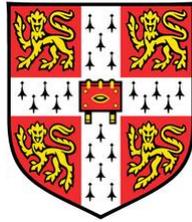


# Accumulation of Somatic Mutations in Normal Human Colonic Epithelium



**Cora Olpe**

Wellcome Trust – MRC Cambridge Stem Cell Institute

CRUK Cambridge Institute

University of Cambridge

This dissertation is submitted for the degree of

*Doctor of Philosophy*



This thesis is dedicated to my family.



## **Declaration**

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. This dissertation contains less than 60,000 words excluding appendices, bibliography, footnotes, tables and equations and has less than 150 figures.

Cora Olpe

August 2019



## Publication contributions

Parts of this thesis were published as:

Nicholson, A. M., **Olpe, C.**, Hoyle, A., ..., Wilkinson, M., Morrissey, E., Winton, D.J. (2018) Fixation and Spread of Somatic Mutations in Adult Human Colonic Epithelium, *Cell Stem Cell*. Elsevier, 22(6), pp. 909-918

The contributions made by Dr Anna Nicholson are included in chapter 1. As outlined in the results chapters (3-5) I validated and applied the clonal mark STAG2 and performed the targeted amplicon sequencing of *KRAS*. I was also involved in manuscript writing and editing and designed the graphical abstract.

During my PhD candidacy I also co-authored the following book:

Olpe, H.R., & **Olpe, C.** Hirnwellness. Alzheimer, Hirnschlag und Depressionen – von den Risiken zu präventiven Möglichkeiten. *Hogrefe*, 2017

This discusses the contribution of lifestyle factors such as diet and exercise to risk of developing Alzheimer's, stroke and depression. It is not part of this thesis.



## Acknowledgements

In many ways, the three years I spent on this project have enabled me to test my limits. My PhD has thus been an exciting journey of scientific but also personal growth. In both cases, this heavily relied on a strong network of people around me. My particular gratitude goes to a number of individuals without whom I would not be the scientist or person I am today.

First of all, I would like to thank my PhD supervisor Doug. It has been a tremendous pleasure working in your laboratory. I am humbled by your ability to immediately spot a weakness in any experiment I propose. Thank you for all the hours of fascinating scientific discussion (including occasional digression into jam making), for your support and for allowing me to explore so many different ideas.

I am also extremely grateful to our collaborators in Oxford, Edward Morrissey and Doran Khamis, who significantly contributed to the success of my project. Their skills and passion for research are inspiring and I have thoroughly enjoyed our numerous meetings. Equally, Richard Kemp in the Winton laboratory has taught me most of what I know about Unix as well as gardening. Thank you for all your patience and teaching. A big thank you also goes to all other members of the Winton lab whom I had the pleasure of working with. I really value all our interesting discussions, both scientific and beyond.

Next, I would like to thank Brian Hendrich and Austin Smith for offering me a place on the Wellcome Trust 1+3 PhD Programme in Stem Cell Biology and Medicine. I have always felt extremely supported by the programme and have hugely enjoyed the scientific freedom that comes with ownership of a grant. A big thank you also goes to my thesis committee, the members of which have provided me with valuable discussion and constructive criticism.

During my undergraduate years as well as the first year of my PhD I was lucky to find a number of inspiring mentors. Thank you, Aleksandra Watson, for always encouraging me to toot my horn and for sharing numerous pearls of wisdom related to science careers. Thank you, Marc de la Roche, for making me feel like a real scientist and for providing coffee and pastries when a failed experiment threw me off my track. Thank you, Anna Nicholson, for all the help and encouragement you provided in my first year and for enduring all the disruptions my wanting to chat must have caused to your work. Finally, thank you so much, Sarah Thorpe, for being an

awesome rotation supervisor, colleague and friend. Lunches at your house were the best antidote to thesis writing crises.

A lot of the experiments I performed during my PhD relied on the expertise of the histopathology core facility at the CRUK Cambridge Institute. Thank you for tirelessly cutting sections for me as well as offering advice on immunohistochemistry protocols and laser capture microdissection. I am particularly grateful to Julia Jones and Cara Brodie – you are stars!

A number of friends from undergraduate times have accompanied me throughout my PhD. Clemency, thank you for all those swims, novel-length text messages and for #3<sup>rd</sup> year. Emilis, thank you for all the road trips, all the eastern European love, hours of whatsapp calls and likes on Instagram. <lemon and key emoji>. Leonie, thank you for hiking, Jass and brunches. Simon, thank you for driving over from the Sanger to have lunch with me countless times and for being my running buddy. Susanna, thank you for being the best college wife, for hosting me all around Europe and for mööööööööh. Tamara, thank you for countless great nights out, catch-ups and a wonderful trip to Istanbul. Xian Yao, thank you for Jailbreak, for Barcelona, Brussels and Amsterdam, for meditation in John's library, for all those Chinese New Year Parties, for Boron chemistry and for letting me batch cook for you.

Equally, the friends I made during my PhD have substantially contributed to making it a fantastic experience. Thank you to all the members of the Corpus Christi College Swim Team for letting me boss you around in the pool and for inspiring me to continue training intensely throughout my PhD. Swimming has, and will always be, a great outlet and an almost meditational way to balance the intellectual challenges associated with research. Thank you, Anne and Carolin for all those girls' evenings of games, gossip and, most importantly, Katzenfleisch. Thank you, Ravi, for boozy science gossip, expensive hipster coffees and impromptu wine tastings. And of course, thank you Katrin for being my thesis writing buddy, for rescuing me from that glitter queue and never judging my bedtime. Die Zimtsterne glänzen nicht.

One should never forget one's roots and mine are in Switzerland from where my great childhood friends continue to support me in a number of ways. Thank you, Rhiannon and Andrea, for your ability to put things into perspective and enabling me to escape the Cambridge bubble when necessary. A huge thank you goes to all my friends from my swim team, the SBO, Kevin, Thomas, Ariane, Bengi, Andrin and Bedi. Your countless visits to Cambridge, our great

swim/underwater sessions as well as numerous social gatherings continue to bring me much happiness and provide me with grounding.

Next, I want to thank my future in-laws. Firstly, Jane and Keith, who hosted me during the most relaxing and productive writing retreat and whose continuous support means so much to me. Also, my lovely future sister-in-law Katie and her husband, Thibaut, for cheering me on, always.

I am lucky to have a wonderful family, by which I have, throughout my life, felt extremely supported. I would like to thank my brother Tobias for inspiring me to start swimming, for countless visits to Cambridge and just generally for being the best brother I could have asked for. I admire the way you tackle life. My parents Beata and Hans Rudolf have always been my biggest supporters and fans. Thank you so much for nurturing my talents from a very early age. Without your dedication, foresight and ambition I would not be where I am today. I am immensely grateful.

Last but certainly not least I am thankful to my fiancé George. Thank you for being that rock in my life, for celebrating the ups and downs of research with me. I admire your strength of character, your kindness, your realism, and your relaxed attitude to life. I look forward to a lifetime of growth, both scientific and personal, by your side.



## Abstract

I investigated stem cell dynamics in normal human colon by detecting somatic variants affecting X-linked and autosomal genes using immunohistochemistry. Applying neutral clonal marks to a large cohort of patients to interpret age-related trends in clone frequencies has established the baseline stem cell dynamics of the tissue.

Analysis of a number of new clonal marks showing biased behaviours suggests that different gene-specific mutations can subvert constraints resulting from the tissue architecture in different ways. In respect of intra-gland competition of stem cell derived clones, a disadvantage is observed for the histone modifier HDAC6, while at the other end of the spectrum, loss of the cohesin member STAG2 strongly advantages affected stem cells.

Subsequent clonal expansion beyond the boundary of a single crypt is recognised by clones occupying multi-crypt patches. Quantification of such events allows the rate of lateral expansion for different mutations to be measured. Moderate effects were found for PTEN, p53 and STAG2, while mutations in the histone demethylase KDM6A generate very large areas of mutant epithelium in aged humans. Further, targeted sequencing revealed dramatic expansion of *KRAS*-mutant clones in histologically normal colon.

Patches may arise from crypt fission and fusion events. Using a clonal mark based on mild periodic acid-Schiff staining, the neutral crypt fission and fusion rates were quantified. Against this baseline, it was found that *KDM6A*-mutant clones expand mainly by fission, while fusion remains at homeostatic levels.

The emerging picture is that of the aged human colon consisting of a mosaic of different mutations. The work presented in this thesis offers detailed insights into the rates at which different gene-specific mutations arising in colonic stem cells can become fixed within individual crypts and undergo subsequent lateral expansion through crypt fission and fusion events. This defines the timeframe taken for cancer drivers to achieve high mutant allele burden within the tissue, which can serve as a basis for cancer prevention strategies.



# Table of contents

Table of contents.....	xv
List of Figures.....	xxi
List of Tables.....	xxiii
Nomenclature.....	xxv
Chapter 1 Introduction.....	1
1.1 What is a stem cell?.....	1
1.2 Adult tissue stem cells.....	3
1.2.1 Experimental methods to study adult stem cells.....	4
1.3 Mammalian intestinal stem cell biology.....	7
1.3.1 Structure of the mammalian intestinal epithelium.....	7
1.3.2 The intestinal stem cell niche.....	8
1.3.3 Mouse studies on intestinal stem cells.....	10
1.4 Defining human intestinal stem cell dynamics.....	15
1.4.1 Early studies on hereditary conditions.....	15
1.4.2 Human lineage tracing.....	15
1.5 Mutations, stem cells and cancer in renewing tissues.....	20
1.5.1 Mutation accumulation and colorectal cancer (CRC).....	20
1.5.2 The Vogelstein model of colorectal cancer progression.....	24
1.5.3 The Big Bang model of CRC initiation.....	26
1.5.4 Mutation acquisition and spread in renewing tissues.....	27
1.6 Summary and aims.....	30
Chapter 2 Materials and methods.....	31
2.1 Human tissue.....	31
2.2 mPAS staining.....	31
2.3 Immunohistochemistry.....	32
2.3.1 Standard protocol.....	32
2.3.2 Adaptations for laser capture microdissection.....	32
2.4 Clonal marks data acquisition and analysis.....	34
2.4.1 Clone and crypt counting.....	34

---

2.4.2	Scoring fusion.....	34
2.4.3	Clone and fufi data analysis .....	34
2.4.4	Patch size frequencies .....	34
2.5	Laser capture microdissection and sequencing .....	35
2.5.1	Laser capture microdissection and sample lysis.....	35
2.5.2	Assessment of FFPE DNA quality .....	35
2.5.3	Sequencing library preparation.....	36
2.5.4	Next-generation sequencing .....	37
2.5.5	Sequencing data analysis.....	38
2.6	Targeted <i>KRAS</i> sequencing .....	39
2.6.1	FFPE DNA sample preparation.....	39
2.6.2	Library preparation .....	40
2.6.3	Targeted <i>KRAS</i> sequencing data analysis .....	41
Chapter 3	Clonal marks discovery.....	45
3.1	Introduction .....	45
3.1.1	The concept behind human clonal marks .....	45
3.1.2	Requirements for clonal marks.....	46
3.1.3	Aims .....	50
3.1.4	Published work .....	50
3.2	Identification of novel X-linked clonal marks .....	51
3.2.1	Selection of X-linked genes .....	51
3.2.2	Antibody testing.....	54
3.2.3	Pre-screening for clonal marks .....	57
3.2.4	Validation and optimisation .....	57
3.3	Identification of cancer-associated clonal marks .....	60
3.3.1	Detection of oncogene activation .....	62
3.3.2	Detection of tumour suppressor loss or gain-of-function .....	64
3.4	A method for next-generation sequencing of laser captured FFPE crypt samples ..	65
3.4.1	Optimisation of antigen retrieval for LCM .....	66
3.4.2	Optimisation of protocol for PCR on FFPE DNA from crypt sections .....	68
3.5	Validation of MAOA and KDM6A by sequencing.....	69
3.5.1	Identification of mutations in MAOA <sup>-</sup> patches .....	69
3.5.2	Identification of mutations in KDM6A <sup>-</sup> patches .....	74
3.6	Summary and Discussion.....	77

---

3.6.1	Identification of four novel X-linked clonal marks .....	77
3.6.2	Identification of two CRC-associated clonal marks .....	79
3.6.3	LCM and sequencing of MAOA <sup>-</sup> and KDM6A <sup>-</sup> crypt sections .....	81
Chapter 4	Inference of intra-crypt behaviours .....	85
4.1	Introduction .....	85
4.1.1	Aim .....	86
4.1.2	Published work .....	86
4.2	General workflow .....	86
4.2.1	Scoring of WPC and PPC frequencies .....	86
4.2.2	A note on sampling and noise .....	88
4.2.3	Inferring intra-crypt bias from $\Delta C_{fix}/C_{part}$ .....	88
4.3	Behaviours associated with X-linked marks .....	90
4.3.1	NONO confirms neutral stem cell dynamics .....	90
4.3.2	HDAC6 loss confers disadvantage .....	92
4.3.3	KDM6A loss confers advantage .....	94
4.3.4	STAG2 loss confers greater advantage than KDM6A .....	98
4.4	Intra-crypt behaviours for cancer-associated marks .....	98
4.4.1	P53 stabilisation is associated with neutral stem cell dynamics .....	100
4.4.2	PTEN loss is associated with neutral stem cell dynamics .....	103
4.5	Comparison of all marks .....	105
4.6	Inference of mutation rates .....	106
4.7	Discussion .....	109
Chapter 5	Expansion beyond the crypt .....	117
5.1	Introduction .....	117
5.1.1	Crypt fission .....	118
5.1.2	Aims .....	120
5.1.3	Published work .....	120
5.2	Quantification of crypt fission using clonal marks .....	121
5.2.1	Gene-specific mutations can promote crypt fission .....	121
5.2.2	Intra-crypt dynamics and crypt fission may be independent .....	124
5.3	<i>KRAS</i> mutations in normal human colonic epithelium .....	126
5.3.1	A targeted amplicon sequencing approach for <i>KRAS</i> exon 2 .....	127
5.3.2	Lateral expansion of <i>KRAS</i> mutations .....	131
5.4	Crypt fusion .....	136

---

5.4.1	The fufi.....	137
5.4.2	Inference of crypt fusion rates.....	138
5.4.3	Fission and fusion duration.....	141
5.5	Summary and Discussion.....	143
Chapter 6	Discussion.....	151
6.1	Investigating mutation accumulation in normal human colon.....	151
6.1.1	Clonal marks for precise quantification in large cohorts.....	152
6.1.2	Targeted sequencing for inference of mutant patch sizes.....	153
6.1.3	Overview of alternative methodologies.....	154
6.2	Gene-specific mutations can subvert normal biology.....	157
6.3	Towards understanding CRC initiation and incidence.....	159
6.3.1	Proposed multiplex method.....	159
6.4	Extrinsic factors.....	160
6.4.1	Niche – stem cell interactions in CRC.....	161
6.4.2	Colitis-associated CRC.....	162
6.4.3	Other extrinsic factors influencing CRC.....	163
6.4.4	Future directions for investigation of extrinsic factors.....	164
6.5	Cancer prevention.....	166
References	.....	169
Appendix A	Patients included.....	A-1
Appendix B	Primers.....	B-1
Appendix C	Mathematical modelling of clonal marks data.....	C-1
C.1	A general note on simulations and the mathematical model.....	C-1
C.2	Statistical inference.....	C-1
C.3	Statistical model.....	C-2
C.4	Continuous labelling of a neutral mutation.....	C-3
C.5	Continuous labelling of a non-neutral mutation.....	C-5
C.6	Fitting the monoclonal clones and partial clones.....	C-6
C.7	Sequential mutations.....	C-7
C.8	Crypt fission and mutation burden.....	C-8
C.9	Statistical model for patch sizes.....	C-9
C.10	Effect of crypt fusion on patch size.....	C-10
Appendix D	Mathematical modelling of KRAS data and comparison to clonal marks....	D-1
D.1	Statistical model for KRAS patch size estimation.....	D-1

---

D.2	Relative expansion coefficient .....	D-3
Appendix E	Mathematical modelling of crypt fusion .....	E-1
E.1	Fusion/Fission events as Poisson processes .....	E-1
E.2	Calculating the fusion rate .....	E-2
E.3	Calculating the event duration .....	E-3



## List of Figures

Figure 1.1 Baking analogy to illustrate the concept of functional stem cells. ....	2
Figure 1.2 Architecture of the mammalian intestinal epithelium. ....	9
Figure 1.3 Continuous labelling data. ....	13
Figure 1.4 Inference of human stem cell dynamics using clonal marks. ....	19
Figure 1.5 Long tail of frequently mutated genes in MSS colorectal cancer. ....	23
Figure 1.6 The Vogelstein model of CRC progression. ....	25
Figure 1.7 Two models for CRC initiation. ....	27
Figure 1.8 Scenarios for second mutation acquisition. ....	29
Figure 2.1 Illustration of PERL script used for analysis of mutations in LCM DNA. ....	39
Figure 3.1 Clonal marks scenarios. ....	47
Figure 3.2 Workflow/decision tree for identification of new X-linked clonal marks. ....	52
Figure 3.3 Schematic representation of spread of gene characteristics. ....	53
Figure 3.4 Antibody screen for novel X-linked clonal marks. ....	55
Figure 3.5 Candidate marks abandoned at validation stage. ....	56
Figure 3.6 Validation of new X-linked clonal marks. ....	59
Figure 3.7 Overview of relationships between selected cancer-associated proteins. ....	61
Figure 3.8 Antibody screen for oncogene-associated clonal marks. ....	63
Figure 3.9 Antibody screen for tumour suppressor-associated clonal marks. ....	65
Figure 3.10 Optimisation of IHC and PCR on FFPE tissue sections and extracted DNA. ....	67
Figure 3.11 Laser capture microdissection and sequencing of multicrypt patches. ....	70
Figure 3.12 Identification of germline SNPs in MAOA. ....	72
Figure 3.13 Identification of mutations in MAOA <sup>-</sup> patches. ....	73
Figure 3.14 Identification of mutations in KDM6A <sup>-</sup> patches. ....	75
Figure 4.1 Workflow for inference of stem cell behaviours. ....	87
Figure 4.2 Sampling and noise in human clonal marks data. ....	89
Figure 4.3 Inference of stem cell behaviour associated with NONO mutation. ....	91
Figure 4.4 Inference of stem cell behaviour associated with HDAC6 mutation. ....	93
Figure 4.5 Inference of stem cell behaviour associated with KDM6A mutation. ....	95

---

Figure 4.6 Loss of KDM6A does not increase H3K27me3 in human colon. ....	96
Figure 4.7 KDM6A <sup>-</sup> WPC accumulation in females and males. ....	97
Figure 4.8 Inference of stem cell behaviour associated with STAG2 mutation. ....	99
Figure 4.9 Inference of stem cell behaviour associated with stabilisation of p53. ....	101
Figure 4.10 P53 stabilisation and HDAC6 loss are not correlated in human colonic crypts. ....	102
Figure 4.11 Inference of stem cell behaviour associated with PTEN mutation. ....	104
Figure 4.12 Stem cell behaviour associated with gene-specific mutations. ....	105
Figure 4.13 Inferred mutation rates scale with coding sequence length. ....	108
Figure 4.14 Adjusting the coding sequence length of KDM6A may rationalise the inferred mutation rate. ....	115
Figure 5.1 Detection of multicrypt patches for clonal marks. ....	122
Figure 5.2 Gene-specific mutations can increase crypt fission rates. ....	124
Figure 5.3 Intra-crypt dynamics and crypt fission may be independent processes. ....	125
Figure 5.4 Effects of intra-crypt bias and fission bias on total mutation burden. ....	126
Figure 5.5 KRAS and mimic amplicons synchronise around the KRAS G12 and G13 context. ....	129
Figure 5.6 Sources of artefactual MAFs and their elimination by the ‘mimic amplicon method’ ....	130
Figure 5.7 Noise and mutation calling in targeted sequencing data. ....	132
Figure 5.8 Identification of KRAS mutations in normal human colonic epithelium. ....	134
Figure 5.9 Ranking advantage conferred by gene-specific mutations. ....	135
Figure 5.10 Proposed process for crypt fission and fusion. ....	136
Figure 5.11 Defining the fufi. ....	137
Figure 5.12 Scoring of patch edge fufis and neighbours. ....	138
Figure 5.13 Identification of mPAS <sup>+</sup> fufis. ....	139
Figure 5.14 Identification of KDM6A-negative fufis. ....	140
Figure 5.15 Crypt fission and fusion rates for mPAS and KDM6A. ....	141
Figure 5.16 Inference of human crypt fission and fusion duration. ....	142
Figure 5.17 Comparison between data and simulations. ....	145
Figure 5.18 Similar efficiency of conversion into malignancy for different KRAS activating mutations. ....	146
Figure 6.1 Clonal marks can be used in green monkey tissue. ....	166
Figure C.1 Effects of fission and fusion on patch size. ....	C-11

## List of Tables

Table 1.1 CRC driver genes with mutation frequency over 5%.....	24
Table 2.1 Primary antibodies tested for IHC.....	33
Table 2.2 Sequence context used to extract mutant reads from sequencing data.....	42
Table 3.1 Intron, exon and total gene length data for candidate marks. ....	54
Table 3.2 Antibodies tested for validation of candidate clonal marks.....	58
Table 3.3 Antibodies screened for colorectal cancer-associated clonal marks. ....	60
Table 3.4 Summary of mutations found in MAOA <sup>-</sup> multicrypt patches.....	74
Table 3.5 Summary of mutations identified in KDM6A <sup>-</sup> patches. ....	76
Table 3.6 Normalised RNA expression for candidate marks. ....	79
Table 4.1 Summary of clonal marks data.....	106
Table 4.2 Inferred mutation rates for clonal marks. ....	107
Table 5.1 Crypt fission rates associated with clonal marks. ....	123
Table 5.2 KRAS mutations detected in normal human colon samples.....	133
Table 6.1 Comparison between different methods for investigation of somatic mutations in human colon. ....	151



# Nomenclature

ACF Aberrant crypt foci

AML Acute myeloid leukaemia

AR Antigen retrieval

CAC Colitis-associated colorectal cancer

CAF Cancer-associated fibroblast

CCO Cytochrome c oxidase

CI Confidence interval

CRC Colorectal cancer

DAB 3,3'-Diaminobenzidine

DTR Diphtheria toxin receptor

ECM Extracellular matrix

FAP Familial adenomatous polyposis

FFPE Formalin-fixed paraffin embedded

G6PD Glucose-6-phosphate dehydrogenase

HDAC6 Histone deacetylase 6

HMCC High-magnification chromoscopic colonoscopy

HPA Human Protein Atlas

- 
- IBD Inflammatory bowel disease
- IGV Integrative Genomics Viewer
- IHC Immunohistochemistry
- ISC Intestinal stem cell
- KDM6A Lysine demethylase 6A
- LCM Laser capture microdissection
- $\lambda_{crypt}$  Replacement rate of stem cells in the crypt
- MAF Mutant allele frequency
- MAOA Monoamine oxidase A
- ME Margin of error
- mPAS Mild periodic acid-Schiff
- MSI Microsatellite instable
- MSS Microsatellite stable
- mtDNA Mitochondrial DNA
- $N_{crypt}$  Number of effective stem cells in a crypt
- NONO Non-POU domain-containing octamer-binding
- PEN Polyethylene naphthalate
- PPC Partially populated crypt
- $P_R$  Probability of replacement
- PRC2 Polycomb repressor complex 2
- PTEN Phosphatase and tensin homolog deleted on chromosome 10

RNI Reactive nitrogen intermediates

ROS Reactive oxygen species

SNP Single nucleotide polymorphism

SNV Single nucleotide variant

STAG2 Stromal antigen 2

TA Transit amplifying

WPC Wholly populated crypt



# Chapter 1 Introduction

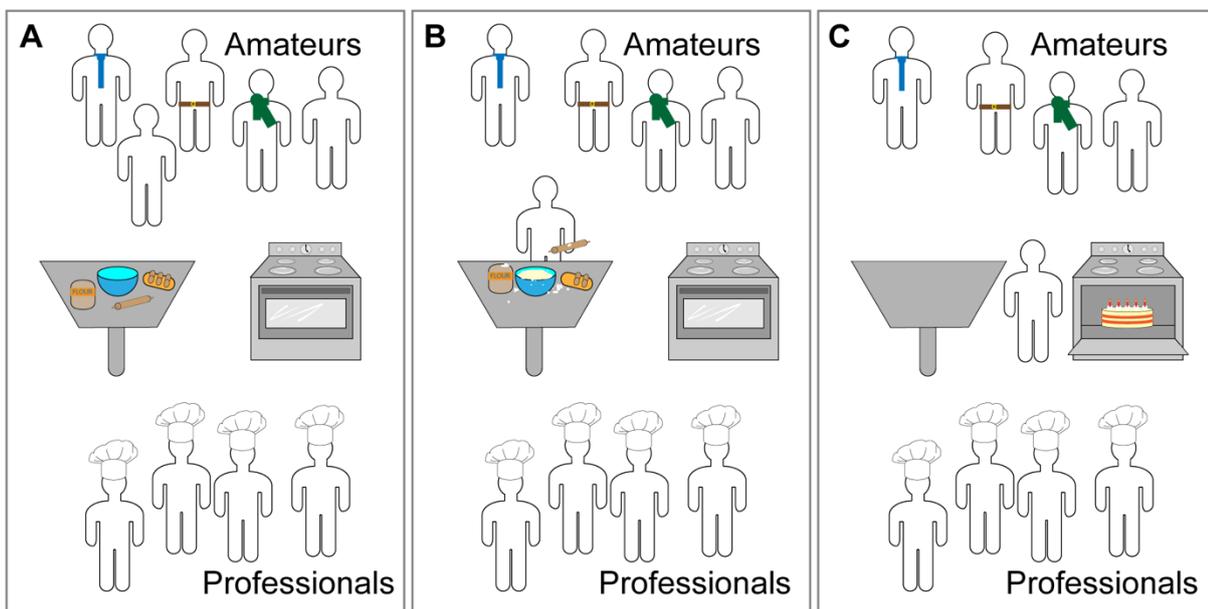
## 1.1 What is a stem cell?

New mammalian life starts with a fertilised egg, the zygote, which is a totipotent stem cell, meaning it is capable of giving rise to any cell type of a complete embryo. Following fertilisation, the zygote undergoes a number of rounds of cell division until it develops into the blastocyst, which consists of the inner cell mass, a ball of cells that will form the embryo proper, and the surrounding trophoblast, which will form the supporting structures (Gilbert, 2000). Cells can be collected from the inner cell mass and grown *in vitro* as embryonic stem cells (Evans & Kaufman, 1981; Martin, 1981).

As the embryo develops further, cells become increasingly specified as they are segregated into different groups that contribute to the growth of individual tissues and organs. These fate choices are governed by precisely timed action of specific transcription factor networks (Gilbert, 2000; Pantazis & Bollenbach, 2012).

From this brief description of early mammalian development, the two main characteristics of stem cells can be extracted. The first is potency, which is the ability to differentiate into specialised cell types. Importantly, different types of stem cells can harbour different degrees of potency. Cells found in the inner cell mass can differentiate into all cell types, except for trophoblast cells, and are thus called pluripotent. Later during development, pools of more restricted, multipotent, stem cells are specified. The second universal characteristic of stem cells is self-renewal, defined as the ability to generate progeny indefinitely (Weissman, 2000). Traditionally, research has tried to classify different types of stem cells according to their gene expression profile. Such grouping relied on specific genes acting as markers for stem cells. It is important to note, however, that the two fundamental features of stem cells are both functional in nature. Therefore, the true ‘stemness’ of a cell can only be defined retrospectively.

A cell that performs the functions attributed to stem cells is a functional stem cell, irrespective of the markers it expresses. A simple analogy can help to illustrate this point (Figure 1.1). Therefore, while markers can be very useful for elucidating the capabilities of certain stem cells as well as the underlying molecular pathways, their use will likely not reveal the full complexity present in a given stem cell-based biological system. This requires the study of functional stem cells, which has been a focus of the Winton laboratory over the past decade and will act as a recurring theme throughout this introduction.



**Figure 1.1 Baking analogy to illustrate the concept of functional stem cells.**

(A) A kitchen (representative of a tissue) contains a table with baking ingredients, an oven and two groups of people, which represent cells: Professional bakers are identified by their hats, which represent putative stem cell markers. Amateurs represent cells that do not express putative stem cell markers, illustrated by the absence of a baker's hat. The assignment is to bake a birthday cake, which here is representative of stem cell function.

(B) One of the amateurs, meaning a cell that does not express the marker, takes on the task.

(C) The result is a beautiful cake. The amateur was thus able to bake a cake, or in other words, the cell that did not express the putative stem cell marker performed the stem cell function. This illustrates that marker expression is not necessarily indicative of stem cell function. Furthermore, it demonstrates that stem cell identity can only truly be verified retrospectively, after a stem cell function has been observed.

---

## 1.2 Adult tissue stem cells

The presence of stem cells is not limited to the developing embryo. Mammalian young are born with a set of spatially defined tissue pools of so-called adult stem cells that support growth into the adult as well as maintain homeostasis throughout life. A special case is the female breast, in which the mammary stem cell population undergoes drastic expansion as the organism enters puberty (Gjorevski & Nelson, 2011).

Adult stem cell pools function in ways that reflect the needs of their resident tissues. A major distinction can be made between renewing and non-renewing tissues. A key example for the latter is the brain. While there are neuronal stem cells, which can divide to generate new neurons in an adult mammal, there is, apart from the olfactory bulb in certain species (Kornack & Rakic, 2001; Winner *et al.*, 2002), no constant turnover of the tissue. Therefore, the stem cells in the brain are found in restricted areas where they can generate new neurons as part of specific processes such as learning (Bergmann, Spalding & Frisé, 2015; Yau, Li & So, 2015).

Quite different are the demands on the stem cell pools of renewing mammalian tissues such as the haematopoietic system, skin and intestine. They need to continuously supply new cells to fuel tissue turnover. Historically, such tissues were described as exhibiting a hierarchical organisation. Stem cells with (seemingly) unlimited self-renewing capacity produce shorter-lived, more differentiated progeny that, after additional replication, ultimately differentiate into the mature cell types of the tissue. Such organisation was first demonstrated in the blood, which lends itself to experimentation due to its easy accessibility. It was thus possible to isolate cells from mouse blood and transplant them into recipient mice whose bone marrow had been removed by irradiation. This identified the cells able to long-term regenerate the entire hematopoietic system as hematopoietic stem cells. Similarly, progenitors were defined as cells that could achieve regeneration, but only for certain lineages (Weissman & Shizuru, 2008). The hematopoietic system was thus defined as a branching hierarchy of stem, progenitor and mature blood cells. Similar hierarchical architecture was found in the skin (Blanpain & Fuchs, 2006), olfactory system (Lander *et al.*, 2009), and epithelium of the intestine (Potten & Loeffler, 1990).

Tiered organisation allows for advantageous features. It enables segregation of stem cells into niches, thereby physically protecting them from hostile environments and providing them with the signalling cues required for their maintenance. In addition, definition of a specific ‘master’ stem cell population allows for tight regulation of tissue growth, homeostasis and repair by modulation of symmetric and asymmetric divisions.

In the past two decades however, a more plastic picture has emerged. In the skin, cells attributed to the progenitor pool have been shown to act functionally as stem cells in stress conditions (Jaks, Kasper & Toftgård, 2010). Lineage plasticity, the ability of progenitors of one lineage to differentiate into functional cells of another, has been demonstrated for a variety of hematopoietic cells (Graf, 2002). From these examples it transpires that while the stem cell pools in renewing adult tissues follow hierarchical organisation, there is a substantial degree of plasticity. This relates back to the functional definition of stem cells discussed earlier. It seems that which cells act functionally as stem cells in adult tissues can be context-dependent and is not necessarily restricted to the cells originally defined to sit at the apex of the hierarchy.

## **1.2.1 Experimental methods to study adult stem cells**

Adult tissue stem cells can be studied both *in vivo* or *in vitro*. The former enables investigation of cells in their physiological context but is more complex, time-consuming, costly and ethically challenging. The latter allows for more high-throughput experimentation that is cheaper and ethically more accepted, however, to what extent results obtained this way are physiologically relevant has to be evaluated on a case-by-case basis.

### **1.2.1.1 *In vivo* methods**

#### **1.2.1.1.1 Experimental organisms**

Historically, studies on adult tissue stem cells relied on lineage tracing, whereby cells are labelled followed by identification and quantification of progeny. This type of experiment thus retrospectively defines the cells that gave rise to progeny as stem cells. Since the early 1990s,

---

genetic recombination has been widely used for such studies in mice. The method is based on expression of site-specific recombination systems such as the Cre-loxP system adapted from bacteriophage P1 (Sternberg & Hamilton, 1981). In a typical experiment Cre recombinase is expressed under control of a tissue or cell-specific promoter. In addition, experimental animals also harbour a reporter gene such as GFP or tdTomato flanked by a loxP-STOP-loxP sequence. This combination results in reporter gene expression in cells in which the promoter is active, as well as in their progeny. As an example, this method was used to demonstrate that Sox9-expressing cells can give rise to all lineages in the hair follicle (Nowak *et al.*, 2008). A more advanced version of this system are inducible Cre<sup>ERT</sup>-lines, in which Cre recombinase is fused to a mutated version of the human oestrogen receptor (ER), which results in cytoplasmic retention until a ligand such as tamoxifen leads to activation (Littlewood *et al.*, 1995; Metzger *et al.*, 1995; Feil *et al.*, 1996). This system allows for temporal regulation of Cre expression, enabling selective activation in adult animals. For example, by crossing Lgr6-Cre<sup>ERT</sup> mice to a reporter strain, Lgr6-expressing cells were identified as stem cells giving rise to all lineages of the adult skin (Snippert, Haegebarth, *et al.*, 2010). Importantly, such experiments require *a priori* knowledge, as a promoter has to be chosen to drive Cre expression. An unbiased alternative is to express Cre<sup>ER</sup> from a promoter active in all cells in the tissue and titrate down the dose of tamoxifen such that only very few cells are marked in a stochastic manner. If progeny is produced long term, the labelled cell was a stem cell. Together with its progeny it forms a clone, the size of which can be used to quantify stem cell behaviour. This approach therefore identifies stem cells in a marker-free, purely functional way. It has been used to describe stem cell behaviour in the mouse mammary gland during embryonic development, puberty and reproductive age (Lloyd-Lewis *et al.*, 2018).

Not all stem cells in a tissue may necessarily proliferate continuously. Any slowly cycling stem cells may be missed by the lineage tracing methods described above. Nucleotide analogues such as BrdU or EdU can be used instead. Following a pulse of such a label, it is rapidly diluted in actively proliferating cells, but is retained much longer in cells that divide more slowly. Due to this phenomenon, pools of slowly cycling stem cells are commonly referred to as label-retaining cells.

Two additional methods to investigate adult stem cell lineages in model organisms are transplantation and direct observation. Using transplantation into damaged organs, the degree of potency of different (stem) cell populations can be defined as the degree to which short- or long-term regeneration is achieved. Direct observation is currently limited to two-photon microscopy, which is restricted to near-body surface imaging, or intravital microscopy, whereby inner organs are observed through a surgically applied window. Combined with fluorescent labelling, these methods enable observation of stem cell divisions in real time, thus removing the need of any inference. A caveat of this method is the limited time window of observation, which may overlook stem cells that remained inactive and may complicate discrimination between different degrees of potency.

#### **1.2.1.1.2 Non-experimental organisms**

In tissues from non-experimental organisms, where labelling and genetic modification are not feasible, experiments rely on post-mortem detection of marks that are naturally acquired through life. One such label is alterations in microsatellites, which are small repetitive regions of the genome that are prone to replication errors. From different degrees of similarity between the patterns of alterations in specific microsatellites, the lineage tree of cells in a tissue can be deduced (Frumkin *et al.*, 2005). DNA methylation patterns can be used in the same way (Kim & Shibata, 2002, 2004). Another naturally occurring label is  $^{14}\text{C}$ . As the levels of this carbon isotope fluctuate in the earth's atmosphere, and this is reflected in the biomass of the planet, the age of a cell can be inferred from its  $^{14}\text{C}$  content. Importantly, human nuclear weapon testing lead to a significant rise in  $^{14}\text{C}$  contents in the 1950s and early 60s. Applied to brains of individuals born in that period, the method dated neurons to be much younger than the individual they belonged to, providing evidence for adult neurogenesis (Spalding *et al.*, 2005).

#### **1.2.1.2 Organoids**

A method to study adult stem cells *in vitro* is by derivation of organoids. These are miniaturised and simplified versions of organs that retain some of the micro-anatomy of the organ of origin but are comprised solely of epithelial cells. Organoid growth requires media supplemented with

modulators of the signalling pathways controlling stem cell activity in the tissue of origin. Importantly, organoids can be disaggregated, with the consequence of each piece growing into a new mini-organ. Such ability to repeatedly generate organoids is commonly used to measure the ‘stemness’ of cells of interest. Furthermore, this platform lends itself to the use of CRISPR/Cas9 technology or inhibitors to study specific signalling pathways. Finally, *in vitro* culture quickly generates large amounts of material for molecular analysis. However, organoids lack physiological context such as stromal and inflammatory cells. In addition, the anatomy of the organ of origin is only partially retained. This limits the scope of biological questions that can be answered with this experimental system.

### 1.3 Mammalian intestinal stem cell biology

One of the best studied model systems for adult mammalian stem cell biology is the intestine, the site of nutrient absorption. It is, similarly to the skin, in contact with the exterior and therefore provides a barrier against damaging chemicals and infectious agents. A consequence of this ongoing exposure to harmful extrinsic factors is the continuous renewal of the intestinal epithelium, which is fuelled by a population of adult stem cells. The next sections will delve deeper into our current understanding of the biology of this tissue and its resident stem cells.

#### 1.3.1 Structure of the mammalian intestinal epithelium

The mammalian intestine is lined by an epithelium consisting of a single layer of cells. In the small intestine, this is arranged into invaginations called crypts of Lieberkühn and finger-like protrusions known as villi (Figure 1.2). It is a very rapidly renewing tissue, which in mice is replaced approximately every 5 days. This remarkable turnover is fuelled by the division of intestinal stem cells (ISCs) residing at the bottom of the crypts. The daughter cells produced by stem cell divisions are fed into a so-called transit-amplifying compartment. There, cells undergo approximately 4-5 rounds of rapid cell division, after which point they terminally differentiate into nutrient-absorbing enterocytes or one of the secretory cell types, which are Paneth cells, goblet cells, tuft cells and enteroendocrine cells (Cheng & Leblond, 1974a, 1974b; Bjerknes & Cheng, 1999; Gerbe *et al.*, 2011). Around 5 days after genesis the cells reach the villus tip and

are subsequently shed into the gut lumen. Paneth cells migrate downward to reach their final destination at the base of the crypt, intermingled between the stem cells. They have a relatively long lifespan of 6-8 weeks (Ireland *et al.*, 2005) and act as niche cells, providing short-range Wnt, Notch and EGF signals to support the stem cells (Sato *et al.*, 2011; Farin *et al.*, 2016). The architecture of the colon is very similar to the small intestine. The two main differences are firstly the flat luminal surface lacking villi and secondly that Paneth cells are nearly absent and substituted by deep secretory cells (Altmann, 1983; Rothenberg *et al.*, 2012; Sasaki *et al.*, 2016) (Figure 1.2).

### 1.3.2 The intestinal stem cell niche

Stem cell niches are defined as the microenvironment necessary to maintain stem cell self-renewal, pluripotency and proliferation as well as direct positioning of differentiating progeny. In the intestine this consists of the extracellular matrix, soluble factors as well as cellular components. The latter comprises epithelial Paneth (small intestine) or deep secretory cells (colon, see above) as well as a plethora of subepithelial mesenchymal cells that line the basal lamina. These are also called stromal cells and include fibroblasts, myofibroblasts, pericytes, endothelial cells and smooth muscle cells (Meran, Baulies & Li, 2017). In addition, the epithelium contains intraepithelial lymphocytes interspersed between epithelial cells as well as populations of other immune cells such as dendritic cells and macrophages that can be found throughout the lamina propria (Van Kaer & Olivares-Villagómez, 2018). These are key to mediating mucosal immunity (Allaire *et al.*, 2018)

Subepithelial stromal cells provide signalling molecules to maintain ISCs. Key among these are Wnts which are the major driver of ISC proliferation. Recently, subepithelial telocytes as well as Gli1-expressing mesenchymal cells have been identified as major sources of Wnt proteins in the intestine (Degirmenci *et al.*, 2018; Shoshkes-Carmel *et al.*, 2018). Subepithelial cells also produce R-spondins, potent Wnt agonists, as well as BMP antagonists such as Noggin and Gremlin, that prevent BMP-mediated differentiation (He *et al.*, 2004; Clevers, Loh & Nusse, 2014; Meran, Baulies & Li, 2017). In contrast, BMP expression increases along the crypt-villus axis, thereby promoting differentiation of cells upon upwards movement (He *et al.*, 2004; Spit,

Koo & Maurice, 2018). ISC and their daughter cells are thus exposed to and directed by a complex array of signals which influences the gradient of stemness and differentiation that exists along the crypt-villus axis.

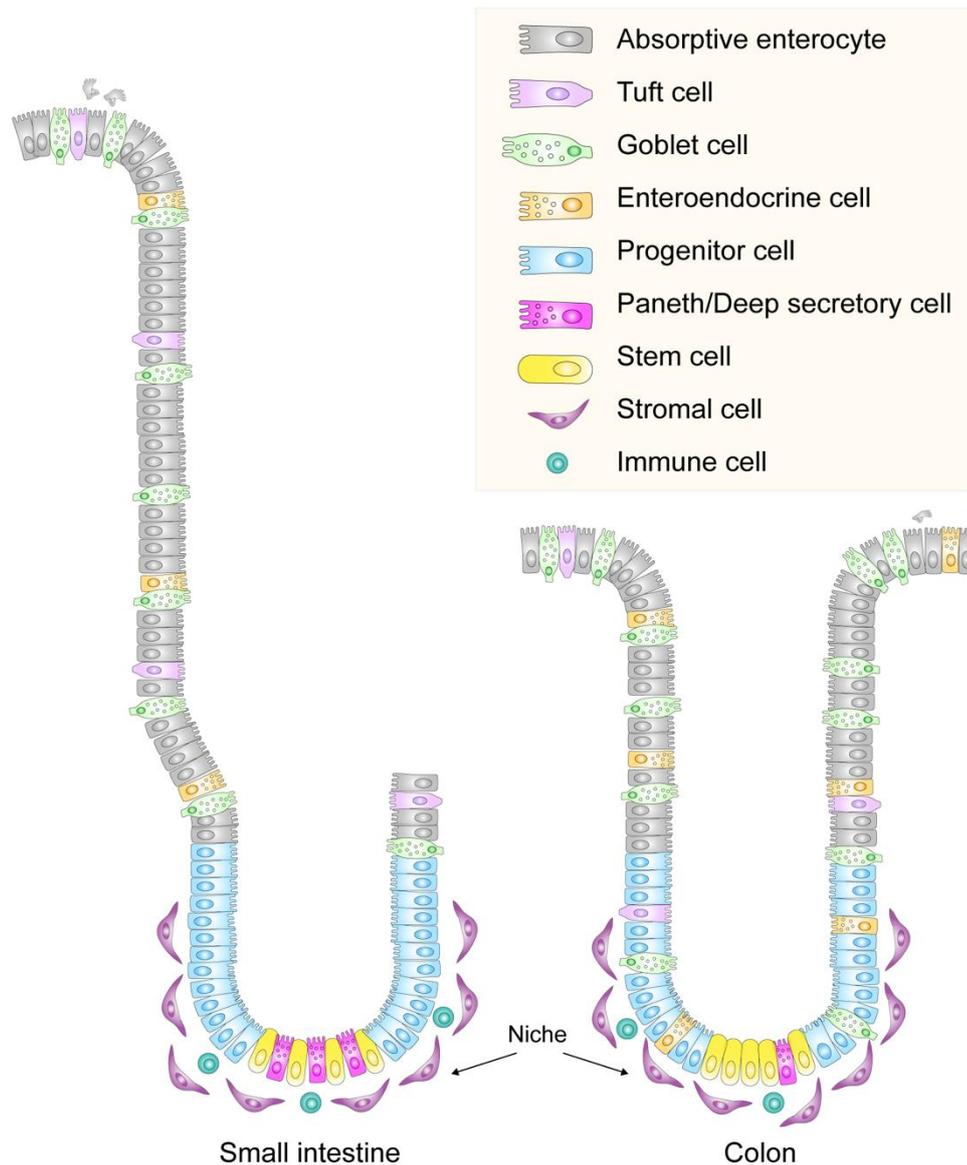


Figure 1.2 **Architecture of the mammalian intestinal epithelium.**

*The small intestine is folded into crypts and villi while the latter are absent in the colon. Stem cells are located at the bottom of crypts. Their division generates progenitors that, as they shift upwards, differentiate into the mature cell types.*

### 1.3.3 Mouse studies on intestinal stem cells

A lot of our current knowledge about ISCs originates from studies using transgenic mouse models. In particular, lineage tracing from specifically marked populations has been used to identify the cells giving rise to long-lived clonal progeny as true stem cells.

#### 1.3.3.1 Intestinal stem cell markers

The first marker that could reliably be used for such experiments was *Lgr5*, which is the receptor for the Wnt signalling agonist R-spondin (Glinka *et al.*, 2011), identified by the Clevers group in 2007 (Barker *et al.*, 2007). Based on its expression, it was estimated that each mouse small intestinal crypt contains around 16 stem cells. *Lgr5*-expressing cells at the bottom of intestinal crypts were found to be positive for the proliferation marker *Ki67* and were shown to incorporate the nucleotide analogue 5-bromodeoxyuridine, which indicates active cycling (Barker *et al.*, 2007). However, it had been known since the late 1970s that intestinal crypts also contain quiescent or slowly cycling stem cells, as indicated by nucleotide analogue retention over periods of around 4 weeks in cells located in a supra-Paneth cell (+4) position within crypts (Potten *et al.*, 1978; Potten, Owen & Booth, 2002). Expression of *Bmi1* (Sangiorgi & Capecchi, 2008), *mTert* (Montgomery *et al.*, 2011) and *Hopx* (Takeda *et al.*, 2011) were all proposed as markers for slowly-cycling intestinal stem cells. Their functional relevance was studied using lineage tracing as well as ablation and injury models. Lineage tracing using *LacZ* or fluorescent reporters revealed infrequent contribution of these quiescent stem cell populations to intestinal homeostasis (Sangiorgi & Capecchi, 2008; Montgomery *et al.*, 2011). However, upon selective ablation of *Lgr5*<sup>+</sup> cells using diphtheria toxin administration in mice expressing the diphtheria toxin receptor in *Lgr5*<sup>+</sup> cells, crypt architecture was maintained. This suggested a role for slowly-cycling stem cells as ‘reserve stem cells’. Intriguingly, removal of *Bmi1*<sup>+</sup> cells with the same system resulted in crypt attrition, suggesting a requirement of *Bmi1*<sup>+</sup> cells for crypt maintenance (Sangiorgi & Capecchi, 2008; Tian *et al.*, 2011). Further, *Lgr5*<sup>+</sup> cells could be eliminated by irradiation and crypt regeneration achieved by proliferation of radiation-resistant reserve stem cell populations (Montgomery *et al.*, 2011; Yan *et al.*, 2012).

---

Taken together these studies suggested the coexistence of two distinct intestinal stem cell populations: One of actively cycling Lgr5-expressing cells and one composed of overlapping subgroups expressing markers such as Bmi1, mTert and Hopx.

However, this model has recently been refined. Using an unbiased split Cre recombinase approach label retaining cells were shown to act as stem cells in an injury setting but normally be fated towards the secretory lineage (Buczacki *et al.*, 2013). The same was found for Bmi1+ cells (Jadhav *et al.*, 2017; Yan *et al.*, 2017). In addition to active and quiescent ISCs, markers have also been proposed for various cell lineages, such as Dll1 (van Es *et al.*, 2012) and Atoh1 for secretory progenitors and Ngn3 (Jenny *et al.*, 2002) for enteroendocrine cells. Interestingly, it has been shown that at least two of the proposed stem cell populations are able to interconvert (Takeda *et al.*, 2011). Moreover, there is evidence that Dll1- and Ngn3-expressing cells, which are predominantly lineage restricted, can also give rise to long-lived multi-lineage clones of cells during homeostasis and post-irradiation (Schonhoff, Giel-Moloney & Leiter, 2004; van Es *et al.*, 2012). Plasticity has also been demonstrated for Paneth cells, which have been shown to contribute to intestinal regeneration in injury or inflammation settings (Schmitt *et al.*, 2018; Yu *et al.*, 2018). The picture is further complicated by the fact that many of the proposed marker genes are expressed by overlapping populations of cells (Muñoz *et al.*, 2012).

Two themes emerge from the above discussion: Firstly, the data obtained from marker-based studies indicate that the stem cell pool exhibits dynamic phenotypic variability. Context seems to determine which cells can and do act as stem cells in the intestine, with a considerable degree of plasticity. Therefore, a more fluid description that takes context (homeostasis, injury, disease, ageing) into account is thus required. Secondly and related to this point is that a marker-free approach is required for capturing of the true functional output of the system. This is in line with findings from other tissues discussed earlier and fits in with the baking analogy from section 1.1. Therefore, intestinal stem cells will from now on be discussed in a marker-free, purely functional way.

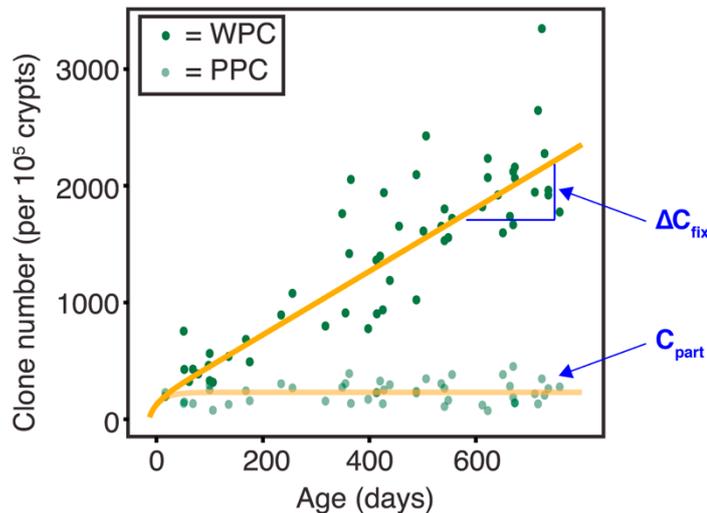
### 1.3.3.2 Mouse intestinal stem cell dynamics

In addition to defining different populations of stem cells, a number of studies in the past decade have addressed the question of how stem cells populate the crypt. It had been known since the late 1980s that mouse crypts become monoclonal by the time the animals are 2 weeks old and that adult intestinal crypts undergo continuous monoclonal conversion such that all cells within a crypt are derived from a single stem cell (Schmidt, Winton & Ponder, 1988; Winton, Blount & Ponder, 1988). However, the precise dynamics of this process remained unquantified for over 20 years. The prevailing model for many years suggested that the number of stem cells in a crypt remained constant through asymmetric stem cell divisions. This assumption was mainly based on a number of studies in *Drosophila*, where asymmetric divisions are important for specifying cell fates during brain development (Betschinger & Knoblich, 2004; Knoblich, 2008; Fürthauer & González-Gaitán, 2009).

This model changed with two landmark studies in 2010, which employed short and long-term pulse-chase lineage tracing experiments in both *Lgr5*- as well as unbiased  $\text{Cre}^{\text{ER}}$  mouse lines. This revealed that monoclonality is achieved through stochastic, unbiased competition between stem cells - a process termed neutral drift (Lopez-Garcia *et al.*, 2010; Snippert, van der Flier, *et al.*, 2010). In this model, the stem cells at the crypt base form an equipotent population of cells that continuously replace each other. Stem cell divisions are balanced by differentiation and loss, such that the overall size of the stem cell pool remains the same. Importantly, while describing the process by which stem cells populate the crypt, these studies did not define the key parameters associated with it, namely the number of functional stem cells and their replacement rate.

This was achieved in a subsequent study by Kozar and colleagues, which confirmed the dynamics of neutral drift as well as precisely quantified both the number of functional stem cells (known as  $N_{\text{crypt}}$ ) and their rate of replacement ( $\lambda_{\text{crypt}}$ ) in the mouse intestine (Kozar *et al.*, 2013). The study was based on the  $(\text{CA})_{30}$  mouse model, in which microsatellite mutation activates reporter gene expression. This leads to continuous appearance of labelled cells throughout the animal's life. When intestinal stem cells are labelled, they can, due to neutral

drift, populate the entire crypt or disappear when non-labelled cells outcompete them. Therefore, the number of crypts fully populated by mutant cells (wholly populated crypts, WPC) increases linearly with age, while the number of crypts partially populated by mutant cells (partially populated crypts, PPC) stays constant as wild type cells readily replace mutant cells (Figure 1.3). The slope of WPC accumulation is hereafter referred to as  $\Delta C_{fix}$  and the constant frequency of PPC as  $C_{part}$ . The age-related behaviour of WPC and PPC was predicted from the neutral drift model, from which it could also be derived that  $\Delta C_{fix}$  depends on the mutation rate and the stem cell replacement rate  $\lambda_{crypt}$ , while  $C_{part}$  also depends on the mutation rate as well as the number of functional stem cells per crypt,  $N_{crypt}$ .



**Figure 1.3 Continuous labelling data.**

Scatter plot showing numbers of PPC and WPC in the mouse colon. Orange lines indicate predictions from modelling.  $\Delta C_{fix}$  and  $C_{part}$  are indicated with blue arrows. Adapted from Kozar et al., 2013.

The mutation rate was calculated from the frequency of single cell clones in the transit amplifying (TA) compartment, which were indicative of *de novo* mutation events, as opposed to belonging to a stem-cell derived clone. Combined with the experimentally derived values for  $\Delta C_{fix}$  and  $C_{part}$  this enabled inference of stem cell dynamics parameters in the mouse colon. The values obtained were: functional stem cells per crypt,  $N_{crypt} = 7$  (95% confidence interval, CI:  $\pm 0.3$ ) and replacement rate,  $\lambda_{crypt} = 0.3$  (95% CI:  $\pm 0.04$ ) per crypt per day. With a stem cell

division rate of around once per day, this suggests that the majority of stem cell divisions in the mouse intestine are symmetric (Kozar *et al.*, 2013).

A further layer of complexity was added by a recent study elegantly combining *Lgr5* genetic labelling with intravital imaging. The technique enabled a detailed description of individual clone lineages and their position along the z-axis of the crypt. This demonstrated that *Lgr5*<sup>+</sup> cells positioned towards the base of the crypt experience a bias towards renewal while the cells near the border are, due to their location, more prone to removal from the niche, resulting in differentiation and loss. However, stem cells can be rearranged at the base of the crypt, which can reassign the fate bias to whichever cells are most central (Ritsma *et al.*, 2014).

In summary, in homeostasis in the mouse intestine, a heterogeneous population of cells near the bottom of the crypt with varying and changing degrees of stem cell potential functions long-term as a smaller population of effective equipotent stem cells (7 in the colon) with dynamics described by neutral drift. Therefore, the output of the neutral drift model quantitatively captures the more complex biological system. Such thorough understanding of homeostatic stem cell dynamics formed the basis for investigations into the effects of perturbations, such as genetic mutations.

### **1.3.3.3 Mutations impacting on neutral drift dynamics**

Using mice harbouring inducible alleles for *KrasG12D* mutation or *Apc* inactivation as models for intestinal tumour initiation, Vermeulen and co-workers demonstrated by lineage-tracing that these mutations confer a marked competitive advantage to intestinal stem cells. A *KrasG12D* mutation or the loss of *Apc* increases the likelihood of a clone becoming fixed. However, the presence of either mutation is not deterministic. Rather, the dynamics governing this process are still stochastic as mutated stem cells are still readily replaced by wild-type stem cells. Interestingly, mutations in the tumour suppressor gene *Tp53* seem to only confer an advantage in an inflammation setting (Vermeulen *et al.*, 2013).

---

## 1.4 Defining human intestinal stem cell dynamics

Studies in genetically modified mice have provided invaluable insights into the homeostatic behaviour of stem cells in the intestinal crypt. Investigating the same processes in human tissue has traditionally been more challenging, as transgenic lineage tracing approaches are clearly not feasible in this setting.

### 1.4.1 Early studies on hereditary conditions

Early observations in humans were based on data from rare hereditary genetic changes. Indeed, the serendipitous finding of an XO/XY mosaic patient who also had Familial Adenomatous Polyposis (FAP, this is a colorectal cancer-predisposing syndrome, meaning that a resection sample was available, the concurrence of these two genetic defects is estimated to occur at a frequency of 1 in 100 million births) showed that human crypts are also monoclonal, as crypts were either Y-positive or -negative, but never mixed. This means that each crypt originates from a stem cell that is either XO (hence Y-negative) or a stem cell that is XY. (Novelli *et al.*, 1996). Novelli and colleagues also studied a cohort of Sardinian women heterozygous for the X-linked glucose-6-phosphate dehydrogenase (G6PD) gene. Due to X-inactivation, cells in these individuals appeared either positive or negative for G6PD activity following histochemical staining. Assessment of intestinal tissue confirmed that human crypts are monoclonal, as individual crypts stained either entirely positively or negatively for G6PD activity (Novelli *et al.*, 2003). Importantly, all these studies were based on germline polymorphisms and clonal evolution that occurs very early in mammalian development. Therefore, they did not provide information about the dynamics and functional stem cell numbers in the adult.

### 1.4.2 Human lineage tracing

In recent years, new methods to conduct lineage tracing in adult human tissue samples have become available. Importantly, they all rely on observation of progeny from mutant stem cells, without specifically identifying the stem cells with markers. Human studies therefore quantify

the output of functional stem cells, without the use of any stem cell markers, in an analogous manner to the continuous labelling approach used by Kozar and colleagues (Kozar *et al.*, 2013).

### 1.4.2.1 Using mitochondrial DNA mutations

A number of studies have utilized mitochondrial DNA (mtDNA) mutations as a marker of clonal expansion of human intestinal stem cells. Mitochondria are the main cellular generators of adenosine triphosphate, the energy currency of the cell. They replicate independently from the host cell and contain their own genomes. Mitochondrial DNA (mtDNA) is more susceptible to mutations than genomic DNA owing to poor repair mechanisms, lack of protective histones and location in a highly oxidative environment (Taylor & Turnbull, 2005). Stochastic somatic mutations within mtDNA can lead to a loss of cytochrome c oxidase (CCO, also known as complex IV) enzyme activity, which can be detected by histochemistry. Somatic mtDNA mutations occur randomly and their incidence increases with age (Michikawa *et al.*, 1999; Taylor *et al.*, 2003). In addition, CCO mutations do not seem to confer any positive or negative selective pressure onto affected cells (Nooteboom *et al.*, 2010; Greaves *et al.*, 2012). Applying this system to ten patient samples ranging in age from 45 to 74 years, Gutierrez-Gonzalez and co-workers observed small intestinal crypts that were composed entirely of CCO<sup>-</sup> or CCO<sup>+</sup> cells as well as crypts containing a mixed population of cells. This supports the notion that human crypts are monoclonal and represents strong evidence for the existence of at least two functional stem cells within the crypt undergoing some form of competition (Gutierrez-Gonzalez *et al.*, 2009). Analogous experiments also demonstrated monoclonal conversion for human colonic crypts (Taylor *et al.*, 2003).

The same methodology was taken a step further by Baker and colleagues who cleverly exploited the characteristic architecture of the crypt to perform measurements on stacks of sections. The group examined changes in cell streams resulting from stem cell replacement events on the walls of colonic crypts and came to the conclusion that human intestinal stem cells conform to neutral drift dynamics with a functional stem cell number of five to six (Baker *et al.*, 2014). However, the use of mtDNA mutations is not without limitations. Each cell contains multiple mitochondria and it is not clear what fraction is mutated when a phenotype is observed. It is

---

also not clear in what way this mixed population of mitochondria is represented in the next generation of cells upon cell division. Modelling of stem cell dynamics with CCO should include a model of mitochondrial dynamics, which is not currently available. Hence, nuclear DNA mutations would represent more reliable markers.

#### 1.4.2.2 Using somatic clonal marks

Instead of mtDNA mutations, the Winton laboratory has recently developed a method based on so-called somatic clonal marks. These are nuclear DNA mutations that can be detected in human colonic tissue sections by (immuno)histochemistry. The method assumes that permanent genetic changes result in altered protein structures or abundance, causing a different staining pattern when specific antibodies or staining methods are used. Somatic clonal marks can be used for continuous labelling studies analogous to those performed by Kozar and colleagues. The following paragraphs describe this work which was recently published (Nicholson *et al.*, 2018).

##### 1.4.2.2.1 mPAS

Mild periodic acid Schiff (mPAS) staining is a well-established clonal mark. This histochemical method distinguishes O-acetylated (stain negative) from non-O-acetylated sialoglycoproteins (stain positive) (Veh *et al.*, 1982). In the human colonic mucosa, individuals exhibit one of three phenotypes: uniformly mPAS positive, uniformly mPAS negative or negative with rare dispersed positive crypts. The scattered pattern was never observed in children, which suggests that this phenotype arises due to spontaneous somatic mutations in a single gene in intestinal stem cells of heterozygous individuals (Fuller *et al.*, 1990). Monogenic inheritance was confirmed by examination of the distribution of the three staining patterns among different racial groups (Campbell *et al.*, 1994). The genetic cause of the phenotype is still unknown. Based on its role in O-acetylation of sialic acids, *CASDI* has emerged as a candidate gene (Arming *et al.*, 2011).

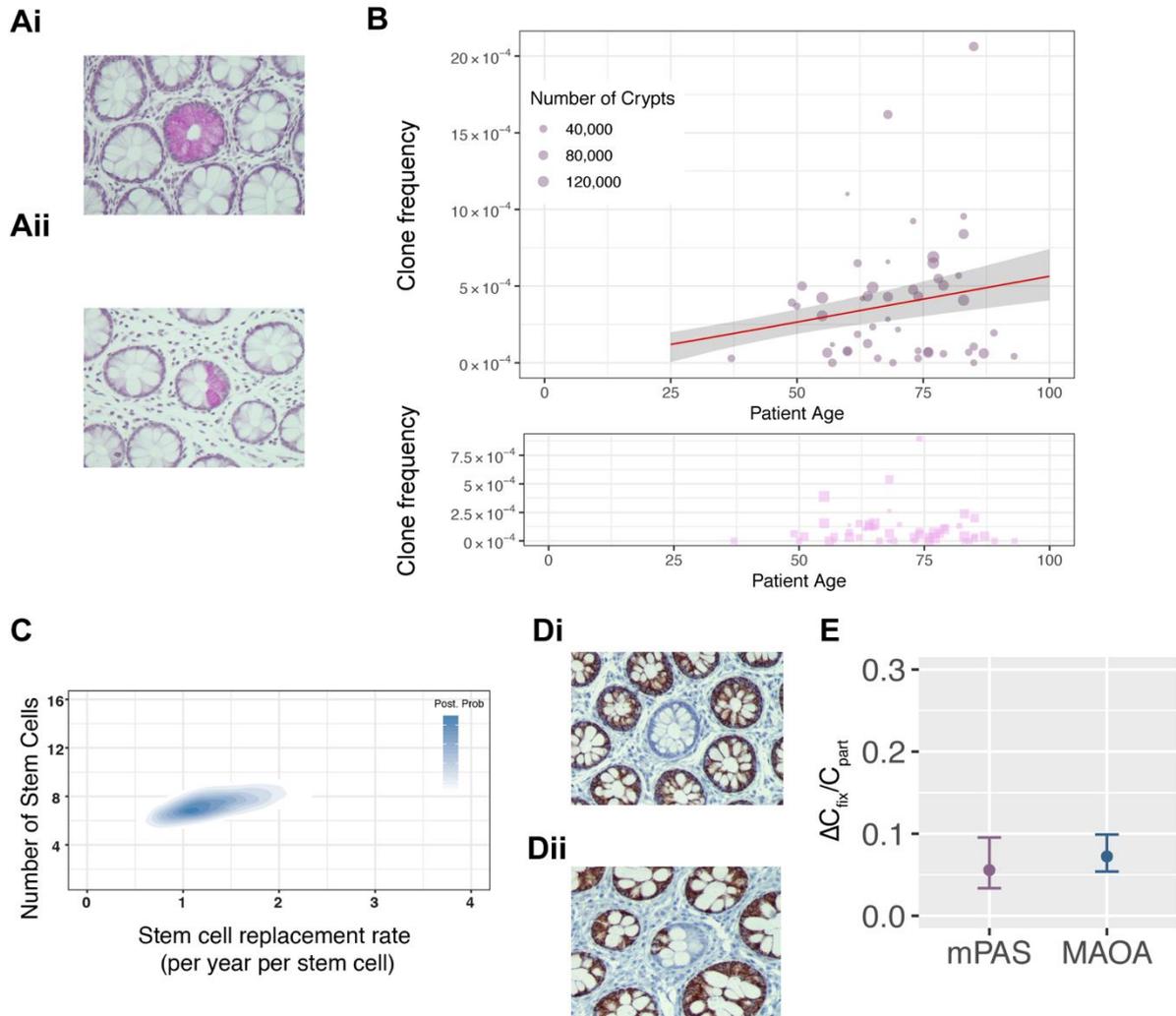
#### 1.4.2.2.2 Inference of human colonic stem cell dynamics using mPAS

Analogous to the continuous labelling approach used in the mouse, inference of human colonic stem cell dynamics using mPAS was based on scoring of frequencies of WPC and PPC in histologically normal colonic epithelial samples from individuals of a wide range of ages. (Figure 1.4A) This revealed an age-related increase in WPC with a slope,  $\Delta C_{fix}$ , of  $5.85 \times 10^{-6}$  crypts per year (95% CI:  $3.87 \times 10^{-6}$ – $9.26 \times 10^{-6}$ ) (Figure 1.4B top panel). As expected, the frequency of PPC ( $C_{part}$ ) remained constant (Figure 1.4B bottom panel). The rate of conversion of PPC to maintain  $\Delta C_{fix}$  indicated that monoclonal conversion of human crypts takes several years (median 6.3). The *de novo* mutation rate was calculated from single cell clones in the TA compartment and, combined with the experimentally derived values of  $\Delta C_{fix}$  and  $C_{part}$ , enabled inference of stem cell dynamics parameters. For the number of stem cells per human colonic crypt,  $N_{crypt}$ , we found 7 (95% CI: 5-10). The replacement rate  $\lambda_{crypt}$ , was calculated as 1.3 stem cell replacements/crypt/year (95% CI: 0.65-2.7) (Figure 1.4C). This is nearly 100-fold slower than the mouse (Kozar *et al.*, 2013).

#### 1.4.2.2.3 Validation using the mark MAOA

The rate of accumulation WPC will vary for different clonal marks because different loci will harbour different somatic mutation rates. For neutral marks, neutral drift will act on the PPC generated by somatic mutations, independently of the rate at which they appear. Hence, as the emergence of PPC is balanced by loss or expansion into WPC to maintain  $\Delta C_{fix}$ , the ratio of  $\Delta C_{fix}/C_{part}$  normalises for mutation rate. It can be used as a metric to describe the intra-crypt dynamics that lead to monoclonal conversion. For the neutral mark mPAS, we found  $\Delta C_{fix}/C_{part}$  to be 0.056 (95% CI: 0.034–0.095). To validate that this mark is indeed neutral, the staining and scoring was repeated for the mark monoamine oxidase A (MAOA Figure 1.4D). This mitochondrial enzyme is involved in catalysing the oxidative deamination of amines, such as dopamine, noradrenaline and serotonin (Bortolato, Chen & Shih, 2008). It is expressed in many tissues, including the intestine (Grimsby *et al.*, 1991) and is not involved in DNA repair or pro-oncogenic processes. Reassuringly, the ratio of  $\Delta C_{fix}/C_{part}$  for MAOA was calculated as 0.072

(95% CI: 0.054–0.099), incredibly similar to that of mPAS. This validated the stem cell dynamics derived using mPAS (Figure 1.4E).



**Figure 1.4 Inference of human stem cell dynamics using clonal marks.**

(A) *En face* tissue section with single mPAS-positive (i) WPC and (ii) PPC.

(B) *Top*: Regression analysis showing  $\Delta C_{fix}$  plotted in red with 95% margin of error in grey  
*Bottom*: Frequency of mPAS<sup>+</sup> PPC plotted against patient age.

(C) Heatmap representing probabilities for the indicated combination of stem cell number and replacement rate with blue indicating high likelihood.

(D) *En face* tissue section with single MAOA-negative (i) WPC and (ii) PPC.

(E) Boxplot showing similar ratio of  $\Delta C_{fix}/C_{part}$  for mPAS and MAOA. Error bars = 95% CI.

Interestingly, the replacement rate calculated by our method is about 100-fold slower than that derived by Baker and colleagues. A possible explanation may be that the ‘plumes’ of clones observed in their study were heavily enlarged by TA compartment proliferation, which may have caused dramatic overestimation of the stem cell replacement rate.

Derivation of the baseline behaviour of human colonic stem cells paves the way for investigations into non-homeostatic scenarios. A key question is, whether mutations can alter intestinal stem cell dynamics to promote tumour initiation, as has been shown in mice (Vermeulen *et al.*, 2013).

## **1.5 Mutations, stem cells and cancer in renewing tissues**

A textbook definition of cancer is that it is a disease of deregulated growth caused by somatic mutations. The generally accepted view is that cancer initiation occurs as a process of sequential accumulation of genetic alterations (Nowell, 1976). The origin of these mutations has recently been the focus of intense debate. One camp argues that cancer incidence can be explained by the number of stem cell divisions, suggesting that mutations occur mostly as a consequence of aberrant DNA replication (such notions are often referred to as the ‘bad luck’ hypothesis) (Tomasetti & Vogelstein, 2015). This is contradicted by a study demonstrating that cancer risk is heavily influenced by environmental and lifestyle factors (Wu *et al.*, 2016).

### **1.5.1 Mutation accumulation and colorectal cancer (CRC)**

Bowel cancer is the 4th most common cancer in the UK, accounting for 12% of all cases in females and males combined. There are currently over 40,000 new diagnoses per year as well as over 16,000 deaths (Cancer Research UK, 2015).

---

### 1.5.1.1 Hereditary colorectal cancer syndromes

About 10% of CRC cases are caused by hereditary cancer syndromes, most frequently Lynch syndrome (previously known as Hereditary Non-Polyposis Colorectal Cancer) and Familial Adenomatous Polyposis (FAP).

Lynch syndrome is caused by inherited mutations in one of the four mismatch repair genes *MLH1*, *MSH2*, *MSH6* and *PMS2* or mutations in the end of the *EPCAM* gene (Snyder & Hampel, 2019). This predisposes affected individuals for colorectal and endometrial cancer and increases their risk for a number of other cancers including ovarian, gastric, urothelial and hepatobiliary cancer (Staffa *et al.*, 2015; Yurgelun & Hampel, 2018; Snyder & Hampel, 2019).

FAP is caused by inherited mutations in APC, a negative regulator of Wnt-signalling, which controls stem cell homeostasis in the intestine (Fevr *et al.*, 2007). The syndrome is characterised by the formation of hundreds to thousands of intestinal polyps with an onset age of about 16 years. A diagnosis of colorectal cancer is essentially inevitable, with a mean age at diagnosis of 39 years. Therefore, patients are recommended to undergo regular colonoscopies from age 10-12 years onwards (Bussey, 1975; Snyder & Hampel, 2019) with total colectomy usually performed once adult (Campos, 2014).

### 1.5.1.2 CRC subtypes

Over the past decade, molecular analysis has revealed the broad spectrum of alterations in CRC. Analysis of genomic sequence, methylome and proteome from tumour samples from over 4,000 patients enabled the CRC subtyping consortium to define four consensus molecular subtypes (CMS1-4) of CRC (Guinney *et al.*, 2015). Strikingly, no single variable is sufficient to subtype a given tumour. This sets CRC apart from most other cancer types, in which classification is achieved by specific characteristics such as overexpression of particular genes. A prime example is breast cancer, whereby expression of the oestrogen receptor, progesterone receptor and human epidermal growth factor receptor 2 define the cancer subtype, which then guides treatment (Dai *et al.*, 2015).

A key criterium used by the CRC subtyping consortium is microsatellite instability (MSI), which can be detected in about 15% of CRC (Guinney *et al.*, 2015). Microsatellites, also known as short tandem repeats, are 1-6 bp long repeating stretches of DNA which, due to their repetitive nature, are prone to high error rate (Ellegren, 2004). Alterations in DNA mismatch repair genes (as they occur in Lynch syndrome described above) result in error-prone DNA replication, which manifests as unstable (length-changeable) microsatellites (Kim & Kang, 2014). It follows that MSI tumours, compared to their microsatellite stable (MSS) counterparts, harbour more and different somatic mutations. As they represent the majority, the following sections will focus on MSS disease.

