Abhd5	Capn5	Dst	Gpat3	Ltbp4	Paqr3	Rhof	Tec
Acnat1	Car12	Dync1h1	Gpr4	Ltf	Paqr5	Rimbp2	Terb1
Acot11	Cars	Ecm1	Gprc5a	Macf1	Paqr8	Rnf123	Tgfa
Acsf2	Ccdc160	Efna3	Gprc5b	Mafg	Pard6b	Rnf223	Tgfb1
Acy1	Ccdc180	Efnb2	Gpt2	Magi2	Parp3	Rtkn	Tln2
Adam22	Cck	Ehd2	Gramd1b	Map3k6	Pde4c	Rusc2	Tmem132b
Aen	Cd44	Emp1	Grhl3	Map7d1	Pde4dip	S100a13	Tmem150b
Agpat4	Cers3	Endod1	Gsap	Mapre3	Pecam1	S100a3	Tmie
Akr1b7	Ces1g	Enho	Gsta1	Mast4	Piezo1	S100g	Tnfrsf9
Aldh4a1	Cidec	Epha2	Gtf3c1	Matn2	Pitpnm2	Sat1	Tpbg
Als2cl	Ckap2	Eps8l3	Hepacam	Mctp2	Pitpnm3	Sdr42e1	Trim29
Amotl1	Clca4a	Ereg	Hmgcll1	Mgst1	Plat	Sesn2	Trpc1
Ank3	Cldn23	Esam	Homer3	Mib2	Plaur	Sh2b2	Tsku
Anln	Cldn4	Etv4	Hspg2	Mst1r	Plcg2	Sh3bgrl2	Ttc22
Ano1	Cnga3	Fam102b	Hykk	Mtg1	Plec	Slc12a4	Ttc39b
Anxa10	Cntfr	Fam19a3	Ick	Mtm1	Plk3	Slc25a24	Tuba8
Anxa3	Col26a1	Fam46b	Ifnlr1	Mvp	Plxna1	Slc25a33	Tubb4a
Aoc3	Col4a2	Fat1	Il17rc	Myh14	Ppfibp2	Slc28a3	Tuft1
Apobr	Coro2a	Fbxo24	Il17re	Myo1e	Ppm1j	Slc2a9	Ubash3b
Apold1	Cpeb2	Fmo5	Il34	Myom3	Ppp1r13l	Slc35g2	Ubr4
Aqp3	Cpt1c	Fosb	Inf2	Myzap	Prkaa2	Slc38a3	Usp46
Aqp7	Creld1	Frmd4b	Iqck	N4bp2l1	Psapl1	Slc44a2	Xkr5
Arhgef2	Crybg2	Fyb2	Itpkb	Nbeal2	Ptpre	Slc7a15	Zbtb38
Arhgef4	Ctsd	Gclc	Itpkc	Nbl1	Ptprr	Slco2a1	Zbtb8b
Arnt2	Cyp1a1	Gjb1	Itpripl2	Ncmap	Pvr	Ssc5d	Zdhhc18
Artn	Cyp27a1	Gjb3	Kazn	Ndnf	Rab11fip5	Stbd1	Zfp462
Atp2b4	Cyp3a11	Gjb4	Kcnd1	Ndrg4	Rab42	Steap1	
Bax	D630045J12Ri	Glis3	Kcnq4	Ndufaf6	Ralgapb	Stk32c	
Bbc3	Dcaf15	Gltp	Kdm7a	Nfatc2	Rasal1	Stox2	

(Figure adapted from Supplementary Table 2 of Tran, Hanse, Habowski et al., 2020)

Table 4.2: ***Identified intestinal differentiation-associated genes with decreased DNA methylation upon DM-aKG treatment are shown.***

Gene list was identified by overlapping genes with decreased DNA methylation upon αKG treatment with intestinal differentiation gene signature.

ANTXR2
ARFGEF1
CPEB2
DKK2
DST
FEZ2
HINT3
ITFG1
RB1CC1
PALLD
PDLIM5
RABGAP1L
KRT20
RNF103
RNF19A
SEC24A
SLC1A3
SLC6A20
MTMR6
MTUS1
NR3C2
TSC22D1
TNIK
TNKS
TPD52
MAPK8
TM4SF20

(Figure adapted from Supplementary Table 3 of Tran, Hanse, Habowski et al., 2020)

Table 4.3: ***Sequences of primers used in the study.***

	Gene (m-mouse, h-human)	Forward primer	Reverse primer
qPCR	m-AXIN2	5'- GCAGCTCAGCAAAAAGGGAAAT-3'	5'- TACATGGGGAGCACTGTCTCGT-3'
qPCR	m-MYC	5'-CTCAGTGGTCTTTCCCTACCCG -3'	5'- TGTCCAACTTGGCCCTCTTGGC-3'
qPCR	m-LGR5	5'- CAAGCCATGACCTTGGCCCTG-3'	5'- TTTCCCAGGGAGTGGATTCTATT-3'
qPCR	m-ASCL2	5'-TAGTGCAGCCTGACCAAATG -3'	5'-AAGTCCTGATGCTGCAACGT -3'
qPCR	m-DKK4	5'-GTACTGGTGACCTTGCTTGGA-3'	5'-CCGTTCATCGTGAAACGCTAAG-3'
qPCR	m-KRT20	5'-TTCAGTCGTCAAAGTTTTCACCG-3'	5'-TCCTATGCGAGCCACTCA-3'
qPCR	h-WNT3	5'- CTCGCTGGCTACCCA ATTTG -3'	5' -AGG CTG TCA TCT ATG GTG GTG -3'
qPCR	h-LGR5	5'-CTC CCA GGT CTG GTG TGT TG-3'	5'- GAG GTC TAG GTA GGA GGT GAA G-3'
qPCR	h-AXIN2	5'-CAA CAC CAG GCG GAA CGA A-3'	5'-GCC CAA TAA GGA GTG TAA GGA CT-3'
qPCR	h-WNT6a	5'- GGC AGC CCC TTG GTT ATG G-3'	5'-CTC AGC CTG GCA CAA CTC G-3'
qPCR	h-CCND2	5’-TTCCCTCTGGCCATGAATTA-3’	5’- TGTAAATGCACAGCTTCTCC-3’
qPCR	h-MYC	5’- TTCGGGTAGTGGAAAACCAG-3’	5’- CAGCAGCTCGAATTTCTTCC-3’
qPCR	h-DKK4	5'-ACATGCAGAAGGAACAACTG-3'	5'-CTCCAAAAGGACTGGCTTAC-3'
qPCR	h-WNT5a	5'-ATT CTT GGT GGT CGC TAG GTA -3'	5'-CGC CTT CTC CGA TGT ACT GC-3'
qPCR	h-WNT10a	5'-GGT CAG CAC CCA ATG ACA TTC -3'	5'-TGG ATG GCG ATC TGG ATG C-3'
qPCR	h-MUC2	5'-GAG GGC AGA ACC CGA AAC C-3'	5'-GGC GAA GTT GTA GTC GCA GAG -3'
qPCR	h-KRT20	5'- GGA CGA CAC CCA GCG TTT AT -3'	5'-CGC TCC CAT AGT TCA CCG TG-3'
qPCR	h-TET1	5'-CAT CAG TCA AGA CTT TAA GCC CT -3'	5'- CGG GTG GTT TAG GTT CTG TTT -3'
MeDIP	MeDIP-hDKK4	5'-TGG CCA GTA TGA TTC ATC CT-3'	5'-AAG TTA GTT CAA AGG GCC AC-3'
ChIP	CHIP-hAXIN2	5'-CGG TTG GCG AAA GTT TGC-3'	5'-GGA CTC GGG AGC CTA AAG GT-3'
ChIP	CHIP-hMYC	5'-GCG TGG GGG AAA AGA AAA-3'	5'-GTC CAG ACC CTC GCA TTA -3'

(Figure adapted from Supplementary Table 5 of Tran, Hanse, Habowski et al., 2020)

REFERENCES

1. Torre, L. A. *et al.* Global cancer statistics, 2012. *CA. Cancer J. Clin.* **65**, 87–108 (2015).
2. Fearon, E. R. Molecular genetics of colorectal cancer. *Annu. Rev. Pathol. Mech. Dis.* **6**, 479–507 (2011).
3. Schepers, A. & Clevers, H. Wnt signaling, stem cells, and cancer of the gastrointestinal tract. *Cold Spring Harb. Perspect. Biol.* **4**, (2012).
4. Drost, J. *et al.* Sequential cancer mutations in cultured human intestinal stem cells. *Nature* **521**, 43–47 (2015).
5. Kinzler, K. W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159–170 (1996).
6. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).
7. de Sousa e Melo, F. *et al.* A distinct role for Lgr5+ stem cells in primary and metastatic colon cancer. *Nature* **543**, 676–680 (2017).
8. Medema, J. P. Targeting the Colorectal Cancer Stem Cell. *N. Engl. J. Med.* **377**, 888–890 (2017).
9. Vermeulen, L. *et al.* Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. *Nat. Cell Biol.* **12**, 468–476 (2010).
10. Pate, K. T. *et al.* Wnt signaling directs a metabolic program of glycolysis and angiogenesis in colon cancer. *EMBO J.* **33**, 1454–1473 (2014).
11. Gao, P. *et al.* C-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. *Nature* **458**, 762–765 (2009).
12. Huang, F., Zhang, Q., Ma, H., Lv, Q. & Zhang, T. Expression of glutaminase is upregulated in colorectal cancer and of clinical significance. *Int. J. Clin. Exp. Pathol.* **7**, 1093–1100 (2014).
13. Kim, M. H. & Kim, H. The roles of glutamine in the intestine and its implication in intestinal diseases. *International Journal of Molecular Sciences* **18**, (2017).
14. Le, A. *et al.* Glucose-independent glutamine metabolism via TCA cycling for proliferation and survival in b cells. *Cell Metab.* **15**, 110–121 (2012).
15. Carey, B. W., Finley, L. W. S., Cross, J. R., Allis, C. D. & Thompson, C. B. Intracellular α-ketoglutarate maintains the pluripotency of embryonic stem cells. *Nature* **518**, 413–416 (2015).
16. Pan, M. *et al.* Regional glutamine deficiency in tumors promotes dedifferentiation through inhibition of histone demethylation. *Nat. Cell Biol.* **18**, 1090–1101 (2016).
17. Tran, T. Q., Lowman, X. H. & Kong, M. Molecular pathways: Metabolic control of histone methylation and gene expression in cancer. *Clin. Cancer Res.* **23**, 4004–4009 (2017).
18. Altman, B. J., Stine, Z. E. & Dang, C. V. From Krebs to clinic: Glutamine metabolism to cancer therapy. *Nature Reviews Cancer* **16**, 619–634 (2016).
19. Denkert, C. *et al.* Metabolite profiling of human colon carcinoma - Deregulation of TCA cycle and amino acid turnover. *Mol. Cancer* **7**, (2008).

20. Schuijers, J. *et al.* Ascl2 acts as an R-spondin/wnt-responsive switch to control stemness in intestinal crypts. *Cell Stem Cell* **16**, 158–170 (2015).

21. Gregorieff, A. & Clevers, H. Wnt signaling in the intestinal epithelium: from endoderm to cancer. *Genes Dev.* **19**, 877–90 (2005).

22. Voronov, E. & Apte, R. N. IL-1 in Colon Inflammation, Colon Carcinogenesis and Invasiveness of Colon Cancer. *Cancer Microenviron.* **8**, 187–200 (2015).

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

24. Tse, J. W. T., Jenkins, L. J., Chionh, F. & Mariadason, J. M. Aberrant DNA Methylation in Colorectal Cancer: What Should We Target? *Trends in Cancer* **3**, 698–712 (2017).

25. Kim, R., Sheaffer, K. L., Choi, I., Won, K. J. & Kaestner, K. H. Epigenetic regulation of intestinal stem cells by Tet1-mediated DNA hydroxymethylation. *Genes Dev.* **30**, 2433–2442 (2016).

26. Ashktorab, H. & Brim, H. DNA Methylation and Colorectal Cancer. *Current Colorectal Cancer Reports* **10**, 425–430 (2014).

27. Morris, L. G. T. *et al.* Recurrent somatic mutation of FAT1 in multiple human cancers leads to aberrant Wnt activation. *Nat. Genet.* **45**, 253–261 (2013).

28. Baehs, S. *et al.* Dickkopf-4 is frequently down-regulated and inhibits growth of colorectal cancer cells. *Cancer Lett.* **276**, 152–159 (2009).

29. TeSlaa, T. *et al.* α-Ketoglutarate Accelerates the Initial Differentiation of Primed Human Pluripotent Stem Cells. *Cell Metab.* **24**, 485–493 (2016).

30. Salz, T. *et al.* HSETD1A regulates wnt target genes and controls tumor growth of colorectal cancer cells. *Cancer Res.* **74**, 775–786 (2014).

31. Sierra, J., Yoshida, T., Joazeiro, C. A. & Jones, K. A. The APC tumor suppressor counteracts β-catenin activation and H3K4 methylation at Wnt target genes. *Genes Dev.* **20**, 586–600 (2006).

32. Ring, A., Kim, Y. M. & Kahn, M. Wnt/Catenin Signaling in Adult Stem Cell Physiology and Disease. *Stem Cell Rev. Reports* **10**, 512–525 (2014).

33. Degos, L. & Wang, Z. Y. All trans retinoic acid in acute promyelocytic leukemia. *Oncogene* **20**, 7140–7145 (2001).

34. Vlachogiannis, G. *et al.* Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science (80-.).* **359**, 920–926 (2018).

35. Dow, L. E. *et al.* Apc Restoration Promotes Cellular Differentiation and Reestablishes Crypt Homeostasis in Colorectal Cancer. *Cell* **161**, 1539–1552 (2015).

36. De Rosa, M. *et al.* Genetics, diagnosis and management of colorectal cancer (Review). *Oncol. Rep.* **34**, 1087–1096 (2015).

37. Moser, A. R. *et al.* ApcMin: A mouse model for intestinal and mammary tumorigenesis. *Eur. J. Cancer* **31**, 1061–1064 (1995).

38. Kamphorst, J. J. *et al.* Human pancreatic cancer tumors are nutrient poor and tumor cells actively

scavenge extracellular protein. *Cancer Res.* **75**, 544–553 (2015).

39. Roberts, E. *et al.* Amino Acids in Epidermal Carcinogenesis in Mice. *Cancer Res.* **9**, (1949).

40. Miyo, M. *et al.* Metabolic Adaptation to Nutritional Stress in Human Colorectal Cancer. *Sci. Rep.* **6**, (2016).

41. Reid, M. A. *et al.* The B55α Subunit of PP2A Drives a p53-Dependent Metabolic Adaptation to Glutamine Deprivation. *Mol. Cell* **50**, 200–211 (2013).

42. Tran, T. Q. *et al.* Tumor-associated mutant p53 promotes cancer cell survival upon glutamine deprivation through p21 induction. *Oncogene* **36**, 1991–2001 (2017).

43. Yang, Y. *et al.* MiR-135 suppresses glycolysis and promotes pancreatic cancer cell adaptation to metabolic stress by targeting phosphofructokinase-1. *Nat. Commun.* **10**, (2019).

44. Lowman, X. H. *et al.* p53 Promotes Cancer Cell Adaptation to Glutamine Deprivation by Upregulating Slc7a3 to Increase Arginine Uptake. *Cell Rep.* **26**, 3051-3060.e4 (2019).

45. Carnero, A. & Lleonart, M. The hypoxic microenvironment: A determinant of cancer stem cell evolution. *BioEssays* **38**, S65–S74 (2016).

46. Sullivan, R., Paré, G. C., Frederiksen, L. J., Semenza, G. L. & Graham, C. H. Hypoxia-induced resistance to anticancer drugs is associated with decreased senescence and requires hypoxia-inducible factor-1 activity. *Mol. Cancer Ther.* **7**, 1961–1973 (2008).

47. Raa, A. *et al.* Hyperoxia retards growth and induces apoptosis and loss of glands and blood vessels in DMBA-induced rat mammary tumors. *BMC Cancer* **7**, (2007).

48. Stuhr, L. E. B. *et al.* Hyperoxia retards growth and induces apoptosis, changes in vascular density and gene expression in transplanted gliomas in nude rats. *J. Neurooncol.* **85**, 191–202 (2007).

49. Stępień, K., Ostrowski, R. P. & Matyja, E. Hyperbaric oxygen as an adjunctive therapy in treatment of malignancies, including brain tumors. *Medical Oncology* **33**, (2016).

50. Mori, Y. *et al.* Somatic mutations of the APC gene in colorectal tumors: Mutation cluster region in the APC gene. *Hum. Mol. Genet.* **1**, 229–233 (1992).

51. Voloshanenko, O. *et al.* Wnt secretion is required to maintain high levels of Wnt activity in colon cancer cells. *Nat. Commun.* **4**, (2013).

52. Schneikert, J., Grohmann, A. & Behrens, J. Truncated APC regulates the transcriptional activity of β-catenin in a cell cycle dependent manner. *Hum. Mol. Genet.* **16**, 199–209 (2007).

53. Pai, R. *et al.* Prostaglandin E2, transactivates EGF receptor: A novel mechanism for promoting colon cancer growth and gastrointestinal hypertrophy. *Nat. Med.* **8**, 289–293 (2002).

54. Janssen, K. P. *et al.* APC and Oncogenic KRAS Are Synergistic in Enhancing Wnt Signaling in Intestinal Tumor Formation and Progression. *Gastroenterology* **131**, 1096–1109 (2006).

55. Jemal, A. *et al.* Global cancer statistics. *CA. Cancer J. Clin.* **61**, 69–90 (2011).

56. Seshagiri, S. *et al.* Recurrent R-spondin fusions in colon cancer. *Nature* **488**, 660–664 (2012).

57. Muzny, D. M. *et al.* Comprehensive molecular characterization of human colon and rectal cancer.

Nature **487**, 330–337 (2012).

58. Krishnamurthy, N. & Kurzrock, R. Targeting the Wnt/beta-catenin pathway in cancer: Update on effectors and inhibitors. *Cancer Treatment Reviews* **62**, 50–60 (2018).

59. Jimeno, A. *et al.* A first-in-human phase I study of the anticancer stem cell agent ipafricept (OMP-54F28), a decoy receptor for wnt ligands, in patients with advanced solid tumors. *Clin. Cancer Res.* **23**, 7490–7497 (2017).

60. O'Rourke, K., Ackerman, S., Dow, L. & Lowe, S. Isolation, Culture, and Maintenance of Mouse Intestinal Stem Cells. *BIO-PROTOCOL* **6**, (2016).

61. van de Wetering, M. *et al.* Prospective Derivation of a Living Organoid Biobank of Colorectal Cancer Patients. *Cell* **161**, 933–945 (2015).

62. Tran, T. Q. *et al.* Glutamine deficiency induces DNA alkylation damage and sensitizes cancer cells to alkylating agents through inhibition of ALKBH enzymes. *PLoS Biol.* **15**, (2017).

63. Liu, X., Ser, Z. & Locasale, J. W. Development and quantitative evaluation of a high-resolution metabolomics technology. *Anal. Chem.* **86**, 2175–2184 (2014).

64. Tong, K. *et al.* Degree of Tissue Differentiation Dictates Susceptibility to BRAF-Driven Colorectal Cancer. *Cell Rep.* **21**, 3833–3845 (2017).

CHAPTER 5

Discussion

Unpublished work:
***Habowski, A.N.**,2020.*

The intestine contains rapidly dividing epithelial stem cells that give rise to mature cells of two lineages: absorptive and secretory. Together multiple mature cell types from these lineages populate the inner lining of the intestine to ensure continual, active absorption of water and nutrients, and maintenance of a crucial barrier layer. In the normal intestine this process is highly regulated and homeostatic proportions of the two lineages persist. When wounding events occur, either acute or chronic, there is disruption in this homeostasis. In extreme cases such as cancer, the wound never heals and homeostasis is permanently disrupted. One of the key components of this homeostasis is maintaining an appropriate balance of stemness. The processes of losing and gaining stemness in the intestine have been further investigated in this dissertation.

NORMAL INTESTINE

Our work in the normal intestine (Chapter 2) investigated transcriptomic and proteomic changes during loss of stemness. One of the most profound discoveries was the predominance of mRNA processing changes between stem cells and progenitors [1]. This indicated that during the first steps in loss of stemness, the most notable change in the transcriptome is not changes in gene expression/mRNA level, but rather, changes in mRNA isoform production via alternative splicing events (predominantly skipped exon) and alternative polyadenylation. Alternative splicing and polyadenylation events can alter the mRNA coding sequence, or can impact transcript stability, localization of protein or mRNA, as well as protein translation rate and stability [2–5]. The idea that these mRNA processing events correlate with routine loss of stemness in adult stem cells is consistent with previous work in embryonic stem cells [6–15]. However, it also suggests that adult somatic stem cells may rely on changes in mRNA processing as a rapid mechanism to facilitate loss of stemness, and importantly by inference, the reversibility of mRNA processes changes might account for plasticity of mature cells.

The mRNA processing events that correlate with rapid loss of stemness in the intestine might be on a faster timescale than gene expression changes regulated by signal transduction pathways. Previous work in embryonic stem cells has identified gene expression changes of several important mRNA processing regulators that influence embryonic stem cell differentiation [7,11–13]. However, for the estimated rapid time scale of changes in stemness in the crypt and its presumably matching rapid time scale of changes in mRNA processing, we suggest that processing events are largely independent of changes in gene expression of RNA processing regulators. Previous work by Pleiss *et al* reported just how rapid highly regulated mRNA splicing can occur [16,17]. In this yeast study, the authors showed that within 2 minutes of amino acid starvation or addition of ethanol, splicing changes were observed in transcripts [16]. And more importantly, the transcripts that were subjected to splicing changes following starvation were distinct from those altered following treatment with ethanol, underscoring the specificity of signaling and stresses that elicit distinct patterns of RNA processing in the transcriptome [16]. In other studies with plants, alternative mRNA splicing enabled rapid adaptation to environmental stress [18]. Based on these extremely rapid responses, we postulate that posttranslational modification of mRNA processing regulators might be the more relevant regulatory events responsible for changes in mRNA processing, rather than changes in the expression of RNA regulators themselves [16–18]. We are currently performing additional proteomic work to further investigate signaling relationships in the various sorted colonic crypt cell populations using targeted proteomics and phosphoproteomics. These approaches, as compared to the global proteomics we performed previously, will enable us to better identify signaling activities of several pathways of interest and, importantly, observe posttranslational protein modifications such as phosphorylation. Reversible phosphorylation of splicing regulators influences splice site selection and localization of splicing regulators and we hypothesize this impacts mRNA processing regulator function during loss of stemness [19].

The mechanism driving changes in mRNA processing in the intestine is currently unknown. If phosphorylation of mRNA processing regulators influences loss of stemness, the standing question is what facilitates this phosphorylation change? We hypothesize that two possible contributing factors are: 1) changes in cell adhesion, and 2) signal transduction alterations based on cell localization. Both of these relate to physical remodeling of cell locations at the base of the crypt that occur due to continued proliferation. As the ~dozen stem cells at the base of the crypt continue to proliferate, stem cells positioned closer to the edge of the stem cell niche are physically pushed out of the niche and trigger rapid loss of stemness. This physical migration occurs as a transient loss of cell adhesion as well as placing the migrating cell in a different microenvironment (i.e. changes in morphogens such as Wnt ligands).

Several studies have shown that changes in migration and cell adhesion are not only influenced by integrin activity but can, in turn, influence integrin outside-in signaling which has the capacity to activate several signaling transduction pathways [20–22]. Additionally, integrin signaling can alter activation of phosphatases/kinases, notably protein phosphatase 2A (PP2A) [23,24]. PP2A family phosphatases are essential for spliceosome assembly and function [25,26]. Many RNA binding proteins that influence splicing (SR proteins for example) undergo altered activity and change subcellular location in response to reversible changes in phosphorylation [27,28]. One intriguing example is SRp38, which functions as a general splicing repressor when dephosphorylated, but when phosphorylated, functions as a sequence-specific splicing activator [29]. Transcription and mRNA processing are tightly linked and phosphorylation of the C-terminal domain of RNA polymerase II influences cleavage and polyadenylation machinery recruitment that can lead to alternative mRNA processing [30]. Changes in phosphatases/kinase activity can impact mRNA splicing and polyadenylation making this a feasible mechanism to rapidly facilitate alternative mRNA processing during loss of stemness in the intestinal crypt.

WOUNDED INTESTINE

In Chapter 3, the immediate response to intestinal wounding and the initiation of recovery is investigated. We observed three dominant activities in the first few days following radiation wounding: 1) physical barrier maintenance, 2) secretion of protective peptides, and 3) increased proliferation to regain lost crypt cells. Responses are rapid because in only four days post-irradiation, biomarkers of cell death are decreasing and single cell RNA-sequencing shows a strong increase in proliferation in TA and absorptive cell populations. Paired with histological analysis, this suggests it is these two populations that might be responsible for most of the repopulation of the damaged crypt base. Interestingly, within this four-day rapid response time frame, we do not see a return of any cells that express biomarkers of stem cells or increases in Wnt signaling. The lack of such signatures seems to indicate that the actual return of a stemness transcriptome is delayed.

Although not previously reported, our data demonstrate that in addition to a loss of stem cells, DCS cells are also quickly eliminated during irradiation. This would suggest that the entire crypt base structure needs to be reformed since it is stem cells and DCS that form the entirety of the stem cell niche. We propose that stem cells arise from the cells that have migrated down to the damage crypt base to repopulate it. Exposure of these “newly arrived” cells to morphogens from the surrounding microenvironment might serve to recreate the stem cell niche. Although we observed a prominent increase in the prevalence of proliferative uncommitted TA and immature absorptive cells, our hypothesis that these populations repopulate the stem cell niche would not preclude the possibility that differentiated cell types might also occupy spaces at the base of the crypt. The phenomenal plasticity of intestinal crypt epithelial cells might be a general property that enables any epithelial cell to de-differentiate into stem cells if given the appropriate environmental cues. If this is true, then the relative abundance of the different cell types closer to the crypt base would dictate which populations are the most likely to restore the stem cell niche. Fundamental to evaluation of this hypothesis will be including an investigation of the intestinal stroma responses to wounding and the microenvironment that may be responsible for driving re-expression of a stem cell transcriptome.

We and others have observed that the small intestine is more sensitive to irradiation than the colon [31,32]. It is suggested that the higher levels of expression of the anti-apoptotic regulator Bcl2 in the colon might counteract p53-dependent apoptosis signaling relative to the small intestine, promoting cell-cycle arrest instead and continued existence of cells containing potentially harmful DNA mutations [33,34]. Without high levels of Bcl2, the small intestine, would exhibit signs of more severe wounding, and cell death, preventing the persistence of cells with mutated DNA[33,34]. Consistent with this, small intestinal cancers are extremely rare in patients, and colorectal cancer all too common. The distinction between the small intestine and colon responses to wounding, and potentially the processes of recovery, could lie at the crux of the explanation for higher rates of colorectal cancer.

