# Functional Analysis of *POLE* Exonuclease Domain Mutations in Cancer




Emily Rayner

St Catherine's College

University of Oxford

A thesis submitted for the degree of

*Doctor of Philosophy*

Hilary Term 2019

**Functional Analysis of *POLE* Exonuclease Domain Mutations in Cancer**

Emily Rayner

St Catherine's College, University of Oxford, D.Phil, Hilary 2019

## Abstract

*POLE* encodes the polymerase and exonuclease domains of the leading strand replicase, DNA polymerase $\epsilon$. Somatic and germline *POLE* exonuclease domain mutations (EDMs) occur in colorectal and endometrial cancer and are associated with tumour ultramutation, a distinct mutation signature and an excellent prognosis. The mechanism by which *POLE* EDMs cause mutagenesis and the molecular consequences of this, however, are unclear.

Previous functional studies in *POLE* EDM knock-in cell and mouse models have been performed, however they all have limitations in their clinical relevance. Furthermore, speculation regarding the timing at which somatic *POLE* EDMs occur in carcinogenesis is based purely on genomic data from established cancers rather than pre-cancers. The aim of this thesis was to study the function of *POLE* EDMs in clinically relevant isogenic human cell-lines and mouse models and determine their timing in colorectal cancer development.

Isogenic colorectal cancer cell lines with correction of the somatic $POLE^{\mathrm{P286R}}$ EDM were generated using CRISPR-Cas9. Correction of the *POLE* mutation resulted in a striking reduction in the *HPRT1* mutation rate (11-75-fold), confirming their causal role in tumour ultramutation. Furthermore, the mutation spectra in *POLE* EDM cell lines showed a significant bias towards *POLE* COSMIC Signature 10 mutations: TCT>TAT, TCG>TTG and TTT>TGT ($P<0.0001$).

Analysis of inducible uterine and intestinal-specific and constitutive germline $Pole^{\mathrm{L424V}}$ EDM knock-in mice revealed no gross morphological or histopathological phenotype, however constitutive mice demonstrated elevated Cleaved Caspase-3 expression in the small intestine ($P<0.05$). Germline $Pole^{\mathrm{L424V}}$ expression increased the colonic polyp burden in $Apc^{\mathrm{Min}}$ mice ($P<0.05$), although there was no significant difference in the small intestinal polyp burden.

Finally, a somatic $POLE^{\mathrm{V411L}}$ EDM was identified in a colorectal pre-cancer with an elevated T-cell infiltrate, increased somatic mutation burden and a bias of *POLE* EDM-associated amino acid substitutions.

Overall, this work confirms that somatic *POLE* exonuclease domain mutations cause tumour ultramutation, a distinct mutation signature and are an early event in cancer. Lastly, *Pole* EDM mice only developed a gross morphological phenotype when crossed with $Apc^{\mathrm{Min}}$ mice, raising interesting questions regarding the mechanism of tumour initiation in cancers with *POLE* exonuclease domain mutations.

# Acknowledgements

I would first of all like to thank Professor Ian Tomlinson for giving me the opportunity to undertake my D.Phil. in his laboratory, providing me with the scope to explore my own line of research in a supportive and intellectually rich environment.

Secondly I would like to thank the every member of the Tomlinson, Leedham, Church and Kearsey Laboratories, both past and present, for providing me with guidance and a friendly environment to work. I can't name everybody, but in particular I would like to thank Hayley, Luke, Annie, Emma, Chiara, Len, Shazia, Matt, Rachael, Claire, Nacho, Nagore, Roland and in the final stages of writing, Sibel your support and company has been immeasurable. I would also like to thank Professor Stephen Kearsey who has always been supportive of my work, providing me with feedback and input throughout my D.Phil.

My mum and dad have been supportive throughout and I cant thank them enough. I am thankful to all of my friends from Falcon Kayaking club, for their welcoming attitude to all and providing a rich social atmosphere.

Finally, this thesis would all not exist at all without the immense support and guidance from Dr David Church. I cant thank you enough for all that you have done, in terms of intellectual, academic and most importantly moral support when I have needed it the most. Thank you.

# Preface

This thesis is a collection of the research work undertaken by the author under the supervision of Professor Ian Tomlinson and Dr David Church, University of Oxford between October 2014 to March 2019.

I declare that all results described in this thesis are original, though the work of others are clearly cited with appropriate references throughout. No part of this research has been submitted for a degree at the Univeristy of Oxford or any other University.

Some of the work described in this thesis has been published in the following papers:

1. Van Gool, I.C., **Rayner, E.**, Osse, E.M., Nout, R.A., Creutzberg, C.L., Tomlinson, I., Church, D.N., Smit, V.T.H.B.M., de Wind, N., Bosse, T., et al. (2018). Adjuvant treatment for *POLE* proofreading domain-mutant cancers: sensitivity to radiotherapy, chemotherapy, and nucleoside analogs. *Clinical Cancer Research* 24, 3197-3203.

2. Temko, D., Van Gool, I.C., **Rayner, E.**, Glaire, M., Makino, S., Brown, M., Chegwidden, L., Palles, C., Depreeuw, J., Beggs, A., et al. (2018). Somatic *POLE* exonuclease domain mutations are early events in sporadic endometrial and colorectal carcinogenesis, determining driver mutational landscape, clonal neoantigen burden and immune response. *Journal of Pathology* 245, 283-296.

3. **Rayner, E.**[*], Van Gool, I.C.[*], Palles, C., Kearsey, S.E., Bosse, T., Tomlinson,

I., and Church, D.N. (2016). A panoply of errors: polymerase proofreading domain mutations in cancer. *Nature Reviews Cancer* 16, 71-81.

4. Domingo, E., Freeman-Mills, L., **Rayner, E.**, Glaire, M., Briggs, S., Vermeulen, L., Fessler, E., Medema, J.P., Boot, A., Morreau, H., et al. (2016). Somatic *POLE* proofreading domain mutation, immune response, and prognosis in colorectal cancer: a retrospective, pooled biomarker study. *The Lancet Gastroenterology & Hepatology* 1, 207-216.

I have additionally contributed to the following publication:

- Findlay, J.M.*, Castro-Giner, F.*, Makino, S.*, **Rayner, E*.**, Kartsonaki, C., Cross, W., Kovac, M., Ulahannan, D., Palles, C., Gillies, R.S., et al. (2016). Differential clonal evolution in oesophageal cancers in response to neo-adjuvant chemotherapy. *Nature Communications* 7, 11111.

# Contents

**Abstract** i

**Acknowledgements** ii

**List of Figures** xiii

**List of Tables** xvii

**List of Abbreviations** xx

**1 Introduction** 1

1.1 Introduction to cancer . . . . . 2

1.2 Mutagenesis and cancer . . . . . 4

1.2.1 DNA repair defects in cancer . . . . . 5

1.2.2 Base-excision repair . . . . . 5

1.2.3 DNA Mismatch repair . . . . . 7

1.3 DNA replication . . . . . 9

1.3.1 The Eukaryotic replication fork . . . . . 9

1.3.2 The DNA polymerases . . . . . 10

1.3.3 Accessory subunits . . . . . 11

1.3.4 Strand-specificity of leading and lagging strand polymerases . 12

1.4 Polymerase proofreading . . . . . 14

1.4.1 Polymerase exonuclease domain mutations in yeast . . . . . 17

1.4.2 Polymerase exonuclease domain mutations in mice . . . . . 18

1.5 Polymerase exonuclease domain mutations in cancer . . . . . 19

1.5.1 Functionality of polymerase exonuclease domain mutations . . 25

1.5.2 Clinicopathological characteristics of polymerase exonuclease domain mutant tumours . . . 27
1.5.3 Molecular characteristics of polymerase exonuclease domain mutant tumours . . . 29
1.5.4 Pathogenicity of polymerase exonuclease domain mutations . . 32
1.6 Outline of Thesis . . . 33

**2 Materials and Methods 35**
2.1 Cell Culture . . . 35
2.1.1 Cell lines . . . 35
2.1.2 Cell maintenance . . . 35
2.1.2.1 Counting cells . . . 36
2.1.3 Transient transfection and siRNA knockdown . . . 37
2.1.4 Lentiviral production and transduction . . . 38
2.1.4.1 Generation of lentiviral particles . . . 38
2.1.4.2 Lentivirus titration . . . 38
2.1.4.3 Transduction of cells . . . 39
2.1.5 Measuring cell viability and survival . . . 40
2.1.5.1 Absorbance and luminescence . . . 40
2.1.5.2 Clonogenic assay . . . 40
2.1.6 Clonal expansion . . . 41
2.1.7 Cell synchronisation . . . 41
2.1.8 Propidium iodide staining . . . 42
2.1.8.1 FACS analysis . . . 42
2.1.9 DNA fibre immunofluorescence . . . 43
2.1.10 Chromosome counts and Fluorescence in situ hybridization . . 44
2.2 DNA analysis . . . 46
2.2.1 DNA extraction . . . 46
2.2.1.1 Formalin-fixed paraffin embedded (FFPE) tissue . . 46
2.2.1.2 Fresh-frozen tissue and cultured cells . . . 46
2.2.1.3 Bacteria . . . 47
2.2.1.4 Blood . . . 48

2.2.1.5 Hot sodium hydroxide and tris (HotSHOT) DNA extraction . . . 48
2.2.2 Determination of DNA quantity and quality . . . 48
2.2.2.1 NanoDrop . . . 48
2.2.2.2 Qubit . . . 49
2.2.2.3 TapeStation . . . 49
2.2.3 DNA purification from agarose gels . . . 49
2.2.4 Polymerase chain reaction (PCR) . . . 50
2.2.4.1 Long-range PCR amplification . . . 51
2.2.4.2 High-fidelity short-range PCR . . . 52
2.2.4.3 Site-directed mutagenesis . . . 52
2.2.5 Microsatellite instability testing . . . 53
2.2.6 *Pla2g2a* genotyping by restriction length polymorphism (RFLP) analysis and Sanger sequencing . . . 54
2.2.7 DNA/cDNA sequencing . . . 57
2.2.7.1 Sanger sequencing . . . 57
2.3 *HPRT1* library preparation . . . 58
2.3.0.1 Illumina MiSeq 3000 . . . 58
2.3.1 TaqMan$^{TM}$ genotyping of point mutations . . . 59
2.4 DNA Cloning . . . 60
2.4.1 Restriction digestion . . . 60
2.4.2 DNA purification . . . 60
2.4.3 DNA ligation . . . 60
2.4.4 Bacterial transformation . . . 60
2.5 Analysis of RNA . . . 61
2.5.1 RNA extraction . . . 61
2.5.2 Reverse transcription of messenger RNA (mRNA) into complementary DNA (cDNA) . . . 62
2.5.3 Real-Time quantitative PCR (RT-qPCR) . . . 63
2.5.3.1 Quantification of TaqMan probe efficiency . . . 64
2.6 Analysis of protein . . . 64
2.6.1 Protein purification and quantification . . . 64

2.6.2 Western blotting . . . 66
2.6.3 Protein immunohistochemistry . . . 67
2.6.3.1 Immunohistochemistry quantification . . . 68
2.6.4 Visualisation of endogenous fluorescent proteins in mouse tissue 68
2.6.4.1 Tissue embedding . . . 68
2.6.4.2 Fresh-frozen tissue sectioning . . . 69
2.6.4.3 Visualisation of fluorescent protein in fresh-frozen tissue 69
2.7 Mouse husbandry and procedures . . . 69
2.7.1 Animal husbandry . . . 69
2.7.2 Mouse lines . . . 70
2.7.3 Ear punches and genotyping . . . 70
2.7.4 Breeding . . . 70
2.7.5 Induction of $ER^{T2}$ Cre-recombinase . . . 71
2.7.6 Necropsy . . . 71
2.7.6.1 Gut and uterus preparation . . . 71
2.7.7 Scoring of macroscopic phenotypes . . . 72
2.7.8 Tissue processing and embedding . . . 72
2.7.9 Sectioning and histological staining of FFPE tissue . . . 72
2.8 Bioinformatics . . . 73
2.8.1 BAM mapping and calling . . . 73
2.8.2 Variant annotation . . . 74
2.8.3 Software . . . 74
2.9 Statistical analysis . . . 74

**3 Functional characterisation of *POLE* exonuclease domain mutations in a cell-based system 76**

3.1 Introduction . . . 76
3.1.1 Existing cellular models of *POLE* exonuclease domain mutations 77
3.1.2 CRISPR-Cas9 . . . 80
3.1.3 Rationale for study and aims for chapter . . . 81
3.2 Methods . . . 82
3.2.1 Design and generation of CRISPR-Cas9n mammalian expression vectors . . . 82

3.3 Results . . . 83
3.3.1 Generation of isogenic *POLE* wildtype cell lines using CRISPR-Cas9n . . . 85
3.3.2 Off-target effects . . . 86
3.3.3 Large scale deletion in *POLE* isogenic cell lines . . . 88
3.3.4 POLE protein expression in isogenic cell lines . . . 92
3.3.5 Mutation rate of Parental and CRISPR-WT isogenic cell lines 92
3.3.5.1 Mutation spectrum of *HPRT1* mutant cell lines . . . 97
3.3.6 Impact of *POLE* exonuclease domain mutations on cell proliferation . . . 103
3.3.7 Impact of *POLE* exonuclease domain mutations on the cell cycle 105
3.3.8 DNA polymerase dynamics in *POLE* exonuclease domain mutant cell lines . . . 107
3.3.9 DNA damage response in *POLE* exonuclease domain mutant cell lines . . . 111
3.3.10 Knockdown of MMR genes in *POLE* isogenic cell lines . . . 112
3.3.10.1 *MSH6* mutations in *POLE* mutant cell lines . . . 112
3.3.10.2 Cell viability of *MSH6* and *MLH1* knockdown *POLE* exonuclease domain mutant cell lines . . . 114
3.3.11 Effect of $POLE^{P286R}$ exonuclease domain mutations on sensitivity to chemotherapeutic agents . . . 117
3.3.12 Overexpression of *POLE* in diploid MSS colorectal cancer cell lines . . . 120
3.3.12.1 Constitutive overexpression of exonuclease domain mutant POLE in CRC cells . . . 121
3.4 Discussion . . . 127
3.4.1 Generation and genetic characterisation of *POLE* isogenic cell lines . . . 128
3.4.2 Causal role of *POLE* exonuclease domain mutation in ultramutation and mutation Signature 10 . . . 129
3.4.3 *POLE* exonuclease domain mutations do not cause a substantial replication defect in CRC cell lines . . . 130

3.4.4 Relationship of *POLE* exonuclease domain mutation with MMR loss . . . 131
3.4.5 Nucleoside analogues may not be a suitable treatment for heterozygous *POLE* exonuclease domain mutant cancers . . . 131
3.4.6 Summary . . . 133

**4 Characterisation of the *Pole*$^{L424V}$ mutation in mouse models 134**
4.1 Introduction . . . 134
4.1.1 Existing *Pole* exonuclease domain mutant mice . . . 135
4.1.2 *Adenomatous Polyposis Coli (Apc)* mutant mouse models . . . 136
4.2 Tissue-specific recombination by Cre-loxP . . . 138
4.3 Aims of chapter . . . 139
4.4 Results . . . 140
4.4.1 Generation of a conditional *Pole* L424V knock-in mouse model 140
4.4.2 Generation of a constitutive *Pole* L424V knock-in mouse model 144
4.4.2.1 Confirmation of *Pole* recombination in *PgkCre*,*Pole*$^{fl(L424V)/+}$ mice . . . 144
4.4.2.2 Phenotypic analysis of *PgkCre*,*Pole*$^{fl(L424V)/+}$ mice . . 147
4.4.3 Molecular analysis of *PgkCre*,*Pole*$^{fl(L424V)/+}$ mice . . . 148
4.4.3.1 Generation of a germline *Pole* L424V knock-in mouse 150
4.4.4 Generation of an intestinal-specific *Pole* L424V knock-in mouse model . . . 151
4.4.4.1 Phenotyping of *Lgr5CreERT2*,*Pole*$^{fl(L424V)/+}$ mice . . 151
4.4.5 Generation and phenotyping of a uterine-specific *Pole* L424V mouse model . . . 153
4.4.6 Investigation of *Pole*$^{L424V/+}$ in an *Apc*-mutant background . . 156
4.4.6.1 Generation of compound mutant *Apc*$^{Min}$,*Pole*$^{L424V/+}$ mice . . . 157
4.4.6.2 Macroscopic analysis of L424V/+ *Apc*$^{Min}$ compound mice . . . 157
4.4.6.3 Modifiers of Min . . . 160
4.5 Discussion . . . 160
4.5.1 *Pole* fl(L424V) is homozygous lethal in mice . . . 161