### **1.5.1.3 The mutational landscape of CRC**

The majority of CRC occurs as a consequence of spontaneous somatic mutations. When the current list of frequencies of somatic mutations found in MSS CRC is plotted, a striking distribution emerges: Apart from a handful of frequently mutated genes, which have been referred to as ‘mutational mountains’ (Wood *et al.*, 2007), there are a number of mutations in the 5-20% mutation frequency range (‘hills’) as well as a remarkably long tail of drivers with low mutation frequencies (Figure 1.5 and Table 1.1) (data from Giannakis *et al.*, 2016 via cBioPortal). This indicates that a broad range of mechanisms may to varying degrees contribute to the transformation of colorectal tissue.

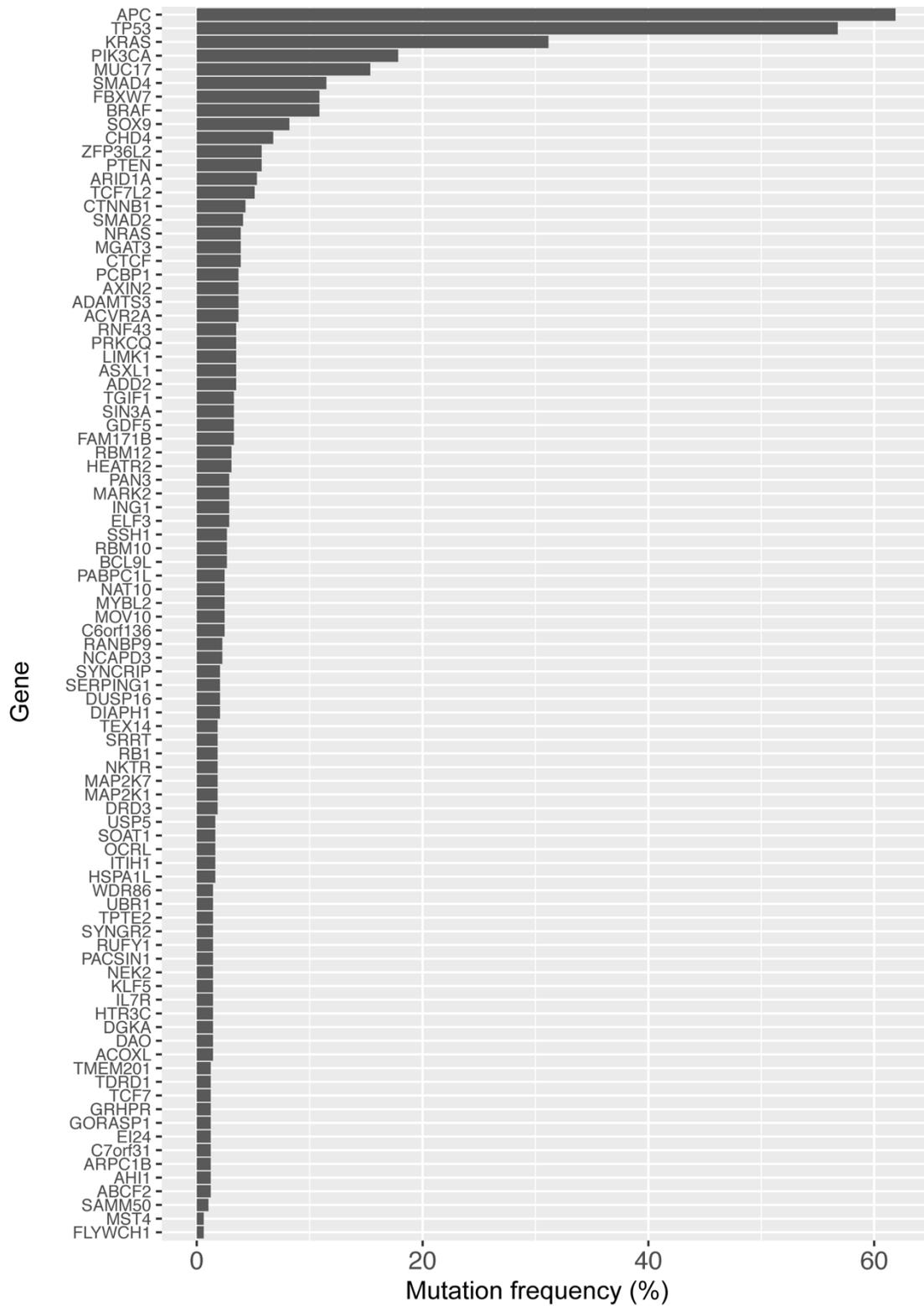


Figure 1.5 Long tail of frequently mutated genes in MSS colorectal cancer. Data obtained from Giannakis et al., (2016) and replotted using R.

Crucially, the mutational landscape of a tumour is merely a representation of the end stage of an evolutionary process occurring on the level of a cellular population. Therefore, these data do not provide insights into the roles and timing of different mutations during transformation. Such information has historically been gained from studying samples of intermediate stages from normal human tissue to malignant tumour.

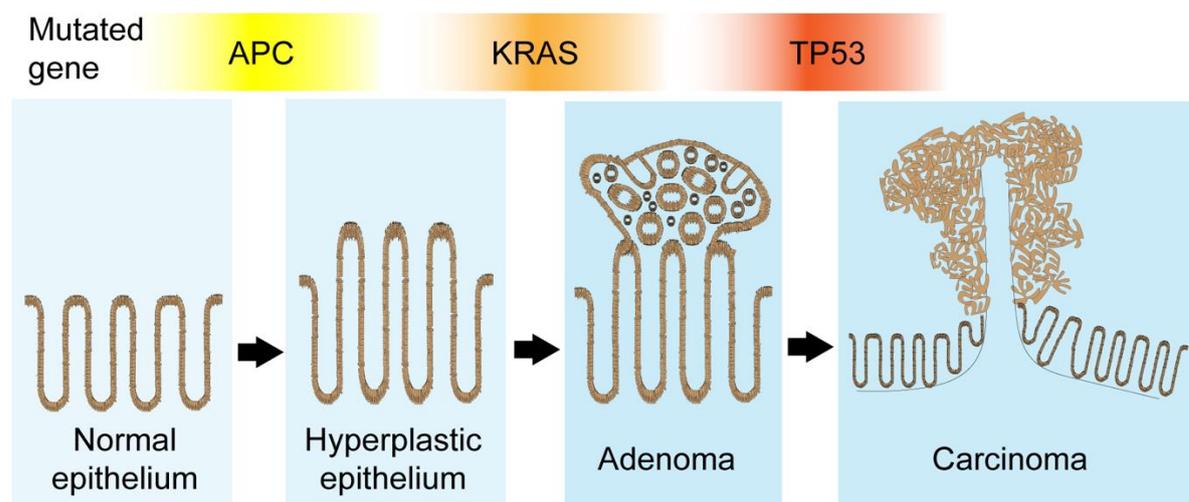
Table 1.1 **CRC driver genes with mutation frequency over 5%.**  
*Data from Giannakis et al., (2016)*

<b>Gene</b>	<b>Mutation frequency (%)</b>
APC	61.9
TP53	56.8
KRAS	31.1
PIK3CA	17.8
MUC17	15.4
SMAD4	11.5
BRAF	10.9
FBXW7	10.9
SOX9	8.2
CHD4	6.8
ZFP36L2	5.7
PTEN	5.7
ARID1A	5.3
TCF7L2	5.1

### 1.5.2 The Vogelstein model of colorectal cancer progression

In the 1980s and 90s, cancer progression was studied by examination of key genetic alterations in early, intermediate and advanced human colorectal lesions (Fearon & Vogelstein, 1990). This generated a number of fundamental observations: Firstly, tumours seemed to be

monoclonal. Secondly, *RAS* mutations were only present in adenomas exceeding 1 cm<sup>3</sup>. Finally, carcinomas exhibited more genetic alterations than early adenomas. Combining these findings led Fearon and Vogelstein to propose a step-wise model, in which CRC progression from benign adenomatous polyps to more advanced and finally invasive cancer and metastasis is driven by sequential accumulation of mutations in specific genes, which perturbs specific signalling pathways. According to this model at each step during progression, the newly acquired mutation generates a subclone with a competitive advantage, enabling it to outcompete its neighbours in a so-called clonal sweep. Due to the frequency of mutations found in human samples *APC* loss was defined as an initiating event, followed by mutations in *KRAS* and then *TP53* (Figure 1.6). Interestingly, their landmark paper also contains the speculation that *RAS*, rather than *APC*, mutations may act as the initiating event in certain tumours. It also proposes the notion that the number of mutations may be more important than their order (Fearon & Vogelstein, 1990). Importantly, a key aspect of the sequential model is that the growth of tumours requires substantial amounts of time, easily decades, which was validated by the association of colorectal cancer and age (Muto, Bussey & Morson, 1975).



**Figure 1.6 The Vogelstein model of CRC progression.**

*Loss of APC in normal epithelium drives a hyperplastic phenotype. Subsequent activating mutation of KRAS causes growth of an adenoma. This is transformed into a carcinoma by acquisition of mutations in TP53.*

### 1.5.3 The Big Bang model of CRC initiation

The Vogelstein model has recently been called into question, as numerous studies in both human tissue and mouse models have generated results that cannot be accommodated by the sequential model. In particular, the aspects of time, tumour heterogeneity and mutation order have exposed the necessity for refinement. From the sequential model with its predetermined sequence of mutations, relatively uniform speed of growth would be predicted for human tumours. However, serial endoscopies showed that different polyps grow at substantially different rates (Pickhardt *et al.*, 2013). Furthermore, synchronous polyps and adenocarcinomas were shown to be clonally related, also demonstrating that progression can occur at very different speeds despite similar genomic changes (Yang *et al.*, 2015). Stronger evidence against the sequential model comes from description of intratumour heterogeneity, which to a large extent has been enabled by advances in genomic sequencing technologies. Therefore, the clonal nature of tumours described in the 1990s is most likely related to the sequencing method. Regional sampling within individual adenomas and adenocarcinomas has revealed substantial intratumour heterogeneity, with subclones found throughout the tumours. This contradicts the notion of clonal sweeps (Sottoriva *et al.*, 2015) and is further supported by sequencing of adenomas and small polyps (Siegmund *et al.*, 2009; Sievers *et al.*, 2017). A final aspect of the sequential model is mutation order, whereby in- or hyperactivation of specific pathways is thought to be required at specific timepoints during cancer progression. Several mouse studies have challenged this view, whereby *PIK3CA* mutation lead to rapid development of invasive adenocarcinomas, circumventing the adenoma stage (Leystra *et al.*, 2012). Equally, simultaneous mutation of *APC*, *KRAS* and *PIK3CA* in mice still resulted in adenoma-carcinoma sequence, demonstrating that the order of mutation acquisition may not be key (Hadac *et al.*, 2015).

These and other findings have led Sottoriva and colleagues to propose the Big Bang model, whereby colorectal cancers arise as a single expansion event when a combination of factors reaches a critical threshold (Sottoriva *et al.*, 2015) (Figure 1.7). In this model, initiation occurs as step-wise mutation accumulation in absence of any phenotype. When a threshold is reached,

meaning transformation has occurred, colorectal cancer grows as a single expansion. During this growth further mutations are acquired, which can be detected as subclones. Of note the initiating mutations seem to determine the tumour phenotype, whereby certain tumours can be ‘born to be bad’, which can include early metastasis (Hu *et al.*, 2019).

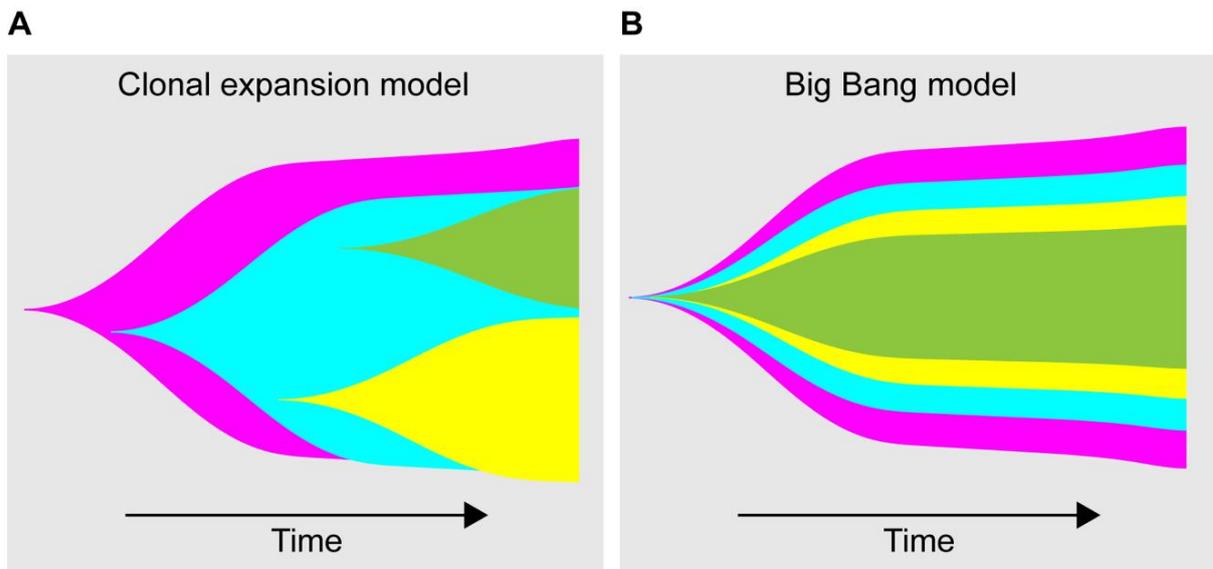


Figure 1.7 **Two models for CRC initiation.**

(A) In the Vogelstein model, sequential clonal sweeps populate the tissue.

(B) According to the Big Bang model, a tumour arises as a single growth when a threshold of factors is exceeded.

#### 1.5.4 Mutation acquisition and spread in renewing tissues

Both the sequential as well as the Big Bang model rely on the acquisition of mutations. To persist within renewing tissues such as the intestine, mutations must occur in functional stem cells. Or, to avoid semantic discussions: to expand into a tumour, the cells carrying the initiating mutations must continuously self-renew as well as produce progeny. Thus, at the moment of initiation, these cells act functionally as stem cells. In line with this notion, stem cells are seen as the cell of origin in cancer of the intestine (Barker *et al.*, 2008; Schwitalla *et al.*, 2013).

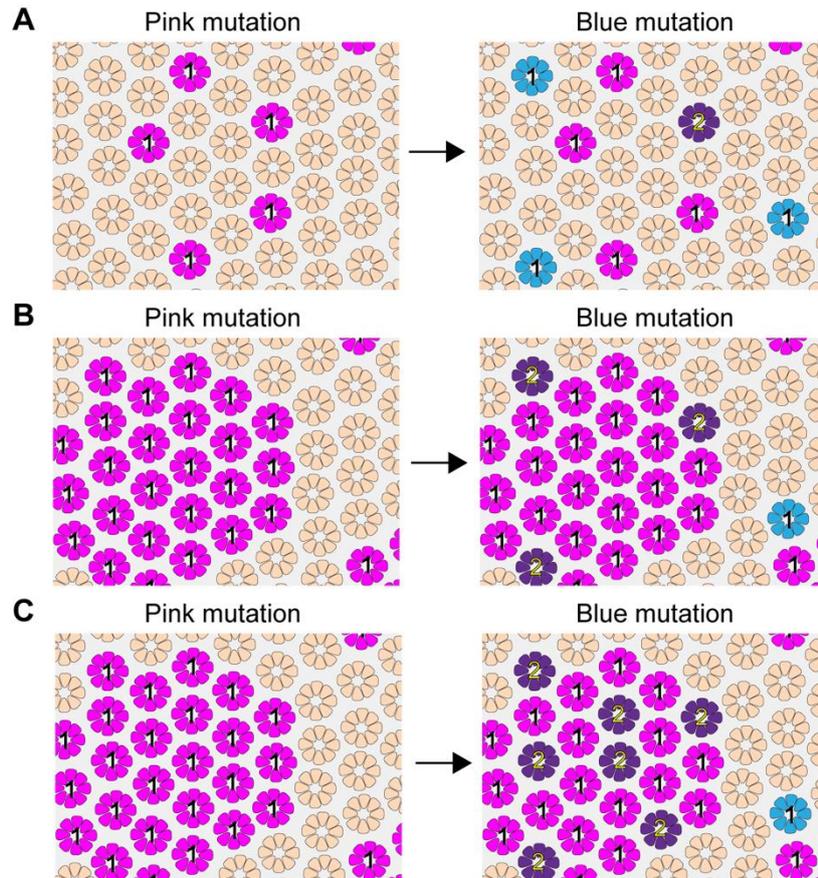
The acquisition of secondary (and subsequent) mutations, which will drive cancer progression, is favoured by expansion of stem cell-derived clones within the tissue, as this increases the pool

of cells available for additional mutation. This effect is further exacerbated by the spread of alterations that predispose for mutation accumulation. Figure 1.8 illustrates this process for the colonic epithelium. The idea of such expansion of mutations, termed ‘field cancerisation’ dates back two decades, when it was found that a proportion of individuals with oral cancer had multifocal tumours, suggestive of an underlying mutated ‘field’ (Garcia *et al.*, 1999). Equally, in the colon, tumours as well as their surrounding tissue were found to harbour the same *KRAS* mutations (Zhu *et al.*, 1997). Similar evidence for field cancerisation has in the meantime also been found in the bladder and prostate (Curtius, Wright & Graham, 2017).

The models of mutation accumulation and field cancerisation predict the presence of mutant clones in non-transformed aged tissues. Indeed, mathematical modelling indicated that at least half of the mutations in cancers of self-renewing tissues originate prior to tumour initiation (Tomasetti, Vogelstein & Parmigiani, 2013). Due to its easy accessibility, patterns of mutant clonal expansion were first documented in the hematopoietic system of aged humans (Welch *et al.*, 2012; Genovese *et al.*, 2014; Xie *et al.*, 2014). More recently, sequencing of microbiopsies of skin and oesophagus from aged individuals have found these tissues to be a mosaic of somatic mutations (Martincorena *et al.*, 2015, 2018). These findings were confirmed by analysis of mutant allele fractions in RNA-sequencing data from a range of normal tissues, which found macroscopic clonal expansions in 95% of samples tested (Yizhak *et al.*, 2019).

Importantly, the size of clonal expansions can fall below the sensitivity of such bulk approaches. Analysis of single cells or organoids from *in vitro* expanded single cells can generate a more comprehensive overview over the mutational landscape of normal adult tissue (Behjati *et al.*, 2014; Blokzijl *et al.*, 2016; Jager *et al.*, 2017; Lodato *et al.*, 2018). Sequencing of organoids derived from single human small intestinal and colonic cells suggests that mutations accumulate in this tissue at a rate of around 40 mutations per year (Blokzijl *et al.*, 2016). However, only a very small number of cells were analysed, therefore rarer mutations as well as the extent of clonal expansions were not quantified. The same holds true for a recent study employing laser capture microdissection and whole exome sequencing to describe the mutational landscape of just over 2,000 normal human colonic crypts. The authors found a large number of somatic mutations, mostly associated with the ageing signature. However, only a handful of mutations

in the cancer most common CRC drivers and no large-scale clonal expansions were detected (Lee-Six *et al.*, 2018). Again, this most likely relates to the low number of crypts probed. To gain more thorough insights into the accumulation of mutations in normal colon, larger areas need to be examined.



**Figure 1.8 Scenarios for second mutation acquisition.**

(A) *Left: The first mutation (pink) hits a number of crypts at random.*

*Right: The second mutation (blue) hits a number of crypts at random. By chance, one crypt ends up with both mutations (purple).*

(B) *Left: The first mutation (pink) hits a crypt and subsequently expands into a field.*

*Right: The second mutation hits a number of crypts at random. Due to the underlying field, several crypts end up with both mutations (purple).*

(C) *Left: The first mutation (pink) hits a crypt and subsequently expands into a field. This mutation predisposes towards accumulation of subsequent mutations.*

*Right: The mutation-promoting field results in an increased number of crypts carrying two mutations compared to situation (B).*

## 1.6 Summary and aims

Stem cells can be defined functionally through their ability to self-renew and differentiate. In adult mammals, key populations of stem cells fuel the turnover of renewing tissues such as the intestine. The functional unit of the intestinal epithelium is the crypt, which contains an equipotent population of functional stem cells that replace each other stochastically through neutral drift. In the mouse, mutations can lead to biases in this process, whereby affected stem cells have an increased probability of fixation within a crypt. CRC is thought to arise from intestinal stem cells that have acquired somatic mutations. According to the current Big Bang model, CRC is initiated when a critical threshold of pro-oncogenic factors is exceeded. In such a scenario, mutations that set cells onto a path towards cancer would originally arise in intestinal stem cells at the bottom of crypts, where they would ride the wave of neutral drift or even subvert it to increase their probability of fixation. Subsequent clonal expansion would occur by crypt division. Eventually, a tumour would be initiated when several driver mutations coincide.

Crucially, both mutation fixation in the crypt and expansion beyond determine the size of the pool of cells that can acquire further hits. Therefore, understanding these processes in normal tissue is key to understanding the earliest events during cancer evolution.

Within this framework the overarching goal of this project was to investigate the means by which mutations populate the normal human colon with age. This encapsulated the following aims:

1. To expand on the tools available for investigating human intestinal stem cell dynamics.
2. To determine the effect of gene-specific mutations on variant fixation within colonic crypts.
3. To investigate the rates and mechanisms of lateral expansion of mutant clones in the human colon.

## **Chapter 2 Materials and methods**

### **2.1 Human tissue**

Normal colon tissue samples were obtained from both Addenbrooke's Hospital Cambridge and Norfolk and Norwich University Hospital under full local research ethical committee approval (Approval document IDs 15/WA/0131 & 17/EE/0265 as well as 06/Q0108/307 & 08/H0304/85, respectively) according to UK Home Office regulations. A total of 351 individuals were included in the study with an age range of 8–93 years (Appendix A). Colectomy specimens were fixed in 10% neutral buffered formalin. From areas of tissue without macroscopically visible disease mucosal sheets were removed from the specimens and embedded *en face* in paraffin blocks. For standard stainings, sections were cut from each sample at 5 µm thickness and mounted onto charged glass slides.

### **2.2 mPAS staining**

Sections were de-waxed and rehydrated on a Leica Multistainer ST5020 before washing in 0.1 M Acetate buffer pH 5.5 at 4 °C for 5 minutes. This was followed by oxidation in 1 mM sodium periodate buffer at 4 °C for 10 minutes before washing in 1% glycerol for 5 minutes. Then, three washes were performed in ultra-pure water for 5 minutes in total, followed by staining in Schiff's reagent for 15 minutes. Sections were then washed again in ultra-pure water and counter-stained in Mayer's Haematoxylin for 40 seconds. After another wash in ultra-pure water and brief blueing in tap water a final rinse in ultra-pure water was performed. This was followed by dehydration, clearing and mounting in DPX on the same Leica Multistainer.

## **2.3 Immunohistochemistry**

### **2.3.1 Standard protocol**

Sections were de-waxed and re-hydrated on a Leica Multistainer ST5020. This was followed by heat-induced epitope-retrieval in citrate buffer (10 mM sodium citrate, pH6) in a medical pressure cooker and washing in PBS. All subsequent washes were 3 x 5 min in PBS-T (PBS with 0.05% Tween-20). After a 15 min incubation with 3% H<sub>2</sub>O<sub>2</sub> in methanol and a wash, slides were incubated with blocking buffer (PBS-T with 10% Donkey Serum, Dako) for 30 min. This and all following incubations were performed at room temperature in a humidified slide box. Slides were then incubated with the primary antibody (Table 2.1) for one hour at room temperature or overnight at 4 °C. After a wash, slides were incubated with the secondary antibody (biotin-SP-conjugated AffiniPure donkey anti-mouse, anti-rabbit or anti-goat, Jackson ImmunoResearch, all 1:500 in PBS-T) for 40 min. Following a wash, slides were incubated with Vectastain® Elite® ABC reagent (Vector Laboratories) for 40 min. This was followed by a final wash and immunoperoxidase detection using a liquid DAB + substrate chromogen system (Dako). Finally, sections were counterstained with Mayer's Haematoxylin, dehydrated and mounted in DPX on the Leica Multistainer.

### **2.3.2 Adaptations for laser capture microdissection**

For staining of sections for laser capture microdissection the protocol was largely identical apart from the following changes: Tissue was cut at 10 µm thickness onto UV-irradiated PEN membrane slides (ZEISS). Heat-induced epitope retrieval was performed in citrate buffer in a water bath at 76 °C for 16 hours. Counterstaining with Mayer's Haematoxylin was performed manually for 15 seconds followed by blueing in tap water for 1 minute and drying at room temperature. Slides were stored at room temperature (short-term) or -20 °C (long-term).

Table 2.1 **Primary antibodies tested for IHC.**

*N/A denotes that no suitable titre was found for staining of human colonic FFPE sections.*

<b>Antigen</b>	<b>Antibody</b>	<b>Supplier</b>	<b>Titre</b>
ADGRG4	HPA017372	HPA	1:50
ATRX	HPA001906	HPA	1:500
BRAFV600E	E19290	Spring Bioscience	1:200
DDX3X	HPA059585	HPA	N/A
DUSP6	ab76310	Abcam	1:100
H3K27me3	ab6002	Abcam	1:1000
HDAC6	HPA003714	HPA	1:200
HDAC6	#7558	CST	1:600
HTATSF1	HPA000504	HPA	1:200
HUWE1	HPA002548	HPA	1:300
HUWE1	ab70161	Abcam	1:400
HUWE1	OALA01854	Aviva	1:300
IDH3G	HPA000425	HPA	N/A
KDM6A	HPA002111	HPA	1:100
KDM6A	#33510	CST	1:200
KRASG12D	#14429	CST	1:250
MAOA	SC-271123	Santa Cruz	1:200
NONO	HPA054559	HPA	1:500
NONO	HPA054094	HPA	1:2000
p-4EBP1	#2855	CST	1:600
p53	ab1101	Abcam	1:1000
PDHA1	HPA063053	HPA	1:700
PTEN	#9559	CST	1:300
PTEN	6H2.1	Sigma-Aldrich	1:200
RBMX	HPA057707	HPA	1:100
RPS6KA	HPA003221	HPA	1:1000
RPS6KA	A302-460A	Bethyl	1:2000
SH3KBP1	HPA003351	HPA	1:500
SLC25A5	HPA071684	HPA	N/A
SMAD4	ab217267	Abcam	1:400
SMC1A	HPA063884	HPA	1:300
SMC1A	ab133643	Abcam	1:400
SMC1A	orb94671	Biorbyt	1:200
STAG2	HPA002857	HPA	1:100
STAG2	LS-B11284	LSBio	1:1000
STARD8	HPA060788	HPA	N/A
XIAP	HPA042428	HPA	1:1000
ZMYM3	HPA003211	HPA	1:500
ZNF75D	HPA004705	HPA	1:50

## 2.4 Clonal marks data acquisition and analysis

### 2.4.1 Clone and crypt counting

WPC and PPC numbers as well as multicrypt patch sizes were manually scored in stained tissue sections using a standard brightfield microscope. Sections were scanned at 20X using a Leica Aperio AT2 scanner and stored as .svs files on a CRUK Cambridge Institute server. Scans were examined online using the Aperio eSlide Manager (Leica) or downloaded and annotated using QuPath (Bankhead *et al.*, 2017). To count the total number of crypts on each section, .svs files were analysed with the DeCryptICS neural network developed by Dr Edward Morrissey and Dr Doran Khamis (<https://github.com/MorrisseyLab/DeCryptICS>, manuscript in preparation).

### 2.4.2 Scoring fusion

To score crypt fusion, the term ‘fufi’ was invented to agnostically refer to structures that could represent a crypt undergoing fusion or fission in *en face* human colonic tissue sections. A fufi was defined as two crypts joined together without any visible gap but with two clearly distinguishable lumina. For stained sections, fufis can consist of two wild type crypts (WT fufi), two mutant crypts (mut fufi) or a WT and a mut crypt (WT-mut fufi). To count the total number of fufis on each section, .svs files were analysed with DeCryptICS followed by manual annotation in QuPath to classify fufis as WT, mut or WT-mut.

### 2.4.3 Clone and fufi data analysis

Data was analysed and plotted using R.

### 2.4.4 Patch size frequencies

For all marks, the total numbers of patches consisting of 2-14 crypts as well as  $\geq 15$  crypts were determined and frequencies calculated from division by total clone numbers (whereby a patch is a clone).

#### **2.4.4.1 All other clonal marks data analyses**

The age-related behaviour of WPC and PPC for HDAC6, KDM6A, MAOA, mPAS, NONO, p53, PTEN and STAG2 was used to infer intra-crypt behaviour as well as crypt fission rates associated with gene-specific mutations. This was performed by Dr Edward Morrissey and is outlined in detail in Appendix C.

The frequencies of WT-mut as well as mut fufis combined with the patch sizes for mPAS and KDM6A were used by Dr Doran Khamis to infer the crypt fusion rates associated with these two clonal marks. KDM6A fufi data was further used to infer the duration of crypt fusion. Details of mathematical models used can be found in Appendix E.

## **2.5 Laser capture microdissection and sequencing**

### **2.5.1 Laser capture microdissection and sample lysis**

On stained FFPE human colonic tissue sections (see 2.3.2), crypts of interest were harvested into lids of 0.2 mm radius PCR tubes using a Leica LMD7000 Laser Microdissection System. 10 µl of Proteinase K solution from the Arcturus® PicoPure® DNA Extraction Kit (ThermoFisher) were added and tubes centrifuged in a mini-centrifuge for PCR tubes. Following lysis for 3h at 65 °C in a standard PCR block the enzyme was inactivated by incubation at 95 °C for 10 min. Samples were kept at 4 °C until PCR.

### **2.5.2 Assessment of FFPE DNA quality**

#### **2.5.2.1 By quantitative PCR**

DNA was extracted from laser capture microdissected crypt sections as described above as well as from FFPE sections using the QIAmp DNA FFPE Tissue Kit (QIAGEN) according to the manufacturer's instructions. Human female genomic DNA (Promega) was used as a control. Quantitative PCR analysis was performed using a TaqMan™ probe for Atoh1, which generates an 83-bp amplicon (Assay ID: Hs00944192\_s1, ThermoFisher) using the supplier's reagents

and protocol in a QuantStudio™ 6 Flex Real-Time PCR system. Each 20 µl reaction contained 10 µl TaqMan™ Universal PCR Mastermix, 1 µl nuclease-free water, 1 µl TaqMan™ probe and 8 µl sample at the relevant DNA concentration.

### **2.5.2.2 By PCR with varying amplicon size**

For assessment of optimal amplicon size, DNA extracted from FFPE sections as described above was diluted to equivalents of 5, 20 and 100 crypts and used for PCR reactions containing primers at 1 µM concentration (Appendix B: Table B.1), 0.5 mM dNTPs (New England BioLabs), 5 µl 5X Phusion® HF Reaction Buffer (New England BioLabs), 1 U of Phusion® High-Fidelity DNA Polymerase (New England BioLabs) and nuclease-free H<sub>2</sub>O (Ambion) to make up a volume of 25 µl. PCR cycling was performed at 95 °C for 2 min for one cycle, followed by 35 cycles at 95 °C for 10 s, 60 °C for 10 s and 72 °C for 15 s. The final cycle was followed by a 5 min extension at 72°C.

## **2.5.3 Sequencing library preparation**

### **2.5.3.1 Pre-amplification**

Due to low quantities of input DNA a pre-amplification PCR was performed. All primers were designed using Primer3 (Rozen & Skaletsky, 2000). They included Fluidigm CS1 and CS2 tags and were ordered from Sigma (Appendix B: Tables B.2 and B.3). Reactions were performed in multiplex primer groups with primers each at 1 µM concentration. For *MAOA*, two primer groups (odd and even numbered primers) were used. For *KDM6A*, three groups were used (group 1: primers 1,4,7...group 2: 2,5,8... group3: 3,6,9...). The LCM DNA sample was diluted such that each PCR reaction contained 10 µl DNA sample. In addition, each reaction contained 0.5 mM dNTPs (New England BioLabs), 5 µl 5X Phusion® HF Reaction Buffer (New England BioLabs), 1 U of Phusion® High-Fidelity DNA Polymerase (New England BioLabs) and nuclease-free H<sub>2</sub>O (Ambion) to make up a volume of 25 µl. PCR cycling was performed at 95 °C for 2 min for one cycle, followed by 35 cycles at 95 °C for 10 s, 60 °C for 10 s and 72 °C for 15 s. The final cycle was followed by a 5 min extension at 72°C. Samples were then treated

---

with ExoSAP-IT enzyme (2  $\mu$ l of enzyme for 5  $\mu$ l of sample, ThermoFisher) at 37 °C for 15 min followed by a 15 min inactivation at 80 °C. Samples were then diluted 1:10 in DNA Suspension Buffer (Teknova) and stored at 4 °C until further processing.

### **2.5.3.2 Individual amplification**

To obtain adequate amounts of DNA for sequencing, the pre-amplified products were further amplified using the Fluidigm Access Array™ according to the supplier's protocol. Primers with Fluidigm CS1 and CS2 tags used in the chip were ordered from Sigma (Appendix B: Tables B.2 & B.3).

### **2.5.3.3 Barcoding**

A unique Fluidigm barcode was added to each sample by PCR in 10  $\mu$ l reactions using the Fast Start High Fidelity PCR System (Roche) containing: 400 nM barcoding primers, 1  $\mu$ l diluted PCR product, 1X Fast Start HF buffer without MgCl<sub>2</sub>, 4.5 mM MgCl<sub>2</sub>, 5% DMSO, 0.2 mM dNTPs each, 0.05 U/ $\mu$ l High fidelity enzyme mix and nuclease-free H<sub>2</sub>O (Ambion) to make up a reaction volume of 10  $\mu$ l. PCR cycling was performed using the protocol: 95°C for 10 min followed by 15 cycles at 95 °C for 15 s, 60 °C for 30 s and 72 °C for 60 s. The final cycle was followed by a 3 min extension at 72 °C.

### **2.5.3.4 Purification and size selection**

Samples were then pooled, purified using a Clean & Concentrator Kit (Zymo Research) and primer dimers eliminated by broad range (200-400 bp) size selection using a PippinBlue (Sage Science). Samples were kept at 4 °C until submission for sequencing.

## **2.5.4 Next-generation sequencing**

Samples were submitted to the in-house genomics core facility for sequencing on the Illumina MiSeq platform. Samples from one Fluidigm Access Array™ were sequenced on one lane using 150 bp paired-end sequencing with 10% PhiX.

### 2.5.5 Sequencing data analysis

Fastq files were converted to .BAM files and analysed using a PERL script (available upon request) by Dr Richard Kemp. Briefly, from the bulk of reads the script first identifies reads corresponding to the amplicons of interest by pulling out reads starting and finishing with the expected sequence as well as containing an expected stretch of sequence in the middle. For these reads, at every nucleotide position outside of the primer sequence, the number of reads corresponding to the reference genome as well as the number of reads containing a base change at that position are recorded. This enables calculation of the noise at every position. Candidate mutations were identified when the mutant allele frequency was either  $>4x$  the mean of the noise at that position or  $>3.29 x$  the standard deviation at that position ( $p \leq 0.001$ ). True mutations were called if they were present in all samples originating from the same patch in serial sections but absent in all wild-type samples from the same sections. To complement this analysis and identify insertions and deletions the script was modified to a version that scans the reads using pairs of four nucleotides corresponding to the reference genome. If a match is found, the base in between is recorded (Figure 2.1). Importantly, if there is a mismatch between one of the quadruplets of nucleotides and the read, the script will not record the base in between. Thus, for a single nucleotide variation, the script will record 4 positions with no reads, followed by a position with reads and then again 4 positions without (Figure 2.1A). For a 1 base deletion, there will be a continuous mismatch for 9 positions (Figure 2.1B). Following this analysis SNPs, SNVs and deletions are easily identified by plotting the total reads against nucleotide position. Dips with a peak represent a SNV or a SNP, dips lacking a peak appear due to a deletion or insertion (Figure 2.1). Again, mutations were considered real if they appeared in all samples corresponding to the same patch in serial sections but were absent in surrounding wild type crypt samples.

This analysis was complemented by manual inspection of reads using the Integrative Genomics Viewer by the Broad Institute (Robinson *et al.*, 2011).



## 2.6.2 Library preparation

### 2.6.2.1 PCR amplification

Primer pairs (Appendix B: Table B.4 ) flanking *KRAS* codons 12 and 13 as well as the mimic area in the mimic gene *PITPNM2* were designed using Primer3 (Rozen & Skaletsky, 2000). Fluidigm CS1 and CS2 tags were added to all primers. For amplification, two methods were used. Initially, 188 samples were amplified in two independent PCR reactions using *KRAS\_2* primers at 1  $\mu$ M concentration as well as 1  $\mu$ l DNA sample, 0.5 mM dNTPs (New England BioLabs), 5  $\mu$ l 5X Phusion® HF Reaction Buffer (New England BioLabs), 0.5 U of Phusion® High-Fidelity DNA Polymerase (New England BioLabs) and nuclease-free H<sub>2</sub>O (Ambion) to make up a volume of 25  $\mu$ l.

For subsequent experiments, the mimic amplicon method as outlined in section 5.3 was used, which required amplification of each DNA sample in two independent duplex PCR reactions containing the *KRAS\_1* and corresponding *MIMIC\_1* primers or *KRAS\_2* and corresponding *MIMIC\_2* primers, respectively. To balance the PCR reactions, *KRAS* and *MIMIC* primers were present at concentrations of 1.2  $\mu$ M and 0.8  $\mu$ M, respectively. Reactions also contained 1  $\mu$ l DNA sample, 0.5 mM dNTPs (New England BioLabs), 5  $\mu$ l 5X Phusion® HF Reaction Buffer (New England BioLabs), 0.5 U of Phusion® High-Fidelity DNA Polymerase (New England BioLabs) and nuclease-free H<sub>2</sub>O (Ambion) to make up a volume of 25  $\mu$ l.

For both methods, PCR cycling was performed as: 95 °C for 2 min for one cycle, followed by 35 cycles at 95 °C for 10 s, 60 °C for 10 s and 72 °C for 15 s. The final cycle was followed by a 5 min extension at 72 °C.

### 2.6.2.2 Barcoding and next-generation sequencing

Prior to barcoding, PCR products were diluted 1:15 in DNA suspension buffer (Teknova). Barcoding to add a unique identifier was then performed as described in section 2.5.3.3. Libraries consisting of barcoded PCR products from 192 samples were then cleaned up using a Clean & Concentrator Kit (Zymo Research), primer dimers eliminated by broad range (200-

---

400 bp) size selection using a PippinBlue (Sage Science) and then submitted for 150 bp paired-end sequencing with addition of 15% PhiX on the Illumina MiSeq platform in the in-house genomics facility.

### 2.6.3 Targeted KRAS sequencing data analysis

Corresponding forward and reverse reads were combined into a single consensus sequence using PANDAseq 2.11 with default options (Masella *et al.* 2012). Artefactual sequences were removed on the basis of not beginning and ending with the forward and reverse gene specific primer sequence and incorrect overall length ( $\pm 3$ bp tolerance).

#### 2.6.3.1 Mutation calling in single amplicon method

This method was used for analysis of the first 188 samples, which were included in our recent publication (Nicholson *et al.*, 2018). Only samples with read number  $\geq 1000$  were considered. The frequency of all nucleotides at all amplicon positions was calculated using a custom PERL script written by Dr Richard Kemp (available on request). A minimum of 10 variant reads was required to estimate mutant allele frequency (MAF) for any given nucleotide position. The mean allele frequency and standard deviation were calculated for all samples on a sequencing run for each particular amplicon position and nucleotide. Mutations were called in a sample if the estimated MAF exceeded either  $4 \times$  the mean allele frequency or the mean allele frequency + 3.209 standard deviations of the mean allele frequency in both KRAS replicates, corresponding to a p-value of  $< 0.001$ . The actual MAF for a particular mutation was calculated by subtracting the mean allele frequency (corresponding to noise) from the estimated MAF.

#### 2.6.3.2 Mutation calling in mimic amplicon method

This method was applied to all subsequent samples. Mutation calling focused on codons 12 and 13 of the KRAS gene and used a custom PERL script written by Dr Richard Kemp. For each DNA sample, reads containing the sequences corresponding to wild type (GGTGGC) as well as all possible mutations of codons 12 and 13 were extracted (Table 2.2). This yielded a MAF for every possible mutation in all four amplicons for every DNA sample and thus also revealed

the noise at every position. *KRAS* mutations were called if 1) >1000 reads were obtained for both *KRAS* amplicons and at least one mimic amplicon and 2) the MAF in both *KRAS* amplicons was > 0.1% (corresponding to at least 10 mutant reads) but found at background levels in the mimic amplicons. These criteria correspond to  $\geq 1.96$  standard deviations or a p-value of < 0.025 for the noisiest nucleotide position (KRASG12D). The actual MAF for a particular mutation was calculated by subtracting the mean allele frequency (corresponding to noise) from the estimated MAF.

**Table 2.2 Sequence context used to extract mutant reads from sequencing data.**  
For each variant, the frequency was calculated by dividing by the total read number.

Context	Nucleotide change	Amino acid change
GAGCTGGTGGCGTA	WT	WT
GAGCTGATGGCGTA	G>A	G12D
GAGCTGCTGGCGTA	G>C	G12A
GAGCTGTTGGCGTA	G>T	G12V
GAGCTAGTGGCGTA	G>A	G12S
GAGCTCGTGGCGTA	G>C	G12R
GAGCTTGTGGCGTA	G>T	G12C
GAGCTGGTGACGTA	G>A	G13D
GAGCTGGTGCCGTA	G>C	G13A
GAGCTGGTGTCGTA	G>T	G13V
GAGCTGGTAGCGTA	G>A	G13S
GAGCTGGTCGCGTA	G>C	G13R
GAGCTGGTTGCGTA	G>T	G13C
GAGCTGGAGGCGTA	T>A	WT
GAGCTGGCGGCGTA	T>C	WT
GAGCTGGGGGCGTA	T>G	WT
GAGCTGGTGGAGTA	G>T	WT
GAGCTGGTGGGGTA	G>C	WT
GAGCTGGTGGTGTGTA	T>C	WT

### **2.6.3.3 Mathematical modelling of KRAS clone behaviour**

The MAFs obtained by analysis of targeted sequencing data described in 2.6.3.1 and 2.6.3.2 were used by Dr Edward Morrissey to model the behaviour of *KRAS*-mutant clones. The mathematical methods used are described in Appendix D.



## **Chapter 3 Clonal marks discovery**

### **3.1 Introduction**

Lineage tracing experiments undertaken in transgenic mice have informed our current understanding of intestinal stem cell dynamics. For human studies however, such approaches are not feasible. Therefore, insights are largely gained by visualisation of spontaneous mutations in samples from cohorts of individuals from a wide range of ages. Conceptually, this type of study is analogous to the continuous labelling approach taken by Kozar and colleagues, which defined the number and replacement rate of effective stem cells in the mouse intestine (Kozar *et al.*, 2013). In our laboratory we use visualisation of genetic alterations in nuclear DNA as a tool for human lineage tracing. We refer to such visualisable events as clonal marks.

#### **3.1.1 The concept behind human clonal marks**

Human clonal marks rely on visualisation of somatic mutations. The method is predicated on genetic changes resulting in a different staining pattern when specific antibodies or histochemical methods are used. Somatic mutations can cause translation of a structurally altered or truncated protein, which abrogates binding of specific antibodies. Mutations may also occur in regulatory elements, leading to up- or downregulation of gene expression. Changes in protein abundance may further be caused by heritable epigenetic events. For most of our marks we use IHC, whereby immunodetection is carried out via horseradish peroxidase-catalysed oxidation of 3,3'-Diaminobenzidine (DAB), resulting in a brown precipitate. Mayer's Haematoxylin is used as a counterstain, which yields blue nuclear staining. An exception is the clonal mark mPAS, which is detected by a different histochemical method (see 2.2) and which we previously used to infer human intestinal stem cell dynamics (Nicholson *et al.*, 2018).

Clonal marks can be positive or negative. In the latter case, genetic alterations cause reduction or loss of immunoreactivity, resulting in mutant clones appearing blue on a brown background. For positive marks, genetic alterations lead to upregulation or stabilisation of a protein that is normally absent or rapidly turned over. Mutations may also introduce very specific amino acid changes that can only be detected by antibodies against that particular mutant form. In both cases, rare dark brown clones are visualised on a pale background. For both types of clonal mark, the same set of technical and biological requirements apply.

### **3.1.2 Requirements for clonal marks**

This section formalises, in increasing order of stringency, the requirements candidate clonal marks need to fulfil in order to be applied.

#### **3.1.2.1 Uniform staining pattern**

The most basic requirement for a candidate clonal mark is that all intestinal epithelial cells stain evenly. Failure to meet this criterion precludes distinction of real negative or positive cells from artefacts caused by the plane of sectioning. Figure 3.1A illustrates this point. Variability of protein expression along the crypt axis can, when the epithelium is not perfectly horizontally aligned within the paraffin block, lead to a variegated staining pattern in which identification of true positives or negatives is not possible. A further exclusion criterium is cell-state or cell-type specific protein expression, as demonstrated in Figure 3.1B. Rare positive or negative cells could in such a case be falsely identified as crypts partially populated by mutant clones, corrupting the data.

Related to uniformity of staining is saturation. For negative marks, strong saturated staining is necessary for unequivocal detection of negative clones. Equally, for positive marks, dark clones are required for identification against the background. Antibody titres are optimised to achieve these saturated outcomes. However, depending on the price of a specific antibody, the required titre may result in a cost per section that exceeds the budget of the project (> £5 per section). Such antibodies are not pursued further.

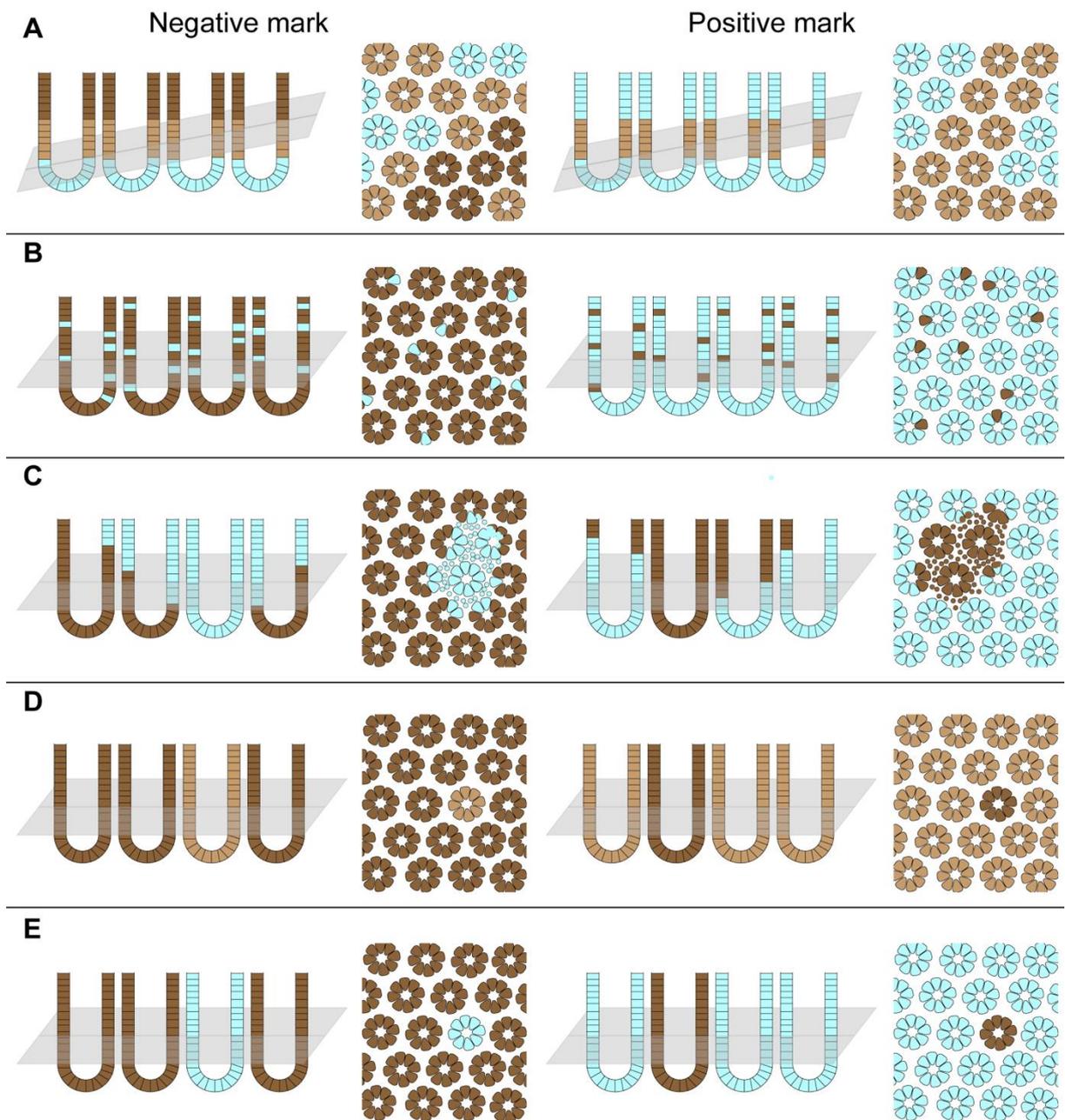


Figure 3.1 **Clonal marks scenarios.**

*A-D are not suitable as clonal marks.*

*A) Protein expression variability along the crypt axis can result in a variegate pattern on transverse sections when the plane of sectioning is not orthogonal to the crypt axis.*

*B) Expression in specific cell types is indistinguishable from partially populated crypts in transverse sections.*

*C) Non-confined clones with stromal immuno-reactivity are challenging to interpret.*

*D) Weak immunoreactivity complicates clone identification.*

*E) Useable clonal marks yield confined clones that are easily identified within the uniformly different background.*

### **3.1.2.2 Confined clones**

Mutations detected as clonal marks are assumed to occur in stem cells at the bottom of the crypt and become fixed by the process of neutral drift. Clones may thus manifest as crypts partially populated by mutant cells or crypts that are entirely mutant. The ability to identify the former provides strong support for the genetic origin of a clonal mark, because it demonstrates the ability to define cellular boundaries separating the clone from wild type epithelium where juxtaposed within a single crypt. Importantly, the biology of the intestinal epithelium precludes spread of a genetic event occurring in intestinal stem cells to non-epithelial cells. Therefore, areas containing epithelial as well as stromal reactivity could be indicative of local insults including infection or damage or could be an artefact relating to tissue fixation or sectioning thickness. A telling characteristic of such artefacts is altered staining that is not confined to cellular boundaries. As these staining patterns are generally indicative of non-genetic events, such marks are abandoned. Of note, stromal reactivity could be a direct reaction to the presence of a mutant crypt, in which case the clonal mark may be useable. However, distinguishing such an effect from artefacts would require mechanistic insights, which exceeds the scope of the project. Therefore, any marks displaying local areas of stromal and epithelial reactivity are not pursued further (Figure 3.1C).

### **3.1.2.3 Distinctive clones**

The use of clonal marks relies on staining of hundreds of samples followed by manual scoring of clones. Their easy identification is a prerequisite for time-efficient data acquisition. It is conceivable that certain mutations, for example of the missense type, will only moderately change the protein structure and thus also immunoreactivity, resulting in clones that are only marginally different from the surrounding wild-type crypts (Figure 3.1D). This may significantly increase the time required for scoring to a point where it is impractical. Further, faint clones may lead to uncertainties in slightly thicker/thinner areas of a section. Therefore, alternative antibodies from different suppliers must be tested to improve on the staining pattern

(Figure 3.1E) (see 3.1.2.4). If this fails to yield more distinctive clones, the protein is not pursued further.

### 3.1.2.4 Independent antibody validation

Screening of large area sections may identify weakly or strongly immuno-reactive epithelium present as confined clones. In both cases, this is indicative of mutations leading to changes in the protein in question. However, in absence of positive or negative controls the specificity of the antibody needs to be validated. Thus, even when screening identifies distinctive clones for a particular candidate mark, serial sections are stained with an independent antibody recognising a different epitope within the protein. If this identifies the same crypts as clones, the mark is validated. Further, when the recognised epitope is conserved across different species, staining of tissues from genetically engineered knock-out animals can provide an additional layer of validation.

### 3.1.2.5 Validation by sequencing

In addition to using an independent antibody, clonal marks could in theory also be validated by targeted sequencing, which would require isolation of clones followed by DNA extraction. However, to date no suitable protocol was available in the Winton laboratory.

### 3.1.2.6 Clone frequency

In practical terms, whether a validated clonal mark is useable or not heavily relies on the frequency at which clones are detected. For the previously used marks MAOA and mPAS, clones occur at frequencies of approximately 1 in 10,000 and 1 in 5,000 crypts, respectively (Nicholson *et al.*, 2018). This range of rates is sufficient for inference of stem cell behaviours, and empirically sets the bar for new marks. Thus, when screening sections for clones, 10 large area sections with > 10,000 crypts each are evaluated. The probability of not finding a clone in 100,000 crypts when they occur at a frequency of 1/10,000 is:

$$\left( \frac{\text{Number of non-clones in sample}}{\text{Number of crypts in sample}} \right)^{\text{Number of crypts screened}} \approx \text{Probability of no clone found}$$

Which, if the numbers used here are inserted gives:

$$\frac{9,999^{100,000}}{10,000} \approx 0.00005$$

This means that for every 100,000 antibodies trialled by screening 100,000 crypts each, only 5 marks occurring at a frequency of 1/10,000 would be missed. Or in other words, if no clones are found in 100,000 crypts, it is very likely that they occur at a frequency below 1/10,000, which, when compared to the known MAOA, would require an impractical amount of sections to be scored for data to be used quantitatively.

### 3.1.3 Aims

We have recently used the clonal marks mPAS and MAOA to define the homeostatic stem cell dynamics in the human colon (Nicholson *et al.*, 2018). To begin to define a repertoire of behaviours associated with gene-specific mutations, more clonal marks were required. The specific aims for this part of the study were thus:

- To identify novel X-linked clonal marks.
- To identify new marks related to frequent mutations in colorectal cancer.
- To develop a method for validation of IHC results by sequencing.

### 3.1.4 Published work

STAG2, the validation of which I performed and describe in this chapter, was included in our recent publication (Nicholson *et al.*, 2018), on which I am second author.

---

## 3.2 Identification of novel X-linked clonal marks

X-linked genes lend themselves as clonal marks, as X-inactivation in females and hemizyosity in males results in both genders presenting as effectively heterozygous for the loci in question. This means that in principle only one ‘hit’ is required for a change in protein levels or structure. Identification of novel X-linked clonal marks followed a sequential filtering process as outlined in Figure 3.2. Importantly, the screen focused on potential negative marks, relating to the left-hand side of Figure 3.1.

### 3.2.1 Selection of X-linked genes

The human X chromosome contains approximately 156 megabases of sequence and is estimated by the Consensus Coding Sequence Project to contain just over 800 protein-encoding genes (Pruitt *et al.*, 2009). This would make testing each and every protein encoded by this chromosome by IHC a very laborious and expensive process. Therefore, a pre-selection had to be performed. Assuming that somatic mutations hit the genome at random, it follows that gene length should correlate with amount of somatic mutations. Based on this hypothesis, genes could initially be filtered according to length. However, not all mutations will have the same impact on protein structure. It is plausible that exonic mutations are more impactful than intronic alterations. Therefore, a high exon-intron ratio may be favourable. Moreover, it is hard to predict the impact of single amino acid changes on antibody binding. It is conceivable that mutations affecting areas hidden within large proteins may not affect antibody binding as much as this could happen in smaller proteins (encoded by smaller genes). Therefore, small proteins should not be excluded from initial screening. As a consequence of these considerations, a variety of genes with different characteristics (long absolute length, high exon-intron ratio, long coding sequence) were selected for an initial screen. In theory, the results of this screen would define selection criteria for future studies. The preliminary list was cross-referenced with the Human Protein Atlas (HPA), where staining intensity has been assessed with IHC on a tissue-by-tissue basis. Only genes with detectable expression in colonic tissue were taken further. The final selection criterion was antibody availability. Importantly, for this project a collaboration was set up with Dr Cecilia Lindskog Bergström at the HPA.

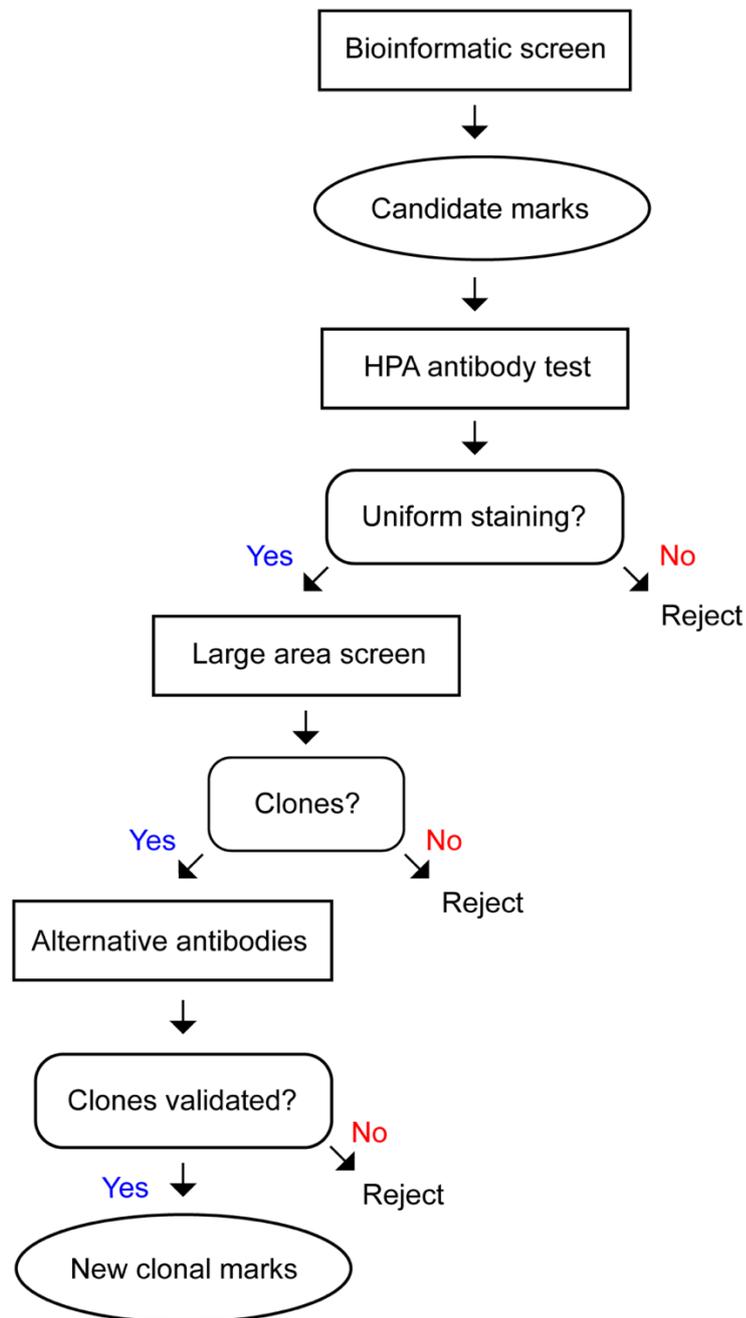


Figure 3.2 **Workflow/decision tree for identification of new X-linked clonal marks.** Rectangular encapsulation represents input of work, rounded boxes are evaluation points, ellipses are outputs.

From the filtered list of X-linked genes she was able to provide samples of development-stage antibodies against 21 X-linked proteins. Table 3.1 lists the selected genes and the characteristics discussed above. The balance between the different characteristics among the genes chosen is illustrated in figure 3.3, which shows the sum of ranking by total gene length, coding sequence length and exon-intron ratio. The figure includes *MAOA* as a comparison as well as *STAG2* which is discussed in 3.2.4.1. As there are 23 genes on the list, an average ranking (position 12) for all three categories would score 36. There is a fairly even spread of genes around this number, showing that for the genes chosen, high ranks in one category are often balanced by low scores in another. For example, the gene *HDAC8* is the third largest, however, this is mainly due to introns, as it has the lowest exon-intron ratio. Combined with the low ranks for the exons, a total score of 47 results. Conversely, *SLC25A5* is the smallest gene but with a very high exon-intron ratio. The combined ranks also add up to 47, despite very different gene characteristics.

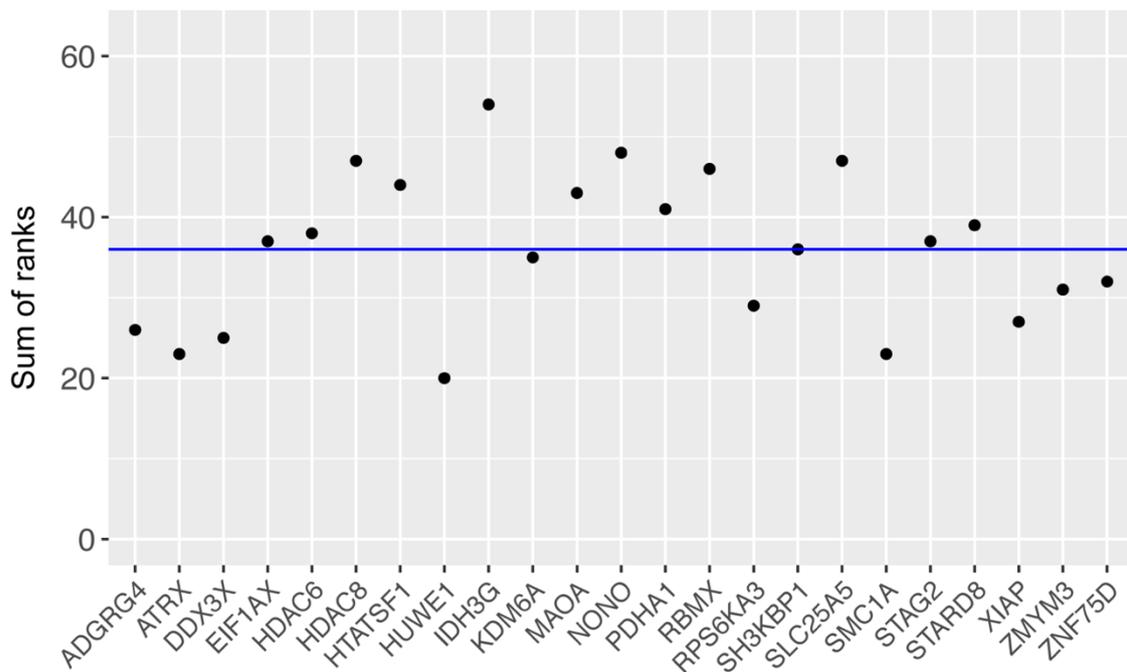


Figure 3.3 **Schematic representation of spread of gene characteristics.**

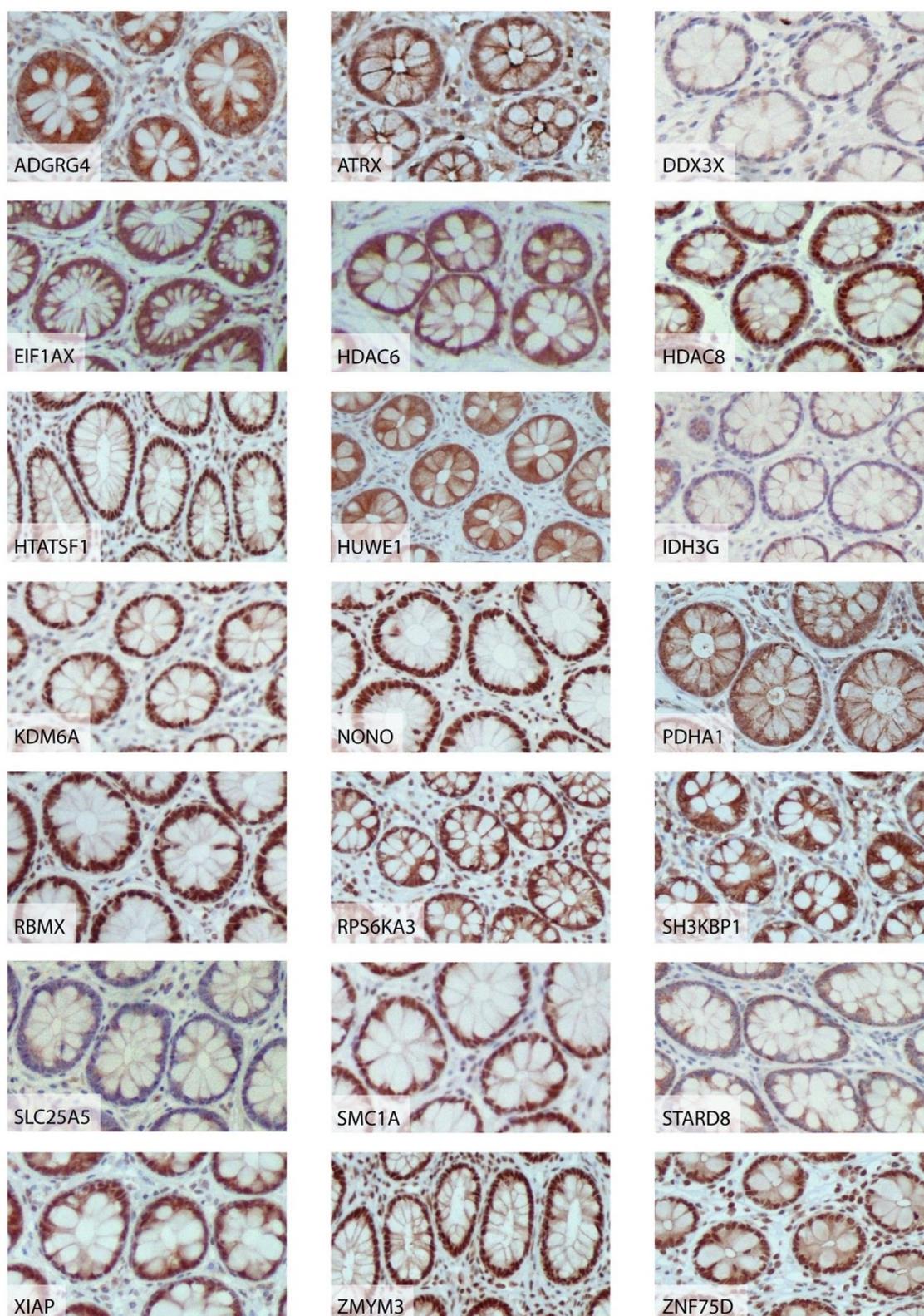
For each gene, a rank from 1-23 was assigned for total gene length, length of exons combined and exon-intron ratio. In each of the three categories, rank 1 corresponds to the largest value among the 23 genes and 23 to the smallest value in that particular category. The three ranks obtained for each of the 21 candidate marks, the previously used mark *MAOA* and the previously identified *STAG2* were then combined to give the final scores. These are plotted in alphabetical order. Blue line = 36, the value obtained for median ranking (12) in all three categories.

### 3.2.2 Antibody testing

For each of the 21 antibodies provided by the HPA, IHC was performed on *en face* human colonic FFPE sections using both ends of the recommended titre range. As this screen was aimed at detecting negative marks, uniformly strong staining was required for an antibody to pass this stage. In four cases, for DDX3X, IDH3G, SLC25A5 and STARD8, this was not achieved (Figure 3.4).

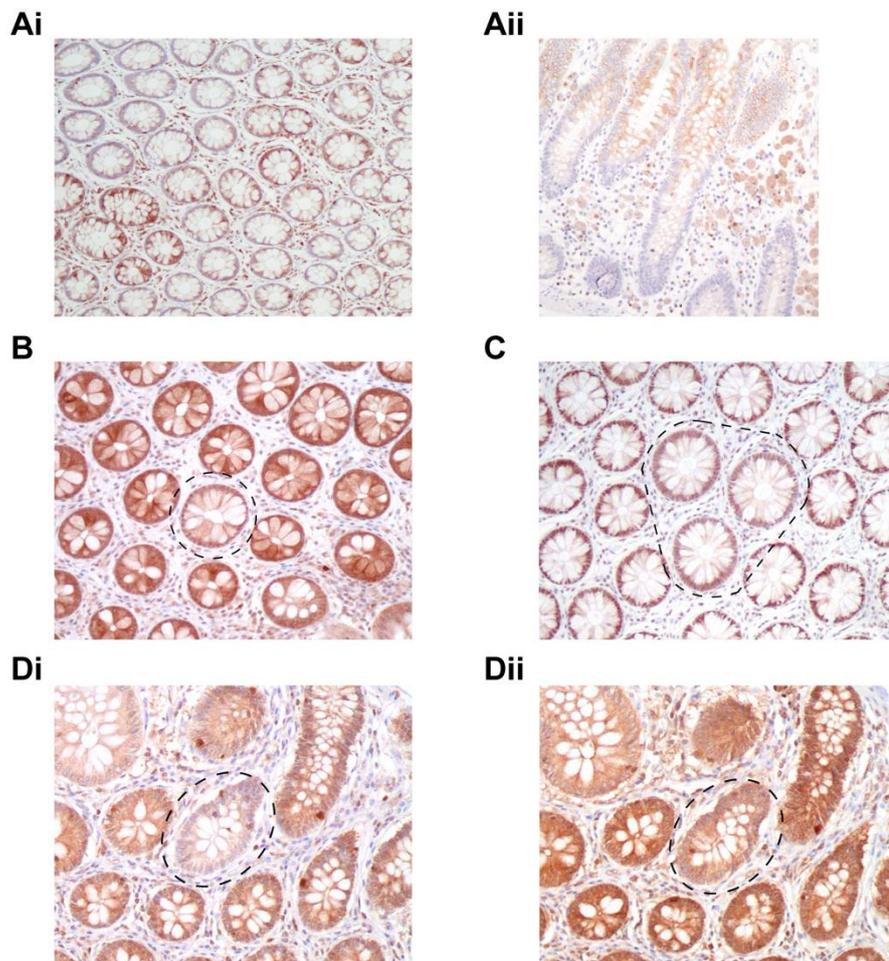
Table 3.1 Intron, exon and total gene length data for candidate marks.

Gene	Introns (bp)	Exons (bp)	Intron/exon ratio	Total length (bp)
ADGRG4	102216	9820	0.096	112036
ATRX	270172	11167	0.041	281339
DDX3X	11478	6907	0.602	18385
EIF1AX	12900	4414	0.342	17314
HDAC6	18792	4196	0.223	22988
HDAC8	241574	1754	0.007	243328
HTATSF1	12245	3019	0.247	15264
HUWE1	139893	14731	0.105	154624
IDH3G	7273	1335	0.184	8608
KDM6A	233654	5438	0.023	239092
MAOA	86588	4015	0.046	90603
NONO	14873	2661	0.179	17534
PDHA1	14329	3484	0.243	17813
RBMX	5253	2012	0.383	7265
RPS6KA3	109012	8111	0.074	117123
SH3KBP1	348896	4728	0.014	353624
SLC25A5	1646	1307	0.794	2953
SMC1A	38796	9784	0.252	48580
STAG2	135762	5218	0.038	140980
STARD8	73228	4841	0.066	78069
XIAP	45362	8591	0.189	53953
ZMYM3	8981	5536	0.616	14517
ZNF75D	52708	5611	0.106	58319



**Figure 3.4 Antibody screen for novel X-linked clonal marks.**  
*Representative images for staining of human FFPE colonic tissue sections.*

For SH3KBP1, IHC on *en face* tissue sections revealed a variegated staining pattern. Inspection of longitudinal sections revealed a gradient of protein expression along the crypt axis, with highest levels at the top of the crypt (Figure 3.5A). The protein was thus rejected as a clonal mark. However, it may find future applications when orientation along the crypt axis is required. For the remaining 16 antibodies a suitable titre was found, enabling progression to the next stage, which was screening for negatively staining crypts.



**Figure 3.5 Candidate marks abandoned at validation stage.**

(A) Representative images of SH3KBP1 staining showing (i) variegated pattern in transverse section and (ii) crypt-axis gradient in longitudinal section.

(B) IHC with antibody against HUWE1, crypt with weak immunoreactivity highlighted.

(C) IHC with antibody against SMC1A, three crypts with weak immunoreactivity highlighted.

(D) RPS6KA3-negative crypts confirmed in serial sections stained with antibodies from (i) HPA and (ii) Bethyl.

### 3.2.3 Pre-screening for clonal marks

For each of the 16 antibodies taken further, at least 10 large (>10,000 crypts) area sections were screened for the presence of confined clones with reduced immuno-reactivity. As discussed in 3.1.2.6. this number was deemed sufficient to identify candidate marks with a sufficiently high event rate. For 10 out of the 16 antibodies screened no weaker staining crypts were found. Several reasons may account for this result. Mutations may alter a protein structure in a way that does not affect the epitopes for the antibodies used here. Mutations may not be tolerated in human colonic epithelium or the antibody used may not harbour the specificity stated.