COLON CANCER

Through our work on colon cancer and the impact of the cellular environment we found the metabolite α-ketoglutarate (downstream of glutamine) promotes differentiation and, consistently, low glutamine levels promote stemness (Chapter 4) [35]. This finding has important therapeutic implications and raises questions over the role of stemness in cancer promotion. Cancer cells are often considered to be glutamine addicted [36], in part due to a reliance on Warburg metabolism and the shunting of glucose for cellular building blocks rather than an optimal amount of energy [37–39]. This results in cancer cells using glutamine as an alternate energy source [36]. Naturally this idea encouraged researchers to develop therapies that target the cancer cell's use of glutamine, in particular the development of glutaminase inhibitors which inhibit the conversion of glutamine to glutamate (downstream of glutamate is α-ketoglutarate) [36,40]. Although inhibition of glutaminase (i.e. low glutamine) might temporarily decrease tumor growth by removing an important nutrient, our work suggests that this lowered glutamine level would actually promote stemness in the remaining cancer cells [35]. Additional evidence for this concern was observed in melanoma studies which observed that glutamine supplementation, which increased downstream levels of α-ketoglutarate, inhibited melanoma growth, prolonged mouse survival, and increased sensitivity to other therapies [41]. This raises questions about the efficacy of targeting glutamine metabolism and the therapeutic usefulness of glutaminase inhibitors.

There are currently 18 cancer clinical trials in various phases of testing using glutaminase inhibitors (*clinicaltrials.gov*). Four of these trials are complete, but no data has been posted yet. Many of these trials are for combination therapies in solid tumors, including colorectal cancer. Several are focused on subsets of patients with specific mutations (IDH, KEAP/NRF2, EGFR, or TNBC breast cancer) whereas others focus on subsets without certain mutations (RAS or BRCA wildtype). Nearly all trials use CB-839 (Telaglenastat), developed by Calithera Bioscience, Inc. CB-839 is a selective, reversible oral inhibitor of glutaminase and plasmid concentration of >300 nM results in >90% glutaminase inhibition [42]. Glutaminase inhibition reduces cell growth, eventually leading to apoptosis [42]. However, with CB-839 treatment, glutaminase is not entirely shutdown allowing for very low levels of glutamate to persist in cancer cells. We, and others, have shown that this can promote stemness in cancer cells [35,41].

The primary reason that traditional cancer therapies are not effective is the recurrence or persistence of therapy-resistant cancer cells, and in a worst-case scenario, metastases that emerge after treatment. Cancer stem cells are thought to be the therapy-resilient cells that cause relapse and metastasis, and thus finding methods to target them is a high priority [43]. Recently, some methods of targeting cancer stem cells have moved into the clinic [44]. One class of therapies are antibodies to cancer stem cell associated surface markers such as Rituximab (CD20), Alemtuzumab (CD52), Bivatuzumab (CD44v6), Adecatumumab/Catumaxomab (Epcam), among many more in clinical trials (targeting: CD3, CD4, CD123, CD124, CD47, CD33) [44]. Another class of therapies is to target cancer stem cell associated signaling pathways including Notch (clinical trials with ɣ-secretase inhibitors and antibodies to DLL4), TGF-β, PI3K, Wnt, and others [44]. Results with signaling pathway inhibitors show variation from tissue to tissue and have had mixed success [44]. Other therapeutic ideas include targeting the cancer stem cell microenvironment or using cancer stem cell directed immunotherapy. Alternatively, rather than targeting stem cells directly, other approaches promote stem cell differentiation, which could increase susceptibility to other therapies [43]. For example, BMP4 can promote differentiation of colorectal cancer stem cells and increase efficacy of 5-FU and oxaliplatin [45]. This notion of combination therapies might be more aligned with the general property of cellular plasticity that we and others have discovered. Single agent targeting might only push cells to adjust and convert to another proliferative and equally lethal cancer cell state. Ultimately, further research needs to be carried out to better understand cancer stem cells, and cancer stemness broadly, in order to find optimal targets that will not interrupt function of normal adult stem cells.

INTESTINAL STEMNESS

It is important to understand and define stemness and appropriate loss of stemness in normal tissue as well as acute wounding, chronic wounding, and cancer so that a complete picture of homeostasis is understood, and aberrant activity is clearly defined. Identification of the signaling pathways that enable cells to survive stresses, that instruct cells to ramp up their proliferative index and that direct cellular plasticity can lead to more clearly defining therapeutic windows, novel potential targets, and the feasibility of targeting stemness or differentiation. The intestinal epithelium is a phenomenal, intricate, and rapidly proliferating tissue. It extends logically that this could lead to complex disease. However, the essential and crucial function of the intestine necessitates a deep dive into the processes that enable normal function, with the hopes of uncovering how to promote this function once again in disease states.

REFERNCES

1. Habowski, A. N. *et al.* Transcriptomic and proteomic signatures of stemness and differentiation in the colon crypt. *Commun. Biol.* **3**, 1–17 (2020).
2. Mayr, C. What are 3′ UTRs doing? *Cold Spring Harb. Perspect. Biol.* **11**, (2019).
3. Berkovits, B. D. & Mayr, C. Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. *Nature* **522**, 363–367 (2015).
4. Wang, Y. *et al.* Mechanism of alternative splicing and its regulation. *Biomed. Reports* **3**, 152–158 (2015).
5. Yang, X. *et al.* Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell* **164**, 805–817 (2016).
6. Salomonis, N. *et al.* Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10514–9 (2010).
7. Yamazaki, T. *et al.* TCF3 alternative splicing controlled by hnRNP H/F regulates E-cadherin expression and hESC pluripotency. *Genes Dev.* **32**, 1161–1174 (2018).
8. Yeo, G. W. *et al.* Alternative Splicing Events Identified in Human Embryonic Stem Cells and Neural Progenitors. *PLoS Comput. Biol.* **3**, e196 (2007).
9. Salomonis, N. *et al.* Alternative Splicing in the Differentiation of Human Embryonic Stem Cells into Cardiac Precursors. *PLoS Comput. Biol.* **5**, e1000553 (2009).
10. Gabut, M. *et al.* An Alternative Splicing Switch Regulates Embryonic Stem Cell Pluripotency and Reprogramming. *Cell* **147**, 132–146 (2011).
11. Lackford, B. *et al.* Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO J.* **33**, 878–889 (2014).
12. Grassi, E. *et al.* Choice of alternative polyadenylation sites, mediated by the rna-binding protein Elavl3, plays a role in differentiation of inhibitory neuronal progenitors. *Front. Cell. Neurosci.* **12**, (2019).
13. de Morree, A. *et al.* Alternative polyadenylation of Pax3 controls muscle stem cell fate and muscle function. *Science (80-.).* **366**, 734–738 (2019).
14. Ji, Z., Lee, J. Y., Pan, Z., Jiang, B. & Tian, B. Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7028–7033 (2009).
15. Fiszbein, A. & Kornblihtt, A. R. Alternative splicing switches: Important players in cell differentiation. *BioEssays* **39**, 1600157 (2017).
16. Pleiss, J. A., Whitworth, G. B., Bergkessel, M. & Guthrie, C. Rapid, Transcript-Specific Changes in Splicing in Response to Environmental Stress. *Mol. Cell* **27**, 928–937 (2007).
17. Yang, L., Park, J. & Graveley, B. R. Splicing from the Outside In. *Molecular Cell* **27**, 861–862 (2007).
18. Laloum, T., Martín, G. & Duque, P. Alternative Splicing Control of Abiotic Stress Responses. *Trends in Plant Science* **23**, 140–150 (2018).

19. Stamm, S. Regulation of alternative splicing by reversible protein phosphorylation. *Journal of Biological Chemistry* **283**, 1223–1227 (2008).

20. Harburger, D. S. & Calderwood, D. A. Integrin signalling at a glance. *J. Cell Sci.* **122**, 159–163 (2009).

21. Kechagia, J. Z., Ivaska, J. & Roca-Cusachs, P. Integrins as biomechanical sensors of the microenvironment. *Nature Reviews Molecular Cell Biology* **20**, 457–473 (2019).

22. Sun, Z., Costell, M. & Fässler, R. Integrin activation by talin, kindlin and mechanical forces. *Nat. Cell Biol.* **21**, (2019).

23. Ivaska, J. *et al.* Integrin α2β1 Promotes Activation of Protein Phosphatase 2A and Dephosphorylation of Akt and Glycogen Synthase Kinase 3β. *Mol. Cell. Biol.* **22**, 1352–1359 (2002).

24. Zheng, M. & McKeown-Longo, P. J. Cell adhesion regulates Ser/Thr phosphorylation and proteasomal degradation of HEF1. *J. Cell Sci.* **119**, 96–103 (2006).

25. Shi, Y., Reddy, B. & Manley, J. L. PP1/PP2A Phosphatases Are Required for the Second Step of Pre-mRNA Splicing and Target Specific snRNP Proteins. *Mol. Cell* **23**, 819–829

26. Kamoun, M., Filali, M., Murray, M. V., Awasthi, S. & Wadzinski, B. E. Protein phosphatase 2A family members (PP2A and PP6) associate with U1 snRNP and the spliceosome during pre-mRNA splicing. *Biochem. Biophys. Res. Commun.* **440**, 306–311 (2013).

27. Naro, C. & Sette, C. Phosphorylation-mediated regulation of alternative splicing in cancer. *International Journal of Cell Biology* **2013**, 15 (2013).

28. Virshup, D. M. & Shenolikar, S. Molecular Cell Review From Promiscuity to Precision: Protein Phosphatases Get a Makeover. doi:10.1016/j.molcel.2009.02.015

29. Feng, Y., Chen, M. & Manley, J. L. Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator. *Nat. Struct. Mol. Biol.* **15**, 1040–1048 (2008).

30. Davidson, L., Muniz, L. & West, S. 3’ end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. *Genes Dev.* **28**, 342–356 (2014).

31. Gândara, R. M. C., Mahida, Y. R. & Potten, C. S. Regional Differences in Stem and Transit Cell Proliferation and Apoptosis in the Terminal Ileum and Colon of Mice After 12 Gy Radiation Oncology. *Radiat. Oncol. Biol.* **82**, e521–e528 (2012).

32. Potten, C. & Grant, H. The relationship between ionizing radiation-induced apoptosis and stem cells in the small and large intestine. *Br. J. Cancer* **78**, 993–1003 (1998).

33. Pritchard, D. M., Potten, C. S., Korsmeyer, S. J., Roberts, S. & Hickman, J. A. Damage-induced apoptosis in intestinal epithelia from bcl-2-null and bax-null mice: Investigations of the mechanistic determinants of epithelial apoptosis in vivo. *Oncogene* **18**, 7287–7293 (1999).

34. Potten, C. S. Radiation, the Ideal Cytotoxic Agent for Studying the Cell Biology of Tissues such as the Small Intestine. in *Radiation Research* **161**, 123–136 (Radiat Res, 2004).

35. Tran, T. Q. *et al.* α-Ketoglutarate attenuates Wnt signaling and drives differentiation in colorectal

cancer. *Nat. Cancer* **1**, 345–358 (2020).

36. Altman, B. J., Stine, Z. E. & Dang, C. V. From Krebs to clinic: Glutamine metabolism to cancer therapy. *Nature Reviews Cancer* **16**, 619–634 (2016).

37. Sprowl-Tanio, S. *et al.* Lactate/pyruvate transporter MCT-1 is a direct Wnt target that confers sensitivity to 3-bromopyruvate in colon cancer. *Cancer Metab.* **4**, 20 (2016).

38. Pate, K. T. *et al.* Wnt signaling directs a metabolic program of glycolysis and angiogenesis in colon cancer. *EMBO J.* **33**, 1454–1473 (2014).

39. Liberti, M. V. & Locasale, J. W. The Warburg Effect: How Does it Benefit Cancer Cells? *Trends in Biochemical Sciences* **41**, 211–218 (2016).

40. Huang, F., Zhang, Q., Ma, H., Lv, Q. & Zhang, T. Expression of glutaminase is upregulated in colorectal cancer and of clinical significance. *Int. J. Clin. Exp. Pathol.* **7**, 1093–1100 (2014).

41. Ishak Gabra, M. B. *et al.* Dietary glutamine supplementation suppresses epigenetically-activated oncogenic pathways to inhibit melanoma tumour growth. *Nat. Commun.* **11**, 1–15 (2020).

42. Guerra, V. A. *et al.* Interim Analysis of a Phase II Study of the Glutaminase Inhibitor Telaglenastat (CB-839) in Combination with Azacitidine in Advanced Myelodysplastic Syndrome (MDS). in *American Society of Hematology* (2019).

43. Turdo, A. *et al.* Meeting the challenge of targeting cancer stem cells. *Front. Cell Dev. Biol.* **7**, 16 (2019).

44. Yang, L. *et al.* Targeting cancer stem cell pathways for cancer therapy. *Signal Transduction and Targeted Therapy* **5**, 1–35 (2020).

45. Lombardo, Y. *et al.* Bone morphogenetic protein 4 induces differentiation of colorectal cancer stem cells and increases their response to chemotherapy in mice. *Gastroenterology* **140**, 297-309.e6 (2011).

APPENDIX A

Isolation of murine large intestinal crypt cell populations with flow sorting

Text adapted from published work:

Habowski, A.N, J.M. Bates, J.L. Flesher, R.A. Edwards, and M.L. Waterman. (2020) Isolation of murine large intestinal crypt cell populations with flow sorting. *Nature Protocol Exchange*. DOI: 10.21203/rs.3.pex-994/v1.

ABSTRACT

Here we present a high-resolution sorting protocol for colon crypt stem cells, their daughter cells and mature, differentiated cell types. We used freshly dissected mouse colons and validated intestinal cell surface markers amenable to Flow Activated Cell Sorting (FACS). This 5-7 hour protocol enables isolation of six distinct cryptal cell populations (Stem, AbsPro, SecPDG, Tuft, Ent, and EEC) from any mouse strain/background (Figure A.1 and A.2) . Downstream analysis of sorted cells (Transcriptomics = bulk RNA-seq and Proteomics = small cell number LC/MS) validated the identity of cell populations. An important strength of this protocol is the independence from any trans-genic labeling of cell types and the flexibility for users to add additional markers for a variety of downstream applications. The absence of proteases during dissociation increases antigen expression resolving the six cell types but also decreases cell yield (Figure A.3). The main steps of this protocol include: Tissue Dissection, Tissue Dissociation, Preparation for FACS, and Performing FACS.

INTRODUCTION

Multiple cell sorting protocols have been optimized to isolate intestinal stem cells, but each lack the resolution to purify daughter cells and differentiated progeny populations [1,2]. For example, the transgenic stem cell lineage marker Lgr5-EGFP enables purification of GFP-bright stem cells, but a mosaic expression pattern of the transgene in the intestine has made it difficult to confidently separate daughter cells from GFP-negative stem cells and differentiated cells [3,4]. Single cell RNA sequencing (scRNA-seq) captures cellular diversity when analyzing mixed cell populations and has been useful for defining intestinal lineage trajectories and diversity of mature cells (for example enterocytes and enteroendocrine cells) [5–9]. However, the low sequencing depth of scRNA-seq misses moderate-to-lowly expressed transcripts and is not compatible with other downstream analysis or applications.

The sorting protocol detailed here enables sorting of colonic crypt cell populations from the large intestine of the mouse, independent of transgene markers. This protocol is compatible with a variety of downstream applications including bulk RNA-seq and mass spectrometry. Importantly, our analysis has validated the identity of the isolated populations, enabling others to use this protocol for FACS analysis of their intestinal system to chart changes in crypt dynamics and populations. For FACS analysis <1/3 mouse colon is more than sufficient for a snapshot of crypt populations, although for sorting, several mice may need to be pooled depending on the downstream application. This protocol is also compatible with additional antibody markers or mice of any strain/gender (including transgenic mice – with compatible fluorophore). We recommend additional markers use FITC-EGFP channel.

REAGENTS AND EQUIPMENT

Reagents:

Rock inhibitor (Y-27632 AdipoGen Life Sciences from Fisher #501146540)

DNAse (Sigma-Aldrich #4716728001)

CD45-BV510 (1:200, Clone 30-F11; BD Biosciences #563891)

CD31-BV510 (1:200, Clone MEC 13.3; BD Biosciences #563089)

CD326-eFluor450 (1:100, Clone G8.8; eBioscience #48-5791-82)

CD44-PerCP-Cy5.5 (1:100, Clone IM7; Thermo Fisher #A26013)

CD24-PECy7 (1:200, Clone M1/69; eBioscience #25-0242-82)

CD117-APC-Cy7 (1:100, Clone 2B8; Thermo Fisher #A15423)

Live/Dead Aqua (Thermo Fisher # L34957)

Other standard lab reagents:

70% Ethanol

PBS (Chilled; 1.5 L)

EDTA (0.4 M)

FBS

Optional Reagents:

TRIzol (if sorting for downstream RNA isolation; cell sort collection solution)

100 mM ammonium bicarbonate (if sorting for downstream mass spectrometry)(3+ L, depending on number of sorts, needed for instrument sheath fluid and cell sort collection solution)

Plasticware:

Petri dishes (4)

10 mL syringe (1)

50 mL conical tubes (21)

1.5 mL Eppendorf tubes (10 +)

FACS tubes (4 +)

100 µm and 40 µm filters (that fit on top of 50 mL conical tube) (4 + of each)

optional = Collection tubes (depending on downstream applications – FACS Tubes, Eppendorf tubes, PCR tubes, ect.)

optional = FACS tubes with 40 µm filter cap (and/or separate cap filters)

Equipment:

Pipets

Vortex

Ice buckets

Centrifuge for conical tubes and for Eppendorf tubes at 4 °C

Rotator at 4 °C (*recommended* Fisherbrand™ Multi-Purpose Tube Rotators #88-861-049)

Dissection Tools:

Dissecting tray with pins

Scissors

Forceps

Blunt point popper needle [attached to 10 mL syringe] (*recommended* ~16-gauge, long tip ~3-4 inch, threaded – Luer lock termination)

PROCEDURE

Preparation

1. Identify mice to be used. This protocol is designed for 4 mice to be simultaneously processed with each mouse serving as individual biological samples. Note: 4 separate samples are the maximum recommended amount for one person.
2. Prepare the following solutions (can be done the night before):
    a. **Dissociation solution**: 8 x 20 mL (in 50 mL conical tubes) chilled dissociation solution.
        i. Make stock
            1. 160 mL PBS
            2. 160 µL of 1000x Rock Inhibitor (Final concentration is 10 µM)
            3. 800 µL of 0.4 M EDTA (Final concentration is 2mM)
        ii. Aliquot 20 mL each into 8 conical tubes (Dissociation #1 and #2 for each sample; recommended labels 'Mouse #1 Dissociation #1' = 'M#1 Dis #1' and 'Mouse #1 Dissociation #2' = 'M#1 Dis #2'), keep chilled at 4°C if overnight, and on ice for immediate use.
    b. **FACS Buffer** (40 mL total)
        i. 38.8 mL PBS
        ii. 1.2 mL FBS (Final concentration is 3%)

        iii. 40 µL of 1000x Rock Inhibitor (Final concentration is 10 µM)

3. Label 50 mL conical tubes – 4 tubes per mouse.
    a. Recommended label for Mouse #1: M#1 – 100f #1; M#1 – 100f #2; M#1 – 40f #1; M#1 – 40f #2 (100f = 100 µm filter; 40f = 40 µm filter).
4. Label 4 petri dishes with mouse ID/#, pour half full with cold PBS, keep on ice.
5. Sacrifice mice and immediate proceed with dissection.

Tissue Dissection

6. Pin mouse on dissection tray, spray abdomen with 70% ethanol, and make incision to open peritoneal cavity. Pin skin back as needed.
7. Gently scoop out the small intestine to the left side. Cut through the pelvis just to the right of the rectum, then find the rectum and cut at the end near the junction with the skin. Avoid cutting blood vessels.
8. Cut the small intestine where it meets the cecum. Gently tug the top of the colon and if cut correctly at the rectum, it will slowly pull free with minimal mesentery tissue.
9. Place the colon in the chilled PBS in the petri dish and keep on ice. Proceed with dissecting remaining mice.
10. Linearize each section of colon – keeping it in PBS on ice:
    a. Snip off the tip of the cecum pouch.
    b. Fill the 10 mL syringe attached to a popper needle with cold PBS and gently flush out the cecum to remove fecal matter.
    c. Refill the syringe as needed and continue to expel PBS while threading the popper needle gently through the colon. Expel all remaining PBS once entire colon is threaded on the popper needle.

    d. Prop the plunger end of the syringe with the popper needle/tissue attached in a firm and stable place (for example wedged against the edge of an ice bucket) with popper needle facing you. Place one point of dissecting scissors in the opening of the popper needle – the syringe and popper needle should stay firmly in place while you hold the scissors. With your free hand, use forceps to pull the intestine off the popper needle, towards the scissors, cutting and linearizing as you pull the intestine towards you.

11. Keep linearized colon in PBS, while proceeding with other colons.

Tissue Dissociation

12. Rinse colon by swirling in clean PBS until a majority of fecal material is removed.
13. Place clean linearized colons in dissociation solution #1 (as described in preparation step 2) and place at a slow rotation (If using recommended Fisherbrand Multi-Purpose Tube Rotator, speed setting of 8 rpm with a horizontal axis of rotation ensuring solution and intestines are rotated from the top of the conical tube to the bottom). Dissociate for 30 min at 4 °C.
14. Remove colon tissue from dissociation solution #1. Cut tissue into small ~3-5 mm pieces, using forceps to dangle the tissue above dissociation tube #2 with edge of tissue resting on tube rim to pull taut and cutting with scissors. Make sure all pieces are immersed in the dissociation solution #2. Place back on rotator at 4 °C for an additional 30 min rotation.
15. Turn on centrifuge (swinging bucket preferred) during this period to allow time to chill to 4°C.
    a. Adjust settings: 500 xg, 5 min, 4 °C, decrease deceleration speed to low setting.

16. After the 30 min dissociation step is completed (1 hr in total), collect conical tubes and shake for 3 minutes *very aggressively* and rapidly (up and down motion).
    a. Solution should be quickly become cloudy with an observable abundance of floating cells.
17. Pour suspension through 100 µm filter into new conical tube (100 µm filter #1 tube – example label M#1 – 100f #1). Rinse filter with cold PBS (final volume of 40-50 mL). Collect tissue chunks trapped in the filter and place back in the dissociation tube. Re-use this filter for the next 100 µm filter for this sample (move the filter to the conical tube labeled 100 µm filter #2).
18. Add ~20 mL of PBS to the dissociation tube that contains tissue chunks and store on ice. It is important to not allow the tissue chunks to get dry.
19. Repeat steps 17-18 for all samples and then immediately spin down all 100 µm filter #1 tubes at 500 xg, 5 min, 4 °C, with decreased deceleration speed.
    a. Note: all 4 collection tubes will need to be processed rapidly in sequence so that centrifugation steps are done together. The best rates of cell survival depend on minimizing the time cell suspensions are sitting and ensuring that when they are sitting it is always on ice.
20. Manually shake tissue in the dissociation solution tubes (now with 20 mL PBS) again for 3 minutes rapidly and *aggressively*.
21. Filter through 100 µm filter into a new tube (100 µm filter #2 tube) using the same 100 µm filter used for each sample previously. Rinse with PBS for a final volume of ~40-50 mL. Tissue chunks trapped in filter can be saved (as back-up) or discarded. Spin down 100 µm filter #2 tube (and collect 100 µm filter #1 tubes from centrifuge).

22. Gently pour off supernatant from the 100 µm filter #1 tube and resuspend in 1 mL of FACS buffer. Mix well with a pipet to achieve a homogenous suspension. Filter the suspension through a 40 µm filter into a new tube (40 µm filter #1 tube). Rinse the filter with PBS filling to 50 mL. Repeat for all samples and save the 40 µm filters for the next step.
23. Repeat step #22 using the suspension from the second shake (100 µm filter #2 tubes) and filter into 40 µm filter #2 tube using the same 40 µm filter saved from step #22 for each sample.
24. Centrifuge the suspension in both 40 µm filter tubes (#1 and #2) at 500 xg, 5 min, 4 °C, with a decreased deceleration speed to protect cell viability.
25. During the centrifugation step, begin a prep of DNAse in a 1.5 mL Eppendorf tube.
    a. 20 µL of 10x buffer (contains MgCl) + 40 µL DNAse I + 140 µL Water.
26. Gently pour off supernatant from the 40 µm filter tubes and resuspend each cell pellet in ~1 mL of FACS buffer, making sure to mix and suspend the cells very well. Transfer the suspension into 1.5 mL Eppendorf tubes (if you end up with a larger volume, it is okay to use additional tubes).

Preparation for FACS

27. Centrifuge at 1,000 xg, 5-10 min, 4 °C, with a soft stop setting on to maintain cell viability. As with all centrifuge steps, repeat if pellet is not good.
    a. The initial centrifugation, when cell suspensions are well mixed, usually needs a full 10 min, subsequent centrifugations require only 5 minutes.
28. Carefully remove and discard supernatant from all tubes using a pipet. Dispense 500 µL of FACS buffer to each tube – if more than one tube was collected per sample, merge the contents of these into one tube with 500 µL FACS buffer total. Add 50 µL of DNAse and

mix. Mechanically mix the DNAse and cell suspension up and down 5-10 times with a P-1000 pipet. Incubate at room temperature for 5 min.

29. Collect the cells by centrifugation, and then remove supernatant carefully. Resuspend cell pellets in a total of ~500 µL of FACS buffer.
30. Add antibodies to the cell suspensions as listed below:
    a. Amount per tube (or recipe for master mix below)

i.	CD117 (cKit) – APC-Cy7	[1:100]	5 µL
ii.	CD326 (Epcam)– eFluor450	[1:100]	5 µL
iii.	CD44 – PerCP-Cy5.5	[1:100]	5 µL
iv.	CD24 – PECy7	[1:200]	2.5 µL
v.	CD31 – BV510	[1:200]	2.5 µL
vi.	CD45 – BV510	[1:200]	2.5 µL

    b. Antibody mastermix (for 4 mice/4 tubes) (22.5 µL per tube)
        i. 20 µL of cKit, Epcam and CD44
        ii. 10 µL of CD24, CD31, CD45
    c. Note: Add single channel antibody controls as needed to establish compensations and sorting gates. Once these gates are established, the protocol is very consistent and these controls are not needed for every sort.
31. Vortex briefly and then incubate for 30 min at 4 °C in the dark.
32. Collect cells by centrifugation (1,000 xg, 5-10 min, 4 °C, with a soft stop setting) and carefully remove supernatant. Add 1 mL fresh FACS buffer, resuspend, and centrifuge again. Remove supernatant (Wash step).
33. Resuspend each suspension thoroughly in ~500 µL - 1mL fresh FACS buffer into 5 mL round bottom tube for sorting (label tube as ‘To Sort’). *NOTE:* It is recommended to start with a

lower resuspension volume for a more concentrated cell suspension enabling faster sorting. Adjust total volume depending on mouse number/condition and cell number since the yield and recommended resuspension volume depends on the quality of the prep.

34. Add Live/Dead Aqua dye (maintain 1 µL Live/Dead per 1 mL of cell suspension), wrap in foil and incubate at room temperature for ~2 min. Incubate at least 5-10 minutes before running samples for FACS but keep samples on ice (it is not necessary to incubate for longer). *NOTE:* The Live/Dead Aqua dye is only good for 2 weeks once reconstituted.
35. Store all samples on ice, protected from light, until time to sort.