4.5.2 *Pole* L424V/+ alone does not induce tumourigenesis in mice . 162
4.5.3 *Pole* L424V increases colonic polyposis in $Apc^{Min}$ mice . . . . 163

**5 When do somatic *POLE* exonuclease domain mutations occur in colorectal cancer? 165**

5.1 Introduction . . . . . 165
5.1.1 Timing of *POLE* mutations in cancer . . . . . 165
5.1.2 Immune response, prognosis and therapies . . . . . 167
5.1.3 Aims of this chapter . . . . . 167
5.2 Methods . . . . . 168
5.3 Results . . . . . 168
5.3.1 Selection of Adenomas to screen for *POLE* proofreading domain mutations . . . . . 169
5.3.2 Immune infiltrate quantification . . . . . 169
5.4 Microsatellite instability testing . . . . . 172
5.4.1 Next generation sequencing of *POLE* EDM adenomas . . . . . 173
5.5 Discussion . . . . . 176
5.5.1 *POLE* exonuclease domain mutations are an early event in colorectal cancer development . . . . . 176
5.5.2 *POLE* exonuclease domain mutant adenomas have an increased T-cell immune infiltrate . . . . . 177
5.5.3 The $POLE^{V411L}$ adenoma shows an amino acid substitution bias 178

**6 Discussion and future work 179**

6.1 Discussion and future work . . . . . 179
6.1.1 Somatic *POLE* exonuclease domain mutations causes ultramutation and mutational bias in cancer . . . . . 180
6.1.2 Germline $Pole^{L424V/+}$ mutation increases in colonic polyp burden in tumour-prone $Apc^{Min}$ mice . . . . . 182
6.1.3 Therapeutic implications of *POLE* exonuclease domain mutations 184
6.2 Ongoing and future work . . . . . 184
6.3 Concluding statement . . . . . 190

**Appendices 192**

**A Primer sequences and PCR conditions** **192**

**B Genes distal to *POLE* on chromosome 12q** **195**

**C Comprehensive list of Lesch Nyhan syndrome variants** **196**

**D Nomenclature of $Pole^{L424V}$ mutant mice** **204**

**E Gross morphological and histological analysis of constitutive *PgkCre*, $Pole^{fl(L424V)/+}$ mice** **205**

**F Gross morphological and histological analysis of inducible *Lgr5CreERT2*, $Pole^{fl(L424V)/+}$ mice** **208**

**G Gross morphological and histological analysis of *PRCre*,$Pole^{fl(L424V)/+}$ mice** **211**

**H Gross morphological analysis of compound mutant $Apc^{Min}$,$Pole^{L424V/+}$ mice** **214**

**I Raw intestinal polyp counts from compound mutant $Apc^{Min}$,$Pole^{L424V/+}$ mice** **216**

**J Molecular Inversion Probe panel genes and target regions** **219**

**References** **220**

# List of Figures


1.1 Adenoma-carcinoma sequence . . . . . 3
1.2 PRALINE multiple alignment of POLE and POLD1 exonuclease domains 15
1.3 Pymol structure of exonuclease active site mutant *Saccharomyces cerevisiae* Pol2 (analogous to POLE) N-terminal domain . . . . . 16
1.4 Lollipop plot of cancer-associated mutations in the exonuclease domains of POLD1 and POLE . . . . . 24
1.5 $Can^r$ mutation frequencies in *POLE* exonuclease domain mutant *S. cerevisiae* . . . . . 27
1.6 Mutation Signature 10 from Alexandrov *et al.* (1) . . . . . 29

2.1 Lenti-X p24 Rapid Titer Kit standard curve . . . . . 39
2.2 Standard FACS parameters for the analysis of propidium iodide-stained cells . . . . . 43
2.3 TapeStation output . . . . . 50
2.4 Microsatellite instability analysis . . . . . 55
2.5 *Pla2g2a* genotyping using restriction fragment length polymorphism (RFLP) analysis and Sanger sequencing . . . . . 56
2.6 28S and 18S ribosomal RNA bands . . . . . 62
2.7 TaqMan reaction probe efficiency standard curve and amplification plot 65
2.8 CB-X™ protein quantification assay standard curve . . . . . 66

3.1 Single guide RNA (sgRNA) 1 and 2 . . . . . 82
3.2 PX335 commercial Cas9 nickase plasmid and sgRNA cloning procedure 84
3.3 Green fluorescent protein transfection efficiency . . . . . 86

3.4 Chromatogram of the heterozygous *POLE* c.857C→G (P286R) mutation and the corrected mutants . . . . . 87
3.5 Chromosome counts of isogenic cell lines . . . . . 90
3.6 Cytogenetic analysis of isogenic cell lines . . . . . 91
3.7 POLE protein expression in isogenic cell lines . . . . . 93
3.8 dNTP synthesis pathways and *HPRT1* assay mechanism . . . . . 94
3.9 Representative images of plating efficiency and 6-TG selection plates for the *HPRT1* assay . . . . . 95
3.10 *HPRT1* mutation rate in isogenic cell lines . . . . . 96
3.11 Schematic of the *HPRT1* sequencing protocol . . . . . 98
3.12 Agarose gel of *HPRT1* amplicons for NGS library preparation . . . . 100
3.13 TapeStation output for the *HPRT1* library . . . . . 100
3.14 Cell proliferation curves and doubling time in isogenic cell lines . . . 104
3.15 Cell cycle analysis of isogenic cell lines . . . . . 106
3.16 DNA replication fork structures generated from the DNA fibre assay . 108
3.17 DNA replication analysis in isogenic cell lines by DNA fibre assay . . 108
3.18 DNA replication fork structures in isogenic cell lines . . . . . 110
3.19 Proportion of stalled replication forks to elongating replication forks following hydroxyurea treatment . . . . . 111
3.20 CHK1 and Phospho-CHK1 protein expression in isogenic cell lines . . 112
3.21 MLH1 and MSH6 protein expression post-siRNA knockdown . . . . . 114
3.22 Cell viability of HCC2998 isogenic cell lines after MSH6 or MLH1 siRNA knockdown . . . . . 115
3.23 Expression of apoptotic and DNA damage response markers in isogenic cell lines . . . . . 116
3.24 5-fluorouracil sensitivity in *POLE* exonuclease domain mutant CRC cell lines . . . . . 117
3.25 Dose-response curves of HCC2998 isogenic cell lines exposed to 5-FU, camptothecin and fludarabine . . . . . 119
3.26 Pilot dose-response curve of HCC2998 isogenic cell lines exposed to cytarabine . . . . . 120
3.27 Amplification, cloning and site-directed mutagenesis of *POLE* cDNA 122

3.28 GFP transfection efficiency and POLE-HA expression in SW480 cells 123
3.29 Screening of SW480 clones transduced with POLE-HA lentivirus . . . 124
3.30 Tetracycline off system and POLE-HA inhibition by doxycycline . . . 125
3.31 *HPRT1* forward mutation assay results from POLE-HA P286R, exo-null and wildtype overexpressing cell lines . . . . . 126
3.32 Plating efficiency POLE-HA P286R, exo-null and wildtype cell lines . 127

4.1 Wnt signalling pathway schematic . . . . . 137
4.2 Human APC protein domains . . . . . 138
4.3 Conservation of *POLE* and the $Pole^{fl(L424V)}$ knock-in targeting construct 141
4.4 Multiplex PCR from two heterozygous $Pole^{fl(L424V)/+}$ and two wildtype mice . . . . . 142
4.5 *Pole* mRNA expression in the small intestine, uterus and kidney . . . 143
4.6 *Pole* recombination PCR and *Pole* L424V expression chromatogram . 145
4.7 Analysis of mCherry protein expression . . . . . 146
4.8 Gastric (Stomach) and neck lymphomas from a *PgkCre*,$Pole^{fl(L424V)/+}$ mouse . . . . . 147
4.9 Representative H&E-staned intrathoracic tumour sections from a 17.6 week *PgkCre*,$Pole^{fl(L424V)/+}$ mouse . . . . . 148
4.10 Molecular analysis of apoptosis and proliferation in morphologically normal 52 week old small intestinal tissue . . . . . 149
4.11 *Lgr5CreERT2* construct and Lgr5-EGFP expression in intestinal stem cells . . . . . 152
4.12 Microscopic analysis of a H&E-stained spleen section from a 52 week old *Lgr5CreERT2*,$Pole^{fl(L424V)/+}$ mouse with splenomegaly . . . . . 152
4.13 Representative images of the uterus and mammary glands from experimental and control mice at 52 weeks . . . . . 154
4.14 Lymphomas and organs with lymphocytic infiltrate in *PRCre*,$Pole^{fl(L424V)/+}$ and *PRCre* mice . . . . . 155
4.15 Uterine hyperplasia from one experimental and one control mouse . . 156
4.16 Colonic polyp burden in $Apc^{Min}$,$Pole^{L424V/+}$ compound mutant mice . 158
4.17 Small intestine polyp burden and size in $Apc^{Min}$,$Pole^{L424V/+}$ compound mutant mice . . . . . 159

4.18 Representative images of *Pla2g2a* genotyping of $Apc^{Min}$,$Pole^{L424V/+}$ and $Apc^{Min}$ mice . . . . . 161

5.1 Somatic *POLE* exonuclease domain variants of unknown significance in early intestinal adenomas . . . . . 170
5.2 $POLE^{V411L}$ mutant adenoma H&E, chromatogram and TaqMan plot 171
5.3 $CD8^{+}$ immunohistochemistry . . . . . 172
5.4 Cumulative % variation in mononucleotide and dinucleotide microsatellite marker peak frequency between gDNA and tDNA . . . . . 173
5.5 Variant allele frequency of genic mutations in the $POLE^{V411L}$ exonuclease domain mutant colorectal adenoma . . . . . 175
5.6 Mutation spectrum of $POLE^{V411L}$ exonuclease domain mutant and $POLE^{WT}$ colorectal adenomas . . . . . 176

6.1 Mutation rate analysis schematic in *POLE* exonuclease domain mutant isogenic cell lines using whole genome sequencing . . . . . 186
6.2 Small intestinal cystic crypts and colonic aberrant crypt foci . . . . . 188
6.3 Measuring mutation accumulation in $Pole^{L424V/+}$ mice using organoid culture and sequencing . . . . . 189

E.1 Representative images of H&E stained FFPE tissue sections from experimental (*PgkCre*,$Pole^{fl(L424V)/+}$) and control ($Pole^{fl(L424V)/+}$ and *PgkCre*) mice. . . . . 207

F.1 Representative images of methylene blue stained intestine and H&E stained intestinal FFPE tissue sections from experimental *Lgr5CreERT2*, $Pole^{fl(L424V)/+}$ and control *Lgr5CreERT2* and $Pole^{fl(L424V)/+}$ mice . . . 210

G.1 Representative images of H&E stained uterus and mammary gland sections from experimental *PRCre*,$Pole^{fl(L424V)/+}$ and control *PRCre* mice at at 24 weeks . . . . . 211

# List of Tables


1.1 Genome stability genes mutated in germline and sporadic cancer . . . 6
1.2 Table of MutS and MutL protein dimers and their function in mismatch repair . . . 8
1.3 DNA polymerases and their function in eukaryotic cells . . . 11
1.4 Germline cancer-associated *POLE* and *POLD1* exonuclease domain mutations . . . 21
1.5 Somatic cancer-associated *POLE* and *POLD1* exonuclease domain mutations . . . 22

2.1 Colorectal cell lines used in this thesis . . . 36
2.2 siRNA reagents and product codes for gene knockdown . . . 37
2.3 Fluorescence in situ hybridization BAC probes. . . . 45
2.4 Standard PCR master mix . . . 51
2.5 Standard PCR reaction conditions . . . 51
2.6 Standard long-range PCR master mix . . . 51
2.7 Standard long-range PCR reaction conditions . . . 52
2.8 Standard High-Fidelity PCR master mix. . . . 52
2.9 Standard High-Fidelity PCR reaction conditions. . . . 52
2.10 QuikChange II XL site-directed mutagenesis master mix. . . . 53
2.11 QuikChange II XL site-directed mutagenesis reaction conditions. . . . 53
2.12 Primers for microsatellite status testing . . . 54
2.13 BigDye Terminator v3.1 master mix. . . . 57
2.14 BigDye Terminator v3.1 PCR conditions. . . . 57
2.15 MiSeq library barcoding PCR conditions. . . . 58

2.16 TaqMan genotyping master mix. . . . . 59
2.17 TaqMan genotyping PCR conditions. . . . . 59
2.18 High Capacity cDNA reverse transcriptase kit master mix. . . . . 62
2.19 High Capacity cDNA reverse transcriptase kit PCR conditions. . . . . 63
2.20 TaqMan Gene expression assay probes . . . . . 63
2.21 TaqMan RT-qPCR thermal cycling conditions. . . . . 64
2.22 Antibodies for western blotting. . . . . 67
2.23 Immunohistochemistry primary and secondary antibodies . . . . . 68
2.24 Mouse strains used in this study . . . . . 70
2.25 Illumina adapter sequences for *HPRT1* gene library preparation. . . . 73
2.26 Software to process and analyse high throughput sequencing data. . . 74
2.27 VEP custom annotation tools. . . . . 74

3.1 Published cell lines with *POLE* exonuclease domain mutations and their limitations. . . . . 79
3.2 Single-stranded oligonucleotide (ssODN) repair template sequence. . . 83
3.3 *POLE* exonuclease domain mutant colorectal cancer cells lines . . . . 83
3.4 Targeting efficiency of the CRISPR-Cas9 nickase in HCC2998 and SNU-81 cell lines. . . . . 86
3.5 Predicted off-target binding sites of sgRNA-1 and sgRNA-2 in within genic regions . . . . . 88
3.6 Single nucleotide variants in the Signature 10 trinucleotide context (TCT>TAT, TCG>TTG, TTT>TGT) in the 5'-3' and 3'-5' direction. Ref is the reference nucleotide and Alt is the mutant nucleotide. . . . 97
3.7 Illumina adapter sequences . . . . . 99
3.8 Contingency table of *POLE* Signature 10 and other mutations in *HPRT1* mutant isogenic cell lines and Lesch-Nyhan syndrome patients . . . . 102
3.9 Somatic *MSH6* mutations in *POLE* exonuclease domain mutant, wild-type and microsatellite unstable colorectal cancer cell lines . . . . . 113

4.1 $\chi^2$ test for homozygous $Pole^{\mathrm{fl(L424V)}}$ embryonic lethality . . . . . 143
4.2 $\chi^2$ test for homozygous $Pole^{\mathrm{(L424V)}}$ embryonic lethality . . . . . 150

A.1 Primers for the generation and genetic characterisation of CRISPR-Cas9n constructs and cell lines . . . . . 192
A.2 *HPRT1* exon sequencing primers . . . . . 193
A.3 Sequencing and site-directed mutagenesis primers for the generation of *POLE* overexpression constructs. . . . . 193
A.4 Mouse genotyping and sequencing primers . . . . . 194
A.5 FFPE DNA *POLE* sequencing primers. . . . . 194

B.1 Genes distal to *POLE* on chromosome 12q . . . . . 195

C.1 A comprehensive list of published Lesch-Nyhan syndrome variants . 196

D.1 Nomenclature of *Pole*-mutant mice used in this study . . . . . 204

E.1 *PgkCre*,$Pole^{\mathrm{fl(L424V)/+}}$ (experimental), *PgkCre* (control), and $Pole^{\mathrm{fl(L424V)/+}}$ (control) mice necropsy . . . . . 205
E.2 *PgkCre*,$Pole^{\mathrm{fl(L424V)/+}}$, $Pole^{\mathrm{fl(L424V)/+}}$ and *PgkCre* mouse histopathology results . . . . . 206

F.1 Reason for sacrifice, age of sacrifice and results from necropsy from *Lgr5CreERT2*, $Pole^{\mathrm{fl(L424V)/+}}$, *Lgr5*-Cre and $Pole^{\mathrm{fl(L424V)/+}}$ mice. . . . 208
F.2 Intestine and tumour histopathology results from *Lgr5CreERT2*,$Pole^{\mathrm{fl(L424V)/+}}$, *Lgr5CreERT2* and $Pole^{\mathrm{fl(L424V)/+}}$ mice. . . . . 209