In six cases, for the antibodies against HDAC6, HUWE1, KDM6A, NONO, RPS6KA3 and SMC1A faintly or entirely negatively staining crypts were identified (Figures 3.5 & 3.6) These were indicative of stable reduction of protein expression in epithelial cells. These antibodies were thus taken further for optimisation and or validation.

### 3.2.4 Validation and optimisation

For each of the six potential marks one or more alternative antibodies were tested (Table 3.2). For HUWE1 and SMC1A none of these validated the fainter staining crypts in serial sections. For RPS6KA3 the weaker staining crypts were confirmed with an alternative antibody used on serial sections but with substantial residual immunoreactivity (Figure 3.5). Hence, these candidate marks were not pursued further.

For KDM6A and HDAC6 alternative antibodies by Cell Signalling Technology (CST) were tested. Both improved the staining quality and lead to clean negatives confirming the HPA antibody result in serial sections (Figure 3.6). For NONO staining with the HPA antibody resulted in very clean negative crypts. These could be confirmed by staining of serial sections with an alternative antibody also provided by the HPA (Figure 3.6).

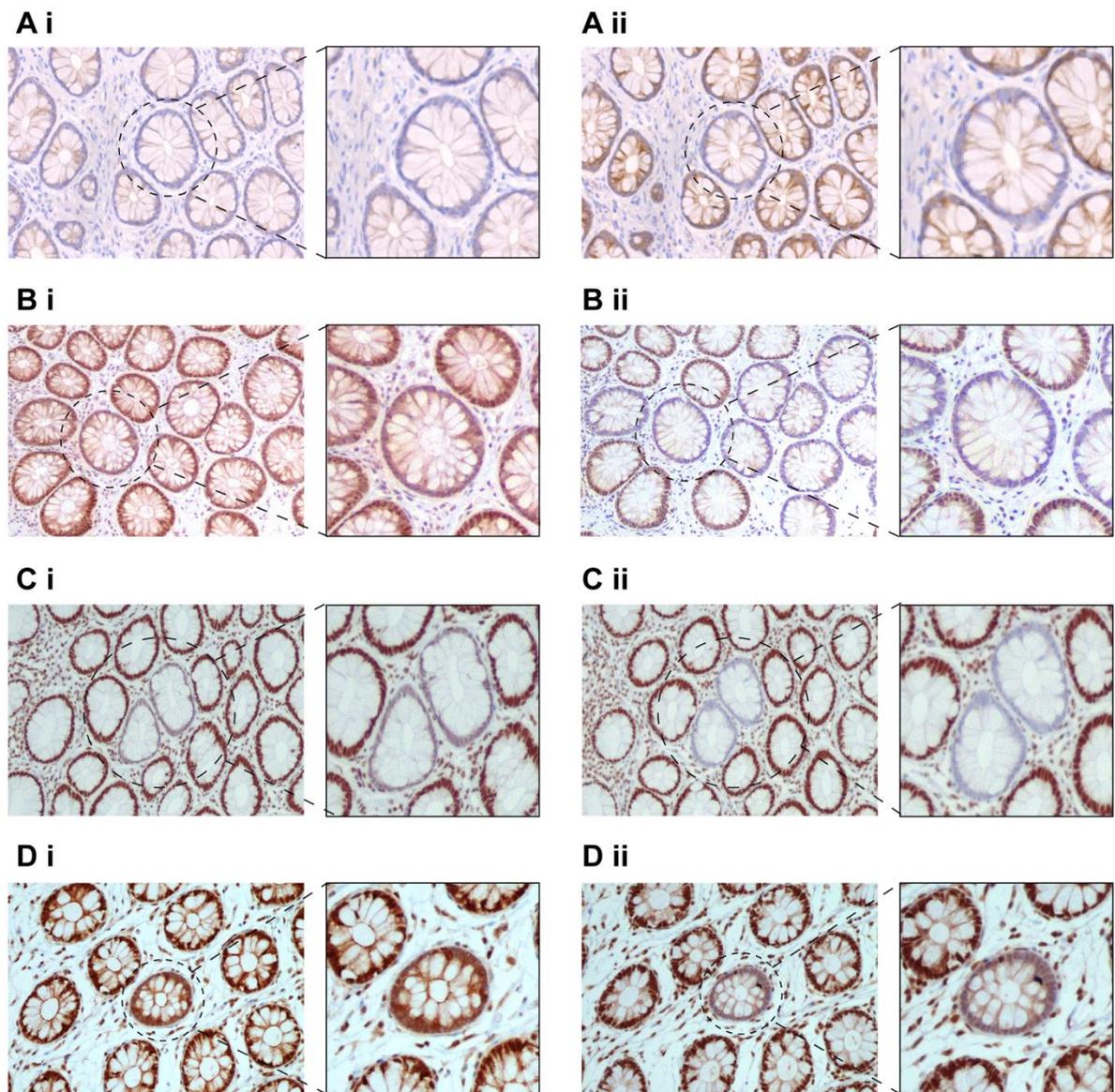
To summarise, all X-linked genes were filtered to generate a list of genes of interest. The HPA provided development-stage antibodies which were evaluated as potential clonal marks. Ultimately, screening of 21 antibodies yielded three new X-linked clonal marks: HDAC6, KDM6A and NONO.

Table 3.2 Antibodies tested for validation of candidate clonal marks.

<b>Antigen</b>	<b>Supplier</b>	<b>Antibody</b>	<b>Result</b>
HDAC6	CST	#7558	validated as clonal mark
HUWE1	Abcam	ab70161	not validated
HUWE1	Aviva	OALA01854	not validated
NONO	HPA	HPA054559	validated as clonal mark
KDM6A	CST	#33510	validated as clonal mark
RPS6KA3	Bethyl	A302-460A	weakly validated
SMC1A	Abcam	ab133643	not validated
SMC1A	Biorbyt	orb94671	not validated
STAG2	LSB	LS-B11284	validated as clonal mark

### 3.2.4.1 Validation of the clonal mark STAG2

Before the screen described above was carried out, Dr Anna Nicholson in the lab performed a smaller scale screen of HPA antibodies. This identified the candidate X-linked mark STAG2, for which weakly negative crypts had been found. Staining of serial sections with an alternative antibody by LifeSpan BioSciences Inc. (LSB, Table 3.2) resulted in cleaner negative crypts, validating STAG2 as a clonal mark (Figure 3.6).



**Figure 3.6 Validation of new X-linked clonal marks.**

*IHC on serial sections with different antibodies. Negative crypts highlighted and enlarged.*

*(A) HDAC6 (i) HPA antibody (ii) CST antibody*

*(B) KDM6A (i) HPA antibody (ii) CST antibody*

*(C) NONO (i) HPA antibody (ii) CST antibody*

*(D) STAG2 (i) HPA antibody (ii) LSB antibody*

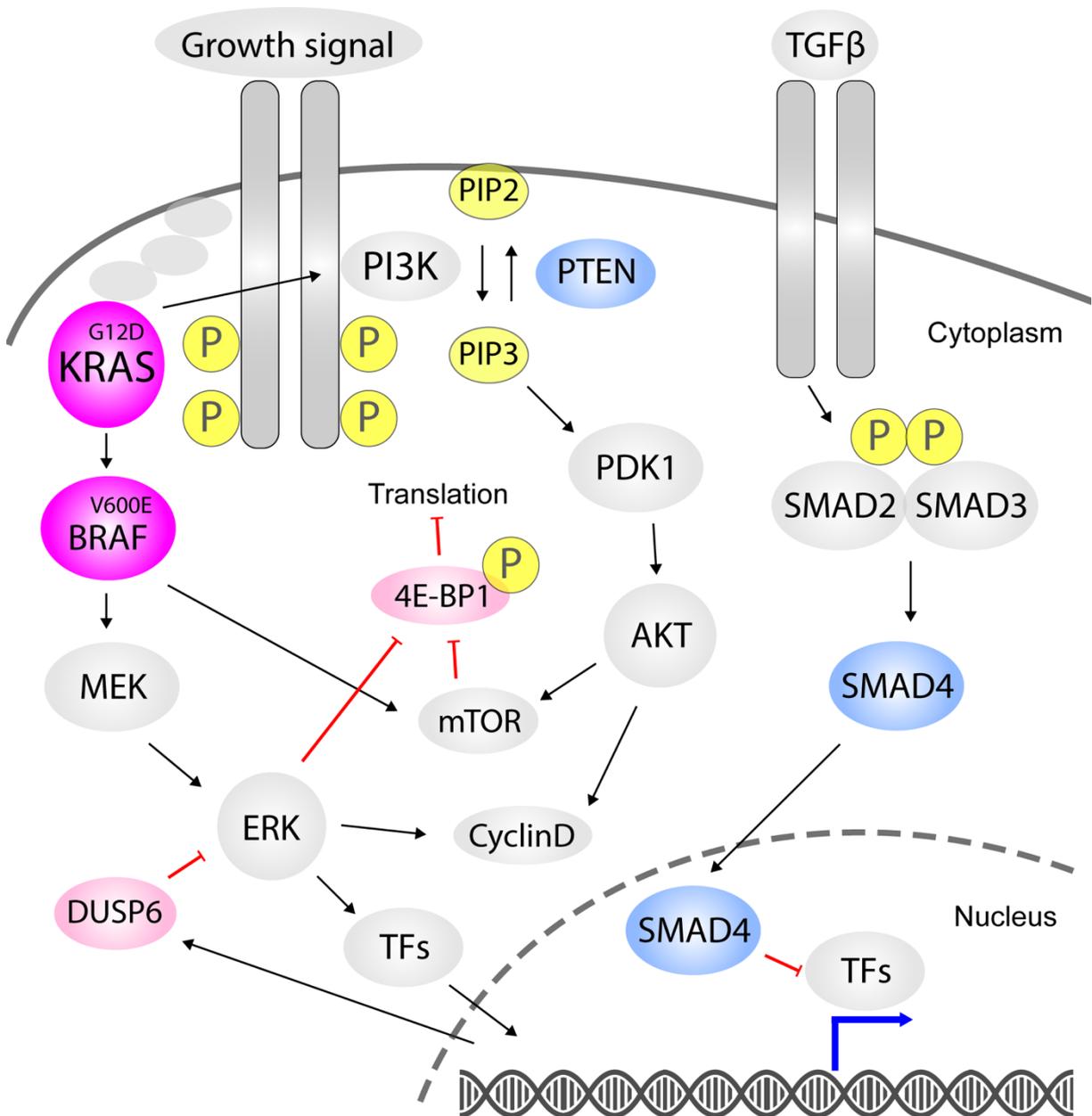
### 3.3 Identification of cancer-associated clonal marks

X-linked marks represent an invaluable tool to study stem cell dynamics in the crypt. As such they enable detailed description of homeostatic stem cell turnover as well as investigation of behaviours associated with different gene-specific mutations. Extension of the method to genes commonly associated with colorectal cancer may enable deeper understanding of the specific mechanisms promoting the earliest stages of the disease. Unfortunately, none of the most common driver genes of colorectal cancer are located on the X chromosome. However, this may not preclude detection of certain specific alterations. This screen was conceptually different from the process that identified new X-linked clonal marks. The latter was aimed at detection of any heritable genetic event that reduced protein abundance in affected cells. Here, the genetic alterations to be identified are well defined, but their detection may not be trivial.

Driver mutations in CRC were discussed in section 1.5.1.3. Table 3.3 summarises the antibodies that were screened for cancer-associated clonal marks. An overview of their relationships in oncogenic signalling is presented in Figure 3.7. Importantly, the two commonly used categories of genes mutated in cancer, oncogenes and tumour suppressors, harbour different challenges for IHC-based detection. For this reason they are discussed separately.

Table 3.3 **Antibodies screened for colorectal cancer-associated clonal marks.**  
*TSG = tumour suppressor gene*

Antigen	Type	Antibody	Supplier	Result
BRAFV600E	oncogene	E19290	Spring Bioscience	no clones detected
DUSP6	surrogate	ab76310	Abcam	no clones detected
KRASG12D	oncogene	#14429	CST	no clones detected
p-4E-BP1	surrogate	#2855	CST	no clones detected
p53	TSG	ab1101	Abcam	positive clones
PTEN	TSG	#9559	CST	negative clones
PTEN	TSG	04-035	Sigma-Aldrich	validation
SMAD4	TSG	ab217267	Abcam	no clones detected



**Figure 3.7 Overview of relationships between selected cancer-associated proteins.**

*KRAS* activated by mutation engages the mitogen-activated protein kinase pathway via *BRAF*-*MEK*-*ERK*. A negative feedback loop around *ERK* exists via *DUSP6*. The pathway can also be activated by mutation of *BRAF*. Activated *KRAS* can further signal through the *PI3K*-*PDK1*-*AKT* pathway. *PTEN* acts as a tumour suppressor, opposing the activity of *PI3K*. *4E-BP1* is an inhibitor of translation and is inhibited by phosphorylation. *SMAD4* is an effector of *TGFβ* signalling, inhibiting gene expression. Proteins screened for are highlighted in colour. Magenta = activated oncogene, pink = surrogate marker, blue = tumour suppressor, TF = transcription factor, yellow = signal relay molecule.

### 3.3.1 Detection of oncogene activation

Oncogenes are usually activated by single nucleotide variants (SNVs) leading to very specific amino acid changes. According to TCGA sequencing data from 2,644 human samples, the four most common specific SNVs in human colon adenomas and adenocarcinomas occur in *KRAS* and *BRAF* (*KRASG12D*, *KRASG12V*, *KRASG13D* and *BRAFV600E*). Antibodies advertised as detecting *KRASG12D* and *BRAFV600E* are commercially available. However, detection of such specific amino acid changes by IHC remains challenging. In addition, it may be possible to detect oncogenic activation by a surrogate marker, meaning a downstream protein the levels of which should increase or decrease as a consequence of the oncogenic signalling. For example, the levels of DUSP6 and phosphorylated 4E-BP1 (p-4E-BP1) should rise upon oncogenic activation of BRAF or KRAS (Figure 3.7) (Rojo *et al.*, 2007; Ekerot *et al.*, 2008; Buffet *et al.*, 2017; Silva *et al.*, 2017; Ahmad *et al.*, 2018; Unni *et al.*, 2018). Of note, activated oncogenes as well as their surrogates would all present as positive clonal marks, the challenges of which are discussed in the introduction and illustrated on the right side of Figure 3.1.

Four antibodies (two for direct detection, two for surrogates) were initially tested in human FFPE colonic tumour sections to optimise titres (Figure 3.8A). This was followed by IHC on normal human FFPE colon sections (Figure 3.8). The antibodies against *KRASG12D* and *BRAFV600E* yielded patchy staining. This result is suggestive of at least partially non-specific immunoreactivity. For DUSP6 and p-4E-BP1 relatively strong staining was obtained, even at increased titres (Figure 3.8B). Screening of large area normal tissue sections did not identify any confined significantly stronger positive clones, which would have validated the use of the antibodies despite background immunoreactivity. Hence, the four antibodies were not pursued further.

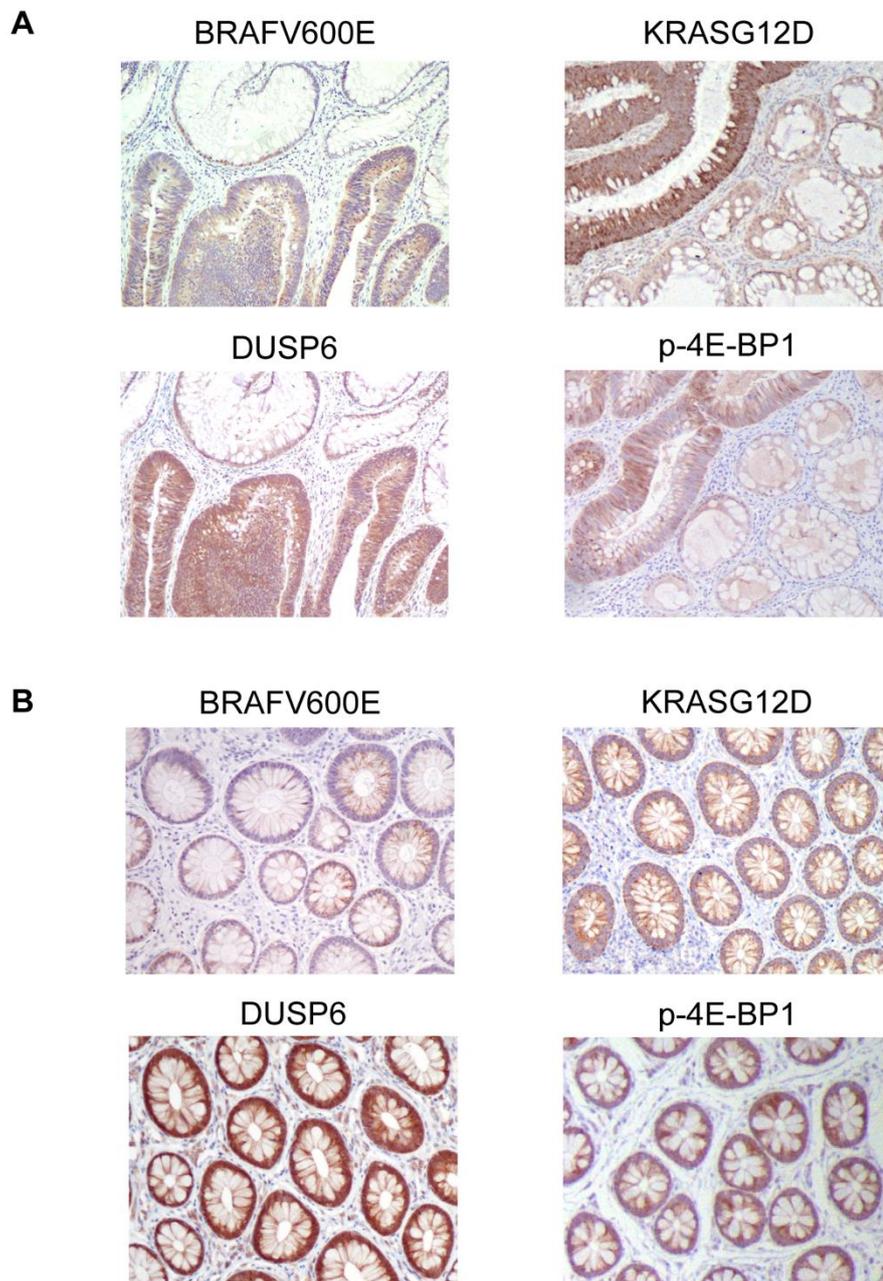


Figure 3.8 **Antibody screen for oncogene-associated clonal marks.**

(A) IHC on human colonic tumour sections using antibodies against the activated oncogenes *KRASG12D* and *BRAFV600E* as well as the surrogates *DUSP6* and *p-4E-BP1*.

(B) IHC on normal human colonic tissue sections with the same antibodies as in (A).

### 3.3.2 Detection of tumour suppressor loss or gain-of-function

Tumour suppressors are usually mutated in a way that abrogates their function. Such loss-of-function mutations should in principle be detectable in the same way as negative X-linked clonal marks. However, as the genes in question are autosomal, both copies may need to be affected for a change to be visualisable by IHC. This implies that the clone frequency may be significantly lower than that observed for X-linked marks. Alternatively, single copy loss may present as weaker but not completely negative clones, which would preclude use as a clonal mark.

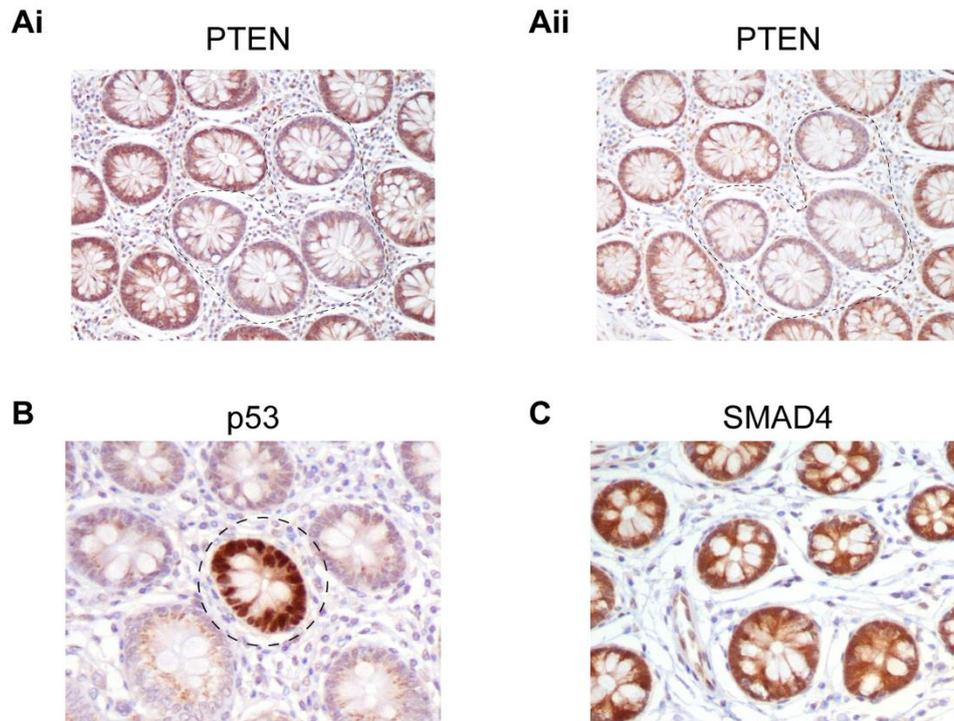
A special case for detection of a tumour suppressor is p53. The levels of this protein are low in normal cells, due to turnover via ubiquitination and proteasomal degradation (Lahav *et al.*, 2004). Two types of p53 alterations occur in cancer: Mutations that abrogate its function as well as SNVs that confer oncogenic functions by means of a dominant gain-of-function (GOF) mechanism. GOF mutant p53 can stabilise and accumulate in the nucleus where it may activate transcription factors to promote tumourigenesis (Brosh & Rotter, 2009; Nakayama & Oshima, 2018). As such, p53 may be detected as a positive mark.

For the three tumour suppressors p53, PTEN and SMAD4, mutations occur in 51.7%, 8.2% and 11.6% of colorectal cancers, respectively (Giannakis *et al.*, 2016, via cBioPortal).

IHC with antibodies against these three proteins (Table 3.3) was performed on normal human FFPE colon sections. Uniformly strong staining was achieved for PTEN and SMAD4. As expected staining was absent for p53, due to low levels of this protein reported in normal cells (Figure 3.9). Screening of at least 10 large area sections (>10,000 crypts each) revealed individual negative crypts for PTEN, indicative of loss-of-function mutations. These were validated by staining of serial sections with an alternative antibody (Figure 3.9A). For p53, rare individual positive crypts were identified, suggestive of stabilisation of the protein (Figure 3.9B). For SMAD4, no negative crypts were identified (Figure 3.9C).

In summary, seven antibodies to visualise common CRC mutations in normal colonic epithelium were tested. Four related to detection of oncogene activation did not yield

satisfactory outcomes. Out of three antibodies aimed at detecting alterations in tumour suppressors, two yielded new clonal marks: p53 and PTEN.



**Figure 3.9 Antibody screen for tumour suppressor-associated clonal marks.**  
 (A) IHC for PTEN on serial sections with antibodies from (i) Sigma-Aldrich (ii) CST. Negative crypt highlighted.  
 (B) IHC with antibody against p53. Crypt with stabilised p53 highlighted.  
 (C) IHC with antibody against SMAD4.

### 3.4 A method for next-generation sequencing of laser captured FFPE crypt samples

IHC-based detection of mutant clones builds on the assumption that the observed staining pattern is caused by a genetic event which changes the abundance or structure of a protein. Validation using alternative antibodies in serial tissue sections provides compelling support for this hypothesis. However, sequencing of DNA from specific crypts harbours the potential to identify the causal mutations in the gene, therefore unequivocally demonstrating the genetic

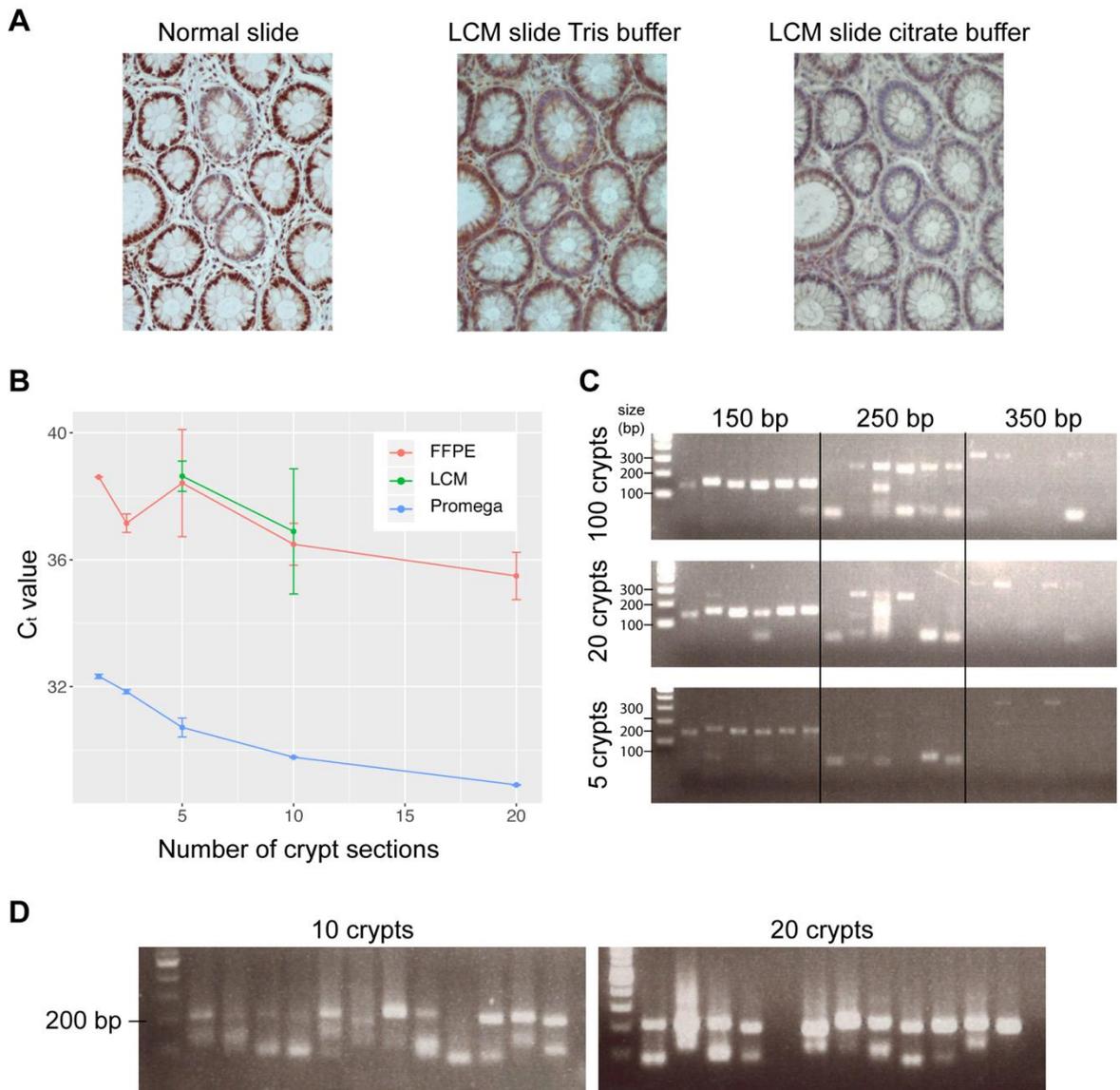
---

origin of a clone. Laser capture microdissection (LCM) is a technique enabling precise isolation of specific cells in a tissue section (Emmert-Buck *et al.*, 1996). High power IR or UV lasers activate a polymer film that expands and surrounds the area of interest while preserving cell morphology, DNA, RNA and protein composition (Espina *et al.*, 2006). Specific cells can thus be isolated from a tissue sample, DNA extracted and subjected to sequencing. A protocol for LCM and sequencing of human FFPE colonic crypt sections was thus devised.

### 3.4.1 Optimisation of antigen retrieval for LCM

The standard protocol used for IHC in the Winton laboratory requires heat-induced antigen retrieval (AR) with citrate buffer at high temperature (126°C) and pressure. However, current laser microdissection systems require special polyethylene naphthalate (PEN) membrane slides, which are not resistant to high temperatures. Hence, three alternative types of AR for IHC of FFPE tissue sections on membrane slides were tested: Pepsin, Tris buffer and citrate buffer. All three have been published for LCM (Eberle *et al.*, 2010).

The first AR method tested was Pepsin at 37 °C for 5-20 minutes. Short incubation times resulted in absence of staining, whereas longer treatment seemed to cause unspecific staining, possibly due to tissue digestion. Subsequently, AR with Tris buffer (pH9) at 60–75 °C for 16 h was attempted. This yielded satisfactory staining (Figure 3.10A middle). However, this method resulted in frequent detachment of the tissue sections from the slides, which may be attributable to the relatively harsh pH conditions. This was prevented by use of the gentler citrate buffer (pH6). Optimal results were obtained with AR with citrate buffer at 75 °C for 16 h (Figure 3.10A right). The method was successfully applied to MAOA, STAG2 and KDM6A. Occasionally, bleb formation underneath the membrane was observed. However, this did not affect the quality of the IHC or subsequent LCM.



**Figure 3.10 Optimisation of IHC and PCR on FFPE tissue sections and extracted DNA.**

(A) Representative images of optimised IHC for *STAG2* on serial sections on membrane slides compared to the standard protocol with a normal slide.

(B) Results from qPCR for *ATOH1* using diluted commercially available DNA (Promega), diluted DNA bulk extracted from FFPE tissue sections (FFPE) and laser capture micro-dissected crypts (LCM).

(C) Products generated by PCR amplification of parts of *MAOA*, *STAG2*, *CASD1* and *LOC107985193* using FFPE DNA diluted to equivalents of 5, 20 and 100 crypts.

(D) Products from individual PCR reactions with one *MAOA* primer pair each, each using the product of a multiplex PCR (with 12 *MAOA* primer pairs) on DNA extracted from 10 or 20 laser captured FFPE crypts as a template. Note that one tube evaporated in the panel on the right.

### 3.4.2 Optimisation of protocol for PCR on FFPE DNA from crypt sections

With a reliable method for IHC on membrane slides in place, the next step was to develop a protocol for obtaining DNA for sequencing from clonal mark-negative crypts. Two challenges were identified: Firstly, the very low amounts of input of DNA available. Secondly, fragmentation of DNA due to formalin fixation. It has been demonstrated that amplicons greater than 500 bp are only very inefficiently amplified from such material (Dietrich *et al.*, 2013).

In a first step, the quality of DNA extracted from the available FFPE tissue was assessed. Varying numbers of transverse crypt sections of 10 µm thickness were laser capture microdissected from stained human FFPE sections and DNA extracted. One crypt section contains an average of 43 cells. One human genome equates to approximately 6 pg of DNA (Gillooly, Hein & Damiani, 2015). Hence one crypt section can be estimated to contain a maximum of 0.258 ng of DNA. For qPCR work-up DNA samples originating from bulk extractions from large area sections with known crypt numbers as well as commercially available human DNA were used. Both were diluted down to a concentration estimated to be similar to the laser capture microdissected crypt sections (using the estimation from above). Analysis by qPCR revealed that bulk or small sample extracted DNA from FFPE tissue is equally amplifiable. However, even though used at the same concentration as determined by spectrophotometry, both these samples exhibited much higher Ct values than the purchased DNA at equivalent concentrations (Figure 3.10B). This suggests that when working with FFPE DNA, the concentration does not inform the amplifiability (or bioavailability) of the DNA in PCR based assays. To gain further insight into the quality of the material, PCR on DNA bulk extracted from FFPE sections, diluted to concentrations ranging from that expected from 100 down to 5 crypt sections per reaction, was performed. Primers were designed for a set of amplicons ranging from 150–350 bp across 6 different genomic loci including the exons of *MAOA*, *STAG2* and *CASD1*. Primers available to generate amplicons of desired length from a locus encoding a non-coding RNA (*LOC107985193*) were also used. This experiment demonstrated that for the template material used here, exceeding 150 bp in amplicon size leads to a marked reduction in PCR product yield (Figure 3.10C).

Next, a test was performed on *MAOA*, as this gene has a relatively short combined cDNA sequence of 1484 bp. A set of 24 primer pairs to tile the exons of *MAOA* in amplicons of approximately 150 bp was designed. A library for sequencing was generated in a two-step process: An initial round of two multiplexed PCR reactions (due to overlap of the amplicons) followed by 24 individual reactions using the first product as a template. Of note, PCR products reach a size of approximately 200 bp due to addition of adapters for next-generation sequencing to the primers. Amplicons were reliably obtained from DNA extracted from 10 FFPE crypt sections (Figure 3.10D).

In summary, a method to generate a library for sequencing of DNA extracted from laser capture microdissected FFPE crypt samples was developed. This included optimisation of IHC on membrane slides and design of a protocol for PCR amplification of 24 small (approx. 150 bp) amplicons from DNA extracted from a sample of 10 laser capture microdissected transverse crypt sections.

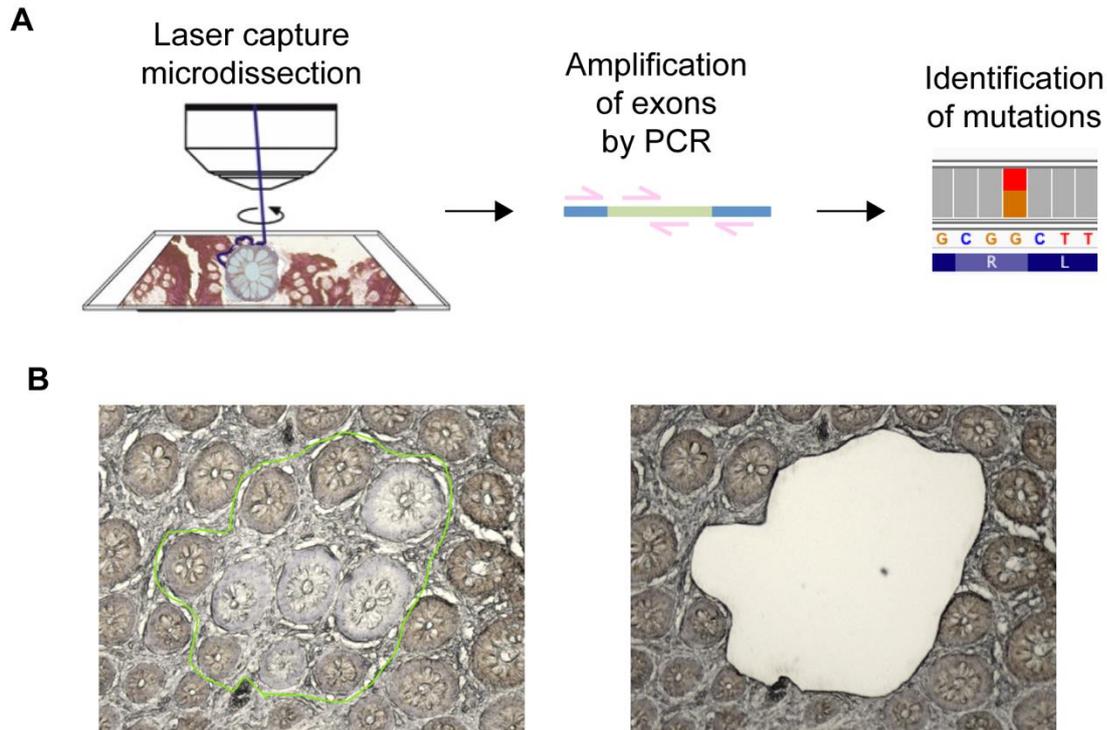
### **3.5 Validation of MAOA and KDM6A by sequencing**

With a suitable protocol in place, LCM followed by next generation sequencing was performed on MAOA- and KDM6A-negative crypts (Figure 3.11). This was aimed at unequivocally demonstrating that clones originate from somatic mutations and are not epigenetic events. Analysis focused on patches of epithelium consisting of several adjacent mutant crypts. Such patches were hypothesised to arise from expansions of single clones via crypt division, which is discussed in detail in chapter 5.

#### **3.5.1 Identification of mutations in MAOA<sup>-</sup> patches**

Serial colonic FFPE tissue sections with previously identified MAOA-negative patches of crypts from 11 tissue blocks were stained for MAOA and mutant patches as well as wild type control crypts laser capture microdissected. Of note, preliminary experiments defined a minimum requirement of 10 crypts for reliable DNA extraction and PCR amplification (Section 3.4). However, MAOA-negative patches mostly consisted of fewer than 10 crypts. Therefore, patches smaller than 10 were complemented with wild-type crypts (Figure 3.11). This would

be reflected in the mutant allele frequency in that the proportion of mutant reads in each sample would be proportional to the fraction of mutant crypts that were captured. DNA was extracted and amplified in 24 amplicons across the coding sequence of *MAOA* by two consecutive PCR reactions and sequenced.



**Figure 3.11 Laser capture microdissection and sequencing of multicrypt patches.**

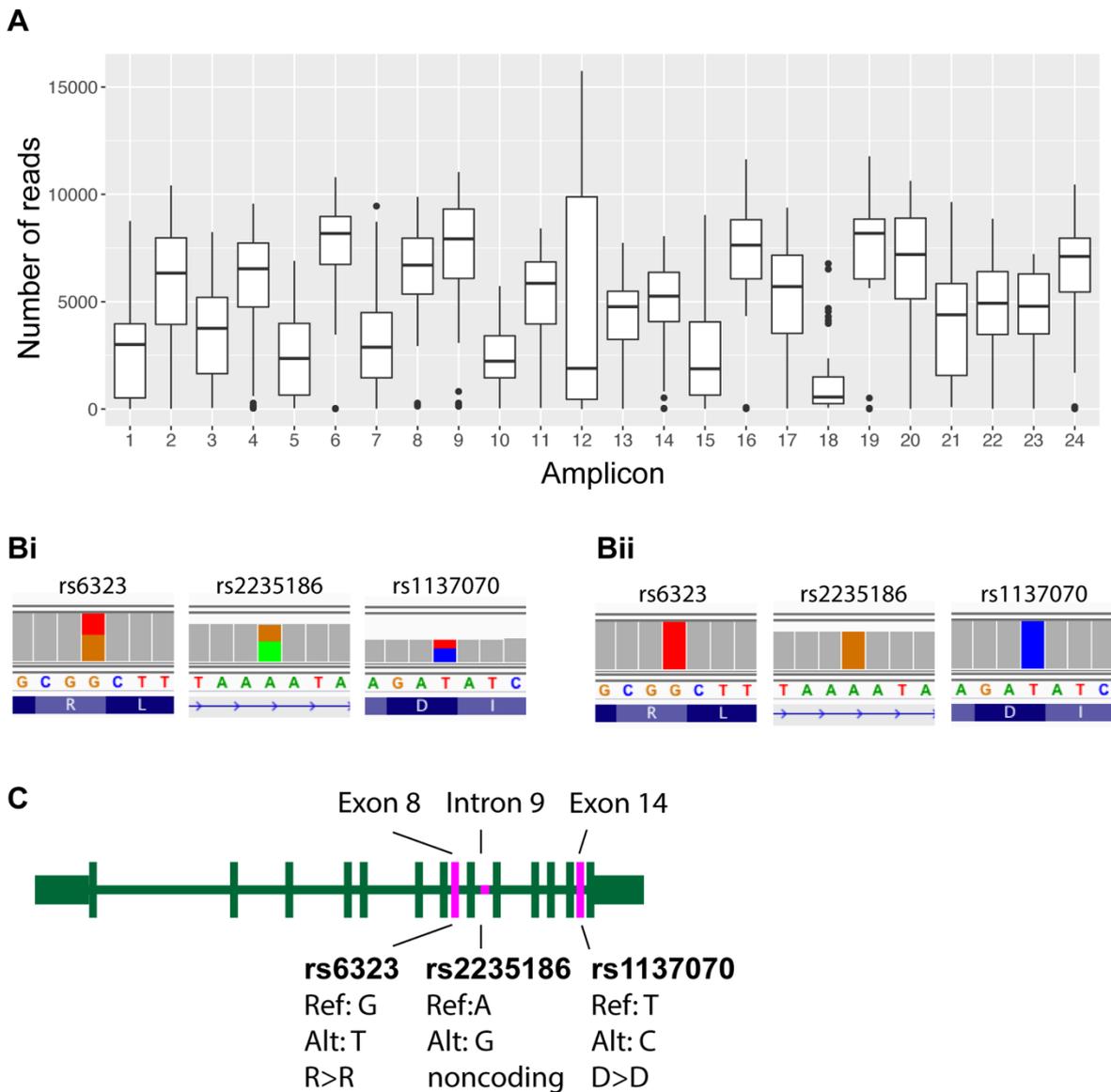
(A) Illustration of general workflow.

(B) Example of *MAOA*<sup>-</sup> patch before and after LCM.

The resulting reads varied in quantity and quality between different samples as well as between the 24 amplicons within each sample. In the majority of cases, >3500 reads per amplicon were obtained (Figure 3.12A). Using the Integrative Genomics Viewer (IGV, Robinson et al. 2011) three common germline single nucleotide polymorphisms (SNPs), rs6323, rs2186 235 and rs1137070 in *MAOA* were identified in the samples (Figure 3.12B & C). None of the identified SNPs result in amino acid changes. Of note, due to the X-linked nature of *MAOA*, male patients always appear as homozygous for either the reference sequence or the SNP.

---

To find mutations in *MAOA*, analysis with IGV was complemented by a PERL script written by Richard Kemp in our laboratory. Using the reference human genome, it scans each read with pairs of four nucleotides and records the base in between (Figure 3.13A). Importantly, if there is a mismatch between one of the quadruplets of nucleotides and the read, the script will not record anything. (Figure 3.13A, second to the left). Thus, for a single nucleotide variant (SNV), the script will record 4 positions with no reads, followed by a position with reads (Figure 3.13A SNV) and then again 4 positions without. For a 1 base deletion, there will be a continuous mismatch for 9 positions (Figure 3.13A rightmost). Following this analysis SNVs and deletions are easily identified by plotting the total reads against nucleotide position. Dips with a peak represent a SNV, dips lacking a peak appear due to a deletion (Figure 3.13B & C). Of note, the script also identifies SNPs, but these were more rapidly detected using IGV. When analysing the data, it was key to bear in mind that *MAOA* is an X-linked gene. Therefore, in samples from a female patient, even if they consisted of only MAOA-negative crypts, the expected allele frequency would be 50%, as the second copy of the X-chromosome would lack the mutation. Further, the expected allele frequency would drop for patches smaller than 10 due to added wild-type crypts. Finally, wild-type stromal cells between mutant crypts would add wild-type reads. Therefore, for mutations, even male samples are not expected to yield 100% mutant reads, as the stroma surrounding MAOA-negative crypts is wild-type. This was not the case for SNPs, which are germline variants and therefore present in both epithelial and stromal cells. Using the methods described above mutations were found in 5 out of 11 laser capture microdissected MAOA-negative patches. 4 different candidate mutations were identified, located in exons 8 and 13 of the gene. These appeared in the same MAOA-negative patch in serial sections but were absent in corresponding wild-type crypt samples (Figure 3.13D). The data is summarised in Table 3.4. Notably, the exact same deletion in exon 13 was found in two independent (different individuals) samples.

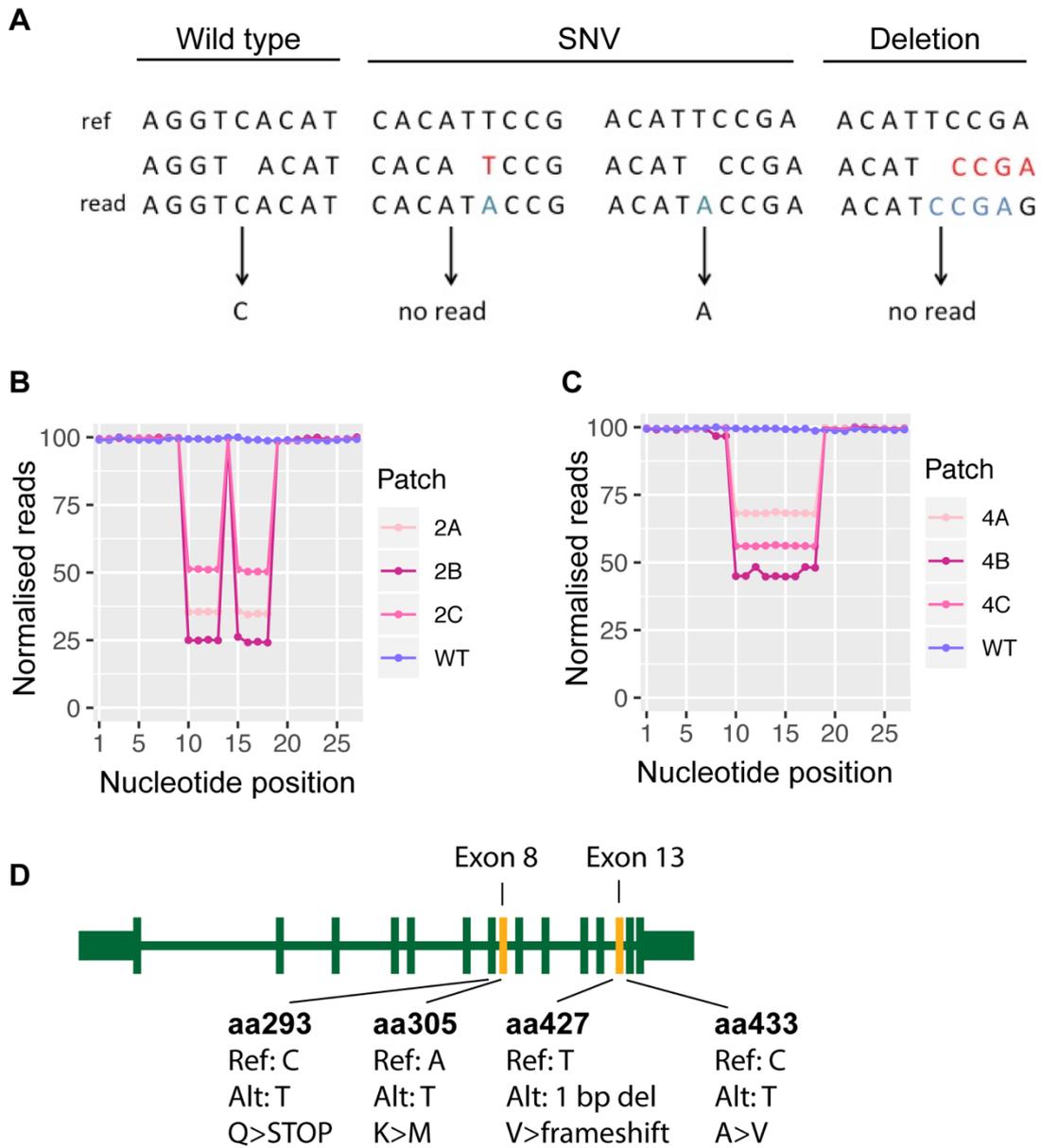


**Figure 3.12 Identification of germline SNPs in MAOA.**

(A) Graph showing read numbers for each of the 24 amplicons covering the exons of MAOA. Error bars = standard deviation.

(B) IGV snapshots of (i) heterozygous SNPs and (ii) homozygous SNPs identified.

(C) Graphical representation of MAOA gene structure with location of identified SNPs highlighted.



**Figure 3.13 Identification of mutations in MAOA<sup>-</sup> patches.**

(A) Illustration of method used to identify SNVs and deletions. Pairs of four nucleotides from the reference genome are compared to reads. If there is a match, the middle nucleotide is recorded. For more detailed schematic see Figure 2.1

(B) Graph showing data for SNV identified in samples 2A-C and corresponding WT sample.

(C) Graph showing data for deletion identified in samples 4A-C and corresponding WT sample.

(D) Graphical representation of MAOA gene structure with location of identified mutations.

Table 3.4 **Summary of mutations found in MAOA<sup>-</sup> multicrypt patches.**

*Ref* = nucleotide in reference genome, *freq* = frequency, *Alt* = alternative nucleotide -> this is the mutation, *Change* = predicted consequence of the mutation. Frequencies derived from frequency of mutant reads in samples.

Gender	Patch	Sample	Exon	Ref	Alt	Alt freq	Expected alt freq	Change
female	1 + 8 WT	1A	8	C	T	5.8	5.6	Q>STOP
female	1 + 8 WT	1B	8	C	T	2.8	5.6	Q>STOP
female	1 + 8 WT	1C	8	C	T	4.8	5.6	Q>STOP
female	1 + 8 WT	1D	8	C	T	3.4	5.6	Q>STOP
male	7	2A	8	A	T	64.2	0	K>M
male	7	2B	8	A	T	74.5	0	K>M
male	7	2C	8	A	T	48.4	0	K>M
female	3 + 5 WT	3A	13	C	T	8.3	18.75	A>V
female	3 + 7 WT	3B	13	C	T	14.3	15	A>V
male	5 + 5 WT	4A	13	T	del	31.4	50	frameshift
male	5 + 5 WT	4B	13	T	del	53.6	50	frameshift
male	4 + 5 WT	4C	13	T	del	43.6	44.4	frameshift
female	5 + 5 WT	5A	13	T	del	25	6.3	frameshift
female	5 + 5 WT	5B	13	T	del	25	11.6	frameshift

### 3.5.2 Identification of mutations in KDM6A<sup>-</sup> patches

The workflow described for MAOA was largely followed for KDM6A, with two alterations: Firstly, much larger patches of KDM6A-negative crypts could be found, eliminating the necessity of adding wild-type crypts to samples. Secondly, the coding sequence of *KDM6A* is a lot longer than that of *MAOA* (cDNA lengths of 4206 bp and 1484 bp, respectively). This meant that all exons of *KDM6A* could not be covered with 24 amplicons of 150 bp size. Amplicons were thus designed to cover loci where mutations had been identified in cancers, as listed in the COSMIC database as well as TCGA. An amplicon covering a mutation communicated by H. Lee-Six from the Sanger Institute was also added (Figure 3.14A).

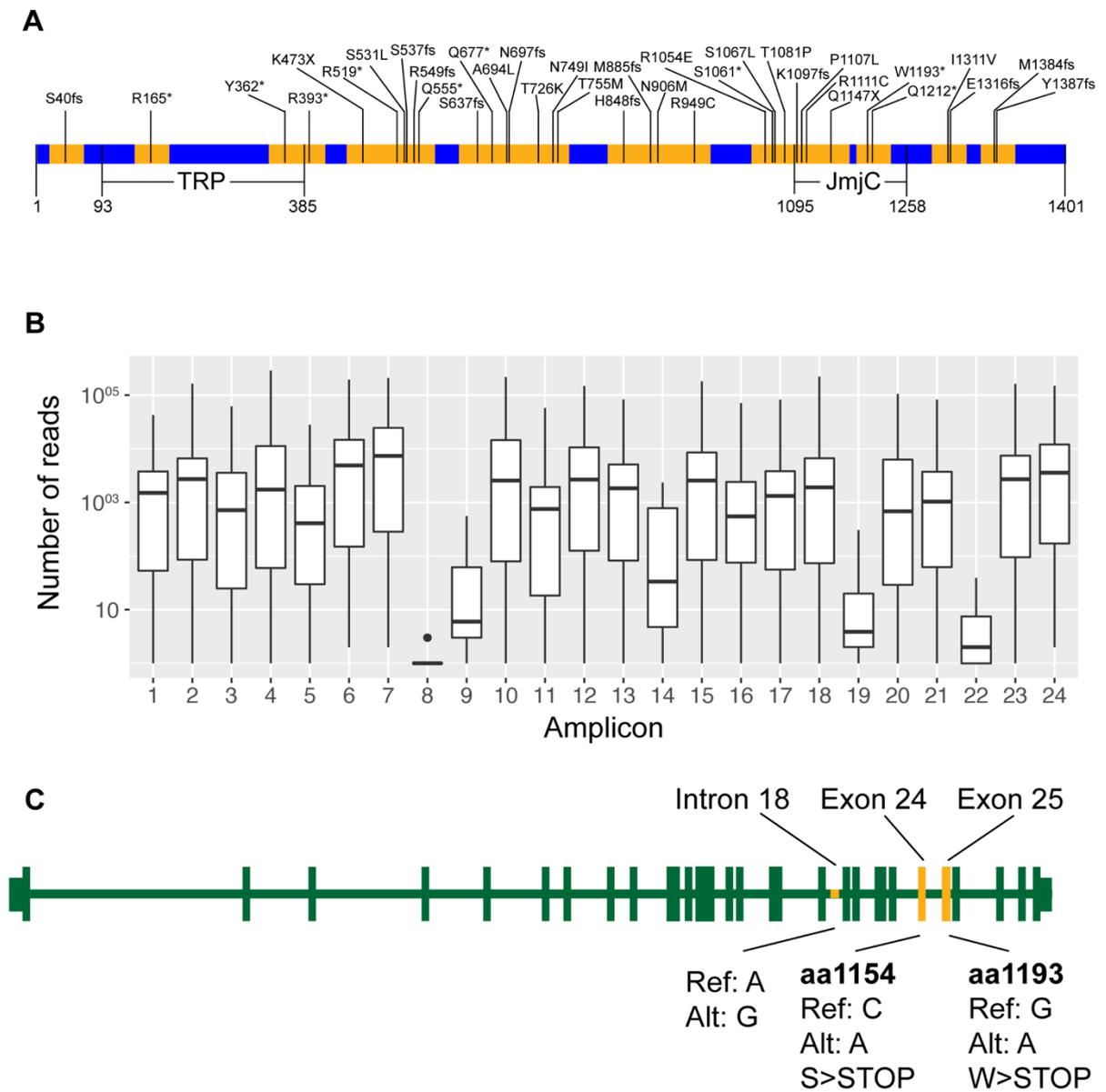


Figure 3.14 **Identification of mutations in KDM6A<sup>-</sup> patches.**

(A) Overview of KDM6A cDNA with annotation of selected mutations from databases. Orange = area covered by amplicons. The domains important for KDM6A function, the tetratricoprotein repeat (TPR) for protein-protein interaction and the JmjC domain for demethylase activity (JmjC) are indicated.

(B) Graph showing read numbers for each of the 24 amplicons covering the exons of KDM6A. Error bars = standard deviation. Logarithmic scale on y-axis.

(C) Graphical representation of KDM6A gene structure with location of identified mutations.

11 KDM6A-negative patches as well as wild-type control samples from the same blocks were laser capture microdissected, DNA extracted, amplified and sequenced. The resulting reads varied in quality across amplicons and samples: >1000 reads were obtained for 40% of samples. Two amplicons did not yield any reads (Figure 3.14B). Two further amplicons yielded very few reads (Figure 3.14B).

To find mutations in *KDM6A*, an adapted version of the script written for *MAOA* by Richard Kemp was applied. This identified mutations in 4 out of the 11 analysed patches. In three cases these were present the same KDM6A-negative patch in serial sections but absent in corresponding wild type samples. The fourth mutation was found at nearly 50% mutant allele frequency in one patch sample but was absent in the serial section as well as corresponding wild type samples. As the staining quality on this block was poor, it is possible that wild type, instead of KDM6A-negative, crypts were laser captured in one case. The mutation calls for the experiment are summarised in Figure 3.14C and Table 3.5. Of note, the mutation found in intron 18 was present in two independent patches. Furthermore, the mutations S1154\* and W1193\* have previously been found in cancer samples (COSMIC database).

**Table 3.5 Summary of mutations identified in KDM6A<sup>-</sup> patches.**

*Ref* = nucleotide in reference genome, *freq* = frequency, *Alt* = alternative nucleotide -> this is the mutation, *Change* = predicted consequence of the mutation. Frequencies derived from frequency of mutant reads in samples.

Gender	Sample	Exon	Ref	Alt	Alt freq	Expected alt freq	Change
male	1A	intron 18	A	G	85.8	100	intronic
male	1B	intron 18	A	G	72.9	100	intronic
male	2A	intron 18	A	G	82.9	100	intronic
male	2B	intron 18	A	G	85.3	100	intronic
male	3A	24	C	A	51	100	S>STOP
male	3B	24	C	A	0	100	N/A
female	4A	25	G	A	15.8	50	W>STOP
female	4B	25	G	A	10.4	50	W>STOP

## 3.6 Summary and Discussion

### 3.6.1 Identification of four novel X-linked clonal marks

The first aim of this part of the study was to develop novel clonal marks for investigation of human intestinal stem cell dynamics. As transgenic approaches are not feasible in humans, lineage tracing experiments rely on the detection of spontaneous somatic mutations. X-linked clonal marks were previously used to define the baseline stem cell dynamics in the human colon (Nicholson *et al.*, 2018). Here, the toolbox of such marks was expanded by means of an antibody screen. Out of 21 antibodies tested, three yielded new clonal marks: HDAC6, KDM6A and NONO. In addition, the previously identified STAG2 was validated.

As described in the introduction to this chapter, identification of novel X-linked clonal marks occurs in a sequential filtering process. It contains a number of decision points at which certain requirements need to be met for a candidate mark to proceed to the next stage. The first criterion is, as these were all candidate negative marks, uniform and strong staining. For four of the antibodies, against DDX3X, IDH3G, SLC25A5 and STARD8 this was not achieved. According to HPA tissue protein expression data, these proteins should be expressed at sufficient levels to be detected by IHC. As the antibodies used were still at development stage, this was likely due to absence of immunoreactivity. For most of the other antibodies tested, uniformly positive staining was achieved but no negative crypts were identified. This may have technical reasons such as lack of appropriate specificity of the antibody or failure of the antibody to detect specific mutations. These hypotheses could be tested by western blotting in wild-type and CRISPR-modified isogenic cell lines. Equally, this might provide biological insights. Potentially, mutations in these genes or their regulatory regions are exceedingly rare. They may confer a strong negative bias to affected stem cells, such that they are rapidly outcompeted by surrounding wild-type cells. Or mutations in these genes may not be tolerated in the colonic epithelium. In two cases, for SMC1A and HUWE1, weakly negative crypts were identified with HPA antibodies but this could not be validated. As the epitopes differ for the antibodies used, it is plausible that very specific mutations are not detected by most of them. Alternatively, the weaker staining seen using the HPA antibodies could have been an artefact.

---

When selecting X-linked candidates for the antibody screen, genes with a range of different characteristics (total gene length, combined exonic length, exon-intron ratio) were chosen (Table 3.1 and Figure 3.3). Analysis of the genes that yielded new marks reveals no unifying feature. In terms of total length, KDM6A and STAG2 rank near the top, while HDAC6 and NONO are considerably shorter. Coding sequence length proved no more indicative, with NONO being considerably shorter than the other four genes. Finally, the exon-intron ratios vary 10-fold between the highest (HDAC6) and the lowest (KDM6A). An approach not considered initially was analysis of RNA-Seq data for human colonic epithelium. This was retrospectively obtained from our collaborator Dr Matthias Zilbauer. Read counts were normalised for exonic sequence length, to account for the fact that longer genes are covered by more reads. This analysis revealed no correlation between normalised read counts and outcome of IHC as reported here (Table 3.6). This is exemplified by the failed marks as well as ADGRG4 (highlighted in red): For DDX3X, IDH3G and SLC25A5 reads are present but protein is not detected while IHC for ADGRG4 yielded strong signal despite no apparent RNA expression. These results may relate to fast RNA or protein turnover or lack of antibody specificity. The identified clonal marks exhibit a range of values for normalised reads. For future studies, a cut-off around 0.1 reads per bp could be considered. However, this will not eliminate negative results stemming from lack of antibody specificity, as illustrated by the high values displayed by four marks that were not detected.

Taken together, these results validate a very broad almost trial-and-error approach to screening for novel X-linked clonal marks. Importantly, only a very small number of genes were considered here. Therefore, a practical cut-off may be useful for larger scale studies.

Table 3.6 **Normalised human colonic RNA expression for candidate marks.**  
*For completeness MAOA and STAG2 are included.*

Gene	Exons (bp)	Mean reads	Reads per bp
SLC25A5	1307	5117	3.915
MAOA	4015	5041	1.256
RBMX	2012	1560	0.775
DDX3X	6907	4441	0.643
NONO	2661	1648	0.619
PDHA1	3484	1160	0.333
IDH3G	1335	442	0.331
STAG2	5218	1541	0.295
RPS6KA3	8111	2053	0.253
EIF1AX	4414	934	0.212
SH3KBP1	4728	909	0.192
HUWE1	14731	2760	0.187
HTATSF1	3019	522	0.173
SMC1A	9784	1269	0.130
KDM6A	5438	635	0.117
XIAP	8591	995	0.116
HDAC6	4196	440	0.105
ATRX	11167	840	0.075
HDAC8	1754	104	0.059
ZMYM3	5536	175	0.032
ZNF75D	5611	148	0.026
STARD8	4841	38	0.008
ADGRG4	9820	<1	<0.001

### 3.6.2 Identification of two CRC-associated clonal marks

X-linked clonal marks enable investigation of stem cell behaviours in human colonic epithelium. However, IHC detection of CRC-associated mutations could provide more specific insights into mechanisms of CRC initiation. Initially, detection of oncogenes and their surrogates was attempted. The antibodies were first tested in tumour tissue, where immunoreactivity was confirmed. This was followed up by IHC on normal colonic tissue. For antibodies aimed at detecting the common activating mutations *KRASG12D* and *BRAFV600E*

the patchy staining pattern obtained in normal colonic tissue sections indicated immunoreactivity with the wild-type protein. Equally, IHC for the surrogate markers DUSP6 and pE4-BP1 yielded strong immunoreactivity. In principle however, background immunoreactivity should not preclude detection of clones with significantly increased or mutant protein, as these would be expected to stand out. However, upon screening of 10 large area (>10,000) sections, no significantly darker crypts were identified. This may indicate non-specificity of the antibody or rarity of the pro-oncogenic genetic alterations in normal human colon. Sequencing-based methods may represent a more high-throughput and technically straightforward way of detecting such specific mutations.

Antibodies were also tested for detection of alterations in the CRC-associated tumour suppressors p53, PTEN and SMAD4. Negative clones were identified for PTEN, indicating genetic events causing protein loss or truncation. The *PTEN* gene is located on the autosomal chromosome 10. Hence the remaining wild-type copy of the gene would be expected to compensate for mutations of one allele. Alternatively, heterozygous mutations could manifest as weakly immunoreactive clones. The observed clones show very low levels of remaining immunoreactivity. As such they may be caused by homozygous mutations. In normal tissue such spontaneous homozygous mutations should occur very infrequently. Their presence may be indicative of positive selection pressure promoting loss of the second copy of the gene. Equally, promoter hypermethylation of *PTEN* has been reported as a frequent mechanism of gene expression reduction (Goel *et al.*, 2004). Bisulfite sequencing has recently been optimised for FFPE DNA (Ludgate *et al.*, 2017). It may be possible to adapt this technique to DNA from laser captured material. Combining it with sequencing this may identify the mechanism of PTEN protein reduction in PTEN-negative crypts.

A special case of detection of an altered tumour suppressor was p53, for which rare positive crypts were identified. Also known as ‘guardian of the genome’, p53 plays a crucial role in the cellular response to genotoxic stress (Lane, 1992). Levels of the protein are usually low due to ubiquitination and degradation (Lahav *et al.*, 2004). However, upon detection of a stress signal such as DNA damage, oncogene activation or hypoxia, p53 is stabilised via post-translational modification. It then exerts its tumour suppressive role by inducing a variety of mechanisms

including cell cycle arrest, apoptosis and senescence (Vogelstein, Lane & Levine, 2000; Vousden & Prives, 2009). Hence, when stabilised p53 is detected, this could be evidence of various stress signals, cancer-promoting mutations such as oncogene activation or loss of MDM2 as well as mutations in the *TP53* gene itself. Importantly, p53 stabilisation was observed in confined clones that look histologically normal. This is suggestive of a genetic event, as opposed to more widespread stress or insults such as infection.

Finally, IHC for SMAD4, which is more commonly mutated in human adenocarcinomas than PTEN, did not identify any negative crypts. As for X-linked marks, several reasons may account for this result. There may be a technical explanation, whereby the epitope recognised by the antibody is unaffected by the most frequent mutations in *SMAD4*. The antibody used here recognises a C-terminal epitope around amino acid residue 500 of the protein. According to the COSMIC database, the most common mutation in SMAD4 in CRC is R361H, which may not affect antibody binding. In addition, *SMAD4* mutations in normal colon may be mostly heterozygous and therefore hard to detect by IHC. Homozygous mutations may be exceedingly rare, not be tolerated in normal human colon or rapidly convert to pathology.

### 3.6.3 LCM and sequencing of MAOA<sup>-</sup> and KDM6A<sup>-</sup> crypt sections

To complement and validate the IHC-based work, a protocol for sequencing of DNA from laser capture microdissected FFPE colonic epithelial sections was devised. Formalin fixation causes DNA fragmentation, so evaluation of the quality of DNA obtained from the FFPE tissue available was paramount. qPCR demonstrated that DNA can be extracted from laser capture microdissection of samples as small as 10 FFPE crypt sections and can be used as a template for PCR. However, for reliable amplification, amplicon size should not exceed 150 bp. This lays the groundwork for further validation of clonal marks by sequencing as well as targeted sequencing approaches where IHC was unsuccessful for mutation detection.

Laser capture microdissection and next-generation sequencing of multicrypt patches revealed that the absence of staining correlates with mutations in *MAOA* and *KDM6A*. Patches contained single mutations that were absent in surrounding crypts, validating the hypothesis that patches arise from crypt division (For a detailed discussion see chapter 5). The fractions of mutant reads

obtained largely corresponded to the expected values. In some cases, a substantially lower frequency was found. This discordance has two likely causes: Firstly, the poor DNA quality may have caused allele dropout leading to an overrepresentation of wild-type DNA molecules in the sample. Also, wild type stroma was microdissected alongside mutant crypts, further lowering the mutant allele frequency in the sample. For *MAOA*, mutations were identified in 5 out of 11 patches. From this small sample size, it appears that exons 8 and 13 may represent mutational “hotspots”. Interestingly, a deletion in exon 13 was present in two independent samples. Sequencing of a larger number of mutant patches could confirm whether this is a particularly common mutation. Further informative experiments would be introduction of the mutations into cell lines by CRISPR or computational modelling of MAOA structure to fully confirm that the mutations are causative of the loss of staining. The antibody used in this study binds to the amino acid residues encoded by exons 13-15. This may explain the mutations found in exon 13. However, it is plausible that a mutation in exon 8 causes an overall change in the MAOA protein structure, hence abolishing antibody binding. In addition, as the entire coding sequence of MAOA was analysed, the lack of mutation identification in 6 out of 11 patches is indicative of alternative mechanisms causing the loss of staining. For example, intronic mutations could generate truncated forms of MAOA via altered splicing. This is supported by the finding of an intronic mutation in a *KDM6A*-negative patch and could be tested by sequencing of introns in MAOA-negative patches. Equally, mutations may affect regulatory elements or genes involved in the regulation of *MAOA* transcription or translation. For example, the transcription factor *GATA2* has been implicated in regulating MAOA expression (Gupta *et al.*, 2015). To test the involvement of transcription factors, IHC could be used, while regulatory sequences could be explored by sequencing. Finally, epigenetic changes could cause the loss of MAOA expression. Potential changes in DNA methylation could be investigated by bisulfite or methylation sequencing.

For *KDM6A*, only partial coverage of the coding sequence was achieved, it thus is not surprising that mutations were only identified in 4 out of 11 sequenced patches. For two of these mutations, W1193\* and S1154\*, entries can be found in the COSMIC database. Intriguingly, both these mutations are located in the JmjC domain of the protein, which is the catalytic domain

---

responsible for demethylation of histones. A third mutation, located in intron 18, was found in two separate KDM6A-negative patches. It is conceivable that this would impair splicing and therefore generate mutant forms of KDM6A protein. Interestingly, the antibody used for detection of KDM6A binds near the amino acid 490 of the protein, N-terminally to the mutations observed here. However, it is plausible that they cause large-scale changes in the 3D structure of the protein, therefore impeding antibody binding. All these hypotheses could be tested using CRISPR/Cas9 in cell lines followed by IHC or Western blotting as well as computational modelling of KDM6A protein structure. Importantly, no mutations were found in 7 out of 11 KDM6A-negative patches. As only parts of the coding sequence were analysed, the experiments presented here should be followed up by sequencing of the remaining exonic sequence. In addition, the same mechanisms discussed for *MAOA* may also regulate *KDM6A* expression and could be investigated in an analogous manner.



## Chapter 4 Inference of intra-crypt behaviours

### 4.1 Introduction

According to the Big Bang model, colorectal cancer is initiated when a number of factors reaches a critical threshold (Sottoriva *et al.*, 2015). One important such factor are mutations, which, to initiate cancer, must accumulate in normal tissues. To persist within renewing tissues such as the intestine, mutations must occur in stem cells. However, each colonic crypt contains a number of effective stem cells that replace each other in a process of neutral drift (Lopez-Garcia *et al.*, 2010; Snippert, van der Flier, *et al.*, 2010; Baker *et al.*, 2014; Nicholson *et al.*, 2018). Therefore, the fate of neutral mutations within the crypt is stochastic and the probability of fixation is proportional to the number of stem cells. Mouse studies have shown that this balance can shift in the presence of pro-oncogenic mutations such as *Kras* activation and loss of *Apc*, whereby affected stem cells are more likely to achieve fixation within a crypt. Importantly, the process remains stochastic (Vermeulen *et al.*, 2013). This mechanism can enable cancer drivers to achieve significantly higher mutant allele frequencies within the intestinal epithelium than would be possible from mutation alone.

In human colon, clonal marks allow for investigation of stem cell dynamics. Using a large set of colonic samples from individuals of a wide range of ages enabled identification of age-related trends in clone accumulation for the neutral marks mPAS and MAOA, which allowed for inference of the number ( $N_{crypt}$ ) and replacement rate ( $\lambda_{crypt}$ ) of functional stem cells in the human colon (Nicholson *et al.*, 2018). From this result the next key question was whether mutations could, as had been seen in the mouse, subvert the dynamics in the crypt to achieve higher probability of fixation and therefore higher mutant allele burden in the epithelium.

### 4.1.1 Aim

The screens described in chapter 3 identified and validated six novel clonal marks: HDAC6, KDM6A, NONO, TP53, PTEN and STAG2. Here, the aim was to use these marks to determine whether different gene-specific mutations could alter the dynamics within human colonic crypts to achieve altered probability of fixation.

### 4.1.2 Published work

The data I obtained for STAG2 was included in our recent publication (Nicholson *et al.*, 2018), on which I am second author.

## 4.2 General workflow

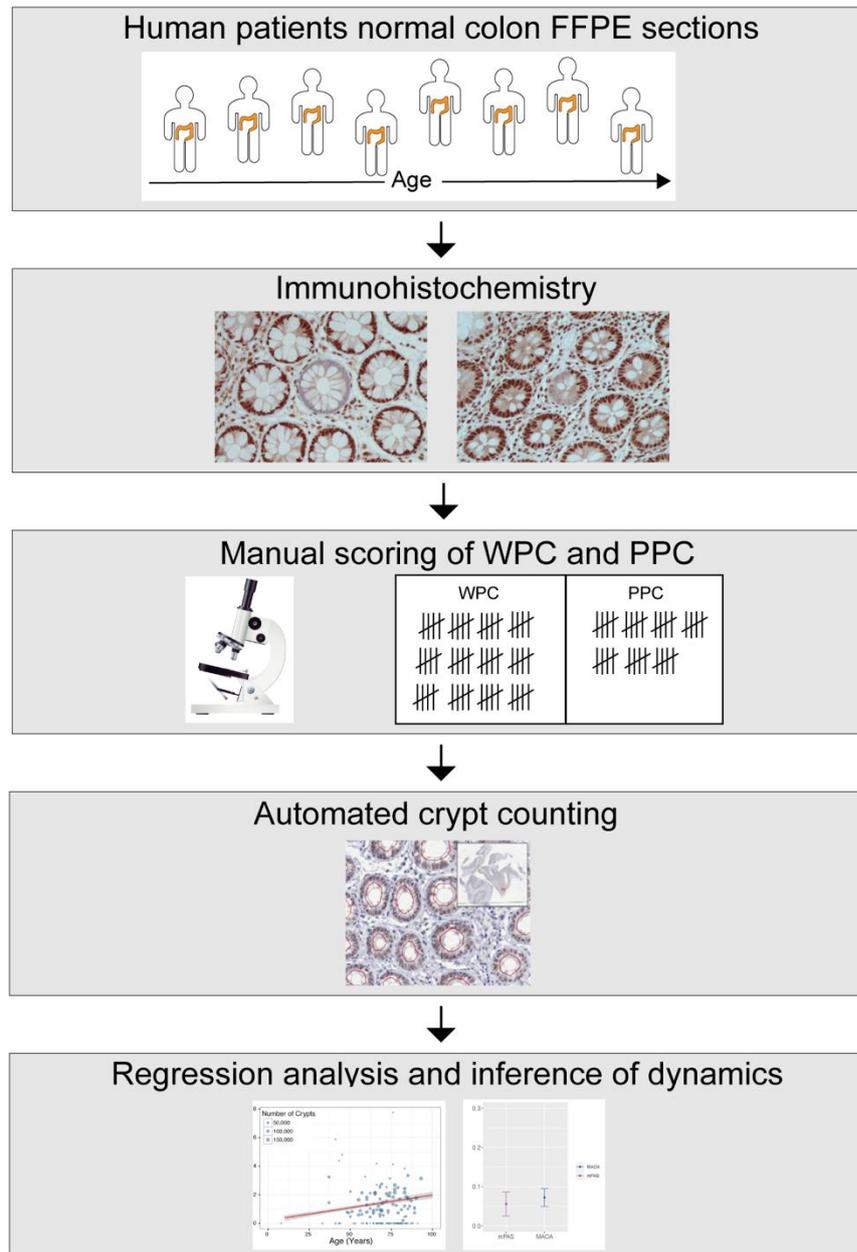
To infer the stem cell behaviours associated with the new clonal marks, the same method that enabled calculation of  $N_{crypt}$  and  $\lambda_{crypt}$  for the human colon using mPAS and MAOA was applied (Figure 4.1).