Performing FACS

36. If only analysis is being performed, no preparation for sorting is needed. If sorting is occurring, prepare tubes/reagents for collection.

    For sorting followed by *RNA isolation* we recommend sorting directly into TRIzol – a step that preserves RNA integrity. Collection tubes can be FACS tubes or Eppendorf tubes. For sorting followed by *mass spectrometry* we changed the sorting machine sheath fluid to 100 mM ammonium bicarbonate and sorted directly into 50 µL of 100 mM ammonium bicarbonate – a step that preserves protein integrity and prevents salt contaminants. We recommend sorting into PCR tubes (that can be fitted inside of Eppendorf tubes) – although this depends on the set-up of downstream mass spectrometry equipment.

    Regardless of downstream applications store collection tubes on ice/chilled prior to, during, and after sorting.
37. Populations to collect (live cells):
    a. Stem Cells

       (CD45-, CD31-, CD326+, CD44highest, CD24-, CD117-)

    b. Absorptive Progenitor [AbsPro]

       (CD45-, CD31-, CD326+, CD44med/+, CD24-, CD117-)

    c. Secretory Progenitor + Deep Crypt Secretory + Goblet(minor) [SecPDG]

       (CD45-, CD31-, CD326+, CD44 high, CD24med, CD117med)

    d. Tuft Cells

       (CD45-, CD31-, CD326+, CD24+, CD117+)

    e. Enterocytes [ENT]

       (CD45-, CD31-, CD326+, CD44-/low, CD24-)

    f. Enteroendocrine Cells [EEC]

       (CD45-, CD31-, CD326+, CD44-/low, CD24+)

38. Follow the gating schema provided in Figure A.4.
39. Recommended sorting parameters:
    a. BD FACS Aria Fusion using a 100 µm nozzle (20 PSI).
    b. Flow rate of 2.0 with a maximum threshold of 5000 events/sec.
    c. Keep sample chamber and collection tubes at 4 °C.

TROUBLESHOOTING

Common Problems

1. *Small pellets during first centrifugation steps (100 µm filter).*

   Solution: This is likely caused by not being aggressive enough with the manual shaking steps. To test how aggressive and effective the shaking is an additional (3rd) round of shaking the tissue in PBS can inform on whether additional cells are recovered. If the cell pellet is much larger than the first two pellets, there is a clear need to shake harder starting in the beginning

2. *Large/decent pellets during 100 µm filter, but small after 40 µm filter.*

   Solution: This is likely due to cells not being released into a true single cell suspension. To address this, one can shake for longer and/or more aggressively during the initial shaking or add some additional gentle shaking of the PBS resuspension prior to using the 40 µm filter. Alternatively, if the problem persists, do not use the 40 µm filter following the 100 µm filter step. Instead, continue with the protocol – including the important DNAse treatment step – and prior to adding Live/Dead stain, filter the cell suspension into a FACS tube with a 40 µm filter cap (depending on the single cell suspension several filter cap might be needed – this can be painstakingly slow but will improve yield).

3. *During Eppendorf centrifugation the cell pellet is poor and a bit fluffy.*

   Solution: If the pellet is small and/or of poor quality always repeat the centrifugation step. Gently flick/vortex the tube to release the pellet and centrifuge again. If possible, a swinging bucket centrifuge can improve quality of pellets.

4. *Low cell viability.*

   Solution: Because of the long duration of this protocol and the fragility of mature epithelial cell types there is an innately low cell viability. Some important things to implement to improve viability include keeping cells on ice/chilled at all times unless protocol specifies otherwise. Additionally, working quickly (immediately after sacrifice) and smoothly during the dissection and linearization and ensuring that the tissue does not dry out are important features of the protocol. Maximally active Rock inhibitor and high quality FBS in the FACS buffer increases viability.

5. *Total cell yield during sorts is very low.*

   Solution: This protocol is not designed for an optimized yield of cells, but rather for a high-quality separation that can distinguish cell populations. Please see section below on improving yield if this feature is specifically important for downstream applications.

6. *During the FACS procedure, cells are clumpy, clog the machine, or do not run at a constant flow rate*.

   Solution: Thoroughly mix the sample with a pipet and/or gently vortex. Dilute the sample with FACS buffer, and pass through a 40 µm filter cap. Since extracellular, extruded DNA from lysed cells is a major reason for cell clumping, make sure the DNAse concentration and treatment time is sufficient for the number of cells. In addition, also make sure the FACS buffer contains FBS to help prevent clumping.

Additional Steps to Improve Cell Yield:

1. Instead of using 100 µm filter followed by 40 µm filter during the initial centrifugation steps, use only the 100 µm filter (or a 70 µm filter instead). Immediately prior to adding the Live/Dead stain, filter the cell suspension into a FACS tube with a 40 µm filter cap (depending on the volume and density of the single cell suspension you might need several caps – this can be painstakingly slow. It will nevertheless improve yield).
2. For all filtering steps (50 mL conical) swirl the pipet tip along the filter and pipet up and down, to help solution pass through. Be sure to add additional PBS/FACS buffer to rinse the filter which will collect additional cells.
3. IMPORTANT: Pre-wet pipet tip with FACS buffer before resuspending any cell solution to prevent cellular adherence to the walls of the tip.
4. Use low-binding pipet tips if available.

RESULTS

Time Taken:

Duration depends on number of mice, below are estimates based on 4 mice (maximum number of mice recommended for one person at a time).

1. Tissue Dissection = 30-1 hr (including sacrificing and linearizing intestine)
    a. With experience, dissection and linearization should take <5 min per mouse.
2. Tissue Dissociation = ~2 hrs
3. Preparation for FACS = ~1.5 hrs
4. Performing FACS = ~ 2 hrs (1-3 hrs)
    a. This is highly dependent of the quality of the prep, number of mice, and intended downstream use (analysis or sorting).

Anticipated Outcome:

If performed correctly, a FACS plots similar to Figure A.4 and A.3b will be observed. Depending on the mice/age/quality of the preparation 5-10,000 Epcam+ cells can be sorted.

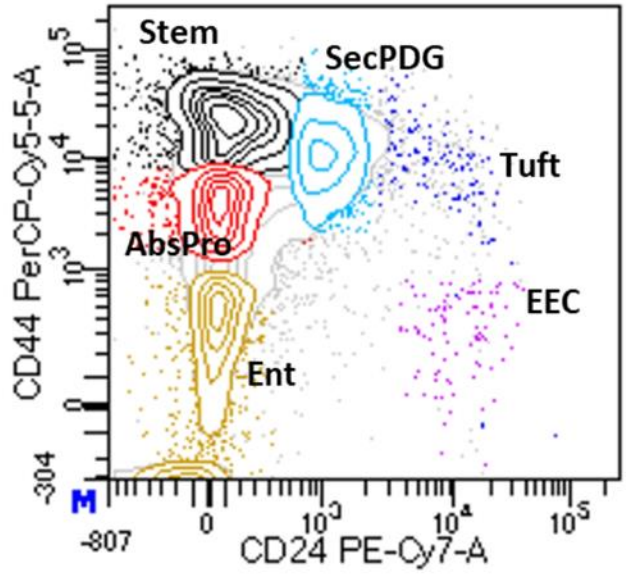


Figure A.1: ***FACS plot of six distinct colonic crypt cell populations based on Cd44 and Cd24 maker expression.***

This is a representative image of C57BL/6N male mouse aged 5-7 weeks. Cell populations are Stem, AbsPro (absorptive progenitors), SecPDG (secretory progenitors + deep crypt secretory cells + minor amount of goblet cells), Tuft, Ent (enterocytes), and EEC (enteroendocrine cells). Goblet cells do not survive this procedure well, only a small signature is observed in SecPDG. The EEC population is dominated by enterochromaffin cells.

(Figure adapted from Fig. 1 of Habowski et al., 2020; Protocol Exchange)

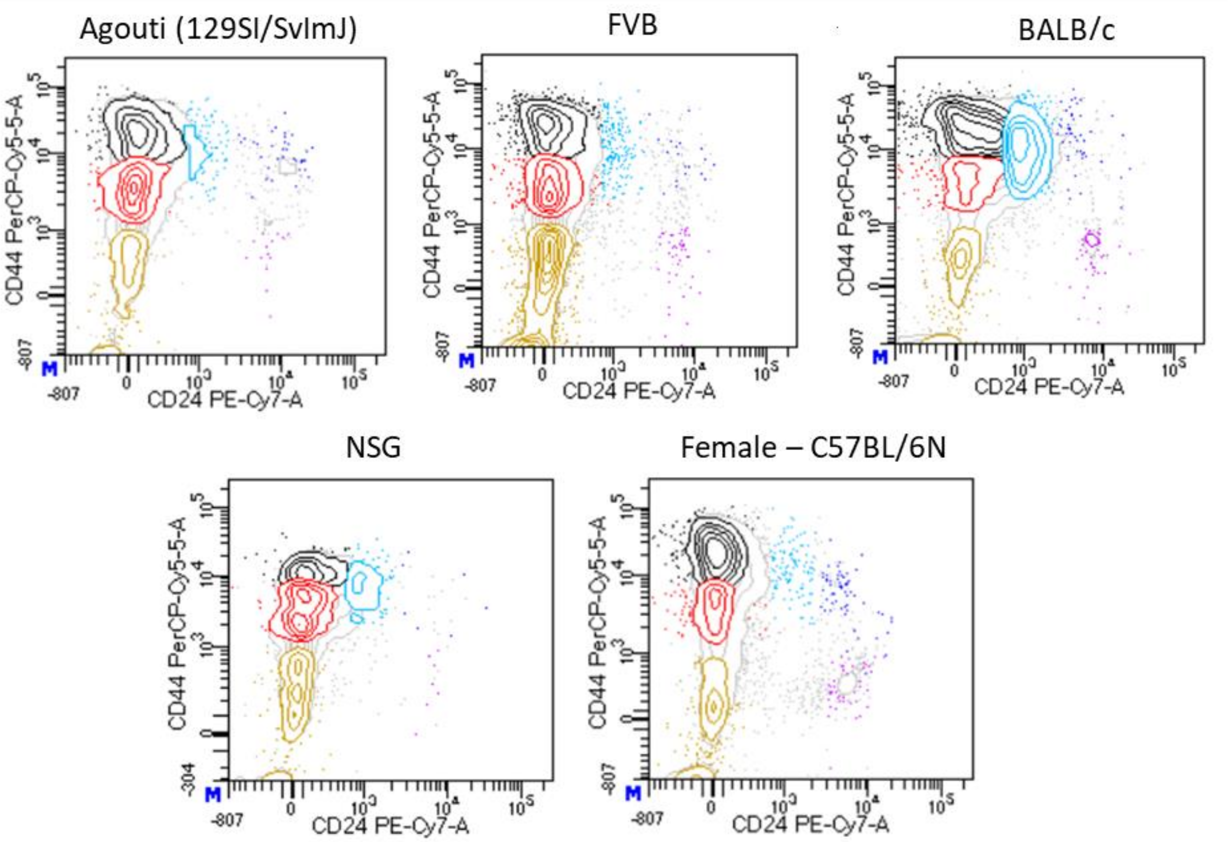
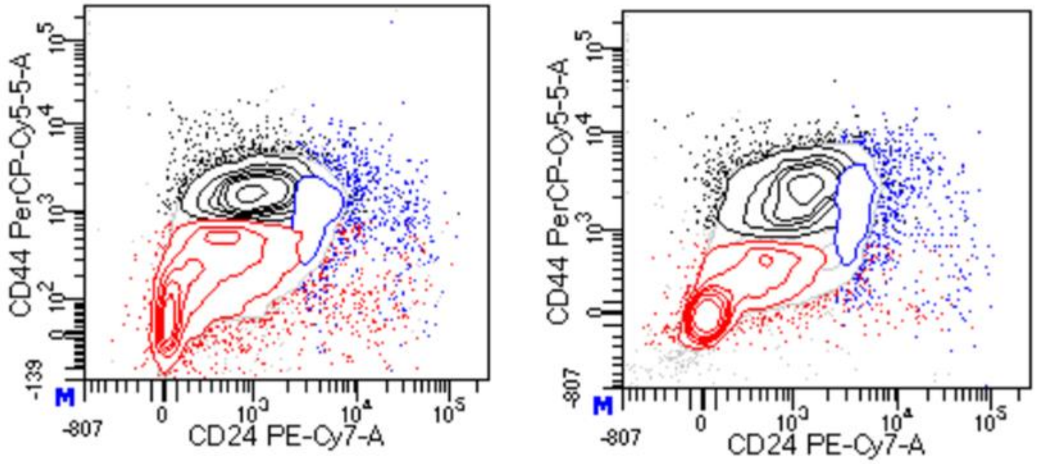


Figure A.2: ***FACS plots demonstrating sorting procedure is effective in diverse mouse strains and both genders.***

FACS plots showing colonic crypt populations in mice of different backgrounds all aged 5-7 weeks and male (unless otherwise specified). Sorting procedure is universal and resolves cell types in the colon of other mouse strains and gender including Agoutti, FVB, Balbc, NSG, and females. Each FACS plot is a representative image from one mouse, n=3 independent sorts.

(Figure adapted from Fig. 2 of Habowski et al., 2020; Protocol Exchange)

a FACS plot resulting from dissociation with TrypLE



b FACS plots resulting from dissociation with **NO** TrypLE (EDTA only)

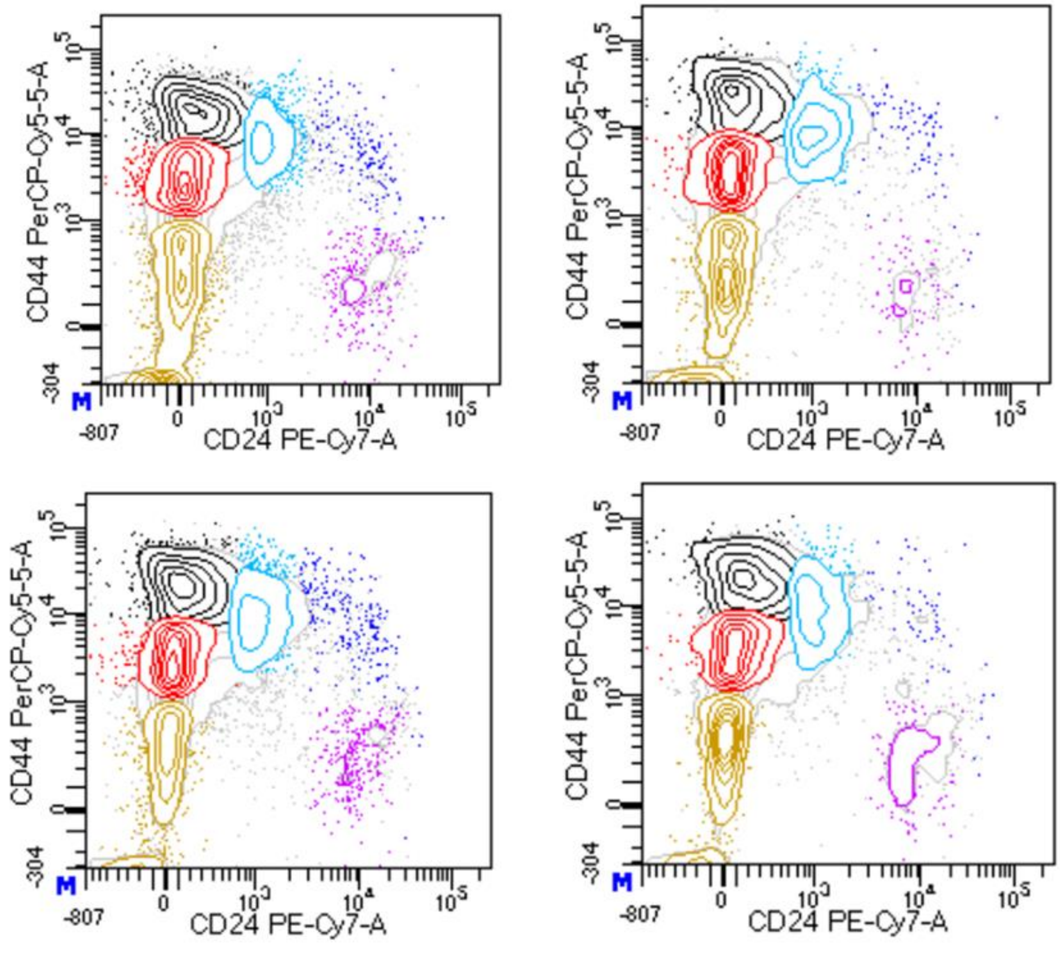
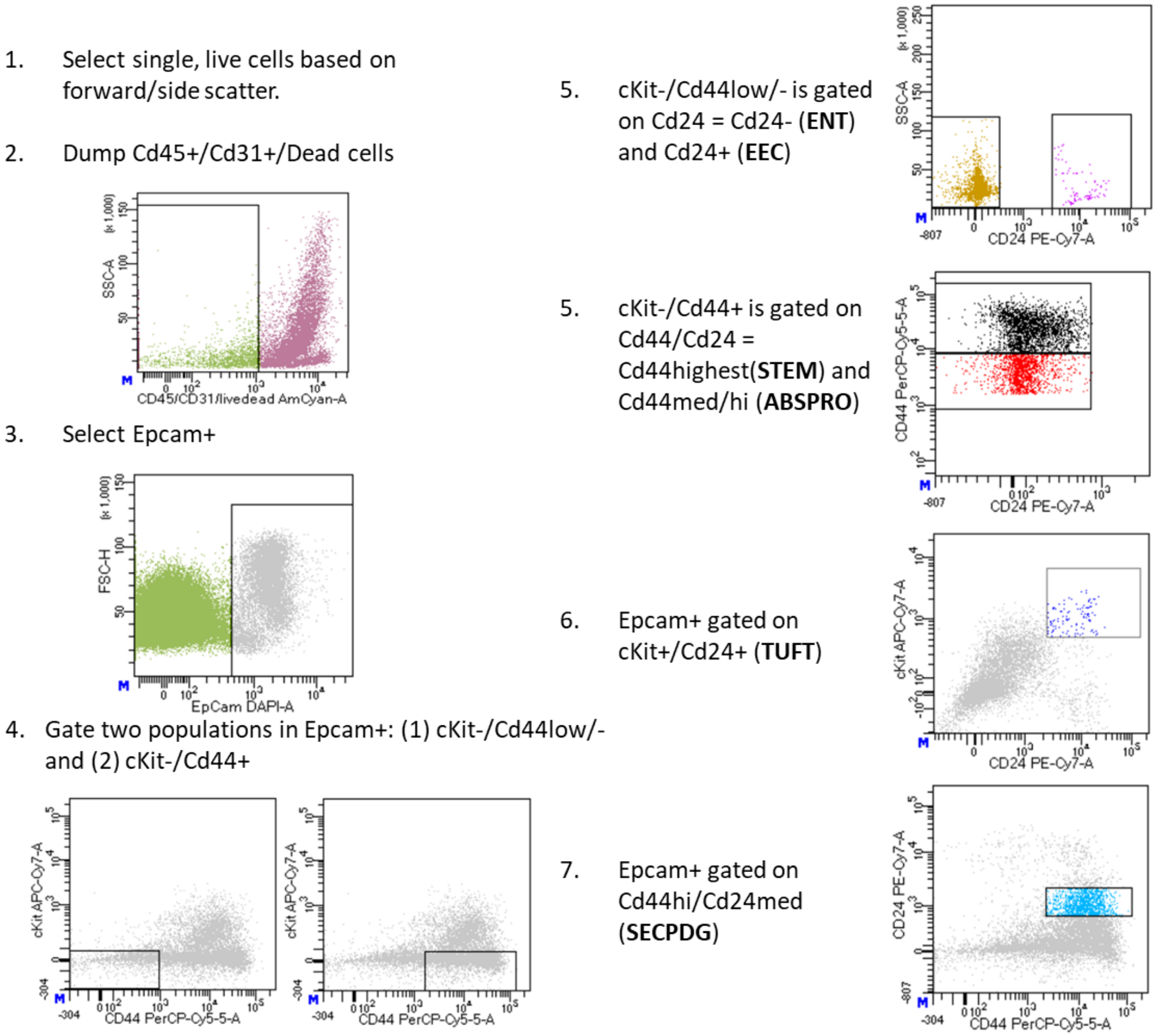


Figure A.3: ***FACS plots of colon crypts dissociated with or without TrypLE protease treatment.***

a When the TrypLE cocktail is used during intestine dissociation, FACS detects decreased Cd44 surface expression and the plots show a compressed population resolution compared to **b** no TrypLE. Each plot is an independent sort from one mouse and is a representative image.

(Figure adapted from Fig. 3 of Habowski et al., 2020; Protocol Exchange)



1. Select single, live cells based on forward/side scatter.
2. Dump Cd45+/Cd31+/Dead cells
3. Select Epcam+
4. Gate two populations in Epcam+: (1) cKit-/Cd44low/- and (2) cKit-/Cd44+
5. cKit-/Cd44low/- is gated on Cd24 = Cd24- (**ENT**) and Cd24+ (**EEC**)
5. cKit-/Cd44+ is gated on Cd44/Cd24 = Cd44highest(**STEM**) and Cd44med/hi (**ABSPRO**)
6. Epcam+ gated on cKit+/Cd24+ (**TUFT**)
7. Epcam+ gated on Cd44hi/Cd24med (**SECPDG**)

Figure A.4: ***FACS gating strategies that define six colonic crypt cell populations.***

As a first step, standard gating is performed to select single, live cells based on forward and side scatter (Step 1). A dump channel then removes dead cells along with immune cells (Cd45+) and endothelial cells (Cd31+) (Step 2). Epcam+ cells (Step 3) are then gated using Cd44, Cd24, and cKit to isolate six distinct populations (Steps 4-7). The resulting populations are Enterocytes (Ent), Enteroendocrine (EEC), Stem cells, Absorptive Progenitors (AbsPro), Tuft, and SecPDG (a mixed population of secretory progenitors, deep crypt secretory cells, with a minor contribution from goblet cells).

(Figure adapted from Fig.4 of Habowski et al., 2020; Protocol Exchange)

REFERENCES

1. Nefzger, C. M. *et al.* A Versatile Strategy for Isolating a Highly Enriched Population of Intestinal Stem Cells. *Stem Cell Reports* **6**, 321–329 (2016).
2. Wang, F. *et al.* Isolation and Characterization of Intestinal Stem Cells Based on Surface Marker Combinations and Colony-Formation Assay. *Gastroenterology* **145**, 383-395.e21 (2013).
3. van der Flier, L. G. *et al.* Transcription Factor Achaete Scute-Like 2 Controls Intestinal Stem Cell Fate. *Cell* **136**, 903–912 (2009).
4. Schuijers, J., Van Der Flier, L. G., Van Es, J. & Clevers, H. Robust cre-mediated recombination in small intestinal stem cells utilizing the Olfm4 locus. *Stem Cell Reports* **3**, 234–241 (2014).
5. Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).
6. Yan, K. S. *et al.* Intestinal Enteroendocrine Lineage Cells Possess Homeostatic and Injury-Inducible Stem Cell Activity. *Cell Stem Cell* **21**, 78-90.e6 (2017).
7. Gehart, H., Van Es, J. H., Dekkers, J. F., Rios, A. & Clevers Correspondence, H. Identification of Enteroendocrine Regulators by Real-Time Single-Cell Differentiation Mapping Data Resources GSE113561 Gehart et al. *Cell* **176**, (2019).
8. Moor, A. E. *et al.* Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis. *Cell* **175**, 1156-1167.e15 (2018).
9. Herring, C. A. *et al.* Unsupervised Trajectory Analysis of Single-Cell RNA-Seq and Imaging Data Reveals Alternative Tuft Cell Origins in the Gut. *Cell Syst.* **6**, 37-51.e9 (2018).

APPENDIX B

Tissue biobanking and human colon organoids

INTRODUCTION TO ORGANOIDS

Human cell lines have enabled phenomenal advances in our understanding of human biology and development of medicine. However, although these model systems are revolutionary and still frequently used today, there are limitations. Cell lines must be transformed in some manner to promote constant proliferation and as a result normal tissue cannot be maintained. Many cancer cell lines were derived from patients decades ago and now, in some biological respects, no longer accurately represent the initial patient tumor. Additionally, the standard manner of culturing cell lines as a monolayer on hard plastic does not recapitulate the human body. For these reasons, and more, improved *in vitro* systems that more accurately mimic the human body have been sought after.

Since identifying limitations of 2D culture systems, researchers have been developing 3D culture systems since the 1970s [1,2]. The first organoids, or 'mini organs', were small intestine organoids initially created and published by the Clevers lab in 2009, following their identification of Lgr5+ stem cells in 2007 [3,4]. Sustained growth of normal un-transformed tissue is enabled by identifying the adult stem cells and optimizing conditions to enable their growth [5]. Importantly, adult stem cells in organoids not only maintain stemness but preserve the ability to form differentiated progeny and distinct lineages when given the appropriate signals. Tumor tissue may not need all of the same growth factors/conditions as its partnered normal tissue but can recapitulate human disease and drug response [6–8]. Importantly organoids are grown in a thick artificial extracellular matrix enabling 3D structures and growth in all directions and requires no adhesion to a plastic surface. The stiffness of the matrix better recapitulates the human body. Researchers have also been uncovering organoid conditions for all types of tissue and disease states [5,9]. Organoids can be rapidly grown directly from patient tissue and many researchers are currently using fresh patient-derived organoids to test drug therapies for individual patients [7,10,11].

Normal human colon organoids are derived from colonic crypts (epithelial layer) and consist primarily of proliferative adult stem cells (Figure B.1a). These stem cells are Wnt dependent and thus a high level of Wnt ligands must be included in the culturing media (see section on conditioned media). Colon tumor organoids frequently have an APC mutation (or other Wnt activating mutation) and are thus grown independent of Wnt ligands (Figure B.1a). Normal crypts and tumor cells are fresh isolated from patient tissue (see protocol section below; Figure B.1b) and within 24 hours of plating small organoid structures are visible (Figure B.2). Normal crypts are not full dissociated to single cells and within a few hours the crypts will close off into a sphere-like structure forming an identifiable organoid by the next day (Figure B.2a-b). Tumors, which are plated as mostly single cells, which have a dense ball structure by the next day which will continue to grow (Figure B.2c-d). Established tumor and normal organoid lines form large sphere structures and proliferate well in optimal conditioned (Figure B.3).

Following training and consultation with the Clevers lab along with assistance from pathologists at UCIMC and St. Joseph Hospital Orange, the Waterman lab has established a human colon organoid biobank and successfully cultured normal human colon organoids and colon tumor organoids. These organoids were used the Nature Cancer, Tran *et al* 2020 publication. We have also demonstrated they are capable of growing tumors in the cecum of an orthotopic xenograft mouse model (Figure B.4). The remainder of this appendix outlines the status of the biobank, protocols for creating organoids lines, maintaining organoids, and creating and testing conditioned media. The hope is this appendix will enable fruitful expansion and continued use of the biobank in future Waterman lab research.