G.1 Reason for sacrifice, age of sacrifice and results from necropsy from *PRCre*,$Pole^{\mathrm{fl(L424V)/+}}$ and *PRCre* mice . . . . . 212
G.2 Histopathology results from *PRCre*,$Pole^{\mathrm{fl(L424V)/+}}$ and *PRCre* mice . . 213

H.1 Gross morphological analysis of $Apc^{\mathrm{Min}}$,$Pole^{\mathrm{L424V/+}}$ and $Apc^{\mathrm{Min}}$ mice 215

I.1 Raw intestinal polyp counts from $Apc^{\mathrm{Min}}$,$Pole^{\mathrm{L424V/+}}$ mutant compound and $Apc^{\mathrm{Min\ mice}}$ . . . . . 217
I.2 Ratio of mononucleotide and dinucleotide marker peaks in gDNA/tDNA for adenomas . . . . . 218

J.1 Genes and gene fusion events covered by the Molecular Inversion Probe (MIP) panel covering 63.1kb . . . . . 219

# List of Abbreviations

| | |
|---|---|
| AA | Anaplastic astrocytoma |
| ABC | Avidin/Biotinylated enzyme complex |
| ANOVA | Analysis of variance |
| Ara-C | Cytarabine |
| APC | Adenomatous Polyposis Coli |
| APE1 | Apurinic endonuclease 1 |
| ATCC | American Type Culture Collection |
| ATP | Adenosine Triphosphate |
| BAC | Bacteriophage |
| BER | Base-excision repair |
| BLCA | Bladder cancer |
| BrC | Breast cancer |
| BrdU | Bromodeoxyuridine |
| BSA | Bovine serum albumin |
| CALMs | cafe-au-lait macules |
| Cas9n | Cas9-nickase |
| CC | Squamous cervical cancer |
| CDK | Cyclin-dependent kinase |
| cDNA | Complementary DNA |
| Chr | Chromosome |
| CIMP | CpG island methylator phenotype |
| CldU | 5-Chloro-2-deoxyuridine |
| CMG | Cdc45, Mcm2-7 and GINS |
| CMMRD | Constitutional biallelic mismatch repair deficiency |
| CMV | Cytomegalovirus |
| CNA | Copy number alteration |
| CORGI | COloRectal Gene Identification |
| COSMIC | Catalogue of Somatic Mutations in Cancer |
| CPT | Camptothecin |
| CRC | Colorectal cancer |
| Cre | cyclization recombination enzyme |
| CRISPR | Clustered regularly interspaced short palindromic repeats |
| Cryo-EM | Cryo-Electron microscopy |

| | |
|---|---|
| Ct | Cycle threshold |
| C/V | Crystal violet |
| DAB | 3,3-diaminobenzidine |
| DAPI | 4,6-diamidino-2-phenylindole dihydrochloride |
| DMEM | Dulbeccos Modified Eagles Medium |
| DMSO | Dimethyl Sulphoxide |
| DNA | Deoxyribonucleic acid |
| dNTP | Deoxynucleoside triphosphate |
| DPX | di-N-butyl phthalate in xylene |
| DSB | Double-strand break |
| DSBR | Double-strand break repair |
| DuC | Duodenal cancer |
| EC | Endometrial cancer |
| ECL | Enhanced chemiluminescence |
| EDM | Exonuclease domain mutation |
| EDTA | Ethylenediamine tetra-acetic acid |
| EEX | Error-induced extinction |
| ELISA | Enzyme-linked immunosorbent assay |
| ENU | N-ethyl-N-nitrosourea |
| EOC | Endometriod ovarian cancer |
| FACS | Fluorescence-activated cell sorting |
| FAP | Familial Adenomatous Polyposis |
| FBS | Fetal bovine serum |
| FFPE | Formalin-Fixed Paraffin-Embedded |
| FISH | Fluorescence in situ hybrisization |
| FITC | Fluorescein isothiocyanate |
| G1 | Gap1 |
| G2 | Gap2 |
| GAPDH | Glyceraldehyde 3-phosphate dehydrogenase |
| GBM | Glioblastoma |
| GC | Gastric cancer |
| gDNA | Genomic DNA |
| GFP | Green fluorescent protein |
| GINS | go, ichi, ni and san |
| GWAS | Genome-wide association study |
| HA | Human influenza hemagglutinin |
| H&E | Haematoxylin and Eosin |
| HT | Hypoxanthine-Thymidine |
| HAT | Hypoxanthine-Aminopeterin-Thymidine |
| HBS | HEPES-buffered saline |
| HDR | Homology-directed repair |
| HEPES | N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid |
| HotSHOT | Hot sodium hydroxide and tris |

| | |
|---|---|
| HPF | Highpower fields |
| HPRT1 | Hypoxanthine-guanine phosphoribosyltransferase |
| HR | Homologous recombination |
| HRP | Horseradish peroxidase |
| HU | Hydroxyurea |
| ICL | interstrand crosslink |
| IdU | 5-Iodo-2-deoxyuridine |
| IGV | Integrative Genomics Viewer |
| IHC | Immunohistochemistry |
| IMDM | Iscove's Modified Dulbecco's Medium |
| Indels | Insertions and deletions |
| IRES | Internal ribosome entry site |
| kDa | Kilo Dalton |
| L-15 | Leibovitz's L-15 Medium |
| LB | Luria-Bertani Media |
| LNS | Lesch-Nyhan syndrome |
| LOF | Loss of function |
| LOH | loss of heterozygosity |
| LoxP | locus of chromosomal crossover in the bacteriophage P1 |
| LP-BER | Long-patch base-excision repair |
| MAF | Minor allele frequency |
| MAP | MUTYH-associated polyposis |
| MCM | Minichromosome maintenance |
| MEF | Mouse embyonic fibroblast |
| MESC | Mouse embryonic stem cell |
| Min | Multiple intestinal neoplasia |
| MIP | Molecular inversion probe |
| MM | Multiple myeloma |
| MMR | Mismatch repair |
| MMRD | Mismatch repair deficiency |
| MMS | Methyl methanesulfonate |
| MOI | Multiplicity of infection |
| mRNA | Messenger RNA |
| MSI | Microsatellite instability |
| MSS | Microsatellite stable |
| NaBut | Sodium butyrate |
| NaOH | Sodium hydroxide |
| NBF | Neutral buffered formalin |
| NER | Nucleotide excision repair |
| NGS | Next Generation Sequencing |
| NHEJ | Non-homologous end joining |
| NLS | Nuclear localisation signal |
| NMR | Nuclear magnetic resonance |
| NR | Not reported |

| | |
|---|---|
| NTD | N-terminal domain |
| OC | Ovarian cancer |
| OCT | Optimal Cutting Temperature |
| ODG | Oligodendroglioma |
| ORC | Origin recognition complex |
| PAAD | Pancreatic cancer |
| PAM | Protospacer adjacent motif |
| PAP | Peroxidase-Antiperoxidase |
| PBS | Phosphate buffered saline |
| PCNA | Proliferating cell nuclear antigen |
| PCR | Polymerase chain reaction |
| PDB | Protein data bank |
| PE | Plating efficiency |
| PGK | Phosphoglycerate kinase |
| PPAP | Polymerase-proofreading-associated polyposis |
| PR | Progesterone Receptor |
| PrC | Prostate cancer |
| Pre-RC | Prereplicative |
| P/S | Penicillin-Streptomycin |
| PVDF | polyvinylidene difluoride |
| RIPA | Radioimmunoprecipitation assay |
| RT-qPCR | Real-time quantitative PCR |
| RFC | Replication factor C |
| RFP | Red fluorescent protein |
| RNA | Ribonucleic acid |
| RPA | Replication Protein A |
| RPMI | Roswell Park Memorial Institute |
| SAC | Spindle-assembly checkpoint |
| SB | Small bowel |
| SD | Standard deviation |
| SDS | Sodium dodecyl sulfate |
| SEM | Standard error of the mean |
| sgRNA | Single-stranded guide RNA |
| shRNA | short hairpin RNA |
| siRNA | small interfering Ribonucleic Acid |
| SNP | Single nucleotide polymorphism |
| S-phase | Synthesis phase |
| SOC | Super Optimal Broth |
| SP-BER | Short-patch base-excision repair |
| SSC | Saline sodium citrate |
| ssODN | single stranded oligonucleotide |
| TALENs | Transcription activator-like effector nuclease |
| TBE | Tris-Borate EDTA |
| TBS | Tris buffered saline |

| | |
|---|---|
| TBST | Tris buffered saline with Tween-20 |
| TCGA | The Cancer Genome Atlas |
| tDNA | Tumour DNA |
| TE | Tris-EDTA |
| TetO | Tetracycline operator |
| TetR | Tetracycine repressor |
| TILs | Tumour-infiltrating lymphocytes |
| TLS | Translesion synthesis |
| TP53 | Tumor protein p53 |
| TRE | Tetracycline-response element |
| TSG | Tumour suppressor gene |
| tTA | tetR plus C-terminus of VP16 |
| UTR | Untranslated region |
| UV | Ultraviolet |
| VAF | Variant allele frequency |
| WCHG | Wellcome Centre for Human Genetics |
| WES | Whole exome sequencing |
| WGS | Whole genome sequencing |
| WT | Wild-type |
| YFP | Yellow-fluorescent protein |
| ZFN | Zinc-finger nuclease |
| 5-FU | 5-fluorouracil |
| 6-TG | 6-thioguanine |
| 8-oxoG | 8-Oxoguanine |

# Chapter 1

# Introduction

Faithful genome replication is crucial to suppress mutagenesis and, in higher eukaryotes, tumourigenesis. Multiple DNA repair pathways operate to minimise replication errors and thus maintain genome stability. Nevertheless, with each cell cycle, errors can arise due to exogenous and endogenous stress which, if uncorrected by these DNA damage repair processes, may cause a damaging mutation or chromosomal changes, leading to cancer. The critical importance of DNA repair pathways for tumour suppression is illustrated by their frequent dysregulation in cancer, for example, mutations in several mismatch repair (MMR) and base-excision repair (BER) genes are known to initiate colorectal cancer. A cause of genomic instability described recently comprises mutations in the exonuclease domain of *POLE* and *POLD1*, the catalytic subunits of DNA Polymerase $\epsilon$ and Polymerase $\delta$, respectively. During this chapter I will review several critical DNA repair pathways and describe how their perturbation leads to cancer. I will then summarise recent evidence regarding the molecular characteristics of *POLE* and *POLD1* mutations and the consequences of these for the development of various cancer subtypes.

## 1.1 Introduction to cancer

Cancer is a complex genetic disease. The accumulation of base-pair substitutions, insertions and deletions (indels), chromosomal rearrangements, copy number alterations (CNAs) and epigenetic changes constitutively activate proto-oncogenes and reduce the activity of tumour suppressor genes (TSGs) which, together with disruption of genome stability genes, provides cells with a selective proliferative advantage. The "hallmarks of cancer" define the phenotypic consequence of these genetic and epigenetic alterations (2, 3) and include replicative immortality, angiogenesis, avoidance of apoptosis, sustained proliferation, evading negative growth signals and ultimately metastasis. The acquisition of some or all of these are accelerated by underlying genomic instability.

TSGs and oncogenes are recurrently mutated in cancers, with patterns that are often found in tumours of the same histological subtype and tissue of origin. This pattern has perhaps been best characterised in colorectal cancer (CRC), partly due to the ease of collecting tissue biopsies through routine colonoscopy. Furthermore, the extensive hereditary component of CRC, predicted to be around 15-30%, including genome-wide association and other genomic contributions (3-5% Mendelian syndromes), has aided the identification of CRC driver genes (4). The development of CRC is thought to progress from normal tissue to pre-malignant cells to adenoma then advanced tumour and finally metastasis through sequential mutations in distinct TSGs and oncogenes, in a process referred to the adenoma-carcinoma sequence (5) (Figure 1.1). The model suggests that a few mutations in TSGs and oncogenes, as little as 15, are sufficient to initiate and drive tumourigenesis (4).

Wood *et al.* showed that a few somatically mutated genes are common to most colorectal cancers, including *APC*, *TP53*, *KRAS*, *PIK3CA*, *FBXW7*, *SMAD4* and *TGFBR2* (6). Somatic inactivating mutations in *APC* are found in 70-80% of colorectal carcinomas (4, 7). Together with *TP53* and *KRAS* mutations these have been suggested to be required for adenoma-carcinoma sequence progression (6). In addition to these key drivers, mismatch repair gene mutations and hypermethylation of the MMR gene *MLH1* are found in 15% of sporadic colorectal cancer cases with microsatellite instability. Germline mismatch repair (MMR) gene mutations cause the Mendelian colorectal cancer disorder, Lynch syndrome. In the past few years,

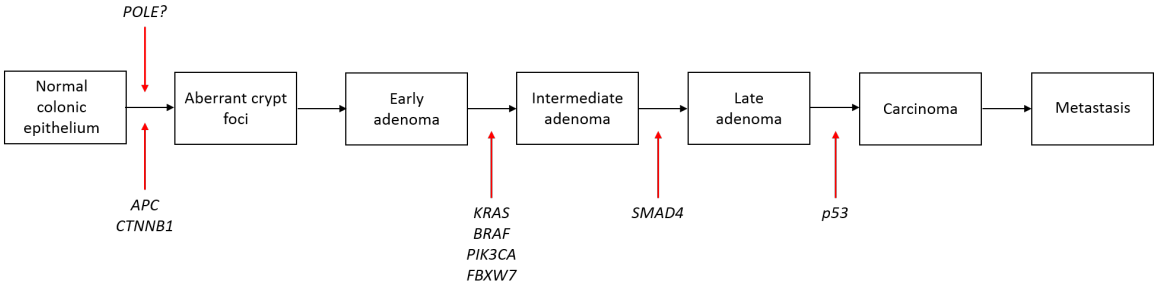



**Figure 1.1:** The adenoma-carcinoma sequence. Black arrows show the progression from pre-cancer to invasive adenocarcinoma. Red arrows represent the sequential accumulation of mutations in colorectal cancer driver genes. (Adapted from Fearon and Vogelstein 1990 (5)).

tumour sequencing has detected extensive genetic diversity and revealed that CRC can take various evolutionary trajectories, depending on the genes mutated.

In 2013, results from whole exome sequencing of 224 colon and rectal cancers were published by The Cancer Genome Atlas (TCGA) (8). The study confirmed Vogelstein and colleagues' earlier observations of common genetic mutations in key driver genes such as *APC*, *TP53*, *PIK3CA* and *KRAS*, amongst others. A multitude of additional sub-clonal mutations were, however, also identified which may suggest a more gradualist (polygenic) model of tumour progression (9).

Sixteen percent of cancers were hypermutated with upwards of 10 mutations/mb, of which 3/4 were microsatellite unstable, mostly with *MLH1* hypermethylation (8). The remaining 1/4 of hypermutated cancers had mismatch repair mutations and mutations in the exonuclease domain of *POLE*. Several tumours had upwards of 100 mutations/mb, termed ultramutation and exhibited extensive mutational heterogeneity. The mutation spectrum differed from non-ultramutated cancers, with mutations common in *ACVR2A*, *APC*, *TGFBR2*, *MSH3*, *MSH6*, *SLC9A9* and *TCF7L2* and *BRAF*, but not *CTNNB1* and *TP53*. Indeed, studies have shown that these cancers do not follow the classic sequence of mutations previously observed in other cancers (Figure 1.1) (10).

The contribution of genomic instability to tumour development has been widely debated and two conflicting hypotheses have been proposed, as discussed below. The first proposes the normal cellular mutation rate is sufficient to induce all of the mutations required to cause carcinogenesis and that DNA replication and repair defects are

found only in a subset of cancers. In this hypothesis, cancer is driven by the selection of cells with a survival or growth advantage, which then proliferate and acquire more mutations (5, 11). The occurrence of recurrent mutations across cancers demonstrates that certain mutations are positively selected for throughout tumour development. Furthermore, it has been proposed that the number of stem cell divisions may be adequate to generate the number of mutations in cancers without increasing the mutation rate (12). The second hypothesis, the "mutator phenotype", proposes that an elevated number of mutations caused by a DNA replication or repair defect lead to further mutations in TSGs and proto-oncogenes (13). Whilst sequencing efforts find recurrent mutations between samples, they have also detected a vast array of sub-clonal and passenger mutations, which appear to impart little or no selective advantage on the cell (14). The prevalence of somatic mutations that do not impart a selective advantage suggests that an increased mutation rate may be the initial driver for acquiring these mutations.