### 4.2.1 Scoring of WPC and PPC frequencies

For each of the clonal marks identified in chapter 3, IHC was performed on FFPE colonic tissue sections from at least 50 individuals from a range of ages and numbers of WPC and PPC manually scored (Figure 4.1). An exception was PTEN, which was developed most recently and for which therefore only 36 individuals are included here.

The total number of crypts on each section was determined by means of a neural network-driven image analysis software package, Deep Crypt Image Classification Segmentation (DeCryptICS), developed by Dr Edward Morrissey and Dr Doran Khamis (<https://github.com/MorrisseyLab/DeCryptICS>, manuscript in preparation) to enable calculation of WPC and PPC frequencies. Plots revealed age-related accumulation of WPC and constant background frequencies of PPC for all new clonal marks. Regression analysis was then

performed on the data to obtain the slope of WPC accumulation,  $\Delta C_{fix}$ , and the frequency of PPC,  $C_{part}$ , from the data (see Appendix C for the mathematical methods used) (Figure 4.1).



**Figure 4.1 Workflow for inference of stem cell behaviours.**

*Normal FFPE colon sections from at least 50 individuals are stained and manually scored. Crypt counts are obtained by means of DeCryptICS. Regression analysis is then performed on clone frequencies to obtain  $\Delta C_{fix}$  and  $C_{part}$ , which are used for inference of stem cell behaviour.*

### 4.2.2 A note on sampling and noise

For all marks, the WPC data showed a large spread. This may lead a critical mind to question the linear regression performed to derive  $\Delta C_{fix}$ . Human data is, due to a large number of genetic as well as environmental factors, notoriously noisy. In addition, when sampling events with a low frequency, such as the clones detected here, the sample size will heavily influence fluctuations in the data. This relationship between sampling and calculated clone frequency is illustrated in Figure 4.2A. Simulations performed by Dr Edward Morrissey show accurate regression fitting despite the noise (Figure 4.2B). Importantly, the Bayesian inference used here to obtain  $\Delta C_{fix}$  takes account of variable estimates in clone frequencies arising from sampling noise.

### 4.2.3 Inferring intra-crypt bias from $\Delta C_{fix}/C_{part}$

The ratio of  $\Delta C_{fix}/C_{part}$  enables inference of biases in stem cell behaviour. This is because  $\Delta C_{fix}$  and  $C_{part}$  are dependent on the rate of stem cell replacement and number of effective stem cells per crypt, respectively, but are also both dependent on the mutation rate. Therefore, the ratio of  $\Delta C_{fix}/C_{part}$  is a metric that describes the dynamics of monoclonal conversion independently of mutation rate. For advantageous mutations an upwards shift as compared to neutral marks will be observed, as PPC are underrepresented due to increased rate of fixation into WPC. Conversely, disadvantageous mutations will decrease  $\Delta C_{fix}/C_{part}$  as decreased probability of clone fixation will lead to WPC underrepresentation.

Additionally, the probability of replacement ( $P_R$ ) associated with a particular mark can be estimated from a full equation containing  $\Delta C_{fix}/C_{part}$ ,  $N_{crypt}$  and  $\lambda_{crypt}$  (see Appendix C for full mathematical model). For neutral marks, this is 0.5, as every stem cell has a 50% chance of replacing its neighbour and a 50% chance of being replaced by the neighbour.

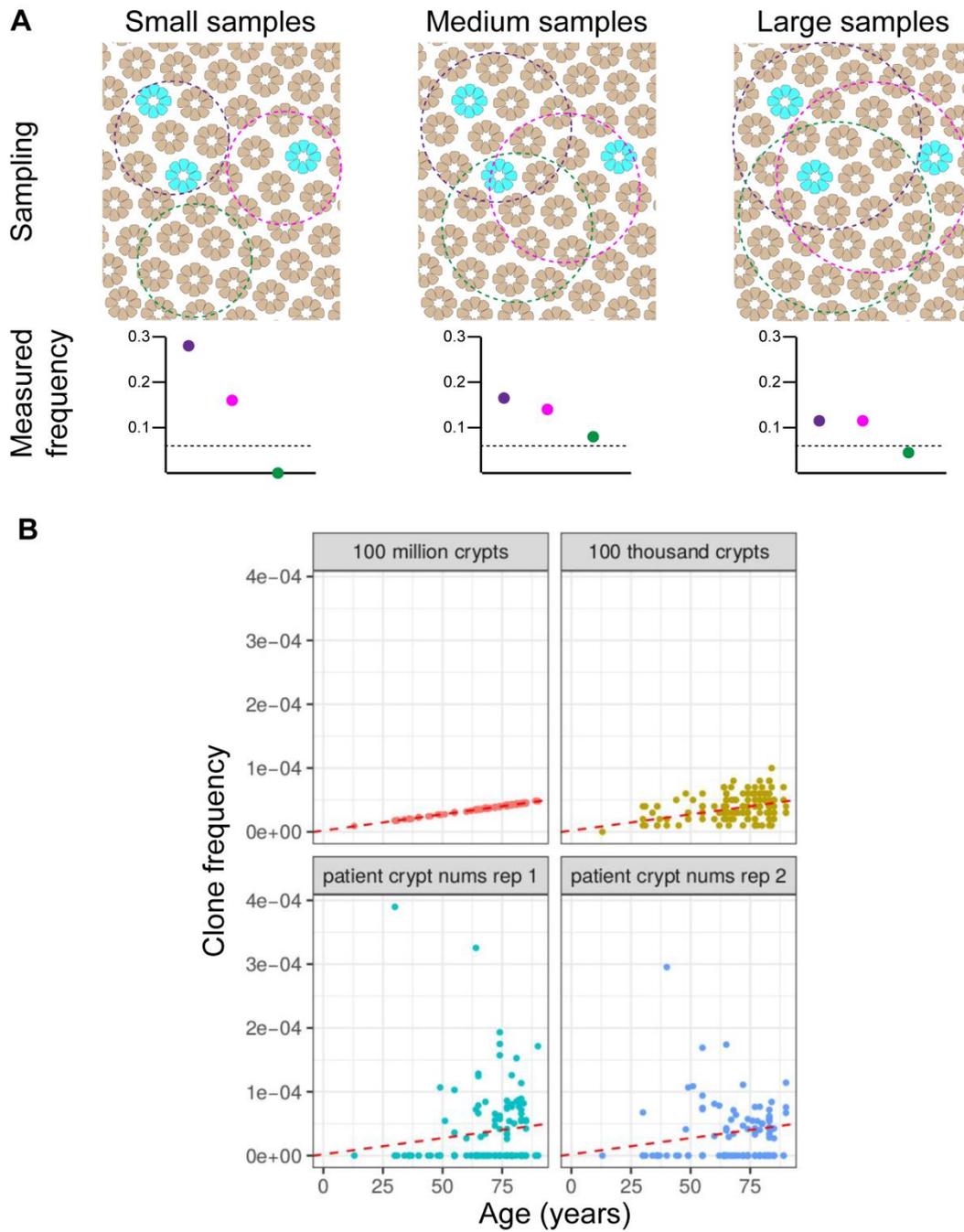


Figure 4.2 **Sampling and noise in human clonal marks data.**

(A) Schematic illustration of how sample size influences the measured clone frequency. Top: Dashed circles represent samples of increasing size. Bottom: Coloured dots represent the clone frequency obtained by the corresponding sample. Dashed line represents the actual clone frequency. Larger samples result in measured clone frequencies approaching the real value.

(B) Simulation of clonal marks data. Top: Scoring of 100 million (left) or 100 thousand (right) crypts per patient. Bottom: Independent simulations of scoring crypt numbers typically used here. Red line:  $C_{fix}$ .

## 4.3 Behaviours associated with X-linked marks

### 4.3.1 NONO confirms neutral stem cell dynamics

The 471 amino acid protein NONO (Non-POU domain-containing octamer binding protein) is a member of the *Drosophila* behaviour/human splicing (DBHS) protein family. The protein is modular in structure, possessing domains for protein as well as nucleic acid binding. This translates into association with a broad spectrum of transcription factors to regulate a variety of aspects of gene expression such as transcriptional activation, repression, initiation, elongation and termination (Knott, Bond & Fox, 2016). NONO has also been found to be involved in the formation of paraspeckles, which are ribonucleoprotein bodies localised in the nucleus of mammalian cells (Fox *et al.*, 2002; Bond & Fox, 2009; Passon *et al.*, 2012). Most recently, NONO was reported to be regulating telomere stability (Petti *et al.*, 2019). Germline mutations in *NONO* have been associated with X-linked mental retardation (Carlston *et al.*, 2019).

To look at the intra-crypt effect of *NONO* mutations, IHC was performed on FFPE colon sections from 78 patients aged 20–93 years. WPC and PPC were then scored (Figure 4.3A) and plotted, revealing an age-related increase of WPC with a  $\Delta C_{fix}$  of  $1.44 \times 10^{-6}$  (95% CI:  $1.04 \times 10^{-6}$ – $2.04 \times 10^{-6}$ ) (Figure 4.3B) and a constant  $C_{part}$  (Figure 4.3C). The value obtained for  $\Delta C_{fix}/C_{part}$  was 0.082 (95% CI: 0.048–0.141), remarkably similar to both mPAS and MAOA (Figure 4.3D). In addition,  $P_R$  was calculated to be 0.546 (95% CI: 0.454–0.678). This indicated that mutations affecting NONO protein levels are probably neutral, meaning they do not alter the competition between stem cells in the crypt.

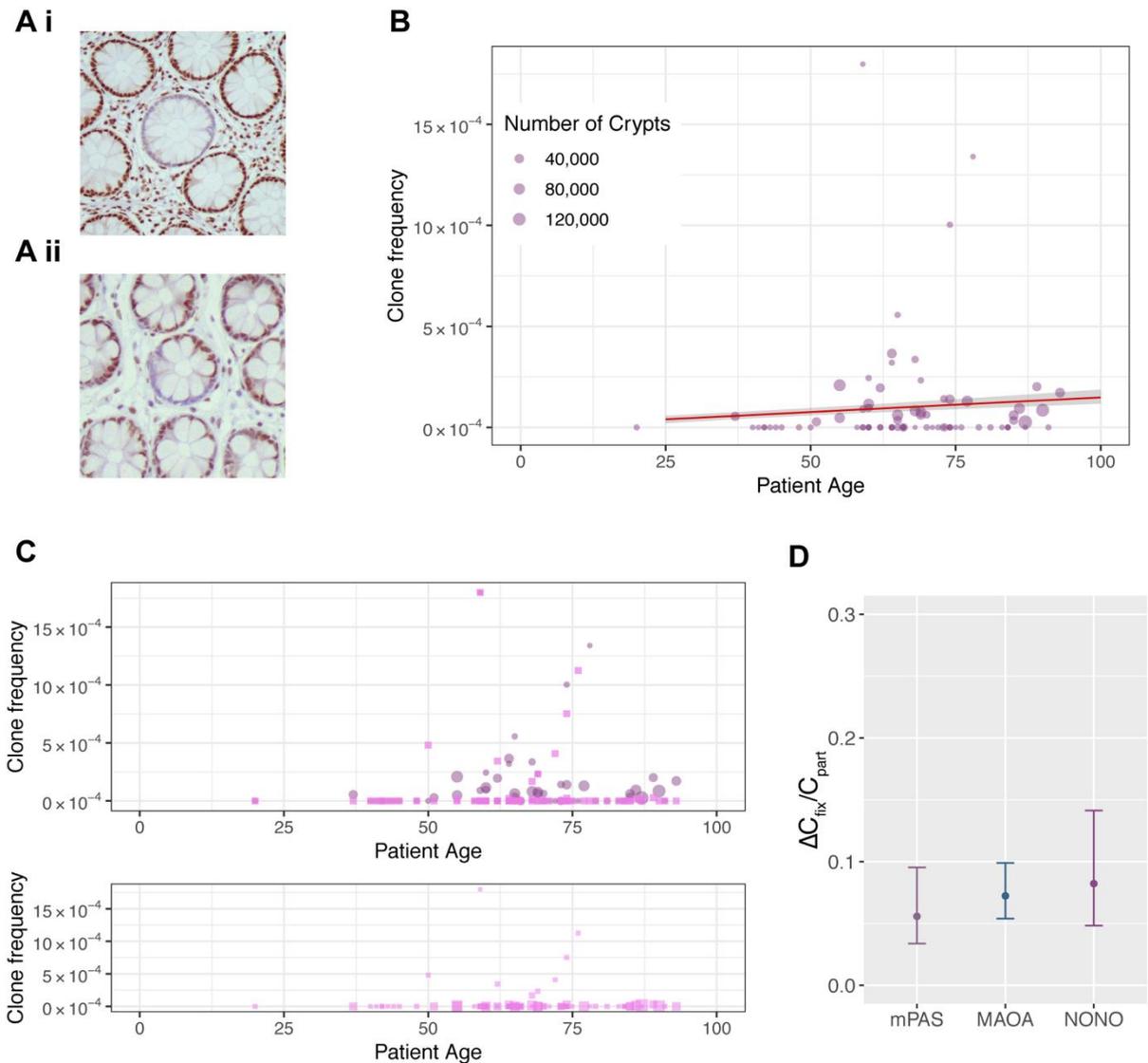


Figure 4.3 **Inference of stem cell behaviour associated with NONO mutation.**

(A) (i) WPC and (ii) PPC in en face tissue sections stained for NONO.

(B) Regression analysis showing  $\Delta C_{fix}$  for NONO ( $1.44 \times 10^{-6}$  per year) with 95% margin of error (ME) shaded in grey.

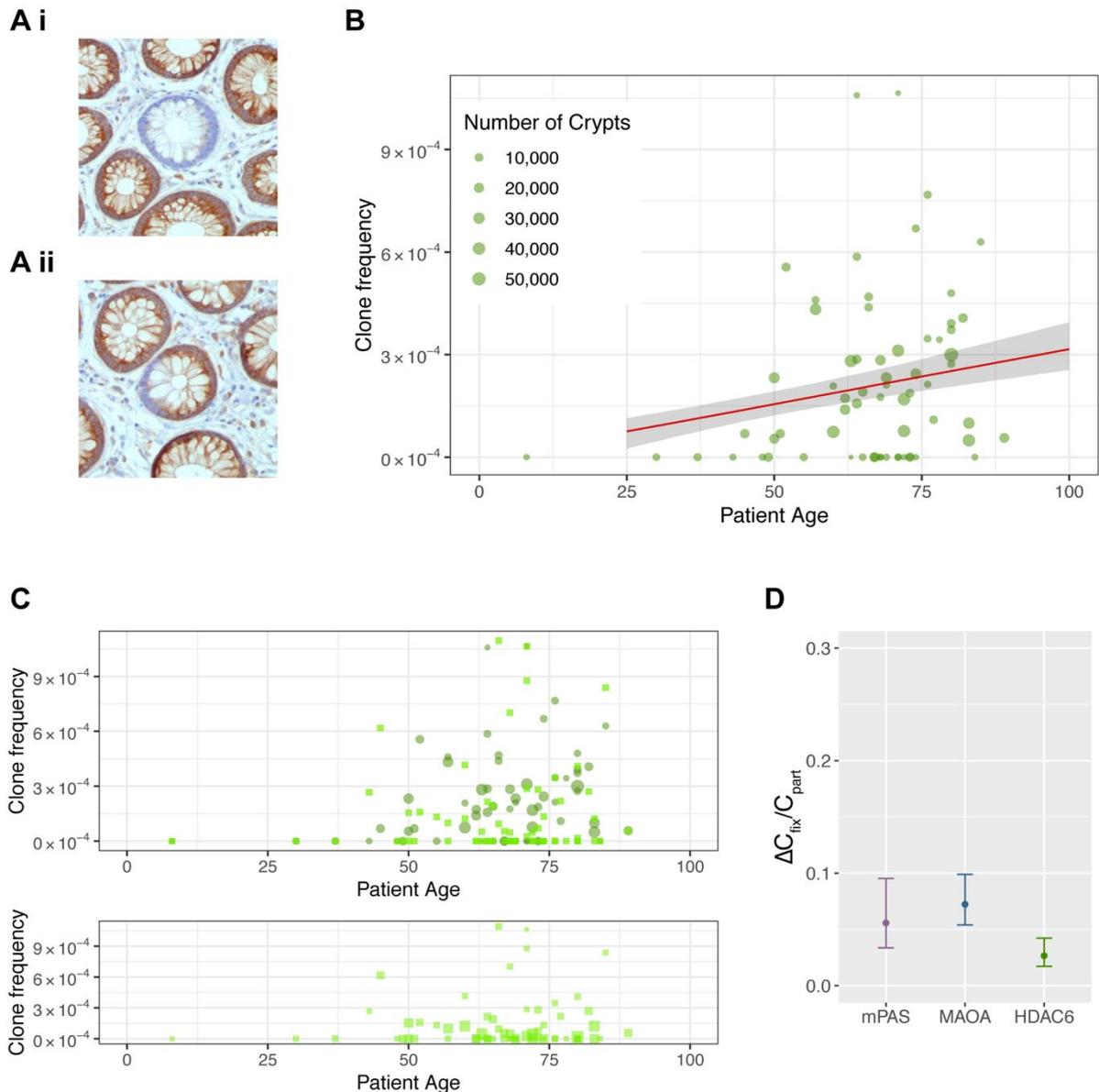
(C) Frequency plot of WPC (circles) and PPC (squares) for NONO clones. Bottom panel shows PPC frequencies alone on condensed y-axis.

(D) Plot showing similar ratio for  $\Delta C_{fix}/C_{part}$  for mPAS, MAOA and NONO. Error bars = 95% CI.

### 4.3.2 HDAC6 loss confers disadvantage

Histone deacetylase 6 (HDAC6) is a 1215 amino acid protein with a counterintuitive name. It is a unique member of the type II HDACs in that due to its nuclear export sequence (NES) as well as cytoplasmic retention motif (SE14) it is actually located in the cytoplasm (Verdel *et al.*, 2000; Bertos *et al.*, 2004). Therefore, contrary to its name, HDAC6 does not deacetylate histones to regulate gene activity in the nucleus but works to maintain the acetylation balance of various non-histone substrates. This includes the cytoskeletal proteins  $\alpha$ -tubulin (the building blocks of microtubules) and cortactin, which means that HDAC6 plays an important role in cell motility (Hubbert *et al.*, 2002; Zhang *et al.*, 2007; Deakin & Turner, 2014). An additional target of HDAC6 is the chaperone HSP90 (Aoyagi & Archer, 2005), which assists protein folding.

IHC for HDAC6 was performed on colonic FFPE samples from 66 individuals of the age range 8-89 years, WPC and PPC scored (Figure 4.4A) and plotted. This revealed an age-related increase in WPC (Figure 4.4B) with a  $\Delta C_{fix}$  of  $3.19 \times 10^{-6}$  (95% CI:  $2.31 \times 10^{-6}$ - $4.65 \times 10^{-6}$ ) and a constant frequency of PPC (Figure 4.4C). Following regression analysis, the ratio of  $\Delta C_{fix}/C_{part}$  revealed a value of 0.027 (95% CI: 0.017–0.042) for HDAC6. This is significantly (p-value = 0.0004) lower than the values obtained for the neutral marks mPAS and MAOA (Figure 4.4D), implying a disadvantage. In line with this finding, a  $P_R$  of 0.432 (95% CI: 0.359–0.494) was calculated.



**Figure 4.4 Inference of stem cell behaviour associated with HDAC6 mutation.**

(A) (i) WPC and (ii) PPC in en face tissue sections stained for HDAC6.

(B) Regression analysis showing  $\Delta C_{\text{fix}}$  for NONO ( $3.19 \times 10^{-6}$  per year) with 95% margin of error (ME) shaded in grey.

(C) Frequency plot of WPC (circles) and PPC (squares) for NONO clones. Bottom panel shows PPC frequencies alone on condensed y-axis.

(D) Plot showing lower ratio for  $\Delta C_{\text{fix}}/C_{\text{part}}$  for HDAC6 as compared to mPAS and MAOA. Error bars = 95% CI.

### 4.3.3 KDM6A loss confers advantage

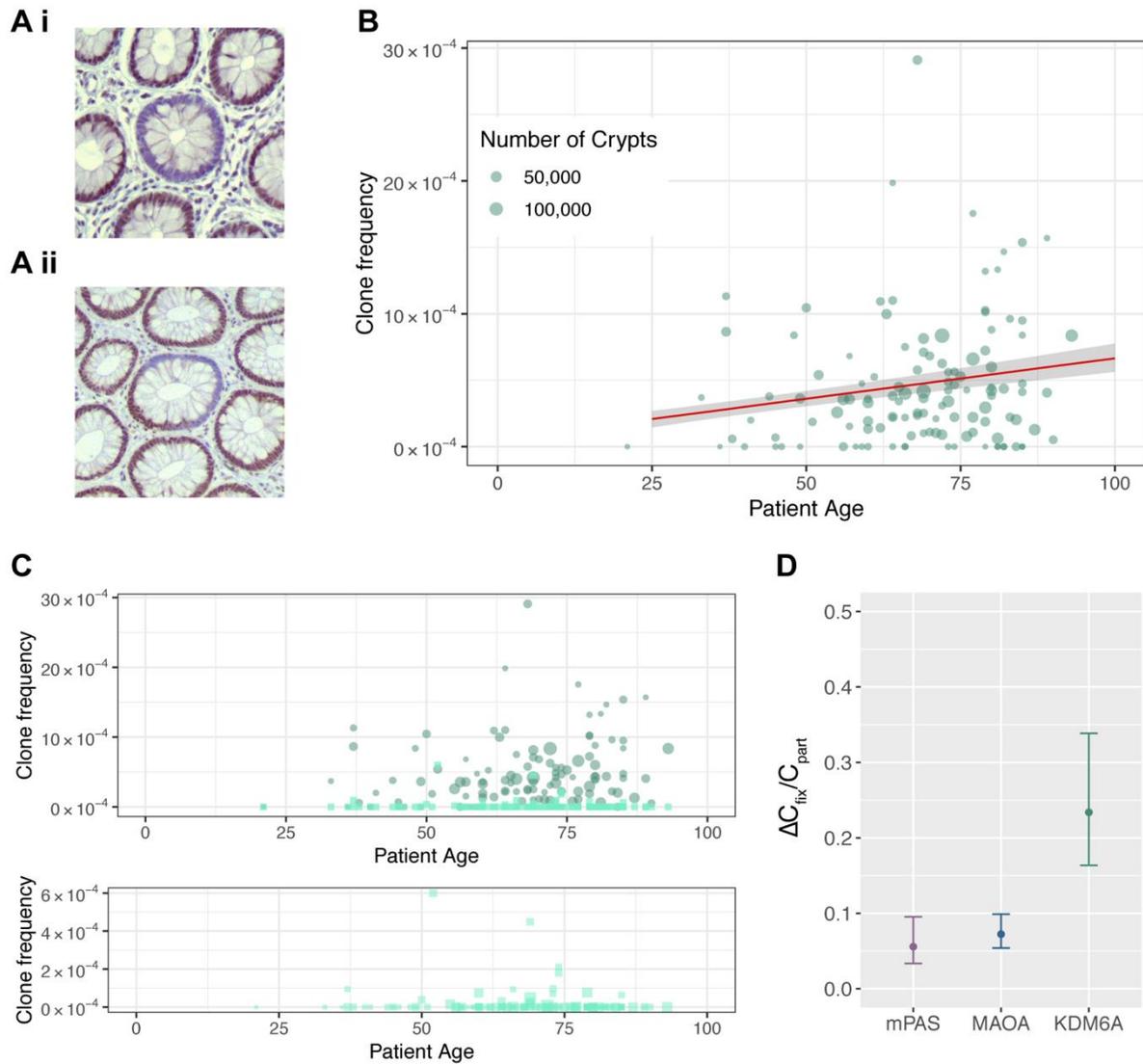
KDM6A (lysine-specific demethylase 6A), also known as UTX (ubiquitously transcribed tetratricopeptide repeat X chromosome) is a 1401 amino acid protein with two paralogs in humans, KDM6B and UTY. All three contain a JmjC-domain which catalyses the demethylation of tri- or dimethylated histone H3. Interestingly, the catalytic activity of UTY is negligible (Walport *et al.*, 2014). KDM6A acts as part of the MLL2/3 or ‘COMPASS’ protein complex (also described as MLL3/4 or ASCOM complex) which transcriptionally activates genes mainly via H3K4 methylation and H3K27me2/me3 demethylation (Cho *et al.*, 2007). It thus antagonises the action the polycomb repressor complex 2 (PRC2), which catalyses H3K27 tri- and dimethylation (Margueron & Reinberg, 2011). The COMPASS complex has also been shown to be involved in chromatin remodelling via the SWI/SNF complex, a function that does not seem to require demethylase activity of KDM6A (Miller, Mohn & Weinmann, 2010). Germline *KDM6A* mutations cause the hereditary Kabuki syndrome, a condition characterised by a broad spectrum of symptoms including distinctive facial features, skeletal abnormalities and mild to moderate intellectual disabilities (Niikawa *et al.*, 1981).

For determination of intra-crypt dynamics, colonic FFPE sections from 120 patients aged 21–93 years were stained, WPC and PPC (Figure 4.5A) scored and plotted, revealing an age-related increase in WPC (Figure 4.5B) with a  $\Delta C_{fix}$  of  $6.04 \times 10^{-6}$  (95% CI:  $4.69 \times 10^{-6}$ – $7.80 \times 10^{-6}$ ) (Figure 1.5B) and constant  $C_{part}$  (Figure 4.5C). The value obtained for  $\Delta C_{fix}/C_{part}$  was 0.234 (95% CI: 0.164–0.339) around 4.5-fold higher than that for neutral marks, indicating a competitive advantage (Figure 4.5D).  $P_R$  was calculated to be 0.760 (95% CI: 0.605–0.999) implying that KDM6A-deficient stem cells have a significantly increased probability of populating the crypt.

#### 4.3.3.1 Loss of KDM6A does not alter H3K27me3 levels in human colon

As KDM6A functions in demethylation of H3K27me3, loss of this protein could in theory lead to an increase in this histone modification. To test whether loss of KDM6A resulted in greater

abundance of H3K27me3 in human colonic epithelium, IHC for H3K27me3 was performed on serial sections. This did not show any increase in H3K27me3 in KDM6A-negative patches (Figure 4.6) (n = 5).



**Figure 4.5 Inference of stem cell behaviour associated with KDM6A mutation.**

(A) (i) WPC and (ii) PPC in *en face* tissue sections stained for KDM6A.

(B) Regression analysis showing  $\Delta C_{\text{fix}}$  for KDM6A ( $6.53 \times 10^{-6}$  per year) with 95% margin of error (ME) shaded in grey.

(C) Frequency plot of WPC (circles) and PPC (squares) for KDM6<sup>-</sup> clones. Bottom panel shows PPC frequencies alone on expanded y-axis.

(D) Plot showing 4.5-fold higher ratio for  $\Delta C_{\text{fix}}/C_{\text{part}}$  for KDM6A as compared to mPAS and MAOA. Error bars = 95% CI.

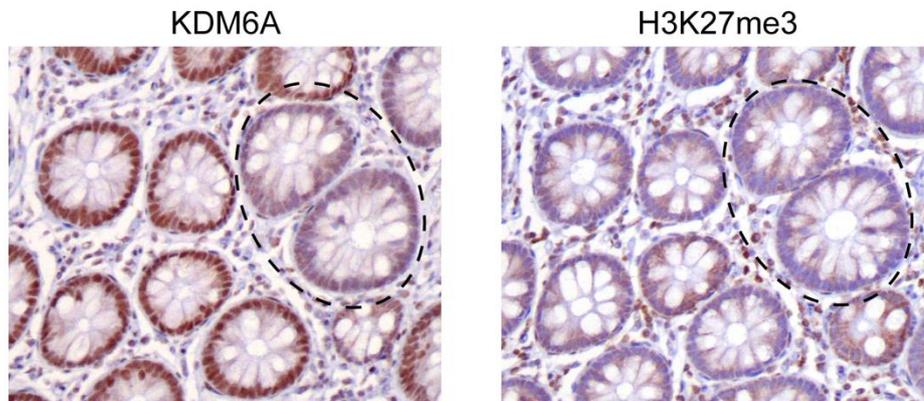


Figure 4.6 **Loss of KDM6A does not increase H3K27me3 in human colon.**

Representative images of IHC for KDM6A and H3K27me3 on serial human colonic FFPE sections. The same two KDM6A<sup>-</sup> crypts are highlighted in both images.

#### 4.3.3.2 KDM6A may not escape X-inactivation in human colon

Choosing X-linked genes for clonal marks is advantageous as X-inactivation in females effectively leaves their cells haploid for these genes, which due to the XY karyotype is always the case for males. Therefore, in both sexes, one ‘hit’ should be sufficient for loss of staining, enabling higher clone frequencies and therefore more robust data than would be obtained from autosomal genes. However, it has been known since the 1990s that some X-linked genes escape X-inactivation in somatic cells (Carrel *et al.*, 1999). Recently, analysis of both whole exome sequencing as well as data from the GETx project found that tumour suppressor genes that escape from X-inactivation (EXITS genes) may contribute to the male bias for certain cancers. Importantly, the list of candidate EXITS genes included *KDM6A* (Dunford *et al.*, 2017). In addition, a recent analysis of RNA-seq data from 29 tissues including the colon found *KDM6A* expression to be higher in females than males, suggestive of incomplete X-inactivation (Tukiainen *et al.*, 2017). Escape from X-inactivation should manifest in our samples as markedly reduced clone frequency in females. However, this is not the case, as there is no marked difference in *KDM6A* WPC frequency between females and males, with remarkably similar values of  $\Delta C_{fix}$  of  $6.36 \times 10^{-6}$  (95% CI:  $4.50 \times 10^{-6}$ – $9.02 \times 10^{-6}$ ) and  $6.44 \times 10^{-6}$  (95% CI:  $4.85 \times 10^{-6}$ – $8.57 \times 10^{-6}$ ), respectively (Figure 4.7). This data suggests that in human colon,

*KDM6A* does not escape X-inactivation. Furthermore, there was no difference in intensity of *KDM6A* staining between female and male colonic samples. Therefore, the higher female expression reported by Tukiainen and colleagues may not be reflected by protein abundance.

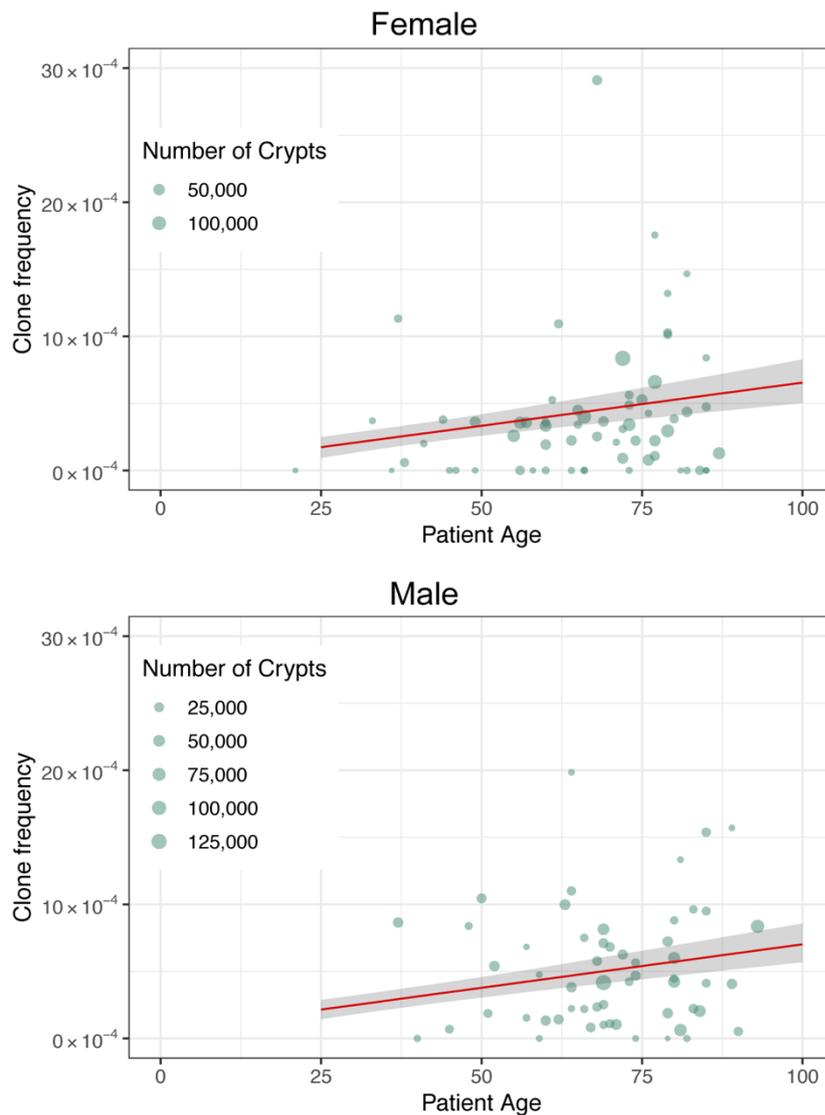


Figure 4.7 ***KDM6A*<sup>-</sup> WPC accumulation in females and males.**

Regression analysis showing very similar  $\Delta C_{fx}$  for *KDM6A*<sup>-</sup> WPC in females ( $6.36 \times 10^{-6}$  per year) and males ( $6.44 \times 10^{-6}$  per year) with 95% margin of error shaded in grey.

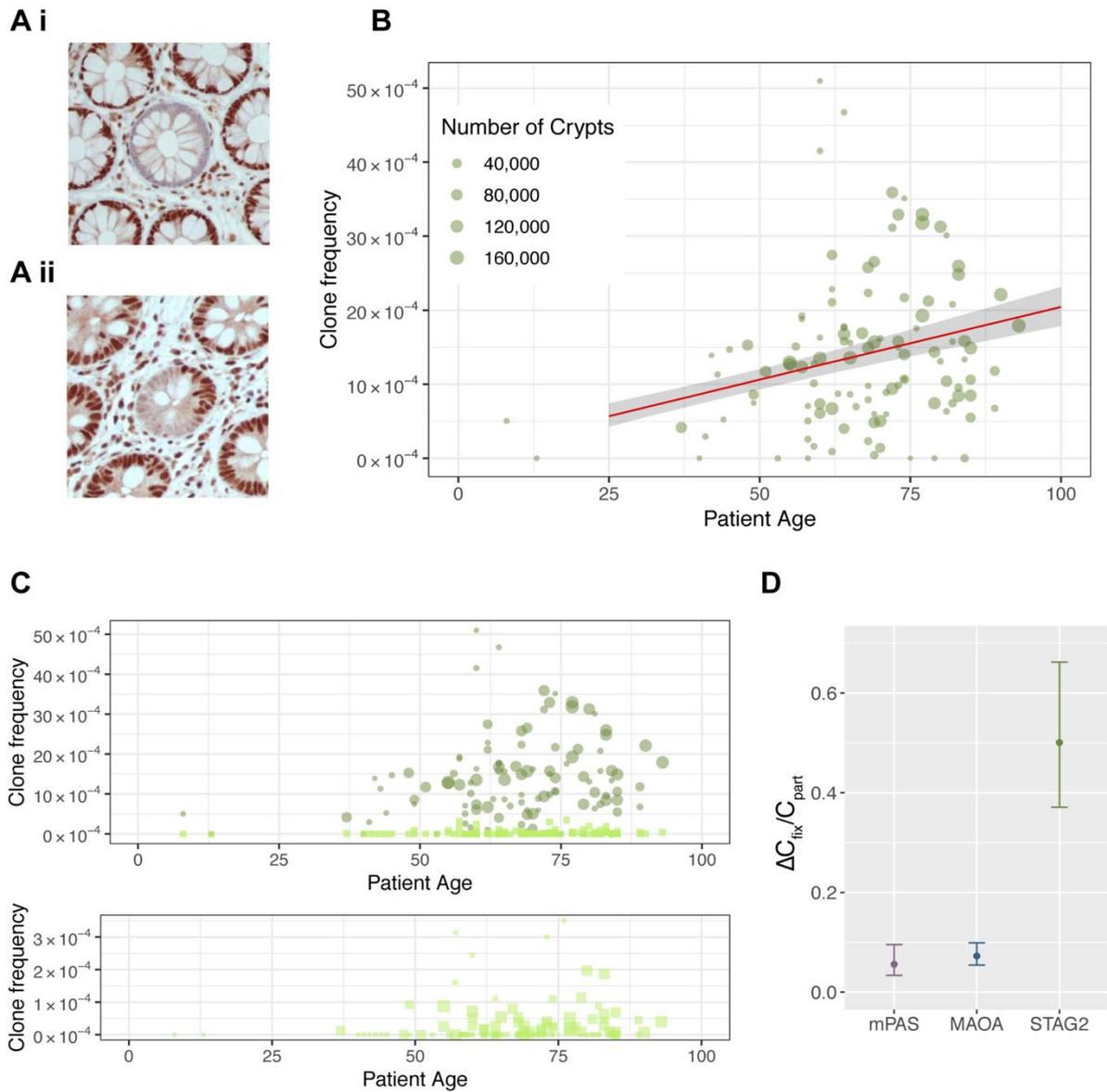
#### 4.3.4 STAG2 loss confers greater advantage than KDM6A

The 1231 amino acid protein STAG2 (stromal antigen 2) is a member of the cohesin complex, which is highly conserved from yeast to human. There are two types of this four-subunit complex. Both contain two structural maintenance proteins SMC1 and SMC3 as well as RAD21 and one of the two STAG homologs 1 or 2. The complexes function to maintain telomere or centromere cohesion, respectively (Canudas & Smith, 2009). Germline mutations in *STAG2* have recently been associated with an X-linked intellectual deficiency syndrome (Soardi *et al.*, 2017).

IHC for STAG2 was performed on colonic FFPE sections from 110 patients ranging in age from 8–93 years. Scoring of WPC and PPC (Figure 4.8A) frequencies revealed an age-related increase for WPC with a  $\Delta C_{fix}$  of  $1.95 \times 10^{-5}$  (95% CI:  $1.56 \times 10^{-5}$ – $2.39 \times 10^{-5}$ ) (Figure 4.8B) and a constant  $C_{part}$  (Figure 4.8C). The ratio of  $\Delta C_{fix}/C_{part}$  was 0.50 (95% CI: 0.371–0.662), a 10-fold increase compared to mPAS, MAOA (Figure 4.8D), as well as nearly 2-fold increase compared to KDM6A, suggesting a strong advantage in stem cell replacement. Indeed,  $P_R$  was calculated to be 0.999 (95% CI: 0.804–0.999), meaning that a STAG2-deficient stem cell will almost certainly populate the crypt.

#### 4.4 Intra-crypt behaviours for cancer-associated marks

The results obtained for X-linked clonal marks indicate that gene-specific mutations in stem cells can result in a range of behaviours in the crypt. Importantly, an advantage was observed for loss of KDM6A and STAG2. Such mechanisms could potentially also exist for known colorectal cancer drivers, enabling them to increase their likelihood of fixation and therefore reaching higher mutant allele burden within the tissue.



**Figure 4.8 Inference of stem cell behaviour associated with STAG2 mutation.**

(A) (i) WPC and (ii) PPC in en face tissue sections stained for STAG2.

(B) Regression analysis showing  $\Delta C_{\text{fix}}$  for STAG2 ( $1.96 \times 10^{-5}$  per year) with 95% margin of error (ME) shaded in grey.

(C) Frequency plot of WPC (circles) and PPC (squares) for STAG2<sup>-</sup> clones. Bottom panel shows PPC frequencies alone on expanded y-axis.

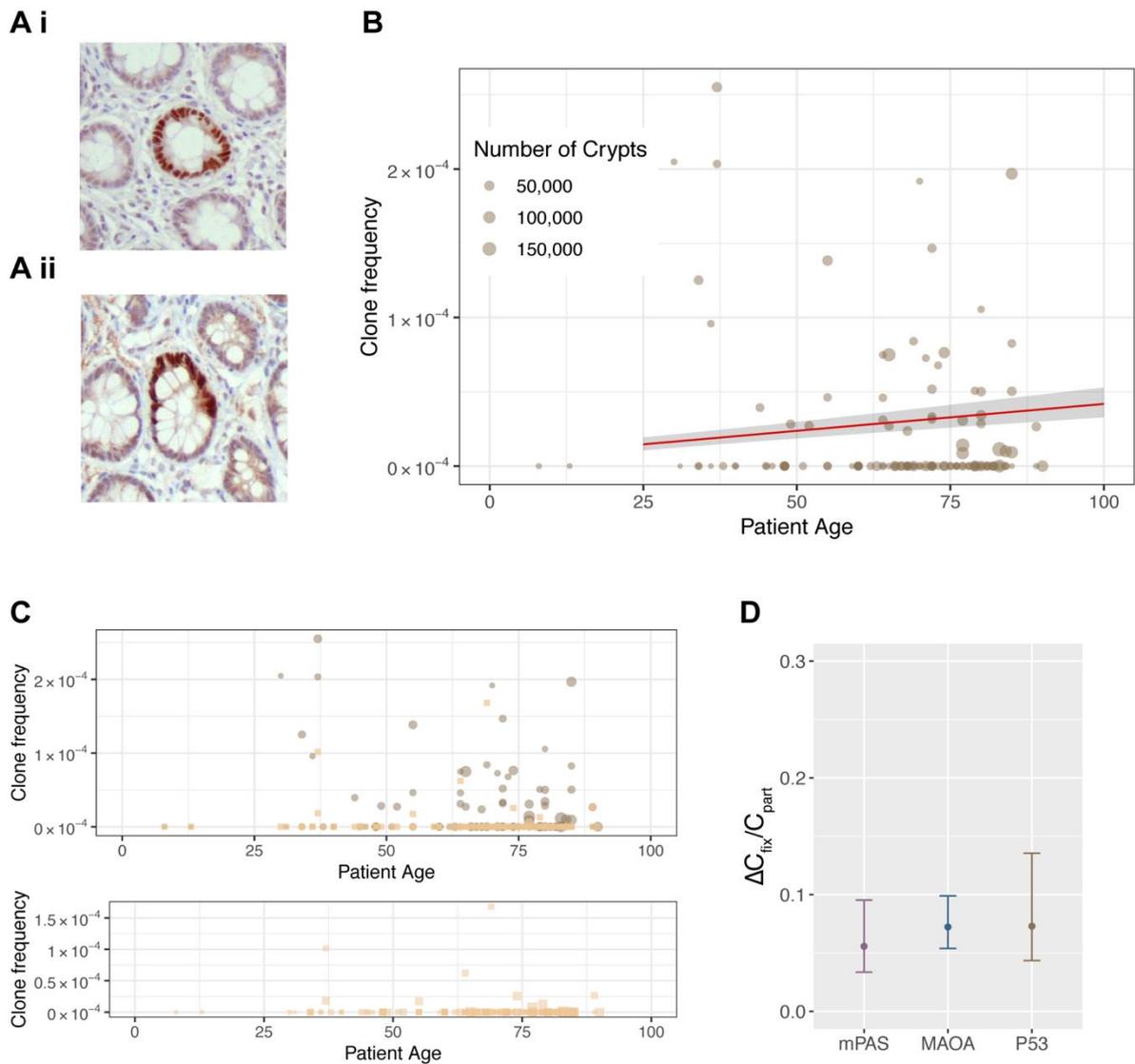
(D) Plot showing 10-fold higher ratio for  $\Delta C_{\text{fix}}/C_{\text{part}}$  for STAG2 as compared to mPAS and MAOA. Error bars = 95% CI.

#### 4.4.1 P53 stabilisation is associated with neutral stem cell dynamics

The human *TP53* gene is located on chromosome 17 and is translated into at least 12 protein isoforms, the study of which forms its own field within cancer biology (Khoury & Bourdon, 2010; Vieler & Sanyal, 2018). P53 levels are normally low in cells due to turnover involving the negative regulator MDM2 (Lahav *et al.*, 2004). However, upon activation via stress signals such as DNA damage, p53 is stabilised and can, as a dimer of dimers, act as a transcription factor to regulate genes involved in processes such as cell cycle arrest, senescence and apoptosis (Kitayner *et al.*, 2006; Vousden & Prives, 2009).

Considering all cancers, *TP53* is the most frequently mutated gene (Zehir *et al.*, 2017). About 56% of MSS stable colorectal cancers harbour mutations in the gene (Giannakis *et al.* 2016 via cBioPortal). Mutations occur across the entire length of the gene and can lead to loss of function as well as expression of gain-of-function mutant versions of the protein (Vogelstein, Lane & Levine, 2000; Vousden & Prives, 2009; Yue *et al.*, 2017).

This clonal mark is unique in that it is not loss but stabilisation of TP53 which is detected. IHC was performed on FFPE colonic sections from 124 individuals aged 8-90 years and scored for WPC and PPC with stabilised p53 (Figure 4.9A). Plotting of WPC frequencies revealed an age-related increase with a  $\Delta C_{fix}$  of  $3.59 \times 10^{-7}$  (95% CI:  $2.68 \times 10^{-7}$ – $4.88 \times 10^{-7}$ ) (Figure 4.9B), while the frequency of PPC stayed constant (Figure 4.9C). The ratio of  $\Delta C_{fix}/C_{part}$  was 0.073 (95% CI: 0.044–0.136) (Figure 4.9D), remarkably similar to mPAS, MAOA. The  $P_R$  was calculated to be 0.531 (95% CI: 0.444–0.664). This implied that stem cells with stabilised p53 behave neutrally in the crypt.



**Figure 4.9 Inference of stem cell behaviour associated with stabilisation of p53.**

(A) (i) WPC and (ii) PPC in en face tissue sections stained for p53.

(B) Regression analysis showing  $\Delta C_{\text{fix}}$  for p53 ( $1.96 \times 10^{-5}$  per year) with 95% margin of error (ME) shaded in grey.

(C) Frequency plot of WPC (circles) and PPC (squares) for STAG2 clones. Bottom panel shows PPC frequencies alone on condensed y-axis.

(D) Plot showing 10-fold higher ratio for  $\Delta C_{\text{fix}}/C_{\text{part}}$  for STAG2 as compared to mPAS and MAOA. Error bars = 95% CI.

#### 4.4.1.1 P53 stabilisation and HDAC6 loss are not correlated in human colon

A number of studies report regulation of p53 by HDAC6 (Ding *et al.*, 2013; Park *et al.*, 2017). To investigate whether such an effect was occurring in human colonic crypts, serial sections were stained for p53 and HDAC6. No stabilisation of p53 was observed in HDAC6-negative crypts nor were lower levels of HDAC6 observed in crypts with stabilised p53 (Figure 4.10, n = 10 sections each). This suggests that HDAC6 is not involved in stabilising p53 in human colonic epithelium.

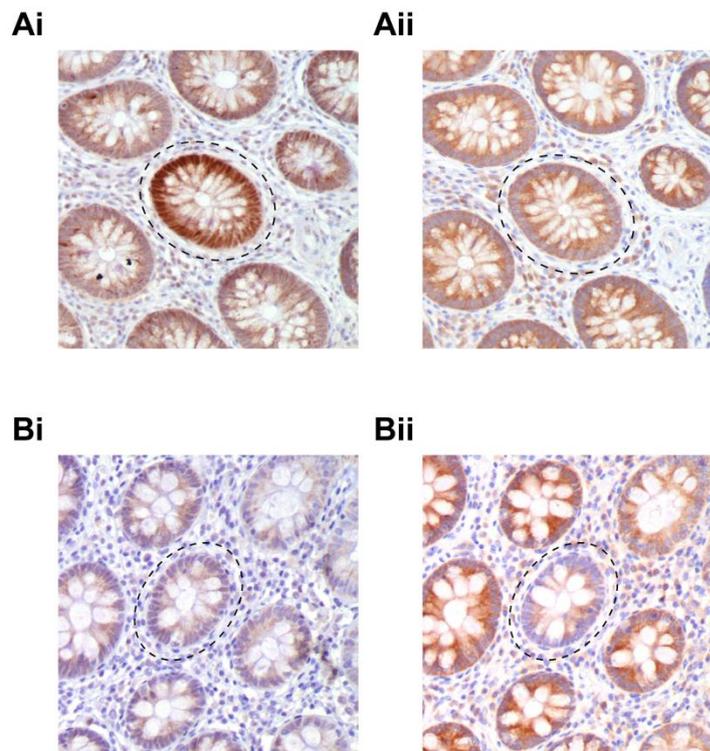


Figure 4.10 **P53 stabilisation and HDAC6 loss are not correlated in human colonic crypts.**

(A) Serial sections stained for (i) p53 and (ii) HDAC6. Crypt with stabilised p53 highlighted.  
(B) Serial sections stained for (i) p53 and (ii) HDAC6. HDAC6-negative crypt highlighted.

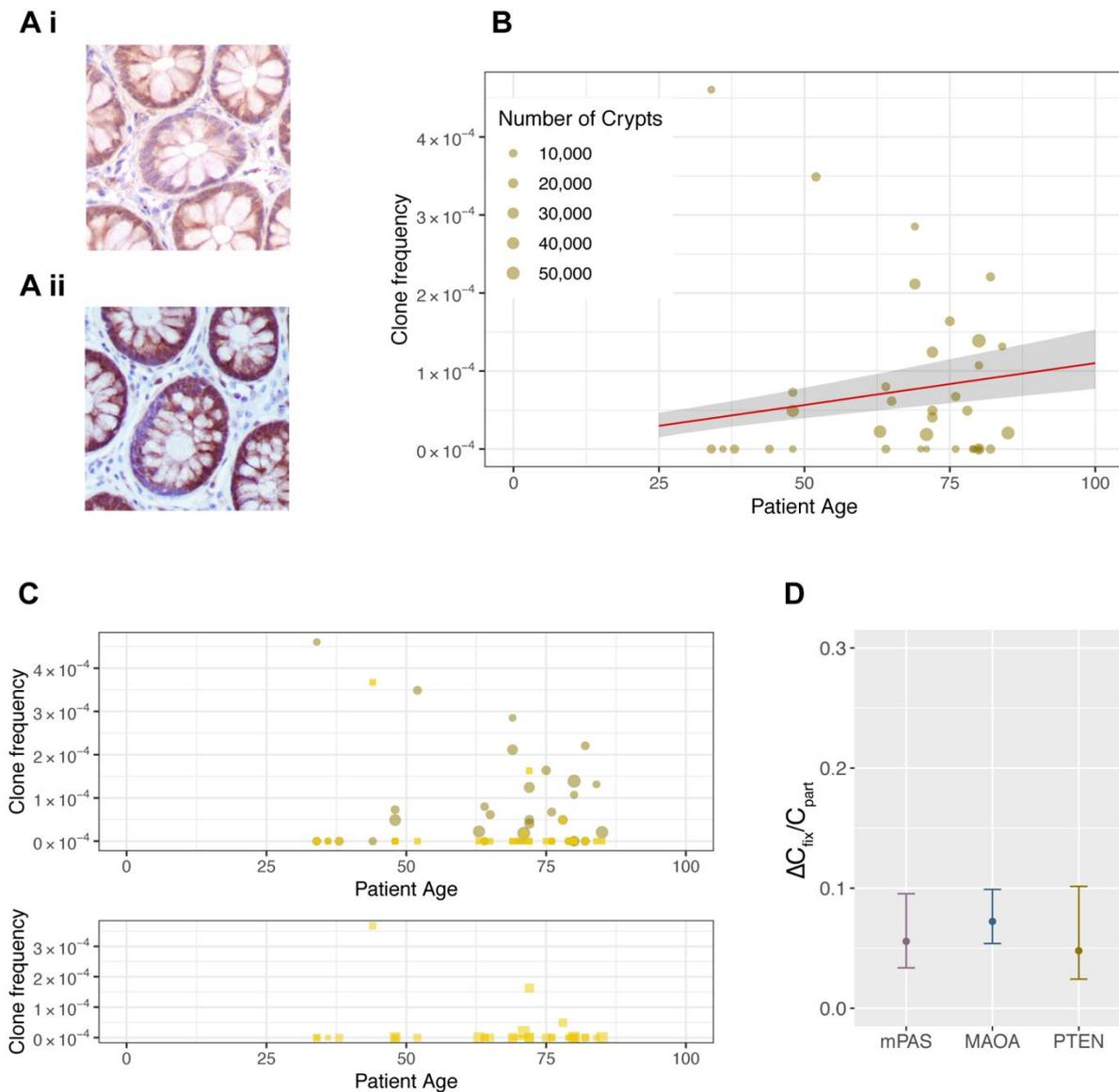
#### 4.4.2 PTEN loss is associated with neutral stem cell dynamics

Phosphatase and tensin homolog deleted on chromosome 10 (PTEN) is a phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase, catalysing the dephosphorylation of PIP3 into PIP2 (Maehama & Dixon, 1998). Therefore, the tumour suppressive role of PTEN is mainly linked to its role as negative regulator of the PI3K-AKT-mTOR signalling axis (Figure 3.7), controlling a number of processes including cell growth, cell cycle progression, and stem cell renewal (Hollander, Blumenthal & Dennis, 2011; Milella *et al.*, 2015; Manning & Toker, 2017; Papa & Pandolfi, 2019). However, independent functions including phosphatase activity that can bring about cell cycle arrest as well as nuclear functions have also been reported (Hlobilkova *et al.*, 2000; Milella *et al.*, 2015).

Germline mutations in *PTEN* are associated with a number of tumour predisposition syndromes collectively known as PTEN Hamartoma Tumours Syndromes. Affected individuals develop cellular overgrowths known as hamartomas in a broad range of tissues including breast, skin and brain (Yehia, Ngeow & Eng, 2019).

Looking broadly across all cancer types, *PTEN* is one of the 10 most frequently mutated genes (Zehir *et al.* 2017 via cBioPortal). As is common for tumour suppressor genes, mutations occur throughout the entire length of the protein. Interestingly, three hotspots can be seen, however these do not correlate with specific cancer types (COSMIC database). In addition to mutation, promoter hypermethylation has been reported to result in PTEN silencing in a number of cancers (Kang, Lee & Kim, 2002; García *et al.*, 2004; Ho *et al.*, 2009). Just over 5% of colorectal cancers harbour PTEN mutations (Giannakis *et al.* 2016 via cBioPortal). Therefore, the gene has been classified as a driver in this disease.

For determination of the effect of PTEN loss on human colonic stem cell dynamics, sections from 36 individuals ranging in age from 34–85 years were stained and WPC and PPC scored (Figure 4.11A). An age-related increase in WPC with a  $\Delta C_{fix}$  of  $1.06 \times 10^{-6}$  (95% CI:  $7.00 \times 10^{-7}$ – $1.61 \times 10^{-6}$ ) (Figure 4.11B) and constant  $C_{part}$  was observed (Figure 4.11C). For the ratio of  $\Delta C_{fix}/C_{part}$  a value of 0.047 (95% CI: 0.024–0.101) was obtained, similar to mPAS and MAOA (Figure 4.11D). Also,  $P_R$  was calculated to be 0.483 (95% CI: 0.397–0.600). This suggests that PTEN loss does not confer any bias to affected stem cells.



**Figure 4.11 Inference of stem cell behaviour associated with PTEN mutation.**

(A) (i) WPC and (ii) PPC in *en face* tissue sections stained for PTEN.

(B) Regression analysis showing  $\Delta C_{fix}$  for PTEN ( $1.06 \times 10^{-6}$  per year) with 95% margin of error (ME) shaded in grey.

(C) Frequency plot of WPC (circles) and PPC (squares) for PTEN<sup>-</sup> clones. Bottom panel shows PPC frequencies alone on condensed y-axis.

(D) Plot showing similar ratio of  $\Delta C_{fix} / C_{part}$  for mPAS, MAOA and PTEN. Error bars = 95% CI.

## 4.5 Comparison of all marks

In summary, staining of normal human FFPE colon sections from a large patient cohort followed by scoring of WPC and PPC enabled inference of stem cell behaviours associated with gene-specific mutations. A range of behaviours was identified, spanning from the negatively biased HDAC6 to strong advantage observed for loss of STAG2. The values obtained for  $\Delta C_{fix}/C_{part}$  range from 0.027 to 0.5, which is a nearly 20-fold difference between HDAC6 and STAG2 (Figure 4.12 and Table 4.1). These results show that, as in the mouse, mutations can change the behaviour of stem cells in the crypt to achieve higher mutant allele burden in the tissue than would arise from mutation alone.

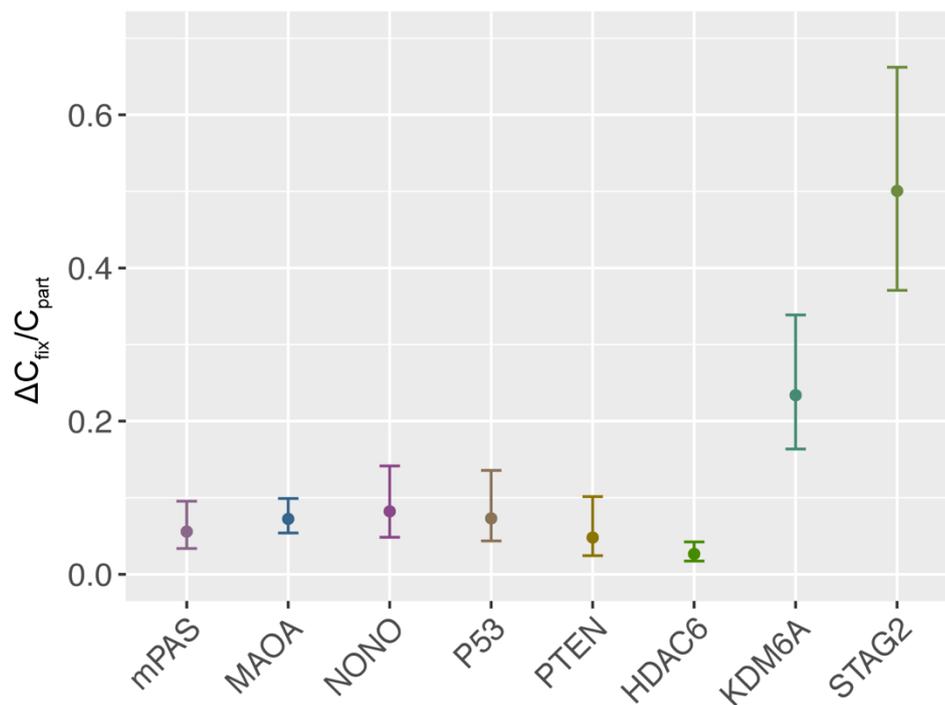


Figure 4.12 **Stem cell behaviour associated with gene-specific mutations.**  
Plot showing  $\Delta C_{fix}/C_{part}$  for all clonal marks. Error bars = 95% CI.

Table 4.1 **Summary of clonal marks data.**

*For completeness this includes the previously used neutral marks mPAS and MAOA.*

Mark	No. of patients	Ages	Mean age	$\Delta C_{fix}$	$\Delta C_{fix}/C_{part}$	P <sub>R</sub>
HDAC6	66	8-89	66	$3.19 \times 10^{-6}$	0.027	0.432
KDM6A	120	21-93	68	$6.04 \times 10^{-6}$	0.234	0.760
MAOA	152	8-93	70	$1.76 \times 10^{-6}$	0.072	0.531
mPAS	50	37-93	70	$5.84 \times 10^{-6}$	0.056	0.499
NONO	78	20-93	66	$1.44 \times 10^{-6}$	0.082	0.546
P53	124	8-90	66	$3.59 \times 10^{-7}$	0.073	0.531
PTEN	36	34-85	67	$1.06 \times 10^{-6}$	0.048	0.483
STAG2	110	8-93	67	$1.96 \times 10^{-5}$	0.501	0.999

## 4.6 Inference of mutation rates

For neutral marks, the frequency of PPC is proportional to the mutation rate. Therefore, the following holds true:

$$\frac{C_{part \ mark 1}}{mutation \ rate \ mark \ 1} = \frac{C_{part \ mark 2}}{mutation \ rate \ mark \ 2}$$

Which can be rearranged to:

$$mutation \ rate \ mark \ 2 = mutation \ rate \ mark \ 1 \times \frac{C_{part \ mark \ 2}}{C_{part \ mark \ 1}}$$

The *de novo* mutation rate for mPAS was previously derived from the ratio of single mPAS-positive cells/total goblet cells (see section 1.4.3.2 and Nicholson *et al.*, 2018). Therefore, using the known mPAS mutation rate of  $4.44 \times 10^{-6}$  mutations/mitosis, the mutation rates for the neutral marks MAOA, NONO, p53 and PTEN could be derived (Table 4.2). For non-neutral marks, the full equation including  $N_{crypt}$ , and  $P_R$  was used (Appendix C) (Table 4.2). Mutation rates are expected to scale with coding sequence length. Plotting of the values confirmed such a trend but revealed HDAC6 and KDM6A as outliers with higher and lower mutation rates than expected, respectively (Figure 4.13A). Focusing on the other marks, a trend towards scaling of mutation rate with coding sequence length can be observed (Figure 4.13B)

Table 4.2 **Inferred mutation rates for clonal marks.**  
*For comparison this includes the previously derived mPAS and MAOA.*

<b>Mark</b>	<b>Mutations per mitosis</b>	<b>95% Confidence interval</b>
HDAC6	$6.85 \times 10^{-6}$	$2.91 \times 10^{-6} - 1.64 \times 10^{-5}$
KDM6A	$9.28 \times 10^{-7}$	$4.29 \times 10^{-7} - 1.99 \times 10^{-6}$
MAOA	$1.03 \times 10^{-6}$	$5.13 \times 10^{-7} - 1.99 \times 10^{-6}$
mPAS	$4.44 \times 10^{-6}$	$2.47 \times 10^{-6} - 7.77 \times 10^{-6}$
NONO	$7.45 \times 10^{-7}$	$3.42 \times 10^{-7} - 1.59 \times 10^{-6}$
P53	$2.09 \times 10^{-7}$	$8.89 \times 10^{-8} - 4.48 \times 10^{-7}$
PTEN	$9.36 \times 10^{-7}$	$3.72 \times 10^{-7} - 2.16 \times 10^{-6}$
STAG2	$1.66 \times 10^{-6}$	$7.89 \times 10^{-7} - 3.34 \times 10^{-6}$

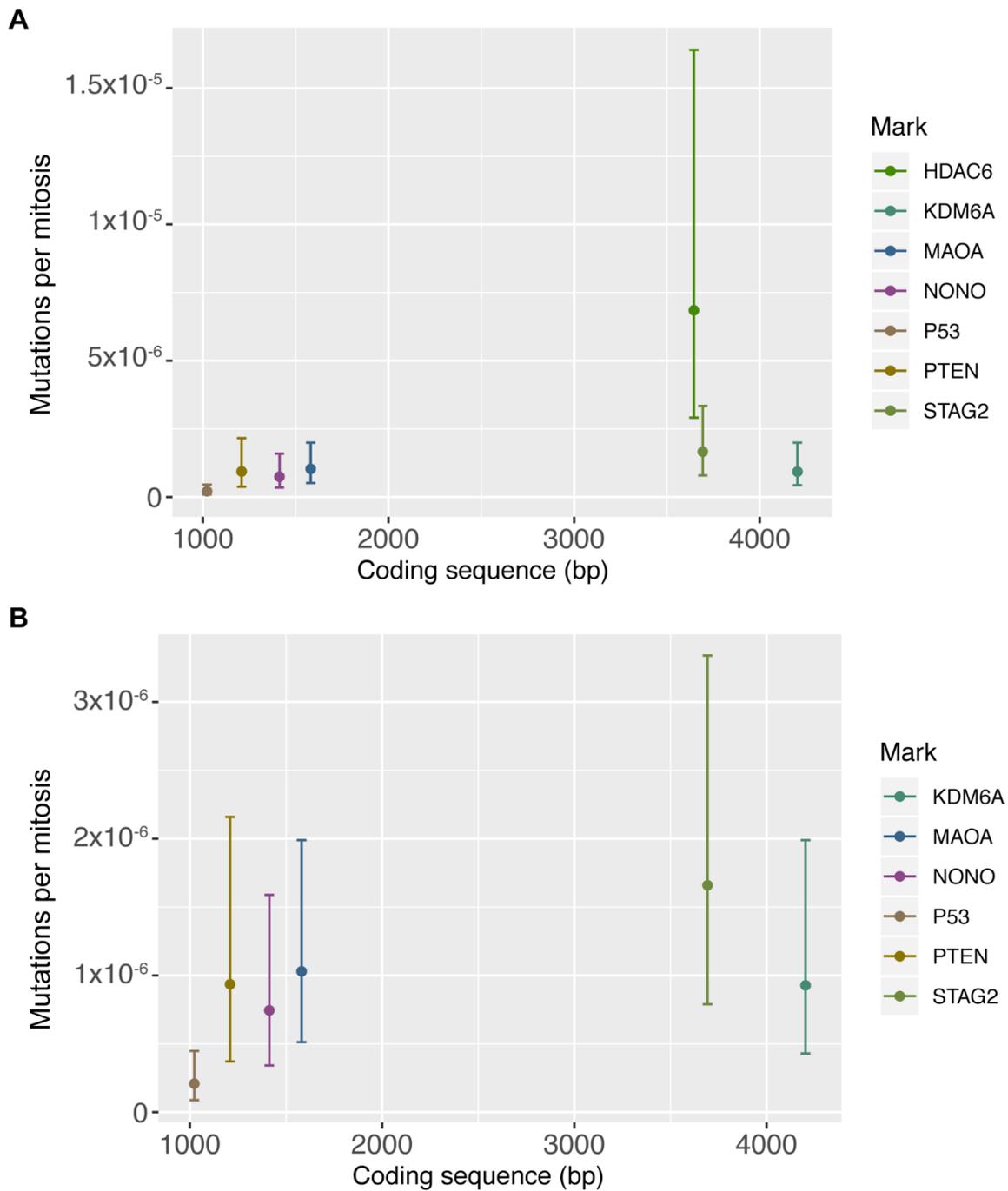


Figure 4.13 **Inferred mutation rates scale with coding sequence length.**

*MPAS is not included as the causative gene remains unknown. Error bars = 95% CI.*

(A) Mutation rates of clonal marks plotted against coding sequence.

(B) Same as (A) but with expanded X-axis, excluding HDAC6.

## 4.7 Discussion

The neutral clonal marks mPAS and MAOA enabled inference of human colonic stem cell dynamics (Nicholson *et al.*, 2018). Here, a novel set of clonal marks, validated in chapter 3, was used to describe a range of human intestinal stem cell behaviours associated with different gene-specific mutations. In a first step, this required staining of FFPE colonic sections from hundreds of individuals of a wide range of ages. For all new marks, WPC and PPC were observed. Scoring revealed an age-related increase in the frequency of WPC, but a constant frequency of the transition forms, PPC, for each clonal mark. The constant value of  $C_{part}$  observed across ages indicates that the rate of *de novo* mutations giving rise to new clones is balanced by loss of PPC either by their resolution to WPC or by clone extinction. The linear accumulation of WPC also indicates a constant process during life. These age-related trends are predicted by neutral drift theory and can only be observed for a given clonal mark if it arises as an irreversible and cell autonomous genetic event. Therefore, the observed age-related behaviours serve as validation for the genetic origin of the phenotypes observed for clonal marks. If the clones were caused by non-genetic events such as inflammation, the distribution of clones would most likely not adhere to this pattern. Furthermore, isolated PPC would not be expected. This is particularly relevant to the clonal detection of stabilised p53, where the observed phenotype could theoretically be caused by a plethora of mechanisms (see below).

An interesting observation was made when WPC frequencies for KDM6A were plotted. Contrary to the published notion that KDM6A escapes X-inactivation, no difference in clone frequency was found between males and females. This suggests that there may be a colon-specific regulatory mechanism at play. RNA-seq data from fresh tissue or human colonic organoids may distinguish mono- from biallelic expression, which could validate this observation.

Following scoring, regression analysis was used to determine the slope for WPC accumulation,  $\Delta C_{fix}$ , and the frequency of PPC,  $C_{part}$ . The ratio between these two values,  $\Delta C_{fix}/C_{part}$ , represents a measure for intra-crypt bias. Importantly, this calculation normalises for mutation rate and thus allows for direct comparison between marks. Figure 4.12 shows the range of values obtained for all novel clonal marks to the neutral marks mPAS and MAOA.

A first striking observation was that three new marks, NONO, p53 and PTEN all exhibited a  $\Delta C_{fix}/C_{part}$  that was remarkably similar to the marks previously described as neutral. Considering that all scoring is done blinded with regards to sample identity, the finding that three additional marks behave in the exact same way as mPAS and MAOA lends further support to the claim that these are neutral.

Loss of NONO has been found to associate with significantly increased tumour size in breast cancer (Traish *et al.*, 1997). A greater number of cancer studies report overexpression, with examples including malignant melanoma (Schiffner *et al.*, 2011), malignant pleural mesothelioma (Vavougiou *et al.*, 2015), breast cancer (Zhu *et al.*, 2015) and neuroblastoma (Liu *et al.*, 2014). Overexpression is also more common than mutation in colorectal cancer, with the COSMIC database reporting 33.28% (610 tested) versus 0.26% (2332 tested) of large intestine samples, respectively. A similar 0.8% mutation frequency was found in a different study (Giannakis *et al.* 2016 via cBioPortal). In light of these findings, the result that NONO loss does not affect stem cell dynamics in the crypt intuitively makes sense. It seems that the protein can exert pro-oncogenic effects when overexpressed, whereas NONO loss is only tumour-promoting in very specific contexts, such as the breast.

HDAC6 loss seems to confer a disadvantage to affected stem cells, such that they have a less than 50% chance of replacing their neighbour. The main alteration of HDAC6 in cancer reported by the COSMIC database is overexpression. As for colorectal cancer, HDAC6 is very rarely mutated (1.6% according to Giannakis *et al.*, via cBioPortal) while CNV gain has been found in about 15% of samples of a small cohort of 26 patients (TCGA data). Mechanistically, HDAC6 seems to be required for Ras-driven oncogenic transformation (Lee *et al.* 2008) in fibroblasts. It has also been reported to promote tumorigenesis via inhibition of p53 (Ding *et al.*, 2013; Park *et al.*, 2017) and to regulate immunosuppressive properties of cancer-associated fibroblasts in breast cancer (Li *et al.*, 2018). Of note, in our tissue p53 stabilisation was not observed in HDAC6-negative crypts. Furthermore, the intra-crypt behaviour of HDAC6-negative and p53-stabilised crypts was significantly different. Therefore, it seems that HDAC6 is not involved in regulation of p53 in human colonic crypts. In line with a potential pro-oncogenic role of HDAC6, selective HDAC6-inhibitors have been reported to slow down the

growth of cancer cell lines *in vitro* (Hideshima *et al.*, 2017; Zhang & Gan, 2017; Hye-Rim *et al.*, 2018) and a number of HDAC6 inhibitors are currently being evaluated in clinical trials or have been approved for the clinic by the FDA (Suraweera, O'Byrne & Richard, 2018). The finding described here, that HDAC6 loss causes a disadvantage within the crypt is consistent with these accounts of growth retardation following HDAC6 inhibition. Considering the effect of HDAC6 inhibition seen in cell lines, the intra-crypt disadvantage may be related to proliferation.