STATUS OF BIOBANK

Number and type of samples

From UCIMC we have received 25 tissues samples in total (23 normal, 2 tumors) with the first collection date of 02/14/2017 (Table B.1). To date we have 57 vials of normal organoids, 6 vials of tumor organoids, and 3 vials of tumor tissue samples frozen in liquid nitrogen from UCIMC (Table B.3). We first begin collected from St. Joseph Hospital Orange on 8/30/2017 and we have received 43 tissues samples in total (22 normal, 19 tumors, and 1 polyp) (Table B.2). The 19 collected tumors all have matching normal. To date we have 215 vials of normal colon organoids, 174 colon tumor organoids, and 55 tumor tissue samples frozen in liquid nitrogen from St. Joseph Hospital Orange (Table B.3).

Naming System

Each tissue that we receive is assigned a name which includes a number, 'N' normal or 'T' tumor, and a 'J' if it is from St. Joseph. For example: N3 (tissue number 3 from UCI, normal) or J-T13 (tissue number 13 from St. Joseph, tumor). Following these numbers is a 'P' for passage and this needs to be recorded on all cryovials and culture flasks. 'P0' vials are isolated crypts/cells that have not yet been passaged. Once these are plated, they are still 'P0' but during the first passage when vials are frozen down these vials are labeled as 'P1'. An example of the names would be J-T23P0 and several passages later J-T23P3. In addition to storing Organoid, tumor tissue chunks are frozen for further genetic analysis or PDX model systems. These are not assigned a 'P' and labeled as 'Tissue'.

IRB approval

Initial work to establish the biobank started in 2017 and with the assistance of Dr. Yung Lyou our IRB was approved. The most recent IRB was approved on 01/10/2018 (Human Subject Assurance Number 00004071). However, St. Joseph Hospital Orange collects tissue with informed consent for general research use (including appropriate Institutional Review Board approval at St. Joseph Hospital Orange) and we have been able to easily receive samples working with their pathology departments - Dr. Shu-Yuan Liao. UCI has recently expanded the Experimental Tissue Resources (ETR) lead by Dr. Delia Tifrea and Dr. Robert Edwards making tissue collection through UCIMC more feasible.

SOLUTIONS

Solutions and protocols are based on material and personal experience/training gained from the Clevers lab. Additional useful references include newly published protocols [11,12] and the original published protocol [7] from the Clevers lab.

AdDF4+

Advanced DMEM/F12 base media (500 mL bottle)

5 mL Glutamax

5 mL PenStrep

5 mL Hepes

1 mL Primocin

Tumor Dissociation Solution (double recipe and use 2 tubes if very large tumor)

10 mL AdDF4+

500 µL for 1.5 mg/mL Collagenase II (Stock is at 30 mg/mL)

20 µL for 20 ug/mL Hyalurodinase (Stock is 10 mg/mL)

10 µL Rock inhibitor (Stock is 1000x)

Normal Colon Chelation Solution (makes enough to be split into 2 tubes with 20 mL)

4.0 mg DTT

200 µL EDTA 0.4M

40 µL Rock inhibitor (1000x)

40 mL -/- PBS

Human Tumor Organoid Media (For 1 mL media)

200 µL 5x Expansion Media

500 µL AdDF4+

200 µL RSPO1 Conditioned Media

100 µL Noggin Conditioned Media

*** for first plating after passaging or isolating new organoid line add Rock inhibitor

Normal Human Colon Organoid Media (For 1 mL media)

200 µL 5x Expansion Media

500 µL WNT3A Condition Media

200 µL RSPO1 Conditioned Media

100 µL Noggin Conditioned Media

***for first plating after passaging or isolating new organoid line add Rock inhibitor

5x Expansion Media (makes 20 mL)

17 mL AdDF4+

2 mL B27 (50x)

250 µL n-acetylcystine (500mM)

1 mL Nicotinamide (1M)

10 µL hEGF (500 µg/mL)

10 µL Gastrin (100 µM)

10 µL A83-01 (5 mM)

10 µL SB202190 (30 mM)

10 µL Prostaglandin (100 µM)

CONDITIONED MEDIA (CM)

Noggin

Cell line: HEK-293 cells stably transfected with: Plasmid is pcDNA3 NEO insert mouse Noggin

C-terminal tag is human IgG1 –Fc (From Hans Clevers' Lab)

Selection: Geneticin (G418) 500 µg/ml **To make conditioned media cells MUST be growing without selection agent**

R-spondin1

Cell line: 293T-HA-RspoI-Fc cells (From Calvin Kuo's Lab)

Selection: Zeocin 300 µg/ml **To make conditioned media cells MUST be growing without selection agent**

Wnt-3a

Cell line: L-Wnt3a (From Hans Clevers' Lab)

Selection: Zeocin 125 µg/ml **To make conditioned media cells MUST be growing without selection agent**

Protocol to make CM (takes several weeks)

1. Thaw vial of cells and plate into T75 flask.
2. Once confluent passage into T225 flask with fresh media (50mL)
3. For Wnt3a flask only – add in 5 mL of pure FBS after 3 days.
4. Collect first batch of CM after appropriate number of days (Noggin ~2-4 days, RSPO ~5 days, Wnt3a must be closer to 7 days). Add fresh media to cells.
5. Centrifuge (2,000 xg 5 min) collected CM and sterile filter (0.22 µm). This can be stored in the fridge for a few days and then pooled with the next batch, or immediately frozen down. Pooling with the next batch creates more consistent CM (but for the first time making it, it is good to test each batch individually to ensure good quality). When freezing down aliquots keep in mind the ratio of CM in organoid media (aka large Wnt3a aliquots, small Noggin aliquots). Make a small aliquot (~3-6 mL) for testing.

6. Noggin can be collected every 1-2 days and RSPO every 3-5 days. Wnt3a needs additional FBS spike on day 3 and then can be collected once a week.
7. Wnt3a (and likely RSPO) cells can continue for ~3 weeks at which point the cells should be discarded. Noggin cells should probably be discarded earlier. Media collected from cells, particularly after the first batch, will likely be acidic (meaning orange or even yellow colored) - this is okay.
8. Avoid freeze-thaw of CM. Store at 4 °C once thawed and use within a few weeks.

Testing CM – Luciferase assay (3-day procedure)

1. Plate HEK293 cells in 6 well dishes for a luciferase assay (following normal luciferase assay protocol) and incubate overnight in regular media (DMEMc).
2. Transfect cells the next day with the appropriate plasmid constructs. For Wnt3a and RSPO CM testing = β-galactosidase + SuperTOP flash (or TOP flash). For Noggin CM testing = β-galactosidase + BRE-luc (pGL3 BRE Luciferase addgene plasmid #45126. This is a BMP/Smad transcriptional reporter and the luciferase activity should be *inhibited* by Noggin).
3. Incubate transfected cells overnight in CM. Wnt3a nor RSPO alone will be sufficient to strongly induce TOPflash, must have both in combination to evaluate effectiveness. Each well in a 6 well contains 2 mL of media. To test RSPO CM batches mix with one batch of Wnt3a and for testing Wnt3a CM mix with one batch of RSPO. An example is shown below. For Noggin testing add 0.5 mL of CM to 1.5 mL of DMEMc.
    a. Example for testing three batches of RSPO. Two plates of 6 well dishes = 12 wells total. Perform duplicates. Use same batch of Wnt3a for all wells.

       Well 1+2 = 2 mL DMEMc

       Well 3+4 = 1.5 mL DMEMc + 0.5 mL Wnt3a

       Well 5+6 = 1 mL DMEMc + 0.5 mL Wnt3a + 0.5 mL RSPO batch 1

Well 7+8 = 1 mL DMEMc + 0.5 mL Wnt3a + 0.5 mL RSPO batch 2

Well 9+10 = 1 mL DMEMc + 0.5 mL Wnt3a + 0.5 mL RSPO batch 3

Well 11+12 = 1.5 mL DMEMc + 0.5 mL RSPO batch 3

4. The next day perform luciferase assay. Examples of results are shown in Figure B.5 Note: L-WRN conditioned media was previously tested (from the L-WRN cell line) and it does not appear to strongly activate Wnt signaling nor promote growth of organoids very well. It does however seem to have good expression of Noggin.

ORGANOID PROTOCOLS

Tissue collection

1. Following surgery, samples are placed in an empty sterile container per standard pathology protocols. This container is stored at 4°C and then transported to pathology for standard clinical pathology work. This transportation step can be sped up if folks in the pathology department know the tissue is ready or if a clinical coordinator is able to help facilitate.
2. Pathologists determine what tissue is considered clinical excess (and would otherwise be discarded). This tissue is saved for our research purposes and is placed in 50mL conical tube containing transport media and stored at 4°C.
    a. Ideally the time from resection of the tissue until it is placed in transport should be less than 2 hours.
    b. Transport media:
        i. AdDF4+ (*preferred*) – this must be made in our lab and transported to the hospital. It is stored at 4°C and good for at least 4 weeks. When tissue collections are infrequent the logistics of keeping the hospital supplied with this transport

media can be a bit challenging (and on more than one occasion old media was used).

    ii. RMPI (or other media the pathology lab has on hand) – this works just fine and I have not noticed much of a difference in yield if prep is done right away. The logistics of this are easier.

3. Tissue will be picked up same day or next day and transported back to lab on ice.
    a. For crypt isolation and organoid cultures, the whole process should ideally be done in 24 hours. Maximum time is ~48 hours and then yield is greatly decreased. The sooner the better.

From a single surgery we usually receive a normal and tumor sample, and it is best to prep these simultaneously. Start with the tumor, and during the dissociation (1hr) begin the normal colon dissection. Ensure all reagents are already prepared and thawed prior to starting (including reagents needed to plate cultures if platting).

Colon tumor isolation

1. (Working in a tissue culture hood) In a large petri dish cut the tumor tissue into pieces of approximately 1-2 mm, save some pieces for RNA/histology/PDX and other intended downstream applications.
    a. For RNA place tissue in 500 µL of TRIzol and homogenize with Precellys.
2. Chop the remaining tumor tissue further using knife/blades until the tumor mass essentially becomes mush and looks a bit viscous.
3. Incubate the tumor mush in a 15 mL conical tube with <u>tumor dissociation solution</u> at 37 °C for 60 min while shaking at 200 rpm. After the first 30 minutes check on the tissue and give the conical a few shakes.

4. Shake vigorously for 1 minute. Filter through 100 µm filter into 50 mL conical. Add 10 mL serum.
5. Add 10 mL media. Centrifuge at 500 xg, 4 °C 5 min then ensure there is a good pellet before removing the supernatant.
6. Add red blood cell lysis to pellet (~ 5 mL), resuspend and wait for a few minutes.
7. Add 10 mL media and split into 15 mL conical tubes for downstream processing (tube for RNA isolation, freezing down cells, and directly plating cells).
8. For freezing down use 2x freezing media (40 % serum, 20 % DMSO, 40 % AdDF4+) mixed 1:1 with cells in AdDF4+. <u>This is labeled as passage 0.</u>
9. For plating, resuspend in Cultrex BME (Clevers lab tumor ratio = 100 µL media + 600 µL BME), keep on ice. Place droplets of suspension on pre-warmed 6 well plate. Allow to solidify for ~1-5 minutes at room temperature. Then gently flip plates upside down and place in incubator at 37 °C and 5% CO2.
10. Wait 15-45 minutes for complete solidification, and then add in <u>human tumor organoid media</u> with Rock Inhibitor.
11. Add fresh human tumor organoid media every 2-3 days and passage organoids after 7-10 days.

Normal colon isolation

1. (Working in a tissue culture hood) Place normal colon tissue in a large petri dish along with a little transport media (to ensure it does not dry out). Spread out the normal tissue with the mucosal side down.
2. Remove the muscle layer, fat, and as much of stroma as possible using surgical scissors and forceps. This will have to be done blinded unless a dissecting scope can be used in the hood. Take no longer than 30 minutes (closer to 15 minutes if possible) and make sure tissue does not dry out - if it starts to get sticky this is a sign it is drying out.
3. Collect a few chunks of tissue to save for RNA, Protein, etc.

    a. For RNA place directly into 500 µL TRIzol and homogenize with Precellys.

4. Put tissue in <u>normal colon chelation solution</u> #1 (50 mL conical), on rotator in cold room for 15 minutes.
5. Transport tissue to new tube of Chelation solution (#2), cutting tissue into small chunks. Place in cold room rotator for 15 minutes.
6. Shake <u>vigorously</u> for 3 minutes. Be very aggressive.
7. Pour supernatant through a 70 µm filter and rinse with 10 mL of serum (or serum containing media). Spin down this supernatant (500 xg 4°C, 5 min). Use forceps to recollect tissue trapped in the filter, place back in conical with ~20 mL of PBS or serum containing media, shake for 3 minutes and then filter again, and spin down. (These can also be pooled before centrifuging).
8. Ensure there is a good pellet before removing supernatant. Repeat centrifugation if needed.
9. Add red blood cell lysis to pellet (~ 5 mL), resuspend and wait for a few minutes.
10. Add ~5-10 mL media and split into 15 mL conical tubes for downstream processing (tube for RNA isolation, freezing down cells, and directly plating cells).
11. Spin down 15 mL conical tubes and then remove supernatant. Keep pellets on ice while processing.
12. For freezing down use 2x freezing media (40 % serum, 20 % DMSO, 40 % AdDF4+) mixed 1:1 with cells in AdDF4+. <u>This is labeled as passage 0</u>.
13. For plating, resuspend in Cultrex BME (Clevers lab normal colon ratio = 100 µL media + 600 µL BME), keep on ice. Place droplets of suspension on pre-warmed 6 well plate. Allow to solidify for ~1-5 minutes at room temperature. Then gently flip plates upside down and place in incubator at 37 °C and 5% CO2.
14. Wait 15-45 minutes for complete solidification, and then add in <u>normal colon organoid media</u>. Typically, crypts close within a few hours.

15. Add fresh complete media every 2-3 days and passage after 7-10 days.

Passaging Organoids

1. Harvest organoids by pipetting up and down with p1000 to release BME from plate (do not remove media).
2. Collect in a 15 mL conical tube.
3. Spin down for 5 min at 500 xg, 4 °C.
    a. This results in a larger pellet and a foggy layer above with a mix of organoids and BME, or just more BME and other cell debris if spun down well.
4. Remove supernatant (and BME if can do so safely), be very gentle.
5. Add 5 mL TrypLE, vortex briefly and incubate at 37 °C.
    a. For normal organoids = only ~ 5 minutes
    b. For tumor organoids = ~15 minutes
6. Add 5 mL of media and mix with a glass pinpoint tip pipet (~3 mixes for normal, 5+ mixes for tumor. Depends on the flow rate)
    a. I made my pipets with a 10 mL glass serological and flamed them to make a small pin prick hole. I bleach them, autoclave them, and re-use them. Others make these out of pastuer pipettes and flame them in the hood while cells are in TrypLE.
    b. This step is very important to break apart the organoids.
7. Spin down for 5 min at 500 xg, 4 °C.
8. Remove supernatant, resuspend pellet, and prep for platting/freezing or downstream analysis. Plating procedure outlined above in isolation procedures.

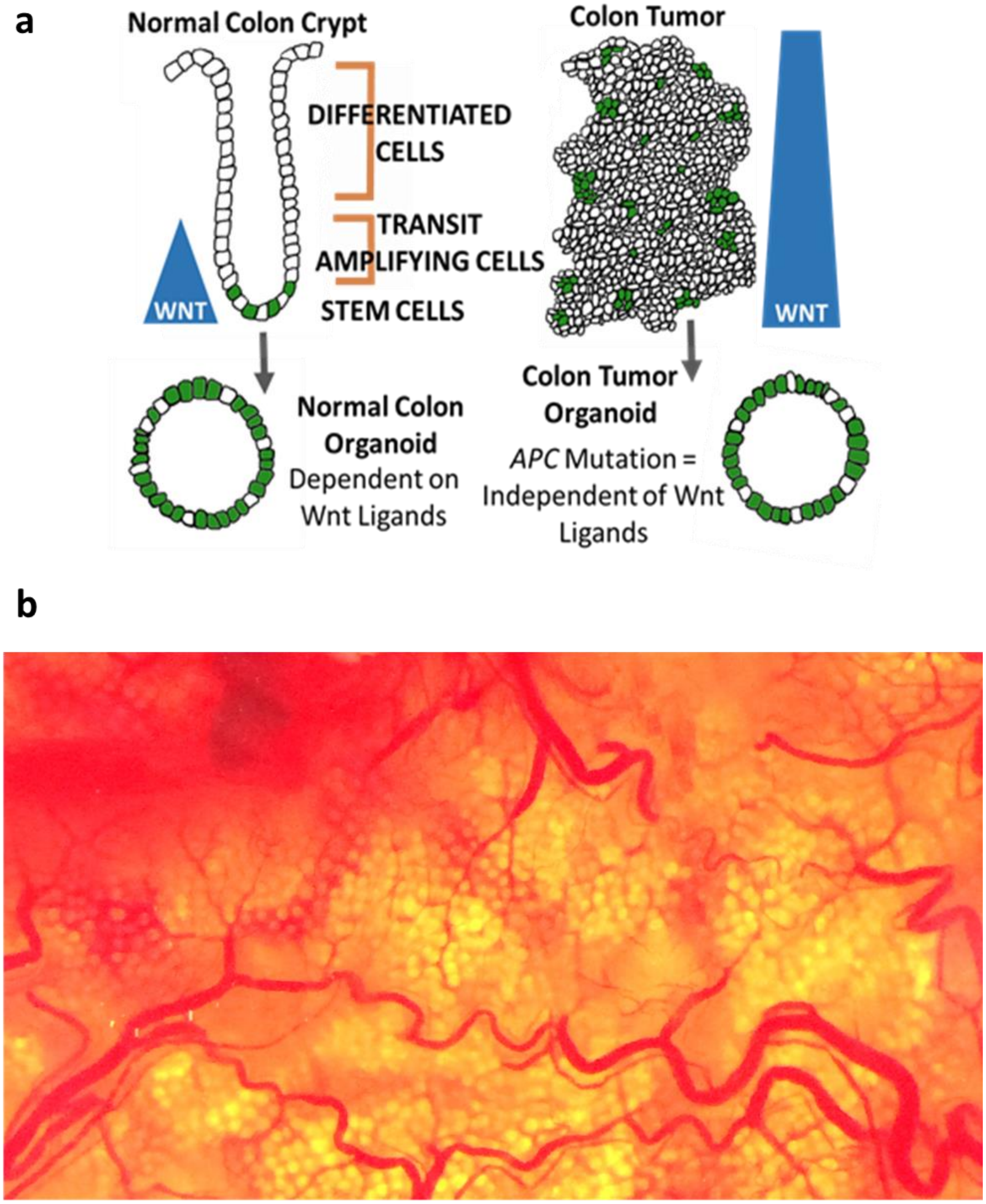
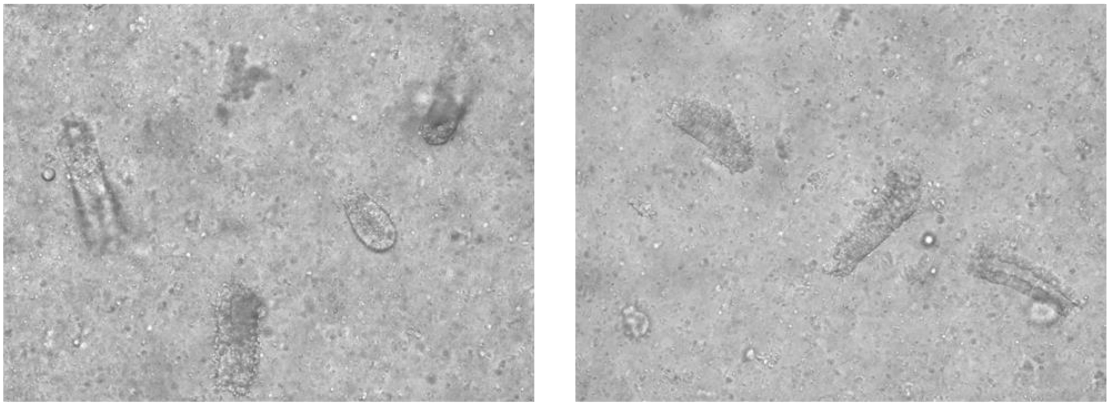


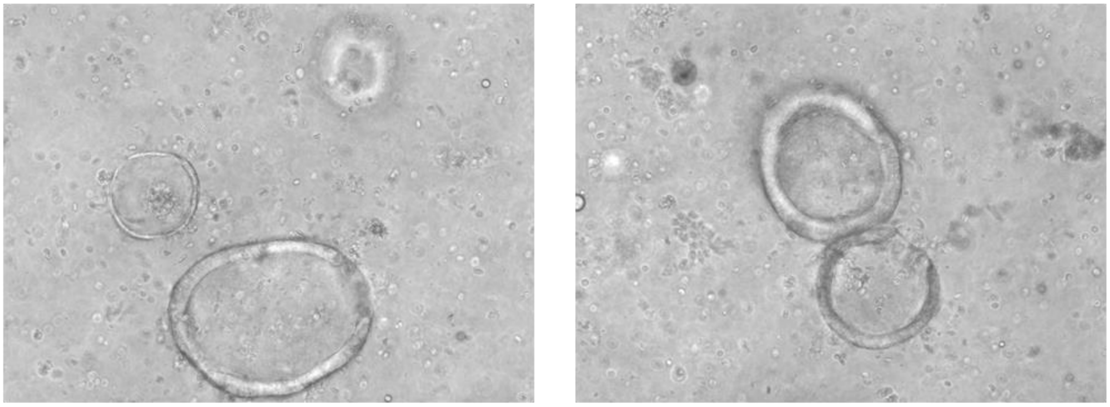
Figure B.1: ***Human colon tissues isolation and organoid generation.***

a Intestinal organoids are isolated from normal colon crypts (epithelial layer) and from colon tumors. Normal organoids are primarily composed of stem cells and are dependent on Wnt ligands for sustained growth. Tumor organoids are not supplemented with Wnt ligand enabling cells with APC mutations (or other Wnt activating mutations) to proliferate. **b** Human stroma during normal colon dissection. The stroma is highly vascularized, underneath this stroma layer the crypt forms a honeycomb-like pattern.

a Normal Colon (0 hrs after platting)



b Normal Colon (24 hrs after platting)



c Colon Tumor (0 hrs after platting)

d Colon Tumor (24 hrs after platting)

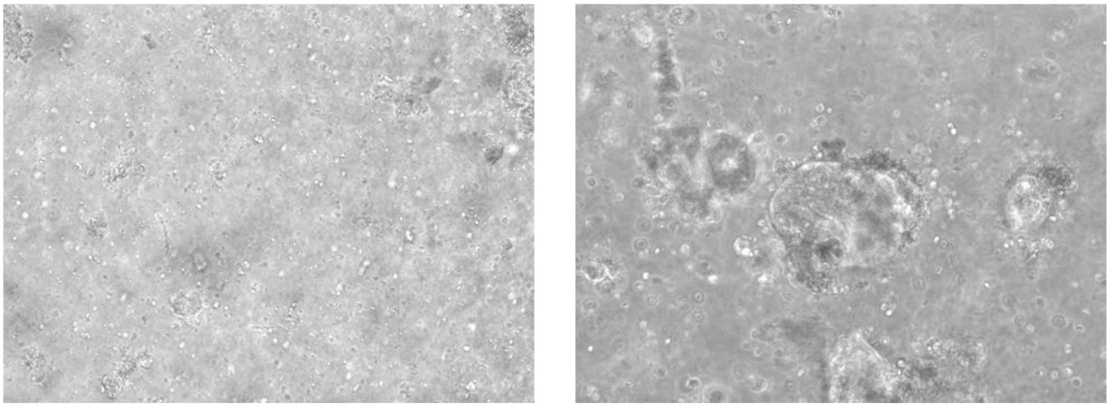
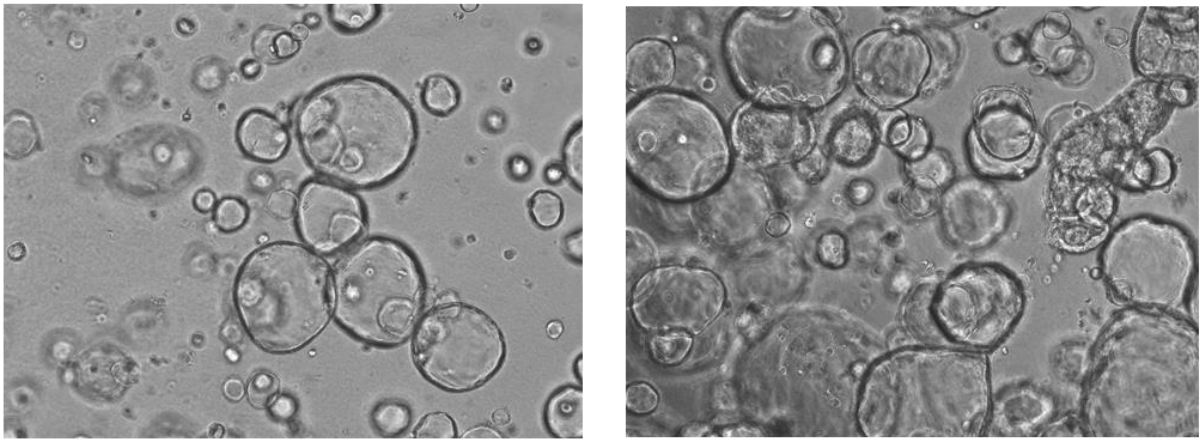
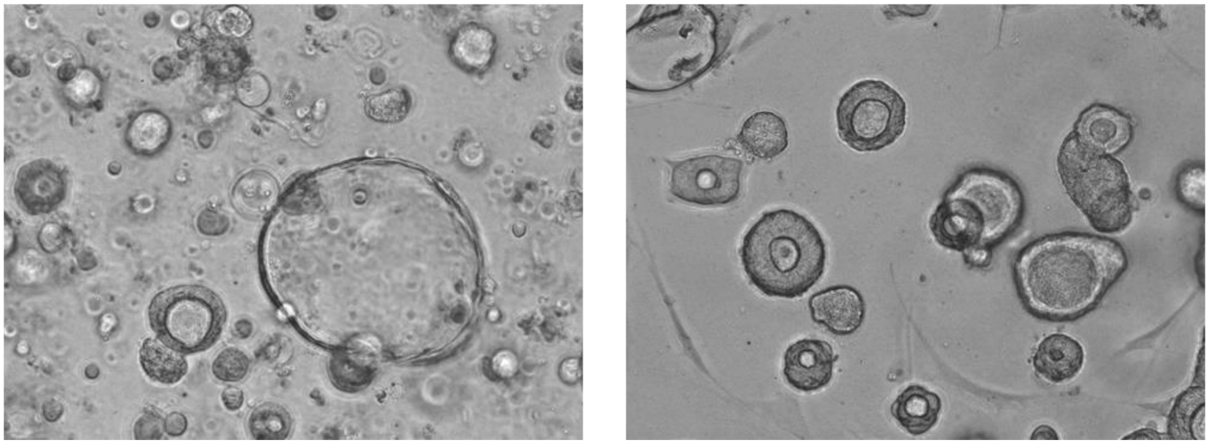


Figure B.2: ***Rapid formation of human colon organoids.***

Organoids are visible within 24 hours of platting colon cells in matrix following dissociation. **a** Dissociated normal colon (with some partially intact crypt sections) immediately after plating compared with **b** 24 hours later (10x). **c** Dissociated tumor (nearly all single cells) immediate after plating (10x) compared with **d** 24 hours later (20x). During the first few passages there is an abundance of debris.



a Normal Colon Organoids



b Colon Tumor Organoids

Figure B.3: ***Established human organoid lines.***

Established organoid lines rapidly proliferate and create identifiable structures. **a** Normal human colon organoids are cystic with a hollow center. The cells are the edge are stretched out and can be rather thin. We think of these as ‘necklace’ organoids. **b** Colon tumor organoids come in a variety of morphologies. ‘Necklace’ organoids can be observed, along with ‘fatty’ (where the cystic center is smaller and cells are oriented from the center to the outside rather than being stretched around like ‘necklace’), and ‘ball’ which do not have a cystic center. Each patient-derived organoid line can have some variation in morphology.