For brevity, I will focus on genes directly involved in accessing the double helix and repairing DNA, but not checkpoint genes or other downstream pathways, as this is beyond the scope of my thesis.

## 1.2 Mutagenesis and cancer

Spontaneous somatic mutations occur at a frequency of only $1x10^{-10}$ mutations/cell per genome duplication (13). This low mutation frequency is due to high fidelity nucleotide selectivity and exonucleolytic proofreading by the replicative DNA polymerases, post-replicative mismatch repair (MMR) and other DNA repair pathways. Disruption of these processes can lead to genomic instability, resulting in microsatellite instablity (MSI) and an increase in base-pair substitutions. The "mutator phenotype" hypothesis, proposed by Larry Loeb in 1974, suggested that an increase in mutation rate is an initiating factor in cancer progression (13). A mutator phenotype can be driven by the dysregulation of genome stability genes involved in DNA replication, repair and recombination due to mutations, epigenetic modifications or an increase in endogenous or exogenous carcinogenic stress such as smoking, radiation or inflammation.

DNA undergoes a vast amount of damage, with over 20,000 damaging events and over 10,000 replication errors/cell per day (15, 16) and the multiple enzymatic steps required to repair the damage before mitosis means that the possibility of a mutator phenotype occurring is not implausible. For example, budding yeast are predicted to have over 100 genes required for the maintenance of genomic stability (17). Data from The Cancer Genome Atlas show that tumours have a wide range of mutation frequencies, from a few 100 in pilocytic astrocytoma, to over 100,000 in melanoma and some uterine and colorectal cancers (1). Sequencing may underestimate the true mutation frequency, as sub-clonal allele frequency is below the threshold for variant detection. In an analysis of 12 types of cancer, a significant correlation was observed between a high mutation rate and mutations in DNA repair genes such as *PRKDC*, *TP53* and *MSH6* (18).

### 1.2.1 DNA repair defects in cancer

In addition to somatic mutations in genome stability pathways, several germline defects in DNA repair genes are associated with Mendelian cancer syndromes. One of the first examples was Xeroderma Pigmentosum, where patients harbour germline defects in the nucleotide excision repair (NER) pathway, predisposing them to skin cancer upon UV exposure (19). Mutations in tumour suppressor genes *BRCA1* and *BRCA2*, involved in homologous recombination of double strand breaks, predispose women to breast, ovarian and uterine cancer (20). Both of these defects cause an increase in chromosomal instability. A list of hereditary cancer syndromes associated with DNA repair gene defects and somatically mutated DNA repair genes are listed in Table 1.1. I will now describe in more detail how somatic and germline mutations in the base-excision repair and mismatch repair machinery are specifically involved in colorectal polyposis.

### 1.2.2 Base-excision repair

Germline mutations in multiple DNA repair genes predispose patients to colorectal cancer and are well characterised. Biallelic substitutions in the DNA glycosylase *MU-*

| Gene | Process | Somatic | Germline |
|---|---|---|---|
| *BRCA1/2* | HR | Yes | Hereditary breast and ovarian cancer syndrome |
| *FANCA, FANCD2, FANCM, FANCN* | HR | Yes | Hereditary breast and ovarian cancer syndrome |
| *RECQL3* | HR | NR | Bloom syndrome |
| *RECQL4* | HR | NR | Rothmund-Thompson syndrome |
| *WRN* | HR | Yes | Werner syndrome |
| *MRE11* | HR | Yes | Nijmegen syndrome |
| *BAP1* | HR | Yes | BAP1-associated syndrome |
| *BARD1* | HR | Yes | NR |
| *RAD51L3* | HR | Yes | NR |
| *RAD52* | HR | Yes | NR |
| *ATM* | HR/NHEJ | NR | Ataxia telangiectasia syndrome |
| *PRKDC* | NHEJ | Yes | NR |
| *MSH2, MSH6, MLH1, PMS2* | MMR | Yes | Lynch syndrome/Constitutional biallelic MMR syndrome |
| *PMS1* | MMR | Yes | NR |
| *XPB, XPD, XPF, XPG* | NER | Yes | Xeroderma pigmentosum |
| *RAD23A* | NER | Yes | NR |
| *MUTYH* | BER | NR | MUTYH-associated polyposis |
| *NTHL1* | BER | NR | NTHL1-associated polyposis |
| *XRCC1* | BER | Yes | NR |
| *BUB1* | SAC | Yes | NR |
| *CENPE* | SAC | Yes | NR |
| *ZW10* | SAC | Yes | NR |
| *TTK* | SAC | Yes | NR |
| *KNTC1* | SAC | Yes | NR |
| *STK12* | SAC | Yes | NR |

**Table 1.1:** Genome stability genes mutated in germline and sporadic cancer. HR, Homologous recombination; NHEJ, Non-homologous end joining; MMR, Mismatch repair; NER, Nucleotide excision repair; BER, Base-excision repair; SAC, Spindle assembly checkpoint; NR, Not reported. Table is adapted from Negrini *et al.* and Romero-Laorden *et al.* (21, 22).

*TYH* were identified in three affected siblings from the same family with colorectal adenomas and carcinomas (23). Shortly after this discovery, homozygous and compound heterozygous mismatch and truncating mutations in MUTYH were found in seven unrelated patients, each with an excess of 100 adenomas and some with colorectal carcinomas (24). The syndrome was named MUTYH-associated polyposis (MAP). MUTYH is one of 11 several specialised DNA glycosylases which initiate the base excision repair (BER) pathway through the detection and excising of DNA damage, such as oxidative lesions, alkylation, cytosine deamination and single-strand breaks (25). Intracellular oxidation and deamination of methylated cytosine can inflict in excess of 9000 base-pairs of DNA damage per cell per day (26).

Glycosylases, along with apurinic endonuclease (APE1) and ligase 1 or ligase III$\alpha$/XRCC1, are involved in short-patch (SP)-BER and FEN1 and Ligase 1 are involved in long-patch (LP)-BER. The gap-filling process is performed by DNA polymerases $\beta$, $\delta$ and possibly polymerase $\epsilon$. MUTYH, along with OGG1, are involved in the repair of 8-Oxoguanine (8-oxoG), one of the most extensively studied oxidative

lesions (27). During DNA synthesis, 8-oxoG pairs preferentially with adenine rather than cytosine, resulting in a 8-oxoG:A mismatch. OGG1 is involved in the excision of 8-oxoG from the template strand and MUTYH is involved in excision of adenine from the nascent strand. If repair is not successful before DNA replication, this leads to a G:C→T:A transversion.

Biochemical data show that attenuation or complete loss of glycosylase function for several of the cancer-associated MUTYH variants (23) and tumour DNA from patients show an excess of C→A transversions (23, 24, 28). *Myh*$^{-/-}$ mice, however, show an increased tumour susceptibility only when crossed with a truncating mutation in *Apc* (*Apc*$^{\mathrm{Min}}$) (29). These data show that MUTYH loss of function predisposes patients to colorectal adenomas and carcinomas, but a secondary mutation in an additional cancer driver gene may be required to initiate tumourigenesis.

Recently, homozygous nonsense mutations in *NTHL1*, another oxidative damage repair DNA glycosylase, were identified in the germline of patients with familial colon, endometrial, prostate, pancreatic and duodenal cancer (30). The mutation spectrum consisted of an increased frequency of C:G→T:A mutations, consistent with the mutation spectrum of *Nthl1* double-knockout mice (31). Tumours had an increased mutation burden (although not hypermutated). NTHL1 repairs a wider spectrum of oxidised lesions than MUTYH, potentially explaining the different mutation spectrum.

### 1.2.3 DNA Mismatch repair

One of the best characterised DNA repair defects in cancer is mismatch repair deficiency. Mismatch repair corrects errors made by the replicative polymerases during genome duplication, comprising either single base mismatches or insertions and/or deletions of one or more base pairs in repetitive (microsatellite) regions due to strand slippage during replication. Seven evolutionarily conserved mismatch repair proteins are involved in this pathway: MLH1, MSH6, MSH2, PMS1, PMS2, MSH3 and MLH3. In humans, these proteins act in dimers to repair different types of replication errors (Table 1.2). Replication errors are detected by either MutS$\alpha$ or MutS$\beta$, which subsequently interact with MutL$\alpha$ to nick the nascent strand in a complex with replication

factor C (RFC). Along with PCNA, the 5'-3' exonuclease Exo1 and possibly Pol $\epsilon$ and $\delta$ excise the mismatch, the DNA is resynthesised and ligated by LigI (32).

| Complex | Protein dimer | Repair specificity |
|---|---|---|
| Muts$\alpha$ | MSH2-MSH6 | Single base pair insertions and deletions |
| MutS$\beta$ | MSH2-MSH3 | Long (up to 16bp) insertions and deletions |
| MutL$\alpha$ | MLH1-PMS2 | Variety of mismatches |
| MutL$\beta$ | MLH1-PMS1 | Unknown |
| MutL$\gamma$ | MLH1-MLH3 | Meiotic recombination |

**Table 1.2:** Table of MutS and MutL protein dimers and their function in mismatch repair. Data adapted from Kunkel and Erie (33).

Approximately 15% of CRC cases have either methylation of *MLH1* or pathogenic somatic mutations in MMR genes and 2-3% of CRC patients have a germline mutation in one or more mismatch repair genes (4). Cancers with MMR deficiency are characterised by an elevated number of indels, particularly in microsatellite regions, a phenomenon termed microsatellite instability (MSI). The increase in indels and single base substitutions can cause frameshift and missense mutations, increasing the potential for pathogenic mutations in TSGs or proto-oncogenes. The first evidence of the association of mismatch repair deficiency with cancer were in patients at least one first degree relative with early-onset colorectal and endometrial cancer, known as Lynch syndrome. Patient tumours were chromosomal stable but, they did display microsatellite instability, similar to that observed previously in yeast models with defects in the mismatch repair genes (34). *MSH2*, *MLH1*, *PMS2* and *MSH6* have all been found to be mutated in the germline of patients with Lynch syndrome (35, 36, 37, 38, 39). Germline mutations in *MLH1* and *PMS2* have also been found in patients with Turcot syndrome which predispose carriers to polyposis and colorectal cancer, medulloblastomas and glioblastoma. Genetic redundancy of some mismatch repair genes means that loss of MSH6, MSH3 and PMS1 are less harmful than loss of MLH1 and MSH2, which form dimers with more than one mismatch repair proteins. Mutations in MSH3 are rarely found in hereditary colorectal cancer.

Somatic MMR deficiency is relatively common in colorectal cancer. Most cases of mismatch repair deficiency (MMRD) in sporadic cancer result from *MLH1* promoter hypermethylation by $BRAF^{V600E}$ gain of function mutations (40, 41). *MLH1* promoter methylation may be associated with the methylation of other CpG dinucleotides

throughout the genome termed the CpG island methylator phenotype (CIMP). Tumours with MMR deficiency display a hypermutation, typically between 10-100 mutation/Mb (8). They also display a diploid karyotype, an increased immune infiltrate and have a good prognosis in early stage tumours (42).

## 1.3 DNA replication

The mechanism of DNA replication is highly conserved from archea, to bacteria and eukaryotes. Two anti-parallel DNA strands within the double helix are synthesised unidirectionally by semi-conservative replication, with one continuously (leading-strand) and one discontinuously (lagging-strand) replicating strand. Discontinuous replication of the lagging strand is replicated in 200-350bp fragments, called "Okazaki fragments" after Reji Okazaki & Tuneko Okazaki (43). The assembly of the replication forks and the initiation of DNA replication is described below.

### 1.3.1 The Eukaryotic replication fork

DNA replisome assembly comprises four main steps: 1) loading of the DNA helicase at origins of replication; 2) formation of the pre-initiation complex; 3) activation of the DNA helicase to begin DNA unwinding; 4) priming and replication of DNA.

The eukaryotic helicase, Minichromosome maintenance (MCM)2-7, is a heterohexameric protein comprising six distinct subunits. MCM2-7 is loaded onto double-stranded DNA at origins marked by the origin recognition complex (ORC)-Cdc6 and Cdt1. A second MCM-Cdt1 complex is then recruited to form a double hexamer surrounding the DNA called the pre-replication (Pre-RC) complex (44, 45). Additional factors, Cdc45 and GINS, are subsequently recruited to activate the helicase. GINS (derived from the Japanese go, ichi, ni and san, meaning five, one, two and three) consists of Sld5, Psf1, Psf2 and Psf3 loaded tightly with DNA Polymerase $\epsilon$ (46). Finally the double hexamer is converted into a single hexamer via post-translational modifications by cyclin-dependent kinases (CDKs) and Dpf4-dependent kinase (DDK) catalysed by Mcm10. The two active helicases and the CMG complex (Cdc45, Mcm2-7 and GINS) translocate along the leading strand in opposite direc-

tions creating single-stranded (ss)DNA coated with RPA ready to be replicated by DNA polymerases (47).

### 1.3.2 The DNA polymerases

Humans encode a total of 16 DNA polymerases, maintaining nuclear and mitochondrial genome stability by DNA replication, repair and translesion synthesis (Table 1.3). The bulk of nuclear DNA replication is carried out by the three multisubunit B-family polymerases, Pol $\epsilon$, Pol $\delta$ and Pol $\alpha$-primase (45, 48, 49, 50, 51). Replication is initiated with the priming of DNA by Pol $\alpha$, consisting of two primase subunits (PRIM1 and PRIM2) and two polymerase subunits (POLA1 and POLA2). RNA priming is coordinated by PRIM1, then POLA1 takes over to continue dNTP extension by approximately 20-30nt (52). The replication fork is then extended by polymerases $\epsilon$ and $\delta$. Unlike Pol $\alpha$, Polymerases $\epsilon$ and $\delta$ contain an active 3'-5' exonuclease domain which proofreads newly synthesised DNA for replication errors (53, 54). Pols $\epsilon$ and $\delta$ are comprised of four subunits in humans, the largest of which contains the catalytic polymerase and exonuclease domains and is encoded by *POLE* and *POLD1* respectively. Both polymerases contain accessory subunits, encoded by *POLD2*, *POLD3* and *POLD4*, and *POLE2*, *POLE3* and *POLE4*, at a 1:1:1:1 stoichiometry with the catalytic subunit (55).

Polymerase $\delta$ is an essential protein in eukaryotes due to its important role in high-fidelity DNA replication, repair of double-strand breaks by homologous recombination and is one of the major polymerases involved in gap-filling during base-excision, nucleotide excision and mismatch repair (56). The catalytic subunit of Pol $\epsilon$ (*POLE*) is twice the size of *POLD1*. The 140kDa N-terminal portion of POLE contains the polymerase and exonuclease domains of the protein and the C-terminal portion contains four zinc finger motifs, mediating the interaction between POLE and the accessory subunits (57). Interestingly, studies in budding yeast have demonstrated that the catalytic subunit of POLE is dispensable for DNA replication and cell viability, but deletion of the non-catalytic C-terminal portion is lethal (58, 59).

Unlike Pol $\delta$, Pol $\epsilon$ is highly processive without the need for the processivity factor PCNA (60). The crystal structure of the *S. cerevisiae* POLE catalytic subunit

revealed a domain that is absent in POLD1 (61). This domain has been termed the "P-domain" and has been hypothesised to enhance polymerase processivity by allowing *POLE* to encircle nascent double-stranded (ds)DNA. Deletion of this domain severely impaired polymerase processivity in cell-free assays (61). Pols $\epsilon$ and $\delta$ have also been implicated in nucleotide excision repair (62, 63, 64), base-excision repair (65, 66, 67), mismatch repair (32, 68, 69) and break-induced repair (BIR) (70, 71) by either excising the mismatch or DNA damage using exonuclease activity or by gap-filling after the DNA damage is removed. Finally Pol $\epsilon$ is involved in cell cycle checkpoint regulation and chromatin remodelling (59, 72, 73).

| Family | Polymerase | *Homo sapiens* | *Saccharomyces cerevisiae* | Accessory Subunits | Function |
|---|---|---|---|---|---|
| A | $\gamma$ | *POLG* | *MIP1* | p55 | Mitochondrial replication and repair |
| | $\theta$ | *POLQ* | - | - | DNA repair |
| | $\nu$ | *POLN* | - | - | ICL repair |
| B | $\alpha$ | *POLA* | *POL1 (CDC17)* | p67, p58, p49 (primase) | DNA replication, DSBR, HR, cell cycle regulation |
| | $\delta$ | *POLD1* | *POL3 (CDC3)* | p12, p50, p68 | DNA replication, repair: MMR, BER, NER, DSBR |
| | $\epsilon$ | *POLE* | *POL2* | p17, p12, p59 | DNA replication, DNA repair: BER, NER, DSBR, HR, S-phase checkpoint |
| | $\zeta$ | *POLZ (REV3)* | *REV3* | Rev7, p50, p68 | TLS, ICL repair, somatic hypermutation |
| X | $\beta$ | *POLB* | - | - | BER, DSBR, meiosis |
| | $\lambda$ | *POLL* | *POLIV (POLX)* | - | DSBR, NHEJ, V(D)J recombination, BER |
| | $\mu$ | *POLM* | - | - | DSBR, NHEJ, V(D)J recombination |
| | TdT | *TdT* | - | - | V(D)J recombination |
| Y | $\eta$ | *POLH* | *RAD30* | - | TLS, somatic hypermutation, HR |
| | $\kappa$ | *POLK (DINB)* | - | - | TLS, NER |
| | $\iota$ | *POLI (RAD30B)* | - | - | TLS |
| | Rev1 | *REV1* | *REV1* | - | Abasic site synthesis |
| Archaeo-eukaryotic primase | PrimPol | *PRIMPOL* | | | de-novo DNA priming and DNA replication, TLS |

**Table 1.3:** DNA polymerases and their function in eukaryotic cells. TLS, Translesion synthesis; DSBR, Double-strand break repair; HR, Homologous recombination; ICL, interstrand crosslink; NHEJ, Non-homologous end joining.