For KDM6A loss an advantage was found, with a  $P_R$  of 0.760. KDM6A is the most frequently mutated histone modifier in human cancer (van Haaften *et al.*, 2009). Mutation frequencies vary significantly between different cancer types. While the MSKCC-IMPACT study observed an average frequency of 3.5% across all cancer types (Cheng *et al.*, 2015), mutation frequencies of up to 40% have been reported in urothelial carcinoma. For human colon, the COSMIC database reports *KDM6A* point mutations in 2.66% of adenocarcinomas (35 out of 1318 tested). For all cancers, missense as well as truncating mutations have been found distributed along the entire length of the gene, indicative of tumour suppressive function of KDM6A. The mechanism of tumour suppression by KDM6A was first studied in T-cell Acute Lymphoblastic Leukaemia, whereby a study demonstrated the requirement for the demethylase catalytic function (Van der Meulen *et al.*, 2015). Loss of KDM6A may thus lead to an increase in H3K27me3 and therefore cause activation of genes that may drive cancer progression. Such a mechanism has been demonstrated in bladder cancer, where KDM6A loss was shown to lead to aberrant activation of PCR2-regulated transcription repression (Ler *et al.*, 2017). Importantly, re-introduction of wild-type KDM6A into mutant cancer cells markedly reduced proliferation, an effect that was accompanied by transcriptional changes (van Haaften *et al.*, 2009). However, IHC for H3K27me3 on serial sections did not reveal any increase in this modification in KDM6A-negative crypts. This may be due to compensation by other H3K27me3 demethylases such as JMJD3 (Agger *et al.*, 2007). Our finding is in line with several reports of loss of KDM6A causing no or at most a modest increase in global H3K27me3 levels (van Haaften *et al.*, 2009; Ler *et al.*, 2017; Gozdecka *et al.*, 2018). However, the absence of a negative or positive control significantly weakens this finding. Validation could be performed by comparing KDM6A<sup>-/-</sup> and tri- or demethylase inhibitor-treated organoids. More recently, a

study demonstrated the redundancy of the demethylase function of KDM6A for tumour suppression in murine acute myeloid leukaemia (AML). Instead, proteomic data indicated a role for protein-protein interactions with tetratricopeptide repeats, leading to activation of specific transcriptional programs. Supportive of this finding, the tumour suppressive role of KDM6A in AML is shared by the catalytically inactive Y-chromosomal homologue UTY (Gozdecka *et al.*, 2018). Therefore, it seems plausible that KDM6A loss can, independently of demethylase activity, lead to general chromatin remodelling which may drive the advantaged phenotype observed. Of note, LCM and sequencing identified mutations in the JmjC domain of KDM6A in 2 out of 11 KDM6A-negative patches (see section 3.5.2). It is thus conceivable that mutations in the JmjC domain shift KDM6A function towards recruitment of transcription factors via tetratricopeptide repeats, which then promotes a gene expression program driving intra-crypt advantage. This hypothesis could be tested by CRISPR/Cas9 in human cell lines or organoids followed by RNA-Seq and later ChIP-Seq.

In the present study, the strongest intra-crypt advantage was associated with STAG2 loss. *STAG2* is the most commonly mutated of the cohesin complex members. Mutations are most common in urothelial carcinoma (16%) (Cosmic database). To a large extent these are truncating mutations, consistent with a role as a tumour suppressor (Hill, Kim & Waldman, 2016). In colorectal cancer, *STAG2* mutations are relatively frequent, featuring in about 4.2% of adenocarcinomas (Giannakis *et al.* 2016 via cBioPortal). The consequence of STAG2 loss of function is still a matter of debate. In line with its role in sister chromatid cohesion, it was initially associated with aneuploidy (Solomon *et al.*, 2011), which can drive cancer progression and could explain the intra-crypt bias observed. However, a number of studies have since challenged this finding (Balbás-Martínez *et al.*, 2013; Kim *et al.*, 2016). A different cancer-promoting mechanism of STAG2 loss may involve genomic rearrangements caused by prolonged association of telomeric repeats during the cell cycle (Daniloski & Smith, 2017). If such a mechanism was at play, the bias observed with STAG2 may be mediated by secondary mutations arising from genetic instability. However, Dr Edward Morrissey performed modelling of such a scenario, which revealed that this would require the rate of subsequent mutation to be increased by an order of  $10^5$  (see Appendix C for mathematical model). Such an

extreme rate is unlikely to occur and it is thus likely that the clone dynamics described here arise directly from STAG2 loss. However, speculation about how STAG2 loss confers stem cells with an advantage is currently uncertain. Very generally, as it normally performs a tumour suppressive role, it makes sense that loss would lead to a pro-oncogenic effect.

It is worth discussing the inferred  $P_R$  of 0.99 for STAG2-deficient stem cells. This is strikingly high, effectively corresponding to certain fixation of mutant stem cell clones within the crypt.  $P_R$  was calculated from the full equation containing  $N_{crypt} = 7$  and  $\lambda_{crypt} = 1.3$  replacements/crypt/year. However, the accelerated fixation of STAG2-negative clones may in theory also result from an increased replacement rate. To test this hypothesis, the size distribution of partial clones (as fractions of the entire crypt) could be analysed. As PPC are underrepresented for advantaged marks, the current dataset would have to be expanded to allow for statistically powered modelling.

Most intriguing was the discovery that both p53 stabilisation and PTEN loss do not appear to confer any bias to affected stem cells. As these are both CRC drivers, an advantage might have been predicted for these marks. P53 can be stabilised by a variety of mechanisms such as DNA damage, hypoxia and oxidative stress. Further, mutant p53 proteins have been reported to be stabilised by a variety of mechanisms including interaction with Hsp90 and BAG protein family members, short isoforms of MDM2 and post-translational modification (Li, Marchenko & Moll, 2011; Zheng *et al.*, 2013; Nguyen *et al.*, 2014; Yue *et al.*, 2016, 2017). The origin of this clonal mark may thus be complex, which complicates discussion of the neutral dynamics observed. LCM of individual crypts could, in theory, identify the mutations causing stabilisation. Importantly, these may not be present in p53 but in one of its interaction partners. Due to the poor quality and yield of DNA obtained from FFPE sections, broad coverage by sequencing of LCM material is currently not feasible. With the advent of in-situ proteomics, investigation of potential effects of post-translational modifications of p53 may become accessible in the future. Therefore, current potential explanations for neutral dynamics for stabilised p53 can be based on two lines of evidence: Firstly, the mutation rate of *TP53* in metastases of colorectal cancer is ~80% (Brannon *et al.*, 2014), while they are rarer in benign colonic adenomatous polyps, with mutation frequencies of 15-30% reported (Baker *et al.*, 1990; Hao *et al.*, 2002; Borrás *et al.*, 2016). This indicates that *TP53* mutations may play a more

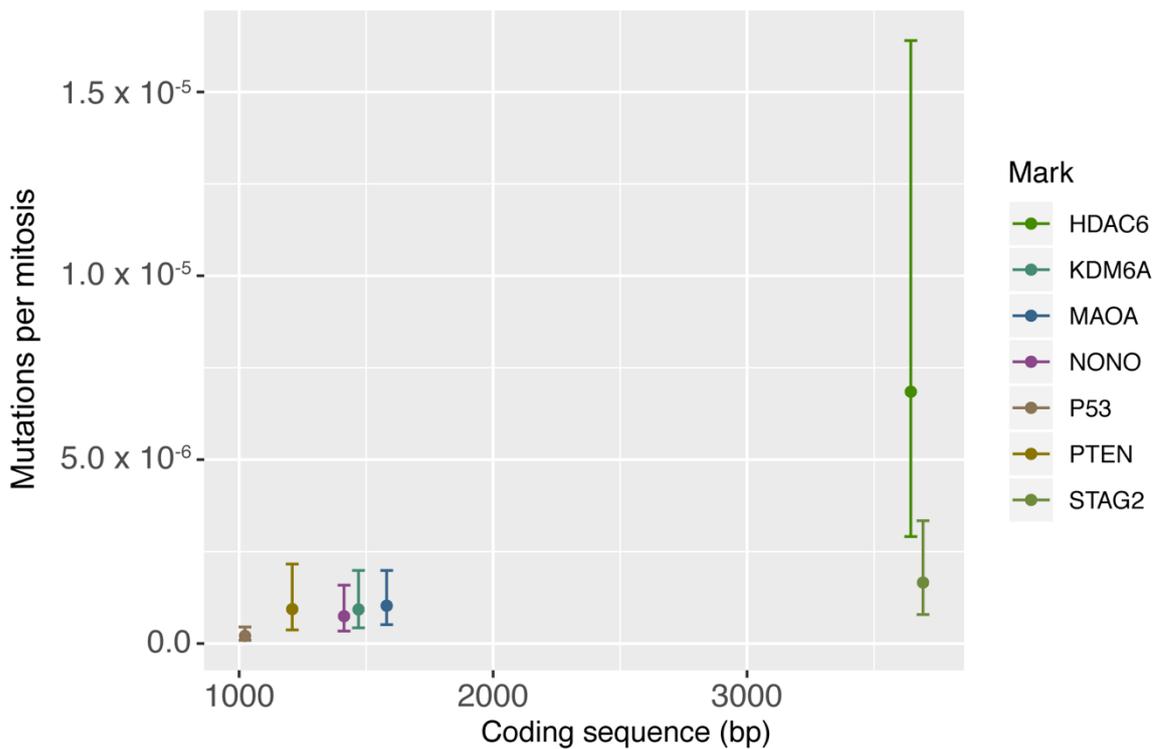
prominent role during later stages of colorectal tumourigenesis, rather than during initiation. Secondly, in the mouse the dominant negative *Tp53* mutation R172H did not confer any bias to affected stem cells. Interestingly however, a  $P_R$  of 0.58, indicating a mild advantage, was found when *Tp53* mutation was induced in a chronic colitis model (Vermeulen *et al.*, 2013). This suggests that the tissue environment may influence the behaviour of p53 mutant clones. It would thus be informative to perform a human p53 study in samples from a cohort of inflammatory bowel disease (IBD) patients.

Due to its tumour suppressive role as a negative regulator of PI3K-AKT-mTOR signalling, loss of PTEN might have been predicted to cause an advantage, potentially via increased proliferation. Indeed, a number of studies report increased stem cell proliferation, however up to date no such effect has been described in the colon (Milella *et al.*, 2015). An important aspect of PTEN biology is dosage. Partial loss can be sufficient for a pro-oncogenic effect (Alimonti *et al.*, 2010). Alternatively, PTEN loss in the absence of further mutations may not cause a strong enough effect on signalling to generate a bias in stem cell dynamics.

In summary, the observed effects of gene-specific mutations on stem cell dynamics can mostly be rationalised by examining mutation frequencies in cancers, putative cancer functions and, in the case of p53, mouse studies. However, these considerations fail to explain the behaviours on a mechanistic level. For such insights, use of the CRISPR/Cas9 system in human colonic organoids combined with BrdU labelling, RNA-seq and proteomics could highlight pathways of interest. Genetically modified organoids could then be tested *in vivo* by xenografting into mice, as recently published by the Sato group (Sugimoto *et al.*, 2018). Alternatively, inducible knock-out mice could be crossed to reporter lines, and used for lineage tracing experiments. These could be followed up with isolation of labelled crypts, RNA-Seq and proteomics.

The data acquired here enabled derivation of mutation rates associated with all clonal marks. Plotting the values revealed a linear trend between coding sequence and mutation rate with two outliers, KDM6A and HDAC6, with markedly lower or higher than expected mutation rates, respectively. Importantly, the mutation rates were calculated from the frequency of clones observed following IHC with an antibody. For most clonal marks, the epitope is C-terminal,

however, for KDM6A this is not the case, with the antibody raised against a peptide around amino acid residue 490 out of 1401. Therefore, it is possible that a number of C-terminal mutations are not detected by IHC, leading to an underestimation of mutation rate. When the coding sequence of KDM6A is shifted to correspond to 490 amino acids, the inferred mutation rate matches that of similarly sized other genes (Figure 4.14). For HDAC6, a potential explanation for the high mutation rate lays in gene structure: Using the open access software RepeatMasker (Smit, Hubley & Green, no date), it was found that *HDAC6* harbours a significantly higher GC-content (50%) than the similarly sized *STAG2* and *KDM6A* (both 39%). GC content has recently been found associated with higher mutation rates in yeast (Kiktev *et al.*, 2018) and may play a role in elevating the mutation rate of *HDAC6*.



**Figure 4.14 Adjusting the coding sequence length of KDM6A may rationalise the inferred mutation rate.**

*Plot identical to 4.13A but with coding sequence of KDM6A altered to reflect location of the epitope recognised by the antibody.*

The findings presented in this chapter significantly impact our understanding of the earliest events during tumour initiation in human colonic crypts. Gene-specific mutations can subvert intra-crypt dynamics to achieve an increased probability of fixation, which elevates their allele frequency in the tissue and increases the probability of second mutation acquisition. It would be interesting to examine whether there are upper limits to advantage and disadvantage within the crypt. For this purpose, more clonal marks would have to be developed. Defining severe disadvantage may be complicated by the difficulty of detecting PPC clones consisting of only a few cells present at low numbers. Extreme advantage, on the other hand, could dramatically reduce the frequency of PPC detected, precluding calculation of  $\Delta C_{fix}/C_{part}$  (as division by zero is not meaningful). A method for lineage tracing in human colonic tissue *in vitro* would aid in overcoming these challenges by enabling capture and therefore quantification of PPC directly after induction, before clone extinction or fixation.

## Chapter 5 Expansion beyond the crypt

### 5.1 Introduction

Due to the dynamics of neutral drift, the fate of neutral somatic mutations arising in stem cells of colonic crypts is stochastic. The probability of fixation of a neutral alteration is thus proportional to the number of stem cells per crypt. However, as indicated in chapter 4, certain gene-specific mutations can subvert intra-crypt dynamics to increase or decrease their probability of fixation. As mutations continuously arise in stem cells throughout life, fixed clones accumulate in the colonic epithelium in a linear, age-related manner. By altering their fixation probability, biased mutations can achieve considerably lower or higher mutant allele burden with time.

As discussed in chapter 1, CRC is thought to initiate in stem cells (Barker *et al.*, 2008; Schwitalla *et al.*, 2013), driven by accumulation of mutations (Fearon & Vogelstein, 1990). According to the current model, a tumour arises when a critical threshold of factors is exceeded in a cell, rather than through sequential sweeps by clones of increasing aggressiveness (Sottoriva *et al.*, 2015). Since several alterations are required for transformation, expansion of clones with a primary mutation increases the population of cells available for acquisition of further hits.

Indeed, a number of studies support the presence of such ‘field cancerisation’ in the human colon. Earliest evidence stems from detection of *KRAS* mutations in tumour adjacent normal samples (Zhu *et al.*, 1997). Further, the discovery of polyclonal tumours suggests the presence of a pre-cancerous field that facilitates adenoma formation (Thirlwell *et al.*, 2010). Also, in a study of synchronous tumours (cases where two or more primary tumours arise in the same individual), methylation patterns were found to be remarkably similar, indicative of an underlying field (Galandiuk *et al.*, 2012).

To generate mutant fields, clones need to spread in the epithelium. Considering the constraints imposed by the architecture of the colon, the most obvious process for this is crypt division.

### **5.1.1 Crypt fission**

Crypt division, usually referred to as fission, drives growth of the intestine during postnatal development (Clarke, 1972; Cummins *et al.*, 2008). Similar mechanisms are conserved across many adult epithelia, such as the lung, kidney and mammary epithelium, which all arise from branching of epithelial tubes (Affolter *et al.*, 2003).

#### **5.1.1.1 Early studies**

The first reports of intestinal crypt fission stemmed from observation of so-called ‘bifurcating crypts’ in mouse and rat intestines. A series of seminal papers published in the 1970s and 1980s described that crypt fission drives the massive increase in crypt numbers that occurs during postnatal growth as well as the recovery from damage by irradiation, chemotherapy and surgical resection (Clarke, 1972; Cairnie & Millen, 1975; Maskens, 1978; Maskens & Duhardin-Loits, 1981; Wright & Al-Nafussi, 1982; Cheng, McCulloch & Bjerknes, 1986). It was also found that while crypt fission is highest in young animals, it continues in adulthood, albeit at a much lower rate (Clair & Osborne, 1986). At the same time, the presence of bifurcating crypts, indicative of low levels of crypt fission, was also reported in adult human colon samples (Cheng *et al.*, 1986). This idea led to the development of the so-called ‘crypt cycle’ model, whereby fission occurs when a crypt reaches a certain threshold size. The authors proposed that such independent crypt cycling caused a slow but steady increase of intestinal size with age. According to their calculations, this would lead to a doubling of crypt numbers every 10-20 years (the cycle time), which they accommodate by proposing increased folding of the mucosa (Cheng *et al.*, 1986; Totafurno, Bjerknes & Cheng, 1987). Importantly, the human model was largely based on mouse data. In addition, inference of the rate of fission from the frequency of bifurcating crypts in static images requires an assumption about duration of the process, which was not available. These factors led to an overestimation of the human crypt fission rate.

### 5.1.1.2 Epigenetic and histochemical analysis

More recently, human crypt fission was investigated by analysis of methylation patterns. *De novo* methylation of CpG rich regions of the genome, so-called CpG islands, is somatically inherited and stochastically acquired at mitosis, leading to age-related increase in number of methylated loci (Ahuja *et al.*, 1998). It follows that comparison of the methylation patterns between crypts may enable derivation of clonal relationships, whereby crypts originating from fission events may display similar methylation patterns while unrelated crypts are, due to the stochastic nature of methylation acquisition, very unlikely to do so. An initial analysis of CpG islands in three genes (*MYOD*, *CSX* and *BGN*) confirmed the previously reported age-related increase in methylation and revealed that methylation patterns varied between crypts (Yatabe, Tavaré & Shibata, 2001). A subsequent study of adjacent crypts also found different methylation patterns, suggesting that crypt fission is rare in human colon (Kim & Shibata, 2002, 2004). Importantly, the technical challenges associated with analysis of methylation patterns restrict the sample size. In the absence of a priori knowledge about crypt relatedness it is not surprising that sampling of low numbers of crypts failed to identify evidence for fission.

This caveat was overcome by combination of (epi)genetic analysis with clonal marking using histochemical detection of loss of CCO activity (see 1.4.2.1). CCO-negative multicrypt patches were shown to be clonal by LCM and sequencing and their size was found to increase with age (Greaves *et al.*, 2006). Surprisingly, the methylation patterns of adjacent CCO<sup>-</sup> crypts were not closely related. Subsequent LCM and methylation pattern analysis followed by mathematical modelling indicated that the dynamic nature of methylation acquisition limits the time window during which clonally related crypts can be identified to 10 years (Graham *et al.*, 2011). The results obtained by Greaves and colleagues can thus be explained by the fission event taking place more than 10 years prior to sample collection. Using methylation patterns is thus likely to generate underestimates of crypt fission rates.

A more potent method was subsequently applied by Baker and colleagues, whereby the human crypt fission rate was derived from mathematical modelling based on the size distribution of CCO<sup>-</sup> patches in a cohort of 20 patients of the age range 42-85 years. The authors calculated a human crypt fission rate of 0.028 ( $\pm$  0.007) divisions per crypt per year, which equals

approximately 3% of crypts undergoing fission per year and a crypt cycle time of about 36 years (Baker *et al.*, 2014). The caveat of this study is the small patient cohort and low numbers of crypts (approximately 137,000 in total) analysed, which may have led to an over- or underestimation. Therefore, a reliable benchmark was to date still lacking.

### 5.1.1.3 Altered crypt fission rates

Increased crypt fission has been found associated with a number of inflammatory conditions such as ulcerative colitis (Cheng *et al.*, 1986). It is also now seen as the main driver of human adenoma growth (Preston *et al.*, 2003), a notion supported by dramatically increased crypt fission rates in patients with FAP (Araki *et al.*, 1995; Campbell *et al.*, 1998; Wasan *et al.*, 1998; Wong *et al.*, 2002; Baker *et al.*, 2014). In the mouse, *KRAS* activating mutations have been found to cause a dramatic up to 30-fold increase in crypt fission rate, therefore driving field cancerisation in the intestinal epithelium (Snippert *et al.*, 2013).

### 5.1.2 Aims

Based on the caveats associated with previous studies, a precise quantification of the homeostatic human colonic crypt fission rate was required. This benchmark would then enable investigation of how gene-specific mutations may subvert this process to achieve field cancerisation in the human colonic epithelium. The specific aims of this part of the present study were thus to:

- Quantify the homeostatic human colonic crypt fission rate.
- Investigate the effect of gene-specific mutations on crypt fission rates.
- Investigate the expansion of *KRAS*-mutant clones in normal human colonic epithelium.

### 5.1.3 Published work

The data we obtained for the fission rates of mPAS and MAOA, my data on fission of STAG2-deficient clones as well as *KRAS* targeted sequencing I performed on samples from 126

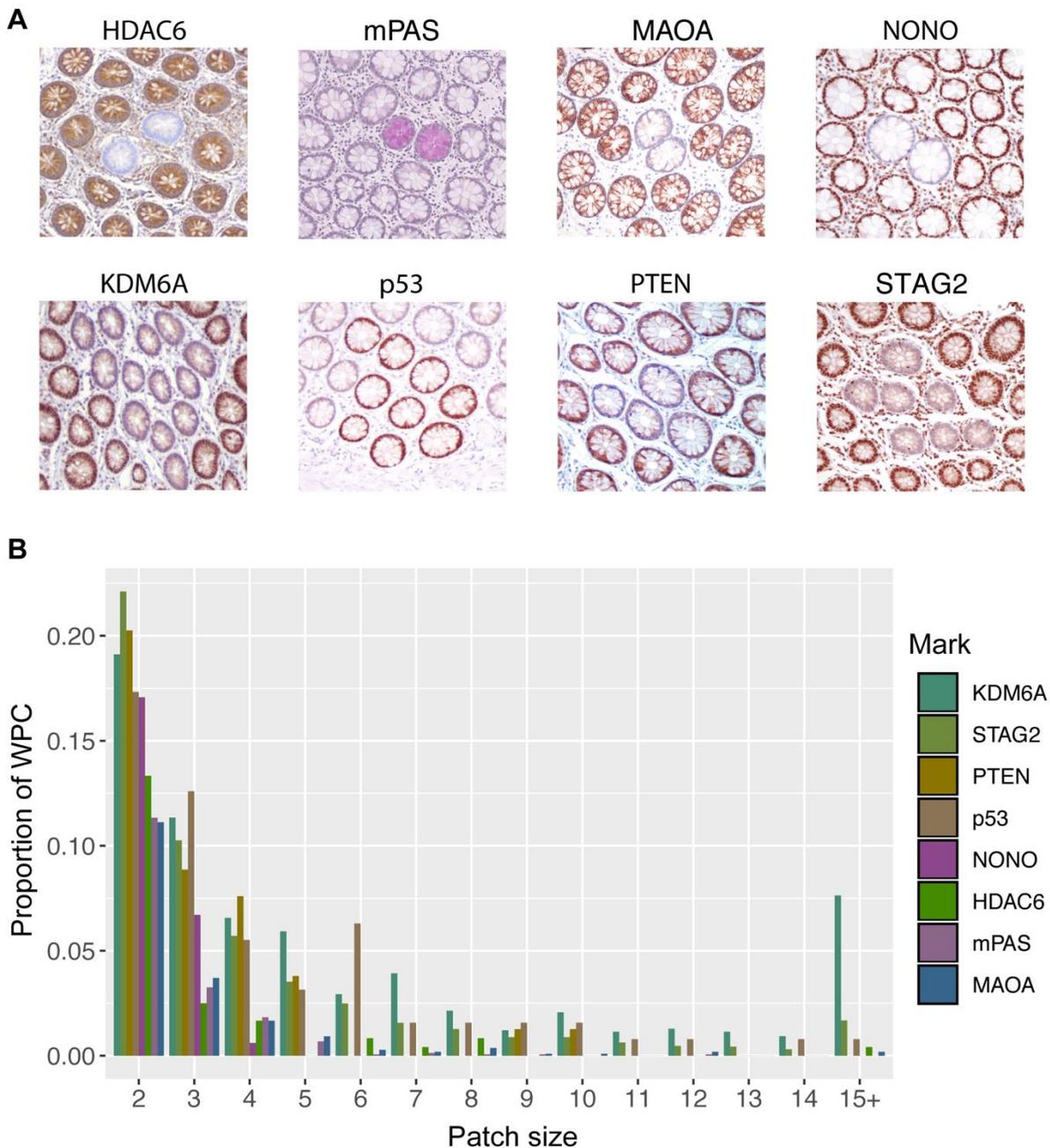
individuals were included in our recent publication (Nicholson *et al.*, 2018), on which I am second author.

## 5.2 Quantification of crypt fission using clonal marks

For all clonal marks patches of mutant epithelium comprising two or more adjacent crypts were observed (Figure 5.1A). For mPAS, MAOA, NONO and HDAC6, patches were mostly composed of two crypts. Larger patches were commonly observed for p53, PTEN, STAG2 and KDM6A. For the latter, patches were commonly composed of more than 15 crypts (Figure 5.1B) with the largest detected patch consisting of 71 crypts. Due to the low frequency of clones observed for all marks, it was assumed that patches arise from the amplification of a single variant crypt by fission. This is also supported by the LCM and sequencing described in chapter 3, whereby no negative patch was found to carry more than one mutation in the causative gene.

### 5.2.1 Gene-specific mutations can promote crypt fission

Combined with the known age of each patient (for patient numbers for each mark refer to Table 4.1), the size distribution of multicrypt patches for the different clonal marks enables modelling of crypt fission rates. This was performed by our collaborator Dr Edward Morrissey and the mathematical methods are described in detail in Appendix C. Importantly, the model used is probabilistic, not deterministic. Initially, once a mutation is fixed within a crypt, this crypt must divide for a patch of size two to arise. Subsequently, both of the two crypts can undergo fission to further increase the patch size. In a deterministic model, both crypts would fission after a certain interval, as every crypt has a deterministic rate at which it fissions. In such a model, the time it takes to expand a crypt into two is always the same. In a stochastic model, however, the time it takes for any given crypt to fission is drawn from a probability distribution governed by the fission rate. This means that the two crypts in a patch of two would most likely not undergo further fission at the same time, whereas in a deterministic model they must.



**Figure 5.1 Detection of multicrypt patches for clonal marks.**

(A) Representative images of multicrypt patches for all clonal marks.

(B) Grouped histogram showing frequency of patch sizes for all marks as proportion of total WPC. 15+ denotes patches composed of 15 crypts or more, combined.

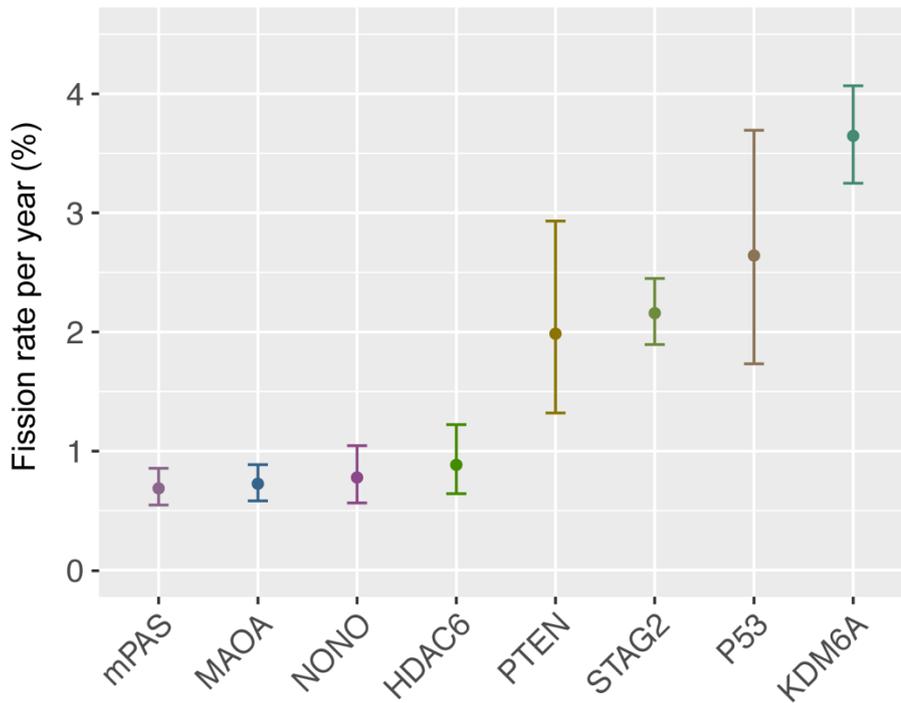
Application of the model to data from all eight clonal marks revealed a spectrum of behaviours.

Four marks, HDAC6, mPAS, MAOA and NONO exhibit very similar fission rates ranging from

0.69% to 0.88% per year (Table 5.1 and Figure 5.2). This was taken to be the homeostatic human colonic crypt fission rate, meaning that about 0.7% of normal crypts undergo fission every year. Importantly, the rate calculated here is very similar to a recent report using the mitochondrial marker CCO in a small cohort of 21 patients, which inferred a rate of 1.1% (Baker *et al.*, 2019). For PTEN, STAG2, p53 and KDM6A, fission rates were found to be 2.0%, 2.2%, 2.6% and 3.7% per year, corresponding to about a 3-fold increase for PTEN, STAG2 and p53 and about a 4.5-fold increase for KDM6A compared to baseline, respectively (Table 5.1 and Figure 5.2). Therefore, as well as conferring an advantage to stem cells within the crypt, mutations can promote lateral expansion to generate large patches within the epithelium.

Table 5.1 **Crypt fission rates associated with clonal marks.**

<b>Mark</b>	<b>Fission rate per year</b>	<b>95% Credible interval</b>
HDAC6	0.88%	0.64 - 1.22
KDM6A	3.65%	3.25 - 4.07
MAOA	0.73%	0.59 - 0.89
mPAS	0.69%	0.55 - 0.86
NONO	0.78%	0.56 - 1.04
P53	2.64%	1.73 - 3.70
PTEN	1.99%	1.32 - 2.93
STAG2	2.16%	1.89 - 2.45



**Figure 5.2 Gene-specific mutations can increase crypt fission rates.**

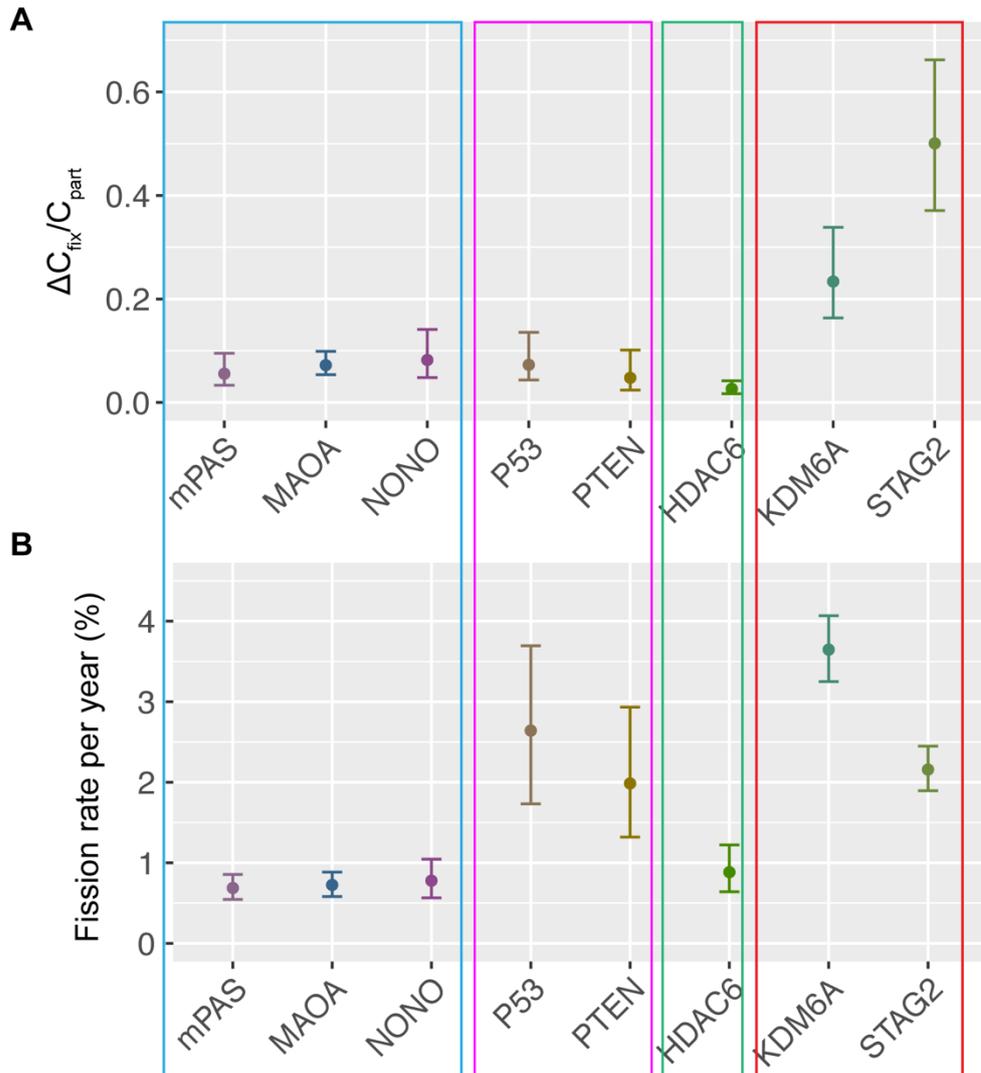
*Plot showing inferred crypt fission rates for all clonal marks in ascending order. Error bars = 95% CI.*

### 5.2.2 Intra-crypt dynamics and crypt fission may be independent

Juxtaposition of the intra-crypt behaviours ( $\Delta C_{fix}/C_{part}$ , Figure 4.10) and the fission rates inferred here (Figure 5.2) revealed a number of patterns for the different clonal marks, some of which were unexpected. There are three marks, mPAS, MAOA and NONO, which seem to be neutral markers for intra-crypt dynamics as well as fission. STAG2 and KDM6A are also consistent in their pattern, in that loss of these genes confers an advantage within the crypt and beyond. HDAC6 loss however, while conferring a disadvantage within the crypt, shows neutral behaviour for fission. P53 stabilisation and PTEN loss on the other hand do not seem to impact on crypt dynamics, but promote fission (Figure 5.3).

The total mutational burden of the human colonic epithelium thus seems to be a result of three independent processes. The first is mutation rate, which dictates the appearance of stem cell

clones. The second is fixation within individual crypts. Under neutral conditions this occurs via neutral drift. However, certain gene-specific mutations can cause a bias, therefore changing the likelihood of fixation. Finally, mutations can spread by crypt fission, which can be accelerated by loss of specific proteins.



**Figure 5.3 Intra-crypt dynamics and crypt fission may be independent processes.**

*Similar behaviour indicated by coloured box. mPAS, MAOA and NONO seem to be completely neutral (blue box), p53 stabilisation and PTEN loss confer a fission bias (pink box), HDAC6 loss confers intra-crypt disadvantage (green box) and loss of KDM6A and STAG2 lead to advantage within the crypt as well as fission acceleration (red box).*

(A) Graph showing  $\Delta C_{\text{fix}}/C_{\text{part}}$  for all marks, equivalent to Figure 4.10

(B) Graph showing fission rates associated with all marks, equivalent to Figure 5.2.

The contributions of these different processes were modelled by Dr Edward Morrissey (Figure 5.4). This shows that an intra-crypt bias alone (increasing  $P_R$  from 0.5 to 0.99) increases the burden of a variant 7-fold. Equally, elevating the fission rate alone can cause a 2-fold increase in mutation burden. Combining the two processes results in a 14-fold increase by the age of 60 (Figure 5.4).

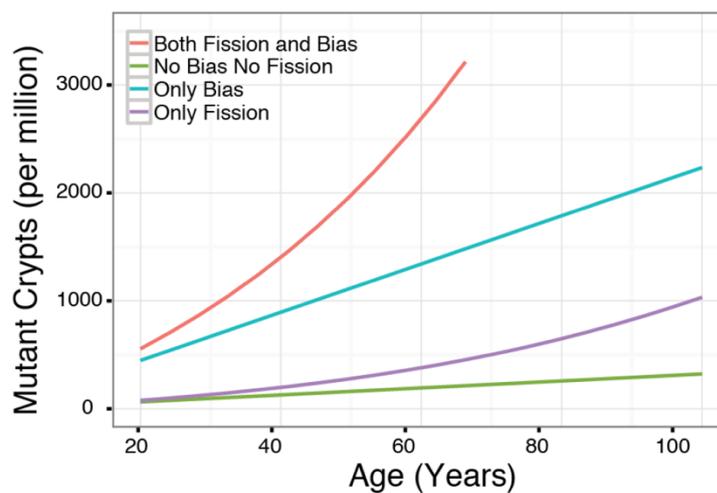


Figure 5.4 **Effects of intra-crypt bias and fission bias on total mutation burden.**

*Simulated number of mutant crypts plotted against age for genes with a common mutation rate of  $2 \times 10^{-6}$  per mitosis. Green = neutral outcome, purple = intra-crypt bias with  $P_R$  elevated to 0.99, blue = 3-fold increase in fission rate, red = combined effect of  $P_R = 0.99$  and 3-fold increase in fission rate, equivalent to *STAG2* mutation.*

### 5.3 *KRAS* mutations in normal human colonic epithelium

Mutations that promote fission accelerate field cancerisation. Indeed, the results described in section 5.2 demonstrate that two alterations commonly found in CRC, p53 stabilisation as well as loss of PTEN promote crypt fission, therefore generating mutant patches in the colonic epithelium. A key driver gene frequently mutated in CRC is *KRAS*. Activating mutations most commonly occur in codons 12 and 13, which harbour the sequences GGT and GGC,

respectively. Specifically, mutations in the second amino acid positions of these two codons, which include the most common KRAS activating mutations G12D (caused by a G>A transversion at position 2), G12V (a G>T transversion at position 2) as well as G13D (a G>A transversion at position 2), can be detected in just over 30% of colorectal cancers (COSMIC data). In fact, the G12D activating mutation is detected in just over 11% of colorectal tumours, closely followed by G12V with 9%, making them the most common specific mutations in CRC. Interestingly, oncogene activation by single nucleotide mutation is several orders of magnitude less frequent than the loss-of-function mutations observed for clonal marks. The G>A transversion responsible for the *KRAS*G12D mutation occurs at a frequency of about  $4 \times 10^{-8}$  per year (Tomasetti, Vogelstein & Parmigiani, 2013). Despite this seemingly low rate, high MAFs have been described in the normal mucosa of individuals not known to have cancer (Dieterle *et al.*, 2004; Kraus *et al.*, 2006; Parsons *et al.*, 2010). *KRAS* activating mutations have also been reported in the adjacent normal epithelium of *KRAS*-mutant tumours, suggesting the presence of an underlying *KRAS*-mutant field (Zhu *et al.*, 1997). In light of the fission bias observed for a number of clonal marks, as well as the finding that *KRAS* activation promotes crypt fission in the mouse (Snippert *et al.*, 2013), an obvious next step was to quantify the behaviour of *KRAS*-mutant clones in the human colonic epithelium. As discussed in chapter 3, antibody-based detection of *KRAS*G12D mutations did not yield satisfactory results. Hence a next-generation sequencing-based approach was taken.

### 5.3.1 A targeted amplicon sequencing approach for *KRAS* exon 2

For identification of *KRAS* activating mutations in human FFPE colonic sections by targeted sequencing, DNA was extracted from FFPE tissue sections, the genomic area of interest amplified by PCR and then sequenced. Characterisation of DNA extracted from FFPE material described in chapter 3 revealed that short (approx. 150 bp) amplicon length is key to successful PCR amplification. However, if a *KRAS* activating mutation occurred in codons 12 and 13 located in exon 2, then one single PCR amplicon will be sufficient to cover the area of interest of 6 base pairs. Importantly, the nature of the DNA material and sequencing method harbour a number of sources of potential false positives that merit consideration.

### 5.3.1.1 Sources of artefactual MAFs in DNA from FFPE material

A well described source of ‘artefactual mutations’ stems from hydrolytic deamination of cytosine residues to uracil. In living cells these are removed by uracil-DNA glycosylase and then restored by base excision repair. In fixed tissue however, they remain unrepaired. PCR amplification of such templates generates artefactual C:G>T:A SNVs as polymerase incorporates an adenine opposite to uracil lesions. Another common effect is depurination, which generates abasic sites at guanine and adenine residues. Polymerases can incorporate adenines opposite of such sites, generating artefactual G:C>T:A and A:T> T:A SNVs (Do & Dobrovic, 2015). During sequencing, these processes will produce a number of artefactual ‘mutant’ reads. Another source of error is PCR amplification. Before sequencing, DNA is amplified by PCR twice, first to generate the amplicons and then to add sample-specific barcodes. These amplifications can introduce errors that will be sequenced and be falsely called as mutations. Finally, the sequencing process in itself occurs by synthesis, which again may introduce errors. The rate of errors introduced by PCR and sequencing is low, however, it can become significant when attempting to call very low MAFs. To account for all these sources of error, a number of control measures were taken.

### 5.3.1.2 Strategy for minimisation of artefactual MAFs

Two primer pairs flanking *KRAS* codons 12 and 13 were designed to generate off-set amplicons. Every DNA sample would thus be amplified in two independent PCR reactions and the base-pairs of interest probed in different sequencing cycles. This means that if an error occurs in a particular sequencing cycle, only half of the amplicons will be affected, allowing for mutation calling in the remaining pool. In addition, primers for generation of two ‘mimic amplicons’ were designed. These flanked an *PITPNM2* gene which contained, at the same position as the *KRAS* amplicons, the sequence context of *KRAS* codons 12 and 13, GGT GGC. In Sanger sequencing data, this leads to perfect synchronisation of the traces in the mimic area (Figure 5.5), while the rest of the trace appears mixed. The mimic amplicons enable discounting of



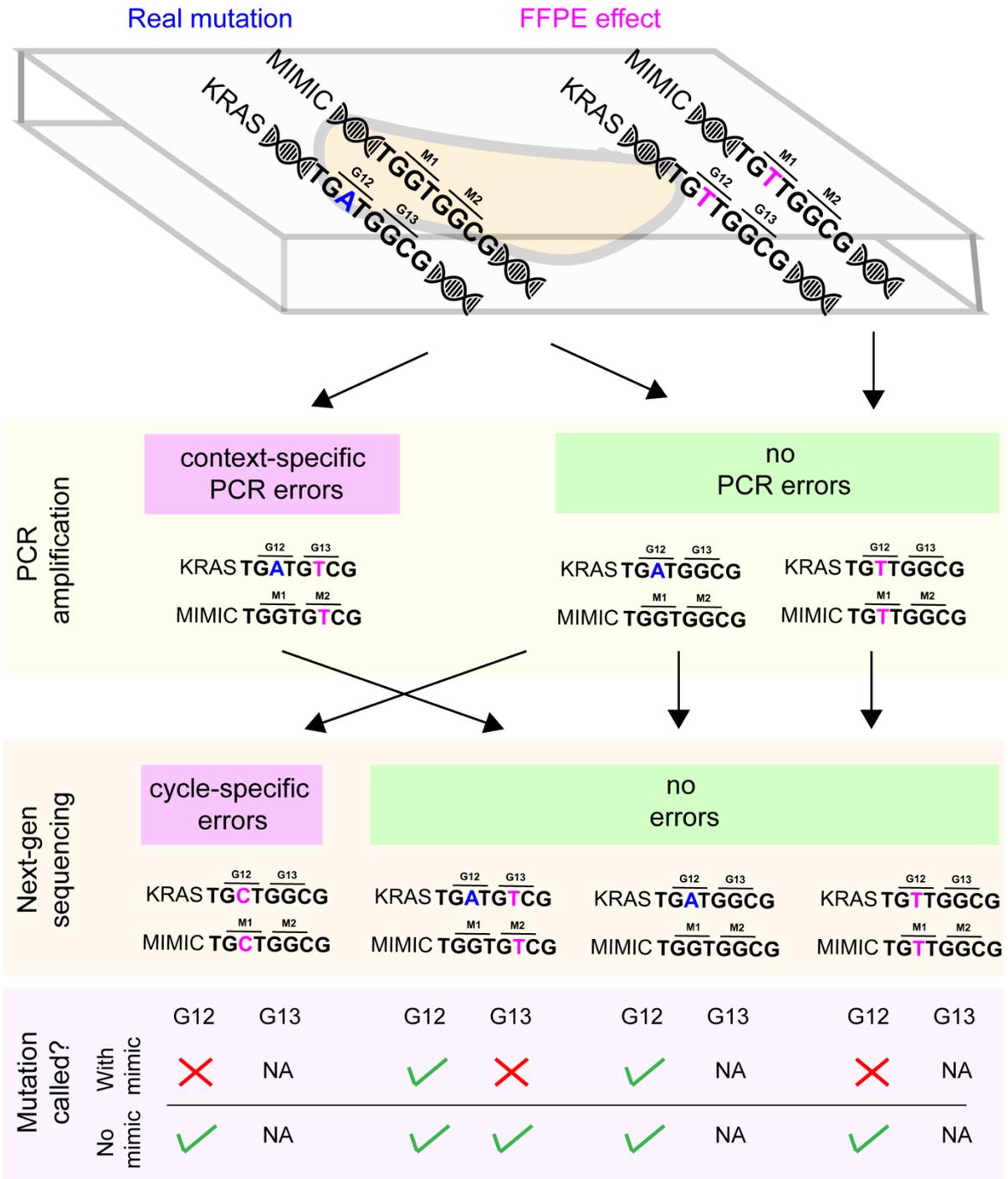


Figure 5.6 Sources of artefactual MAFs and their elimination by the ‘mimic amplicon method’.

Artefacts can be introduced through fixation as well as during various stages of amplification. The pink box outlines scenarios where without mimics, a number of artefactual mutations would be falsely identified as real (red “X”).

Every DNA sample was thus amplified in two multiplex PCR reactions, each containing a *KRAS* and the corresponding ‘mimic’ primer pair. Figure 5.6 outlines the different scenarios for artefactual mutation generation and how the ‘mimic amplicon method’ enables discounting of artefactual mutations. Mutations were thus only called if they are present in both *KRAS* amplicons but remained at background noise levels in the mimic amplicons.

### 5.3.2 Lateral expansion of *KRAS* mutations

The method described in 5.3.1 was applied to DNA from sections from 256 patients ranging in age from 20–91 years. FFPE-related artefacts were readily detectable in the samples. In particular, a high frequency of artefactual G:C>A:T SNVs were observed at a frequency of about 0.05%, which is most likely the effect of deamination of cytosines. Comparatively low frequencies of about 0.01%-0.02% were found for other transversions (Figure 5.7A). As described above, the mimic amplicons enabled discrimination between real and artefactual MAFs. Compellingly, true *KRAS* mutations can easily be visualised in a plot as standing out in pairs (representing the two *KRAS* amplicons) against the noise observed in all four amplicons (Figure 5.7B).

*KRAS* activating mutations were found in 35 out of 256 patients tested, corresponding to 13.7% of the cohort. 7 types of SNVs leading to amino acid changes at codons 12 and 13 were detected, with G12D occurring at the highest frequency, followed by G12V and G12C (Table 5.2). Detected MAFs were in the range of 0.12%–2.35% (Figure 5.8A). From the MAFs as well as the known crypt numbers obtained from DeCryptICS image analysis of serial sections for each sample, the mutant patch sizes could be estimated (Figure 5.8B).



**Figure 5.7 Noise and mutation calling in targeted sequencing data.**

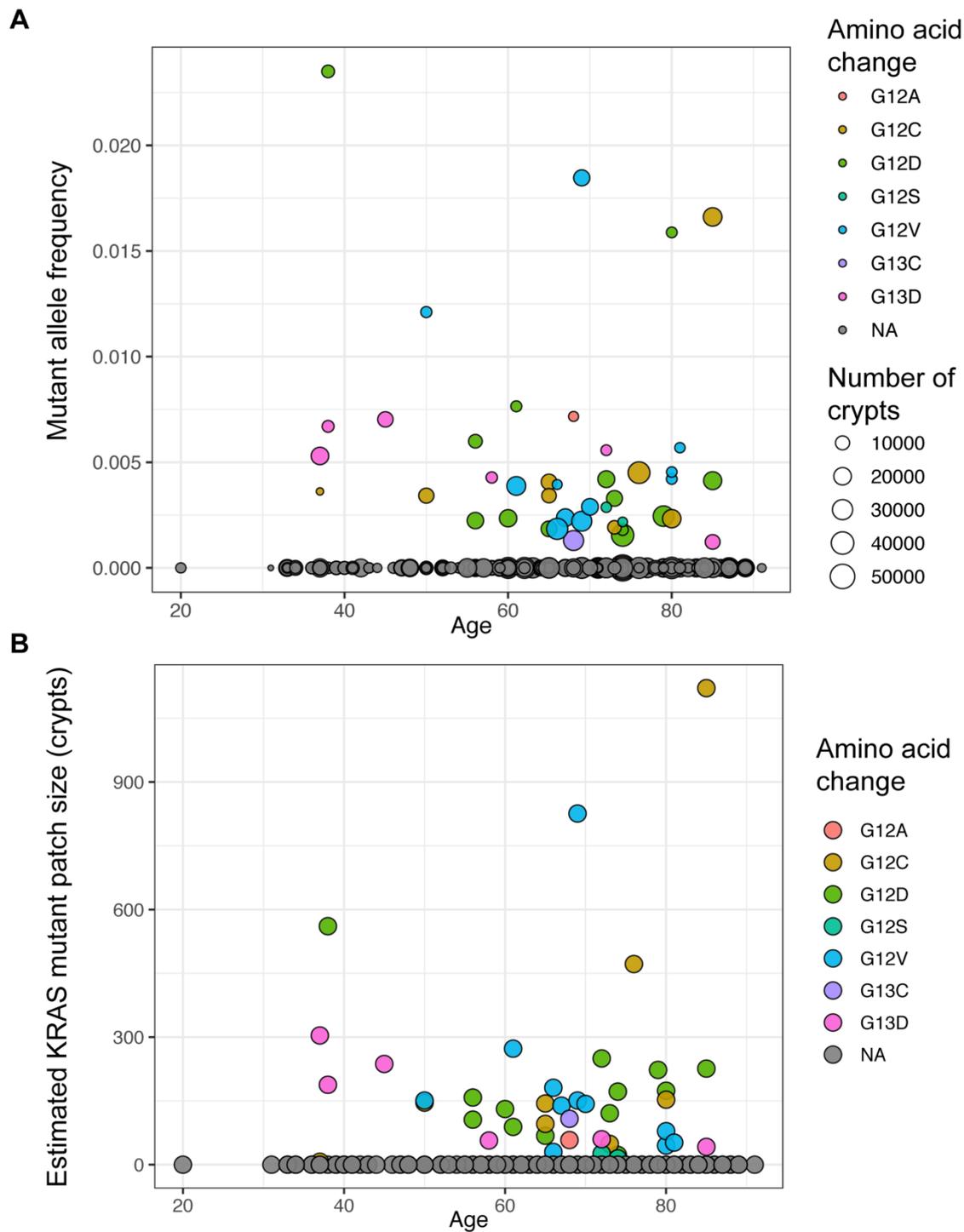
(A) Graph showing frequency of background noise observed for different nucleotide transversions observed at position 2 of codon 12 of KRAS. Error bars = standard deviation.

(B) Graph showing MAFs for KRASG12D in KRAS and mimic amplicons. Samples with identified mutations in both KRAS amplicons but not in mimic amplicons circled in red.

Table 5.2 **KRAS mutations detected in normal human colon samples.**

Amino acid change	Number of samples	Frequency (%)
G12D	13	31.0
G12V	11	26.2
G12C	8	19.0
G13D	6	14.3
G12S	2	4.8
G13C	1	2.4
G12A	1	2.4

Equally, these were used by Dr Edward Morrissey to model the process of *KRAS* mutation accumulation and expansion by fission in the human colon. Briefly, the model requires an estimate for  $\Delta C_{fix}$ , which influences the number of samples that carry mutations as well as the fission rate, which determines how many patches will be large enough to be detected by our method. Using vague priors, the model then finds optimal values for  $\Delta C_{fix}$  and the *KRAS*-mutant fission rate that generate the mutant patches observed (For details on mathematical models used see Appendix D). This analysis revealed a 17-fold increase in lateral expansion of *KRAS*-mutant crypts. Simulating this elevated fission rate of about 12% per year demonstrated that the mutational burden results from a small number of somatic clones that expand dramatically. In contrast, clonal marks have a much higher event rate, but due to the more modest expansion (fission rates of up to 4% for *KDM6A*), the overall mutant allele burden in the tissue is lower (Figure 5.9). This can be expressed with a lifetime expansion coefficient,  $C_{exp}$ , which is obtained by normalising for mutation rate and describing the combined effects of intra-crypt dynamics and subsequent fission over time.  $C_{exp}$  thus represents a way to rank mutations according to their ability to expand in the human colonic epithelium. The highest  $C_{exp}$  was found for *KRAS* followed by *KDM6A* and *STAG2* mutations, which have an average lifetime expansion coefficient that is 2400-, 19- and 13- fold greater than the neutral clonal marks, respectively (Figure 5.9).



**Figure 5.8 Identification of KRAS mutations in normal human colonic epithelium.**

(A) Mutant allele frequency data from 256 individuals plotted against age. 35 individuals displayed detectable mutations.

(B) Estimated KRAS-mutant patch sizes from 256 individuals plotted against age.

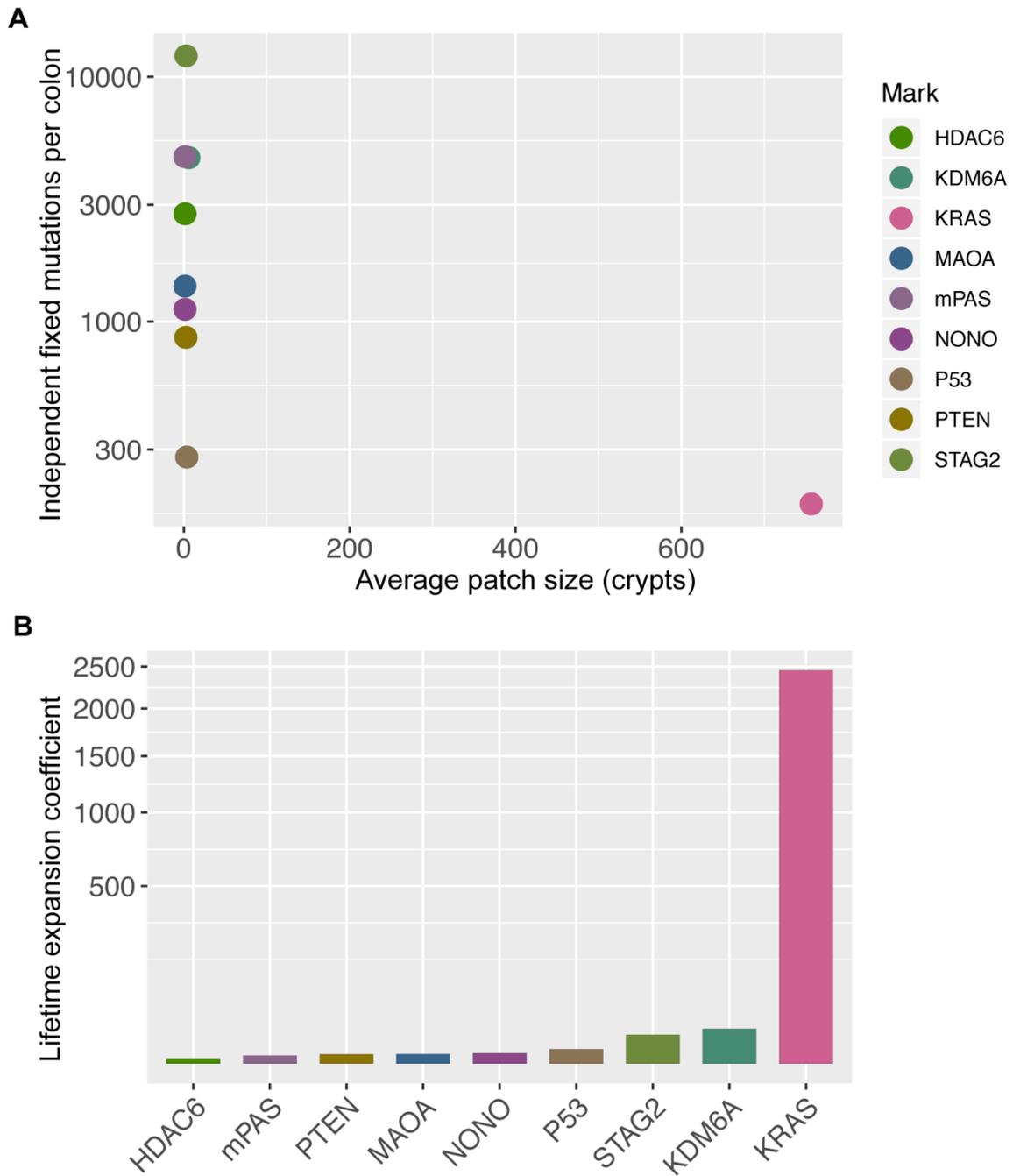


Figure 5.9 **Ranking advantage conferred by gene-specific mutations.**

(A) Average patch size of each clonal mark plotted against inferred total number of clones per colon.

(B) Lifetime expansion coefficient normalised to the neutral mark mPAS (=1).

## 5.4 Crypt fusion

The findings presented here suggest that in the human colon, about 0.7% of crypts undergo fission every year. In addition, mutations can dramatically elevate this rate by several-fold. However, the length of the human intestine does not seem to increase significantly with age (Hounnou *et al.*, 2002). Therefore, a process that balances crypt fission must occur. Potential mechanisms could be crypt death as well as crypt fusion. As for the dramatic effects observed for lateral expansion of *KRAS*-mutant crypts, it seems plausible that, as mutations drive fission to increase the crypt density in a defined area, these start to merge at the borders of the mutant patch to relieve local overcrowding. The idea of crypt fusion is at least a decade old, with Kim and Shibata first speculating about such a process in the discussion of their methylation paper (Kim & Shibata, 2004). Recently, crypt fusion has been described by intravital microscopy in the mouse small intestine (Bruens *et al.*, 2017). This work showed that, at least in the mouse, the processes of crypt fission and fusion occur via opposite mechanisms (Figure 5.10). To investigate fusion in human samples, a method to quantify the rate of this process in *en face* colonic sections was devised.

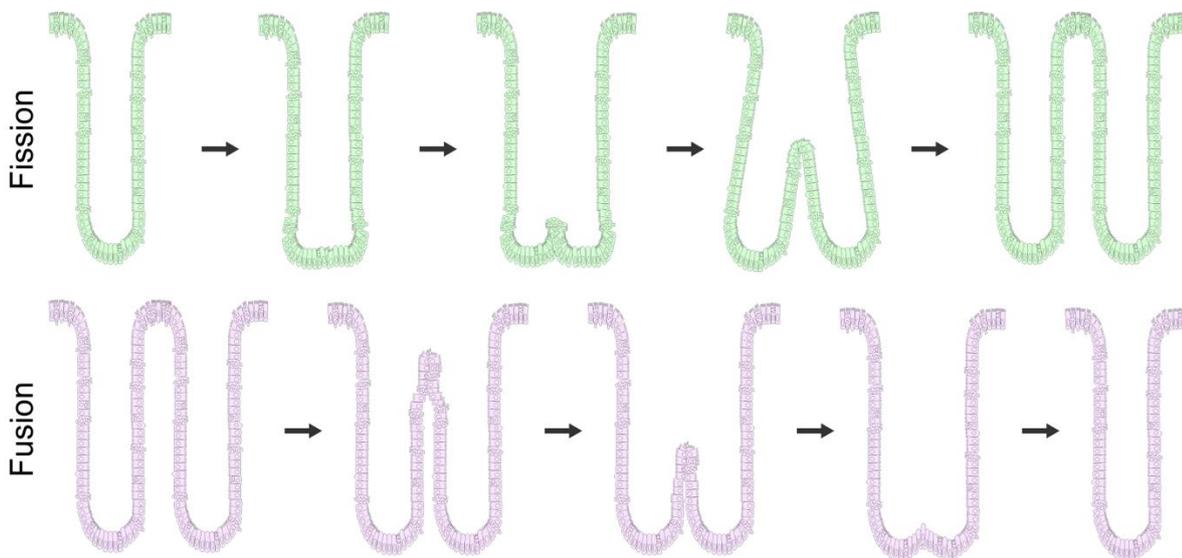


Figure 5.10 **Proposed process for crypt fission and fusion.**

*Mouse intravital microscopy data suggests that crypt fission occurs as a ‘bottom-up’ process, whereas fusion is proposed to be a ‘zipping-down’ between crypts.*

5.4.1 The fufi

In human colon samples, quantification of crypt fission processes historically relied on the observation and quantification of so-called ‘bifid crypts’ – crypts that contain two ‘arms’ indicative of ongoing fission in longitudinal sections (Cummins *et al.*, 2008). In *en face* colonic tissue sections, these structures manifest as two connected circles of cells, previously called ‘8-shaped’ crypts (Bruens *et al.*, 2017). Up to date, no markers that discriminate crypt fission or fusion have been identified. This means that 8-shaped crypts could be snapshots of fusion or fission events. The term ‘fufi’ (from FUSion or FISSION) is thus proposed here, as a means of agnostically referring to these structures. Fufis were defined stringently as two connected crypts without any discernible gap but with two clearly identifiable lumina (Figure 5.11). The latter is key to discerning fufis from potential artefacts caused by oblique sectioning of crypts (Figure 5.11, rightmost schematic).

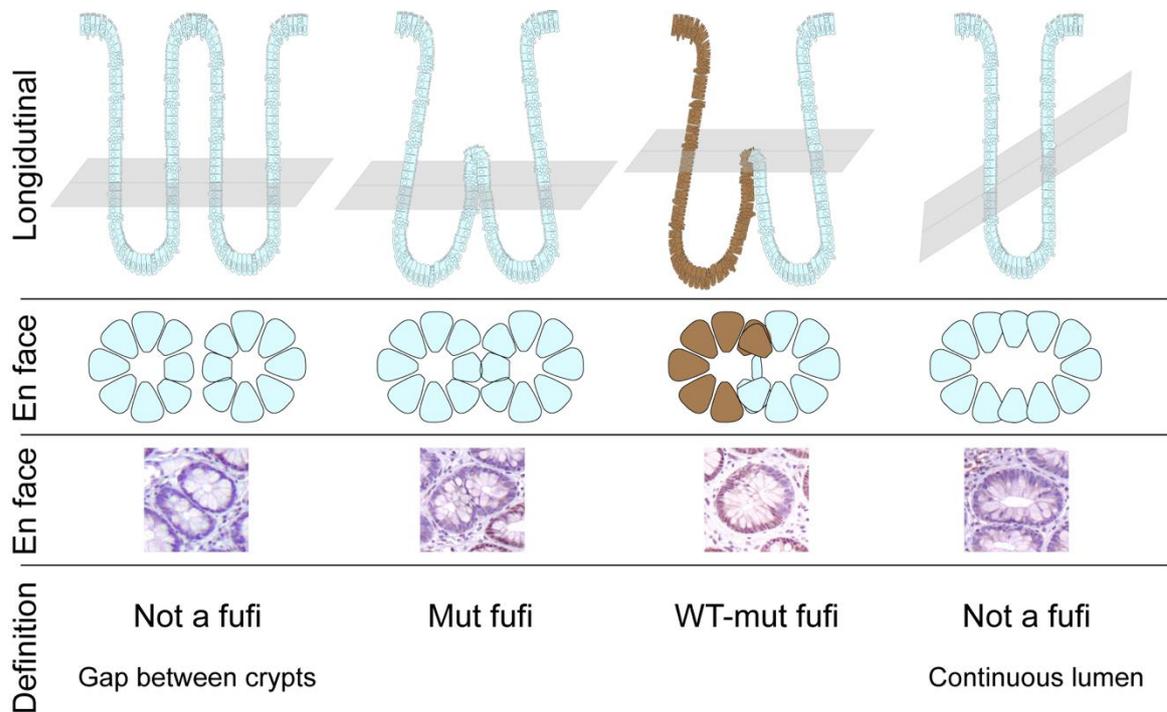
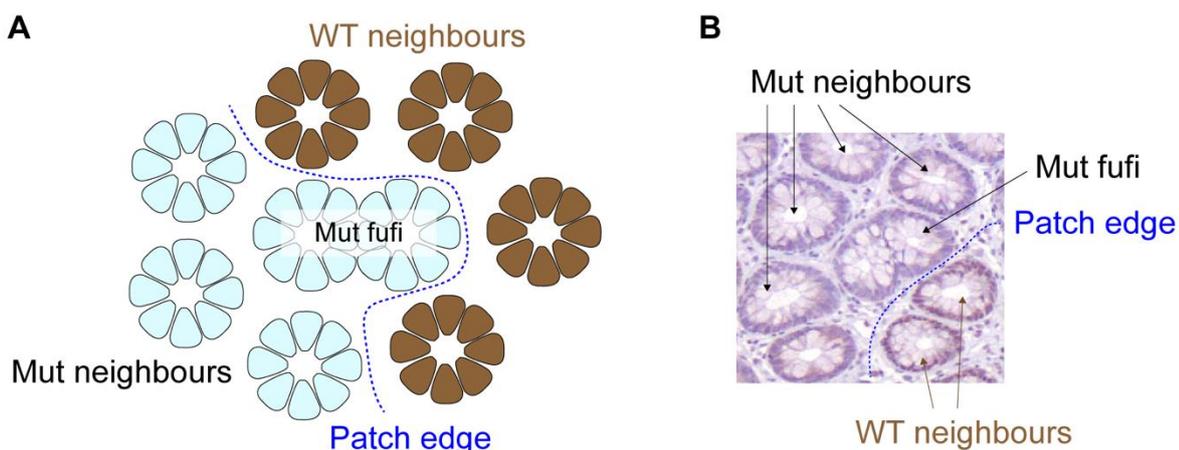


Figure 5.11 **Defining the fufi.** Schematic representations as well as images of real examples with a focus on the key criteria of crypt separation and double lumen.

To infer fusion rates for gene-specific mutations, fufis comprising at least one mutant crypt were analysed. There are two types of such marked fufis: Those comprising two mutant crypts (mut fufis) and those containing one wild-type and one mutant crypt (WT-mut fufis) (Figure 5.11). Importantly, WT-mut fufis were always found exactly half WT and half mutant and were thus assumed to represent crypt fusion events. Mut fufis however, could indicate a fission or fusion event

### 5.4.2 Inference of crypt fusion rates

The mathematical analysis for inference of fusion rates was performed by Dr Doran Khamis and details can be found in Appendix E. Briefly, the observed patch sizes are the output of fission and fusion events. Therefore, the fission rates derived from the distribution of patch sizes (section 5.2.1) were an important prerequisite for the calculation of fusion rate. Importantly, mathematical modelling by Dr Edward Morrissey showed that patch size is dominated by fission (Appendix C: Figure C.1). Therefore, fusion rates could not be derived from patch sizes but required quantification of fufis. Of note, analysis was restricted to fufis neighbouring at least one wild-type crypt, hereafter referred to as patch edge fufis (Figure 5.12).



**Figure 5.12 Scoring of patch edge fufis and neighbours.**

(A) Schematic representation of patch edge mut fufi with four WT and three mut neighbours.  
 (B) Representative image of *KDM6A*<sup>-</sup> patch edge fufi with WT and mut neighbours highlighted.

Of note, single fufis surrounded entirely by WT crypts were scored as patches comprising one crypt. Scoring the amount and type (WT or mut) of crypts neighbouring patch edge fufis (Figure 5.12) enabled calculation of an average number of neighbouring WT and mut crypts for mut fufis. This was used to assign the probability of fusion and fission identity to mut fufis. Combined with the known fission rate, this enabled calculation of the mark-specific fusion rate (for mathematical details see Appendix E).

#### 5.4.2.1 Inference of the neutral crypt fusion rate from mPAS fufis

For inference of the homeostatic human crypt fusion rate, mPAS was chosen, as it displays neutral behaviour but exhibits a relatively high event rate, which expedited quantification. WT, mut and WT-mut fufis were quantified manually in sections from 81 patients of the age range 33-89 years (Figure 5.13). In total, 26,455 fufis (26,391 WT and 64 mPAS<sup>+</sup> fufis) were scored.

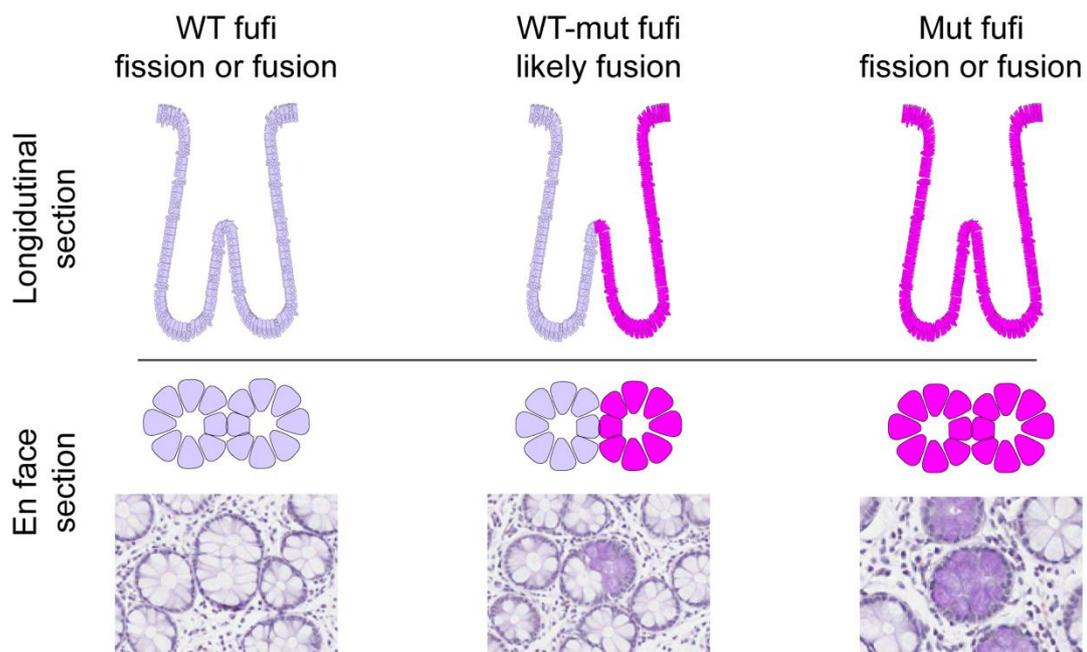


Figure 5.13 **Identification of mPAS<sup>+</sup> fufis.**

*Schematic illustration as well as representative images from human colonic tissue sections.*

*Left: WT fufi – not used for fusion rate.*

*Middle: WT-mut mPAS<sup>+</sup> fufi surrounded by WT crypts.*

*Right: Mut fufi with WT and mut neighbours.*

Combined with the known patch size distribution this enabled calculation of the homeostatic crypt fusion rate (for mathematical models used see Appendix E), which was found to be 0.77% per year (95% CI: 0.46-0.28) (Figure 5.15). Interestingly this is remarkably similar to the mPAS fission rate calculated earlier (0.69% per year) (Section 5.2.1 and Figure 5.2), suggesting that in homeostasis, human colonic crypt fission and fusion balance each other.

### 5.4.2.2 Inference of the fusion rate for KDM6A-negative crypts

The next important question was whether the local increase in fission that can result from gene-specific mutations is balanced by an increase in fusion, potentially as a mechanism to relieve local overcrowding. To answer this question, KDM6A was chosen, as it is the mark causing the strongest increase in crypt fission. WT, mut and WT-mut fufis were quantified in sections from 90 patients aged 13–93 years (Figure 5.14). In total, 6285 fufis (6227 WT and 58 KDM6A-) were scored.

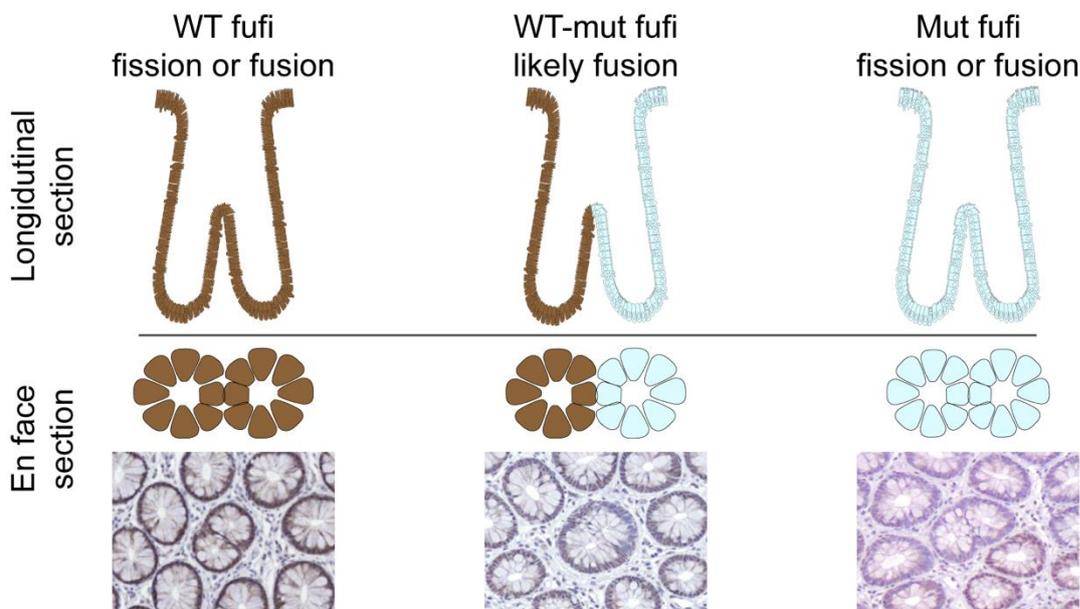


Figure 5.14 **Identification of KDM6A-negative fufis.**

*Schematic illustration as well as representative images from human colonic tissue sections.*

*Left: WT fufi – not used for fusion rate.*

*Middle: WT-mut KDM6A- fufi surrounded by WT crypts.*

*Right: mut fufi with WT and mutant neighbours.*

Modelling revealed a fusion rate of 0.61% per year (95% CI: 0.28–1.03) for KDM6A-negative crypts, remarkably similar to the fission and fusion rate derived for mPAS (Figure 5.15). This indicates that the local increase in crypt fission that results from loss of KDM6A to about 4% per year is not balanced by an increase in crypt fusion. Therefore, fission and fusion may be independent processes, driven by separate mechanisms.