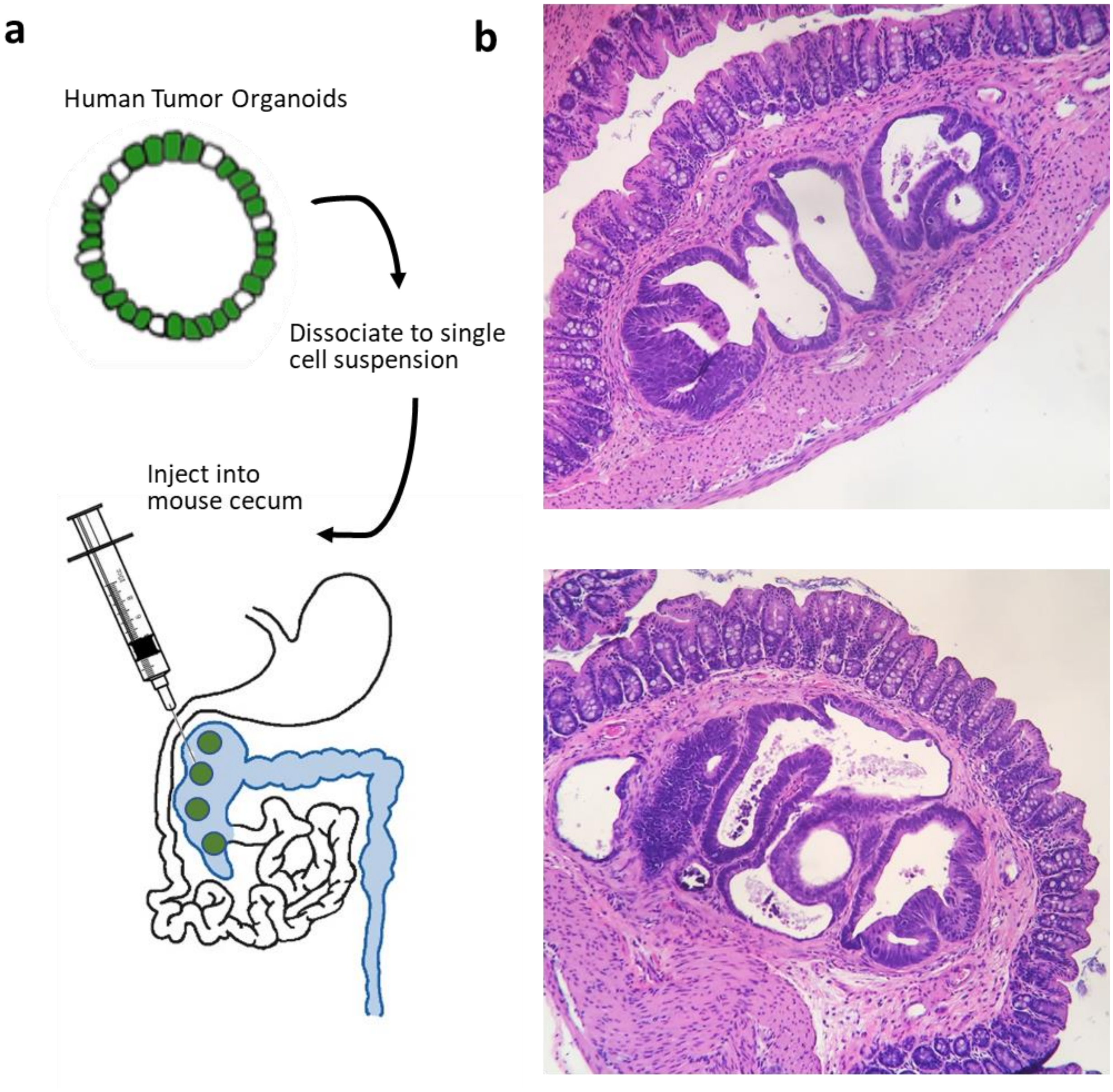


Figure B.4: ***Orthotopic xenografts using human tumor organoids***

a Schema of creating orthotopic xenografts. Human tumor organoids are dissociated into single cells and then carefully injected into the mouse cecum between the muscle and epithelial layer. This results in **b** tumor growth in the stroma. These tumors are not advanced and are well constrained by stroma. The epithelial layer is towards the top made up or repeating crypts. Despite being injected as single cells, the organoids-derived tumor cells formed an organoid-like structure with hallow cystic area in the center.

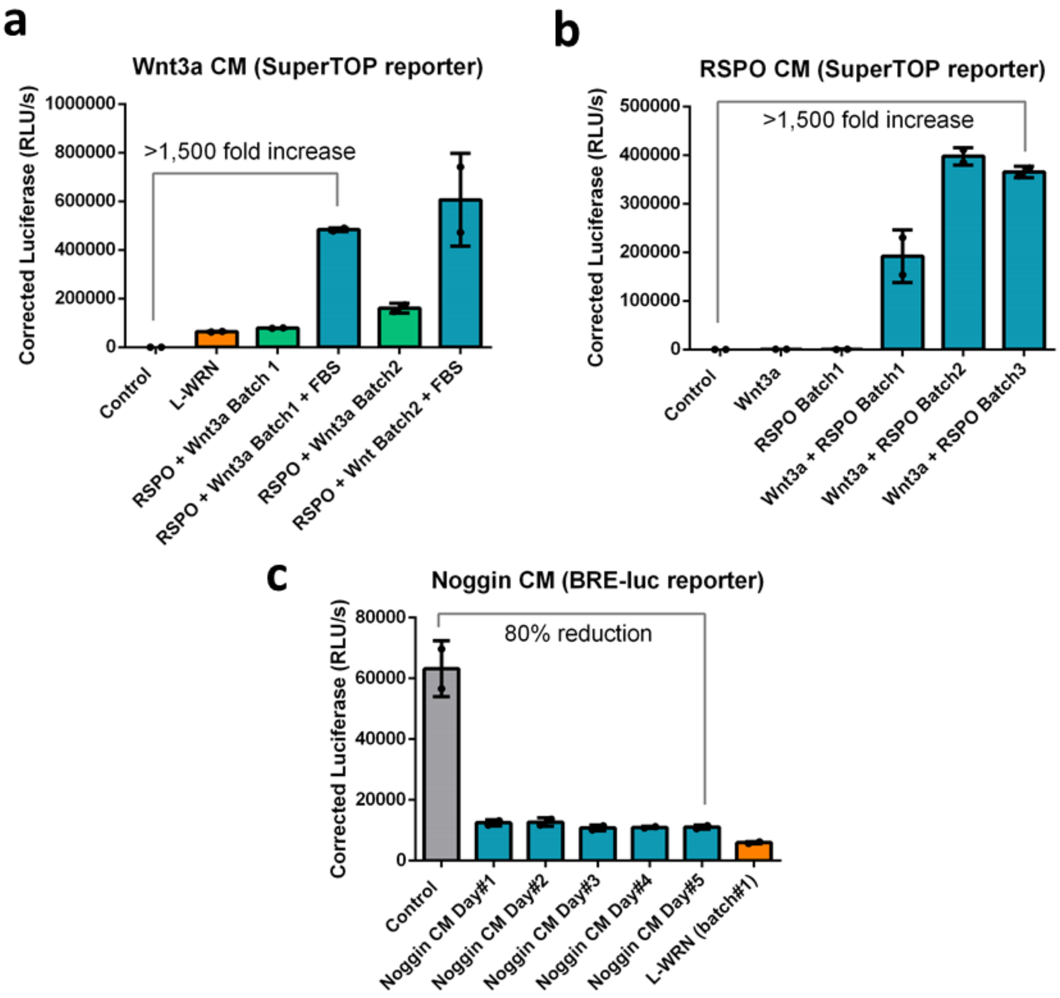


Figure B.5: ***Evaluation of conditioned media with luciferase assay***

A luciferase assay is performed by transfecting HEK293 cells with β-galactosidase and Wnt signaling SuperTOP Flash luciferase reporter (for Wnt3a or RSPO CM) or SMAD/Bmp BRE-luciferase reporter (for Noggin CM). **a** It is critical that Wnt3a CM is made with a FBS spike-in (+FBS) as side-by-side batches with and without FBS show a striking difference. A good batch of Wnt3a-RSPO should result in a >1,500-fold increase in luciferase activity. L-WRN does not strongly activate Wnt signaling. **b** Wnt3a and RSPO alone do not strongly activate Wnt signaling. In this example, RSPO Batch1 is not as active as Batches 2 and 3. **c** These batches of Noggin CM, collected on sequential days, results in upwards of 80% reduction in BRE-luciferase activity (50% is likely sufficient). L-WRN results in 90% reduction (but as shown in **a**, lacks activation of Wnt signaling). The control condition in **a**, **b**, and **c** are HEK293 incubated in 100% DMEMc. CM treatments are 25% for each CM. RSPO + Wnt3a = 50%, resulting in only 50% DMEMc (fresh-nutrient rich media). However, this does not seem to impact the growth or transfection of the HEK293 cells since the incubation is only for 24 hours and β-galactosidase recordings are consistent with the control. Biological duplicates were used for each CM, corrected RLU/s are normalized to β-galactosidase values for the same well.

Table B.1: ***UCIMC normal and tumor tissue collection log.*** *(two pages)*

UCIMC - Colon Tissue Receipt List/Log				Tissue/Organoids labeled as TissueType/number/passage/number; example N1P0 (the first freeze down before platting is P0, the second freeze down - followng 1 split- is P1. A few may not follow this perfectly so double check)							
Tissue Number	Tissue Type	Tissue Sample Name	Date of Surgery	Scheduled time of Surgery	Date of Receipt	Time of Receipt	Date of Organoid Isolation	Time of Isolation	Tissue ID	Notes	Outcome
1	Normal	N1	42780	7:15 AM	2/14/2017	2:30 PM	2/15/2017	10:00 AM	5952 or 5967	Rob delievered to lab. sample had very bad diverticulitis	No oranoids
2	Normal	N2	42783	7:15 AM	2/17/2017	4:30 PM	2/17/2017	5:30 PM	5952 or 5967	Rob delievered to lab.	no organoids - contaminated
3	Normal	N3	42793	Morning	2/27/2017	5:00 PM	2/27/2017	6:00 PM	5970	Drove to the hospital to pick it up, meet Anne Sawyers. Pathology finished and placed in media at 3:48pm. ID = TB5970. During isolation split into two groups (two different dissociations) of N3a and N3b	Saw crypts directly after plating, organoids didn't make it past a few days
4	Normal	N4	42814	Morning	3/20/2017	3:30 PM	3/20/2017	4:20 PM	5995	Drove to the hospital to pick it up, meet Anne Sawyers. Pathology Finished and placed in media at 3:28pm. During Isolation split into two groups (two different dissociations) of N4a and N4b. Patient had early stages of diverticulitis (little indents). Plated at 7pm. Next day plate "a" was contaminated. On thursday plate "b" was contaminated. Likely sample is contaminated.	no organoids - contaminated
5	Normal	N5	42814	2pm?	3/21/2017	2:00 PM	3/21/2017	6:00 PM	3225	Rob delivered to me - very large section! Ended up using half for oraganoid culture and other half to try the tissue direct drug treatment.	unsuccessful RNA prep from tissue treatment. No viable organoids
6	Normal	N6	42815	Morning	3/21/2017	6:45 PM	3/21/2017	9:00 PM	6003	Rob also delievered to me - smaller sample, harder time seeing crypts so not as confident with the isolation (microscope light died).	Grew for a little while but eventually no viable organoids
7	Normal	N7	42829	Morning	4/5/2017	9:30 AM	4/5/2017	8:25 PM	6019	Put in transport media at 6:33pm on 4/4/2017 - still sat out from a pretty long while (also placed in older tranport media 2.17.17). Drove and picked up from the hospital in the morning, couldn't start prep until later. Very bad diverticulitis, very little normal crypt area.	Nothing viable
8	Normal	N8	42829	3:00 PM	4/5/2017	9:30 AM	4/5/2017	8:45 PM	6020	Put in transport media (dated 3.21.17) at 9:27am on 4/5/2017 - this means it sat out overnight - didn't get anything from this prep. Drove and picked up from the hospital in the mornin, couldn't start prep until later. Mild diverticulitis.	Nothing viable
9	Normal	N9	42857	7:15 AM	5/2/2017	5:00 PM	5/2/2017	5:10 PM	?	Rob delivered - patient had crohn's disease. Rob said the illium was very inflammed but the cecum look more normal (which is this). Transport media dated 3.21.17	L-WRN + 10uM CHIR treated is growing very well! Other conditions not.
10	Normal	N10	42864	Morning	5/9/2017	3:45 PM	5/9/2017	8:37 PM	?	Rob dropped off - Younger patient with ulcerative colitis, said this part was pretty normal though. Transport media dated 3.21.17	Started off okay (not great) and then kind of tapered off - grown in CHIR
11	Normal	N11	42901	Morning	6/15/2017	2:15 PM	6/15/2017	4:00 PM	S17-6844	Haik called me to tell me that had normal tissue from a colon cancer patient, Yung clarified this was okay to collect without consent and with our current IRB. Drove to the hospital to pick it up. Very small piece of tissue - not sure if it was even enough to do crypt isolation. Haik said from patient to media was probably about 2.5 hours so this should be a good sample! Older transport media though.	Did RNA isolation on total tissue

12	Normal	N12	42906	?	6/21/2017	5:00 PM	6/21/2017	6:30 PM	?	Haik notified me about this tissue and he was not notified the day of the actual surgery. Time from patient to media was 3 hours. Very very large sample!	RNA isolation from Tissue, Crypts, and plated organoids
13	Normal	N13	42907	Morning	6/21/2017	5:00 PM	6/21/2017	6:55 PM	254875	Haik notified me about this tissue and it was less than two hours from patient to media (at 1:40pm). Good size	RNA isolation from Tissue, Crypts, and plated organoids
14	Normal	N14	42907	Afternoon	6/21/2017	5:00 PM	6/21/2017	7:15 PM	2348779	Haik notified me about this tissue and it was patient to media in 30 minutes! Pretty small sample and it was a very difficult dissection and muscle was stiff, fiberous tissue.	RNA isolation from Tissue, and plated one well of organoids
15	Normal	N15	42940	Morning?	7/24/2017	5:00 PM	7/24/2017	5:05 PM	?	Rob got the tissue at 2:40pm. Was kept in older transport media (3.21.17). Pretty bad diverticulitis and was mostly a large inflammed area with a little bit of normal epithelium. Rob delivered the tissue and since only a small amount of usable tissue, just did a whole tissue RNA prep (no organoids).	Very good RNA prep >350ug of RNA total
16	Normal	N16	42949	Morning?	8/2/2017	5:40 PM	8/2/2017	6:37 PM	6153	Got a text from Haik and picked up from Anne Sawyers. Tube was old transport media from 3.21.17. Time colon was put in media was 15:00, Haik said specimen was <2hrs. Very large normal sample. Performed direct tissue RNA isolation and isolated crypts for RNA isolation as well.	
17	Normal	N17	42969	Morning	8/22/2017	4:30 PM	8/22/2017	5:45 PM	?	Ghadi picked up at UCMC for me, not sure how long it was before being placed in media. Media was old! (2.10.17) Did whole tissue RNA isolation (A+B).	
18	Normal	N18	42984	Afternoon?	9/6/2017	6:00 PM	9/6/2017	6:40 PM	?	Rob dropped off on his way home. Didn't know tissue was coming. Was in newer transport media dated 7.19.17. Nice sample, not sure why is was surgically removed, on the smaller side.	RNA isolation
19	Normal	N19	43007	midday	9/29/2017	4:40 PM	9/29/2017	5:10 PM	026?	Picked up from UCIMC, was ready at 3pm but was finishing another prep. Was in old transport media 7.19.17 and was a very small section. Only did a tissue RNA isolation since not enough tissue to do anything else.	
20	Normal	N20	43024	Morning	10/16/2017	~5:30pm	10/16/2017	7:00 PM	?	Was ready for pickup about 3pm but I had class. Ghadi went and picked up the sample and brought it back for me. Was in old transport media 7.19.17. Pretty small normal but did have everything in tact. Ended up doing tissue RNA isolation for each section (epithelial layer, mucosal layer, fat layer, and muscle layer).	
21	Normal	N21	43025	Morning	10/17/2017	4:00 PM	10/17/2017	4:55 PM	?	Ghadi picked up this sample after his shift. Sample was ready around 3, Ghadi brought it to be at 4:30pm. Was in transport media from 7.19.17. Smaller sample but enough to collect tissue for RNA isolation from all layers (as in N20) and do a small organoid prep.	
22	Normal	N22	43724	?	9/17/2019	9:15 AM	9/17/2019	9:30 AM	N2	Delia brought tissue samples to UCI next day. Linzi received samples 9:15 AM. Normal tissue was thinner and Delia said it was just mucosa (looked mostly like yellow fat). Linzi cut up mucosa into small pieces and then incubated in chelation solution 1X for 25 min (not 2X). Barely had any cells pellet out of fatty opaque solution so froze one cryovial with all the cells in the pellet.	
23	Tumor	T23	43724	?	9/17/2019	9:15 AM	9/17/2019	3:00 PM	T2	Delia brought tissue samples to UCI next day. Linzi received samples 9:15 AM. Tumor was dissected around 3pm. Tumor sample was very fatty and mucosal like, with one large white hard fibrous shell inside was necrotic core (caseous necrosis = looked like cornmeal). RNA preps of the bulk tissue contained pieces of this white shell. Also saved some tissue in FBS/DMSO for PDX. Surprisingly for the little remaining tissue, there was a big cell pellet and Linzi froze 6 cryovials of p0 cells and 2 eppitubes for RNA.	
24	Normal	N24	43738	7:15 AM	10/1/19	12:00 PM	10/1/19	1:45 PM	none given	Delia put tissue samples on ice at 4pm on day of surgery and Pathology brought from UCIMC to Edwards lab next day. Linzi picked up samples after class and immediately dissociated. Normal was much better this time, tiny piece (1 in x 2 in) but had some muscle and submucosa still intact. Mucosa looked as normal too. Since such a small piece, cut into small pieces and incubated in EDTA solution once for 30 min. Counted with countess and cell viability was ~45% and total live cells collected ~ 4 million (frozen in 4 cryovials, 1 million each)	
25	Tumor	T25	43738	7:15 AM	10/1/19	12:00 PM	10/1/19	2:15 PM	none given	Delia put tissue samples on ice at 4pm on day of surgery and Pathology brought from UCIMC to Edwards lab next day. Linzi picked up samples after class and immediately dissociated. Tumor was much better this time, hard fibrousy chunk in between fat and mucosa, so trimmed tumor sections so wouldn't have normal cells contaminating. Saved some tumor pieces for RNA, none for PDX. DNase treated and filtered well. Counted with countess and cell viability was ~65% and total live cells collected ~ 6.5 million (frozen in 6 cryovials, 1 million each)	

Table B.2: *St. Joseph Hospital Orange normal and tumor tissue collection log.* *(three pages)*

St. Joseph - Colon Tissue Receipt List/Log				Tissue/Organoids labeled as J-TissueType/number/passage/number; example J-N1P0. The J is to distinguish it as from St. Joseph's Hospital rather than UCIMC								
Tissue Number	**Tissue Type**	**Tissue Sample Name**	**Date of Surgery**	**Surgery Time**	**Date of Receipt**	**Time of Receipt**	**Date of Organoid Isolation**	**Time of Isolation**	**Tissue ID**	**Notes on Prep**	**NOTES from Shu**	**Other Notes**
1 *used in Thai's Paper	Tumor	J-T1	8/30/2017	am	8/30/2017	noon	8/30/2017	1:00 PM	Case #1 tube 1-B	Picked up from St. Joseph's hospital. Very large tumor! 2 Cryovials of whole tissue (90% FBS + 10% DMSO), 2 homogenized tissue samples in trizol, and two tubes of dissociation (froze down 5 vials of dissociated tissue), plated full plate of tumor organoids.	Normal tissue and tubulovillous adenoma, cecum	CASE #1 Tubulovillous adenoma (cecum), serrated polyp appendix Normal tissue + adenoma MMR: No loss 74/Male
2	Normal	J-N2	8/30/2017	am	8/30/2017	noon	8/30/2017	4:00 PM	Case #1 tube 1-A	Three large pieces, dissected each one at a time and then dissociated together with one refresh of dissociation media after 15 min. Homogenized some tissue for RNA and kept three vials of dissociated cells and plated a full plate of organoids.		
3	Polyp	J-P3	8/30/2017	am	8/30/2017	noon	8/30/2017	4:00 PM	Case #1 tube 1-C	Very small - just homogenized and placed in trizol for RNA isolation		
4	Normal	J-N4	9/26/2017	am	9/26/2017	1:30 PM	9/26/2017	2:45 PM	Case #2	Picked up from hospital. Enormous amount of tisssue! Wow! I took out the four largest pieces, Ghadi practiced on a piece and I gave the left over to George (I think he was going to freeze/embed them for staining or other work - not sure). Each of the four pieces alone I would have been happy with (and were larger than what I often get from UCIMC). Normal tissue looked good. Shu had it ready for me a little after 10:30am but I could make it there until later. I made 4 samples in tissue RNA, 4 samples of dissociated cell RNA, and 10 frozen down sample of dissociated cells.	Normal colon (carcinoma was removed by polypectomy). Carcinoma shows loss of MLH1 and PMS2 nuclear expression.	CASE#2 Normal colon , no carcinoma in the resection specimen Previous polyp removed containing adenocarcinoma MMR was performed on adenocarcinoma. MLH1 & PMS2: loss. MSH2 & MSH6: intact. BRAF mutation, not detected
5	Tumor	J-T5	9/28/2017	midday	9/29/2017	11:30 AM	9/29/2017	12:00 PM	Case #3	Wasn't able to make it to the hospital the day before, but picked the samples up in the morning and was able to process them then. Decent sized tumor and was able to collect tissue, tissue for RNA, and cells for RNA as well as plate organoids. Shu later emailed to say "Turn out this is a MSH6 deficient tumor with high expression of CA9."	Normal colon and adenocarcinoma with loss nuclear staining of MSH6	CASE #3 Normal colon with adenocarcinoma of cecum MMR = MSH6: loss; MSH2, MLH1 & PMS2: intact
6	Normal	J-N6	9/28/2017	midday	9/29/2017	11:30 AM	9/29/2017	12:30 PM	Case #3	Normal to J-T5 case. Rather large sample - also was able to prep tissue, tissue for RNA, and cells for RNA as well as plate.		
7	Tumor	J-T7	10/19/2017	morning	10/19/2017	noon	10/19/2017	12:30 PM	Case #4	Yung was at UCIMC in the morning and was able to pick it up and bring it to me in lab. Decent amount of tumor tissue, not one lump but many small chunks. Plated for organoids, froze down many cells and RNA samples from tissue and cells		All wells contaminated with fungus + after 2 days in culture. Might be the sample? CASE #4 Normal colon with adenocarcinoma MMR: no loss
8	Normal (and polyp?)	J-N8	10/19/2017	morning	10/19/2017	noon	10/19/2017	12:30 PM	Case #4	Yung also picked this sample up at the same time. Large amount of tissue. Prepped tissue, tissue for RNA, and cell for RNA as well as plated. While dissecting this tissue I found a small polyp looking lump. It was a bit firmer than the rest of the tissue and slighty red/pink in coloration compared to the normal epithelium. Looked like a poly so I collected it individually for RNA.		
9 *used in Thai's Paper	Tumor	J-T9	11/7/2017	mid day	11/8/2017	9:00 AM	11/8/2017	6:00 PM	Case #5	This tissue was ready about 5pm on 11/7 but I was unable to pick it up that day. Yung was able to pick it up for me the next morning on his way in to work and then I finally got the chance to prep it in the evening. Transport media was dated 9.26. The tumor was pretty small kind of chunk peices so I didn't freeze down in FBS since there wasn't much. I did collect tissue for RNA and freeze down cells.		CASE #5 Normal colon with adenocarcinoma MMR: no loss 48/female
10	Normal	J-N10	11/7/2017	mid day	11/8/2017	9:00 AM	11/8/2017	6:15 PM	Case #5	Yung pick up at the same time as J-T9. Pretty large sample and so I kept tissue for RNA and also had plenty to freeze down crypts and RNA from crypts. Didn't plate organoids from either of these since I will be gone for the holiday. Will have to plate the frozen stock when I get back.		
11	Normal	J-N11	11/14/2017	morning	11/14/2017	4:30 PM	11/14/2017	5:30 PM	Case #6	Ran out of transport media so Shu put these in Hank's buffer-Ghadi picked up on his way from UCIMC to main campus. Had three pieces of tissue, didn't use the one tissue that had the worst diverticuli and looked a bit more inflammed. Kept tissue for RNA and also frozen down crypts and had crypt RNA.		CASE #6 Diverticulitis, no cancer
12	Normal	J-N12	11/14/2017	afternoon	11/14/2017	4:30 PM	11/14/2017	5:45 PM	Case #9? 7 seems more likely	This is a different patient but the tube said case #9? It might have been meant to be a 7? None the less this was a different patient then the surgery from in the morning. Also tranported in Hank's buffer. Ghadi picked up from the hospital on his way back to campus. saved tissue RNA, crypts frozen and crypt RNA. Sample was several smaller strips and was a bit hard to get a good dissection.	patient had a malignant polyp but it was removed and this was the normal that was left	CASE #7 Normal colon with adenocarcinoma (mucinous type) MMR: no loss