### 1.3.3 Accessory subunits

POLD1, POLD2 and POLD3 (Pol3, Pol31 and Pol32 in *S. cerevisiae*) interact with the processivity factor PCNA to stimulate DNA replication (74, 75, 76, 77) and are sufficient to drive DNA replication *in vitro* (76). In recent years, the accessory subunits have also been shown to play an important role in DNA damage response. The smallest subunit, POLD4, was shown to be degraded in response to DNA damage, leaving a $\delta^3$ complex (78). Although Pol $\delta^3$ has reduced processivity, it has increased discrimination against mutagenic bypass of 8-oxoG and other small base-adducts *in vitro* (79), thus reducing mutagenesis under genotoxic stress. *S. cerevisiae* Pol31 and Pol32 are also accessory subunits of the B-family translesion synthesis polymerase, Polymerase $\zeta$ (80). Deletion of Pol32 leads to DNA damage hypersensitivity, consistent with a role in DNA repair and/or lesion bypass. Germline *POLD3* single nucleotide polymorphisms (SNPs) are associated with an increased risk of developing

colorectal cancer in humans (81) and POLD3 depletion is known to cause a high frequency of genome duplication due to chromosomal instability (70, 82). Finally, reduced expression of POLD4 may play a role in genomic instability in lung cancer (83).

The second subunit of Pol $\epsilon$, POLE2 (Dpb2 in *S. cerevisiae*) is essential to cell viability. Studies in budding yeast show that Dpb2 structurally connects the yeast POLE homologue, Pol2 to GINS, a component of CMG helicase, to drive replication (84, 85, 86, 87). It has been demonstrated that mutations to the cysteine-rich inter-zinc amino acids in the C-terminus of Pol2 reduced DNA replication, increasing sensitivity to the DNA damaging agent methyl methanesulfonate (MMS) (88). It is possible that these mutations reduce zinc binding and weaken dimerisation of Dpb2 to Pol2 (89). The non-essential third and fourth subunit increase binding of the polymerase to dsDNA, increasing polymerase processivity and 3'-5' exonuclease DNA degradation (45, 90). The smallest subunit, POLE4, was shown to be important for S-phase progression in Drosophila and *Pole4*$^{-/-}$ mice have impaired development and reduced lymphocyte maturation (91, 92).

### 1.3.4 Strand-specificity of leading and lagging strand polymerases

The study of error-prone Pols $\epsilon$ and $\delta$ in *S. cerevisiae* and *Schizosaccharomyces pombe* support a model where Pol $\epsilon$ replicates the leading strand and Pol $\delta$ replicates the lagging strand. The first evidence that these two polymerases replicate alternate DNA strands came from a study in 3'-5' polymerase proofreading-deficient *S. cerevisiae*, where the mutation spectrum of the reporter gene *URA3*, placed either side of a defined origin, displayed a strand-specific mutation spectrum (93). Mutations in the polymerase domain of *S. cerevisiae* Pol $\epsilon$ and Pol $\delta$ (*pol2-M644G* and *pol3-L612M* respectively) exhibited a biased mutation spectrum towards G:A→C:T mismatches in *pol3-L612M* mutants and T:A→A:T in *pol2-M644G* mutants in an $\triangle$*msh2* deficient background (94, 95). The mutation spectra at the *URA3* locus was used to assign Pol $\epsilon$ and Pol $\delta$ in replicating the leading and lagging strand respectively (96, 97). These data were supported by the strand-specific bias of excess ribonucleotides incorporated

by these mutagenic polymerases, in an RNase H2-deficient background, to minimise ribonucleotide repair (98, 99).

Studies in reporter genes only look at strand specificity near one origin at a single locus, surveying less than 0.01% of the yeast genome. Yeast have well-mapped origins of replication, meaning that next generation sequencing can be used to sequence the entire yeast genome for strand-specific base-pair substitutions and for ribonucleotide incorporation (100, 101, 102, 103, 104). Deep sequencing studies confirm that in most part, Pol $\epsilon$ does replicate the leading strand and Pol $\delta$ replicates the lagging strand (103, 104). There was, however, subtle variation in the strand-specificity of Pol $\delta$, where it occasionally synthesised the leading strand in unperturbed cells, particularly near origins of replication (101, 102). *S. cerevisiae* containing catalytically dead Pol $\epsilon$ plus mutagenic *pol3-L612M* were viable, slow-growing and had aberrant nuclei, but had a reduced ribonucleotide strand bias (102). This shows that Pol $\delta$ is able to replicate the leading and lagging strand in the absence of Pol $\epsilon$ catalytic activity, albeit less efficiently than Pol $\epsilon$.

There is growing evidence that in certain circumstances, such as fork restart after replication stress or during DNA repair, Pol $\delta$ can replicate the leading strand (105). This is particularly pertinent, as in 2015, a publication challenging the role of Pol $\epsilon$ in leading replication was published by Johnson *et al.* where the authors provided evidence that Pol $\delta$ replicates both the leading and lagging strand and the role of Pol $\epsilon$ is to proofread errors made by Pol $\delta$ (106). The authors suggested that previous results from polymerase mutant yeast resulted from differential mismatch repair of nucleotide substitutions on the leading and lagging strands. It has been demonstrated that in budding yeast, however, Pol $\epsilon$ cannot proofread errors made by Pol $\delta$, but Pol $\delta$ can proofread errors on both strands (107) and does not, therefore, support the hypothesis that Pol $\epsilon$ is merely a proofreading polymerase for Pol $\delta$.

Several recent kinetic and structural studies have directed Pol $\epsilon$ towards a role in leading strand replication. Biochemical and structural studies show that Pol $\epsilon$ directly associates with the CMG helicase, targeting it to the leading stand during DNA replication (108, 109, 110) and if Pol $\delta$ binds first, Pol $\epsilon$ quickly replaces it (108). Importantly, Pol $\epsilon$ can replicate DNA without the need for a the PCNA processivity factor, using its intrinsic processivity domain, unlike Pol $\delta$, making it ideally suited

for continuous replication on the leading strand (60, 61, 74, 111).

## 1.4 Polymerase proofreading

The principle replicative DNA polymerases $\epsilon$, $\delta$ and $\gamma$ replicate the bulk of nuclear and mitochondrial DNA and contain intrinsic 3'→5' exonuclease activity. Nucleotide selectivity has an error rate of $10^{-4}$-$10^{-5}$ depending on the polymerase (112) and polymerase proofreading increases replication fidelity by around 100-1000-fold (113, 114, 115). The exonuclease domain of Pol $\epsilon$ and $\delta$ contains five conserved motifs, ExoI, ExoII, ExoIII, ExoIV and ExoV (116) and have 23% sequence identity and 37% similarity (117) (Figure 1.2). ExoI contains two highly conserved amino acids (D316 and E318 in *POLD1* and D275 and E277 in *POLE*) which are essential for exonuclease activity and form the active site (Figure 1.3). These amino acids coordinate two divalent metal ions that catalyse excision of the 3'-terminal nucleotide (118, 119). ExoII and III each contain a single conserved aspartate residue that coordinates $Mg^{2+}$ residues that are also involved in catalysis (53, 120) (Figure 1.2).

Misincorporation of a nucleotide at the nascent strand terminus results in helix distortion and polymerase pausing. The polymerase and exonuclease domain are separated by approximately 30-40Å, so the primer must separate from the polymerase active site and translocate the 3' end to the exonuclease active site (118). Cryo-electron microscopy (Cryo-EM) and nuclear magnetic resonance (NMR) analysis of the *Escherichia coli* replisome revealed that the mismatch distorts the DNA and causes fraying of the newly synthesised double helix, which enables the primer terminus to reach the exonuclease active site, initiating excision of the mismatch. Polymerase to exonuclease switching is a continuous process without dissociation from the replication fork, reducing the opportunity for nuclear degradation by DNA endonucleases (121).

Approximately 1.5% of the genome is replicated by Pol $\alpha$-primase, which contains no intrinsic exonuclease activity, with a fidelity of $10^{-4}$. Mismatched bases introduced by Pol $\alpha$ have been shown to be excised by Pol $\delta$ but not Pol $\epsilon$ (124) as inactivation of the Pol $\delta$ exonuclease domain increases the mutation frequency of a mutagenic Pol $\alpha$ synergistically. This supports the hypothesis of Pol $\delta$ as the lagging strand

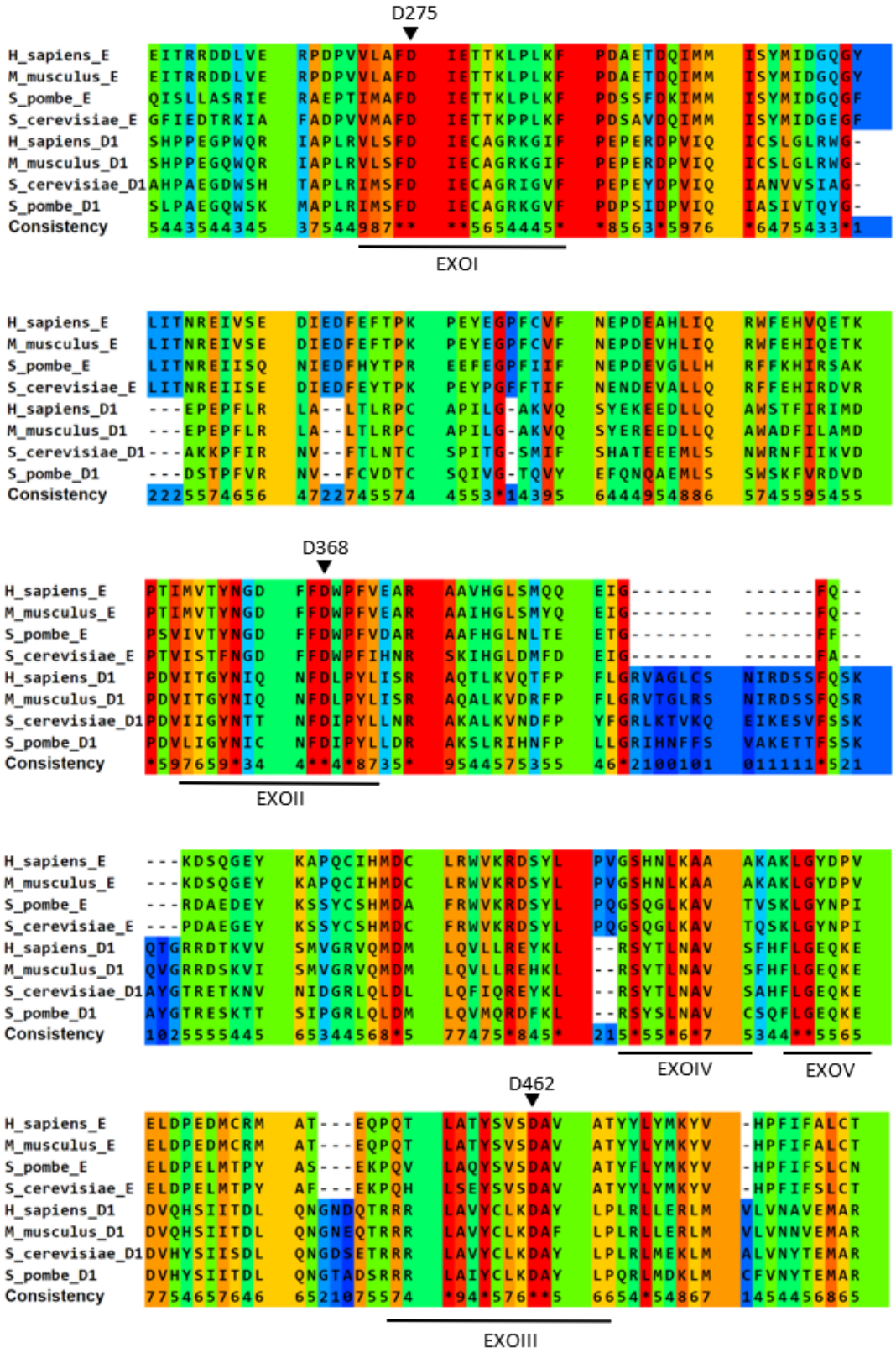



**Figure 1.2:** PRALINE (122) multiple alignment of POLE and POLD1 exonuclease domains. Exo I, II and III active site aspartate residues are labelled.

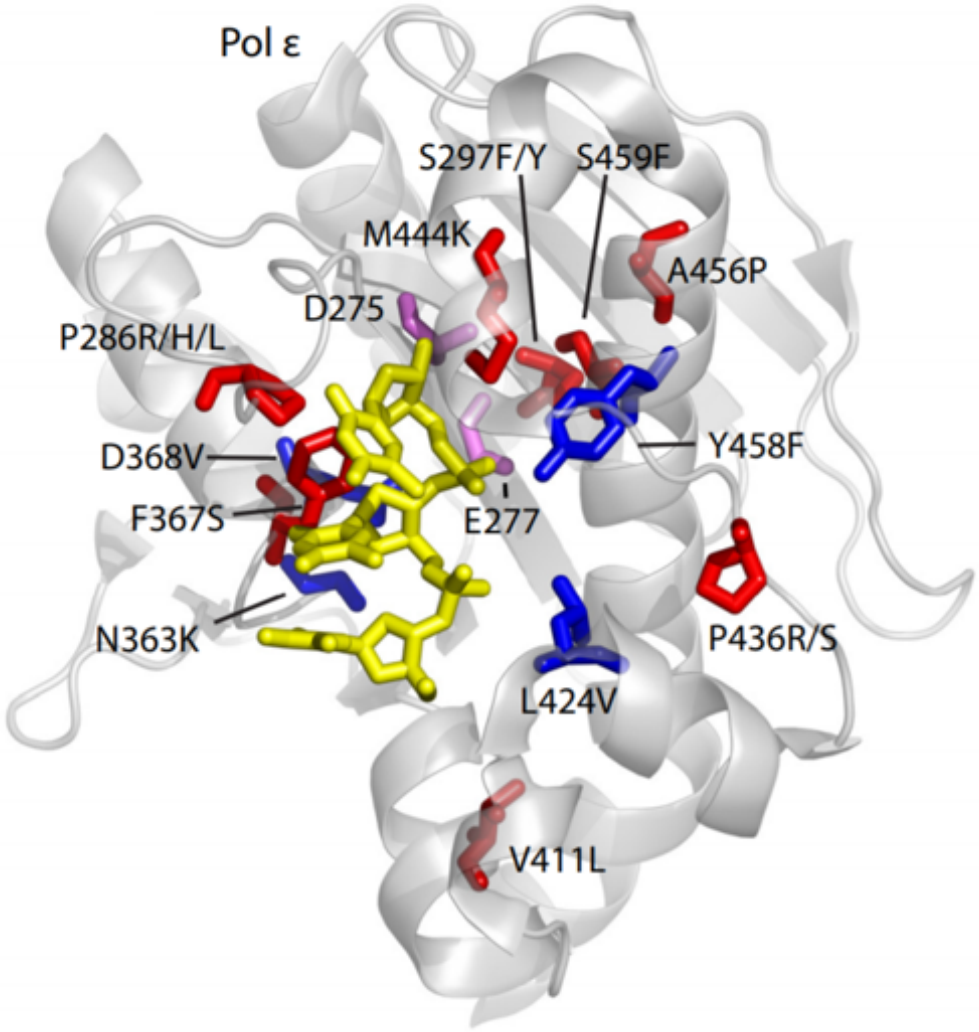



**Figure 1.3:** Pymol structure of the *Saccharomyces cerevisiae* Pol2 (analogous to POLE) N-terminal domain (Protein Data Bank identifier: PBD ID 4M8O50). Somatic exonuclease domain mutations are highlighted in red and germline mutations are highlighted in blue. Active site residues D275/E277 are highlighted in purple. Single-stranded DNA from the aligned bacteriophage T4 polymerase complex (PDB ID 1NOY117) is shown in yellow. POLE exonuclease domain mutations L424V and P436R and V411L, which are both somatic and germline are coloured according to which is more frequent (See Figure 1.4 for mutation frequency plots). Figure from Rayner *et al.* (123).

polymerase as errors are corrected during Okazaki maturation.