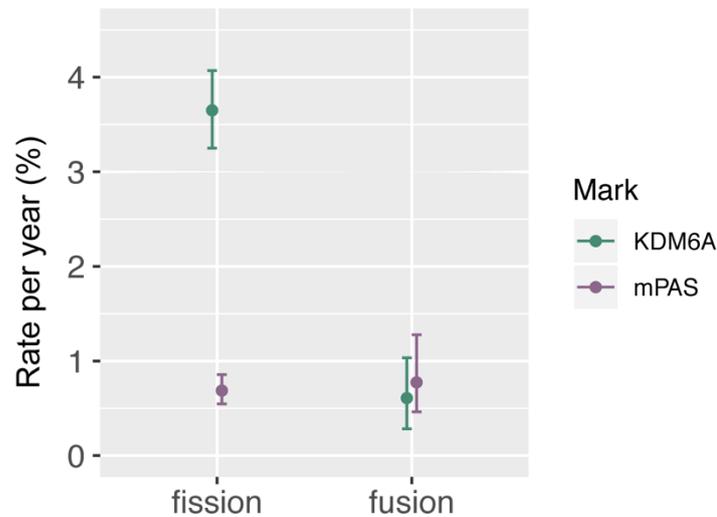
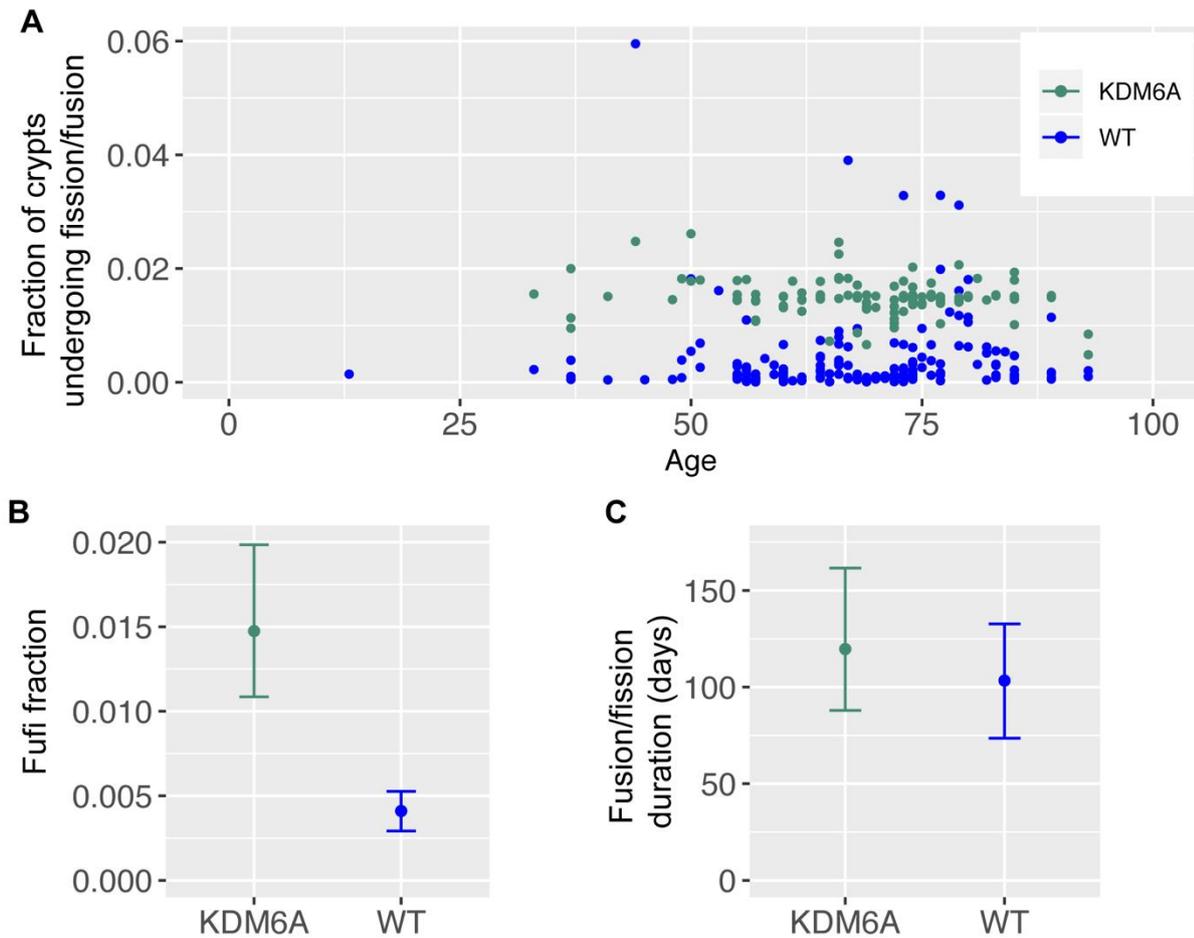


Figure 5.15 **Crypt fission and fusion rates for mPAS and KDM6A.**  
*Error bars = 95% CI.*

### 5.4.3 Fission and fusion duration

In the mouse, intravital microscopy revealed that the dynamic part of crypt fission and fusion spans around 3-5 days, whereby both processes take the same amount of time. The processes cannot currently be directly measured in humans, however, an estimation can be derived. Briefly, assuming equal duration of human crypt fission and fusion, they can be modelled as two simultaneously occurring processes. Therefore, the frequency of fufis at any given moment in time is a snapshot combining ongoing fission and fusion events. It follows that the frequency of fufis in a large cohort of individuals of a wide range of ages can be used to derive the duration of the process (For details on mathematical models used see Appendix E). To rapidly generate a large dataset, the neural network DeCryptICS was trained for fufi detection using the criteria described in 5.4.1. Fufi frequencies (defined as the frequency of fufis as a fraction of total crypts) were obtained for KDM6A-stained colonic tissue sections from 89 patients ranging in

age from 13-93 years. This revealed significant spread between individuals (Figure 5.16A & B).



**Figure 5.16 Inference of human crypt fission and fusion duration.**

(A) Plot showing frequency of fufis as fraction of total crypt number for KDM6A-positive (WT) and KDM6A-negative (KDM6A) crypts.

(B) Plot showing mean fufi frequency for KDM6A-positive (WT) and KDM6A-negative (KDM6A) crypts. Error bars = 95% CI.

(C) Plot of inferred fusion and fission duration for KDM6A-positive (WT) and KDM6A-negative (KDM6A) crypts. Error bars = 95% CI.

Importantly, to explore a potential difference between the fufi duration in KDM6A-negative crypts and the bulk crypt population (KDM6A-positive, hereafter referred to as wild-type, WT),

the dataset was divided accordingly. Of note, the mean calculated for WT crypts included 1,804,580 crypts and 6227 fufis, whereas the KDM6A-negative dataset was much smaller with 5070 crypts and 58 fufis. Mathematical modelling by Dr Doran Khamis revealed the fission and fusion duration to be 103 days (95% CI: 73–133) for WT crypts and 120 days (95% CI: 88–162) for KDM6A-negative crypts (Figure 5.16C). There is no significant difference between these two values, indicating that fission and fusion take about the same amount of time in a mutant as well as a wild-type setting. Interestingly, the duration calculated here is about 30-fold slower than in the mouse.

## 5.5 Summary and Discussion

Lateral expansion of mutations into fields increases the pool of cells available for acquisition of a ‘second hit’ and thus promotes cancer initiation. In the mouse colon, this process occurs via crypt fission, and can be accelerated by pro-oncogenic mutations such as activation of *KRAS*. Here, lateral expansion was investigated in human colon using clonal marks.

The distribution of patch sizes across the large cohort of patients of different ages enabled calculation of human colonic crypt fission rates associated with gene-specific mutations. The neutral rate, inferred from mPAS, MAOA and NONO, was found to be 0.7% per crypt per year. A 3-fold increase in crypt fission was found for *STAG2*, *PTEN* and *p53* and an even stronger 4.5-fold effect for *KDM6A*. This demonstrates that gene-specific mutations can increase human colonic crypt fission to generate mutant fields. Combined with biased intra-crypt dynamics this can significantly increase the mutant allele burden in the epithelium.

To investigate the effect of *KRAS* activating mutations a targeted amplicon sequencing approach was taken. Importantly, a mimic amplicon strategy was devised, which ensured the elimination of artefactual mutations stemming from tissue fixation, PCR amplification and sequencing. Noise was particularly prominent for G>A transversions, which is most likely due to deamination of cytosines. Treatment of DNA samples with uracil DNA glycosylase prior to PCR has been shown to reduce such artefacts (Kim *et al.*, 2017) and should be attempted for future experiments.

*KRAS* activating mutations were identified in 13.7% of samples. However, the DNA sequenced was bulk extracted from FFPE sections. Therefore, a MAF of 1% may represent a patch of 10 mutant crypts in a section with 500 crypts, while it may indicate a patch of 400 mutant crypts in a section with 20,000 crypts. Therefore, the sensitivity of the method is inversely correlated to section size. It is thus likely that a number of small *KRAS*-mutant patches were not detected in large sections, as the resulting MAF in the bulk sample would not exceed our detection threshold. This relationship between section size and mutation detection is visible in differences between the experimental data and simulation performed by Dr Edward Morrissey, whereby the former seems to lack predicted low frequency MAFs (Figure 5.17). Potentially, these are present in the samples but not detected with the current method. In line with this hypothesis is the close correspondence between KDM6A data and simulation. As KDM6A is a clonal mark, the sensitivity of mutation detection is at single crypt level and independent of section size.

For future studies, higher sensitivity could be achieved by subsampling. For example, by dividing larger sections into samples of 1000 crypts, patches of 2 crypts would yield MAFs of 0.1% and would thus be detected. In addition, the simulation reveals an overrepresentation of large *KRAS*-mutant patches in young (<50 years old) individuals in the targeted sequencing data (Figure 5.17). It will be interesting to obtain clinical data for these individuals to examine potential effects of lifestyle, therapies or genetic background.

Regardless of these challenges, the data yielded insights into the spread of *KRAS* activating mutations in normal human colonic epithelium. The most frequent mutation identified was G12D followed by G12V and G12C. The first two are in line with the frequencies of mutations found in CRC, whereby *KRAS*G12D and G12V mutations are the two most frequent specific alterations, followed by BRAFV600E and *KRAS*G13D (COSMIC database). Of note, modelling by Dr Edward Morrissey revealed that the frequencies of different *KRAS* activating mutations identified here mostly scale with the proportion of colorectal cancers they are found in (Figure 5.18, data from cBioPortal). This suggests that they all harbour a similar efficiency of conversion into malignancy.

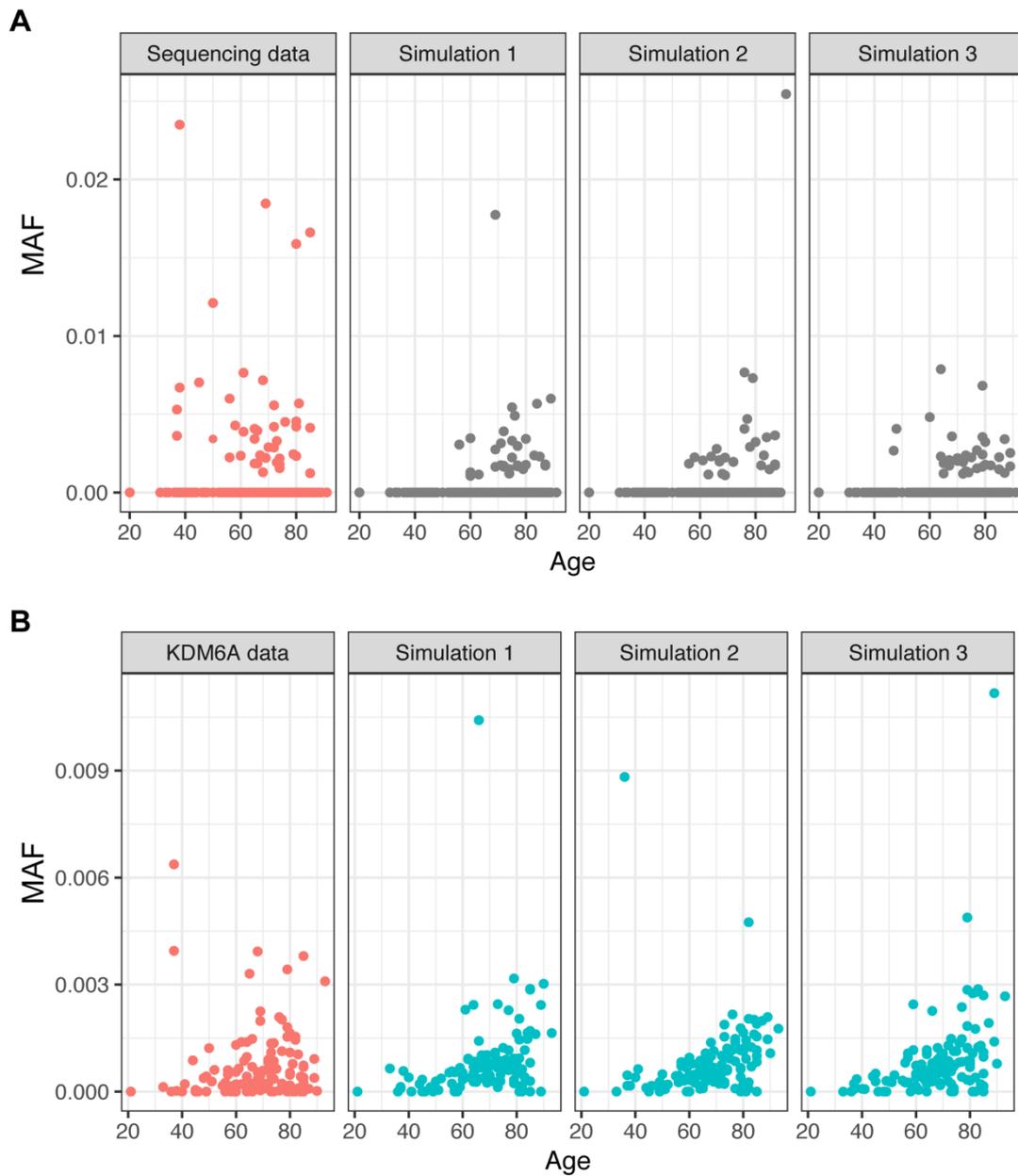
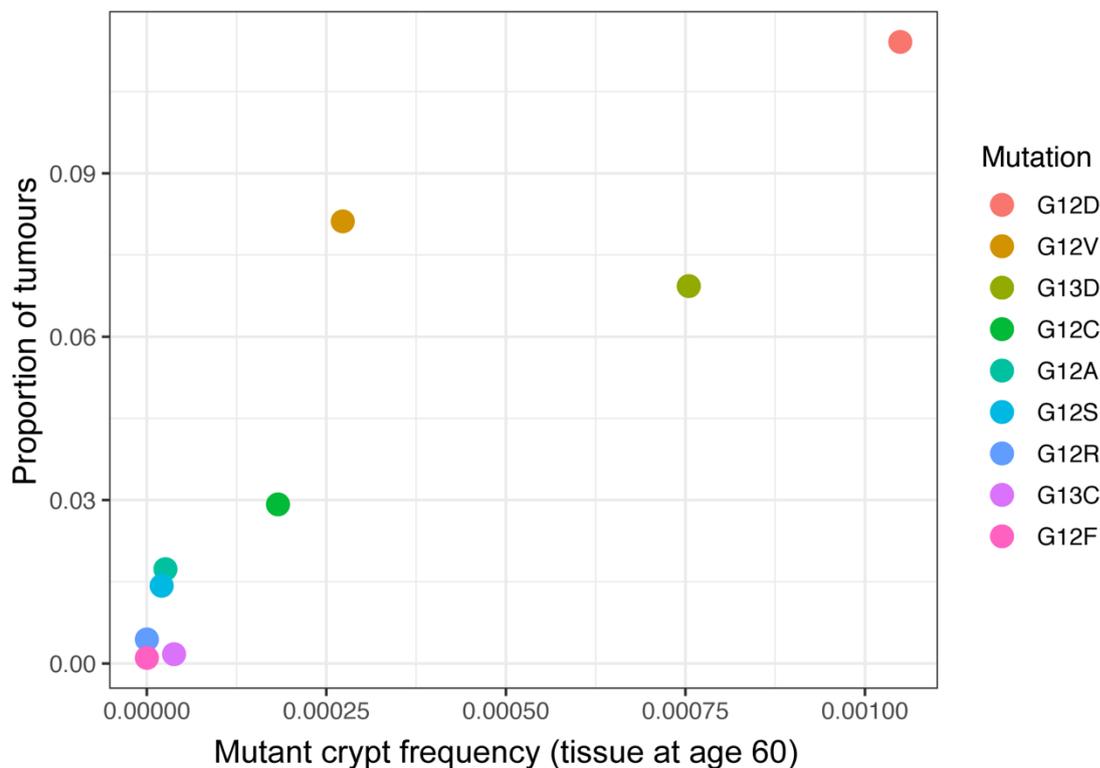


Figure 5.17 **Comparison between data and simulations.**

Figure kindly provided by Dr Edward Morrissey.

(A) Comparison between experimental data (left) and three simulations for *KRAS* MAFs.

(B) Comparison between experimental data (left) and three simulations for *KDM6A*.



**Figure 5.18 Similar efficiency of conversion into malignancy for different KRAS activating mutations.**

*Modelled frequency of crypts with KRAS activating mutations plotted against the proportion of tumours they are found in (data from cBioPortal).*

While for the majority of individuals (157 out of 256), only one FFPE tissue block was sequenced, for 99 between two to seven independent samples were analysed. Interestingly, 6 out of these 99 individuals harboured more than one mutation, found in different tissue samples. In two of these, the same KRAS activating mutation was found in two independent samples. It is possible that these represent one single large expansion, however, in absence of spatial information this remains elusive. In the remaining four cases two different KRAS activating mutations were found in the samples, suggesting independent mutational events. It will be interesting to refer back to clonal marks data to determine whether these individuals exhibit high clone frequencies, which may be indicative of a mutator phenotype. Clinical data as well as investigation of mismatch repair status may aid in further interpretation of the findings.

The mutant patch sizes calculated from MAFs were in the range of 7-1121 crypts. Due to the low frequency of *KRAS* mutations and the fact that blocks usually only contained one type of alteration, it can be assumed that these mutant crypts correspond to single large mutant areas rather than a number of smaller patches. This hypothesis could be tested by subdivision of sections followed by sequencing, as described above. Equally, in-situ hybridisation or even IHC could visualize patches in serial sections. However, to date we have not been able to adapt the BaseScope™ method to our human FFPE colonic tissue sections.

The subject of *KRAS* mutations in colonic epithelium is intimately linked to aberrant crypt foci (ACF). First described in rodents treated with azomethane (Bird, 1987) and shortly after in human colonic tissue samples (Roncucci *et al.*, 1991) these are single or groups of crypts characterised by elevation, oval lumina and thicker epithelial lining. ACF can be further divided into hyperplastic and dysplastic ACF. The former show increased proliferation and are thought to represent an intermediate stage between normal and adenoma (Morimoto *et al.*, 2002). The latter are characterised by an irregular outline, increased overall and nuclear size and loss of polarity and are often classified as microadenomas (Gupta *et al.*, 2017). Importantly, a number of human studies have found a high frequency of *KRAS* activating mutations in hyperplastic ACF, with the frequency of such alterations observed in different studies ranging from 16.3-85% of ACF examined (Otori *et al.*, 1995; Takayama *et al.*, 1998, 2001; Quintanilla *et al.*, 2019). Consistently, dysplastic ACF show a lower frequency of *KRAS* activating mutations than hyperplastic ACF (Takayama *et al.*, 1998, 2001). ACF can be detected in humans by high-magnification chromoscopic colonoscopy (HMCC), whereby the absorptive dye methylene blue enables identification due to a darker appearance (Kudo *et al.*, 1996). As ACF are thought to represent precursors to colorectal cancer (Kukitsu *et al.*, 2008; Orlando *et al.*, 2008), they have been extensively investigated as biomarkers to identify individuals at high risk of developing CRC, including patients with ulcerative colitis and FAP (Roncucci *et al.*, 1991; Kukitsu *et al.*, 2008). However, the correlation between ACF and tumour burden is still a matter of debate (Takayama *et al.*, 1998; Kukitsu *et al.*, 2008; Orlando *et al.*, 2008; Wargovich, Brown & Morris, 2010; Gupta *et al.*, 2017; Quintanilla *et al.*, 2019), an issue most likely related to the variability in endoscopic analyses and classification as well as the complication that by far not all ACF progress to cancer (Ghosh *et al.*, 2018). Considering the high frequency of *KRAS*

mutations found in ACF and their potential use as predictive biomarkers for CRC, it will be very informative to systematically detect such structures in our tissue samples. Ideally, DeCryptICS could be trained for this purpose by a histopathologist. In a first step it will be important to determine whether the *KRAS* activating mutations identified here correlate with the presence of ACF in serial sections. Identified (*KRAS*-mutation positive and negative) ACF could then be further investigated with IHC (to look for loss of mismatch repair genes) or LCM and sequencing of other CRC-associated genes to characterise them further. Combining this genetic information with histopathological description of the tissue (ACF or histologically normal *KRAS*-mutant epithelium) and clinical patient data such as high BMI, presence of inflammatory conditions and tumour grade, may form the basis for more sophisticated classification of different ACF and generation of biomarkers that can be detected by HMCC. This may then enable refined patient stratification, whereby detection of certain high CRC risk-associated features may trigger a chemoprevention or more rigorous surveillance regime. In terms of the former, the work presented here defines the timeframe taken for *KRAS*-mutant patches to expand in the human colonic epithelium. Blocking or even slowing down of this process to the point where large expansion would be shifted beyond normal human lifespan could therefore represent a viable CRC prevention strategy, in particular for tumours that initiate from *KRAS* mutant fields. Short-term MEK inhibition (MEKi) with the MAPK inhibitor CI-1040 (PD184352) has been shown to revert the hyperplastic phenotype observed upon induction of *KRASG12D* mutations in the mouse intestine (Feng *et al.*, 2011). The effect of a regime of intermittent MEKi could first be tested in human organoids xenografted into mice, a method recently published by the Sato group (Sugimoto *et al.*, 2018).

The dramatic increase in crypt fission observed for KDM6A loss and *KRAS* activation lead to the hypothesis that such an effect may drive crypt fusion events, which could relieve local overcrowding caused by increased crypt density. Importantly, an accurate rate of fission could be estimated as the outcome of it is observed by the age-related changes in patch sizes. As modelling revealed that fusion does not impact on this distribution, the rate of this process could be derived. To do so, the term ‘fufi’, was introduced as a means of agnostically referring to crypts captured during the fission or fusion process in *en face* tissue sections. By quantifying

---

mPAS<sup>+</sup>-containing fufis in sections from 89 patients, around 0.77% of crypts were calculated to undergo fusion per year, a value remarkably similar to the homeostatic fission rate derived from mPAS, MAOA and NONO patches. This suggests that as part of homeostasis, akin to the mouse, human crypts undergo fission and fusion events. Interestingly, a very recent study used the histochemical detection of CCO-deficiency to quantify the human crypt fusion rate. In a similar way to the method described here, Baker and colleagues quantified CCO<sup>-</sup>-containing fufis in 13 patient samples to infer a neutral crypt fusion rate of 1.1% per year (range of medians: 0.2–2.4) (Baker *et al.*, 2019). This value is very similar to the fusion rate calculated here, albeit with much larger error bars reflecting the very small patient cohort. Importantly, the model used by Baker and colleagues assumed a balance between fusion and fission rates before performing calculations while the approach taken here did not.

Next, about 0.6% of KDM6A-negative crypts were calculated to undergo fusion per year. This value is remarkably similar to the neutral fusion and fission rates calculated for mPAS. It indicates that loss of KDM6A, while causing a dramatic increase in crypt fission, does not impact on crypt fusion. Therefore, the hypothesis that a local increase in crypt fission rate is balanced by accelerated fusion does not hold true for KDM6A. It will be interesting to investigate whether the same outcome is also found for other gene-specific mutations associated with biased fission such as STAG2, p53, PTEN and KRAS.

The duration of crypt fission and fusion was inferred from fufi frequencies for KDM6A-negative and positive crypts. Interestingly, fission and fusion both seem to occur more frequently in the mouse small intestine, where at any given point in time about 4% of crypts are fissioning and the same percentage fusing. In the human colon it is below 0.5%. For the duration of fission and fusion 103 days for KDM6A-positive and 120 days for KDM6A-negative crypts were calculated. This suggests that KDM6A loss does not alter the duration of fission and fusion. Again, it will be interesting to see if the same holds true for other clonal marks. Fission and fusion durations had also been inferred by Baker and colleagues, who analysed samples from 10 individuals and obtained a value of 41 days, with a range of medians of 4-41 weeks (Baker *et al.*, 2019). Their much larger range can be attributed to the much smaller sample size. Regardless of this limitation, the similar values for neutral crypt fusion rate and duration obtained by Baker and colleagues support the findings presented here.

Considering local overcrowding caused by elevated fission rates, it is plausible that, since such clones are rare, the epithelium as a whole is able to absorb the pressure caused by local increases in crypt density. Specifically, crypts may rearrange within the epithelium, which may be accompanied by stromal remodelling. More intriguing however, is the question of potential intra- and intercellular mechanisms driving the different processes. Crypt fission in the mouse small intestine has previously been described to be fuelled by proliferation as well as positioning of Paneth cells (Snippert *et al.*, 2013; Langlands *et al.*, 2016). However, the molecular mechanism remains elusive. Even less is known about the intracellular and intercellular pathways that might drive fusion. To gain insights into potential pathways regulating the process, RNA-Seq or proteomics experiments could be performed on fufis directly isolated from fresh mouse tissue. The fission bias observed for KDM6A could then be investigated using inducible KDM6A(*Isl*) mice crossed to reporter lines. Combined with BrdU labelling these experiments could also show whether proliferation plays a role in accelerating fission in KDM6A-deficient crypts. The same methodologies could also be applied to investigate the fission bias observed for loss of STAG2 and PTEN as well as stabilisation of p53.

Clonal marks as well as targeted sequencing can provide detailed insights into lateral expansion of mutant clones in the human colonic epithelium. Both methods used rely on detection of specific variants, from which associated behaviours can be inferred. However, the neural network-based tool DeCryptICS might in the future enable a more agnostic approach, whereby serial sections annotated for clonal marks and fufis could be aligned. Areas with increased clone numbers for different marks may indicate underlying local genetic instability while local increases in fufi frequency may be suggestive of a hidden fission-promoting mutant field. The areas could then be laser capture microdissected and, if large enough, whole genome sequenced to find the genetic cause of the phenotype.

## Chapter 6 Discussion

### 6.1 Investigating mutation accumulation in normal human colon

In recent years, a paradigm shift in the general model for cancer initiation has formed the basis for investigations into accumulation of mutations in normal renewing tissues. A number of methodologies are available to examine mutation accumulation in human colonic epithelium. They are summarised in Table 6.1 and discussed in detail in the following sections.

Table 6.1 **Comparison between different methods for investigation of somatic mutations in human colon.**

Method	Advantages	Disadvantages
Clonal marks	Large sample size High resolution	Labour-intensive Material limiting
Targeted sequencing	Large sample size	Detection limit for small expansions No intra-crypt/tissue context
Organoid sequencing	High resolution	Small sample size No intra-crypt/tissue context
LCM single crypt sequencing	High resolution	Labour-intensive Small sample size No intra-crypt context
RNA-Seq data analysis	Large sample size	Detection limit for small expansions No intra-crypt/tissue context
CCO histochemistry	Large sample size High resolution	Not adaptable to other mutations
In situ hybridisation	Large sample size High resolution	Expensive Material limiting

### 6.1.1 Clonal marks for precise quantification in large cohorts

Here, the fixation and spread of mutations in normal human colon was investigated using immunohistochemical detection of somatic variants. By means of an antibody screen the arsenal of such clonal marks was expanded to include, in addition to the previously identified mPAS and MAOA, the six marks HDAC6, KDM6A, NONO, p53, PTEN and STAG2. These were used for IHC on normal human FFPE colonic samples from a large cohort of individuals of a wide range of ages, followed by interpretation of age-related trends in clone frequencies.

This method harbours a number of features that uniquely enable a quantitative description of how gene-specific mutations can achieve high mutant allele burden in the colonic epithelium. Analysis of sections at 20X magnification allows for single-cell resolution within tissue context. Therefore, clonal expansions can be precisely quantified, both on an intra-crypt as well as tissue level. The smallest clones detected here comprised fewer than 10 cells, while the largest KDM6A-negative patch comprised 71 crypts. The simplicity of clone identification enables relatively high throughput analysis of tens of thousands of crypts per individual, totalling several million crypts for every mark. In the future, this will be further accelerated by improvements in the neural-network based tool DeCryptICS. The large sample size ensures that data generated is robust, which is confirmed by statistical analysis and simulation. Finally, as tissue is stored long term embedded in paraffin wax, it is readily accessible for testing of new hypotheses on serial sections. An example of this is IHC for H3K27me3 to investigate potential effects of loss of KDM6A.

Clonal marks also carry limitations. Immunohistochemical detection of gene-specific mutations is limited by the availability of specific antibodies. The antibody screen performed here identified around 20% of antibodies as lacking specificity or immunoreactivity. A particular challenge is detection of activated oncogenes such as KRASG12D and BRAFV600E, where significant cross-reactivity with wild type protein was observed. Initially, clonal marks were limited to X-linked genes, whereby due to male hemizyosity and X-inactivation in females, mutation of one allele is sufficient for visualisation. However, certain autosomal genes such as *PTEN* were also found to yield negative clones. Whether these correspond to mono- or biallelic

loss is not easily determined by IHC. More generally, identification of genetic alterations causing the loss of staining has to date only been achieved for MAOA and KDM6A. It will be informative to attempt the same for the other marks. In addition, the gene causing the mPAS phenotype needs to be identified. Finally, each IHC requires an individual tissue section. Most of the FFPE colonic tissue blocks obtained by the Winton laboratory yield between 10 to 30 sections, in theory enabling investigation with the same number of clonal marks. However, a number of sections are used for antibody screens as well as testing of other methods of mutation detection. Attempts to multiplex the method by dual-colour (red and brown) staining have to date been hampered by challenges in clone detection. Therefore, currently a maximum of around 20 marks per tissue block applies.

### 6.1.2 Targeted sequencing for inference of mutant patch sizes

As visualisation by IHC was not successful for KRAS activating mutations, a targeted sequencing approach was taken, whereby next-generation sequencing was performed on PCR products originating from amplification of DNA extracted from FFPE tissue sections. Importantly, a mimic region was also amplified and sequenced, which enabled estimation of background noise caused by formalin fixation and tissue storage effects. This resulted in detection of KRAS activating mutations in approximately 13% of individuals. MAFs were converted to mutant patch sizes, which revealed dramatic expansion of *KRAS*-mutant clones. The method is easily performed on hundreds of samples, and future studies will benefit from recent technological advances such as automated FFPE DNA extraction (Haile *et al.*, 2017). The targeted sequencing approach developed here requires 1/20 of the DNA from one section (in the range of 5-20 ng per reaction) for analysis of mutations in KRAS exon 2. Therefore, each section could be probed for at least 20 genes, a number which could be increased further by multiplexing. It thus uses considerably less material than mutation detection by IHC, which requires one tissue section per gene. However, the targeted sequencing approach has two major drawbacks: Firstly, spatial context is lost. Mutant crypt numbers can be inferred from MAFs, however, whether these comprise one single large expansion or several smaller patches remains elusive. As KRAS activating mutations are rare and blocks were only ever found to contain one type (G12D, G12V etc), the former is likely. IHC or in situ hybridisation on serial sections

could validate this hypothesis. Secondly, the detection threshold for MAFs is approximately 0.1%, independently of sample size. The minimum detectable mutant patch size within each analysed section is thus proportional to the total number of crypts contained within it. This means that in large sections, small clonal expansions are currently not detected. For future studies this variability could be reduced by subsampling of larger sections.

### **6.1.3 Overview of alternative methodologies**

A number of alternative methods can be used to address questions relating to accumulation of somatic mutations in normal human colonic epithelium. These include sequencing-based approaches as well as histochemical and in situ hybridisation methods. However, a number of factors prevent alternative methods from producing the insights gained here.

#### **6.1.3.1 Sequencing-based approaches**

Studies based on sequencing can be largely subdivided into two types of strategy: Either a small number of crypts were analysed in great detail or large tissue areas examined more broadly.

##### **6.1.3.1.1 In depth sequencing of small samples**

This is exemplified by sequencing of organoids derived from single cells (Blokzijl *et al.*, 2016) and analysis of laser capture microdissected crypts from histologically normal human colon (Lee-Six *et al.*, 2018). Blokzijl and colleagues analysed whole genome sequencing data from 45 organoid cultures derived from single cells from 19 donors of the age range 3-87 years. This offered detailed insights into mutational processes associated with ageing of human intestine and included a description of the most frequent mutational signatures found in the tissue. However, as organoids are grown from single cells after tissue dissociation, intra-crypt dynamics as well as clonal expansion could not be investigated. This was achieved by Lee-Six and colleagues, who whole genome sequenced just over 2000 crypts isolated by LCM to identify mutational signatures. This was complemented by targeted sequencing for 90 putative CRC driver genes. Hotspot mutations were detected in the cancer drivers *PIK3CA*, *ERBB2*,

---

*ERBB3* and *FBXW7* and truncating mutations in a number of genes including *TP53*. However, no *KRAS* mutations were identified. Further, apart from two mutations in *STAG2* and loss of chromosome 10q, which contains *PTEN*, Lee-Six and colleagues did not detect any recurring mutations in the genes used for clonal marks here. In addition, clonal expansions beyond patches of two adjacent crypts were not found. These differences can largely be attributed to sample size. The clones detected using clonal marks are rare, occurring at frequencies of about 1:10,000 crypts. It is thus not surprising that, by sequencing just over 2000 crypts, Lee-Six and colleagues mostly failed to detect mutations in the genes used for clonal marks. In addition, by sampling only around 50 crypts per patient, the chance of capturing entire clonal patches is small, a challenge exacerbated by incomplete coverage of the genome. The same holds true for *KRAS* activating mutations, for which MAFs of 0.2-1.8% were found in approximately 13% of normal colonic samples tested here. As 2000 crypts roughly corresponds to one of the hundreds of samples sequenced here, the chance of finding a *KRAS* activating mutation Lee-Six et al. had in their samples was the frequency of positive samples in our cohort, which is 13%.

In summary, while the studies by Blokzijl et al., and Lee-Six et al., yielded insights into mutational signatures associated with ageing of the human colon, they failed to detect rare mutations and were not equipped to investigate intra-crypt mechanisms as well as clonal expansion processes.

#### 6.1.3.1.2 **Bulk sequencing**

A bulk sequencing-based approach, on the other hand, was taken by Yizhak and colleagues, who developed a method to detect mutations in RNA-Seq data. Applied to normal samples from the Genotype-Tissue Expression (GTEx) Project, this identified macroscopic lesions in most individuals (Yizhak *et al.*, 2019). Mutations were particularly frequent in sun-exposed skin, oesophagus and lung, while large expansions in known CRC drivers were not identified in normal colon gene expression data. Most likely, this is linked to sensitivity which in turn relates to sample size. While clone expansion in the skin might be so dramatic that it can be detected even when diluted out within a large sample, multicrypt patches may be rarer and relatively smaller in the colon. When DNA from too many cells is analysed, MAFs of small clonal expansions are reduced to below detection threshold. Therefore, to detect mutations by bulk

---

sequencing, the sample size needs to be scaled according to the expected clone size. This was not possible for Yizhak *et al.*, who analysed previously acquired data. Therefore, to detect small clonal expansions in human colon by sequencing, a method half way between the targeted KRAS sequencing approach and the laser capture and sequencing of MAOA and KDM6A-negative patches would be most appropriate.

### 6.1.3.2 Histochemical detection of CCO mutations

As described in chapter 1, somatic mutations in the mitochondrially encoded gene *CCO* can be detected by histochemistry. This method was previously used to quantify the behaviour of human colonic stem cells (Baker *et al.*, 2014). More recently, Baker and colleagues quantified CCO-negative crypt containing fufis in samples from 13 individuals to infer a neutral human crypt fission and fusion rate of 0.01 per year, with a range of 0.002–0.024 (Baker *et al.*, 2019). This is very similar to the result obtained with mPAS here, albeit with a much wider confidence interval, caused by a small sample size. Analysis of CCO-deficient crypts can thus yield the same insights into neutral intra-crypt dynamics as well as crypt fission and fusion events as can be obtained from clonal marks data. However, the method is limited to the *CCO* gene. Therefore, biases in these processes remain unquantifiable.

### 6.1.3.3 In situ hybridisation

An alternative to histochemical detection of mutations is in-situ nucleic acid hybridisation. As discussed in chapter 5, a recent development in this area has been the BaseScope™ technology. This method relies on hybridisation of so-called Z-probes to target RNAs followed by several amplification steps. Specificity is very high, enabling discrimination between wild type and mutant transcripts, down to single nucleotide level. This has been successfully used in human tumour sections for detection of various oncogene-activating mutations such as *KRASG12D* or *BRAFV600E* (Baker *et al.*, 2017). However, BaseScope™ is currently very expensive, costing around £60 per section. It is therefore presently not feasible to use this method for screening of hundreds of sections. Instead, it may aid in validation of sequencing results.

A slightly cheaper alternative is RNAScope®, which relies on detection of RNA by hybridisation of at least three Z-probes per transcript. It cannot detect SNVs but could be used for detection of surrogate markers. Mouse data suggests upregulation of *Notum* to be indicative of loss of *Apc* (Owen Sansom, personal communication). This provides motivation for investigation of *APC* mutation detection by RNAScope® for *NOTUM* in human tissue.

Analogous to histochemical detection, in situ hybridisation approaches harbour the disadvantage is that each iteration uses up a tissue section. Therefore, tissue thickness limits the number of gene-specific mutations that can be probed for.

## 6.2 Gene-specific mutations can subvert normal biology

Clonal marks and targeted sequencing were used here to describe how gene-specific mutations can achieve high mutant allele frequencies within the human colonic epithelium. Importantly, this process is intimately linked to the homeostatic behaviour of the resident stem cell population, which can be described by neutral drift theory (Lopez-Garcia *et al.*, 2010; Snippert, van der Flier, *et al.*, 2010; Baker *et al.*, 2014; Nicholson *et al.*, 2018). Each human colonic crypt contains an equipotent population of stem cells that all harbour the same probability of fixation. The total colonic MAF for neutral mutations is thus proportional to the mutation rate. However, certain gene-specific mutations can increase the probability of fixation of stem cell-derived clones within the crypt by biased stem cell replacement. Such an effect was found for loss of KDM6A and, to a greater extent, STAG2, whereby stem cells with mutations in the latter were found to nearly always populate their resident crypt. Biased replacement is thus a mechanism by which mutations can achieve a higher mutant allele burden than would be possible from mutation rate alone. Simply describing the mutational burden of a tissue by sequencing or otherwise would fail to capture this mechanism, as this only describes the final outcome and would thus overestimate the gene-specific mutation rate.

Once fixed within a crypt, mutant clones can expand in the epithelium by crypt fission. Importantly, the present study precisely quantified the homeostatic rate of crypt fission in the human colon. Against this baseline, it was found that loss of STAG2, KDM6A and PTEN as well as stabilisation of p53 and activation of KRAS can accelerate crypt fission to generate

large mutant fields within the colonic epithelium. The most dramatic effects were found for loss of KDM6A and activation of KRAS, with clones comprising tens and hundreds of crypts, respectively. This demonstrates how gene-specific mutations can contribute to a field cancerisation effect, which substantially increases the pool of cells available for subsequent mutations. Importantly, for description of this process only the mutation rate and the final MAF are required.

The potential effect of crypt fusion was also investigated. Strikingly, the baseline crypt fusion rate of human colonic crypts was very similar to the fission rate, suggesting a symmetric balance between the two processes that may act to conserve tissue size. A crypt cycle, as proposed in the 1980s (Totafurno, Bjerknes & Cheng, 1987), thus seems to exist, but the original hypothesis requires addition of crypt fusion. Also, the cycle duration is likely much higher than the 10-20 years estimated by Totafurno and colleagues, with the data presented here corresponding to approximately 150 years. Quantifying the fusion rate for KDM6A-negative crypts revealed that some mutations can alter this balance by promoting fission but not fusion. This suggests that the two processes may be driven by independent mechanisms. Of note, mathematical modelling by Dr Edward Morrissey revealed that increasing the fusion rate for KDM6A would only marginally limit the spread of mutant clones.

For every gene-specific mutation, the intra-crypt dynamics combined with the effect on expansion by crypt fission result in the total mutational burden within the colonic epithelium. Understanding the rates of these processes enables ranking of different mutations irrespective of the specific cellular mechanism through which they arise. This may inform cancer prevention strategies to restrict the spread of pro-oncogenic mutations, which may limit cancer risk. Importantly, these insights cannot be gained from simple description of the total mutational burden in aged human colon, as this does not generate any insights into the mechanisms and timings of mutation fixation and spread.

## 6.3 Towards understanding CRC initiation and incidence

Approximating the frequency of clones detected for the X-linked clonal marks to 1:10,000 crypts and the human genome to 20,000 protein encoding genes, it follows that every colonic crypt in a 60-year old individual carries on average two mutations causing loss of function of two different alleles. Furthermore, alterations known to drive CRC are present in normal colon – 1 in 10,000 crypts harbours stabilised p53, roughly the same frequency have lost *PTEN* and just over 10% of people carry *KRAS* activating mutations in patches comprising several hundred crypts. The colonic epithelium of aged humans is thus a mosaic of different mutations. In line with these findings, a patchwork of mutations was also reported by Lee-Six et al., (2018) and Blokzijl et al. (2016). The former study claims that 1% of crypts harbour a CRC driver mutation. Taking all these observations into account, the obvious question that arises is: Why is bowel cancer not more common?

To answer this question, a comprehensive description of how common CRC drivers spread in the normal human colonic epithelium is required. The present study was focused on describing the ways in which mutations can achieve high mutant allele frequency in the human colonic epithelium. In terms of CRC driver genes however, only three alterations were considered: stabilisation of p53, loss of *PTEN* and activation of *KRAS* via mutations in codons 12 and 13.

### 6.3.1 Proposed multiplex method

Considering the large tail of genes commonly mutated in CRC (Figure 1.6) multiplied by the fact that most genes exhibit at least a number of mutational hotspots means that capturing the full landscape of CRC-associated mutations in normal human colon by *in situ* detection would not be feasible. Equally, from the moderate expansion effects seen for *PTEN* and p53 it can be predicted that a lot of the clones in question will be very small, comprising only a few crypts. Even in a relatively small section with 10,000 crypts, this would generate a MAF below 0.1% and therefore fail to exceed the detection threshold for the mimic amplicon targeted sequencing method. Therefore, a refined hybrid methodology may be called for. For genes harbouring a few mutational hotspots, multiplex sequencing may represent a straightforward way of detection. To increase sensitivity, sections could be cut into several pieces, effectively

---

representing an intermediate method between single crypt sequencing and the bulk approach used for *KRAS* in chapter 5. For tumour suppressor genes exhibiting mutations along the entire length, immunohistochemical detection may still be the most suitable method. Alternatively, surrogate markers may be revisited. While attempts to use IHC detection of DUSP6 and p-4EBP were not successful here, detection of upregulation of *NOTUM* by RNAScope® may act a reliable indicator of loss of *APC*, the most common alteration in CRC. In addition to gene loss and point mutation, copy number changes also need consideration. For this purpose, the recently developed iFISH, a DNA hybridisation method which promises to target an ever-expanding number of genomic loci, may be used (Gelali *et al.*, 2019).

Taken together, such studies will enable a comprehensive description of how key CRC driver mutations spread in the normal human colon. Comparing these mutant allele burdens to those found in tumours will aid in understanding the probability of conversion for different cancer drivers in the colon. For example, mutations that are very common in tumours but cannot be found in normal tissue could be indicative of certain transformation.

## 6.4 Extrinsic factors

Cancer is a disease caused by somatic mutations and it therefore makes sense that studies have so far focused on understanding mutation accumulation in normal tissues. To a large extent, this is thought to occur as part of normal ageing, whereby erroneous DNA replication generates mutations. However, mammalian tissues are not just a collection of isolated cells whose biology is entirely dictated by intrinsic processes. They are part of a multicellular organism and therefore heavily influenced by the physiological state of the cells in their immediate surroundings as well as the body as a whole and the environment it lives in. Cell extrinsic factors can heavily influence the state of tissues and the risk of resident cells progressing to cancer. There are in principle two ways in which extrinsic factors can affect the development of tumours: They can act as mutagens or they can generate a pro- or anti-tumorigenic milieu. Two classic examples for the former are UV radiation from the sun and carcinogens inhaled in air, which have the strongest effect on the exposed tissues – the skin and the lung. Indeed, the

---

number of mutations found in sun exposed skin is considerably higher than in non-sun exposed skin (Yizhak *et al.*, 2019). Cigarette smoke is thought to contain up to 250 carcinogens (Hussain *et al.*, 2019). Apart from lung cancer, smoking is associated with a number of cancers including oesophageal, stomach, pancreas, kidney and bladder cancer (Hussain *et al.*, 2019). Perhaps more intriguing and certainly more complex is the second way in which extrinsic factors can promote or hamper the development of cancer, namely by generating a pro-tumourigenic setting, within cells or the organism as a whole. Instead of directly causing mutations, these factors may increase the proliferation of cells leading to abnormal growths, alter the numbers of stem cells in tissues or influence the immune system such that it can more or less effectively eliminate cancerous cells. A compelling example for such a mechanism is HPV infection causing hyperproliferation in the human female cervix (Burd, 2003). The next sections will discuss pro-tumorigenic extrinsic factors in the intestine.

#### 6.4.1 Niche – stem cell interactions in CRC

As discussed in the introduction, the self-renewal, pluripotency and proliferation of intestinal stem cells at the crypt base is heavily regulated by their niche. While this complex arrangement of cellular and extracellular components is key to maintenance of homeostasis as well as repair following injury, it can also play a key role in the development of cancer, thereby acting as an extrinsic factor. Early during tumourigenesis, the stroma is thought to be anti-tumourigenic, however, this changes as more alterations are acquired (de Visser, Eichten & Coussens, 2006). Additionally, transformation is associated with progressive loss of normal niche-dependency in ISCs (Fujii *et al.*, 2016). Rather, tumour cells rely on a so-called tumour microenvironment, a combination of altered surrounding stroma and infiltrating immune cells, which then plays a key role in supporting tumour cell proliferation, survival and metastasis. Importantly, pro-tumourigenic signalling is bidirectional, whereby stromal cells can be modulated by transformed epithelial cells and vice versa (Peddareddigari, Wang & Dubois, 2010). For example, in homeostasis hedgehog signalling provided by the epithelium regulates the stroma (Kosinski *et al.*, 2010), while adenomatous polyps have recently been shown to secrete IL-33, which stimulates their niche (Maywald *et al.*, 2015). More generally, cancer cells can release signals to activate so-called tumour-associated macrophages, which in turn secrete VEGF to

---

promote angiogenesis and facilitate epithelial-mesenchymal transition of epithelial cells, thus contributing to metastasis (Barbera-Guillem *et al.*, 2002; Pollard, 2004; Green *et al.*, 2009; Jedinak, Dudhgaonkar & Sliva, 2010). Equally, cancer-associated fibroblasts (CAFs), which are thought to arise from local tissue fibroblasts (Mueller *et al.*, 2007), can produce Wnts to promote a cancer stem cell phenotype (Aizawa *et al.*, 2019). Furthermore, CAFs at the tumour edge can set the stage for blood vessel generation and invasion of adenocarcinomas (Sivridis, Giatromanolaki & Koukourakis, 2005). In addition, cancer cells can instruct CAFs to produce metalloproteinases, which causes ECM remodelling and promotes invasion (Shay, Lynch & Fingleton, 2015). In summary, CRC progression is driven by a complex interplay between transformed epithelial cells and their tumour microenvironment.

#### 6.4.2 Colitis-associated CRC

A key example for the potent tumour-promoting effect of extrinsic factors is the association between IBD, which is the umbrella term for the chronic inflammatory conditions ulcerative colitis and Crohn's disease, and CRC. IBD is thought to result from aberrant immune responses to microflora triggered by various environmental insults in the context of susceptibility provided by the genetic background of the host (Strober, Fuss & Mannon, 2007). It is characterised by recurrent cycles of inflammation, ulceration and repair and associated with significantly increased risk for CRC (Eaden, Abrams & Mayberry, 2001; Canavan, Abrams & Mayberry, 2006), which in this scenario is termed colitis-associated CRC (CAC) (Feagins, Souza & Spechler, 2009). First studies into potential mechanisms were performed in mice, whereby it was found that a single injection of azoxymethane resulted in tumours when administered on a colitis background (Okayasu *et al.*, 1996). Since then, investigation of human CAC tumours as well as mouse models have provided more insights into putative mechanisms. Importantly, while sporadic CRC is usually initiated by alterations in Wnt signalling, such as APC mutations, this is not the case for CAC (Terzić *et al.*, 2010). Instead, chronic inflammation seems to drive the activation of Wnt target genes in the absence of APC mutations (Berg *et al.*, 1996; Oguma *et al.*, 2008). In addition, chronic inflammation seems to cause DNA damage directly and indirectly, as inflammatory cells such as activated macrophages and neutrophils

are potent sources of damaging reactive oxygen species (ROS) and reactive nitrogen intermediates (RNI) (Meira *et al.*, 2008; Westbrook *et al.*, 2009). Furthermore, inflammatory cytokines may prime epithelial cells to ROS and RNI production (Goodman *et al.*, 2004; Shaked *et al.*, 2012). This inflammation-associated genotoxicity is exemplified by the observation of p53 mutations in inflamed but otherwise normal human mucosa (Kraus & Arber, 2009). Inflammatory signalling may also stimulate hyperproliferation of non-initiated stem cells, thereby increasing the risk of transformation as well as cause tumour-promoting epigenetic changes and inhibit apoptosis (Atreya & Neurath, 2008; Chiba, Marusawa & Ushijima, 2012; De Lerma Barbaro *et al.*, 2014). Importantly, chronic inflammation also affects the stroma, whereby stromal cells react to immune cell-derived cytokines, which can generate a pro-tumorigenic milieu (discussed above) (Fritsch *et al.*, 1997; Lang *et al.*, 2009; Owens & Simmons, 2013). This is exemplified by the finding that activation of intestinal stromal cells is sufficient to initiate intestinal neoplasia in a mouse model of Crohn's disease (Roulis *et al.*, 2011). Finally, the repeated cycles of intestinal injury and repair (by crypt fission) that occur in colitis can favour the expansion of mutant clones, therefore causing field cancerisation, which promotes tumour initiation (Kuraishy, Karin & Grivennikov, 2011).

In summary, while the accumulation of somatic mutations in epithelial (stem) cells is undoubtedly key to tumour initiation, the cells in their immediate surroundings can have a potent effect and may in some cases even initiate cancer in the absence of driver mutations.

### 6.4.3 Other extrinsic factors influencing CRC

#### 6.4.3.1 The microbiome

Importantly, IBD and CRC are both linked to alterations in the microbiome, which is the collection of over 100 billion bacteria found in the human intestine, the largest proportion of which resides in the colon (Drewes, Housseau & Sears, 2016; Chen, Pitmon & Wang, 2017). Interestingly, specific strains of bacteria can be associated with advanced colorectal adenomas and carcinomas (Zackular *et al.*, 2013; Feng *et al.*, 2015). Animal studies have revealed potential mechanisms of CRC promotion by the microbiome, which include modulation of the

---

host's immune system, production of cancer-associated metabolites and release of genotoxic factors (Dahmus *et al.*, 2018).

#### 6.4.3.2 Diet and obesity

Red meat consumption, vitamin D deficiency, obesity and Western diet have all been tied to increased CRC risk in humans (Shaukat *et al.*, 2017; Turner & Lloyd, 2017; Zheng *et al.*, 2017; Bultman, 2018; McCullough *et al.*, 2018). Mouse studies have shed some light on potential mechanisms. For high fat diet these may include increased numbers and tumorigenicity of stem cells as well as chronic inflammation, the potential effects of which were discussed above (van der Heijden *et al.*, 2015; Beyaz *et al.*, 2016). The haemoglobin in red meat is thought to cause DNA damage (Bastide, Pierre & Corpet, 2011), while vitamin D is thought to inhibit cell proliferation, an effect that was recently demonstrated in human organoids (Fernández-Barral *et al.*, 2019). Interestingly, one study suggests that the tumour-promoting effect of high fat diet can be reduced in mice by walnut consumption, potentially by reducing systemic inflammation (Guan *et al.*, 2018).

#### 6.4.4 Future directions for investigation of extrinsic factors

Considering the (non-exhaustive) description of tumour-promoting interactions between epithelial cells and their stromal niche as well as the immune system given above, the work presented in this thesis took an integrationist approach, whereby potential effects associated with intercompartmental signalling between stem cells and the stroma were not investigated on a mechanistic level. The possibility that gene-specific mutations in epithelial cells, as observed for clonal marks, alter their environment which in turn affects stem cell dynamics, cannot be excluded. The absence of marked increases in immune infiltration around clonal patches is an indication that such mechanisms, if at play, may only be marginal. However, the subject requires more investigation. Importantly, our *en face* FFPE colon tissue does not represent a suitable system to study niche-epithelium interactions on a mechanistic level. First of all, a large fraction of the signalling occurs via soluble factors, which are not visualisable in this material. Furthermore, specific stromal cells can be found at distinct locations along the crypt axis

---

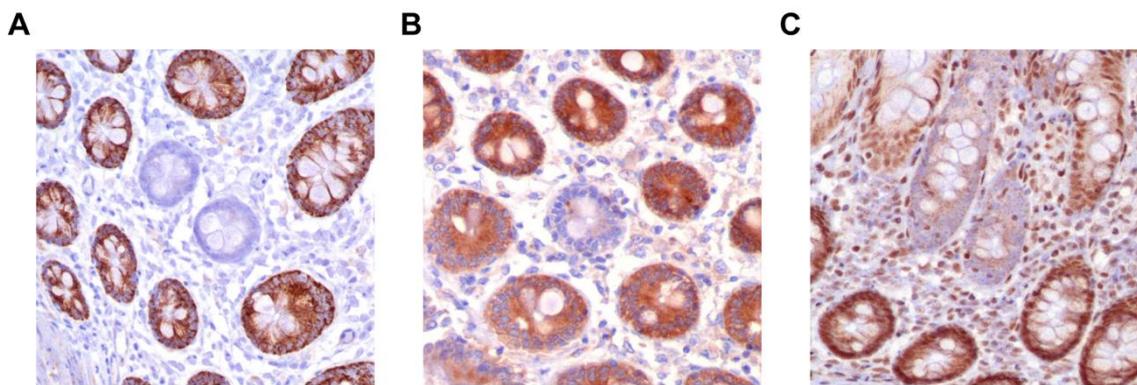
(Halpern *et al.*, 2019), the study of which requires longitudinal sections. Therefore, a different system may be more suitable to investigate the effect of gene-specific mutations on the stroma or immune compartment. For example, organoids modified by CRISPR/Cas9 could be co-cultured directly with cells of interest, seeded on a mesenchymal cell monolayer or interactions investigated indirectly in a transwell assay, all of which are published methods (Pastuła *et al.*, 2016; Pastuła & Marcinkiewicz, 2019).

While mechanisms of stem cell-niche interplay may not be investigated on a mechanistic level with clonal marks, insights into how extrinsic (lifestyle) factors affect stem cell behaviour and mutation accumulation may be gained. Importantly, to date little is known about how extrinsic factors such as obesity, inflammation and an altered microbiome may affect intestinal stem cell behaviour and therefore promote cancer in humans. For lifestyle factors such as BMI and smoking, the current patient cohort could be stratified according to these variables. For inflammatory conditions as well as microbiome studies, a cohort of patients would need to be recruited. Using the neutral clonal marks mPAS, MAOA and NONO the association of different extrinsic factors with changes in intestinal stem cell dynamics could be investigated. This includes intra-crypt dynamics as well as fission and fusion. Considering all marks, significantly higher clone frequencies may be indicative of changes in mutation rates. Elevated rates of fission and fusion may provide a mechanism for accelerated field cancerisation. Of note, one well-established such connection exists between increased rates of crypt fission and colitis (Cheng *et al.*, 1986). This may contribute to the extensive field cancerisation found in patients with Crohn's Ileocolitis.

Importantly, the effects of extrinsic factors on stem cell behaviours may be subtle. Considering the spread of human data, they may only be detectable in very large patient cohorts of several hundred people. The Winton laboratory currently holds anonymised clinical data for over 150 individuals in the patient cohort and is acquiring new informative samples at a rate of approximately 100 per year, so analysis of stratified patient data will be possible within the next few years.

As an alternative approach to investigating the effect of extrinsic factors on intestinal stem cell dynamics, we are using the clonal marks MAOA, HDAC6 and STAG2 in FFPE colonic tissue

sections from a colony of green monkeys (*Chlorocebus aethiops*) in St. Kitt's and Nevis. The animals are genetically outbred but all adhere to the same lifestyle. Therefore, comparison of the spread of human data with that of monkey data may enable quantification of the individual contribution of intrinsic (DNA replication errors) and extrinsic (lifestyle) factors to mutation accumulation in the colon.



**Figure 6.1 Clonal marks can be used in green monkey tissue.**

*Representative images of IHC on green monkey FFPE colonic tissue sections.*

*(A) Two adjacent MAOA-negative crypts in en face section.*

*(B) HDAC6-negative crypt in en face section.*

*(C) Two adjacent STAG2-negative crypts in obliquely sectioned tissue.*

Taken together, these studies will start to reveal the individual contributions of different lifestyle factors on human intestinal stem cell behaviour, which will aid in understanding their contribution to cancer development.

## 6.5 Cancer prevention

A thorough understanding of the mechanisms of lateral spread of mutations as well as the contribution of different extrinsic factors to mutation accumulation and expansion may inform cancer prevention strategies. A number of evidence-based recommendations have to varying

---

extents already reached the general public including the beneficial effects of reducing red meat consumption and maintaining a healthy weight. Research may also identify the most beneficial timing of specific chemopreventive interventions. For example, a drug that reduces crypt fission may be given to individuals at a certain age such that their theoretical onset age of CRC is shifted beyond natural human lifespan.

Importantly, all these endeavours are counteracted by the inevitable imperfection of DNA replication. Even in the healthiest of individuals, somatic mutations accumulate with age. It may be possible to reduce their acquisition and spread by lifestyle modification and interventions, but not entirely prevent it. In a sense, this can be seen as the cost of evolution, as it is based on DNA replication errors.

This thesis started with a description of the beginning of life with the most potent of stem cells, the zygote. The work presented in the results chapters contributes to a deeper understanding of the processes that may prevent as well as accelerate mutation accumulation in the human colonic epithelium. Now, by discussing the potential impact of extrinsic factors and the inevitability of mutation accumulation, the work has been put into a wider context. I would like to conclude by adapting a famous quote from 20<sup>th</sup> century evolutionary biologist Theodosius Dobzhansky: Stem cell biology makes a lot more sense in the light of evolution.



## References

- Affolter, M., Bellusci, S., Itoh, N., Shilo, B., Thiery, J.-P., & Werb, Z. (2003) 'Tube or Not Tube: Remodeling Epithelial Tissues by Branching Morphogenesis', *Developmental Cell*, 4(1), pp. 11–18.
- Agger, K., Cloos, P. A. C., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A. E., & Helin, K. (2007) 'UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development', *Nature*, 449, p. 731.
- Ahmad, M. K., Abdollah, N. A., Shafie, N. H., Yusof, N. M., & Razak, S. R. A. (2018) 'Dual-specificity phosphatase 6 (DUSP6): a review of its molecular characteristics and clinical relevance in cancer', *Cancer biology & medicine*, 15(1), pp. 14–28.
- Ahuja, N., Li, Q., Mohan, A. L., Baylin, S. B., & Issa, J.-P. J. (1998) 'Aging and DNA Methylation in Colorectal Mucosa and Cancer', *Cancer Research*, 58(23), pp. 5489–5494.
- Aizawa, T., Karasawa, H., Funayama, R., Shiota, M., Suzuki, T., Maeda, S., Suzuki, H., Yamamura, A., Naitoh, T., Nakayama, K., & Unno, M. (2019) 'Cancer-associated fibroblasts secrete Wnt2 to promote cancer progression in colorectal cancer', *Cancer medicine*. 2019/08/29. John Wiley and Sons Inc., 8(14), pp. 6370–6382.
- Alimonti, A., Carracedo, A., Clohessy, J. G., Trotman, L. C., Nardella, C., Egia, A., Salmena, L., Sampieri, K., Haveman, W. J., Brogi, E., Richardson, A. L., Zhang, J., & Pandolfi, P. P. (2010) 'Subtle variations in Pten dose determine cancer susceptibility', *Nature genetics*, 42(5), pp. 454–458.
- Allaire, J. M., Crowley, S. M., Law, H. T., Chang, S.-Y., Ko, H.-J., & Vallance, B. A. (2018) 'The Intestinal Epithelium: Central Coordinator of Mucosal Immunity', *Trends in Immunology*, 39(9), pp. 677–696.

- Altmann, G. G. (1983) 'Morphological observations on mucus-secreting nongoblet cells in the deep crypts of the rat ascending colon.', *American Journal of Anatomy*, 167(1), pp. 95–117.
- Aoyagi, S. & Archer, T. K. (2005) 'Modulating molecular chaperone Hsp90 functions through reversible acetylation', *Trends in Cell Biology*, 15(11), pp. 565–567.
- Araki, K., Ogata, T., Kobayashi, M., & Yatani, R. (1995) 'A morphological study on the histogenesis of human colorectal hyperplastic polyps', *Gastroenterology*, 109(5), pp. 1468–1474.
- Arming, S., Wipfler, D., Mayr, J., Merling, A., Vilas, U., Schauer, R., Schwartz-Albiez, R., & Vlasak, R. (2011) 'The human Cas1 protein: A sialic acid-specific O-acetyltransferase?', *Glycobiology*, 21(5), pp. 553–564.
- Atreya, R. & Neurath, M. (2008) 'Signaling molecules: the pathogenic role of the IL-6/STAT3 trans signaling pathway in intestinal inflammation and in colonic cancer.', *Current Drug Targets*, 9(5), pp. 369–374.
- Baker, A.-M., Cereser, B., Melton, S., Fletcher, A. G., Rodriguez-Justo, M., Tadrous, P. J., Humphries, A., Elia, G., McDonald, S. A. C., Wright, N. A., Simons, B. D., Jansen, M., & Graham, T. A. (2014) 'Quantification of Crypt and Stem Cell Evolution in the Normal and Neoplastic Human Colon', *Cell Reports*, 8(4), pp. 940–947.
- Baker, A.-M., Huang, W., Wang, X.-M. M., Jansen, M., Ma, X.-J., Kim, J., Anderson, C. M., Wu, X., Pan, L., Su, N., Luo, Y., Domingo, E., Heide, T., Sottoriva, A., Lewis, A., Beggs, A. D., Wright, N. A., Rodriguez-Justo, M., Park, E., Tomlinson, I., & Graham, T. A. (2017) 'Robust RNA-based in situ mutation detection delineates colorectal cancer subclonal evolution', *Nature Communications*, 8(1), p. 1998.
- Baker, A.-M., Gabbutt, C., Williams, M. J., Cereser, B., Jawad, N., Rodriguez-Justo, M., Jansen, M., Barnes, C. P., Simons, B. D., McDonald, S. A. C., Graham, T. A., & Wright, N. A. (2019) 'Crypt fusion as a homeostatic mechanism in the human colon', *Gut*, p. gutjnl-

2018-317540.

Baker, S. J., Preisinger, A. C., Jessup, J. M., Paraskeva, C., Markowitz, S., Willson, J. K. V., Hamilton, S., & Vogelstein, B. (1990) 'p53 Gene Mutations Occur in Combination with 17p Allelic Deletions as Late Events in Colorectal Tumorigenesis', *Cancer Research*, 50(23), pp. 7717–7722.

Balbás-Martínez, C., Sagrera, A., Carrillo-de-Santa-Pau, E., Earl, J., Márquez, M., Vazquez, M., Lapi, E., Castro-Giner, F., Beltran, S., Bayés, M., Carrato, A., Cigudosa, J. C., Domínguez, O., Gut, M., Herranz, J., Juanpere, N., Kogevinas, M., Langa, X., López-Knowles, E., Lorente, J. A., Lloreta, J., Pisano, D. G., Richart, L., Rico, D., Salgado, R. N., Tardón, A., Chanock, S., Heath, S., Valencia, A., Losada, A., Gut, I., Malats, N., & Real, F. X. (2013) 'Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy', *Nature genetics*, 45(12), pp. 1464–1469.

Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., McQuaid, S., Gray, R. T., Murray, L. J., Coleman, H. G., James, J. A., Salto-Tellez, M., & Hamilton, P. W. (2017) 'QuPath: Open source software for digital pathology image analysis', *Scientific Reports*, 7(1), p. 16878.

Barbera-Guillem, E., Nyhus, J. K., Wolford, C. C., Friece, C. R., & Sampsel, J. W. (2002) 'Vascular Endothelial Growth Factor Secretion by Tumor-infiltrating Macrophages Essentially Supports Tumor Angiogenesis, and IgG Immune Complexes Potentiate the Process', *Cancer Research*, 62(23), pp. 7042 LP – 7049.

Barker, N., van Es, J. H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegbarth, A., Korving, J., Begthel, H., Peters, P. J., & Clevers, H. (2007) 'Identification of stem cells in small intestine and colon by marker gene Lgr5', *Nature*, 449(7165), pp. 1003–1007.

Barker, N., Ridgway, R. A., van Es, J. H., van de Wetering, M., Begthel, H., van den Born, M., Danenberg, E., Clarke, A. R., Sansom, O. J., & Clevers, H. (2008) 'Crypt stem cells as the cells-of-origin of intestinal cancer', *Nature*, 457(7229), pp. 608–611.

Bastide, N. M., Pierre, F. H. F., & Corpet, D. E. (2011) 'Heme Iron from Meat and Risk of

- Colorectal Cancer: A Meta-analysis and a Review of the Mechanisms Involved', *Cancer Prevention Research*, 4(2), pp. 177 LP – 184.
- Behjati, S., Huch, M., van Boxtel, R., Karthaus, W., Wedge, D. C., Tamuri, A. U., Martincorena, I., Petljak, M., Alexandrov, L. B., Gudem, G., Tarpey, P. S., Roerink, S., Blokker, J., Maddison, M., Mudie, L., Robinson, B., Nik-Zainal, S., Campbell, P., Goldman, N., van de Wetering, M., Cuppen, E., Clevers, H., & Stratton, M. R. (2014) 'Genome sequencing of normal cells reveals developmental lineages and mutational processes', *Nature*, 513(7518), pp. 422–425.
- Berg, D. J., Davidson, N., Kühn, R., Müller, W., Menon, S., Holland, G., Thompson-Snipes, L., Leach, M. W., & Rennick, D. (1996) 'Enterocolitis and colon cancer in interleukin-10-deficient mice are associated with aberrant cytokine production and CD4(+) TH1-like responses', *The Journal of clinical investigation*, 98(4), pp. 1010–1020.
- Bergmann, O., Spalding, K. L., & Frisén, J. (2015) 'Adult Neurogenesis in Humans', *Cold Spring Harbor Perspectives in Biology*, 7(7).
- Bertos, N. R., Gilquin, B., Chan, G. K. T., Yen, T. J., Khochbin, S., & Yang, X.-J. (2004) 'Role of the Tetradecapeptide Repeat Domain of Human Histone Deacetylase 6 in Cytoplasmic Retention', *Journal of Biological Chemistry*, 279(46), pp. 48246–48254.
- Betschinger, J. & Knoblich, J. A. (2004) 'Dare to Be Different: Asymmetric Cell Division in *Drosophila*, *C. elegans* and Vertebrates', *Current Biology*, 14(16), pp. R674–R685.
- Beyaz, S., Mana, M. D., Roper, J., Kedrin, D., Saadatpour, A., Hong, S.-J., Bauer-Rowe, K. E., Xifaras, M. E., Akkad, A., Arias, E., Pinello, L., Katz, Y., Shinagare, S., Abu-Remaileh, M., Mihaylova, M. M., Lamming, D. W., Dogum, R., Guo, G., Bell, G. W., Selig, M., Nielsen, G. P., Gupta, N., Ferrone, C. R., Deshpande, V., Yuan, G.-C., Orkin, S. H., Sabatini, D. M., & Yilmaz, Ö. H. (2016) 'High-fat diet enhances stemness and tumorigenicity of intestinal progenitors', *Nature*, 531(7592), pp. 53–58.
- Bird, R. P. (1987) 'Observation and quantification of aberrant crypts in the murine colon

- treated with a colon carcinogen: Preliminary findings', *Cancer Letters*, 37(2), pp. 147–151.
- Bjerknes, M. & Cheng, H. (1999) 'Clonal analysis of mouse intestinal epithelial progenitors', *Gastroenterology*, 116(1), pp. 7–14.
- Blanpain, C. & Fuchs, E. (2006) 'Epidermal stem cells of the skin', *Annual review of cell and developmental biology*, 22, pp. 339–373.
- Blokzijl, F., de Ligt, J., Jager, M., Sasselli, V., Roerink, S., Sasaki, N., Huch, M., Boymans, S., Kuijk, E., Prins, P., Nijman, I. J., Martincorena, I., Mokry, M., Wiegerinck, C. L., Middendorp, S., Sato, T., Schwank, G., Nieuwenhuis, E. E. S., Verstegen, M. M. A., van der Laan, L. J. W., de Jonge, J., IJzermans, J. N. M., Vries, R. G., van de Wetering, M., Stratton, M. R., Clevers, H., Cuppen, E., & van Boxtel, R. (2016) 'Tissue-specific mutation accumulation in human adult stem cells during life', *Nature*, 538(7624), pp. 260–264.
- Bond, C. S. & Fox, A. H. (2009) 'Paraspeckles: nuclear bodies built on long noncoding RNA', *The Journal of Cell Biology*, 186(5), pp. 637–644.
- Borras, E., San Lucas, F. A., Chang, K., Zhou, R., Masand, G., Fowler, J., Mork, M. E., You, Y. N., Taggart, M. W., McAllister, F., Jones, D. A., Davies, G. E., Edelmann, W., Ehli, E. A., Lynch, P. M., Hawk, E. T., Capella, G., Scheet, P., & Vilar, E. (2016) 'Genomic Landscape of Colorectal Mucosa and Adenomas', *Cancer prevention research (Philadelphia, Pa.)*, 9(6), pp. 417–427.
- Bortolato, M., Chen, K., & Shih, J. C. (2008) 'Monoamine oxidase inactivation: From pathophysiology to therapeutics', *Advanced Drug Delivery Reviews*, 60(13–14), pp. 1527–1533.
- Brannon, A. R., Vakiani, E., Sylvester, B. E., Scott, S. N., McDermott, G., Shah, R. H., Kania, K., Viale, A., Oschwald, D. M., Vacic, V., Emde, A.-K., Cercek, A., Yaeger, R., Kemeny, N. E., Saltz, L. B., Shia, J., D'Angelica, M. I., Weiser, M. R., Solit, D. B., & Berger, M. F. (2014) 'Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions', *Genome biology*, 15(8), p. 454.
- Brosh, R. & Rotter, V. (2009) 'When mutants gain new powers: news from the mutant p53

field', *Nature Reviews Cancer*, 9, p. 701.

Bruens, L., Ellenbroek, S. I. J., van Rheenen, J., & Snippert, H. J. (2017) 'In Vivo Imaging Reveals Existence of Crypt Fission and Fusion in Adult Mouse Intestine', *Gastroenterology*, 153(3), pp. 674-677.e3.

Buczacki, S. J. A., Zecchini, H. I., Nicholson, A. M., Russell, R., Vermeulen, L., Kemp, R., & Winton, D. J. (2013) 'Intestinal label-retaining cells are secretory precursors expressing Lgr5', *Nature*, 495(7439), pp. 65–69.