13 *used in Thai's Paper	Tumor	J-T13	12/5/2017	Morning	12/5/2017	3:00 PM	12/5/2017	4:00 PM	Case #8	Shu contacted me that tissue was ready for pick-up at 1pm. I brought her additional transport media but these samples were just in RPMI. Lot of small tumor chunks - but a large amount! Saved some tissue in FBS/DMSO, as well as RNA prep from tissue and isolated cells. Froze down cells for plating as organoids later.		CASE #8 Normal colon with adenocarcinoma MMR: no loss 69/female
14	Normal	J-TN14	12/5/2017	Morning	12/5/2017	3:00 PM	12/5/2017	4:30 PM	Case #8	Picked-up along with corresponding tumor. Lot of normal tissue! Froze down tissue RNA, isolate cell RNA and cells for organoid plating later.		
15	Tumor	J-T15	1/19/2018	Morning	1/19/2018	2:30 PM	1/19/2018	2:45	Case #10	Tissue was ready at 1pm and Ghadi was able to pick up at the hospital for me, and brought to be in a bag on ice. Started on Tumor right away - pretty small amount of tumor but did save some who tumor for PDX, made tissue RNA, and froze some single cell for organoids. Was in transport media dated last 12/5 but should still be good.		
16	Normal	J-N16	1/19/2018	Morning	1/19/2018	2:30 PM	1/19/2018	3:30	Case #10	Tissue was picked up with the tumor by Ghadi. Two larger pieces of tissue, with muscle still attached. Used one of the pieces and allowed Sam to practice a dissection. Made some tissue RNA, crypt RNA, and organoids preps.		** Check with SHU what case # this is -- might have gotten skipped **
17	Tumor	J-T17	2/2/2018	Morning	2/2/2018	12:30 PM	2/2/2018	1:40 PM	Shu wrote on the tube Case #9?	Came in late and picked up at the hospital on my way in. Very small tumor sample, did save some for pdx, tissue RNA, and some cells. Not very much though. Transport media was 12.5.17.	95 year old patient!	CASE #9 Normal colon with adenocarcinoma MMR= MLH1 & PMS2: loss, MSH2 &MSH6: intact BRAF mutation, detected
18	Normal	J-N18	2/2/2018	Morning	2/2/2018	12:30 PM	2/2/2018	2:15 PM	Shu wrote on the tube Case #9?	Picked up with partner tumor. Large size of normal and was able to save tissue RNA, crypt RNA, and many vials of cells.	95 year old patient!	
19	Normal	J-N19	4/10/2018	Morning	4/10/2018	5:30 PM	4/10/2018	7:00 PM	Shu wrote Case #10 - which we already had?	Shu emailed me in the morning and said samples were ready (all three), I couldn't pick them up today but Ghadi was able to do so. Had Hanks Media (exp. 3.19). Decent sized normal, had a large blood clot/bruised area (black/blue-ish) made sure to fully remove this section. Collect Tissue RNA, Cell RNA, and froze organoids.		
20	Tumor	J-T20	4/10/2018	Morning	4/10/2018	5:30 PM	4/10/2018	7:15 PM	Shu wrote Case #10 - which we already had?	Part of the previous sample. Also in Hanks Media. Decent sized tumor chunks, froze down 5 vials with FBS/DMSO. Also collected tissue RNA, Cell RNA, and froze organoids.	CASE #10 Normal colon with adenocarcinoma. MMR = MLH1: loss, MSH2: no loss, MSH6: no loss, PMS2: loss. BRAF mutation, detected. NRAS mutation- NRAS Exon 4 mutation and KRAS mutation, not detected.	
21	Normal	J-N21	4/10/2018	Morning	4/10/2018	5:30 PM	4/10/2018	7:45 PM	Case 11	Part of the previous sample delivery, had my old mediat though (transport media dated 12.5.17 - not idea). Pretty large normal, collected tissue RNA, Cell RNA, and froze organoids.	THIS WAS ACTUALLY NOT NORMAL - Was treated with Chemotherpy. Waiting to hear back from her what chemotherapy was used.	CASE #11 Status post chemotherapy for colon adenocarcinoma case (specimen of carcinoma not available) Specimen submitted was normal colon, No residual carcinoma
22	Normal	J-N22	5/10/2018	Midday	5/10/2018	5:00 PM	5/10/2018	6:30 PM	Case 12	Ghadi picked up both samples today and they were in Hanks Buffer. Shu texted and said the samples would be ready at 4pm. Ghadi drove and picked them up. Large amount of normal. Collected tissue RNA, Crypt RNA, and frozen organoids.		After two attempts have not had successful organoid growth
23 *used in Thai's Paper	Tumor	J-T23	5/10/2018	Midday	5/10/2018	5:00 PM	5/10/2018	6:00 PM	Case 12	Ghadi picked up both samples today and they were in Hanks Buffer. Shu texted and said the samples would be ready at 4pm. Ghadi drove and picked them up. Decent sizes tumor. Collected frozen tissue for PDX, tissue RNA, cell RNA, and frozen organoids.	CASE #12 Normal colon with adenocarcinoma with liver metastasis MMR: no loss. KRAS mutation, detected (c.38G>A(p.G13D). NRAS Exon 4 mutation, not detected. 83/male	
24	Normal	J-N24	7/3/2018	Morning	7/3/2018	2:00 PM	7/3/2018	4:30 PM	Case 13	I drove to the hospital to pick up this sample (Ghadi is on vacation). Decent sized normal tissue but noted the normal was pretty bloody.Collected tissue RNA, Cell RNA, and frozen organoids.		
25	Tumor	J-T25	7/3/2018	Morning	7/3/2018	2:00 PM	7/3/2018	5:30 PM	Case 13	I drove to the hospital to pick up this sample (Ghadi is on vacation). Decent sized tumor but noted it was very bloody and mushy (compared to the tumor I often get that are more fiberous).Collected frozen tissue for PDX, tissue RNA, Cell RNA, and frozen organoids.	48 year old patient with tubular adenoma with dysplasia said Shu. She was excited about this one and thought it was good.	CASE #13 Normal colon with tubular adenoma. MMR: no loss One attempted did not yield good organoids
26	Normal	J-N26	7/5/2018	Morning	7/5/2018	2:00 PM	7/5/2018	5:35 PM	Case 14	I drove to the hospital to pick-up samples. This was a very large normal so I did a joint prep with Linzi (aka cut all tissue in half and she prepped half). Was in colon transport media dated in May. Collected tissue RNA, Cell RNA, and frozen organoids.Noted softer less defined muscle.		
27 *used in Thai's Paper	Tumor	J-T27	7/5/2018	Morning	7/5/2018	2:00 PM	7/5/2018	6:30 PM	Case 14	I drove to the hospital to pick-up samples. This was a very large tumor so I did a joint prep with Linzi (aka cut all tissue in half and she prepped half). Was in colon transport media dated in May. Collected frozen tissue for PDX, tissue RNA, Cell RNA, and frozen organoids. Not very bloody.	She mentioned this was a large tumor!	CASE #14 Normal and cecum adenocarcinoma MMR: no loss. 72/male

28	Normal	J-N28	10/17/2018	Morning	10/17/2018	11:40 AM	10/17/2018	12:30 PM	Case 15	I drove to hospital and picked up from Shu. Linzi helped with the prep. Was in old transport media. Collected tissue RNA, cell RNA, and froze down cells.		
29	Tumor	J-T29	10/17/2018	Morning	10/17/2018	11:40 AM	10/17/2018	12:30 PM	Case 15	Picked up with normal. Linzi helped. Prep was in Hanks buffer. Collected frozen tissue for PDX, tissue RNA, cell RNA, and froze organoids. Not bloody and actually really large tumor! Lot of pieces.	Tech accidentally put this in formalin and it was about a minute before Shu realized. She did 10x rinses in Hanks but there is a chance these might just not grow.	CASE #15 Normal colon and adenocarcinoma. MMR= MLH1 & PMS2: loss, MSH2 & MSH6: no loss. BRAF mutation, detected (c.1799T>A (p.V600E)
30	Normal	J-N30	2/14/2019	Morning	2/14/2019	11:30 AM	2/14/2019	2:15 PM	?Not labeled on tube must be 16	Shu contacted me in the morning and said tissue would be ready around 11am. Ghadi was able to pick up the tissue. Linzi helped me with the dissection and we did it side by side. The normal was one large piece. Decent amout of blood, some blood clots. During isolation tissue/cells were bloody.		
31	Tumor	J-T31	2/14/2019	Morning	2/14/2019	11:30 AM	2/14/2019	1:45 PM	?Not labeled on tube must be 16	Pick-ed up with matching normal. Shu wrote "Dysplastic adenoma or carcinoma". Tissue was softer, mushier and pretty bloddy.	CASE 16 Normal colon and adenoma with focal high grade dysplasia MMR: no loss	
32	Normal	J-N32	5/2/2019	mid day	5/2/2019	4:20 PM	5/2/2019	5:00 PM	Case 17	Two very large nice sections of normal with healthy nice smooth looking mucosal layer. Linzi helped with dissection. Normal cell tubes labeled as "bulk". Tissue was ready at 2pm ish and Ghadi was able to pick it up after his class and dropped it off in lab.		
33	Tumor	J-T33	5/2/2019	mid day	5/2/2019	4:20 PM	5/2/2019	4:45 PM	Case 17	Very small tumor, did not save for PDX. Not bloody at all, more fibrousy/cartilage like, several smaller chunks, still got a pretty decent cell yield.	CASE 17 Normal colon and adenocarcinoma MMR: no loss	
34	Normal	J-N34	5/7/2019	Mid day	5/7/2019	4:00 PM	5/7/2019	5:40 PM	Case 18	Linzi picked up with the corresponding tumor, tissue was very fatty, the stroma/submucosa was rather odd and harder to dissect cleanly. Did not get much of a pellet at all and had a lot of fat just floating on the top. Isolate was very milky colored. proceeded anyways, but never ended up getting a good yeild. Maybe this tissue just neededto be dissociated for much longer with more DTT to get thru the musuc and fat? Hard to say. Saved tissue RNA and just a few vials of organoids.		
35	Tumor	J-T35	5/7/2019	mid day	5/7/2019	4:00 PM	5/7/2019	5:20 PM	Case 18	Linzi picked up samples from the hospital, tumor was very cartilage-like and many smaller chunks. Not very bloody. Got pretty good yeild, froze down tissue RNA, tissue for PDX, and organoids	35 year old - she called this an interesting case. On the tube it looks like she wrote cecum carcinoma?	CASE 18 Normal colon and adenocarcinoma. (35 y) MMR= MSH2: loss, MSH6: no loss, MLH1 & PMS2: no loss. MSI: high -- BAT-25, BAT-26, MONO-27, NR-21, NR-24: unstable
36	Normal	J-N36	8/20/2019		8/21/2019	7:30 AM	8/21/2019	3:00 PM	Case 19	Linzi picked up samples the next morning from Shu. Normal colon tissue looked normal		
37	Tumor	J-T37	8/20/2019		8/21/2019	7:30 AM	8/21/2019	6:00 PM	Case 19	Linzi picked up samples the next morning from Shu. This sample was labeled adenoma and was small bloody pieces of tissue. Linzi tried to dissociate to single cells for FACS sorting but very thick (did not use DNase) so not sortable. Forgot to freeze pieces for PDX.		
38	Normal	J-N38	8/20/2019		8/21/2019	7:30 AM	8/21/2019	3:00 PM	Case 20	Linzi picked up samples the next morning from Shu. Normal colon mucosa was black in color, very obvious and different than what I have seen before.		
39	Tumor	J-T39	8/20/2019		8/21/2019	7:30 AM	8/21/2019	6:00 PM	Case 20	Linzi picked up samples the next morning from Shu. Linzi tried to dissociate to single cells for FACS sorting but cell suspension was very thick (did not use DNase) so not sortable. Forgot to freeze pieces for PDX		
40	Normal	J-N40	9/12/19		9/13/19	7:15 AM	9/13/19	12:00 PM	Case 21	Linzi picked up samples the next morning and left on ice in deli fridge until dissection. Normal colon had thick muscle and was very fatty/yellow in submucosa. Pieces of mucosa blended with trizol for RNA (had to precellys twice to get tissue to fully blend). Froze down p0 cells		
41	Tumor	J-T41	9/12/19		9/13/19	7:15 AM	9/13/19	5:00 PM	Case 21	Linzi picked up samples the next morning and left on ice in deli fridge until dissection. Tumor was very white and cartilage-like and tons of smaller chunks. Saved chunks in 90%FBS/10%DMSO for PDX (3), chunks snap frozen in cryovials and stored at -80C (3), and chunks in trizol for bulk RNA (2), and froze down p0 cells.	Shu said this was aggressive cancer via text, on 10/9/19 when picking up case 22 tissue, Shu mentioned that the aggressive case was interesting becuase it expressed all the stem markers she probed for and none of the differentiated markers (most cancer cases are differentiated and form gland structures) and did not form gland structures. Shu was interested if the organoids grew differently (faster?) but we have not yet tested	
42	Normal	J-N42	10/9/19	early AM?	10/9/19	11:45 AM	10/9/19	1:00 PM	Case 22	Linzi picked up tissue from St. Joseph's as soon as it was available. Normal sample contained two large pieces of colon tissue. Dissociation went well, saved p0 cells for RNA and freezing. Also saved mucosa for tissue RNA. Collected 9.3 million p0 cells from mucosa.		
43	Tumor	J-T43	10/9/19	early AM?	10/9/10	11:45 AM	10/9/10	1:50 PM	Case 22	Linzi picked up tissue from St. Joseph's as soon as it was available. Received lots of smaller chunks of cartilage-like tissue. Not bloody. Saved pieces for PDX and for tissue RNA. Used DNase and filtered well. Collected 11 million total p0 cells from tumor.		

Table B.3: ***Catalog of current vials in biobank***

From St. Joseph Hospital Orange

Organoid Name	# of vials
Normal Colon	
J-N2P0 Organoid	3
J-N4P0 Organoid	8
J-N4P1 Organoid	5
J-N6P0 Organoid	7
J-N6P1 Organoid	5
J-N8P0 Organoid	5
J-N10P0 Organoid	8
J-N11P0 Organoid	3
J-N12P0 Organoid	3
J-N14P0 Organoid	4
J-N16P0 Organoid	5
J-N18P0 Organoid	6
J-N19P0 Organoid	8
J-N21P0 Organoid	12
J-N22P0 Organoid	14
J-N24P0 Organoid	8
J-N26P0 Organoid	13
J-N26P1 Organoid	4
J-N26P3 Organoid	8
J-N26P4 Organoid	7
J-N26P5 Organoid	5
J-N28P0 Organoid	15
J-N30P0 Organoid	16
J-N32P0 Organoid	13
J-N34P0 Organoid	4
J-N36P0 Organoid	5
J-N38P0 Organoid	5
J-N40P0 Organoid	7
J-N42P0 Organoid	9

Tissue	# of vials
J-T1 Tissue	2
J-T5 Tissue	3
J-T13 Tissue	6
J-T15 Tissue	2
J-T17 Tissue	2
J-T20 Tissue	5
J-T23 Tissue	4
J-T25 Tissue	4
J-T27 Tissue	6
J-T29 Tissue	8
J-T31 Tissue	4
J-T35 Tissue	3
J-T41 Tissue	3
J-T43 Tissue	3

Organoid Name	# of vials
Colon Tumor	
J-T1P0 Organoid	3
J-T1P1 Organoid	9
J-T5P0 Organoid	3
J-T7P0 Organoid	8
J-T9P0 Organoid	4
J-T9P2 Organoid	3
J-T13P0 Organoid	9
J-T13P1 Organoid	5
J-T13P2 Organoid	3
J-T15P0 Organoid	2
J-T17P0 Organoid	2
J-T20P0 Organoid	5
J-T20P1 Organoid	3
J-T23P0 Organoid	7
J-T23P3 Organoid	1
J-T23P4 Organoid	1
J-T23P8 Organoid	2
J-T23P9 Organoid	2
J-T25P0 Organoid	7
J-T27P0 Organoid	12
J-T27P1 Organoid	1
J-T27P2 Organoid	5
J-T27P3 Organoid	1
J-T29P0 Organoid	15
J-T31P0 Organoid	12
J-T33P0 Organoid	6
J-T35P0 Organoid	15
J-T37P0 Organoid	6
J-T39P0 Organoid	8
J-T41P0 Organoid	5
J-T43P0 Organoid	9

From UCIMC

Name	# of vials
Normal Colon Organoid	
N1P0 Organoid	2
N2P0 Organoid	2
N3aP0 Organoid	3
N3bP0 Organoid	3
N3aP1 Organoid	3
N3bP1 Organoid	3
N4aP0 Organoid	3
N4bP0 Organoid	2
N5P0 Organoid	3
N6P0 Organoid	2
N7P0 Organoid	1
N8P0 Organoid	2
N9P0 Organoid	3
N10P0 Organoid	2
N12P0 Organoid	2
N12P1 Organoid	2
N12P2 Organoid	2
N13P0 Organoid	2
N13P1 Organoid	2
N21P0 Organoid	2
N22P0 Organoid	1
T23P0 Organoid	6
N24P0 Organoid	4
Colon Tumor Organoid	
T25P0 Organoid	6
UCIMC Tumor Tissue	
T23 Tissue	3

REFERENCES

1. Simian, M. & Bissell, M. J. Organoids: A historical perspective of thinking in three dimensions. *J. Cell Biol.* **216**, 31–40 (2017).

2. Lancaster, M. A. & Knoblich, J. A. Organogenesis in a dish: Modeling development and disease using organoid technologies. *Science (80-.).* **345**, 1247125–1247125 (2014).

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

4. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003–1007 (2007).

5. Kaushik, G., Ponnusamy, M. P. & Batra, S. K. Concise Review: Current Status of Three-Dimensional Organoids as Preclinical Models. *Stem Cells* **36**, 1329–1340 (2018).

6. Sato, T. *et al.* Long-term Expansion of Epithelial Organoids From Human Colon, Adenoma, Adenocarcinoma, and Barrett's Epithelium. *Gastroenterology* **141**, 1762–1772 (2011).

7. Van De Wetering, M. *et al.* Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell* **161**, 933–945 (2015).

8. Drost, J. *et al.* Sequential cancer mutations in cultured human intestinal stem cells. *Nature* **521**, 43–47 (2015).

9. Clevers, H. Modeling Development and Disease with Organoids. *Cell* **165**, 1586–1597 (2016).

10. Fatehullah, A., Hui Tan, S. & Barker, N. Organoids as an in vitro model of human development and disease. *Nat. Cell Biol.* **18**, 246–254 (2016).

11. Driehuis, E., Kretzschmar, K. & Clevers, H. Establishment of patient-derived cancer organoids for drug-screening applications. *Nat. Protoc.* 1–30 (2020). doi:10.1038/s41596-020-0379-4

12. Pleguezuelos-Manzano, C. *et al.* Establishment and Culture of Human Intestinal Organoids Derived from Adult Stem Cells. *Curr. Protoc. Immunol.* **130**, (2020).

APPENDIX C

GECO: Gene expression clustering optimization app for non-linear data visualization of patterns

Text adapted from work In Revision:

Habowski, A.N., T.J. Habowski, and M.L. Waterman. GECO: Gene expression clustering optimization app for non-linear data visualization of patterns. *In Revision.*

ABSTRACT

Background

Due to continued advances in sequencing technology, the limitation in understanding biological systems through an -omics lens is no longer the generation of data, but the ability to analyze it. Importantly, much of this rich -omics data is publicly available waiting to be further investigated. Although many code-based pipelines exist, there is a lack of user-friendly and accessible applications that enable rapid analysis or visualization of data.

Results

GECO (Gene Expression Clustering Optimization; www.theGECOapp.com) is a minimalistic GUI app that utilizes non-linear reduction techniques to rapidly visualize expression trends in many types of biological data matrices (such as bulk RNA-seq or proteomics). The required input is a data matrix with samples and any type of expression level of genes/protein/other unique ID. The output is an interactive t-SNE or UMAP analysis that clusters genes (or proteins/unique IDs) based on expression patterns across samples enabling visualization of trends. Customizable settings for dimensionality reduction, normalization, along with visualization parameters including coloring and filters, ensure adaptability to a variety of user uploaded data.

Conclusions

This local and cloud-hosted web browser app enables investigation of any -omic data matrix in a rapid and code-independent manner. With the continued growth of available -omic data, the ability to quickly evaluate a dataset, including specific genes of interest, is more important than ever. GECO is intended to supplement traditional statistical analysis methods and is particularly useful when visualizing clusters of genes with similar trajectories across many samples (ex: multiple cell types, time course, dose response).

Users will be empowered to investigate their -omic data with a new lens with the potential to uncover genes of interest and previously unseen patterns.

BACKGROUND

The next generation sequencing revolution has resulted in the production of an enormous amount of data [1,2]. While much of this data is available in public repositories or supplementary manuscript material, there remains a bottleneck in a broader public analysis of the data. Thus, the ability to further our understanding of the world thru an -omics lens is limited not by the production of data, or even its accessibility, but by our ability to analyze it. Although others have developed pipelines to aid in re-analyzing publicly available data [3], it is important to develop analysis pipelines for quick and easy use of already available data matrices to encourage their broad utilization. Currently, there are numerous bioinformatic pipelines to statistically analyze -omic data, however the majority are dependent on being able to run code, an expertise lacking for many biologists. Thus there is a great need for GUI (graphical user interface) based programs as well [4–6]. An easy to use data analysis tool which also facilitates data exploration, can lead to new insights. Additionally, since many publications are accompanied by already processed data matrices, a rapid and user-friendly way to analyze these data-matrices is informative and necessary.

Many classic differential expression analyses result in outputs of tables of genes with statistics, volcano plots, or heat maps showing strongly differentially expressed genes between samples [7–9]. Although these analyses are useful, they also make it difficult to visualize the data globally and identify cohorts of genes that might be behaving in a similar manner across samples. Identifying these cohorts of genes can lead to investigation of impacted gene programs or classes of ontology that might be overlooked when sorting through genes by significance. Additionally, many of the bulk RNA-seq pipelines for differentially expressed genes cater to paired analysis – generally between a control and experimental samples. This can make comparisons of time courses or across a cohort of samples challenging. Although there have been specialized pipelines for the analysis of time courses, in many cases these pipelines are still outperformed by pairwise analysis [10,11]. There is a need for analyses that can visualize gene patterns and trends across all samples at the same time.

The increased quantity of sequencing data and the rise of single cell sequencing data has been reliant on more complex bioinformatic analyses, which has further encouraged a merge of the fields of computer science and biology [2,12,13]. Several unsupervised approaches have been borrowed from machine-learning such as PCA (Principal component analysis), t-SNE (t-distributed stochastic neighbor embedding), and UMAP (Uniform manifold approximation and projection). PCA is a mathematical approach that uses a linear dimensionality reduction method to investigate data relatedness [14,15]. In essence, PCA reduces the data to eigenvectors showing how related data points are to one another. The dominant two principle components can usually separate data based on the largest variance. Although PCA can rapidly reduce complex data, visualizing highly dimensional data with PCA is limited [14,15]. Non-linear dimensionality reduction using probabilistic approaches, such as t-SNE [16] and UMAP [17], better enable visualization of complex-multidimensional data in a low dimensional space. Although these techniques were developed by computer scientists for machine learning applications, they have found a prominent home in analyzing the growing expanse of single cell -omic data [18,19]. These non-linear dimensionality reduction techniques better preserve the complexity of the data and importantly, the closeness of data points can be used to draw conclusions on the relatedness between these points. Previous publications have shown the value and usefulness of non-linear over linear dimension reduction and the ability to customize and optimize the parameters [19–21].

Here we present GECO (Gene Expression Clustering Optimization), a minimalistic GUI app that utilizes non-linear reduction techniques to visualize expression trends in biological data matrices (such as bulk RNA-seq or proteomics). The required input is a data matrix with samples and any type of expression level of genes/protein/other unique ID. The output is an interactive t-SNE or UMAP that clusters genes (or proteins/unique IDs) based on expression patterns across samples enabling visualization of trends. Each data point on the plot is one gene/protein/other unique ID with the expression pattern across all samples used to determine its position and location relative to other data points. Features of GECO include:

- User-friendly Streamlit run app accessed through a cloud-hosted website (no code, downloading, or installation needed).
- Option to run Streamlit locally on user's computer with network host capability for temporary sharing.
- Customizable parameters for t-SNE and UMAP generation (optional PCA initial reduction).
- Optional GPU driven clustering for t-SNA and UMAP generation.
- Save function for t-SNE and UMAP enabling re-opening of a saved interactive session (important for stochastic analysis like t-SNE and UMAP where each run will yield variation and a different cluster shape).
- Flexible data type input.
- Optional normalization techniques, filtering, and threshold cutoff.
- Incorporation of curated marker genes, gene searching, and highlighting function.
- Autogenerated bar plot, correlation clustermap (with significance calculated), and heatmap expression of selected genes.
- Generation of downloadable gene list based on clustering and filtering.
- Large selection of colors, inversion and log of scale functions, and .png generation of plots to facilitate user flexibility based on needs/preferences.

IMPLEMENTATION

Architecture

All code for GECO was written in Python 3.7 and uses Streamlit (https://www.streamlit.io/) and Plotly (https://plotly.com/) for GUI and interactive data visualization. Streamlit is a new open source app framework and was chosen for its relative simplicity to implement a graphical interface to the python back end code. All source code, install files, and install directions for GECO are available on github (https://github.com/starstorms9/geco). GECO is intended to be usable without any programming knowledge. A cloud-hosted website version of GECO thru *Streamlit for Teams* (currently in beta form) can be accessed at www.theGECOapp.com. In order to run GECO locally, step-by-step installation instructions are available in on github. README documentation is provided towards the end of this appendix, including step-by-step instructions for analysis and all optional features (also available on github).

User Interface

The Streamlit interface utilizes three main tabs: 1- a landing page that documents usage (README), 2- a data loading, processing, and dimensionality reduction page (Figure C.1), and 3- a reduced data visualization page (Figure C.2). On each page, the sidebar provides access to the majority of the controllable parameters and the main screen shows the results. At the top of the sidebar interface is an assigned Session ID number which the user should save because uploading this ID number later allows the user to re-access the current session including the uploaded datasets and saved plots. Sharing this Session ID is also an easy way to allow collaborators to explore shared datasets.

In the data visualization tab, the reduced dimensionality data can be investigated with a variety of customizable options. The visualization options were developed by investigating various datasets with known trends and features and finding ways to highlight these features most clearly. This strategy helped to facilitate identifying similar trends in new and unexplored datasets. For example, normalizing to a specific control type and selecting entries that have a high fold change relative to that control quickly highlights entries that were most affected by a given condition. The visualization tab also allows a user to select entries of interest and generate a bar graph, correlation clustermap, and heatmap to compare that subset of entries to each other.

Data Input and Output

In order to readily accommodate a wide variety of input data from disparate sources, GECO has a system for automatic data cleaning to ensure that the data put into the dimensionality reduction algorithms is properly formatted. During testing, any issues encountered with loading test datasets were used to develop automatic solutions. For example, it was found that many datasets contain a significant quantity of entries with all 0's or entries with some non-numeric characters, entries which can distort the output of the processing algorithms. GECO provides simple options to remove these entries. Additionally, naming conventions of samples and bio reps is highly varied and so a system was implemented to recognize and group similarly named samples into coherently labeled sets.

After the data has been uploaded and processed through the dimensionality reduction algorithms, it can be saved and then visualized. Options are also available to manually enter or upload a comma separated list of entries of particular interest which are then marked prominently on the plot so that similar entries can be readily identified. Once specific groups of interest have been identified, they can then be downloaded along with their relevant reduced dimensionality parameters for further analysis externally.