### 1.4.1 Polymerase exonuclease domain mutations in yeast

Mutations inactivating the exonuclease active site of *S. cerevisiae POLE* and *POLD1* (*pol2-4* and *pol3-01* respectively), result in a significant increase in mutation rate of 10 and 100-fold respectively (54, 125, 126). These data suggest that Pol$\delta$ may play a more important role in DNA proofreading than Pol $\epsilon$. This difference in mutation rate between *Pol* $\epsilon$ and Pol $\delta$ was demonstrated to be balanced by increased mismatch repair activity on the lagging strand (127). As a result, partial loss of MMR activity due to *MSH6* or *PMS1* deletion is lethal in *POL3-01* strains, whilst only causing attenuated growth in *pol2-4* strains (128, 129). Haploid *pol2-4/pol3-01* strains exhibit error-induced extinction (EEX), but diploids are viable and show a synergistic increase in mutation rate of approximately 2000-fold (93). Such synergy is also observed in *pms1* knockout haploid *pol2-4* and *pol3-01* mutants, suggesting that the mutation rate of exonuclease deficient yeast may well be higher than predicted as some errors are repaired by post-replicative MMR. Such a catastrophic increase in mutation rate could result in error-induced extinction, as was observed in polymerase exonuclease domain mutant strains crossed with *msh2*$\Delta$ and *mlh1*$\Delta$ strains with mismatch repair deficiency (128).

The mutation spectrum differs between Pols $\epsilon$ and $\delta$ exonuclease domain mutant strains. *pol2-4* mutants harbour mostly single base substitutions whilst *pol3-01* have an equal number of single base substitutions and frameshift mutations (130, 131). The *in vivo pol2-4* mutation spectrum consisted predominantly of C$\rightarrow$A transversions and C$\rightarrow$T transitions (37% of total mutations) (132).

In reality, due to the multifunctional role of DNA polymerases, the association with exonuclease active site mutations with an increased mutation rate is likely to be more complex than just loss of exonuclease activity alone. For example, studies in *S. cerevisiae* show that the mutator phenotype observed in *pol2-4* exonuclease deficient cells is abolished by the deletion of the S-phase checkpoint kinase *DUN1* that modulates cellular dNTP levels (133, 134). Furthermore, *S. cerevisiae* exonuclease deficient Pol $\epsilon$ holoenzyme displays increased strand-displacement synthesis (displacement of

newly synthesised dsDNA) *in vitro* (135), resulting potentially in genomic instability when Pol $\epsilon$ comes into contact with an oncoming DNA fork or during DNA repair.

### 1.4.2 Polymerase exonuclease domain mutations in mice

Mice are important model systems to study the impact of DNA repair defects on mutagenesis and cancer (29, 136). Mouse models harbouring homozygous alanine substitutions at the Pold1 (*Pold1*$^{exo/exo}$) and Pole (*Pole*$^{exo/exo}$) exonuclease active site, abolishing exonuclease activity, were tumour-prone and had an elevated mutation rate (137, 138, 139, 140). Interestingly, tumours only developed in homozygous exonuclease-deficient mice, despite an elevation in the mutation rate in tissue samples and mouse embryonic fibroblasts (MEFs) from heterozygous mice. Mutation rate does not appear to correlate with survival, as *Pole*$^{exo/exo}$ mice developed tumours later than *Pold1*$^{exo/exo}$ mice despite *Pole*$^{exo/exo}$ mice having the higher mutation rate (139). These results are also in contrast to those obtained from exonuclease-deficient budding yeast, where *pol3-01* strains have a higher mutation rate than *pol2-4* strains (93).

A study of the germline mutation rate in *Pold1* exonuclease deficient (exo-null) mice again showed an increase compared to wildtype controls. In addition, *Pold1* exo-null mice developed abnormal phenotypes throughout generations, such as shortened limbs, hydrocephaly, weight loss, decreased reproductive capacity and higher rate of postnatal death (141). The accumulation of mutations in the germline throughout generations, may have a detrimental impact to the long term survival of the colony although this has not been validated. Mismatch repair deficient (*Mlh1*$^{\Delta/\Delta}$) mice with loss of Pol $\epsilon$ or Pol $\delta$ exonuclease activity exhibit embryonic lethality. These data are consistent with the observation in budding yeast that loss of polymerase exonuclease activity and post-replicative MMR by *mlh1* or *msh2* deletion exceeds an error threshold required for survival (128, 139).

In addition to the variation in mutation rate and survival, the tumour spectrum differed markedly between exonuclease-deficient *Pold1* and *Pole* mice, with *Pold1*$^{exo/exo}$ mice developing primarily thymic lymphomas and carcinomas of the tail skin and lung and over 50% of *Pole*$^{exo/exo}$ mice developing intestinal adenomas and

carcinomas in addition to hystocytic sarcomas and nodal lymphoma. This difference in tumour spectrum may be explained by the differential repair of lesions by polymerases, for example in *S. cerevisiae*, Pol $\delta$ was shown to proofread errors made by Pol $\epsilon$ during TLS of thymine dimers (142), a common lesion produced by UV radiation. This may explain the increased susceptibility to skin cancer in *Pold1* exonuclease deficient mice. Mice with both *Pold1* and *Pole* exonuclease deficiency were viable and fertile but did show a synergistic decrease in survival compared to *Pold1* or *Pole* exo-null mice (139).

## 1.5 Polymerase exonuclease domain mutations in cancer

Despite evidence that Pol $\epsilon$ and $\delta$ exonuclease domain mutations are associated with an increased mutation rate and tumourigenesis in mice (139), evidence that polymerase exonuclease domain mutations (EDMs) were associated with human malignancy was lacking. The first evidence that *POLE* EDMs were associated with cancer was the discovery of the *POLE* exonuclease domain mutation, F367S, in a rectal tumour by Yoshida *et al.* (143). The authors did not, however, validate its pathogenicity or perform functional analysis.

In 2012, whole genome sequencing (WGS) and linkage analysis of highly selected patients with a family history of multiple colorectal adenomas and carcinomas, but with no known mutations in any cancer predisposition genes, were found to harbour germline mutations in the exonuclease domains of *POLE* (L424V) and *POLD1* (S478N) (Table 1.4) (144). Germline *POLE* and *POLD1* exonuclease domain mutations were associated with a dominant inheritance pattern and carriers had high risk of developing multiple colorectal polyps and adenomas in addition to early-onset CRC. Consequently, the disease was called polymerase proofreading-associated polyposis (PPAP) (117). Interestingly $POLD1^{S478N}$ was also shown to predispose patients to endometrial cancer and brain tumours (144).

*POLE* and *POLD1* exonuclease domain mutant tumours were ultramutated, microsatellite stable and showed no evidence of secondary inactivating mutations or loss

of heterozygosity (LOH). Since the initial report in 2012, several other studies have sequenced unrelated families and discovered germline *POLE*$^{L424V}$ (145, 146, 147, 148) and *POLD1*$^{S478N}$ (146) mutations. An additional seven *POLE* and five *POLD1* germline variants have been reported to date, with most variants co-segregating with affection status (Table 1.4) (149, 150, 151, 152, 153, 154, 155).

Shortly after the association of germline *POLE* and *POLD1* EDMs with ultramuted CRC by Palles *et al.*, whole exome sequencing (WES) of 224 sporadic colorectal carcinomas by the TCGA discovered a subset of ultramutated tumours (>100 mutations/Mb) harbouring heterozygous somatic missense mutations in the exonuclease domain of *POLE* (3% of the total cohort) (8). In parallel, another CRC WES project reported 2/74 (3%) ultramutated, MSS cases with *POLE* EDMs (156). The following year, *POLE* EDMs were discovered in 7-12% of ultramutated, MSS endometrial cancers (157, 158). Recurrent amino acid changes included p.Pro286Arg/His (*POLE*$^{P286R/H}$), p.Val411Leu (*POLE*$^{V411L}$) and p.Ser459Phe (*POLE*$^{S459F}$) (Table 1.5). The mutated *POLE* residues in CRC overlap with those in EC, however changes such as p.Ser297Phe (*POLE*$^{S297F}$) are enriched in EC (Table 1.5).

The somatic and germline *POLE* and *POLD1* mutations listed in Tables 1.5 and 1.4 are located at highly conserved residues either in or just outside the exonuclease motifs. When mapped to the crystal structures of *S. cerevisiae* Pol2 and Pol3, they are situated within the DNA binding cleft and are predicted to reduce exonuclease activity (Figure 1.3) (144). The exception is the germline and somatic variant, *POLE* V411L, which lies outside of the DNA binding cleft (123). Fluctuation analysis of selected germline variants in budding and fission yeast show that the exonuclease domain variants increase the mutation rate compared to wildtype strains (144, 149, 153, 159) (Table 1.4). The most abundant germline variant, *POLE*$^{L424V}$ has been reported in over 20 families to date and is also found both as a *de novo* germline variant and in sporadic cancers, suggesting that it is a mutation hotspot. Barbari *et al.* hypothesised that the GC-rich sequence surrounding the the L424 codon may make the DNA more susceptible to $\beta$ hairpin loop formation, which is known to be poorly replicated by Pols $\epsilon$ and $\delta$ (197).

Other recurrent mutations include *POLD1*$^{S478N}$ and *POLD1*$^{L474P}$ (144, 146, 147,

| Protein change (nucleotide substitution) | Evidence in support of pathogenicity | | | | | Number of carriers[±] | Mean age at diagnosis (range)[#] | Percentage of variant carriers with indicated phenotype | | | | | | | | Other cancers reported in carriers | Refs |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Structural mapping | PhyloP score* | Yeast fluctuation assay | Biochemical evidence for functional effect[‡] | Segregates with affection status | | | CRC | EC | BrC | DuC | OC | GBM | Duodenal A or P | Colonic A or P | | |
| **Pol$\delta$** | | | | | | | | | | | | | | | | | |
| D316G (947A→G) | Exo I motif active site | 1.93 | Yes(159) | Yes(160) | Yes | 2 | 51 (4457) | 50 | 100 | 50 | 0 | 0 | 0 | 0 | 50 | 0 | (150) |
| D316H (946G→C) | Exo I motif active site | 1.16 | Yes(159) | Yes(160) | Yes | 2 | 61 (5864) | 50 | 0 | 50 | 0 | 0 | 0 | 0 | 100 | Mesothelioma | (150) |
| P327L (981C→G) | Flanking exo I motif | 2.16 | Yes(161)[∓] | Yes(162)[∓] | NR | 1 | 70 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | (144) |
| R409W (1225C→T) | Flanking exo II motif | 2.26 | NR | NR | NR | 1 | 32 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | (150) |
| L474P (1421T→C) | Exo IV motif | 1.92 | Yes(159) | Yes(162)[∓] | Yes | 19 | 40 (23-52) | 38 | 54 | 8 | 0 | 0 | 0 | 0 | 30 | GIST | (147, 150, 151) |
| S478N (1433G→A) | Exo IV motif | 1.19 | Yes(144) | NR | Yes | 11 | 36 (26-52) | 45 | 36 | 0 | 0 | 0 | 0 | 0 | 91 | Anaplastic astrocytoma | (144, 146) |
| **Pol$\epsilon$** | | | | | | | | | | | | | | | | | |
| W347C (1041G→T) | Outside exo motifs | 2.74 | Yes(149) | NR | No | 12 | 49 (14-74) | 0 | 0 | 0 | 0 | 0 | 0 | NR | NR | PrC, and cutaneous and uveal melanoma | (149) |
| N363K (1089C→A) | Exo II motif active site | 4.9 | NR | NR | Yes | 12 | 41 (28-56) | 75 | 17 | 0 | 0 | 25 | 0 | NR | 83 | PAAD | (154) |
| D368V (1103A→T) | Exo II motif active site | 5.13 | NR | Yes(160) | NR | 1 | 47 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | (146) |
| V411L (1231G→C/T) | Flanking exo IV motif | 2.66 | Yes(163) | Yes(162) | *De novo* | 1 | 14 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | (155) |
| L424V (1270C→G) | Exo IV motif active site | 2.66 | Yes(159) | Yes(162) | Yes. 2*de novo* carriers | 48 | 39 (15-60) | 61 | 2 | 2 | 2 | 2 | 4 | 19 | 92 | ODG and neuroendocrine carcinoma | (144, 145, 146, 147) |
| P436S (1306C→T) | Within exo IV motif | 3.53 | NR | NR | *de novo* | 1 | 31 | 100 | 0 | 0 | 0 | 0 | 0 | 100 | 100 | 0 | (145) |
| Y458F (1373A→T) | Exo III motif active site | 4.97 | NR | Yes(160) | Yes | 14 | 48 (38-63) | 62 | 0 | 0 | 15 | 8 | 0 | NR | 62 | PAAD | (152) |
| V474I (1420G→A) | Outside exo motifs | 6.03 | NR | NR | NR | 1 | 47 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | (153) | |

**Table 1.4:** Germline cancer-associated *POLE* and *POLD1* exonuclease domain mutations. A or P, adenomas or polyps; BrC, breast cancer; CRC, colorectal cancer; DuC, duodenal carcinoma; EC, endometrial cancer; GBM, glioblastoma; NR, not reported; OC, ovarian cancer; ODG, oligodendroglioma; PAAD, pancreatic adenocarcinoma; PrC, prostate cancer. *PhyloP (phylogenetic conservation) scores were calculated per nucleotide using alignment of 46 vertebrates dbNSFPv23. If a variant mapped to the third position of a codon the average PhyloP score for the codon is displayed. [‡]Data from functional studies of B family polymerases. [±]Reported as of September 2018. [#]Mean age at diagnosis in years refers to cancer or adenoma diagnosis, whichever was earliest. [∓]Functional studies of the corresponding residue in Pol $\epsilon$.