Buffet, C., Hecale-Perlemoine, K., Bricaire, L., Dumont, F., Baudry, C., Tissier, F., Bertherat, J., Cochand-Priollet, B., Raffin-Sanson, M.-L., Cormier, F., & Groussin, L. (2017) 'DUSP5 and DUSP6, two ERK specific phosphatases, are markers of a higher MAPK signaling activation in BRAF mutated thyroid cancers', *PloS one*, 12(9), pp. e0184861–e0184861.

Bultman, S. J. (2018) 'A Reversible Epigenetic Link between Obesity and Cancer Risk', *Trends in Endocrinology & Metabolism*, 29(8), pp. 529–531.

Burd, E. M. (2003) 'Human papillomavirus and cervical cancer', *Clinical microbiology reviews*, 16(1), pp. 1–17.

Bussey, H. (1975) 'Familial polyposis coli: Family studies, histopathology, differential diagnosis and results of treatment', in. Baltimore: Johns Hopkins University Press, pp. 47–49.

Cairnie, A. B. & Millen, B. H. (1975) 'Fission of crypts in the small intestine of the irradiated mouse.', *Cell Tissue Kinet.*, 8(2), pp. 189–96.

Campbell, F., Appleton, M. A. C., Fuller, C. E., Greeff, M. P., Hallgrimsson, J., Katoh, R., Ng, O. L. I., Satir, A., Williams, G. T., & Williams, E. D. (1994) 'Racial variation in the O-acetylation phenotype of human colonic mucosa', *The Journal of Pathology*, 174(3), pp. 169–174.

Campbell, F., Geraghty, J. M., Appleton, M. A. C., Williams, E. D., & Williams, G. T. (1998) 'Increased stem cell somatic mutation in the non-neoplastic colorectal mucosa of patients with

familial adenomatous polyposis', *Human Pathology*, 29(12), pp. 1531–1535.

Campos, F. G. (2014) 'Surgical treatment of familial adenomatous polyposis: dilemmas and current recommendations', *World journal of gastroenterology*, 20(44), pp. 16620–16629.

Canavan, C., Abrams, K. R., & Mayberry, J. (2006) 'Meta-analysis: colorectal and small bowel cancer risk in patients with Crohn's disease', *Alimentary Pharmacology & Therapeutics*. John Wiley & Sons, Ltd (10.1111), 23(8), pp. 1097–1104.

Cancer Research UK (2015) *Cancer Research UK*. available at:<https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer2015>.

Canudas, S. & Smith, S. (2009) 'Differential regulation of telomere and centromere cohesion by the Scc3 homologues SA1 and SA2, respectively, in human cells', *The Journal of Cell Biology*, 187(2), pp. 165–173.

Carlston, C. M., Bleyl, S. B., Andrews, A., Meyers, L., Brown, S., Bayrak-Toydemir, P., Bale, J. F., & Botto, L. D. (2019) 'Expanding the genetic and clinical spectrum of the NONO-associated X-linked intellectual disability syndrome', *American Journal of Medical Genetics Part A*, 179(5), pp. 792–796.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017) 'Stan: A Probabilistic Programming Language', *Journal of Statistical Software; Vol 1, Issue 1 (2017)* .

Carrel, L., Cottle, A. A., Goglin, K. C., & Willard, H. F. (1999) 'A first-generation X-inactivation profile of the human X chromosome', *Proceedings of the National Academy of Sciences*, 96(25), pp. 14440–14444.

Chen, J., Pitmon, E., & Wang, K. (2017) 'Microbiome, inflammation and colorectal cancer', *Seminars in Immunology*, 32, pp. 43–53.

Cheng, D. T., Mitchell, T. N., Zehir, A., Shah, R. H., Benayed, R., Syed, A., Chandramohan, R., Liu, Z. Y., Won, H. H., Scott, S. N., Brannon, A. R., O'Reilly, C., Sadowska, J., Casanova, J., Yannes, A., Hechtman, J. F., Yao, J., Song, W., Ross, D. S., Oultache, A.,

- Dogan, S., Borsu, L., Hameed, M., Nafa, K., Arcila, M. E., Ladanyi, M., & Berger, M. F. (2015) 'Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology', *The Journal of molecular diagnostics : JMD*, 17(3), pp. 251–264.
- Cheng, H., Bjercknes, M., Amar, J., & Gardiner, G. (1986) 'Crypt production in normal and diseased human colonic epithelium', *The Anatomical Record*, 216(1), pp. 44–48.
- Cheng, H. & Leblond, C. P. (1974a) '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(4), pp. 537–561.
- Cheng, H. & Leblond, C. P. (1974b) 'Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine I. Columnar cell', *American Journal of Anatomy*, 141(4), pp. 461–479.
- Cheng, H., McCulloch, C., & Bjercknes, M. (1986) 'Effects of 30% intestinal resection on whole population cell kinetics of mouse intestinal epithelium', *The Anatomical Record*, 215(1), pp. 35–41.
- Chiba, T., Marusawa, H., & Ushijima, T. (2012) 'Inflammation-Associated Cancer Development in Digestive Organs: Mechanisms and Roles for Genetic and Epigenetic Modulation', *Gastroenterology*. Elsevier, 143(3), pp. 550–563.
- Cho, Y.-W., Hong, T., Hong, S., Guo, H., Yu, H., Kim, D., Guszczynski, T., Dressler, G. R., Copeland, T. D., Kalkum, M., & Ge, K. (2007) 'PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex', *The Journal of biological chemistry*, 282(28), pp. 20395–20406.
- Clair, W. H. S. & Osborne, J. W. (1986) 'Crypt fission in the small intestine of the rat', *The British Journal of Cancer. Supplement*, 7, pp. 39–41.
- Clarke, R. M. (1972) 'The effect of growth and of fasting on the number of villi and crypts in

the small intestine of the albino rat', *Journal of anatomy*, 112(Pt 1), pp. 27–33.

Clevers, H., Loh, K. M., & Nusse, R. (2014) 'An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control', *Science*, 346(6205), p. 1248012.

Cummins, A. G., Catto-Smith, A. G., Cameron, D. J., Couper, R. T., Davidson, G. P., Day, A. S., Hammond, P. D., Moore, D. J., & Thompson, F. M. (2008) 'Crypt Fission Peaks Early During Infancy and Crypt Hyperplasia Broadly Peaks During Infancy and Childhood in the Small Intestine of Humans', *Journal of Pediatric Gastroenterology and Nutrition*, 47(2), pp. 153–157.

Curtius, K., Wright, N. A., & Graham, T. A. (2017) 'An evolutionary perspective on field cancerization', *Nature Reviews Cancer*, 18, p. 19.

Dahmus, J. D., Kotler, D. L., Kastenberg, D. M., & Kistler, C. A. (2018) 'The gut microbiome and colorectal cancer: a review of bacterial pathogenesis', *Journal of gastrointestinal oncology*, 9(4), pp. 769–777.

Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J., & Shi, B. (2015) 'Breast cancer intrinsic subtype classification, clinical use and future trends', *American journal of cancer research*, 5(10), pp. 2929–2943.

Daniloski, Z. & Smith, S. (2017) 'Loss of Tumor Suppressor STAG2 Promotes Telomere Recombination and Extends the Replicative Lifespan of Normal Human Cells', *Cancer research*, 77(20), pp. 5530–5542.

Deakin, N. O. & Turner, C. E. (2014) 'Paxillin inhibits HDAC6 to regulate microtubule acetylation, Golgi structure, and polarized migration', *The Journal of cell biology*, 206(3), pp. 395–413.

Degirmenci, B., Valenta, T., Dimitrieva, S., Hausmann, G., & Basler, K. (2018) 'GLI1-expressing mesenchymal cells form the essential Wnt-secreting niche for colon stem cells', *Nature*, 558(7710), pp. 449–453.

Dieterle, C. P., Conzelmann, M., Linnemann, U., & Berger, M. R. (2004) 'Detection of Isolated Tumor Cells by Polymerase Chain Reaction-Restriction Fragment Length

- Polymorphism for K-ras Mutations in Tissue Samples of 199 Colorectal Cancer Patients’, *Clinical Cancer Research*, 10(2), pp. 641–650.
- Dietrich, D., Uhl, B., Sailer, V., Holmes, E. E., Jung, M., Meller, S., & Kristiansen, G. (2013) ‘Improved PCR Performance Using Template DNA from Formalin-Fixed and Paraffin-Embedded Tissues by Overcoming PCR Inhibition’, *PLOS ONE*, 8(10), p. e77771.
- Ding, G., Liu, H.-D., Huang, Q., Liang, H.-X., Ding, Z.-H., Liao, Z.-J., & Huang, G. (2013) ‘HDAC6 promotes hepatocellular carcinoma progression by inhibiting P53 transcriptional activity’, *FEBS Letters*, 587(7), pp. 880–886.
- Do, H. & Dobrovic, A. (2015) ‘Sequence Artifacts in DNA from Formalin-Fixed Tissues: Causes and Strategies for Minimization’, *Clinical Chemistry*, 61(1), pp. 64–71.
- Drewes, J. L., Housseau, F., & Sears, C. L. (2016) ‘Sporadic colorectal cancer: microbial contributors to disease prevention, development and therapy’, *British Journal Of Cancer*, 115(3), pp. 273–280.
- Dunford, A., Weinstock, D. M., Savova, V., Schumacher, S. E., Cleary, J. P., Yoda, A., Sullivan, T. J., Hess, J. M., Gimelbrant, A. A., Beroukhim, R., Lawrence, M. S., Getz, G., & Lane, A. A. (2017) ‘Tumor suppressor genes that escape from X-inactivation contribute to cancer sex bias’, *Nature genetics*, 49(1), pp. 10–16.
- Eaden, J. A., Abrams, K. R., & Mayberry, J. F. (2001) ‘The risk of colorectal cancer in ulcerative colitis: a meta-analysis’, *Gut*, 48(4), pp. 526–535.
- Eberle, F. C., Hanson, J. C., Killian, J. K., Wei, L., Ylaya, K., Hewitt, S. M., Jaffe, E. S., Emmert-Buck, M. R., & Rodriguez-Canales, J. (2010) ‘Immunoguided Laser Assisted Microdissection Techniques for DNA Methylation Analysis of Archival Tissue Specimens’, *The Journal of Molecular Diagnostics : JMD*, 12(4), pp. 394–401.
- Ekerot, M., Stavridis, M. P., Delavaine, L., Mitchell, M. P., Staples, C., Owens, D. M., Keenan, I. D., Dickinson, R. J., Storey, K. G., & Keyse, S. M. (2008) ‘Negative-feedback regulation of FGF signalling by DUSP6/MKP-3 is driven by ERK1/2 and mediated by Ets

- factor binding to a conserved site within the DUSP6/MKP-3 gene promoter', *The Biochemical journal*, 412(2), pp. 287–298.
- Ellegren, H. (2004) 'Microsatellites: simple sequences with complex evolution', *Nature Reviews Genetics*, 5(6), pp. 435–445.
- Emmert-Buck, M. R., Bonner, R. F., Smith, P. D., Chuaqui, R. F., Zhuang, Z., Goldstein, S. R., Weiss, R. A., & Liotta, L. A. (1996) 'Laser Capture Microdissection', *Science*, 274(5289), pp. 998–1001.
- van Es, J. H., Sato, T., van de Wetering, M., Lyubimova, A., Yee Nee, A. N., Gregorieff, A., Sasaki, N., Zeinstra, L., van den Born, M., Korving, J., Martens, A. C. M., Barker, N., van Oudenaarden, A., & Clevers, H. (2012) 'Dll1+ secretory progenitor cells revert to stem cells upon crypt damage', *Nat Cell Biol*, 14(10), pp. 1099–1104.
- Espina, V., Wulfkühle, J. D., Calvert, V. S., VanMeter, A., Zhou, W., Coukos, G., Geho, D. H., Petricoin, E. F., & Liotta, L. A. (2006) 'Laser-capture microdissection', *Nature Protocols*, 1(2), pp. 586–603.
- Evans, M. J. & Kaufman, M. H. (1981) 'Establishment in culture of pluripotential cells from mouse embryos', *Nature*, 292(5819), pp. 154–156.
- Farin, H. F., Jordens, I., Mosa, M. H., Basak, O., Korving, J., Tauriello, D. V. F., de Punder, K., Angers, S., Peters, P. J., Maurice, M. M., & Clevers, H. (2016) 'Visualization of a short-range Wnt gradient in the intestinal stem-cell niche', *Nature*, 530(7590), pp. 340–343.
- Feagins, L. A., Souza, R. F., & Spechler, S. J. (2009) 'Carcinogenesis in IBD: potential targets for the prevention of colorectal cancer', *Nature Reviews Gastroenterology & Hepatology*, 6(5), pp. 297–305.
- Fearon, E. R. & Vogelstein, B. (1990) 'A genetic model for colorectal tumorigenesis', *Cell*, 61(5), pp. 759–767.
- Feil, R., Brocard, J., Mascrez, B., LeMeur, M., Metzger, D., & Chambon, P. (1996) 'Ligand-activated site-specific recombination in mice', *Proceedings of the National Academy of Sciences of the United States of America*, 93(20), pp. 10887–10890.

- Feng, Q., Liang, S., Jia, H., Stadlmayr, A., Tang, L., Lan, Z., Zhang, D., Xia, H., Xu, Xiaoying, Jie, Z., Su, L., Li, Xiaoping, Li, Xin, Li, J., Xiao, L., Huber-Schönauer, U., Niederseer, D., Xu, Xun, Al-Aama, J. Y., Yang, H., Wang, Jian, Kristiansen, K., Arumugam, M., Tilg, H., Datz, C., & Wang, Jun (2015) 'Gut microbiome development along the colorectal adenoma–carcinoma sequence', *Nature Communications*, 6, p. 6528.
- Feng, Y., Bommer, G. T., Zhao, J., Green, M., Sands, E., Zhai, Y., Brown, K., Burberry, A., Cho, K. R., & Fearon, E. R. (2011) 'Mutant KRAS promotes hyperplasia and alters differentiation in the colon epithelium but does not expand the presumptive stem cell pool', *Gastroenterology*, 141(3), pp. 1003-1013.e10.
- Fernández-Barral, A., Costales-Carrera, A., Buira, S. P., Jung, P., Ferrer-Mayorga, G., Larriba, M. J., Bustamante-Madrid, P., Domínguez, O., Real, F. X., Guerra-Pastrián, L., Lafarga, M., García-Olmo, D., Cantero, R., Del Peso, L., Batlle, E., Rojo, F., Muñoz, A., & Barbáchano, A. (2019) 'Vitamin D differentially regulates colon stem cells in patient-derived normal and tumor organoids', *The FEBS Journal*. John Wiley & Sons, Ltd (10.1111), n/a(n/a).
- Fevr, T., Robine, S., Louvard, D., & Huelsken, J. (2007) 'Wnt/ $\beta$ -Catenin Is Essential for Intestinal Homeostasis and Maintenance of Intestinal Stem Cells', *Molecular and Cellular Biology*, 27(21), pp. 7551–7559.
- Fox, A. H., Lam, Y. W., Leung, A. K. L., Lyon, C. E., Andersen, J., Mann, M., & Lamond, A. I. (2002) 'Paraspeckles: A Novel Nuclear Domain', *Current Biology*, 12(1), pp. 13–25.
- Fritsch, C., Simon-Assmann, P., Kedinger, M., & Evans, G. S. (1997) 'Cytokines modulate fibroblast phenotype and epithelial-stroma interactions in rat intestine', *Gastroenterology*, 112(3), pp. 826–838.
- Frumkin, D., Wasserstrom, A., Kaplan, S., Feige, U., & Shapiro, E. (2005) 'Genomic variability within an organism exposes its cell lineage tree', *PLoS computational biology*, 1(5), pp. e50–e50.

- Fujii, M., Shimokawa, M., Date, S., Takano, A., Matano, M., Nanki, K., Ohta, Y., Toshimitsu, K., Nakazato, Y., Kawasaki, K., Uraoka, T., Watanabe, T., Kanai, T., & Sato, T. (2016) 'A Colorectal Tumor Organoid Library Demonstrates Progressive Loss of Niche Factor Requirements during Tumorigenesis', *Cell Stem Cell*, 18(6), pp. 827–838.
- Fuller, C. E., Davies, R. P., Williams, G. T., & Williams, E. D. (1990) 'Crypt restricted heterogeneity of goblet cell mucus glycoprotein in histologically normal human colonic mucosa: a potential marker of somatic mutation.', *British Journal of Cancer*, 61(3), pp. 382–384.
- Fürthauer, M. & González-Gaitán, M. (2009) 'Endocytosis, asymmetric cell division, stem cells and cancer: Unus pro omnibus, omnes pro uno', *Molecular Oncology*, 3(4), pp. 339–353.
- Galandiuk, S., Rodriguez-Justo, M., Jeffery, R., Nicholson, A. M., Cheng, Y., Oukrif, D., Elia, G., Leedham, S. J., McDonald, S. A. C., Wright, N. A., & Graham, T. A. (2012) 'Field cancerization in the intestinal epithelium of patients with Crohn's ileocolitis', *Gastroenterology*, 142(4), pp. 855-864.e8.
- García, J. M., Silva, J., Peña, C., Garcia, V., Rodríguez, R., Cruz, M. A., Cantos, B., Provencio, M., España, P., & Bonilla, F. (2004) 'Promoter methylation of the PTEN gene is a common molecular change in breast cancer', *Genes, Chromosomes and Cancer*, 41(2), pp. 117–124.
- Garcia, S. B., Park, H. S., Novelli, M., & Wright, N. A. (1999) 'Field cancerization, clonality, and epithelial stem cells: the spread of mutated clones in epithelial sheets', *The Journal of Pathology*, 187(1), pp. 61–81.
- Gelali, E., Girelli, G., Matsumoto, M., Wernersson, E., Custodio, J., Mota, A., Schweitzer, M., Ferenc, K., Li, X., Mirzazadeh, R., Agostini, F., Schell, J. P., Lanner, F., Crosetto, N., & Bienko, M. (2019) 'iFISH is a publically available resource enabling versatile DNA FISH to study genome architecture', *Nature Communications*, 10(1), p. 1636.
- Genovese, G., Kähler, A. K., Handsaker, R. E., Lindberg, J., Rose, S. A., Bakhoum, S. F., Chambert, K., Mick, E., Neale, B. M., Fromer, M., Purcell, S. M., Svantesson, O., Landén,

- M., Höglund, M., Lehmann, S., Gabriel, S. B., Moran, J. L., Lander, E. S., Sullivan, P. F., Sklar, P., Grönberg, H., Hultman, C. M., & McCarroll, S. A. (2014) 'Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence', *The New England journal of medicine*, 371(26), pp. 2477–2487.
- Gerbe, F., van Es, J. H., Makrini, L., Brulin, B., Mellitzer, G., Robine, S., Romagnolo, B., Shroyer, N. F., Bourgaux, J.-F., Pignodel, C., Clevers, H., & Jay, P. (2011) 'Distinct ATOH1 and Neurog3 requirements define tuft cells as a new secretory cell type in the intestinal epithelium', *The Journal of Cell Biology*, 192(5), pp. 767–780.
- Ghosh, S., Gupta, B., Verma, P., Vishnubathla, S., Pal, S., Dash, N. R., Gupta, S. D., & Das, P. (2018) 'Topographic, histological and molecular study of aberrant crypt foci identified in human colon in different clinical groups', *Intestinal research*. 2018/01/18. Korean Association for the Study of Intestinal Diseases, 16(1), pp. 116–125.
- Giannakis, M., Mu, X. J., Shukla, S. A., Qian, Z. R., Cohen, O., Nishihara, R., Bahl, S., Cao, Y., Amin-Mansour, A., Yamauchi, M., Sukawa, Y., Stewart, C., Rosenberg, M., Mima, K., Inamura, K., Nosho, K., Nowak, J. A., Lawrence, M. S., Giovannucci, E. L., Chan, A. T., Ng, K., Meyerhardt, J. A., Van Allen, E. M., Getz, G., Gabriel, S. B., Lander, E. S., Wu, C. J., Fuchs, C. S., Ogino, S., & Garraway, L. A. (2016) 'Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma', *Cell Reports*, 17(4), p. 1206.
- Gilbert, S. (2000) 'Early Mammalian Development', in *Developmental Biology*. 6th editio. Sunderland (MA): Sinauer Associates.
- Gillooly, J. F., Hein, A., & Damiani, R. (2015) 'Nuclear DNA Content Varies with Cell Size across Human Cell Types', *Cold Spring Harbor perspectives in biology*, 7(7), pp. a019091–a019091.
- Gjorevski, N. & Nelson, C. M. (2011) 'Integrated morphodynamic signalling of the mammary gland', *Nature Reviews Molecular Cell Biology*, 12, p. 581.
- Glinka, A., Dolde, C., Kirsch, N., Huang, Y.-L., Kazanskaya, O., Ingelfinger, D., Boutros,

M., Cruciat, C.-M., & Niehrs, C. (2011) 'LGR4 and LGR5 are R-spondin receptors mediating Wnt/ $\beta$ -catenin and Wnt/PCP signalling', *EMBO reports*, 12(10), pp. 1055–1061.

Goel, A., Arnold, C. N., Niedzwiecki, D., Carethers, J. M., Dowell, J. M., Wasserman, L., Compton, C., Mayer, R. J., Bertagnolli, M. M., & Boland, C. R. (2004) 'Frequent Inactivation of PTEN by Promoter Hypermethylation in Microsatellite Instability-High Sporadic Colorectal Cancers', *Cancer Research*, 64(9), pp. 3014–3021.

Goodman, J. E., Hofseth, L. J., Hussain, S. P., & Harris, C. C. (2004) 'Nitric oxide and p53 in cancer-prone chronic inflammation and oxyradical overload disease', *Environmental and Molecular Mutagenesis*. John Wiley & Sons, Ltd, 44(1), pp. 3–9.

Gozdecka, M., Meduri, E., Mazan, M., Tzelepis, K., Dudek, M., Knights, A. J., Pardo, M., Yu, L., Choudhary, J. S., Metzakopian, E., Iyer, V., Yun, H., Park, N., Varela, I., Bautista, R., Collord, G., Dovey, O., Garyfallos, D. A., De Braekeleer, E., Kondo, S., Cooper, J., Götting, B., Bullinger, L., Northcott, P. A., Adams, D., Vassiliou, G. S., & Huntly, B. J. P. (2018) 'UTX-mediated enhancer and chromatin remodeling suppresses myeloid leukemogenesis through noncatalytic inverse regulation of ETS and GATA programs', *Nature Genetics*, 50(6), pp. 883–894.

Graf, T. (2002) 'Differentiation plasticity of hematopoietic cells', *Blood*, 99(9), pp. 3089–3101.

Graham, T. A., Humphries, A., Sanders, T., Rodriguez-Justo, M., Tadrous, P. J., Preston, S. L., Novelli, M. R., Leedham, S. J., McDonald, S. A. C., & Wright, N. A. (2011) 'Use of Methylation Patterns to Determine Expansion of Stem Cell Clones in Human Colon Tissue', *Gastroenterology*, 140(4), pp. 1241-1250.e9.

Greaves, L. C., Preston, S. L., Tadrous, P. J., Taylor, R. W., Barron, M. J., Oukrif, D., Leedham, S. J., Deheragoda, M., Sasieni, P., Novelli, M. R., Jankowski, J. A. Z., Turnbull, D. M., Wright, N. A., & McDonald, S. A. C. (2006) 'Mitochondrial DNA mutations are established in human colonic stem cells, and mutated clones expand by crypt fission', *Proceedings of the National Academy of Sciences of the United States of America*, 103(3), pp. 714–719.

Greaves, L. C., Elson, J. L., Nooteboom, M., Grady, J. P., Taylor, G. A., Taylor, R. W., Mathers, J. C., Kirkwood, T. B. L., & Turnbull, D. M. (2012) 'Comparison of Mitochondrial Mutation Spectra in Ageing Human Colonic Epithelium and Disease: Absence of Evidence for Purifying Selection in Somatic Mitochondrial DNA Point Mutations', *PLoS Genet*, 8(11), p. e1003082.

Green, C. E., Liu, T., Montel, V., Hsiao, G., Lester, R. D., Subramaniam, S., Gonias, S. L., & Klemke, R. L. (2009) 'Chemoattractant signaling between tumor cells and macrophages regulates cancer cell migration, metastasis and neovascularization', *PloS one*. Public Library of Science, 4(8), pp. e6713–e6713.

Grimsby, J., Chen, K., Wang, L. J., Lan, N. C., & Shih, J. C. (1991) 'Human monoamine oxidase A and B genes exhibit identical exon-intron organization.', *Proceedings of the National Academy of Sciences*, 88(9), pp. 3637–3641.

Guan, F., Tabrizian, T., Novaj, A., Nakanishi, M., Rosenberg, D. W., & Huffman, D. M. (2018) 'Dietary Walnuts Protect Against Obesity-Driven Intestinal Stem Cell Decline and Tumorigenesis', *Frontiers in Nutrition*, p. 37.

Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., Bot, B. M., Morris, J. S., Simon, I. M., Gerster, S., Fessler, E., De Sousa E Melo, F., Missiaglia, E., Ramay, H., Barras, D., Homicsko, K., Maru, D., Manyam, G. C., Broom, B., Boige, V., Perez-Villamil, B., Laderas, T., Salazar, R., Gray, J. W., Hanahan, D., Tabernero, J., Bernards, R., Friend, S. H., Laurent-Puig, P., Medema, J. P., Sadanandam, A., Wessels, L., Delorenzi, M., Kopetz, S., Vermeulen, L., & Tejpar, S. (2015) 'The consensus molecular subtypes of colorectal cancer', *Nature Medicine*, 21(11), pp. 1350–1356.

Gupta, B., Das, P., Ghosh, S., Manhas, J., Sen, S., Pal, S., Sahni, P., Upadhyay, A. D., Panda, S. K., & Gupta, S. D. (2017) 'Identification of High-Risk Aberrant Crypt Foci and Mucin-Depleted Foci in the Human Colon With Study of Colon Cancer Stem Cell Markers', *Clinical Colorectal Cancer*, 16(3), pp. 204–213.

- Gupta, V., Khan, A. A., Sasi, B. K., & Mahapatra, N. R. (2015) 'Molecular mechanism of monoamine oxidase A gene regulation under inflammation and ischemia-like conditions: key roles of the transcription factors GATA2, Sp1 and TBP', *Journal of Neurochemistry*, 134(1), pp. 21–38.
- Gutierrez-Gonzalez, L., Deheragoda, M., Elia, G., Leedham, S. J., Shankar, A., Imber, C., Jankowski, J. A., Turnbull, D. M., Novelli, M., Wright, N. A., & McDonald, S. A. C. (2009) 'Analysis of the clonal architecture of the human small intestinal epithelium establishes a common stem cell for all lineages and reveals a mechanism for the fixation and spread of mutations', *The Journal of Pathology*, 217(4), pp. 489–496.
- van Haaften, G., Dalgliesh, G. L., Davies, H., Chen, L., Bignell, G., Futreal, P. A., *et al.* (2009) 'Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer', *Nature genetics*, 41(5), pp. 521–523.
- Hadac, J. N., Leystra, A. A., Paul Olson, T. J., Maher, M. E., Payne, S. N., Yueh, A. E., Schwartz, A. R., Albrecht, D. M., Clipson, L., Pasch, C. A., Matkowskyj, K. A., Halberg, R. B., & Deming, D. A. (2015) 'Colon Tumors with the Simultaneous Induction of Driver Mutations in APC, KRAS, and PIK3CA Still Progress through the Adenoma-to-carcinoma Sequence', *Cancer prevention research (Philadelphia, Pa.)*. 2015/08/14, 8(10), pp. 952–961.
- Haile, S., Pandoh, P., McDonald, H., Corbett, R. D., Tsao, P., Kirk, H., MacLeod, T., Jones, M., Bilobram, S., Brooks, D., Smailus, D., Steidl, C., Scott, D. W., Bala, M., Hirst, M., Miller, D., Moore, R. A., Mungall, A. J., Coope, R. J., Ma, Y., Zhao, Y., Holt, R. A., Jones, S. J., & Marra, M. A. (2017) 'Automated high throughput nucleic acid purification from formalin-fixed paraffin-embedded tissue samples for next generation sequence analysis', *PLOS ONE*, 12(6), p. e0178706.
- Halpern, K. B., Massalha, H., Zwick, R. K., Moor, A. E., Castillo-Azofeifa, D., Rozenberg, M., Farack, L., Egozi, A., Miller, D. R., Averbukh, I., Harnik, Y., Weinberg-Corem, N., de Sauvage, F. J., Amit, I., Klein, O. D., Shoshkes-Carmel, M., & Itzkovitz, S. (2019) 'Lgr5+ telocytes are a signaling hub at the intestinal villus tip', *bioRxiv*, p. 850909.
- Hao, X. P., Frayling, I. M., Sgouros, J. G., Du, M. Q., Willcocks, T. C., Talbot, I. C., &

- Tomlinson, I. P. M. (2002) 'The spectrum of p53 mutations in colorectal adenomas differs from that in colorectal carcinomas', *Gut*, 50(6), pp. 834–839.
- He, X. C., Zhang, J., Tong, W.-G., Tawfik, O., Ross, J., Scoville, D. H., Tian, Q., Zeng, X., He, X., Wiedemann, L. M., Mishina, Y., & Li, L. (2004) 'BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt- $\beta$ -catenin signaling', *Nature Genetics*, 36(10), pp. 1117–1121.
- van der Heijden, R. A., Bijzet, J., Meijers, W. C., Yakala, G. K., Kleemann, R., Nguyen, T. Q., de Boer, R. A., Schalkwijk, C. G., Hazenberg, B. P. C., Tietge, U. J. F., & Heeringa, P. (2015) 'Obesity-induced chronic inflammation in high fat diet challenged C57BL/6J mice is associated with acceleration of age-dependent renal amyloidosis', 5, p. 16474.
- Hideshima, T., Mazitschek, R., Qi, J., Mimura, N., Tseng, J.-C., Kung, A. L., Bradner, J. E., & Anderson, K. C. (2017) 'HDAC6 inhibitor WT161 downregulates growth factor receptors in breast cancer', *Oncotarget*, 8(46), pp. 80109–80123.
- Hill, V. K., Kim, J.-S., & Waldman, T. (2016) 'Cohesin mutations in human cancer', *Biochimica et biophysica acta*, 1866(1), pp. 1–11.
- Hlobilkova, A., Guldberg, P., Thullberg, M., Zeuthen, J., Lukas, J., & Bartek, J. (2000) 'Cell Cycle Arrest by the PTEN Tumor Suppressor Is Target Cell Specific and May Require Protein Phosphatase Activity', *Experimental Cell Research*, 256(2), pp. 571–577.
- Ho, C.-M., Lin, M.-C., Huang, S.-H., Huang, C.-J., Lai, H.-C., Chien, T.-Y., & Chang, S.-F. (2009) 'PTEN promoter methylation and LOH of 10q22–23 locus in PTEN expression of ovarian clear cell adenocarcinomas', *Gynecologic Oncology*, 112(2), pp. 307–313.
- Hollander, M. C., Blumenthal, G. M., & Dennis, P. A. (2011) 'PTEN loss in the continuum of common cancers, rare syndromes and mouse models', *Nature Reviews Cancer*, 11(4), pp. 289–301.
- Hounnou, G., Destrieux, C., Desme, J., & Bertrand, P. (2002) 'Anatomical study of the length of the human intestine', *Surgical and Radiologic Anatomy*, 24(5), pp. 290–294.

- Hu, Z., Ding, J., Ma, Z., Sun, R., Seoane, J. A., Scott Shaffer, J., Suarez, C. J., Berghoff, A. S., Cremolini, C., Falcone, A., Loupakis, F., Birner, P., Preusser, M., Lenz, H.-J., & Curtis, C. (2019) 'Quantitative evidence for early metastatic seeding in colorectal cancer', *Nature Genetics*, 51(7), pp. 1113–1122.
- Hubbert, C., Guardiola, A., Shao, R., Kawaguchi, Y., Ito, A., Nixon, A., Yoshida, M., Wang, X.-F., & Yao, T.-P. (2002) 'HDAC6 is a microtubule-associated deacetylase', *Nature*, 417, p. 455.
- Hussain, A., Dulay, P., Rivera, M. N., Aramouni, C., & Saxena, V. (2019) 'Neoplastic Pathogenesis Associated with Cigarette Carcinogens', *Cureus*, 11(1), pp. e3955–e3955.
- Hye-Rim, W., Hyun-Wook, R., Dong-Hee, S., Soo-Keun, Y., Hoon, L. D., & Hee, K. S. (2018) 'A452, an HDAC6-selective inhibitor, synergistically enhances the anticancer activity of chemotherapeutic agents in colorectal cancer cells', *Molecular Carcinogenesis*, 57(10), pp. 1383–1395.
- Ireland, H., Houghton, C., Howard, L., & Winton, D. J. (2005) 'Cellular inheritance of a Cre-activated reporter gene to determine paneth cell longevity in the murine small intestine', *Developmental Dynamics*, 233(4), pp. 1332–1336.
- Jadhav, U., Saxena, M., O'Neill, N. K., Saadatpour, A., Yuan, G.-C., Herbert, Z., Murata, K., & Shivdasani, R. A. (2017) 'Dynamic Reorganization of Chromatin Accessibility Signatures during Dedifferentiation of Secretory Precursors into Lgr5+ Intestinal Stem Cells', *Cell Stem Cell*, 21(1), pp. 65-77.e5.
- Jager, M., Blokzijl, F., Sasselli, V., Boymans, S., Janssen, R., Besselink, N., Clevers, H., van Boxtel, R., & Cuppen, E. (2017) 'Measuring mutation accumulation in single human adult stem cells by whole-genome sequencing of organoid cultures', *Nature Protocols*, 13(1), pp. 59–78.
- Jaks, V., Kasper, M., & Toftgård, R. (2010) 'The hair follicle—a stem cell zoo', *Experimental Cell Research*, 316(8), pp. 1422–1428.
- Jedinak, A., Dudhgaonkar, S., & Sliva, D. (2010) 'Activated macrophages induce metastatic

behavior of colon cancer cells', *Immunobiology*, 215(3), pp. 242–249.

Jenny, M., Uhl, C., Roche, C., Duluc, I., Guillermin, V., Guillemot, F., Jensen, J., Kedinger, M., & Gradwohl, G. (2002) 'Neurogenin3 is differentially required for endocrine cell fate specification in the intestinal and gastric epithelium', *The EMBO Journal*. Oxford, UK, 21(23), pp. 6338–6347.

Van Kaer, L. & Olivares-Villagómez, D. (2018) 'Development, Homeostasis, and Functions of Intestinal Intraepithelial Lymphocytes', *The Journal of Immunology*, 200(7), pp. 2235 LP – 2244.

Kang, Y., Lee, H., & Kim, W. (2002) 'Promoter methylation and silencing of PTEN in gastric carcinoma.', *Lab Invest.*, 82(3), pp. 285–91.

Khoury, M. P. & Bourdon, J.-C. (2010) 'The isoforms of the p53 protein', *Cold Spring Harbor perspectives in biology*, 2(3), pp. a000927–a000927.

Kiktev, D. A., Sheng, Z., Lobachev, K. S., & Petes, T. D. (2018) 'GC content elevates mutation and recombination rates in the yeast *Saccharomyces cerevisiae*', *Proceedings of the National Academy of Sciences*, 115(30), pp. E7109–E7118.

Kim, J.-S., He, X., Orr, B., Wutz, G., Hill, V., Peters, J.-M., Compton, D. A., & Waldman, T. (2016) 'Intact Cohesion, Anaphase, and Chromosome Segregation in Human Cells Harboring Tumor-Derived Mutations in STAG2', *PLoS genetics*, 12(2), pp. e1005865–e1005865.

Kim, J. H. & Kang, G. H. (2014) 'Molecular and prognostic heterogeneity of microsatellite-unstable colorectal cancer', *World journal of gastroenterology*, 20(15), pp. 4230–4243.

Kim, K.-M. & Shibata, D. (2002) 'Methylation reveals a niche: stem cell succession in human colon crypts', *Oncogene*, 21(35), pp. 5441–5449.

Kim, K.-M. & Shibata, D. (2004) 'Tracing ancestry with methylation patterns: most crypts appear distantly related in normal adult human colon', *BMC gastroenterology*, 4, p. 8.

Kim, S., Park, C., Ji, Y., Kim, D. G., Bae, H., van Vrancken, M., Kim, D.-H., & Kim, K.-M.

- (2017) 'Deamination Effects in Formalin-Fixed, Paraffin-Embedded Tissue Samples in the Era of Precision Medicine', *The Journal of Molecular Diagnostics*, 19(1), pp. 137–146.
- Kitayner, M., Rozenberg, H., Kessler, N., Rabinovich, D., Shaulov, L., Haran, T. E., & Shakked, Z. (2006) 'Structural Basis of DNA Recognition by p53 Tetramers', *Molecular Cell*, 22(6), pp. 741–753.
- Knoblich, J. A. (2008) 'Mechanisms of Asymmetric Stem Cell Division', *Cell*, 132(4), pp. 583–597.
- Knott, G. J., Bond, C. S., & Fox, A. H. (2016) 'The DBHS proteins SFPQ, NONO and PSPC1: a multipurpose molecular scaffold', *Nucleic acids research*, 44(9), pp. 3989–4004.
- Kornack, D. R. & Rakic, P. (2001) 'The generation, migration, and differentiation of olfactory neurons in the adult primate brain', *Proceedings of the National Academy of Sciences*, 98(8), pp. 4752–4757.
- Kosinski, C., Stange, D. E., Xu, C., Chan, A. S., Ho, C., Yuen, S. T., Mifflin, R. C., Powell, D. W., Clevers, H., Leung, S. Y., & Chen, X. (2010) 'Indian hedgehog regulates intestinal stem cell fate through epithelial-mesenchymal interactions during development', *Gastroenterology*. 2010/06/11, 139(3), pp. 893–903.
- Kozar, S., Morrissey, E., Nicholson, A. M., van der Heijden, M., Zecchini, H. I., Kemp, R., Tavaré, S., Vermeulen, L., & Winton, D. J. (2013) 'Continuous Clonal Labeling Reveals Small Numbers of Functional Stem Cells in Intestinal Crypts and Adenomas', *Cell Stem Cell*, 13(5), pp. 626–633.
- Kraus, M. C., Seelig, M. H., Linnemann, U., & Berger, M. (2006) 'The balanced induction of K-ras codon 12 and 13 mutations in mucosa differs from their ratio in neoplastic tissues.', *International Journal of Oncology*, 29(4), pp. 957–964.
- Kraus, S. & Arber, N. (2009) 'Inflammation and colorectal cancer', *Current Opinion in Pharmacology*, 9(4), pp. 405–410.
- Kudo, S., Tamura, S., Nakajima, T., Yamano, H., Kusaka, H., & Watanabe, H. (1996) 'Diagnosis of colorectal tumorous lesions by magnifying endoscopy', *Gastrointestinal*

*Endoscopy*. Elsevier, 44(1), pp. 8–14.

Kukitsu, T., Takayama, T., Miyanishi, K., Nobuoka, A., Katsuki, S., Sato, Y., Takimoto, R., Matsunaga, T., Kato, J., Sonoda, T., Sakamaki, S., & Niitsu, Y. (2008) ‘Aberrant Crypt Foci as Precursors of the Dysplasia-Carcinoma Sequence in Patients with Ulcerative Colitis’, *Clinical Cancer Research*, 14(1), pp. 48 LP – 54.

Kuraishy, A., Karin, M., & Grivennikov, S. I. (2011) ‘Tumor promotion via injury- and death-induced inflammation’, *Immunity*, 35(4), pp. 467–477.

Lahav, G., Rosenfeld, N., Sigal, A., Geva-Zatorsky, N., Levine, A. J., Elowitz, M. B., & Alon, U. (2004) ‘Dynamics of the p53-Mdm2 feedback loop in individual cells’, *Nature Genetics*, 36(2), pp. 147–150.

Lander, A. D., Gokoffski, K. K., Wan, F. Y. M., Nie, Q., & Calof, A. L. (2009) ‘Cell Lineages and the Logic of Proliferative Control’, *PLOS Biology*, 7(1), p. e1000015.

Lane, D. P. (1992) ‘p53, guardian of the genome’, *Nature*, 358(6381), pp. 15–16.

Lang, M., Schlechtweg, M., Kellermeier, S., Brenmoehl, J., Falk, W., Schlalmerich, J., Herfarth, H., Rogler, G., & Hausmann, M. (2009) ‘Gene expression profiles of mucosal fibroblasts from strictured and nonstrictured areas of patients with Crohn’s disease’, *Inflammatory bowel diseases*, 15, pp. 212–223.

Langlands, A. J., Almet, A. A., Appleton, P. L., Newton, I. P., Osborne, J. M., & Näthke, I. S. (2016) ‘Paneth Cell-Rich Regions Separated by a Cluster of Lgr5+ Cells Initiate Crypt Fission in the Intestinal Stem Cell Niche’, *PLoS biology*, 14(6), pp. e1002491–e1002491.

Lee-Six, H., Ellis, P., Osborne, R. J., Sanders, M. A., Moore, L., Georgakopoulos, N., Torrente, F., Noorani, A., Goddard, M., Robinson, P., Coorens, T. H. H., O’Neill, L., Alder, C., Wang, J., Fitzgerald, R. C., Zilbauer, M., Coleman, N., Saeb-Parsy, K., Martincorena, I., Campbell, P. J., & Stratton, M. R. (2018) ‘The landscape of somatic mutation in normal colorectal epithelial cells’, *bioRxiv*.

- Lee, Y.-S., Lim, K.-H., Guo, X., Kawaguchi, Y., Gao, Y., Barrientos, T., Ordentlich, P., Wang, X.-F., Counter, C. M., & Yao, T.-P. (2008) 'The Cytoplasmic Deacetylase HDAC6 Is Required for Efficient Oncogenic Tumorigenesis', *Cancer Research*, 68(18), pp. 7561–7569.
- Ler, L. D., Ghosh, S., Chai, X., Thike, A. A., Heng, H. L., Siew, E. Y., Dey, S., Koh, L. K., Lim, J. Q., Lim, W. K., Myint, S. S., Loh, J. L., Ong, P., Sam, X. X., Huang, D., Lim, T., Tan, P. H., Nagarajan, S., Cheng, C. W. S., Ho, H., Ng, L. G., Yuen, J., Lin, P.-H., Chuang, C.-K., Chang, Y.-H., Weng, W.-H., Rozen, S. G., Tan, P., Creasy, C. L., Pang, S.-T., McCabe, M. T., Poon, S. L., & Teh, B. T. (2017) 'Loss of tumor suppressor KDM6A amplifies PRC2-regulated transcriptional repression in bladder cancer and can be targeted through inhibition of EZH2', *Science Translational Medicine*, 9(378), p. eaai8312.
- De Lerma Barbaro, A., Perletti, G., Bonapace, I. M., & Monti, E. (2014) 'Inflammatory cues acting on the adult intestinal stem cells and the early onset of cancer (review)', *International journal of oncology*. 2014/06/10. D.A. Spandidos, 45(3), pp. 959–968.
- Leystra, A. A., Deming, D. A., Zahm, C. D., Farhoud, M., Olson, T. J. P., Hadac, J. N., Nettekoven, L. A., Albrecht, D. M., Clipson, L., Sullivan, R., Washington, M. K., Torrealba, J. R., Weichert, J. P., & Halberg, R. B. (2012) 'Mice expressing activated PI3K rapidly develop advanced colon cancer', *Cancer research*. 2012/04/23, 72(12), pp. 2931–2936.
- Li, A., Chen, P., Leng, Y., & Kang, J. (2018) 'Histone deacetylase 6 regulates the immunosuppressive properties of cancer-associated fibroblasts in breast cancer through the STAT3–COX2-dependent pathway', *Oncogene*, 37(45), pp. 5952–5966.
- Li, D., Marchenko, N. D., & Moll, U. M. (2011) 'SAHA shows preferential cytotoxicity in mutant p53 cancer cells by destabilizing mutant p53 through inhibition of the HDAC6-Hsp90 chaperone axis', *Cell death and differentiation*, 18(12), pp. 1904–1913.
- Littlewood, T. D., Hancock, D. C., Danielian, P. S., Parker, M. G., & Evan, G. I. (1995) 'A modified oestrogen receptor ligand-binding domain as an improved switch for the regulation of heterologous proteins', *Nucleic acids research*, 23(10), pp. 1686–1690.
- Liu, P. Y., Erriquez, D., Marshall, G. M., Tee, A. E., Polly, P., Wong, M., Liu, B., Bell, J. L., Zhang, X. D., Milazzo, G., Cheung, B. B., Fox, A., Swarbrick, A., Hüttelmaier, S., Kavallaris,

- M., Perini, G., Mattick, J. S., Dinger, M. E., & Liu, T. (2014) 'Effects of a Novel Long Noncoding RNA, lncUSMycN, on N-Myc Expression and Neuroblastoma Progression', *JNCI: Journal of the National Cancer Institute*, 106(7).
- Lloyd-Lewis, B., Davis, F. M., Harris, O. B., Hitchcock, J. R., & Watson, C. J. (2018) 'Neutral lineage tracing of proliferative embryonic and adult mammary stem/progenitor cells', *Development*, 145(14), p. dev164079.
- Lodato, M. A., Rodin, R. E., Bohrsen, C. L., Coulter, M. E., Barton, A. R., Kwon, M., Sherman, M. A., Vitzthum, C. M., Luquette, L. J., Yandava, C. N., Yang, P., Chittenden, T. W., Hatem, N. E., Ryu, S. C., Woodworth, M. B., Park, P. J., & Walsh, C. A. (2018) 'Aging and neurodegeneration are associated with increased mutations in single human neurons', *Science*, 359(6375), pp. 555–559.
- Lopez-Garcia, C., Klein, A. M., Simons, B. D., & Winton, D. J. (2010) 'Intestinal Stem Cell Replacement Follows a Pattern of Neutral Drift', *Science*, 330(6005), pp. 822–825.
- Ludgate, J. L., Wright, J., Stockwell, P. A., Morison, I. M., Eccles, M. R., & Chatterjee, A. (2017) 'A streamlined method for analysing genome-wide DNA methylation patterns from low amounts of FFPE DNA', *BMC medical genomics*, 10(1), p. 54.
- Maehama, T. & Dixon, J. E. (1998) 'The Tumor Suppressor, PTEN/MMAC1, Dephosphorylates the Lipid Second Messenger, Phosphatidylinositol 3,4,5-Trisphosphate', *Journal of Biological Chemistry*, 273(22), pp. 13375–13378.
- Manning, B. D. & Toker, A. (2017) 'AKT/PKB Signaling: Navigating the Network', *Cell*, 169(3), pp. 381–405.
- Margueron, R. & Reinberg, D. (2011) 'The Polycomb complex PRC2 and its mark in life', *Nature*, 469(7330), pp. 343–349.
- Martin, G. R. (1981) 'Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells', *Proceedings of the National Academy of Sciences of the United States of America*, 78(12), pp. 7634–7638.

- Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D. C., Fullam, A., Alexandrov, L. B., Tubio, J. M., Stebbings, L., Menzies, A., Widaa, S., Stratton, M. R., Jones, P. H., & Campbell, P. J. (2015) 'Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin', *Science*, 348(6237), pp. 880–886.
- Martincorena, I., Fowler, J. C., Wabik, A., Lawson, A. R. J., Abascal, F., Hall, M. W. J., Cagan, A., Murai, K., Mahbubani, K., Stratton, M. R., Fitzgerald, R. C., Handford, P. A., Campbell, P. J., Saeb-Parsy, K., & Jones, P. H. (2018) 'Somatic mutant clones colonize the human esophagus with age', *Science*, 362(6417), pp. 911–917.
- Maskens, A. (1978) 'Histogenesis of colon glands during postnatal growth', *Acta Anat (Basel)*, 100(1), pp. 17–26.
- Maskens, A. P. & Duhardin-Loits, R. M. (1981) 'Kinetics of tissue proliferation in colorectal mucosa during post-natal growth.', *Cell Tissue Kinet.*, 14(5), pp. 467–77.
- Maywald, R. L., Doerner, S. K., Pastorelli, L., De Salvo, C., Benton, S. M., Dawson, E. P., Lanza, D. G., Berger, N. A., Markowitz, S. D., Lenz, H.-J., Nadeau, J. H., Pizarro, T. T., & Heaney, J. D. (2015) 'IL-33 activates tumor stroma to promote intestinal polyposis', *Proceedings of the National Academy of Sciences of the United States of America*. 2015/04/27. National Academy of Sciences, 112(19), pp. E2487–E2496.
- McCullough, M. L., Zoltick, E. S., Weinstein, S. J., Fedirko, V., Wang, M., Smith-Warner, S. A., *et al.* (2018) 'Circulating Vitamin D and Colorectal Cancer Risk: An International Pooling Project of 17 Cohorts', *JNCI: Journal of the National Cancer Institute*, 111(2), pp. 158–169.
- Meira, L. B., Bugni, J. M., Green, S. L., Lee, C.-W., Pang, B., Borenshtein, D., Rickman, B. H., Rogers, A. B., Moroski-Erkul, C. A., McFaline, J. L., Schauer, D. B., Dedon, P. C., Fox, J. G., & Samson, L. D. (2008) 'DNA damage induced by chronic inflammation contributes to colon carcinogenesis in mice', *The Journal of clinical investigation*. American Society for Clinical Investigation, 118(7), pp. 2516–2525.
- Meran, L., Baulies, A., & Li, V. S. W. (2017) 'Intestinal Stem Cell Niche : The Extracellular Matrix and Cellular Components', *Stem Cells International*, 2017.

Metzger, D., Ali, S., Bornert, J.-M., & Chambon, P. (1995) 'Characterization of the Amino-terminal Transcriptional Activation Function of the Human Estrogen Receptor in Animal and Yeast Cells', *Journal of Biological Chemistry*, 270(16), pp. 9535–9542.

Van der Meulen, J., Sanghvi, V., Mavrakis, K., Durinck, K., Fang, F., Matthijssens, F., Rondou, P., Rosen, M., Pieters, T., Vandenberghe, P., Delabesse, E., Lammens, T., De Moerloose, B., Menten, B., Van Roy, N., Verhasselt, B., Poppe, B., Benoit, Y., Taghon, T., Melnick, A. M., Speleman, F., Wendel, H.-G., & Van Vlierberghe, P. (2015) 'The H3K27me3 demethylase UTX is a gender-specific tumor suppressor in T-cell acute lymphoblastic leukemia', *Blood*, 125(1), pp. 13–21.

Michikawa, Y., Mazzucchelli, F., Bresolin, N., Scarlato, G., & Attardi, G. (1999) 'Aging-Dependent Large Accumulation of Point Mutations in the Human mtDNA Control Region for Replication', *Science*, 286(5440), pp. 774–779.

Milella, M., Falcone, I., Conciatori, F., Cesta Incani, U., Del Curatolo, A., Inzerilli, N., Nuzzo, C. M. A., Vaccaro, V., Vari, S., Cognetti, F., & Ciuffreda, L. (2015) 'PTEN: Multiple Functions in Human Malignant Tumors', *Frontiers in oncology*, 5, p. 24.

Miller, S. A., Mohn, S. E., & Weinmann, A. S. (2010) 'Jmjd3 and UTX play a demethylase-independent role in chromatin remodeling to regulate T-box family member-dependent gene expression', *Molecular cell*, 40(4), pp. 594–605.

Montgomery, R. K., Carlone, D. L., Richmond, C. A., Farilla, L., Kranendonk, M. E. G., Henderson, D. E., Baffour-Awuah, N. Y., Ambruzs, D. M., Fogli, L. K., Algra, S., & Breault, D. T. (2011) 'Mouse telomerase reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells', *Proceedings of the National Academy of Sciences*, 108(1), pp. 179–184.

Morimoto, L. M., Newcomb, P. A., Ulrich, C. M., Bostick, R. M., Lais, C. J., & Potter, J. D. (2002) 'Risk Factors for Hyperplastic and Adenomatous Polyps', *Cancer Epidemiology Biomarkers & Prevention*, 11(10), pp. 1012 LP – 1018.

- Mueller, L., Goumas, F. A., Affeldt, M., Sandtner, S., Gehling, U. M., Brilloff, S., Walter, J., Karnatz, N., Lamszus, K., Rogiers, X., & Broering, D. C. (2007) 'Stromal fibroblasts in colorectal liver metastases originate from resident fibroblasts and generate an inflammatory microenvironment', *The American journal of pathology*. 2007/10/04. American Society for Investigative Pathology, 171(5), pp. 1608–1618.
- Muñoz, J., Stange, D. E., Schepers, A. G., van de Wetering, M., Koo, B.-K., Itzkovitz, S., Volckmann, R., Kung, K. S., Koster, J., Radulescu, S., Myant, K., Versteeg, R., Sansom, O. J., van Es, J. H., Barker, N., van Oudenaarden, A., Mohammed, S., Heck, A. J. R., & Clevers, H. (2012) 'The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent "+4" cell markers', *The EMBO journal*. Nature Publishing Group, 31(14), pp. 3079–3091.
- Muto, T., Bussey, H. J., & Morson, B. C. (1975) 'The evolution of cancer of the colon and rectum.', *Cancer*, 36(6), pp. 2251–70.
- Nakayama, M. & Oshima, M. (2018) 'Mutant p53 in colon cancer', *Journal of molecular cell biology*. Oxford University Press, 11(4), pp. 267–276.
- Nguyen, T.-A., Menendez, D., Resnick, M. A., & Anderson, C. W. (2014) 'Mutant TP53 posttranslational modifications: challenges and opportunities', *Human mutation*, 35(6), pp. 738–755.
- Nicholson, A. M., Olpe, C., Hoyle, A., Thorsen, A.-S., Rus, T., Colombé, M., Brunton-Sim, R., Kemp, R., Marks, K., Quirke, P., Malhotra, S., ten Hoopen, R., Ibrahim, A., Lindskog, C., Myers, M. B., Parsons, B., Tavaré, S., Wilkinson, M., Morrissey, E., & Winton, D. J. (2018) 'Fixation and Spread of Somatic Mutations in Adult Human Colonic Epithelium', *Cell Stem Cell*, 22(6), pp. 909-918.e8.
- Niikawa, N., Matsuura, N., Fukushima, Y., Ohsawa, T., & Kajii, T. (1981) 'Kabuki make-up syndrome: A syndrome of mentalretardation, unusual facies, large and protruding ears, and postnatal growth deficiency', *The Journal of Pediatrics*, 99(4), pp. 565–569.
- Nooteboom, M., Johnson, R., Taylor, R. W., Wright, N. A., Lightowers, R. N., Kirkwood, T. B. L., Mathers, J. C., Turnbull, D. M., & Greaves, L. C. (2010) 'Age-associated mitochondrial DNA mutations lead to small but significant changes in cell proliferation and apoptosis in

- human colonic crypts', *Aging Cell*, 9(1), pp. 96–99.
- Novelli, M., Cossu, A., Oukrif, D., Quaglia, A., Lakhani, S., Poulsom, R., Sasieni, P., Carta, P., Contini, M., Pasca, A., Palmieri, G., Bodmer, W., Tanda, F., & Wright, N. (2003) 'X-inactivation patch size in human female tissue confounds the assessment of tumor clonality', *Proceedings of the National Academy of Sciences*, 100(6), pp. 3311–3314.
- Novelli, M. R., Williamson, J. A., Tomlinson, I. P. M., Elia, G., Hodgson, S. V, Talbot, I. C., Bodmer, W. F., & Wright, N. A. (1996) 'Polyclonal Origin of Colonic Adenomas in an XO/XY Patient with FAP', *Science*, 272(5265), pp. 1187–1190.
- Nowak, J. A., Polak, L., Pasolli, H. A., & Fuchs, E. (2008) 'Hair follicle stem cells are specified and function in early skin morphogenesis', *Cell stem cell*, 3(1), pp. 33–43.
- Nowell, P. C. (1976) 'The clonal evolution of tumor cell populations', *Science*, 194(4260), pp. 23–28.
- Oguma, K., Oshima, H., Aoki, M., Uchio, R., Naka, K., Nakamura, S., Hirao, A., Saya, H., Taketo, M. M., & Oshima, M. (2008) 'Activated macrophages promote Wnt signalling through tumour necrosis factor-alpha in gastric tumour cells', *The EMBO journal*. 2008/05/29. Nature Publishing Group, 27(12), pp. 1671–1681.
- Okayasu, I., Ohkusa, T., Kajiura, K., Kanno, J., & Sakamoto, S. (1996) 'Promotion of colorectal neoplasia in experimental murine ulcerative colitis', *Gut*, 39(1), pp. 87–92.
- Orlando, F. A., Tan, D., Baltodano, J. D., Khoury, T., Gibbs, J. F., Hassid, V. J., Ahmed, B. H., & Alrawi, S. J. (2008) 'Aberrant crypt foci as precursors in colorectal cancer progression', *Journal of surgical oncology*, 98(3), pp. 207–213.
- Otori, K., Sugiyama, K., Hasebe, T., Fukushima, S., & Esumi, H. (1995) 'Emergence of Adenomatous Aberrant Crypt Foci (ACF) from Hyperplastic ACF with Concomitant Increase in Cell Proliferation', *Cancer Research*, 55(21), pp. 4743 LP – 4746.
- Owens, B. M. J. & Simmons, A. (2013) 'Intestinal stromal cells in mucosal immunity and

homeostasis', *Mucosal Immunology*, 6(2), pp. 224–234.

Pantazis, P. & Bollenbach, T. (2012) 'Transcription factor kinetics and the emerging asymmetry in the early mammalian embryo', *Cell Cycle*, 11(11), pp. 2055–2058.

Papa, A. & Pandolfi, P. P. (2019) 'The PTEN-PI3K Axis in Cancer', *Biomolecules*, 9(4), p. 153.

Park, S.-Y., Phorl, S., Jung, S., Sovannarith, K., Lee, S., Noh, S., Han, M., Naskar, R., Kim, J.-Y., Choi, Y.-J., & Lee, J.-Y. (2017) 'HDAC6 deficiency induces apoptosis in mesenchymal stem cells through p53 K120 acetylation', *Biochemical and Biophysical Research Communications*, 494(1), pp. 51–56.

Parsons, B., Marchant-Miros, K., Delongchamp, R., Verkler, T., Patterson, T., McKinzie, P., & Kim, L. (2010) 'ACB-PCR quantification of K-RAS codon 12 GAT and GTT mutant fraction in colon tumor and non-tumor tissue.', *Cancer Invest.*, 28(4), pp. 364–375.

Passon, D. M., Lee, M., Rackham, O., Stanley, W. A., Sadowska, A., Filipovska, A., Fox, A. H., & Bond, C. S. (2012) 'Structure of the heterodimer of human NONO and paraspeckle protein component 1 and analysis of its role in subnuclear body formation', *Proceedings of the National Academy of Sciences of the United States of America*, 109(13), pp. 4846–4850.

Pastuła, A., Middelhoff, M., Brandtner, A., Tobiasch, M., Höhl, B., Nuber, A. H., Demir, I. E., Neupert, S., Kollmann, P., Mazzuoli-Weber, G., & Quante, M. (2016) 'Three-Dimensional Gastrointestinal Organoid Culture in Combination with Nerves or Fibroblasts: A Method to Characterize the Gastrointestinal Stem Cell Niche', *Stem cells international*. 2015/11/30. Hindawi Publishing Corporation, 2016, p. 3710836.

Pastuła, A. & Marcinkiewicz, J. (2019) 'Cellular Interactions in the Intestinal Stem Cell Niche', *Archivum immunologiae et therapeuticae experimentalis*. 2018/09/21. Springer International Publishing, 67(1), pp. 19–26.

Peddareddigari, V. G., Wang, D., & Dubois, R. N. (2010) 'The tumor microenvironment in colorectal carcinogenesis', *Cancer microenvironment : official journal of the International Cancer Microenvironment Society*. Springer Netherlands, 3(1), pp. 149–166.

- Petti, E., Buemi, V., Zappone, A., Schillaci, O., Broccia, P. V., Dinami, R., Matteoni, S., Benetti, R., & Schoeftner, S. (2019) 'SFPQ and NONO suppress RNA:DNA-hybrid-related telomere instability', *Nature Communications*, 10(1), p. 1001.
- Pickhardt, P. J., Kim, D. H., Pooler, B. D., Hinshaw, J. L., Barlow, D., Jensen, D., Reichelderfer, M., & Cash, B. D. (2013) 'Assessment of volumetric growth rates of small colorectal polyps with CT colonography: a longitudinal study of natural history', *The Lancet. Oncology*. 2013/06/07, 14(8), pp. 711–720.
- Pollard, J. W. (2004) 'Tumour-educated macrophages promote tumour progression and metastasis', *Nature Reviews Cancer*, 4(1), pp. 71–78.
- Potten, C. S., Hume, W. J., Reid, P., & Cairns, J. (1978) 'The segregation of DNA in epithelial stem cells', *Cell*. Elsevier, 15(3), pp. 899–906.
- Potten, C. S. & Loeffler, M. (1990) 'Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons for and from the crypt', *Development*, 110(4), pp. 1001–1020.
- Potten, C. S., Owen, G., & Booth, D. (2002) 'Intestinal stem cells protect their genome by selective segregation of template DNA strands', *Journal of Cell Science*, 115(11), pp. 2381 LP – 2388.
- Preston, S. L., Wong, W.-M., Chan, A. O.-O., Poulson, R., Jeffery, R., Goodlad, R. A., Mandir, N., Elia, G., Novelli, M., Bodmer, W. F., Tomlinson, I. P., & Wright, N. A. (2003) 'Bottom-up Histogenesis of Colorectal Adenomas', *Cancer Research*, 63(13), pp. 3819–3825.
- Pruitt, K. D., Harrow, J., Harte, R. A., Wallin, C., Diekhans, M., Maglott, D. R., Searle, S., Farrell, C. M., Loveland, J. E., Ruef, B. J., Hart, E., Suner, M.-M., Landrum, M. J., Aken, B., Ayling, S., Baertsch, R., Fernandez-Banet, J., Cherry, J. L., Curwen, V., Dicuccio, M., Kellis, M., Lee, J., Lin, M. F., Schuster, M., Shkeda, A., Amid, C., Brown, G., Dukhanina, O., Frankish, A., Hart, J., Maidak, B. L., Mudge, J., Murphy, M. R., Murphy, T., Rajan, J., Rajput, B., Riddick, L. D., Snow, C., Steward, C., Webb, D., Weber, J. A., Wilming, L., Wu, W., Birney, E., Haussler, D., Hubbard, T., Ostell, J., Durbin, R., & Lipman, D. (2009) 'The

consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes', *Genome research*, 19(7), pp. 1316–1323.

Quintanilla, I., López-Cerón, M., Jimeno, M., Cuatrecasas, M., Zabalza, M., Moreira, L., Alonso, V., Rodríguez de Miguel, C., Muñoz, J., Castellvi-Bel, S., Llach, J., Castells, A., Balaguer, F., Camps, J., & Pellisé, M. (2019) 'Rectal Aberrant Crypt Foci in Humans Are Not Surrogate Markers for Colorectal Cancer Risk', *Clinical and Translational Gastroenterology*, 10(6).

Ritsma, L., Ellenbroek, S. I. J., Zomer, A., Snippert, H. J., de Sauvage, F. J., Simons, B. D., Clevers, H., & van Rheenen, J. (2014) 'Intestinal crypt homeostasis revealed at single-stem-cell level by in vivo live imaging', *Nature*, 507(7492), pp. 362–365.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011) 'Integrative genomics viewer', *Nature biotechnology*, 29(1), pp. 24–26.

Rojo, F., Najera, L., Lirola, J., Jiménez, J., Guzmán, M., Sabadell, M. D., Baselga, J., & Cajal, S. R. y (2007) '4E-Binding Protein 1, A Cell Signaling Hallmark in Breast Cancer that Correlates with Pathologic Grade and Prognosis', *Clinical Cancer Research*, 13(1), pp. 81–89.

Roncucci, L., Stamp, D., Medline, A., Cullen, J. B., & Robert Bruce, W. (1991) 'Identification and quantification of aberrant crypt foci and microadenomas in the human colon', *Human Pathology*, 22(3), pp. 287–294.

Rothenberg, M. E., Nusse, Y., Kalisky, T., Lee, J. J., Dalerba, P., Scheeren, F., Lobo, N., Kulkarni, S., Sim, S., Qian, D., Beachy, P. A., Pasricha, P. J., Quake, S. R., & Clarke, M. F. (2012) 'Identification of a cKit(+) colonic crypt base secretory cell that supports Lgr5(+) stem cells in mice', *Gastroenterology*, 142(5), pp. 1195-1205.e6.

Roulis, M., Armaka, M., Manoloukos, M., Apostolaki, M., & Kollias, G. (2011) 'Intestinal epithelial cells as producers but not targets of chronic TNF suffice to cause murine Crohn-like pathology', *Proceedings of the National Academy of Sciences of the United States of America*. 2011/03/14. National Academy of Sciences, 108(13), pp. 5396–5401.

- Rozen, S. & Skaletsky, H. (2000) 'Primer3 on the WWW for general users and for biologist programmers.', *Methods in Molecular Biology*, 132, pp. 365–86.
- Sangiorgi, E. & Capecchi, M. R. (2008) 'Bmi1 is expressed in vivo in intestinal stem cells', *Nat Genet*, 40(7), pp. 915–920.
- Sasaki, N., Sachs, N., Wiebrands, K., Ellenbroek, S. I. J., Fumagalli, A., Lyubimova, A., Begthel, H., van den Born, M., van Es, J. H., Karthaus, W. R., Li, V. S. W., López-Iglesias, C., Peters, P. J., van Rheenen, J., van Oudenaarden, A., & Clevers, H. (2016) 'Reg4<sup>+</sup> deep crypt secretory cells function as epithelial niche for Lgr5<sup>+</sup> stem cells in colon', *Proceedings of the National Academy of Sciences of the United States of America*, 113(37), pp. E5399–E5407.
- Sato, T., van Es, J. H., Snippert, H. J., Stange, D. E., Vries, R. G., van den Born, M., Barker, N., Shroyer, N. F., van de Wetering, M., & Clevers, H. (2011) 'Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts', *Nature*, 469(7330), pp. 415–418.
- Schiffner, S., Zimara, N., Schmid, R., & Bosserhoff, A.-K. (2011) 'p54 nrb is a new regulator of progression of malignant melanoma ', *Carcinogenesis*, 32(8), pp. 1176–1182.
- Schmidt, G. H., Winton, D. J., & Ponder, B. A. (1988) 'Development of the pattern of cell renewal in the crypt-villus unit of chimaeric mouse small intestine', *Development*, 103(4), pp. 785–790.
- Schmitt, M., Schewe, M., Sacchetti, A., Feijtel, D., van de Geer, W. S., Teeuwssen, M., Sleddens, H. F., Joosten, R., van Royen, M. E., van de Werken, H. J. G., van Es, J., Clevers, H., & Fodde, R. (2018) 'Paneth Cells Respond to Inflammation and Contribute to Tissue Regeneration by Acquiring Stem-like Features through SCF/c-Kit Signaling', *Cell Reports*, 24(9), pp. 2312-2328.e7.
- Schonhoff, S. E., Giel-Moloney, M., & Leiter, A. B. (2004) 'Neurogenin 3-expressing progenitor cells in the gastrointestinal tract differentiate into both endocrine and non-endocrine cell types', *Developmental Biology*, 270(2), pp. 443–454.

- Schwitalla, S., Fingerle, A. A., Cammareri, P., Nebelsiek, T., Göktuna, S. I., Ziegler, P. K., Canli, O., Heijmans, J., Huels, D. J., Moreaux, G., Rupec, R. A., Gerhard, M., Schmid, R., Barker, N., Clevers, H., Lang, R., Neumann, J., Kirchner, T., Taketo, M. M., van den Brink, G. R., Sansom, O. J., Arkan, M. C., & Greten, F. R. (2013) 'Intestinal Tumorigenesis Initiated by Dedifferentiation and Acquisition of Stem-Cell-like Properties', *Cell*, 152(1), pp. 25–38.
- Shaked, H., Hofseth, L. J., Chumanevich, A., Chumanevich, A. A., Wang, J., Wang, Y., Taniguchi, K., Guma, M., Shenouda, S., Clevers, H., Harris, C. C., & Karin, M. (2012) 'Chronic epithelial NF- $\kappa$ B activation accelerates APC loss and intestinal tumor initiation through iNOS up-regulation', *Proceedings of the National Academy of Sciences of the United States of America*. 2012/08/14. National Academy of Sciences, 109(35), pp. 14007–14012.
- Shaukat, A., Dostal, A., Menk, J., & Church, T. R. (2017) 'BMI Is a Risk Factor for Colorectal Cancer Mortality', *Digestive Diseases and Sciences*, 62(9), pp. 2511–2517.
- Shay, G., Lynch, C. C., & Fingleton, B. (2015) 'Moving targets: Emerging roles for MMPs in cancer progression and metastasis', *Matrix biology: journal of the International Society for Matrix Biology*. 2015/01/31, 44–46, pp. 200–206.
- Shoshkes-Carmel, M., Wang, Y. J., Wangenstein, K. J., Tóth, B., Kondo, A., Massassa, E. E., Itzkovitz, S., & Kaestner, K. H. (2018) 'Subepithelial telocytes are an important source of Wnts that supports intestinal crypts', *Nature*.
- Siegmund, K. D., Marjoram, P., Tavaré, S., & Shibata, D. (2009) 'Many colorectal cancers are "flat" clonal expansions', *Cell Cycle*. Taylor & Francis, 8(14), pp. 2187–2193.
- Sievers, C. K., Zou, L. S., Pickhardt, P. J., Matkowskyj, K. A., Albrecht, D. M., Clipson, L., Bacher, J. W., Pooler, B. D., Moawad, F. J., Cash, B. D., Reichelderfer, M., Vo, T. N., Newton, M. A., Larget, B. R., & Halberg, R. B. (2017) 'Subclonal diversity arises early even in small colorectal tumours and contributes to differential growth fates', *Gut*. 2016/09/08. BMJ Publishing Group, 66(12), pp. 2132–2140.
- Silva, J. M., Bagni, R. K., Ruggero, D., & McCormick, F. (2017) 'Abstract B11: Oncogenic KRAS regulates 4E-BP1, a repressor of cap-dependent translation, independently of growth factor activity', *Cancer Research*, 77(6 Supplement), p. B11.

Sivridis, E., Giatromanolaki, A., & Koukourakis, M. I. (2005) 'Proliferating fibroblasts at the invading tumour edge of colorectal adenocarcinomas are associated with endogenous markers of hypoxia, acidity, and oxidative stress', *Journal of clinical pathology*. Copyright 2005 *Journal of Clinical Pathology*, 58(10), pp. 1033–1038.

Smit, A. F. A., Hubley, R., & Green, P. (no date) 'RepeatMasker Open-4.0.'

Snippert, H. J., Haegebarth, A., Kasper, M., Jaks, V., van Es, J. H., Barker, N., van de Wetering, M., van den Born, M., Begthel, H., Vries, R. G., Stange, D. E., Toftgård, R., & Clevers, H. (2010) '&em&gt;Lgr6&lt;/em&gt; Marks Stem Cells in the Hair Follicle That Generate All Cell Lineages of the Skin', *Science*, 327(5971), pp. 1385 LP – 1389.

Snippert, H. J., van der Flier, L. G., Sato, T., van Es, J. H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A. M., van Rheenen, J., Simons, B. D., & Clevers, H. (2010) 'Intestinal Crypt Homeostasis Results from Neutral Competition between Symmetrically Dividing Lgr5 Stem Cells', *Cell*, 143(1), pp. 134–144.

Snippert, H. J., Schepers, A. G., van Es, J. H., Simons, B. D., & Clevers, H. (2013) 'Biased competition between Lgr5 intestinal stem cells driven by oncogenic mutation induces clonal expansion', *EMBO reports*, p. e201337799.

Snyder, C. & Hampel, H. (2019) 'Hereditary Colorectal Cancer Syndromes', *Seminars in Oncology Nursing*, 35(1), pp. 58–78.

Soardi, F. C., Machado-Silva, A., Linhares, N. D., Zheng, G., Qu, Q., Pena, H. B., Martins, T. M. M., Vieira, H. G. S., Pereira, N. B., Melo-Minardi, R. C., Gomes, C. C., Gomez, R. S., Gomes, D. A., Pires, D. E. V., Ascher, D. B., Yu, H., & Pena, S. D. J. (2017) 'Familial STAG2 germline mutation defines a new human cohesinopathy', *NPJ Genomic Medicine*, 2(1), p. 7.