Algorithms

Three core algorithms were implemented for dimensionality reduction: PCA, t-SNE, and UMAP. Existing implementations of these algorithms were available as open source python modules. Due to the generally long processing time and high degree of parallelization possible with the t-SNE algorithm in particular, a CUDA based implementation called t-SNE-CUDA [22] was used for GECO. This t-SNE implementation is approximately ~50 times faster than standard CPU based algorithms and allows for rapid exploration of the effects that various hyperparameters such as perplexity and learning rate have on the final output. However, GPU enabled implementations of t-SNE are currently only available on Linux based systems and so a backup CPU based implementation is automatically switched to when the program is run on other systems. PCA alone was insufficient to visualize the data clearly but it is used as a preprocessor before the t-SNE algorithm runs in order to reduce the number of variables and make the calculation time for the t-SNE tenable. UMAP is another popular non-linear reduction technique and is implemented here as it captures global correlations and structure more accurately compared to t-SNE which primarily focuses on local structure. UMAP performance speed also far outperforms t-SNE (when run without t-SNE-CUDA) and is recommended for faster dimensionality reduction times.

There are two important normalization options that can be applied to the data before running the dimensionality reduction algorithm: 1) normalize per row and 2) normalize to type. To normalize per row, every entry is scaled down by the sum of that row. This strategy ensures that the algorithm focuses only on the relative pattern for a given entry instead of just the overall magnitude of that entry. For gene expression data in particular, without normalizing per row the resulting reduced dimensionality plots are often simply aligned according to the overall expression levels and ignore more interesting but subtle expression patterns that are shared by genes that are expressed in similar ways (Figure C.3). Likewise, normalizing every entry to a selected type (e.g. control) prior to reducing the data ensures that the reduction algorithm focuses on the patterns that change relative to the control instead of looking at global patterns.

RESULTS AND DISCUSSION

Example Usage 1: Colon Crypt Cell Types

The inner layer of the colon contains epithelial cells in a crypt structure including proliferating stem cells. These stem cells give rise to daughter cells which then further differentiate into mature cells. Previously, bulk RNA-seq was performed on sorted crypt populations including stem cells, immediate daughter cells (AbsPro, SecPDG), and more mature differentiated cells (Tuft, Ent, and EEC) [23]. A UMAP plot generated with GECO from this dataset and colored by assigned type shows a gene expression trajectory of stem related genes which transition to those associated with more differentiated cell types (Figure C.4a). All cell types have assigned genes that are well expressed (Figure C.4b). Coloring the data points (genes) by stem expression reveals the clustering of highly expressed stem-associated genes to one region (Figure C.4c). Further, coloring by stem enrichment shows a smaller region where the genes are highly expressed in stem, and less expression in other cell types (Figure C.4d). Filtering the genes displayed on the plot with a minimum expression level (500 normalized counts) further shows the region of genes enriched in stem cells (Figure C.4e). Zooming in on a region of interest and adding a filter for a 1.5-fold cutoff for stem enrichment reveals clustering of stem-associated genes (Figure C.4f). Stem cell marker genes (n=27) previously identified using traditional statistical differential expression methods (DESeq2) are clustered in this region as well [23].

In this example, GECO enables visualization of genes that are enriched in different cell types in the colon crypt. The clustering of genes assigned to each cell type follow the natural trajectory of stem → daughter cells → differentiated cells. Previously identified stem markers overlap well with genes remaining following filtering of those enriched in stem cells. Although the GECO plot in Figure C.4f displays many genes of potential interest, additional stringent filtering could also be applied to decrease the data points.

However, in this case GECO is able to rapidly reveal the trajectory of gene expression changes in these cell types and identify stem-associated genes.

Example Usage 2: Infection Time Course of F. nucleatum

Fusobacterium nucleatum (*F. nucleatum*) is a pathogen that frequently contributes to periodontal diseases. Previous work investigated the impact of *F. nucleatum* infection on human gingival fibroblasts using a time course of bulk RNA-seq [24]. A GECO generated UMAP plot colored by assigned type (Figure C.5a) reveals the genes are clustered tightly based on expression at different time points during injection (0, 2, 6, 12, 24, and 48hr). This is further evident when the continuous color setting is used and a clear trajectory from 0hr (control) to 48hr post infection emerges (Figure C.5b). Altering the color setting to expression fold enrichment for 0hr (Figure C.5c) versus 48hr (Figure C.5d) highlights the genes that are most highly expressed at 0hr (top left corner of the plot), compared to genes that are more highly expressed at 48hrs (bottom right corner of the plot). Genes can be identified that are enriched at specific time points or that gradually increase or decrease over the duration of the infection. Figure C.5e shows several examples of selected genes graphed using GECO where each data point was a bio-replicate from the uploaded dataset.

To identify a small cohort of genes elevated early during *F. nucleatum* infection, the UMAP plot was colored based on enrichment at 6hr (Figure C.6a) and then restricted to a minimum 1.5-fold cutoff or greater (Figure C.6b). This highlights a small region of genes that are elevated in the first several hours of infection compared to all other time points. This selection was further filtered to find ~25 highly expressed genes (minimum expression cutoff of 150). This gene list was then printed to the screen and four genes with trends of interest are marked with red circles on the UMAP plot and displayed in GECO-generated bar graphs. The gene list includes *CXCL3* and *ICAM1* which are sharply induced at 2hr from the start of infection and then gradually decline, *TWIST2* which peaks at 6hr, and *NINJ1* which is elevated at 2 and 6hr. Displaying these genes, along with those displayed in Figure C.5, in a GECO generated correlation clustermap shows a significant correlation between this cohort of 4 genes (*CXCL3, ICAM1, TWIST2,* and *NINJ1)* (Figure C.7a). *COL1A1* and *SOD2* show inverse trends in Figure C.5e and as expected the clustermap reveals they are significantly anti-correlated (Figure C.7a).

GECO is therefore useful to visualize gene expression changes across multiple samples such as a time course, and can be used to define cohorts of genes with matching gene expression trends. In this dataset there is a clear trajectory of genes that are elevated in the 0hr-control samples or at each time point (ex: *TNFAIP3*), whereas other genes gradually change over the time course and peak at one datapoint (ex: *COL1A1* and *SOD2*). In the later case, these are often genes that are difficult to uncover with traditional statistical differential expression analysis (particularly paired analysis), but when looking at global trends such as those that GECO enables, these genes can be uncovered along with other genes that behave in a similar pattern.

CONCLUSION

GECO is a minimalistic Streamlit GUI app that utilizes non-linear reduction techniques to visualize expression trends in biological data matrices. This app enables investigation of any -omic data matrix in a rapid and code-independent manner. With the continued growth of available -omic data, the ability to quickly evaluate a dataset, including specific genes of interest, is more important than ever. GECO is intended to supplement more traditional statistical analysis methods and is particularly useful when visualizing clusters of genes with similar trajectory across many samples (ex: multiple cell types, time course, dose response). With a variety of options for dimensionality reduction, normalization methods, and visualization (coloring), along with thorough step-by-step instructions, users will be empowered to investigate their -omic data with a new lens with the potential to uncover genes of interest and previously unseen patterns.

GECO AVAILABILITY AND REQUIREMENTS

Project name: GECO

Project home page: www.theGECOapp.com & https://github.com/starstorms9/geco

Operating system(s): Linux, Windows, Mac

Programming language: Python 3.7+

Other requirements: Streamlit, Plotly, Scipy, Pandas, Seaborn, Umap-Learn, t-SNE-CUDA, numpy

License: MIT License

Any restrictions to use by non-academics: none

GECO README

Welcome to GECO (Gene Expression Clustering Optimization; www.TheGECOapp.com), the straightforward, user friendly Streamlit app to visualize and investigate data patterns with non-linear reduced dimensionality plots. Although developed for bulk RNA-seq data, GECO can be used to analyze any .csv data matrix with sample names (columns) and type (rows) [type = genes, protein, any other unique ID]. The output is an interactable and customizable t-SNE/UMAP analysis and visualization tool. The visualization is intended to supplement more traditional statistical differential analysis pipelines (for example DESeq2 for bulk RNA-seq) and to confirm and/or reveal new patterns.

If questions or issues arise please contact Amber Habowski at Habowski@uci.edu or post an issue on the github issues page here.

GECO Video demonstration in 3 minutes

***Quick Guide to Getting Started* ---**

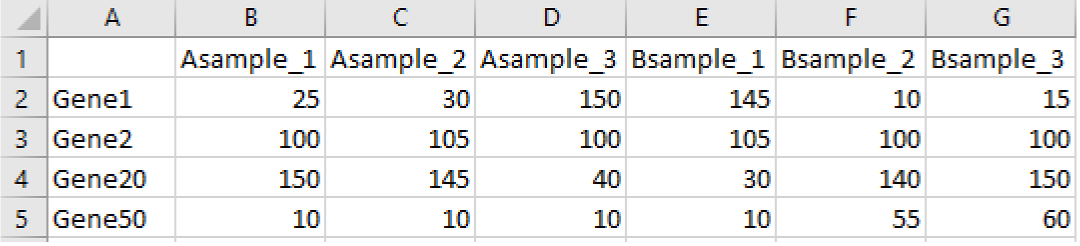
1. Upload data file and verify that GECO has interpreted the sample names/bio-replicates.
2. Select the reduction parameters to be used for the analysis.
3. Click the 'Run UMAP/t-SNE reduction' button at the bottom of the parameters sidebar.
4. Once a plot is generated, save it by clicking ‘Save data file’ at the bottom of the sidebar.
5. Proceed to the ‘Plot reduced data’ mode to visualize the saved plot.

File Upload **--**

(required) The Data Matrix:

- Must be supplied as a .csv file.
- The first column should contain the unique IDs for this dataset (genes, isoforms, protein, or any other identifier) which will be renamed ‘geneid’ in the program. Each unique ID should have ‘expression’ data listed in each row that corresponds to each sample.
- Sample names must be listed at the top of each columns, with biological replicates being indicated by ‘_#’ following each sample name. Biological replicates are averaged during the analysis and the number of biological replicates does not need to match between samples. For example, for two samples (‘Asample’ and ‘Bsample’) with three biological replicates each the column names should be assigned as shown in the example below.
    - If no ‘_#’ columns are found for a given sample name, but there are duplicated column names, they will automatically have sample numbers appended.
- The file should not have any index column (1,2,3,4) or other columns with additional information.
- 'NA' entries will be interpreted as unknowns and those entire rows will be removed
    - Also any of: ['-1.#IND', '1.#QNAN', '1.#IND', '-1.#QNAN', '#N/A N/A', '#N/A', 'N/A', 'n/a', 'NA', '', '#NA', 'NULL', 'null', 'NaN', '-NaN', 'nan', '-nan', '']

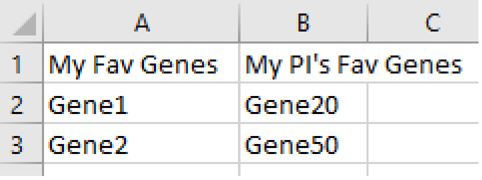
   - If there is a need to include rows with non-numeric values, the NA values must be imputed manually and replaced. For example, this can be achieved by "zero-ing" non-numeric values , replacing NA values with average of samples, etc.



	A	B	C	D	E	F	G
1		Asample_1	Asample_2	Asample_3	Bsample_1	Bsample_2	Bsample_3
2	Gene1	25	30	150	145	10	15
3	Gene2	100	105	100	105	100	100
4	Gene20	150	145	40	30	140	150
5	Gene50	10	10	10	10	55	60

(optional) Curated Markers List.

- Must be supplied as a .csv file.
- Each column should start with a descriptive title (that will appear in a drop-down list). Below the title will be a list of unique IDs (that overlap with the provided data matrix). A minimal example is provided below.
- Multiple curated lists can be provided by listing them next to each other, one per column. Do not skip columns, and do not use different excel sheet tabs.
- GECO is case sensitive, so make sure capitalization of gene entry text is consistent with the provided Data Matrix.



	A	B	C
1	My Fav Genes	My PI's Fav Genes	
2	Gene1	Gene20	
3	Gene2	Gene50	

***Usage Instructions* ---**

Session ID Information:

At the top of the sidebar is information on the current session ID and a box to input a previously saved session ID. This enables a previous session (including uploaded data and saved plots) to be reloaded for further investigation. This is also a great way to share saved analysis with collaborators if using the cloud-hosted version of GECO.

Generate reduced dimensionality data:

1. Start with the ‘Generate reduced data’ mode (click the ‘>’ in the top left to reveal the sidebar menu with Select Mode option.)
2. Upload data matrix in the format described above (“File Upload”). Review the inferred sample information from the upload – if the information is not correct, update the .csv file accordingly.
3. Set Reduction Run Parameters:
    a. See info here for setting t-SNE parameters
    b. See info here for setting UMAP parameters
    c. Three optional boxes can be checked:
        i. Remove entries with all zeros
        ii. Normalize per gene (row) – this function divides each entry of a given row by the sum of the entire row and thus allows for investigation of trends across samples *independent* of overall expression level.
        iii. Normalize to type– This function normalizes each row to a specified sample type (a drop-down menu will allow you to select a type for normalization) and allows for investigation of trends relating to the fold change compared to the selected type.
4. Run the reduction. This could take a bit of time depending on the size of the data. Active analysis is visualized by a ‘running’ icon in the top right corner of the app.
5. Preliminary data and manipulation are available. Enter a filename and save the data.
    a. Note that using the same filename as one that already exists will overwrite the file.

Visualize the data:

6. Switch to ‘Plot reduced data’ mode
7. Select the desired saved dataset from the drop-down list under ‘Load Data’.
8. Each dot/data point on the t-SNE/UMAP corresponds to one gene/protein/other unique ID depending on the input data. Using the ‘Color Data’ pull down menu the coloring of these data points can be altered in several ways:
    a. Expression of s*ample name/type* = Average expression/value based on each sample type – one at a time.
    b. Assigned type =Each data point is assigned the color of the sample type that has the highest expression/value for that gene/protein/other unique ID.
    c. Average expression of assigned type = The color scale is set to a range of expression values and each data point is colored for the average of the sample type that has the highest expression.
    d. Enrichment in type (select) = The color scale is set to a range of fold change of the selected sample (chosen in the dropdown) over maximum sample type expression. The color corresponding to the highest expression value will mark data points that are highest in the selected sample relative to other sample types.
9. Additional color/display options:
    a. Log Scale [only for some color data options] - transforms the scale of colors and is useful if there are prominent outliers overshadowing other data points.
    b. The Continuous Color Scale option [only for some color data options] - uses a continuous color gradient across discrete types to show transitions from one sample to the next. This analysis mode is particularly useful for time course or drug treatment when sample types are related.

   c. Reverse Color Scale switches the order of the colors.
   d. Sequential/Discrete Color Scales changes the overall colors used.

10. Additional filtering steps:
    a. Min expression of assigned type to show = removes data points with low expression values based on the number in the filter. Filters are based on the ‘assigned type’ which is the sample with the maximum expression.
    b. Min expression of *selected type* to show [only available for “Expression of *sample name/type”* color option] = removes data points below a specified threshold for the sample type selected to colorize the data.
    c. Fold change of selected type over average of other types [only available for the “Enrichment in type (select)” color option] = removes data points below a specified cut-off between selected type and other types.
11. ‘Gene Markers to Show’ will highlight the specified genes/proteins/unique IDs.
    a. Curated marker lists can be uploaded as a .csv file (genes in a column with descriptive header – files with multiple columns are accepted; see description in file upload format). The descriptive header will be used to select which marker gene list should be displayed. These lists will be highlighted on the plot with an ID dot (which is a large dot with a black outline).
    b. The Gene IDs input list will also be highlighted with an ID dot (the ID dot color will be different than the Curated list ID dots, if present).
    c. This is case sensitive and must match the given dataset – if a gene/protein/unique ID is not found it will not be displayed (no warning/error message), but other matching IDs from the list will be displayed.

    d. The ID dot will be a large circle appearing on the plot behind the data point. Depending on the color scale and density of data points, adjustments may need to be made to see the ID dot (filtering, zooming in, etc.). A key ‘Genes’ and/or ‘Markers’ will appear on the bottom left corner of the plot. By clicking on either of these terms, you will hide the circles while preserving the gene list.

12. Once at least one gene has been correctly entered in the Gene ID box, a bar graph will appear below the t-SNE/UMAP and plot the gene across the samples. If more than one gene has been entered the one displayed in the bar graph can be changed using the drop down ‘Gene to bar plot’ menu.
13. Once at least two genes have been correctly entered in the Gene ID box, two additional displays will appear:
    a. A clustermap showing the correlations of the selected genes. This displays all of the genes and calculates a correlation coefficient (significance of correlation or anti-correlation is shown with an asterisk).
    b. An expression heatmap of the specified genes across all samples. This is by default normalized by gene but the box that appears can be un-checked to disable this.
14. Under the ‘Filtered Gene Download’ there is a ‘Get all genes’ button that will print a list of all of the genes/proteins/unique IDs in the dataset. To print a specified cluster of genes type in the window of x and y coordinates and all genes from that specific range will be printed (with any plotting filters applied). Once the gene list is printed to the screen it can be downloaded (add the correct .csv extension to the end of the file name before opening).
15. The gene displayed can be adjusted by filtering as specified earlier or by zooming in/out on the plot. Zooming in can be performed by highlighting an area or hovering over the plot until a set of buttons appears in the top right corner (including +/- buttons). Double clicking on the plot returns

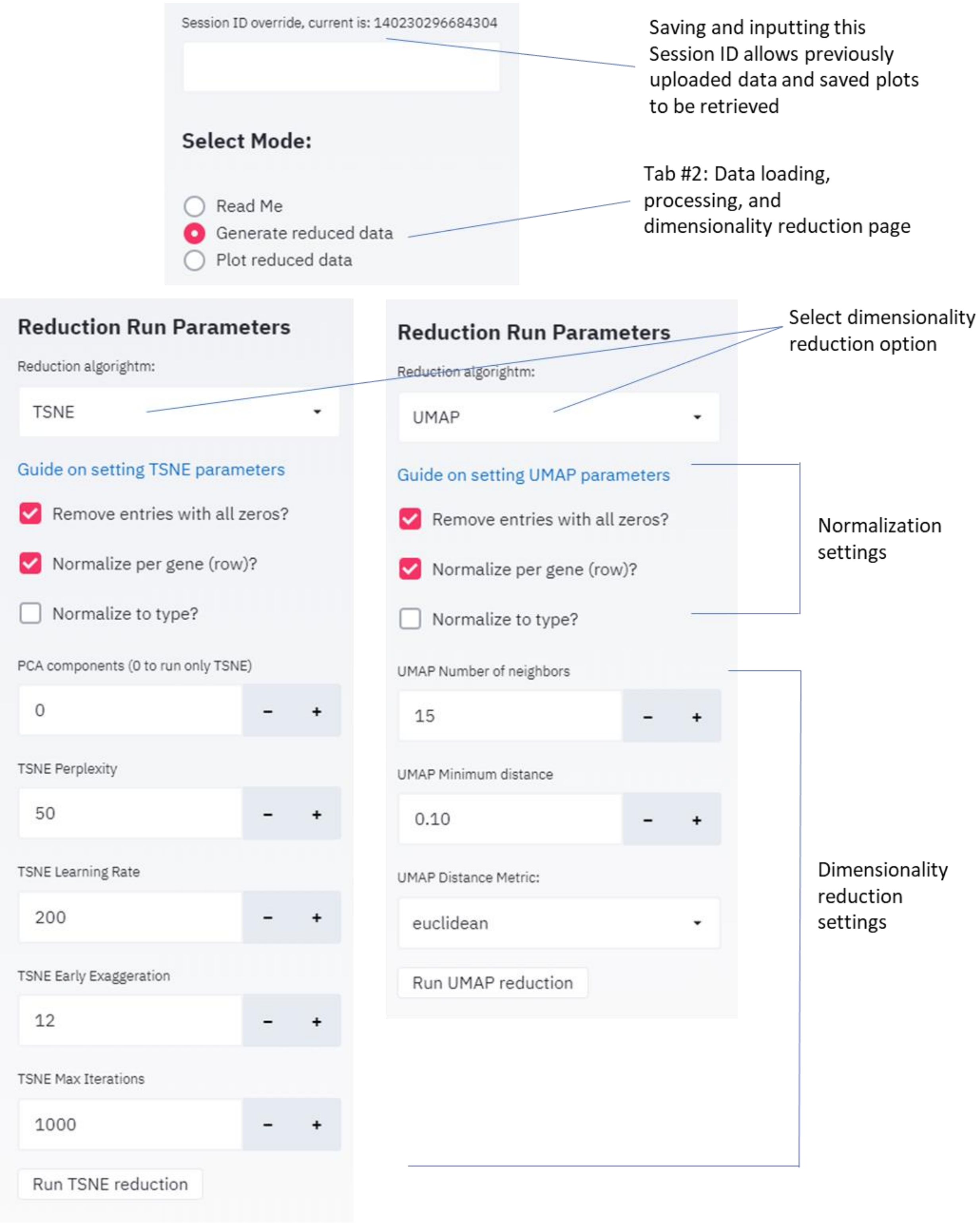
to the default view. Hovering over a data point will display the gene and information on the expression in samples and the current color scale information.

Troubleshooting/FAQ --

1. File uploader utility says ‘files are not allowed’
    - Check that the file ends in a .csv and is a simple comma separated table.
2. Odd persistent issues with the app
    - You can soft reboot the app by hitting ‘c’ and ‘clear cache’ and then hit ‘r’ to reload.
2. The normalization options for ‘Normalize per gene (row)’ and ‘Normalize to type’ are not appearing as options when generating reduced data.
    - These two options are only available when more than two sample types are included in the data matrix. When these normalizations are used on only two sample types it results in a useless analysis.
3. TSNE takes a long time to run when using GECO locally, how can I make it faster?
    - To reduce the runtime of the t-SNE algorithm a GPU can be used. This requires a CUDA enabled graphics card (most Nvidia GPU’s), a Linux based system, and a more complex installation. However, using a GPU will reduce the t-SNE runtime down to only a few seconds for even very large datasets.
4. Given the same settings and dataset, why do each generated t-SNE and UMAP not look the same?
    - Both UMAP and t-SNE are stochastic reductions which means there is a level of randomness to where datapoints initially fall into the plotted space, however the relationships and trends between datapoints will be consistent.
5. What is the best way to save GECO generated data?

- It is recommended to save the current session ID number as the cloud hosted web browser page can get reloaded/cleared and it does eventually time-out after a long duration of use.
- Plots can also be downloaded as a .png to a local folder using the camera capture button just above the top right side of the plot (hover the mouse over the plot if these buttons are not visible).
- Once a gene list is generated a link will appear enabling the full list to be generated. Once this download occurs add the necessary '.csv' extension to the file name before opening.

6. What correlation metric is used for the clustermap?
   - A Pearson r correlation test is used. More info here.
7. What are some suggested color scales?
   - Blackbody
   - Electric
   - Jet
   - Thermal



Session ID override, current is: 140230296684304

Saving and inputting this Session ID allows previously uploaded data and saved plots to be retrieved

Select Mode:

○ Read Me
● Generate reduced data
○ Plot reduced data

Tab #2: Data loading, processing, and dimensionality reduction page

Reduction Run Parameters

Reduction algorightm: TSNE

Guide on setting TSNE parameters

☑ Remove entries with all zeros?
☑ Normalize per gene (row)?
☐ Normalize to type?

PCA components (0 to run only TSNE): 0

TSNE Perplexity: 50

TSNE Learning Rate: 200

TSNE Early Exaggeration: 12

TSNE Max Iterations: 1000

Reduction Run Parameters

Reduction algorightm: UMAP

Guide on setting UMAP parameters

☑ Remove entries with all zeros?
☑ Normalize per gene (row)?
☐ Normalize to type?

UMAP Number of neighbors: 15

UMAP Minimum distance: 0.10

UMAP Distance Metric: euclidean

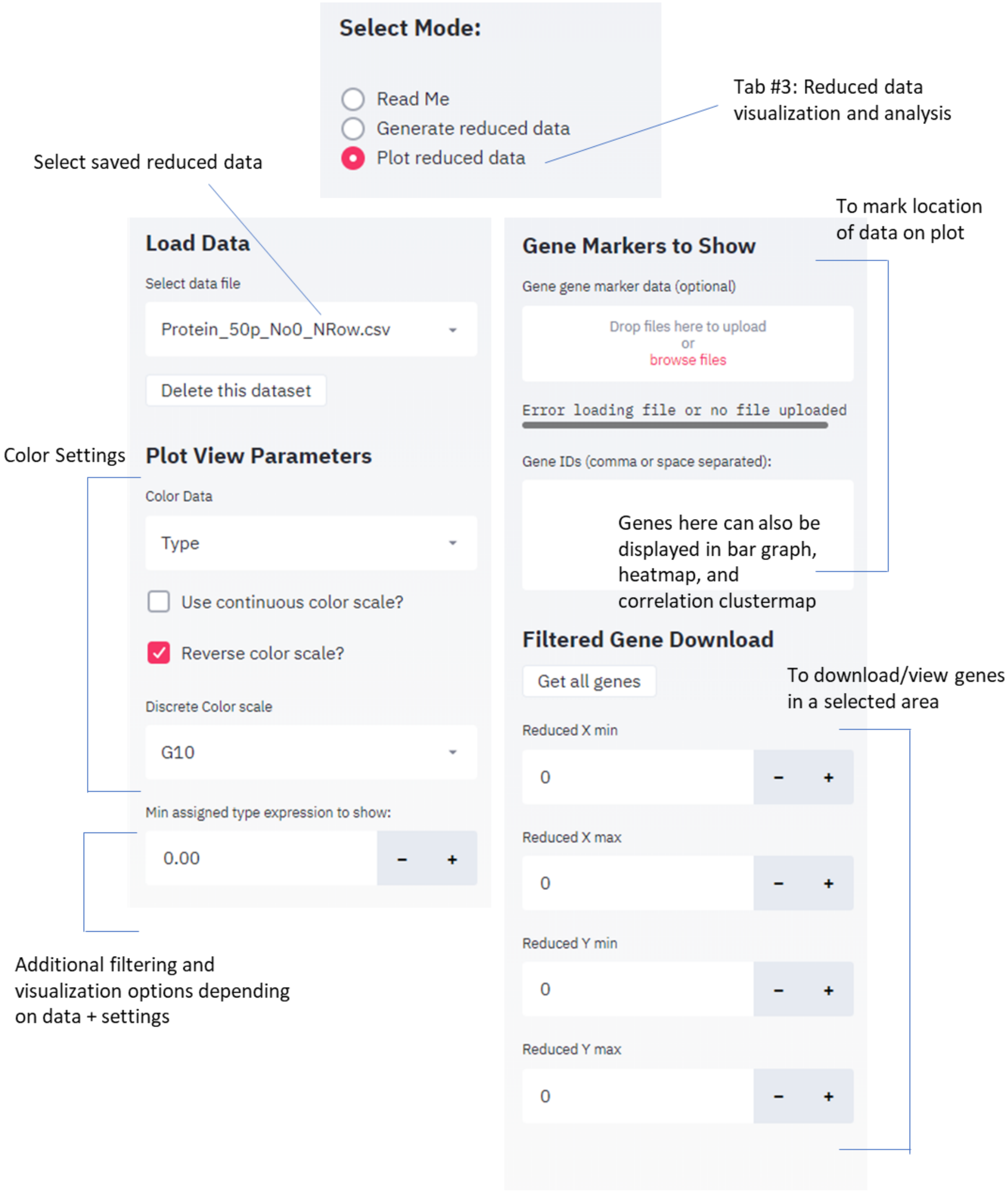
Select dimensionality reduction option

Normalization settings

Dimensionality reduction settings

Figure C.1: ***GECO app interface to generate reduced date.***

Once a data matrix is uploaded in the main window and samples are identified, this side bar is used to select the desired reduction type (t-SNE or UMAP). Normalization options include default settings of removing entries with all zeros and row normalization. Normalizing to a selected type is optional. Parameter options for t-SNE include initial PCA reduction, perplexity, learning rate, early exaggeration, and iteration number. UMAP parameters include number of neighbors, minimum distance, and distance metric. Standard default settings automatically appear, but links to t-SNE and UMAP parameter guides are provided to aid in exploration and customization.