| Protein change (Nucleotide substitution) | Evidence in support of pathogenicity | | | | Molecular characteristics* | | | Tumour type(s) & references |
|---|---|---|---|---|---|---|---|---|
| | Structural mapping | PhyloP score‡ | Yeast fluctuation assay | Biochemical evidence ± for functional effect | Median number of mutations/exome (range) | % MSS | % tumours with ≥20% C→A substitutions | |
| **Polδ** | | | | | | | | |
| D316G (947A→G) | Exo I motif active site | 1.93 | Yes(159) | Yes(160) | 4,346 | NR | NR | EC(14) |
| C319Y (956G→A) | Within exo I motif | 2.38 | Yes(159) | NR | 7,349 (41-14,654) | 100 | NR | GBM(164), MM(14) |
| S478N (1443G→A) | Exo IV motif | 1.19 | Yes(144) | NR | 5,699 | 100 | NR | CRC(165), EC(14) |
| **Polε** | | | | | | | | |
| P286R/H/L (857C→G/A/T) | Flanking exo I motif | 2.57 | Yes(161, 163) | Yes(162) | 5,399 (735-14,323) | 98 | 100 | CRC(8, 14, 156, 161, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174), EC(14, 157, 158, 171, 171, 172, 174, 175, 176, 177, 178, 179, 180), GBM(171, 181), EOC(171, 182), BrC(14), PAAD(14, 171), BLCA(174, 183), PrC(174) |
| S297F/Y (890C→T/A) | Flanking exo I motif | 2.67 | NR | NR | 6,472 (1,518-12,071) | 83 | 86 | CRC(14, 166, 169, 184), EC(14, 158, 171, 172, 177, 178, 179, 180), GBM(164, 171, 181), AA(171), EOC(185), CC(14), Melanoma(171) |
| E311D (931G→T) | Flanking Exo II motif | 6.17 | NR | NR | 248 | NR | 100 | CRC(166), BLCA(14) |
| F367S (1100T→C) | Exo II active site | 2.19 | Yes(163) | Yes(160, 162) | 6,872 (2,919-10,825) | 100 | 100 | CRC(8, 143, 166), EC(14) |
| V411L (1231G→C/T) | Flanking exo IV motif | 2.66 | Yes(163) | Yes(162) | 6,834 (957-14,074) | 100 | 92 | CRC(8, 165, 166, 168, 171, 173, 174, 186), EC(14, 77, 157, 158, 172, 174, 175, 176, 177, 178, 179, 180, 187, 188, 189, 190), GBM(14, 171, 181), EOC(171, 182), GC(191, 192), Medulloblastoma(193), BLCA(173, 174), PrC(171, 174), Penile(174), CC(14), KC(171) |
| L424V/I (1270C→G/T) | Exo IV active site | 2.66 | Yes(159, 163) | Yes(160, 162) | 6,476 (162-12,218) | 75 | 100 | CRC(194), EC(157, 195), BrC(14), Lung(171, 174), GBM(171), AA(196) Liver(171) |
| P436R/H (1307C→G/1307C→A) | Exo V motif | 3.53‡ | Yes(163) | Yes(160) | 7,290 (6,616-11,902) | 100 | 100 | CRC(8, 166, 171), EC(14, 177, 188), GBM(164, 171, 196), BLCA(171) |
| M444K (1331T→A) | Flanking exo V motif | 2.15 | NR | NR | 1,202 | 100 | 100 | CRC(171), EC(157, 187), GBM(171) |
| A456P/V (1366G→C/1367C→T) | Within exo III motif | 2.61 | NR | NR | 4,528 (1,326-6,910) | 100 | 100 | CRC(14, 165, 166, 171, 186, 194), EC(157, 158, 171, 177, 188), GBM(14, 171), EOC(171), CC(171) |
| S459F (1367C→T) | Within exo III motif | 2.52 | Yes(163) | Yes(160, 162) | 4,835 (1,737-9,943) | 86 | 100 | CRC(14, 165, 166, 171, 173, 174, 194), EC(14, 171, 187), GBM(171, 181), AA(164) |
| S461P/T/L (1381A→G/1382C→T) | Within exo III motif | 4.97 | NR | Yes(164) | 7,580 (5,490-8,670) | 100 | NR | CRC(194), EC(14), GBM(164, 171), Neuroectodermal tumour(164) |

**Table 1.5:** Somatic cancer-associated *POLE* and *POLD1* exonuclease domain mutations. AA, anaplastic astrocytoma; BLCA, bladder cancer; BrC, breast cancer; CC, squamous cell cervical carcinoma; CRC, colorectal cancer; EC, endometrial cancer; EOC, endometrioid ovarian carcinoma, GBM, glioblastoma; GC, gastric cancer; MM, multiple myeloma; PAAD, pancreatic adenocarcinoma, PrC, prostate cancer; MSS, microsatellite stable; NA, not applicable; NR, not reported. *Data from WGS, WES or whole *POLE* exonuclease domain targeted sequencing. ‡PhyloP (phylogenetic conservation) scores were calculated per nucleotide using the alignment of 46 vertebrates dbNSFPv23. For P436R the variant mapped to the third position of a codon so the average PhyloP score for the codon is displayed. ±Data from studies of B family polymerases.

150, 151). The pathogenicity of two germline variants, *POLE*$^{W347C}$ and *POLE*$^{V474I}$, are, however, uncertain. *POLE*$^{W347C}$ displays low penetrance inheritance (with six unaffected family members) and the tumour spectrum consists predominantly of cutaneous malignant melanoma and uveal malignant melanoma (45% of carriers) rather than CRC, which occurred in a wildtype control (149). *POLE*$^{V474I}$ has only been identified in one patient, albeit with a family history of colorectal polyps and carcinomas and is located at the distal end of the exonuclease domain (153). Despite this, both variants increase the mutation rate in *S. pombe* (149, 153).

Many germline EDMs map to the exonuclease domain active sites, including *POLE* L424V, N363K , D368V, Y458F and *POLD1* D316G/H and are some of the most highly penetrant *POLE* and *POLD1* variants (123). Carriers of germline *POLE* exonuclease domain mutations tend to develop CRC as well as duodenal cancer, ovarian cancer and brain cancers, whilst carriers of germline *POLD1* EDMs tend to develop CRC and EC (Table 1.4). Other tumour types observed include cancers of the breast, pancreas, prostate and skin.

Since the identification of somatic *POLE* EDMs in ultramutated CRC and EC (8, 156, 157, 158), multiple WGS, WES and targeted sequencing studies have estimated the frequency of pathogenic EDMs in these cancers to be 0.5-3% and 7-12% respectively (166, 171, 177, 179, 187, 198). The tumour spectrum associated with somatic *POLE* EDMs includes brain (glioblastoma and anaplastic astrocytoma), ovarian, pancreatic, stomach, breast, kidney, bladder and prostate cancer amongst others (14, 162, 164, 173, 174, 176, 181, 185) (Table 1.5). Just under half of all *POLE* EDMs are P286R/H/L and V411L mutations (Figure 1.4 and Table 1.5), followed by recurrent S297F, A456P and S459F mutations (162). *POLE* V411L is found in the widest variety of cancer subtypes, ranging from glioblastoma to ovarian, gastric and kidney cancer, medullablastoma to penile cancer. Some variants, including V411L, L424V and P436S are found both in hereditary and sporadic cancers (Figure 1.4).

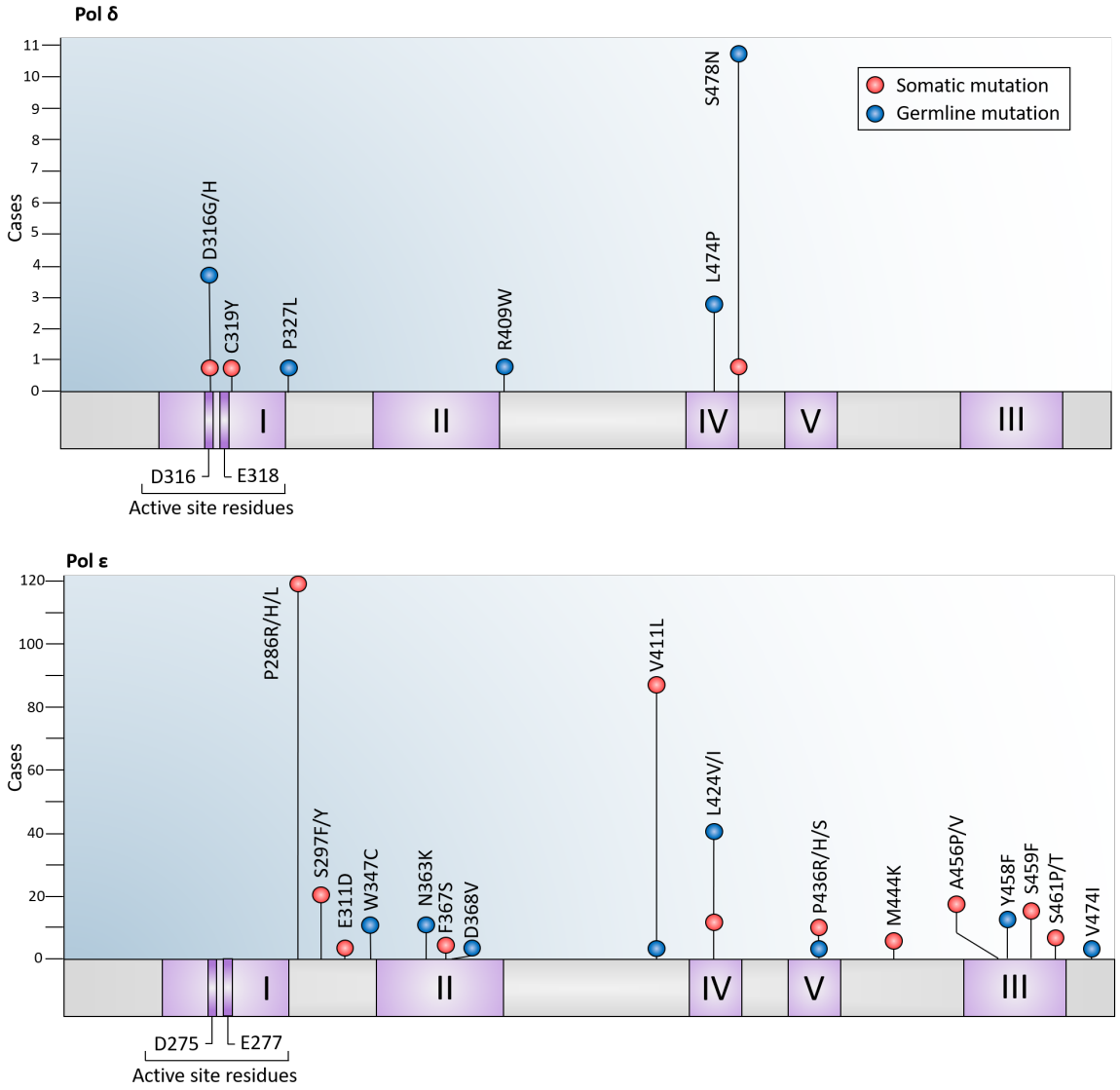



**Figure 1.4:** Lollipop plot of cancer-associated mutations in the exonuclease domains of POLD1 and POLE, adapted from Rayner *et al.* (123). Schematics include conserved exonuclease motifs (I to V), exonuclease active site residues, and the frequency of somatic (red) and germline (blue) exonuclease domain mutations from sample where the entire exonuclease domain has been sequenced.

Interestingly, cancer-associated *POLD1* exonuclease domain mutations are rarely somatic, with only five reported cases (three different variants) found to date (Table 1.5). Somatic *POLD1* C319Y has been discovered in two tumours. The first tumour was from a patient who was diagnosed with an ultramutated MSS glioblastoma at 12.5 years, who was found to have constitutional biallelic mismatch repair deficiency (CMMRD) due to a compound heterozygous mutation in the mismatch repair gene

*MSH6* (c.3984_3987dupGTCA/c.3959_3962delCAAG) (164). C319Y has also been identified in a multiple myeloma, however this tumour was not ultramutated, with just 41 mutations detected in the exome (199). Interestingly, the two other somatic *POLD1* variants, D316G and S478N, were observed in an ultramutated CRC and EC respectively are both known to be pathogenic in the germline (14, 165) (Table 1.5). An additional substitution at D316 (D316N) was discovered in a highly mutated gastric cancer (192) and two ultramutated endometrial cancers (14). The gastric cancer has a high number of insertions and deletions and the variant allele frequency (VAF) of the *POLD1* variant was 0.12, suggesting that it is sub-clonal. The VAF of the D316N allele in the two endometrial cancers was approximately 0.4, however one of the endometrial cancers harboured the *POLE* V411L hotspot mutation, potentially explaining the ultramutator phenotype. As a result, the pathogenicity of the *POLD1* D316N variant is unclear.

This section has highlighted the association of several *POLE* and *POLD1* exonuclease domain mutations with the onset of ultramutated cancer. The mechanism by which these mutations affect polymerase proofreading and how this leads to cancer, particularly colorectal and endometrial cancer, is poorly understood. In the next section, I will review the current literature on the functional consequences of *POLE* exonuclease domain mutations.

### 1.5.1 Functionality of polymerase exonuclease domain mutations

The functional consequences of cancer-associated *POLE* exonuclease domain mutations have been investigated using a variety of complementary experiments in model organisms and in cell-free assays. The purified N-terminal catalytic subunit of hPOLE (which has been shown to have comparable catalytic activity to the Pol $\epsilon$ yeast holoenzyme (94, 200)) was used to compare the relative exonuclease activity of seven *POLE* EDMs (P286R/H, F367S, V411L, L424V/I and S459F) *in vitro*. POLE$^{\mathrm{P286R/H}}$ and POLE$^{\mathrm{S459F}}$ substantially reduced, but did not abolish, exonuclease activity and the remaining four *POLE* variants had variable degrees of exonuclease activity from 17-42% (162, 200). Despite *POLE*$^{\mathrm{V411L}}$ being the second most recurrent cancer-associated

*POLE* EDM (Figure 1.4), POLE$^{V411L}$ reduced the exonuclease activity by 3-fold, compared to 10-fold with POLE$^{P286R}$ and 20-fold with POLE$^{S459F}$ (162). Variants at amino acid POLE$^{L424}$ are associated with both somatic (L424V/I) and germline (L424V) cancers (123). Although POLE$^{L424V}$ is more recurrent in cancer, it retains 34% exonuclease activity compared to 17% with POLE$^{L424I}$ (162, 171). The specific amino acid change may differentially affect the exonuclease activity, so caution should be exercised when categorising a variant as pathogenic based solely on the specific codon. It is also possible that variants that have a milder functional impact on exonuclease activity may be more likely to be tolerated in the germline (discussed in Section 1.5.4).

In addition to a reduction in exonuclease activity, cancer-associated *POLE* EDMs increase mutation rates. POLE P286R/H, F367S, L424V/I, S459F and S461P have all been shown to significantly increase *in vitro lacZ* mutation frequency compared to wildtype POLE (162, 164). Interestingly, despite being outside of the exonuclease motifs, haploid and diploid *S. cerevisiae* strains harbouring the equivalent mutation to the highly recurrent human POLE$^{P286R}$ had an exceptionally strong mutator phenotype compared to haploid and diploid *pol2-4* exonuclease deficient strains (Figure 1.5) (161).

The equivalent *POLE* S459F, F367S, P286RH, P436R and L424V mutations in haploid budding yeast show a moderately elevated mutation rate above exonuclease-null yeast, possibly due to an effect beyond loss of proofreading alone (163) (Figure 1.5). Diploid yeast, heterozygous for cancer-associated *POLE* exonuclease domain mutations, have a considerable increase in mutation rate compared to exonuclease-null strains (163). Surprisingly haploid strains, containing the *POLE*$^{V411L}$ analogue had a mutation rate similar to that of *POLE* wildtype strains despite *POLE*$^{V411L}$ being highly recurrent in ultramutated cancer (123, 163) (Table 1.5). Unlike other *POLE* exonuclease domain mutations, V411L is located outside of the DNA binding cleft and is not predicted to contact incoming DNA, therefore the mechanism of *POLE*$^{V411L}$ mutagenesis may differ from other *POLE* exonuclease domain mutants. When comparing common cancer-associated *POLE* variants there is no correlation between the reduction in *in vitro* exonuclease activity and *in vivo* mutation rates, supporting the hypothesis that pathogenicity may be driven by other functional al-

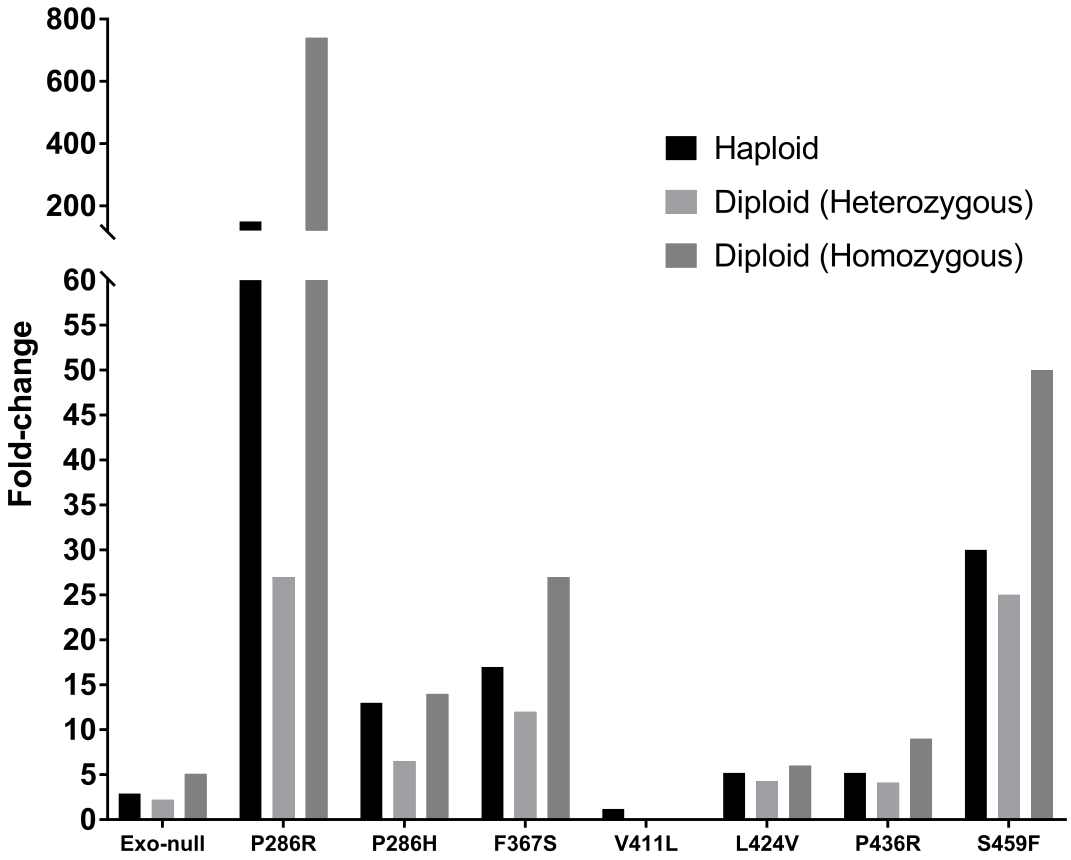



**Figure 1.5:** Fold increase in Can$^{r}$ mutation frequency in *POLE* exonuclease domain mutant haploid, diploid heterozygous and diploid homozygous *S. cerevisiae* (161, 163). No data has been published for diploid V411L strains.

terations, beyond a reduction in exonuclease activity (162, 163).