Solomon, D. A., Kim, T., Diaz-Martinez, L. A., Fair, J., Elkahloun, A. G., Harris, B. T., Toretsky, J. A., Rosenberg, S. A., Shukla, N., Ladanyi, M., Samuels, Y., James, C. D., Yu, H., Kim, J.-S., & Waldman, T. (2011) 'Mutational Inactivation of *STAG2* Causes

- Aneuploidy in Human Cancer’, *Science*, 333(6045), pp. 1039–1043.
- Sottoriva, A., Kang, H., Ma, Z., Graham, T. A., Salomon, M. P., Zhao, J., Marjoram, P., Siegmund, K., Press, M. F., Shibata, D., & Curtis, C. (2015) ‘A Big Bang model of human colorectal tumor growth’, *Nature genetics*, 47(3), pp. 209–216.
- Spalding, K. L., Bhardwaj, R. D., Buchholz, B. A., Druid, H., & Frisén, J. (2005) ‘Retrospective Birth Dating of Cells in Humans’, *Cell*, 122(1), pp. 133–143.
- Spit, M., Koo, B.-K., & Maurice, M. M. (2018) ‘Tales from the crypt: intestinal niche signals in tissue renewal, plasticity and cancer’, *Open biology*. The Royal Society, 8(9), p. 180120.
- Staffa, L., Echterdiek, F., Nelius, N., Benner, A., Werft, W., Lahrmann, B., Grabe, N., Schneider, M., Tariverdian, M., von Knebel Doeberitz, M., Bläker, H., & Kloor, M. (2015) ‘Mismatch Repair-Deficient Crypt Foci in Lynch Syndrome – Molecular Alterations and Association with Clinical Parameters’, *PLOS ONE*, 10(3), p. e0121980.
- Sternberg, N. & Hamilton, D. (1981) ‘Bacteriophage P1 site-specific recombination: I. Recombination between loxP sites’, *Journal of Molecular Biology*, 150(4), pp. 467–486.
- Strober, W., Fuss, I., & Mannon, P. (2007) ‘The fundamental basis of inflammatory bowel disease’, *The Journal of clinical investigation*. American Society for Clinical Investigation, 117(3), pp. 514–521.
- Sugimoto, S., Ohta, Y., Fujii, M., Matano, M., Shimokawa, M., Nanki, K., Date, S., Nishikori, S., Nakazato, Y., Nakamura, T., Kanai, T., & Sato, T. (2018) ‘Reconstruction of the Human Colon Epithelium In Vivo’, *Cell Stem Cell*, 22(2), pp. 171-176.e5.
- Suraweera, A., O’Byrne, K. J., & Richard, D. J. (2018) ‘Combination Therapy With Histone Deacetylase Inhibitors (HDACi) for the Treatment of Cancer: Achieving the Full Therapeutic Potential of HDACi’, *Frontiers in oncology*. Frontiers Media S.A., 8, p. 92.
- Takayama, T., Katsuki, S., Takahashi, Y., Ohi, M., Nojiri, S., Sakamaki, S., Kato, J., Kogawa, K., Miyake, H., & Niitsu, Y. (1998) ‘Aberrant Crypt Foci of the Colon as Precursors of Adenoma and Cancer’, *New England Journal of Medicine*. Massachusetts Medical Society, 339(18), pp. 1277–1284.

- Takayama, T., Ohi, M., Hayashi, T., Miyanishi, K., Nobuoka, A., Nakajima, T., Satoh, T., Takimoto, R., Kato, J., Sakamaki, S., & Niitsu, Y. (2001) 'Analysis of K-*ras*, APC, and  $\beta$ -catenin in aberrant crypt foci in sporadic adenoma, cancer, and familial adenomatous polyposis', *Gastroenterology*. Elsevier, 121(3), pp. 599–611.
- Takeda, N., Jain, R., LeBoeuf, M. R., Wang, Q., Lu, M. M., & Epstein, J. A. (2011) 'Interconversion Between Intestinal Stem Cell Populations in Distinct Niches', *Science*, 334(6061), pp. 1420–1424.
- Taylor, R. W., Barron, M. J., Borthwick, G. M., Gospel, A., Chinnery, P. F., Samuels, D. C., Taylor, G. A., Plusa, S. M., Needham, S. J., Greaves, L. C., Kirkwood, T. B. L., & Turnbull, D. M. (2003) 'Mitochondrial DNA mutations in human colonic crypt stem cells', *The Journal of Clinical Investigation*, 112(9), pp. 1351–1360.
- Taylor, R. W. & Turnbull, D. M. (2005) 'Mitochondrial DNA mutations in human disease', *Nat Rev Genet*, 6(5), pp. 389–402.
- Terzić, J., Grivennikov, S., Karin, E., & Karin, M. (2010) 'Inflammation and Colon Cancer', *Gastroenterology*. Elsevier, 138(6), pp. 2101-2114.e5.
- Thirlwell, C., Will, O. C. C., Domingo, E., Graham, T. A., McDonald, S. A. C., Oukrif, D., Jeffrey, R., Gorman, M., Rodriguez-Justo, M., Chin-Aleong, J., Clark, S. K., Novelli, M. R., Jankowski, J. A., Wright, N. A., Tomlinson, I. P. M., & Leedham, S. J. (2010) 'Clonality Assessment and Clonal Ordering of Individual Neoplastic Crypts Shows Polyclonality of Colorectal Adenomas', *Gastroenterology*, 138(4), pp. 1441-1454.e7.
- Tian, H., Biehs, B., Warming, S., Leong, K. G., Rangell, L., Klein, O. D., & de Sauvage, F. J. (2011) 'A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable', *Nature*, 478(7368), pp. 255–259.
- Tomasetti, C., Li, L., & Vogelstein, B. (2017) 'Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention', *Science*, 355(6331), pp. 1330–1334.

- Tomasetti, C. & Vogelstein, B. (2015) 'Variation in cancer risk among tissues can be explained by the number of stem cell divisions', *Science*, 347(6217), pp. 78–81.
- Tomasetti, C., Vogelstein, B., & Parmigiani, G. (2013) 'Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation', *Proceedings of the National Academy of Sciences of the United States of America*, 110(6), pp. 1999–2004.
- Totafurno, J., Bjerknes, M., & Cheng, H. (1987) 'The crypt cycle. Crypt and villus production in the adult intestinal epithelium', *Biophysical journal*, 52(2), pp. 279–294.
- Traish, A. M., Huang, Y.-H., Ashba, J., Pronovost, M., Pavao, M., McAneny, D. B., & Moreland, R. B. (1997) 'Loss of Expression of a 55 kDa Nuclear Protein (nmt55) in Estrogen Receptor-Negative Human Breast Cancer.', *Diagnostic Molecular Pathology*, 6(4), pp. 209–221.
- Tukiainen, T., Villani, A.-C., Yen, A., Rivas, M. A., Marshall, J. L., MacArthur, D. G., *et al.* (2017) 'Landscape of X chromosome inactivation across human tissues', *Nature*, 550(7675), pp. 244–248.
- Turner, N. D. & Lloyd, S. K. (2017) 'Association between red meat consumption and colon cancer: A systematic review of experimental results', *Experimental biology and medicine*, 242(8), pp. 813–839.
- Unni, A. M., Harbour, B., Oh, M. H., Wild, S., Ferrarone, J. R., Lockwood, W. W., & Varmus, H. (2018) 'Hyperactivation of ERK by multiple mechanisms is toxic to RTK-RAS mutation-driven lung adenocarcinoma cells', *eLife*, 7, p. e33718.
- Vavougios, G. D., Solenov, E. I., Hatzoglou, C., Baturina, G. S., Katkova, L. E., Molyvdas, P. A., Gourgoulanis, K. I., & Zarogiannis, S. G. (2015) 'Computational genomic analysis of PARK7 interactome reveals high BBS1 gene expression as a prognostic factor favoring survival in malignant pleural mesothelioma', *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 309(7), pp. L677–L686.
- Veh, R. W., Meessen, D., Kuntz, H. D., & May, B. (1982) 'A new method for histochemical demonstration of side chain substituted sialic acids.', *Colonic Carcinogenesis*, pp. 355–365.

- Verdel, A., Curtet, S., Brocard, M.-P., Rousseaux, S., Lemerrier, C., Yoshida, M., & Khochbin, S. (2000) 'Active maintenance of mHDA2/mHDAC6 histone-deacetylase in the cytoplasm', *Current Biology*, 10(12), pp. 747–749.
- Vermeulen, L., Morrissey, E., van der Heijden, M., Nicholson, A. M., Sottoriva, A., Buczacki, S., Kemp, R., Tavaré, S., & Winton, D. J. (2013) 'Defining Stem Cell Dynamics in Models of Intestinal Tumor Initiation', *Science*, 342(6161), pp. 995–998.
- Vieler, M. & Sanyal, S. (2018) 'p53 Isoforms and Their Implications in Cancer', *Cancers*, 10(9), p. 288.
- de Visser, K. E., Eichten, A., & Coussens, L. M. (2006) 'Paradoxical roles of the immune system during cancer development', *Nature Reviews Cancer*, 6(1), pp. 24–37.
- Vogelstein, B., Lane, D., & Levine, A. J. (2000) 'Surfing the p53 network', *Nature*, 408(6810), pp. 307–310.
- Vousden, K. H. & Prives, C. (2009) 'Blinded by the Light: The Growing Complexity of p53', *Cell*, 137(3), pp. 413–431.
- Walport, L. J., Hopkinson, R. J., Vollmar, M., Madden, S. K., Gileadi, C., Oppermann, U., Schofield, C. J., & Johansson, C. (2014) 'Human UTY(KDM6C) Is a Male-specific N(ε)-Methyl Lysyl Demethylase', *The Journal of Biological Chemistry*, 289(26), pp. 18302–18313.
- Wargovich, M. J., Brown, V. R., & Morris, J. (2010) 'Aberrant crypt foci: the case for inclusion as a biomarker for colon cancer', *Cancers*. MDPI, 2(3), pp. 1705–1716.
- Wasan, H. S., Park, H.-S., Liu, K. C., Mandir, N. K., Winnett, A., Sasieni, P., Bodmer, W. F., Goodlad, R. A., & Wright, N. A. (1998) 'APC in the regulation of intestinal crypt fission', *The Journal of Pathology*, 185(3), pp. 246–255.
- Weissman, I. L. (2000) 'Stem cells: units of development, units of regeneration, and units in evolution.', *Cell*, 100(1), pp. 157–168.

- Weissman, I. L. & Shizuru, J. A. (2008) 'The origins of the identification and isolation of hematopoietic stem cells, and their capability to induce donor-specific transplantation tolerance and treat autoimmune diseases', *Blood*, 112(9), pp. 3543–3553.
- Welch, J. S., Ley, T. J., Link, D. C., Miller, C. A., Larson, D. E., Wilson, R. K., *et al.* (2012) 'The origin and evolution of mutations in acute myeloid leukemia', *Cell*, 150(2), pp. 264–278.
- Westbrook, A. M., Wei, B., Braun, J., & Schiestl, R. H. (2009) 'Intestinal mucosal inflammation leads to systemic genotoxicity in mice', *Cancer research*, 69(11), pp. 4827–4834.
- Winner, B., Cooper-Kuhn, C. M., Aigner, R., Winkler, J., & Kuhn, H. G. (2002) 'Long-term survival and cell death of newly generated neurons in the adult rat olfactory bulb', *European Journal of Neuroscience*, 16(9), pp. 1681–1689.
- Winton, D. J., Blount, M. A., & Ponder, B. A. J. (1988) 'A clonal marker induced by mutation in mouse intestinal epithelium', *Nature*, 333(6172), pp. 463–466.
- Wong, W.-M., Mandir, N., Goodlad, R. A., Wong, B. C. Y., Garcia, S. B., Lam, S.-K., & Wright, N. A. (2002) 'Histogenesis of human colorectal adenomas and hyperplastic polyps: the role of cell proliferation and crypt fission', *Gut*, 50(2), pp. 212 LP – 217.
- Wood, L. D., Parsons, D. W., Jones, S., Lin, J., Sjöblom, T., Leary, R. J., Shen, D., Boca, S. M., Barber, T., Ptak, J., Silliman, N., Szabo, S., Dezso, Z., Ustyanksky, V., Nikolskaya, T., Nikolsky, Y., Karchin, R., Wilson, P. A., Kaminker, J. S., Zhang, Z., Croshaw, R., Willis, J., Dawson, D., Shipitsin, M., Willson, J. K. V, Sukumar, S., Polyak, K., Park, B. H., Pethiyagoda, C. L., Pant, P. V. K., Ballinger, D. G., Sparks, A. B., Hartigan, J., Smith, D. R., Suh, E., Papadopoulos, N., Buckhaults, P., Markowitz, S. D., Parmigiani, G., Kinzler, K. W., Velculescu, V. E., & Vogelstein, B. (2007) 'The Genomic Landscapes of Human Breast and Colorectal Cancers', *Science*, 318(5853), pp. 1108–1113.
- Wright, N. A. & Al-Nafussi, A. (1982) 'The kinetics of villus cell populations in the mouse small intestine. II. Studies on growth control after death of proliferative cells induced by cytosine arabinoside, with special reference to negative feedback mechanisms.', *Cell Tissue Kinet.*, 15(6), pp. 611–21.

- Wu, S., Powers, S., Zhu, W., & Hannun, Y. A. (2016) 'Substantial contribution of extrinsic risk factors to cancer development', *Nature*, 529(7584), pp. 43–47.
- Xie, M., Lu, C., Wang, J., McLellan, M. D., Johnson, K. J., Wendl, M. C., McMichael, J. F., Schmidt, H. K., Yellapantula, V., Miller, C. A., Ozenberger, B. A., Welch, J. S., Link, D. C., Walter, M. J., Mardis, E. R., Dpersio, J. F., Chen, F., Wilson, R. K., Ley, T. J., & Ding, L. (2014) 'Age-related mutations associated with clonal hematopoietic expansion and malignancies', *Nature medicine*, 20(12), pp. 1472–1478.
- Yan, K. S., Chia, L. A., Li, X., Ootani, A., Su, J., Lee, J. Y., Su, N., Luo, Y., Heilshorn, S. C., Amieva, M. R., Sangiorgi, E., Capecchi, M. R., & Kuo, C. J. (2012) 'The intestinal stem cell markers *Bmi1* and *Lgr5* identify two functionally distinct populations', *Proceedings of the National Academy of Sciences*, 109(2), pp. 466 LP – 471.
- Yan, K. S., Gevaert, O., Zheng, G. X. Y., Anchang, B., Probert, C. S., Larkin, K. A., Davies, P. S., Cheng, Z.-F., Kaddis, J. S., Han, A., Roelf, K., Calderon, R. I., Cynn, E., Hu, X., Mandleywala, K., Wilhelmy, J., Grimes, S. M., Corney, D. C., Boutet, S. C., Terry, J. M., Belgrader, P., Ziraldo, S. B., Mikkelsen, T. S., Wang, F., von Furstenberg, R. J., Smith, N. R., Chandrakesan, P., May, R., Chrissy, M. A. S., Jain, R., Cartwright, C. A., Niland, J. C., Hong, Y.-K., Carrington, J., Breault, D. T., Epstein, J., Houchen, C. W., Lynch, J. P., Martin, M. G., Plevritis, S. K., Curtis, C., Ji, H. P., Li, L., Henning, S. J., Wong, M. H., & Kuo, C. J. (2017) 'Intestinal Enteroendocrine Lineage Cells Possess Homeostatic and Injury-Inducible Stem Cell Activity', *Cell stem cell*, 21(1), pp. 78-90.e6.
- Yang, C.-Y., Tseng, J.-Y., Chen, C.-F., Chou, T.-Y., Gao, H.-W., Hua, C.-L., Lin, C.-H., Lin, J.-K., & Jiang, J.-K. (2015) 'Genome-wide copy number changes and CD133 expression characterized distinct subset of colon polyps: differentiation between incidental polyps and cancer-associated polyps', *International Journal of Colorectal Disease*, 30(12), pp. 1617–1626.
- Yatabe, Y., Tavaré, S., & Shibata, D. (2001) 'Investigating stem cells in human colon by using methylation patterns', *Proceedings of the National Academy of Sciences of the United*

---

*States of America*, 98(19), pp. 10839–10844.

Yau, S., Li, A., & So, K.-F. (2015) ‘Involvement of Adult Hippocampal Neurogenesis in Learning and Forgetting’, *Neural plasticity*, 2015, p. 717958.

Yehia, L., Ngeow, J., & Eng, C. (2019) ‘PTEN-opathies: from biological insights to evidence-based precision medicine.’, *The Journal of Clinical Investigation*, 129(2), pp. 452–464.

Yizhak, K., Aguet, F., Kim, J., Hess, J. M., Kübler, K., Grimsby, J., Frazer, R., Zhang, H., Haradhvala, N. J., Rosebrock, D., Livitz, D., Li, X., Arich-Landkof, E., Shores, N., Stewart, C., Segrè, A. V., Branton, P. A., Polak, P., Ardlie, K. G., & Getz, G. (2019) ‘RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues’, *Science*, 364(6444), p. eaaw0726.

Yu, S., Tong, K., Zhao, Y., Balasubramanian, I., Yap, G. S., Ferraris, R. P., Bonder, E. M., Verzi, M. P., & Gao, N. (2018) ‘Paneth Cell Multipotency Induced by Notch Activation following Injury’, *Cell stem cell*. 2018/06/07, 23(1), pp. 46-59.e5.

Yue, X., Zhao, Y., Huang, G., Li, J., Zhu, J., Feng, Z., & Hu, W. (2016) ‘A novel mutant p53 binding partner BAG5 stabilizes mutant p53 and promotes mutant p53 GOFs in tumorigenesis’, *Cell discovery*, 2, p. 16039.

Yue, X., Zhao, Y., Xu, Y., Zheng, M., Feng, Z., & Hu, W. (2017) ‘Mutant p53 in Cancer: Accumulation, Gain-of-Function, and Therapy’, *Journal of molecular biology*, 429(11), pp. 1595–1606.

Yurgelun, M. B. & Hampel, H. (2018) ‘Recent Advances in Lynch Syndrome: Diagnosis, Treatment, and Cancer Prevention’, *American Society of Clinical Oncology Educational Book*, (38), pp. 101–109.

Zackular, J. P., Baxter, N. T., Iverson, K. D., Sadler, W. D., Petrosino, J. F., Chen, G. Y., & Schloss, P. D. (2013) ‘The Gut Microbiome Modulates Colon Tumorigenesis’, *mBio*, 4(6), pp. e00692-13.

Zehir, A., Benayed, R., Shah, R. H., Syed, A., Middha, S., Berger, M. F., *et al.* (2017) ‘Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of

10,000 patients', *Nature Medicine*, 23(6), pp. 703–713.

Zhang, G. & Gan, Y.-H. (2017) 'Synergistic antitumor effects of the combined treatment with an HDAC6 inhibitor and a COX-2 inhibitor through activation of PTEN', *Oncology Reports*, 38(5), pp. 2657–2666.

Zhang, X., Yuan, Z., Zhang, Y., Yong, S., Salas-Burgos, A., Koomen, J., Olashaw, N., Parsons, J. T., Yang, X.-J., Dent, S. R., Yao, T.-P., Lane, W. S., & Seto, E. (2007) 'HDAC6 modulates cell motility by altering the acetylation level of cortactin', *Molecular cell*, 27(2), pp. 197–213.

Zheng, J., Zhao, M., Li, J., Lou, G., Yuan, Y., Bu, S., & Xi, Y. (2017) 'Obesity-associated digestive cancers: A review of mechanisms and interventions', *Tumor Biology*, 39(3), p. 1010428317695020.

Zheng, T., Wang, J., Zhao, Y., Zhang, C., Lin, M., Wang, X., Yu, H., Liu, L., Feng, Z., & Hu, W. (2013) 'Spliced MDM2 isoforms promote mutant p53 accumulation and gain-of-function in tumorigenesis', *Nature communications*, 4, p. 2996.

Zhu, D., Keohavong, P., Finkelstein, S. D., Swalsky, P., Bakker, A., Weissfeld, J., Srivastava, S., & Whiteside, T. L. (1997) 'K-ras Gene Mutations in Normal Colorectal Tissues from K-ras Mutation-positive Colorectal Cancer Patients', *Cancer Research*, 57(12), pp. 2485–2492.

Zhu, Z., Zhao, X., Zhao, L., Yang, H., Liu, L., Li, J., Wu, J., Yang, F., Huang, G., & Liu, J. (2015) 'p54nrb/NONO regulates lipid metabolism and breast cancer growth through SREBP-1A', *Oncogene*, 35(11), pp. 1399–1410.





# Appendix A Patients included

*Patient ID = Number assigned in the Winton laboratory for anonymisation.*

*Columns 4–15 denote different experimental parts.*

*TRUE = Patient used, FALSE = Patient not used*

Patient_ID	Age	Sex	HDAC6	KDM6A	MAOA	mPAS	NONO	P53	PTEN	STAG2	KRAS	mPAS fusion	KDM6A fusion	Fusion duration
CDA_00002	59	M	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
CDA_00003	85	F	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE						
CDA_00005	60	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00008	84	M	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00010	60	M	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00013	52	M	FALSE	TRUE	FALSE	FALSE	FALSE							
CDA_00016	66	F	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE
CDA_00019	66	M	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00021	88	F	FALSE	TRUE	FALSE	FALSE	FALSE							
CDA_00022	66	M	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
CDA_00024	87	F	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
CDA_00025	66	M	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00026	78	M	FALSE	TRUE	FALSE	FALSE	FALSE							
CDA_00027	76	M	FALSE	TRUE	FALSE	FALSE	FALSE							
CDA_00028	84	M	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00029	78	F	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
CDA_00030	64	F	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00031	79	M	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CDA_00033	68	F	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
CDA_00034	73	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00035	69	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
CDA_00037	91	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00039	84	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CDA_00041	72	F	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00043	76	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00046	77	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00048	77	M	FALSE	TRUE	TRUE	FALSE	FALSE							
CDA_00049	78	F	FALSE	TRUE	FALSE	FALSE	FALSE							
CDA_00050	73	M	FALSE	TRUE	FALSE	FALSE								
CDA_00051	79	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00052	85	M	FALSE	TRUE	TRUE	FALSE	FALSE							
CDA_00054	64	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE						
CDA_00058	58	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE						
CDA_00060	60	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE

CDA_00061	73	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CDA_00063	76	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00064	52	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CDA_00065	67	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00066	84	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CDA_00067	88	F	FALSE	TRUE	TRUE	FALSE	FALSE							
CDA_00068	81	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
CDA_00069	68	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00070	88	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00071	62	F	FALSE	TRUE	TRUE	FALSE	FALSE							
CDA_00073	76	F	FALSE	TRUE	FALSE	FALSE	FALSE							
CDA_00074	79	M	FALSE	TRUE	TRUE	FALSE								
CDA_00075	78	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00076	81	M	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE
CDA_00077	66	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00078	43	M	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CDA_00080	64	M	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE
CDA_00081	70	F	FALSE	FALSE	TRUE	FALSE								
CDA_00082	76	F	FALSE	TRUE	TRUE	FALSE	FALSE							
CDA_00083	61	M	FALSE	TRUE	FALSE	FALSE								
CDA_00084	45	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
CDA_00086	83	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00087	79	F	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE
CDA_00088	41	F	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE
CDA_00089	82	F	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE
CDA_00090	72	M	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CDA_00091	86	M	FALSE	TRUE	FALSE	FALSE	FALSE							
CDA_00093	43	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00095	81	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CDA_00096	80	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
CDA_00097	77	F	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE
CDA_00099	87	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00100	66	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CDA_00102	83	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00104	72	F	FALSE	TRUE	FALSE	FALSE	FALSE							
CDA_00105	67	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00107	84	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CDA_00108	85	F	FALSE	TRUE	FALSE	FALSE	FALSE							
CDA_00109	59	M	FALSE	TRUE	FALSE	FALSE	FALSE							
CDA_00111	78	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00112	33	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CDA_00113	62	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00115	53	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CDA_00116	68	M	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CDA_00163	63	M	FALSE	FALSE	FALSE	TRUE	FALSE							
CRA_00117	73	M	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
CRA_00118	74	M	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
CRA_00119	83	M	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
CRA_00121	60	M	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
CRA_00125	75	M	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
CRA_00126	70	M	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
CRA_00127	85	M	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE

CRA_00128	70	F	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE
CRA_00129	93	M	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE
CRA_00130	55	F	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE
CRA_00131	62	M	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
CRA_00132	74	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
CRA_00133	68	F	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
CRA_00134	64	F	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE
CRA_00182	68	M	FALSE	FALSE	TRUE	FALSE								
CRA_00183	69	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
CRA_00184	64	F	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
CRA_00185	86	M	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE						
CRA_00186	66	M	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CRA_00294	89	M	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
CRA_00295	85	F	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
CRA_00296	50	M	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CRA_00297	50	M	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE
CRA_00298	76	F	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE
CRA_00299	37	F	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE
CRA_00300	60	F	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE
CRA_00301	69	M	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE
CRA_00302	83	M	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE
CRA_00303	57	F	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
CRA_00304	60	M	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
CRA_00305	51	M	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE
CRA_00306	76	F	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE
CRA_00307	62	F	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE
CRA_00308	73	F	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
CRA_00309	80	F	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
CRA_00310	78	F	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
CRA_00311	62	M	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE
CRA_00312	67	M	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
CRA_00313	57	M	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
CRA_00314	89	M	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
CRA_00315	60	F	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
CRA_00316	85	M	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
CRA_00317	69	M	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
CRA_00318	48	M	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
CRA_00319	89	F	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CRA_00347	68	M	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
CRA_00348	56	F	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE
CRA_00349	37	M	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
CRA_00350	70	M	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
CRA_00351	68	F	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
CRA_00353	83	F	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
CRA_00354	83	F	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
CRA_00355	65	F	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
CRA_00356	82	M	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
CRA_00357	79	M	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
CRA_00358	87	F	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CRA_00359	69	M	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CRA_00360	65	F	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE
CRA_00361	65	F	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE
CRA_00362	85	F	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
CRA_00363	72	M	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE

CRA_00364	74	M	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
CRA_00365	74	F	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
CRA_00366	77	F	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
CRA_00367	82	M	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
CRA_00370	69	M	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE
CRA_00371	60	F	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
CRA_00372	81	M	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
CRA_00373	73	F	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
CRA_00374	84	M	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
CRA_00409	77	F	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE
CRA_00410	55	F	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
CRA_00411	90	M	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
CRA_00412	68	F	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE
CRA_00413	65	F	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE
CRA_00414	72	F	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE
CRA_00415	85	M	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE
CRA_00418	81	M	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
CRA_00419	46	F	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE
CRA_00420	67	M	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CRA_00421	77	M	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CRA_00422	30	M	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
CRA_00423	31	M	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CRA_00424	55	F	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CRA_00425	36	M	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE
CRA_00426	64	M	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE
CRA_00427	67	F	TRUE	FALSE	TRUE	TRUE								
CRA_00428	73	F	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE
CRA_00429	63	M	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE
CRA_00430	69	M	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE
CRA_00431	75	F	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE
CRA_00432	73	M	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CRA_00433	80	F	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
CRA_00434	76	F	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE
CRA_00435	72	M	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
CRA_00436	84	F	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE
CRA_00437	55	M	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CRA_00438	76	F	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
CRA_00439	71	M	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
CRA_00440	65	M	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE
CRA_00441	69	M	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
CRA_00442	64	M	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE
CRA_00443	82	M	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE
CRA_00444	59	F	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
CRA_00445	61	M	FALSE	TRUE	FALSE	FALSE	FALSE							
CRA_00446	44	M	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CRA_00447	78	F	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE
CRA_00448	48	F	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE
CRA_00449	79	F	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE						
CRA_00450	40	M	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CRA_00451	84	M	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
CRA_00452	33	M	FALSE	TRUE	FALSE	FALSE	FALSE							
CRA_00453	34	M	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE





NDX_00334	73	F	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE						
NDX_00335	75	F	FALSE	TRUE	TRUE	FALSE	FALSE							
NDX_00336	41	F	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE						
NDX_00337	74	F	FALSE	TRUE	FALSE	FALSE	FALSE							
NDX_00338	65	F/M	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
NDX_00339	74	F	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
NDX_00341	64	F	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE						
NDX_00342	42	F	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE						
NDX_00343	74	M	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE						
NDX_00344	64	M	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
NDX_00369	68	F	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE
NDX_00370	68	M	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
NDX_00371	76	F	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
NDX_00372	58	F	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
NDX_00373	58	F	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE						
NDX_00374	60	M	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
NDX_00375	74	F	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
NDX_00376	64	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
NDX_00377	65	F	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
NDX_00378	62	M	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
NDX_00379	62	M	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
NDX_00380	59	M	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
NDX_00381	57	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
NDX_00382	65	M	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE						
NDX_00383	71	F	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
NDX_00384	8	M	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
NDX_00385	43	M	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
NDX_00386	68	F	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
NDX_00387	58	F	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE						
NDX_00388	69	F	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE						
NDX_00389	63	F	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE
NDX_00390	79	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
NDX_00393	78	M	FALSE	TRUE	FALSE	TRUE	TRUE							
NDX_00394	73	F	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE
NDX_00396	80	F	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
NDX_00397	75	M	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
NDX_00398	59	M	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE
NDX_00399	84	M	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
NDX_00400	65	F	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
NDX_00401	70	M	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
NDX_00402	44	F	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
NDX_00403	64	M	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
NDX_00404	81	F/M	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
NDX_00405	59	M	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
NDX_00406	74	M	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE
NDX_00407	71	M	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE						
NDX_00408	41	M	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
NRA_00345	72	F	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
NRA_00346	49	F	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE
NRA_00368	79	F	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
NRA_00369	72	F	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE
NRA_00416	77	F	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE
NRA_00417	46	F	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE







## Appendix B Primers

Table B.1 Primers used for FFPE DNA PCR amplicon size test

Primer names include gene name and size of resulting amplicon. LOC = non-coding RNA 107985193.

Primer	Sequence (5' -> 3')
MAOA_150_F1	ACCCATCAGTTACTCCTTCCC
MAOA_150_R1	GGGATTAAGCTGGGAGTTTCT
MAOA_150_F2	TAGCAGGGCCTTGAATCTGT
MAOA_150_R2	GATAGTGCCCAGAGTCACCA
STAG2_150_F1	GGAGAAGAAGACACAGTTGGATG
STAG2_150_R1	TTCTGTGAGGCATTTAGGGAAAA
STAG2_150_F2	CCTATGCTCGCACAACCTATGAG
STAG2_150_R2	GGAAGCCACACATCCTCTCT
CASD1_150_F1	ACCTGGAAACCCTATGCTCAA
CASD1_150_R1	TGCAGCTATACATGCCAACC
LOC_150_F1	TCGTCTGCTTCATCCTCCTC
LOC_150_R1	GCCTAACATGCTTGGACCAC
MAOA_250_F1	TGCAAGTCTTAGGTTGGTTGC
MAOA_250_R1	TCAGTAATGGGTCATGTGCAA
MAOA_250_F2	AAGACATGTAGGGTTGGGGC
MAOA_250_R2	CAGAACACCCTGCTCTAACCT
STAG2_250_F1	GACTCTAAGGCCAGGTCAGG
STAG2_250_R1	GGAGGTGAGTTGTGGTGTCT
STAG2_250_F2	GCCTAATCATTCTCCCTGACCT
STAG2_250_R2	TGGTGTCAAAAATCCATTCCCTC
CASD1_250_F1	GGTTAGAGGAAGACAAAAGTGGA
CASD1_250_R1	CCTCAGTCCACACTTTGATACAC
LOC_250_F1	AGCTTACCTCTTTGTCTCTTCT
LOC_250_R1	CAACCTCAAAGTATCACGTGGA
MAOA_350_F1	TTCTTCAGAAATTGAATCCTTG
MAOA_350_R1	CCTGGGAGAAAGCAAAATCA
MAOA_350_F2	TCCCGGAGTATCAGCAAAG
MAOA_350_R2	CATGAGAGACCCCAAACAC
STAG2_350_F1	TCCGAATATTTTTGGTGCATT

STAG2_350_R1	CAGAGCCTTGATGAGTGCTG
STAG2_350_F2	TCTGAAGGAATGCTATGGTATGAA
STAG2_350_R2	TTGTCAAGGGTCATAGACACAA
CASD1_350_F1	CTTTGGGAAGCTTTGCGTAAAA
CASD1_350_R1	CGATTCAGGAAGATGTAAGCCA
LOC_350_F1	TCAGGAATGATGGTCTACGTGA
LOC_350_R1	TCTCAGCTCTATTCCGTGAGT

**Table B.2 Primers used for amplification of MAOA.**

*Number = amplicon number, F = forward primer, R = reverse primer. Sequence includes Fluidigm CS adapters.*

Primer	Sequence (5' -> 3')
1_F	ACACTGACGACATGGTTCTACACTCCTGTGCCTACGACCC
1_R	TACGGTAGCAGAGACTTGGTCTGTTCCCCTACCCCTCACTG
2_F	ACACTGACGACATGGTTCTACATAAGCATTGAATGTTACGTTGCT
2_R	TACGGTAGCAGAGACTTGGTCTATATGTTCTTCCCTCCAACCCTGT
3_F	ACACTGACGACATGGTTCTACATGCTGCCAACTCTTGACTG
3_R	TACGGTAGCAGAGACTTGGTCTCTGTGATTCCAGTGGTGCC
4_F	ACACTGACGACATGGTTCTACATTTATAGGGGAGATGTAAGGCAC
4_R	TACGGTAGCAGAGACTTGGTCTGTTGGTCCCACATAAGCTCC
5_F	ACACTGACGACATGGTTCTACATGATTACGTAGATGTTGGTGGAG
5_R	TACGGTAGCAGAGACTTGGTCTAATGCTTCCCTATCAGTCACTT
6_F	ACACTGACGACATGGTTCTACACATTTCCGGGGCGCCTTTC
6_R	TACGGTAGCAGAGACTTGGTCTTGGGTCATGTGCAAACCTGAAC
7_F	ACACTGACGACATGGTTCTACACCGAGAAGAGGTGGCAGTTA
7_R	TACGGTAGCAGAGACTTGGTCTACTTTGTCCAGCAGATTTTGTCA
8_F	ACACTGACGACATGGTTCTACAGAGGCTCAACATGCTGACAA
8_R	TACGGTAGCAGAGACTTGGTCTGTTGTTAAGAAAAGCACTGCAGT
9_F	ACACTGACGACATGGTTCTACAACATGACATTCTCTGACTCCTGT
9_R	TACGGTAGCAGAGACTTGGTCTCCCCGCACTGCTTCACATA
10_F	ACACTGACGACATGGTTCTACAATGTGACCTCTGAGCCTCAC
10_R	TACGGTAGCAGAGACTTGGTCTTGGCCTGGGAGAAAGCAAAA
11_F	ACACTGACGACATGGTTCTACATTTTCTTCTTGGGCTTTCAT
11_R	TACGGTAGCAGAGACTTGGTCTAGTGACAGGATGGTTCAGCT
12_F	ACACTGACGACATGGTTCTACAAAGTGAGCGAACGGATAATGG
12_R	TACGGTAGCAGAGACTTGGTCTACTGCAAAGACACAAGGCC

13_F	ACACTGACGACATGGTTCTACACACAAAGACTGCAGCTCACAT
13_R	TACGGTAGCAGAGACTTGGTCTTGCACTTAATGACAGCTCCCA
14_F	ACACTGACGACATGGTTCTACAGATCCCTCCGACCTTGACTG
14_R	TACGGTAGCAGAGACTTGGTCTACATGAATAATAGCAGCCTACCC
15_F	ACACTGACGACATGGTTCTACATGTGTGTATGGGTGTCTCTGA
15_R	TACGGTAGCAGAGACTTGGTCTCACCTGCTCTAACCTACCC
16_F	ACACTGACGACATGGTTCTACAGAGGAGGAGGGTCTGAGAGA
16_R	TACGGTAGCAGAGACTTGGTCTCCTTATGTAGCTTAGCAAGTCGA
17_F	ACACTGACGACATGGTTCTACATTCTTGCCCCGAAAGCTG
17_R	TACGGTAGCAGAGACTTGGTCTGTGAAGTACTGTGTTTTTCCTCAT
18_F	ACACTGACGACATGGTTCTACATACCCATCAGTTACTCCTTCCC
18_R	TACGGTAGCAGAGACTTGGTCTGGGATTAAGCTGGGAGTTTC
19_F	ACACTGACGACATGGTTCTACATTCTTCCCCACTGAACTGC
19_R	TACGGTAGCAGAGACTTGGTCTAGGTCTTGGATTAATTGGCGT
20_F	ACACTGACGACATGGTTCTACACCCACCTTCCCAAGTAACTC
20_R	TACGGTAGCAGAGACTTGGTCTTTCCTGCTTACCTCCCTAGC
21_F	ACACTGACGACATGGTTCTACATACATGGAAGGGGCAGTTGA
21_R	TACGGTAGCAGAGACTTGGTCTATGAGTGAGGGGCAGAGAAA
22_F	ACACTGACGACATGGTTCTACAGCAGGGCCTTGAATCTGTAG
22_R	TACGGTAGCAGAGACTTGGTCTAGATAGTGCCAGAGTCACC
23_F	ACACTGACGACATGGTTCTACATTTTGCTCATGATCTGTGTTC
23_R	TACGGTAGCAGAGACTTGGTCTGTACAGCACAAACCCAGG
24_F	ACACTGACGACATGGTTCTACATGGCCTGCTGAAGATCATTG
24_R	TACGGTAGCAGAGACTTGGTCTTGACACAGCCTTTAAACTTGTC

**Table B.3 Primers used for amplification of KDM6A.**

*Number = amplicon number, F = forward primer, R = reverse primer. Sequence includes Fluidigm CS adapters.*

<b>Primer</b>	<b>Sequence (5' -&gt; 3')</b>
1_F	ACACTGACGACATGGTTCTACACGCTTTCGGTGATGAGGAAA
1_R	TACGGTAGCAGAGACTTGGTCTCCGTACCTGTCCAGTCCG
2_F	ACACTGACGACATGGTTCTACATCTTTCAGGGCAATTAAGCATT
2_R	TACGGTAGCAGAGACTTGGTCTACAACCTACCTTTAAACTAGACTCA
3_F	ACACTGACGACATGGTTCTACAGTACAATTGGACCATGGCCA
3_R	TACGGTAGCAGAGACTTGGTCTAGTGCAGAGGTATTACTACAATT
4_F	ACACTGACGACATGGTTCTACACAGGATGCCATTAATGCTACTT
4_R	TACGGTAGCAGAGACTTGGTCTTCTGGGAAATATGTGGCTTT

5_F	ACACTGACGACATGGTTCTACAATGCTGTGTCACATCCTCCA
5_R	TACGGTAGCAGAGACTTGGTCTACTTGTGTTGCTACCTCTACTCCT
6_F	ACACTGACGACATGGTTCTACATGACAGATGAGACCAACAGGA
6_R	TACGGTAGCAGAGACTTGGTCTCAGGCTGAGAGACGCTAGG
7_F	ACACTGACGACATGGTTCTACACTGCCTACAACTCAGTCTCTG
7_R	TACGGTAGCAGAGACTTGGTCTCAGAAAAGGGTCCATTGGCC
8_F	ACACTGACGACATGGTTCTACATAACCGCACAAACCTGACCA
8_R	TACGGTAGCAGAGACTTGGTCTTCTCTCAAAGTGTATAAAACCCAGT
9_F	ACACTGACGACATGGTTCTACACGACCTCTCTTCCACTGG
9_R	TACGGTAGCAGAGACTTGGTCTAATGCCTTGTTGCCACCTG
10_F	ACACTGACGACATGGTTCTACAGGCTGCTCTCAATCACCTCT
10_R	TACGGTAGCAGAGACTTGGTCTGCAGTGTGTTAGGTGTCTC
11_F	ACACTGACGACATGGTTCTACAGAGACACCTAACAGCACTGC
11_R	TACGGTAGCAGAGACTTGGTCTTCCCATCAACAAGGCAGAGA
12_F	ACACTGACGACATGGTTCTACAGCCATTTCAACAGCAACACC
12_R	TACGGTAGCAGAGACTTGGTCTGGGGCTCTGAGATTCTTCCA
13_F	ACACTGACGACATGGTTCTACAGGAAGAATCTCAGAGCCCCA
13_R	TACGGTAGCAGAGACTTGGTCTCACACTAACCTGCATGCCTT
14_F	ACACTGACGACATGGTTCTACAATGGACTTGTGCAAATGCCTAGTAA
14_R	TACGGTAGCAGAGACTTGGTCTTGGAGGTGGACATTTATCCAACAA
15_F	ACACTGACGACATGGTTCTACATGTTTTCTGAGATCTAACCACA
15_R	TACGGTAGCAGAGACTTGGTCTCAAGGCCACGTATTACTGTAACA
16_F	ACACTGACGACATGGTTCTACATGTAGAACACTAACTAGACTGCT
16_R	TACGGTAGCAGAGACTTGGTCTACACAGTATTAGAAACATGCCTTTT
17_F	ACACTGACGACATGGTTCTACAGTTCTGGGAGGAGGAGGAAA
17_R	TACGGTAGCAGAGACTTGGTCTAGCACAGGATAACTCTTTGCA
18_F	ACACTGACGACATGGTTCTACAAAACCTCCACAGGTATTTGTAGC
18_R	TACGGTAGCAGAGACTTGGTCTCCAACATGGCTTAGAAGATTTCC
19_F	ACACTGACGACATGGTTCTACAGTGGAAGTTGCAGCTACATGA
19_R	TACGGTAGCAGAGACTTGGTCTTGCTCCCTGGAACTTTCATG
20_F	ACACTGACGACATGGTTCTACAACCGTGTGCTAACCAATTGC
20_R	TACGGTAGCAGAGACTTGGTCTACAAACCATTACAGTCACCT
21_F	ACACTGACGACATGGTTCTACAGGAGCTTCTTAATGTAGTTGATCC
21_R	TACGGTAGCAGAGACTTGGTCTGCTGAATAAACCTATACTGGAAC
22_F	ACACTGACGACATGGTTCTACACTAATGGGTTCTTGGTGGCC
22_R	TACGGTAGCAGAGACTTGGTCTTGAACCAATGAACAGTGCC
23_F	ACACTGACGACATGGTTCTACAGCTGGTCACAAATAATTTCTCCC

23_R	TACGGTAGCAGAGACTTGGTCTTGAGCTGGTTCTTCTTTTGTC
24_F	ACACTGACGACATGGTTCTACAACCTTGGAACCTTTGTGGTGCT
24_R	TACGGTAGCAGAGACTTGGTCTCACTGCTGCTTCATAACCCA

**Table B.4 Primers used for amplification of area around *KRAS* codons 12 and 13 as well as the mimic gene *PITPNM2*.**

*Number = amplicon number, F = forward primer, R = reverse primer. Sequence includes Fluidigm CS adapters.*

Primer	Sequence (5' -> 3')
KRAS_1_F	ACACTGACGACATGGTTCTACATAAGGCCTGCTGAAAATGACT
KRAS_1_R	TACGGTAGCAGAGACTTGGTCTATGGTCCTGCACCAGTAATATG
MIMIC_1_F	ACACTGACGACATGGTTCTACATAGCACCCAGCCAGCTTG
MIMIC_1_R	TACGGTAGCAGAGACTTGGTCTATGACCACCCATGAAATATGAGCT
KRAS_2_F	ACACTGACGACATGGTTCTACAGGTGGAGTATTTGATAGTGTATTAACC
KRAS_2_R	TACGGTAGCAGAGACTTGGTCTTAGCTGTATCGTCAAGGCAC
MIMIC_2_F	ACACTGACGACATGGTTCTACACCTGCTGTTCCCTACAAAGCTG
MIMIC_2_R	TACGGTAGCAGAGACTTGGTCTAGGCTTCTCCCGTCTAAGGA







# **Appendix C Mathematical modelling of clonal marks data**

Analysis of clone data was performed by Dr Edward Morrissey, who wrote the following sections about the mathematical methods used. They were also published as part of our recent publication (Nicholson *et al.*, 2018).

## **C.1 A general note on simulations and the mathematical model**

Generally, we make use of a mathematical model (described below) that models the acquisition of a mutation, the competition of the mutant stem cell with the other stem cells and, once fixed, the fission of the mutant crypt.

Additionally, the model was challenged with a more complex scenario to study whether more complexity is warranted. This was done using simulations that encode the same assumptions as the mathematical model but with additional behaviours. Specifically, the effect of double hits on the clonal dynamics was modelled. Double hit simulations require just one tracking the individual cells and how many mutations each one has. All the simulations were coded in python using the numba library for speed.

## **C.2 Statistical inference**

All data fitting was done using the statistical models described in the fitting sections below and sampled from using Rstan (Carpenter *et al.*, 2017). Rstan was run using 5 chains of 10,000 iterations and a thinning of 5. The default parameters were used for the sampler, though where

necessary, the models were reparametrised and run parameters adapted. Convergence was checked using the scale reduction factor provided by Rstan.

Within chapters 4 and 5 estimates are presented as credible intervals (CI) or alternatively as a margin of error (ME) expressed as a median and 1.96 times the standard deviation of the posterior. For cases where new parameters, are calculated that are functions of the inferred parameters we apply the function to all the posterior mcmc samples and present the median and 1.96 times the standard deviation of the transformed samples.

For some of the cases below, Gaussians were used to model the population variability of a parameter defined in the  $[0, 1]$  range, for these cases the range of the parameter was specified in Stan.

### C.3 Statistical model

Patients were selected based on tissue block size that so as to be able to estimate a mutation rate per block. In some cases we had several such blocks for the same patient, which we used within the statistical model to estimate the within patient variability and experimental error. A hierarchical model was used as follows, assuming we measure  $k_{i,b}$  TA clones for patient  $i$  in block  $b$ , the number of goblet cells measured is  $G_{i,b}$  and the mutation rate for patient  $i$  is  $\alpha_i$  the counts are distributed as

$$q_{i,b} \sim \text{Normal}(\alpha_i, \sigma_{\text{error}})$$

$$k_{i,b} \sim \text{Binomial}(G_{i,b}, q_{i,b})$$

We calculate the distribution of the mutation rate in the patient population as

$$\alpha_i \sim \text{Normal}(\mu_\alpha, \sigma_\alpha)$$

The priors used were:

$$\mu_\alpha \sim \text{Beta}(1/2, 1/2)$$

$$\sigma_{\alpha} \sim \text{Beta}(1/2, 1/2)$$

$$\sigma_{\text{error}} \sim \text{Beta}(1/2, 1/2)$$

## C.4 Continuous labelling of a neutral mutation

Here we describe the continuous labelling model that can be found in Kozar et al., (2013). It has been shown that crypts are maintained by an equipotent population of stem cells at the crypt base that constantly replace each other in a stochastic fashion (Lopez-Garcia *et al.*, 2010; Snippert, van der Flier, *et al.*, 2010). The equations that govern the change in clone size with time assume we start tracking the progeny of a clone of size 1 stem cell at  $t=0$ . The probability of a crypt having clone of size  $n$  (for  $0 < n < N$ ) at time  $t$  is:

$$P_n(t) = \frac{2}{N} \sum_{m=1}^{N-1} \sin\left(\frac{\pi m}{N}\right) \sin\left(\frac{\pi m n}{N}\right) e^{-4\lambda \sin^2\left(\frac{\pi m}{2N}\right)t}$$

Here  $n$  is the number of stem cells that make up the clone,  $N$  is the total number of stem cells in the crypt base and  $\lambda$  is the rate of stem cell replacement. For the probability of the clone being of maximum size, i.e. a monoclonal crypt:

$$P_N(t) = \frac{2}{N} \sum_{m=1}^{N-1} (-1)^{m+1} \cos^2\left(\frac{\pi m}{2N}\right) \left(1 - e^{-4\lambda \sin^2\left(\frac{\pi m}{2N}\right)t}\right)$$

For our case if we are tracking mutationally tagged clones. If we take the mutation rate to be  $\alpha$  the rate at which a crypt will get a mutationally activated clone will be

$$\kappa = \alpha \lambda N$$

If we write down the stochastic master equation for this:

$$\frac{dQ_0}{dt} = -\kappa Q_0$$

$$\frac{dQ_1}{dt} = \kappa Q_0$$

We can solve and get

$$Q_1(t) = (1 - e^{-\kappa t})$$

As the mutation rate is very low we can use a Taylor expansion to get

$$Q_1(t) \approx \kappa t$$

New clones of size one stem cell are appearing continuously over time, assuming the mutation has no effect on the stem cell dynamics, the clone size will evolve according to the equations above. To model the probability of clone size over time we can use the integral

$$C_n(t) = \int_0^t \frac{dQ_1}{d\tau}(\tau) P_n(t - \tau) d\tau$$

Which assumes that the clones that disappear due to stem cell competition have a negligible effect on  $Q_0$ .

Solving for the non-monoclonal clones and pooling them to get the partial clone prediction we get:

$$C_{\text{partial}} = \alpha \frac{N(N-1)}{2} - \frac{\alpha}{2} \sum_{n,m=1}^{N-1} \frac{\sin\left(\frac{\pi m}{N}\right) \sin\left(\frac{\pi mn}{N}\right)}{\sin^2\left(\frac{\pi m}{2N}\right)} e^{-4\lambda \sin^2\left(\frac{\pi m}{2N}\right)t}$$

For the monoclonal clones we get

$$C_{\text{monoclonal}} = \alpha \lambda t - \frac{\alpha}{2} \sum_{m=1}^{N-1} \frac{(-1)^{m+1}}{\tan^2\left(\frac{\pi m}{2N}\right)} \left(1 - e^{-4\lambda \sin^2\left(\frac{\pi m}{2N}\right)t}\right)$$

The effect of the exponential term is quickly lost, leading to a constant term for the partials and a linear function for the monoclonals.

## C.5 Continuous labelling of a non-neutral mutation

Vermeulen et al., (2013) showed that certain mutations can affect the clonal dynamics. Furthermore, they showed that these altered dynamics could be parameterized by introducing a replacement probability,  $P_R$ . The equations for the non-monoclonal and monoclonal clones are as follows:

$$R_n(t) = \frac{2}{N} \left( \frac{\beta}{\gamma} \right)^{\frac{1}{2}(n-1)} \sum_{m=1}^{N-1} k_{m,n} e^{-h_m t}$$

$$R_N(t) = \frac{2\beta}{N} \left( \frac{\beta}{\gamma} \right)^{\frac{1}{2}(N-2)} \sum_{m=1}^{N-1} \frac{k_{m,N-1}}{h_m} (1 - e^{-h_m t})$$

Where the following shorthand has been used:

$$\gamma = 2\lambda(1 - P_R)$$

$$\beta = 2\lambda P_R$$

$$k_{m,n} = \sin\left(\frac{\pi m}{N}\right) \sin\left(\frac{\pi m n}{N}\right)$$

$$h_m = 4\sqrt{\gamma\beta} \sin^2\left(\frac{\pi m}{2N}\right) + \gamma + \beta - 2\sqrt{\gamma\beta}$$

While the drift dynamics are different to the neutral case, the dynamics of the appearance of the initial mutations are the same; therefore we can derive the continuous labelling equations in the same way:

$$\widehat{C}_n(t) = \int_0^t \frac{dQ_1}{d\tau}(\tau) R_n(t - \tau) d\tau$$

Which leads to

$$\widehat{C}_{partial}(t) = \frac{2\kappa}{N} \sum_{m,n=1}^{N-1} \left(\frac{\beta}{\gamma}\right)^{\frac{1}{2}(n-1)} \frac{k_{m,n}}{h_m} (1 - e^{-h_m t})$$

$$\widehat{C}_{monoclonal}(t) = \frac{2\beta\kappa}{N} \left(\frac{\beta}{\gamma}\right)^{\frac{1}{2}(N-2)} \sum_{m=1}^{N-1} \frac{k_{m,N-1}}{h_m} \left(t - \frac{1}{h_m} (1 - e^{-h_m t})\right)$$

For the sake of brevity we do not expand the equations, however it is worth noting that much like the neutral mutations, after a short initial period the monoclonals follow a linear equation and the partials converge to a constant value. For both the neutral and non-neutral cases the equations are proportional to the mutation rate, meaning that the ratio of the slope of the monoclonal accumulation over the partials gives a value that is independent of the mutation rate. This can be used as a way of comparing the clonal dynamics for different mutations.

## C.6 Fitting the monoclonal clones and partial clones

As the probability of a crypt containing a monoclonal clone at time  $t$  is a linear function we fit the following model to the monoclonal data:

$$p_i = a_i(t_i - t_0)$$

$$k_i^{mono} \sim \text{Binomial}(C_i, p_i)$$

Where  $k_i^{mono}$  is the number of monoclonal crypts found for patient  $i$ ,  $C_i$  is the number of crypts in the tissue sample,  $t_i$  is the patient age,  $a_i$  is the slope of the monoclonal accumulation for patient  $i$  and  $t_0$  is the x-axis intercept. As we expect the mutation rate to have some variation

between individuals, as well as the drift parameters, we allow each patient to have its own slope, using a hierarchical model

$$a_i \sim \text{Normal}(\mu_a, \sigma_a)$$

The priors on the parameters are as follows

$$\mu_a \sim \text{Gamma}(10^{-2}, 10^{-2})$$

$$\sigma_a \sim \text{Gamma}(10^{-2}, 10^{-2})$$

$$t_0 \sim \text{Normal}(0, 10)$$

Note how we are allowing  $t_0$  to be negative. While the stem cell dynamics equations suggest that the y-intercept should be negative, and as such the x-intercept should be positive it is possible that clones might arise during development that would increment the y-intercept allowing for the x-intercept to become negative. We choose a value that encompasses  $\sim 20$  years to either side of the origin to allow a wide range of values, however restricting implausible values.

We follow a similar analysis for the partial clones.

$$k_i^{\text{partial}} \sim \text{Binomial}(C_i, b_i)$$

$$b_i \sim \text{Normal}(\mu_b, \sigma_b)$$

With priors

$$\mu_b \sim \text{Gamma}(10^{-2}, 10^{-2})$$

$$\sigma_b \sim \text{Gamma}(10^{-2}, 10^{-2})$$

## C.7 Sequential mutations

The mutation of *STAG2*, a gene that when mutated is associated with chromosomal instability, was found to have a biased behaviour. The fact that *STAG2* is associated to chromosomal instability raises the question of whether the biased behaviour is the consequence of further

unmeasured mutations enabled by the chromosomal instability or directly caused by *STAG2*. To find which might be the most likely scenario we run simulations where we assume that a first neutral mutation raises the mutation rate of a second mutation that biases drift.

The simulation uses the Gillespie Algorithm to simulate a single crypt with  $N$  stem cells, each of which starts with no mutations and can acquire a first mutation which doesn't change the drift dynamics, however the mutant cells now have an enhanced probability of a second mutation which does lead to a bias. The simulation produces two outputs, the monoclonal and partial crypts for the first mutation, regardless of whether or not they have the second mutation (this would be what we measure with *STAG2*) and also outputs the full and partial crypts with both mutations (as you can't have mutation 2 without 1).

If we can only measure mutation 1, as happens with *STAG2*, in order to see altered dynamics caused by mutation 2 the mutation has to occur while mutation 1 has not yet become monoclonal, otherwise we would measure no difference.

## C.8 Crypt fission and mutation burden

We model crypt fission as a Yule-Furry pure birth process. The general solution to this process is:

$$\widehat{F}_n(t) = \binom{n-1}{n-n_0} e^{-\rho n_0 t} (1 - e^{-\rho n_0 t})^{n-n_0}$$

Where  $n_0$  is the patch size at time  $t = 0$  and  $\rho$  is the rate of crypt fission. In order to calculate the patch size distribution over time given that the monoclonal crypts appear following a known function we can use a similar calculation as for the continuous labelling equations. We fix  $n_0 = 1$  and integrate:

$$F_n(t) = \int_0^t \frac{dC_{monoclonal}}{d\tau}(\tau) \widehat{F}_n(t - \tau) d\tau$$

Ignoring the exponential term from  $C_{monoclonal}$  which has a negligible effect, we find

$$F_n(t) = \Delta C_{monoclonal} \frac{(1 - e^{-\rho t})^n}{\rho n}$$

Here  $\Delta C_{monoclonal}$  is the slope of the monoclonal accumulation. This equation also holds for mutations that affect clonal drift. We use this equation to estimate the mutant burden per million crypts used in the main text:

$$B(t) = 10^6 \sum_{n=1}^{\infty} n F_n(t)$$

## C.9 Statistical model for patch sizes

The patch size equation depends on the slope of the monoclonals, which we can infer from the monoclonal data. However, in order to minimise the uncertainty in the crypt fission estimation, we calculate the equation for the relative distribution of patch sizes that does not depend on the slope of the monoclonals:

$$f_n(t) = \frac{F_n(t)}{C_{monoclonal}(t)} = \frac{(1 - e^{-\rho t})^n}{\rho n t}$$

This is the same equation used by Baker et al., (2014). We also apply a correction for the confounding effect of two unrelated clones randomly being found next to each other and counted as a patch. If a tissue sample has  $k$  clones,  $C$  crypts and each crypt has  $\delta$  neighbouring crypts the proportion of clones that form random doublets will be:

$$D = \sum_{i=1}^{k-1} \frac{\delta}{k} \frac{i}{C-i} \approx \delta \frac{k-1}{2C}$$

We do not calculate the probability of patches larger than two appearing due to chance as the probability of these events will be negligible. When fitting the model to the data we add  $D$  to  $f_2$  and subtract  $D$  from  $f_1$ .

As a first step for the fitting we filter samples with no clones as we are fitting the relative patch size. Again, we use a hierarchical model to account for patient-to-patient variability. If  $g_i$  is a

vector of measured patch sizes,  $t_i$  is the age of the patient,  $\rho_i$  is the fission rate for that patient we have

$$g_i \sim \text{Multinomial}(f(\rho_i, t_i))$$

$$\rho_i \sim \text{Normal}(\mu_\rho, \sigma_\rho)$$

Where  $f$  is the vector of probabilities of each patch size calculated from the fission equation and corrected as specified above. The priors used for the population parameters are

$$\mu_\rho \sim \text{Gamma}(10^{-2}, 10^{-2})$$

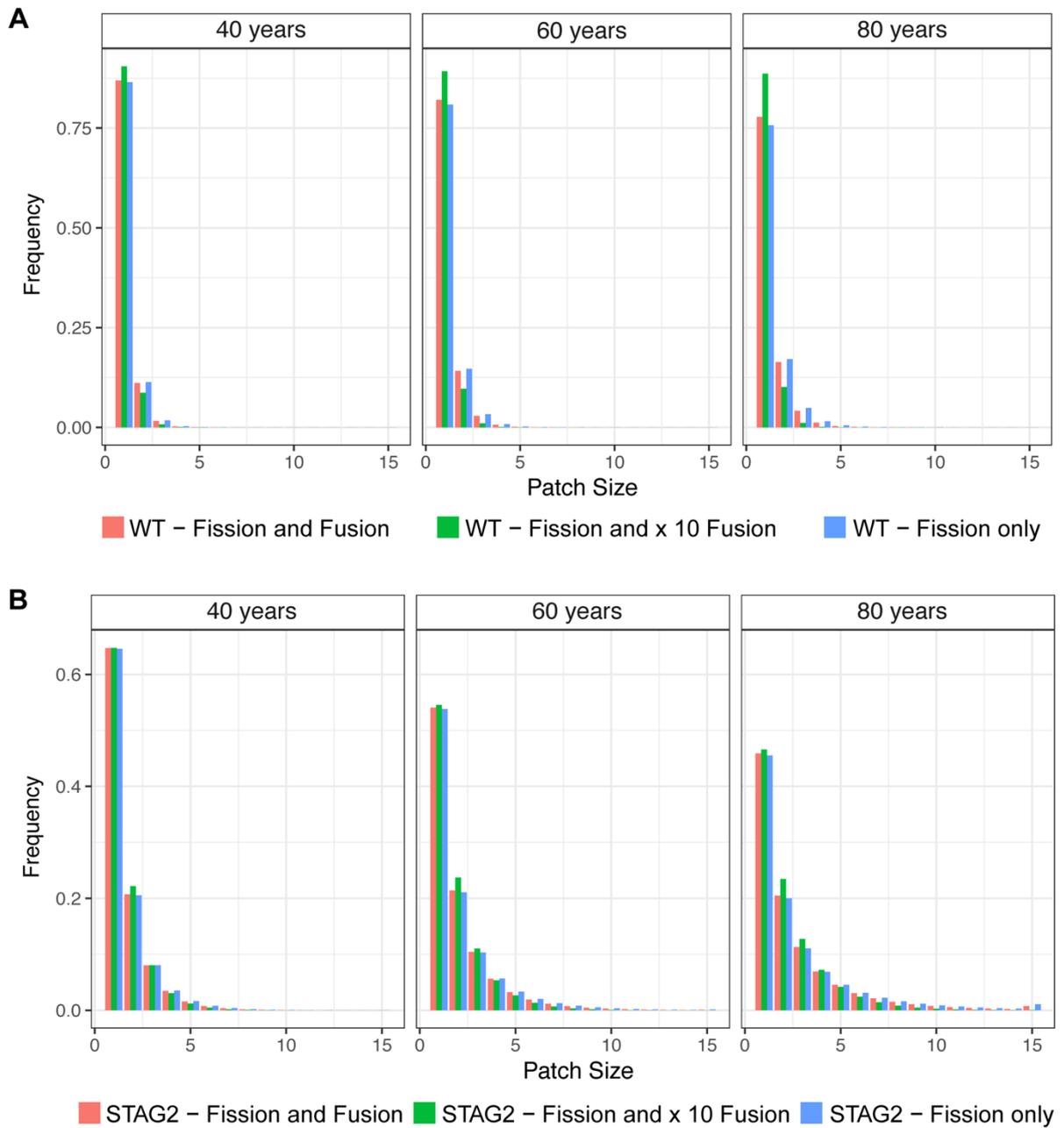
$$\sigma_\rho \sim \text{Gamma}(10^{-2}, 10^{-2})$$

## C.10 Effect of crypt fusion on patch size

A recent study has shown that crypts not only undergo fission, where a crypt divides into two crypts, but they can also fuse with a neighbouring crypt thus combining the stem cell pools. The study found that fission and fusion are balanced, both occurring at the same rate.

At the clonal level fusion can cause a mutant crypt to join with a non-mutant producing a partially mutant crypt or two mutant crypts can join forming a single mutant crypt. This introduces a spatial aspect to the model, which complicates an analytical approach. To assess the effect of fusion we implement a stochastic simulation algorithm which uses the gillespie algorithm. The simulation models a field of crypts and implements the mutation process, stem cell drift, fission and fusion, including the spatial aspects as well as the two types of fusion events described above.

The simulations showed that relative patch size is dominated by fission, with fusion having a very modest effect (Figure C.1).



**Figure C.1 Effects of fission and fusion on patch size.**

Figure by Dr Edward Morrissey, modified from Supplementary Figure 6 in (Nicholson et al., 2018).

(A) Output of mathematical model for effects in WT crypts.

(B) Output of mathematical model for effects in STAG2-deficient crypts.







# Appendix D Mathematical modelling of KRAS data and comparison to clonal marks

Mathematical modelling to infer the behaviour of *KRAS*-mutant clones from targeted sequencing data was performed by Dr Edward Morrissey who wrote the following sections. These were also partly published in our most recent publication (Nicholson *et al.*, 2018).

## D.1 Statistical model for KRAS patch size estimation

To analyse the *KRAS* mutant allele frequency data we first convert it to mutation burden. To do so we note that if in a section of tissue we have  $m$  mutant crypts,  $C$  total crypts and  $n$  cells per crypt the ratio of mutant copies of a gene to total copies of the gene will be

$$f_{allele} = \frac{mn}{2Cn}$$

which means that the allele frequency is half of the mutation burden.

In order to model this data we can use the equation for patch sizes derived earlier, namely

$$F_n(t) = \Delta C_{monoclonal} \frac{(1 - e^{-\rho t})^n}{\rho n}$$

Which gives us the probability of finding a patch of size  $n$  at age  $t$ . The model has two parameters the fission rate  $\rho$  and the monoclonal accumulation rate  $\Delta C_{fix}$ . These are the two parameters we wish to infer from the data.

The statistical fitting must account for the fact that there is a detection limit below which there may be clones but we cannot detect them. We set up the statistical model so that if the mathematical model predicts that there should be a patch but we measure none, as long as it is below the specified detection threshold, it does not penalise the fit.

We first take the measured allele frequency and convert them to mutation burden, we then use the number of crypts from that sample to convert the burden into patch size. We know how many crypts we have in the sample from the image analysis.

We cannot directly use the patch size equation as we need to accommodate the fact that we have a range of possible patch sizes of which each patient will only have one, also the probability of not detecting a patch will need to be calculated depending on the values of the parameters.

We model each patient sample as a multinomial with three categories, probability that a crypt has no detectable clone  $q_0$ , probability  $q_1$  that we see a patch of size  $n$  (where  $n$  is the observed patch size) and  $q_2$  the probability of all the remaining patch sizes, used to normalise the multinomial  $q_2 = 1 - (q_0 + q_1)$ . We calculate  $q_0$ , which incorporates the detection threshold as

$$q_0 = \sum_{n=0}^{n_{limit}} F_n(t)$$

Here  $n_{limit}$  is the largest patch size that would not be detected. We calculate the probability of no clone with

$$F_0 = 1 - \Delta C_{monoclonal} t$$

The likelihood will be

$$n_i \sim \text{Multinomial}(q)$$

Where  $q$  is the vector described above and  $n_i$  is a vector of 3 counts for patient  $i$ : total crypts, zero or one if there is a patch and 0 for the third category.

The priors used for the two parameters are:

$$\rho \sim \text{half-normal}(0, 0.5)$$

$$\Delta C_{fix} \sim \text{half-normal}(0, 10^{-4})$$

We also need to scale  $\Delta C_{fix}$  by the number of mutations we look at, which are 12.

## D.2 Relative expansion coefficient

In order to derive a metric for each mutation that allows comparison of the ability of the mutation to spread through the tissue we calculate the burden of a mutation averaged over the lifetime of the individual. We then calculate the ratio of average burden between a given mutation and the wild-type parameters. By fixing the mutation rate to the same value for both average burden estimates, the mutation rate disappears from the ratio.

$$I^{mutant} = \frac{\frac{1}{100} \int_0^{100} B^{mutant}(t) dt}{\frac{1}{100} \int_0^{100} B^{WT}(t) dt}$$

We refer to this value as a relative expansion coefficient ( $C_{exp}$ ). The values used in the main text were calculated numerically using the burden equation described in section C.8.







# Appendix E Mathematical modelling of crypt fusion

This was performed by Dr Doran Khamis who wrote the following paragraphs about the mathematical methods used.

## E.1 Fusion/Fission events as Poisson processes

For individual crypts, fusion/fission events occur at a rate  $\rho$  and have a duration  $\Delta\tau$ . If we take a snapshot of a piece of tissue at a time  $t_0$  we see all fusion/fission events that occurred in the window  $[t_0 - \Delta\tau, t_0]$ . Calculating the average number of events per crypt,  $X$ , in a time  $\Delta\tau$  over many snapshots is the same as calculating the probability of an event for a single crypt in the window  $\Delta\tau$  (as we can only have a single event in any time window equal to the event duration). The number of events for a single crypt follows a Poisson distribution,

$$X_1 \sim Poi(\rho \Delta\tau).$$

We want to observe events on the edge of mutant patches such that we can differentiate fission events from fusion events. For a patch with edge length  $N$  (that is,  $N$  crypts define the patch perimeter, each with at least one wild-type crypt as a neighbour), the number of crypts undergoing fusion or fission is distributed as

$$X_N \sim Poi(N \rho \Delta\tau).$$

For a given patch with edge length  $N$ , then, the probability of zero events in a window  $\Delta\tau$  is

$$p\{X_N = 0\} = e^{-N\rho\Delta\tau} \sim 1 - N\rho\Delta\tau + \mathcal{O}(N^2\rho^2\Delta\tau^2),$$

where  $N\rho$  is considered small compared with  $\Delta\tau$  such that we may define a parameter  $\varepsilon = N\rho\Delta\tau$  where  $\varepsilon \ll 1$ . Correspondingly, the probability of seeing at least one event in a window  $\Delta\tau$  is

$$p\{X_N \geq 1\} = 1 - p_0 \sim N\rho\Delta\tau + \mathcal{O}(\varepsilon^2). \quad (1)$$

This equation can be applied to either fusion events or fission events separately by changing the event rate  $\rho$  to  $\rho_{fu}$  or  $\rho_{fi}$ , the event rates of fusion and fission, respectively, assuming that the event duration  $\Delta\tau$  is approximately equal for fission and fusion.

## E.2 Calculating the fusion rate

Let the number of partially mutant (partial) and fully mutant (monoclonal) fufi events observed on the edge of mutant patches be  $n_p$  and  $n_m$  respectively. These numbers are combined over many different mutant patches and tissue samples, with a total patch edge length of  $N$ . To calculate the fusion rate  $\rho_{fu}$  given the fission rate  $\rho_{fi}$ , we use the following observations and assumptions:

1. All fission events are monoclonal (mut fufis).
2. Not all monoclonal events are fissions.
3. All partial events (WT-mut fufis) are fusions.
4. Not all fusions are partials (WT-mut fufis).

The first and third assumptions stem from the belief that the time scale over which monoclonal conversion occurs in a crypt is short compared to the time between fusion/fission events. The second and fourth observations are alternative statements of the fact that fusion at the patch edge can happen inwards: a mutant crypt on the patch edge can fuse with a mutant crypt within the patch, hence creating a monoclonal event. We define the parameter  $\chi$  to be the proportion of fusion events that are ‘‘inwards’’ and hence monoclonal. Then the ratio  $n_p/n_m$  may be written as

$$n_p / n_m = n_{fu}(1 - \chi) / (n_{fi} + n_{fu}\chi), \quad (2)$$

where  $n_{fu}$  and  $n_{fi}$  are the number of fusion and fission events, respectively. We do not know  $n_{fu}$  and  $n_{fi}$  *a priori*, however.

We may recast the probability  $p\{X_N \geq 1\}$  from eq. (1) as the number of observed events over sample size,  $n_{events}/N$ , for fusion and fission, we get

$$n_{fu} / N \sim N \rho_{fu} \Delta\tau \quad \text{and} \quad n_{fi} / N \sim N \rho_{fi} \Delta\tau,$$

where we have assumed the duration of a fusion event is approximately equal to that of a fission event. Thus, we find the approximate equivalence

$$n_{\bar{n}} / n_{fu} \sim \rho_{fi} / \rho_{fu}$$

between the ratios of numbers and rates of events. Using this, we may rewrite eq. (2) as

$$n_p / n_m = (1 - \chi) / (\rho_{fi} / \rho_{fu} + \chi),$$

and subsequently find an expression for the fusion rate:

$$\rho_{fu} = \rho_{fi} / [(n_m / n_p) (1 - \chi) + \chi].$$

The simplest model for the parameter  $\chi$  is to assume unbiased (isotropic) fusion, such that the proportion of fusions that are monoclonal is simply

$$\chi = (N_t - N_w) / N_t,$$

where  $N_t$  and  $N_w$  are the total number of neighbours and number of wild-type neighbours of a given fusion event, respectively.

### E.3 Calculating the event duration

The duration of a fusion event is assumed to be the same as that of a fission event. To find this duration, we treat fusion and fission as two Poisson processes occurring simultaneously. Thus,

$$p\{X_1 = 0\} = \exp\{-(\rho_{fi} + \rho_{fu})\Delta\tau\},$$

and  $p\{X_N \geq 1\} = 1 - p\{X_1 = 0\}$ . We can recast the probability  $p\{X_N \geq 1\}$  as a proportion of fufi events over total crypts,  $n_f / N_c$  and then rearrange for the duration:

$$\Delta\tau = -\log(1 - n_f / N_c) / (\rho_{fi} + \rho_{fu}).$$

We observe the total number of crypts  $N_c$  and the number of events  $n_f$ , but we do not know how many of the  $n_f$  are fission or fusion. We choose to approach this problem by splitting the data into wild type and mutant to find two event durations. This leads to two expressions for the bulk population mean durations:

$$\Delta\tau^{WT} = -\log(1 - n_f^{WT} / N_c^{WT}) / (\rho_{fi}^{WT} + \rho_{fu}^{WT}),$$

for fusion/fission events in wild type crypts, where a superscript  $WT$  denotes wild type and

$$\Delta\tau^{\text{Mut}} = -\log(1 - n_f^{\text{Mut}} / N_c^{\text{Mut}}) / (\rho_{fi}^{\text{Mut}} + \rho_{fu}^{\text{Mut}}),$$

for fusion/fission events in mutant crypts.

We can capture the variability in the system further by using a Bayesian hierarchical model for the fufi occurrence. We model the fufi counts on a given slide,  $\tilde{n}_f$ , as a binomial distribution

$$\tilde{n}_f \sim \text{Binom}(\tilde{N}, p_f),$$

where  $\tilde{N}$  is the crypt count of the slide and event probability  $p_f$  is drawn from a Gaussian distribution:

$$p_f \sim \text{Norm}(\mu_f, \sigma_f).$$

Thus patient-to-patient variability in fufi occurrence, assumed to derive from differences in baseline fission and fusion rate, is captured as variation under a single normal distribution with mean  $\mu_f$  and variance  $\sigma_f$ . We use Markov Chain Monte Carlo sampling to infer the hyper-parameters defining the distribution on the population fufi fraction  $p_f$ , then use the sampling draws to calculate the mean and 95% confidence intervals for the fufi duration.