Select Mode:

○ Read Me
○ Generate reduced data
● Plot reduced data

Load Data

Select data file: Protein_50p_No0_NRow.csv

Delete this dataset

Plot View Parameters

Color Data: Type

☐ Use continuous color scale?

☑ Reverse color scale?

Discrete Color scale: G10

Min assigned type expression to show: 0.00

Gene Markers to Show

Gene gene marker data (optional)

Drop files here to upload
or
browse files

```
Error loading file or no file uploaded
```

Gene IDs (comma or space separated):

Filtered Gene Download

Get all genes

Reduced X min: 0

Reduced X max: 0

Reduced Y min: 0

Reduced Y max: 0

Tab #3: Reduced data visualization and analysis

Select saved reduced data

To mark location of data on plot

Color Settings

Genes here can also be displayed in bar graph, heatmap, and correlation clustermap

To download/view genes in a selected area

Additional filtering and visualization options depending on data + settings

Figure C.2: ***GECO app interface for plotting and investigating reduced data.***

This tab and interface enable investigation of a previously saved plot of reduced data. The data displayed and color settings can be adjusted and filtered. Optional gene marker lists can be uploaded or input in the Gene ID box to be highlighted on the plot or further investigated in bar graph, clustermap, or heatmap. A gene list from a region of interest can be generated using the coordinate system.

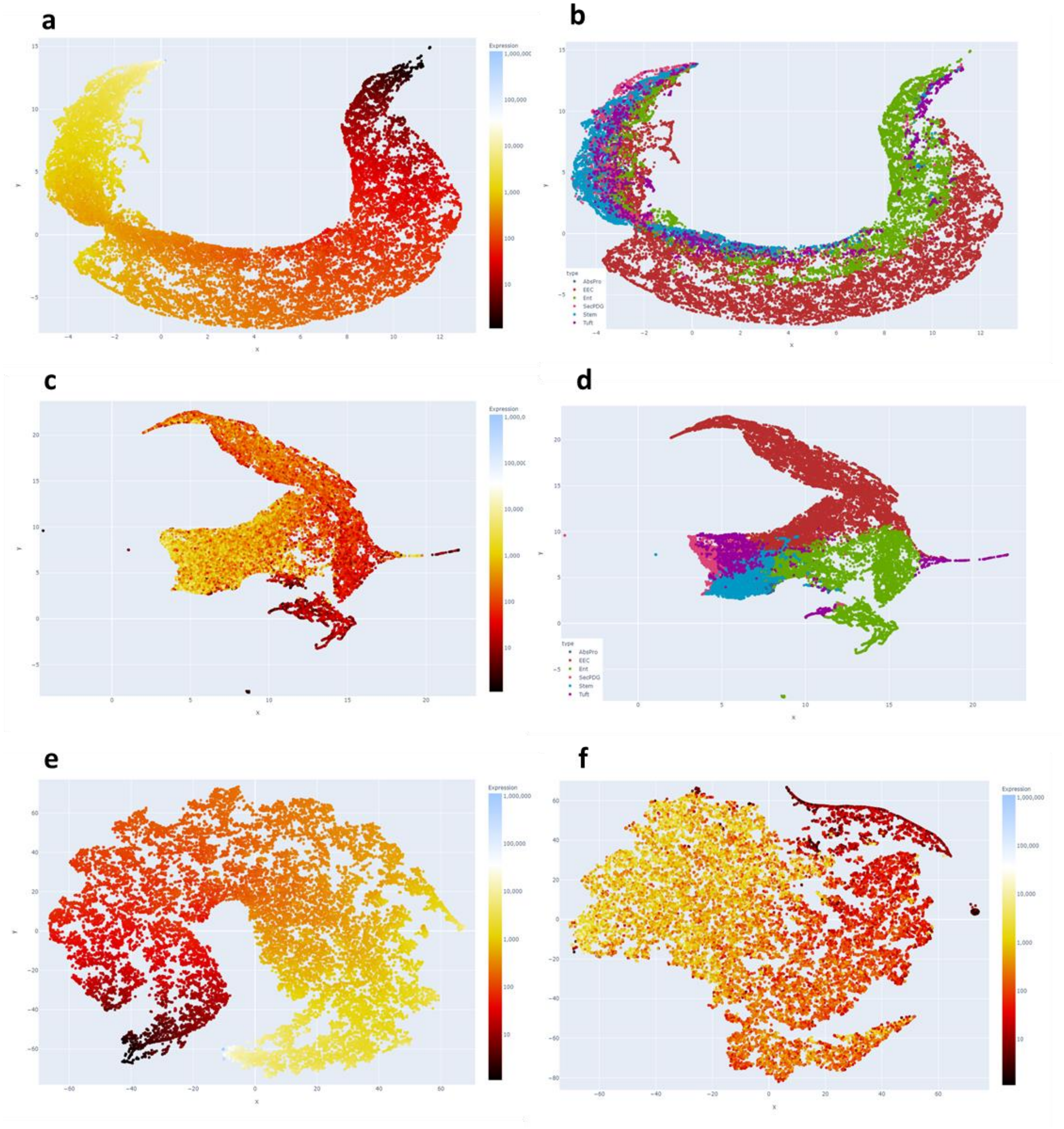


Figure C.3: ***Data normalization influences clustering and usefulness of UMAP and t-SNE.***

UMAP generated without row normalization (removal of zeros, number of neighbors = 15; minimum distance = 0.10; distance metric = Euclidean) colored by **a** average expression of assigned type or **b** assigned type. Assigned type is the sample which has the highest expression of a given gene. UMAP generated with the same parameters in addition to row normalization colored by **c** average expression of assigned type or **d** assigned type. t-SNE generated plot (PCA = 0; perplexity = 50; learning rate = 200; early exaggeration = 12; max iterations = 1000) **e** without row normalization and **f** with row normalization. Previously published bulk RNA-seq of colon crypt cell types [23] was used to generate UMAP and t-SNE clustering.

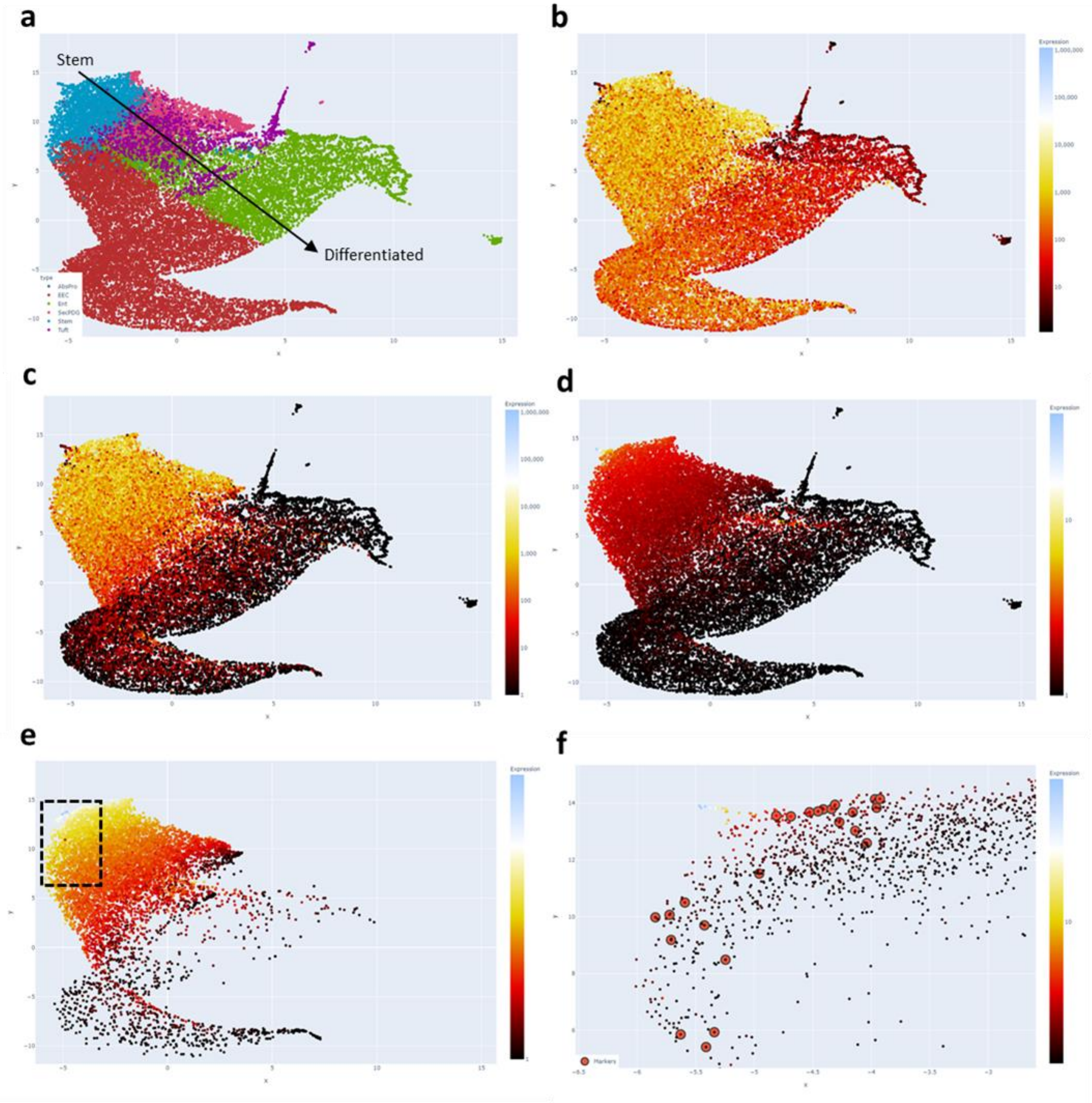


Figure C.4: ***Investigation of intestinal stem genes using GECO***.

UMAP generated plot of colon crypt cell types with the following settings: row normalization, removal of zeros, number of neighbors = 35, minimum distance = 0.5, distance metric = Manhattan. Data points (genes) colored by **a** type, **b** average expression of assigned type, **c** stem expression, **d** stem enrichment, **e** stem enrichment with 500 (normalized counts) minimum expression level with box showing zoomed in region displayed in **f**. **f** Zoomed in region with stem enrichment coloring and a 1.5-fold cutoff. Highlighted in red circles are 27 genes that were previously identified as being statistically differentially expressed and enriched in stem cells in this dataset [23].

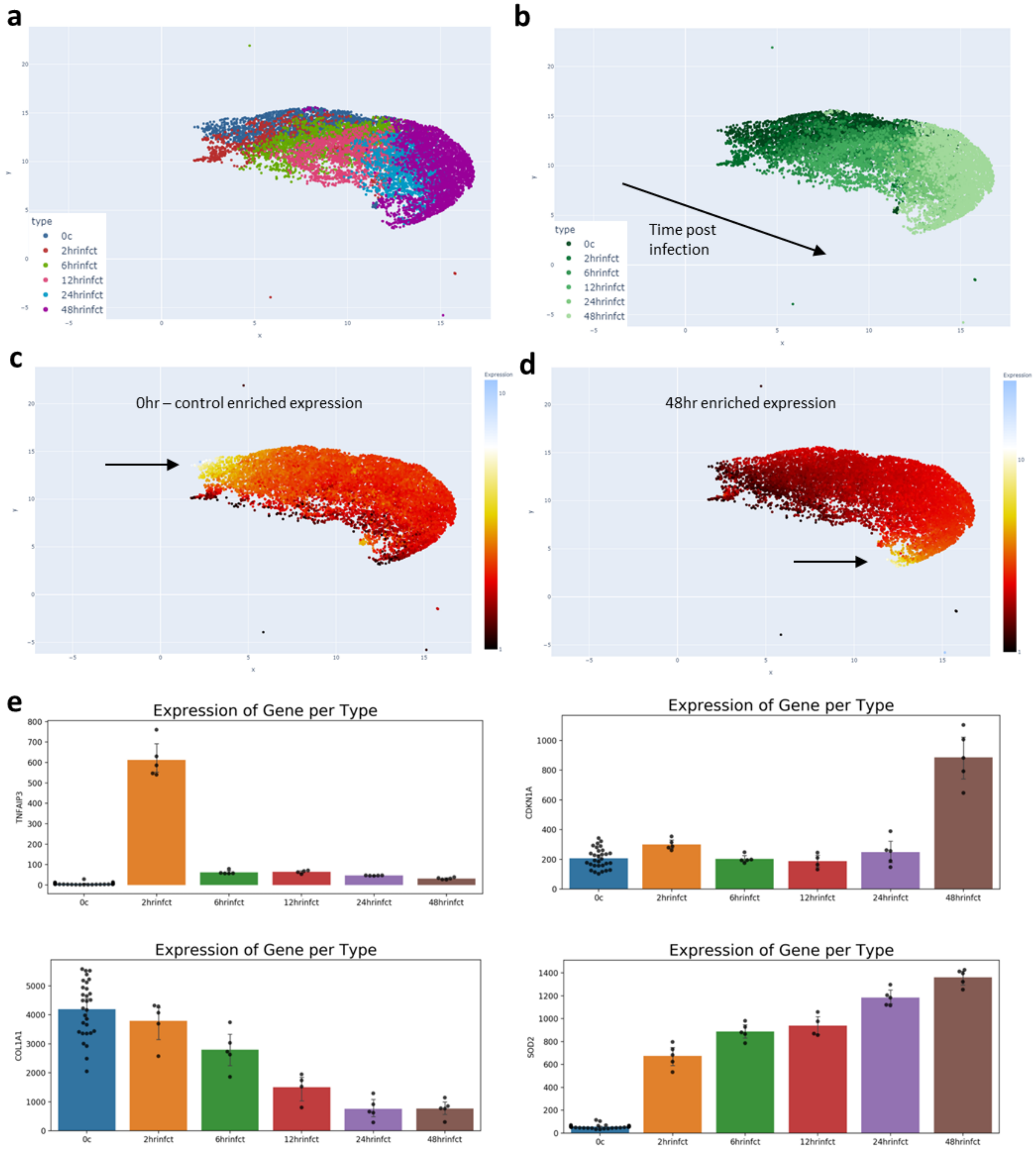


Figure C.5: ***Gene expression patterns during infection time course.***

UMAP generated plot of *F. nucleatum* infection at time points 0, 2, 6, 12, 24, and 48hrs with the following settings: row normalization, removal of zeros, number of neighbors = 15, minimum distance = 0.10, distance metric = Cosine, minimum expression = 1. Data points (genes) colored by **a** type, **b** type with continuous color setting, and expression of enrichment in **c** 0hr and **d** 48hr. **e** Selected genes of interest graphed using GECO. Previously published bulk RNA-seq of *F. nucleatum* infection time course [24] was used to generate UMAP clustering.

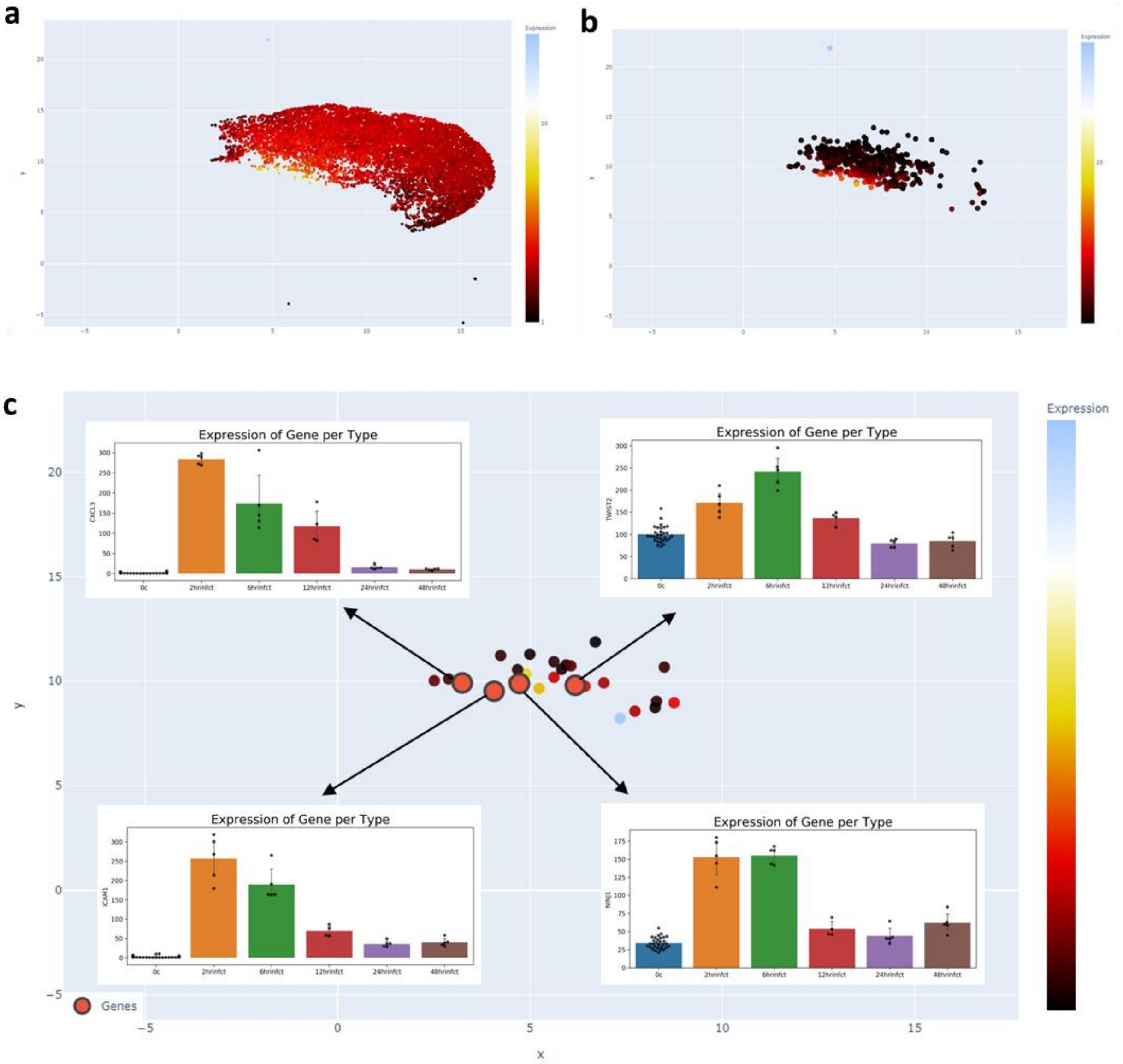


Figure C.6*: Identification of genes of interest during an infection using GECO.*

UMAP generated plot of *F. nucleatum* infection at time points 0, 2, 6, 12, 24, and 48hrs [24] with the following settings: row normalization, removal of zeros, number of neighbors = 15, minimum distance = 0.10, distance metric = Cosine. Data points (genes) colored by **a** expression of enrichment in 6hr with minimum expression = 1, **b** additional 1.5-fold cutoff filter, and **c** 150 minimum expression cutoff filter. **c** Small subset of remaining genes following filtering with 4 genes highlighted with a red dot (*TWIST2*, *ICAM1*, *CXCL3*, and *NINJ1*) that are enriched in 6hr and 2hr infection time points. Bar graphs generated in GECO show bio-replicates and expression of genes across all samples.

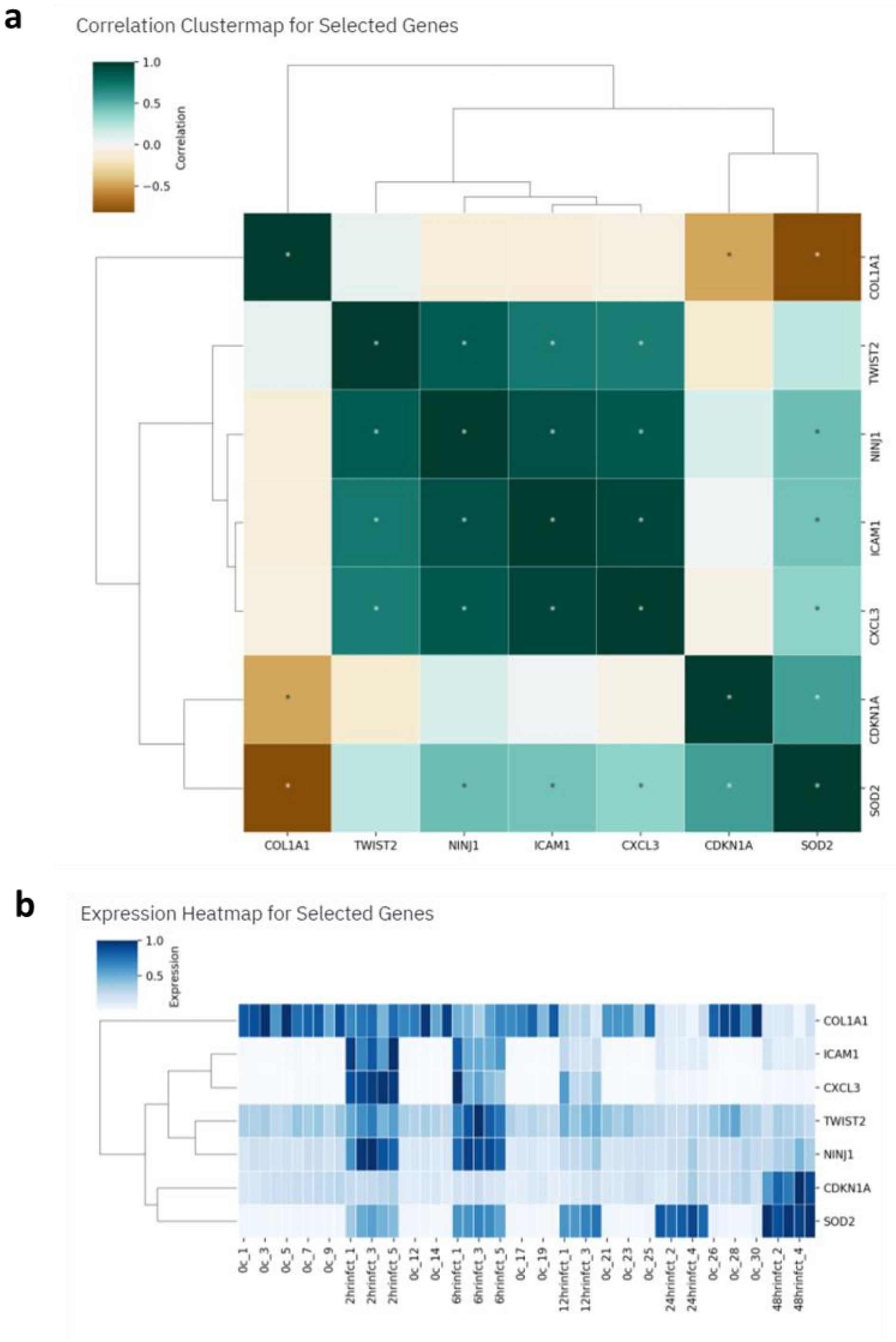


Figure C.7: ***GECO enables further visualization of genes of interest using clustermaps and heatmaps.***

a Correlation of *F. nucleatum* infection related genes identified in Figure C.5 and C.6 are displayed in a GECO generated clustermap. Dark teal boxes with an asterisk are significantly positively correlated in expression across all samples (ex: *ICAM1* and *CXCL3*), whereas dark brown with an asterisk are anti-correlating across all samples (ex: *COL1A1* and *SOD2*). **b** Expression of selected genes in a GECO generated heatmap.

REFERENCES

1. Kchouk, M., Gibrat, J. F. & Elloumi, M. Generations of Sequencing Technologies: From First to Next Generation. *Biol. Med.* **09**, (2017).

2. Muir, P. *et al.* The real cost of sequencing: Scaling computation to keep pace with data generation. *Genome Biol.* **17**, 53 (2016).

3. Mahi, N. Al, Najafabadi, M. F., Pilarczyk, M., Kouril, M. & Medvedovic, M. GREIN: An Interactive Web Platform for Re-analyzing GEO RNA-seq Data. *Sci. Rep.* **9**, 1–9 (2019).

4. Perrin, H. *et al.* OMICtools: a community-driven search engine for biological data analysis. *arXiv Prepr. arXiv1707.03659* (2017).

5. Henry, V. J., Bandrowski, A. E. & Pepin, A.-S. OMICtools: an informative directory for multi-omic data analysis. *Database* **2014**, (2014).

6. Conesa, A. *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biology* **17**, 1–19 (2016).

7. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

8. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).

9. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).

10. Spies, D., Renz, P. F., Beyer, T. A. & Ciaudo, C. Comparative analysis of differential gene expression tools for RNA sequencing time course data. *Brief. Bioinform.* **20**, 288–298 (2019).

11. Oh, S., Song, S., Grabowski, G., Zhao, H. & Noonan, J. P. Time series expression analyses using RNA-seq: A statistical approach. *Biomed Res. Int.* **2013**, 16 (2013).

12. Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Experimental and Molecular Medicine* **50**, 96 (2018).

13. Poirion, O. B., Zhu, X., Ching, T. & Garmire, L. Single-cell transcriptomics bioinformatics and computational challenges. *Frontiers in Genetics* **7**, 163 (2016).

14. Lever, J., Krzywinski, M. & Altman, N. Principal component analysis. *Nature Methods* **14**, 641–642 (2017).

15. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **374**, 20150202 (2016).

16. Van Der Maaten, L. & Hinton, G. *Visualizing Data using t-SNE*. *Journal of Machine Learning Research* **9**, (2008).

17. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv Prepr. arXiv1802.03426* (2018).

18. Kobak, D. & Berens, P. The art of using t-SNE for single-cell transcriptomics. *Nat. Commun.* **10**, 1–14 (2019).

19. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–47 (2019).

20. Li, W., Cerise, J. E., Yang, Y. & Han, H. Application of t-SNE to human genetic data. *J. Bioinform. Comput. Biol.* **15**, 1750017 (2017).

21. Bushati, N., Smith, J., Briscoe, J. & Watkins, C. An intuitive graphical visualization technique for the interrogation of transcriptome data. *Nucleic Acids Res.* **39**, 7380–7389 (2011).

22. Chan, D. M., Rao, R., Huang, F. & Canny, J. F. t-SNE-CUDA: GPU-Accelerated t-SNE and its Applications to Modern Data. *arXiv:1807.11824* (2018).

23. Habowski, A. N. *et al.* Transcriptomic and proteomic signatures of stemness and differentiation in the colon crypt. *Commun. Biol.* **3**, 1–17 (2020).

24. Kang, W. *et al.* Time-Course Transcriptome Analysis for Drug Repositioning in Fusobacterium nucleatum-Infected Human Gingival Fibroblasts. *Front. Cell Dev. Biol.* **7**, 204 (2019).