### 1.5.2 Clinicopathological characteristics of polymerase exonuclease domain mutant tumours

Most patients with germline *POLE* or *POLD1* exonuclease domain mutations typically present with multiple colorectal polyps, adenomas and carcinomas, with no evidence of upper gastrointestinal tumours (144). Carriers of *POLE* EDMs exhibit a wider tumour spectrum than *POLD1* carriers which include breast cancer, ovarian cancer, glioblastoma and pancreatic cancer (117) (Table 1.4). Disease typically presents at an early age (<50 years) in patients with PPAP and patients demonstrate

no other atypical characteristics (117, 144, 150) (with the exception of cafe-au-lait macules (CALMs) in a young patient with a *de novo* germline V411L mutation (155)). Somatic *POLE* exonuclease domain mutations are associated with ultramutation in CRC, EC, GBM and other tumour types including ovarian, lung, pancreatic and breast cancer (Table 1.5).

Patients with somatic *POLE* exonuclease domain mutant tumours are generally diagnosed at a younger age (166, 167, 172, 181, 182, 201, 202), possibly due to rapid tumour growth after *POLE* EDM acquisition (164). Colorectal carcinomas with *POLE* mutations are more frequently right sided in location and are significantly more likely to occur in men (166, 167). *POLE* exonuclease domain mutant tumours are generally high grade and present multiple bizarre multinucleated giant cells with high mitotic activity, fragmented or dispersed chromatin and severe nuclear atypia (175, 181, 201, 202, 203). Giant cells are thought to originate from repeated failure of mitosis/cytokinesis, which is paradoxical to the chromosomal stability of *POLE* exonuclease domain mutant tumours (204).

A notable characteristic of exonuclease domain mutant tumours is a high prevalence of tumour-infiltrating lymphocytes (TILs) comprising of $CD8^+$ cells and enrichment of genes encoding immunosuppressive checkpoints such as PD1 and PDL-1 (166, 181, 205). Ultramutation is predicted to increase the number of neoantigens in *POLE* mutant cancers, resulting in an anti-tumour response. The association of *POLE* exonuclease domain mutations with a good prognosis in EC was first suggested by the TCGA (157) and has since been reproduced in numerous follow-up studies (175, 177, 178, 179, 180, 188, 206, 207). A good prognosis has been suggested in CRC (166, 167) and GBM (181), although not formally confirmed. Caution must be taken, as many of these patients received adjuvant radiotherapy and/or chemotherapy therefore a differential response to post-operative treatment cannot be excluded. It is worth noting that several studies have published prognostic data in *POLE* exonuclease mutant cancers but they did not select for confirmed pathogenic *POLE* exonuclease domain mutations (194, 208, 209). Further studies are required to confirm whether there remains a good prognosis in the absence of post-operative therapy and whether this is also holds true in CRC and GBM. These results will ultimately have wider implications for patient treatment.

### 1.5.3 Molecular characteristics of polymerase exonuclease domain mutant tumours

The most striking molecular characteristic in tumours harbouring germline and somatic *POLE* EDMs is their ultramutation, often exceeding 100 mutations/mb. These mutations comprise single base substitutions, predominantly resulting in missense mutations, with relatively few frameshifts (8, 144, 156, 158, 162). Whilst all types of somatic base pair substitutions are elevated in *POLE* EDM tumours, they exhibit a highly specific mutation signature. These include a 30-fold increase in C:G→T:A transitions in the context of TCG, a 100-fold relative increase in C:G→A:T transversions in the context of TCT and a strong relative increase in T:A→G:C transversions in the context of TTT. "Signature 10" published by Alexandrov *et al.* was associated with *POLE* exonuclease domain mutant tumours (Figure 1.6) (1, 144, 158, 162).

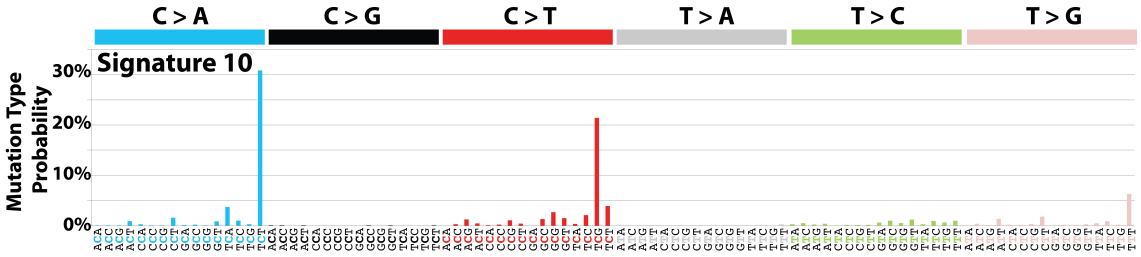



**Figure 1.6:** Probability of base pair substitutions in a 96-channel trinucleotide context in samples with *POLE* exonuclease domain mutations. This signature was categorised as Signature 10 (1). Figure from Alexandrov *et al.* (1).

The specific trinucleotide mutation context results in a bias of amino acid changes. These include a higher frequency of serine to tyrosine or isoleucine caused by TCT→TAT and TCG→TTG substitutions respectively and arginine (coded by the reverse complement to serine) to isoleucine or glutamine substitutions due to AGA→ATA and CGA→CAA mutations respectively (162). Interestingly, the most abundant amino acid substitutions found in *POLE* EDM tumours are glutamic acid to stop codons due to GAA or GAG preceded by an adenine (AGA→ATA). The over representation of specific amino acid substitutions result in recurrent hotspot mutations in cancer driver genes including: PI3K catalytic subunit-$\alpha$ (PIK3CA)-R88Q; PTEN-R130Q; adenomatous polyposis coli (APC)-R1114X and APC-Q1338X; MSH6-E946X and MSH6-E1322X; F-box and WD repeat domain-containing 7 (FBXW7)-E369X; and

p53-R213X (8, 10, 144, 156, 158, 162). Some of these mutations differ from the conventional cancer-driver mutation hotspots. For example $APC^{R1114X}$, although predicted to be functionally suboptimal, is significantly enriched in TCGA colorectal cancers with *POLE* exonuclease domain mutations. Is is plausible that these functionally sub-optimal mutations are tolerated due to the ultramutator phenotype conferred by *POLE* exonuclease domain mutations (10, 144).

Importantly, the bias of context-specific base pair substitutions and amino acid changes exhibit strand-specificity in somatic *POLE* exonuclease domain mutant tumours, consistent with a role of Pol $\epsilon$ in leading strand replication. Interestingly, tumours and budding yeast with the second most abundant *POLE* exonuclease domain mutation, V411L, exhibit a variation in mutation signature compared to other cancer-associated *POLE* EDMs, with a lower frequency of TCT→TAT mutations and a higher relative frequency of T→C mutations. These data add to the evidence that the mechanism of *POLE* V411L mutagenesis may differ from other cancer-associated *POLE* exonuclease domain mutations (158, 163, 210).

The mechanism behind the specific mutagenesis of TCT→TAT and TCG→TTG trinucleotides is currently unestablished. It has been demonstrated that germline mutations in the DNA glycosylase *MUTYH* result in an increase in somatic C→A mutations in tumours due to a lack of excision of mispaired adenine opposite 8-oxoG (Section 1.2.2). It is plausible that exonuclease domain mutant Pol $\epsilon$ or $\delta$ are more likely to insert an adenine opposite 8-oxoG, potentially overwhelming the BER pathway. The replicative polymerases Pol $\epsilon$ and $\delta$ have been shown to partake in the gap-filling stage of long-patch base-excision repair in association with PCNA (66, 211, 212, 213, 214). Exonuclease deficient POLE is able to gap-fill more efficiently than wildtype POLE, suggesting the BER DNA gap may be filled preferentially by an error-prone polymerase (215). Repeated mispairing of adenine opposite 8-oxoG by exonuclease domain mutant Pol $\epsilon$ following MUTYH excision may increase the propensity of C→A mutations in *POLE* mutant tumours. Organs with high oxidative stress due to rapid cell turnover, chronic inflammation and immune response may be the reason why *POLE* and *POLD1* exonuclease domain mutations are more prevalent in the intestine and endometrium. Additionally, transversions are more poorly repaired than transitions (127), therefore an increase in G:A mispairs may

also overwhelm the MMR system resulting in an increase in C→A mutations (162).

The second most abundant signature mutation in *POLE* exonuclease domain mutant cancers are C→T mutations flanked by T and G, which reflects a CpG island methylation-associated signature. 5-methylcytosine is deaminated more readily than unmethylated cytosine, making C→T mutations particularly abundant in cancers and inherited diseases, as well as contributing to mutation hotspots in genes such as *TP53* (26, 216). The high proportion of somatic C→T mutations in *POLE* mutant tumours may result from the ineffective repair of deaminated 5-methylcytosine during base-excision repair. It has recently been suggested that mispairing of adenine opposite methylated cytosine (due to its structural similarity to thymine) and the inability of exonuclease domain mutant Pol $\epsilon$ to recognise and/or repair the mismatch may explain the increase in C→T mutations (217). In *POLE* EDM tumours there is an enrichment of CpG→TpG on the leading strand, around origins of replication, which again suggests a role for Pol $\epsilon$ as the leading strand replicase (217). Cytosine deamination and oxidative damage are the two major types of spontaneous DNA lesion in cells, at around 100-500 events per day (26). It is plausible that if Pol $\epsilon$ and $\delta$ are involved in the repair of these lesions, that an increase in base-pair substitutions associated with these DNA damage types will increase dramatically.

Germline and sporadic *POLE* and *POLD1* exonuclease domain mutant cancers display very few copy number alterations and most are microsatellite stable despite harbouring multiple somatic mutations in MMR genes such as *MSH6* (8, 123, 144, 157, 158, 162). As has been described in previous Sections 1.4.2 and 1.4.1, complete loss of MMR with *POLE* exonuclease active site mutations are lethal in *S. cerevisiae* and mice (128, 139). A few cancers, however, have been found to have concurrent somatic mismatch repair deficiency (MMRD) and pathogenic *POLE* EDMs (148, 164, 171, 205, 210). Most cases of concurrent MMRD and *POLE* EDMs include loss of PMS2 or MSH6, which result in a milder mutator phenotype than loss of MSH2 and MLH1, due to the genetic redundancy of the MMR pathway (8, 128, 157).

A study of ultramutated paediatric malignant brain tumours with constitutional MMRD revealed a subset of cases with a secondary *POLE* exonuclease domain mutation (164). Interestingly, tumours were predominantly microsatellite stable despite showing loss of the corresponding mutant MMR gene. Therefore it is plausible that

the mechanism of *POLE* mutagenesis may be accentuated by partial loss of mismatch repair without causing microsatellite instability. The loss of MLH1, MSH2 or an MSI phenotype may result in error-induced extinction by exceeding an error threshold in human tumours, such as we observe in model organisms (128). A recent study has shown that concurrent MSI and *POLE* EDMs result in a distinct mutation signature resembling Signature 14 (1), that shares some features with *POLE* Signature 10 and defective DNA mismatch repair Signature six, but also has distinct mutations such as G[C→A]T that may normally be repaired by polymerase proofreading or mismatch repair (210). This signature distinctly resembles the the signature in CMMRD *POLE* exonuclease domain mutant brain tumours (171).

### 1.5.4 Pathogenicity of polymerase exonuclease domain mutations

The suggestion that the pathogenicity of cancer associated *POLE* and *POLD1* exonuclease domain mutations may be due to functional alterations other than, or in addition to, loss of proofreading, makes it challenging to distinguish true pathogenic variants from neutral passengers. Since the initial report of germline *POLE* and *POLD1* exonuclease domain mutations in multi-generational familial colorectal and endometrial cancer, a further 12 variants have been reported in families, in addition to recurrent cases of the same mutations in seemingly unrelated families. The following criteria have been proposed, to assign pathogencity of a germline *POLE* or *POLD1* exonuclease domain mutation: segregation of the variant with the affection status of the patient; conservation of the amino acid between POLE and POLD1; validation of an increased mutation rate or decreased exonuclease activity in model organisms and cell-free assays (162, 163); sequencing of tumours for an ultramutator phenotype (123).

The TCGA database lists 121 *POLE* and 136 *POLD1* unique somatic exonuclease domain mutations (14). The association of *POLE* exonuclease domain mutations with a good prognosis makes for accurate classification of somatic pathogenic *POLE* exonuclease domain variants critical in patient management (Section 1.5.2). Based on combined TCGA sequencing data of *POLE* exonuclease domain mutant cancers,

Shinbrot *et al.* suggested criteria for identifying pathogenic *POLE* EDMs which include: ultramutation (often exceeding 100 mutations/mb); increased propensity of C→A mutations over 20%, particularly in the context of TCT; low number of somatic indels (approximately 2%) and; residues that are recurrently mutated in cancer (162). Whilst these criteria define most pathogenic *POLE* mutations validated to date, other *POLE* EDMs may have a different mechanism of pathogenicity. New candidate polymerase exonuclease domain mutations will need to be functionally analysed on a case by case basis.

A study of over 81,000 tumours revealed that for several *POLE* cancer associated mutations, ultramutation is dependent on the specific amino acid change. For example, P286H, S459C and P436L mutations do not result in an ultramutator phenotype (171). Caution must be taken, however, as somatic L424V mutations do not result in ultramutation in the cancers studied, but has shown to be pathogenic in germline cancers and functional assays (144, 159, 162). Finally, the use of *in silico* prediction tools such as SIFT (218), Polyphen (219) and MutationTaster (220) and mapping on the 3D structure of yeast enzymes, can evaluate the pathogenicity of variants and should be included in functional analyses.

## 1.6 Outline of Thesis

As described above, cancer-associated *POLE* exonuclease domain mutations are associated with striking molecular and clinicopathological characteristics in tumours. The aims of this work were to determine the functional and mechanistic effects of somatic and germline *POLE* exonuclease domain mutations and to determine if *POLE* mutations are an early event in cancer by:

- Generating and functionally analysing clinically relevant *POLE* exonuclease domain mutations in isogenic colorectal cancer cell lines by correcting a somatic $POLE^{P286R}$ mutation to wildtype using CRISPR-Cas9n. With these isogenic cell lines I sought to determine if cancer-associated *POLE* exonuclease domain mutations have a direct causal role in ultramutation and whether other cellular functions were affected (Chapter 3).

- Generating and phenotypically characterising an inducible knock-in germline $Pole^{L424V}$ mutant mouse model. I studied the phenotype in clinically-relevant tissues (intestine and uterus) by crossing the mice with tissue-specific cre-recombinases. Constitutive Pole L424V mouse models were generated and crossed with $Apc^{min}$ to determine whether the *Pole* mutation modified the multiple polyposis phenotype (Chapter 4).

- Sequencing a cohort of colorectal pre-cancers (adenomas) for *POLE* exonuclease domain mutations. *POLE* exonuclease domain mutant adenomas underwent molecular analyses to determine immune infiltrate and mutation burden to compare with the molecular characteristics observed in advanced *POLE* exonuclease domain mutant colorectal carcinomas (Chapter 5).

# Chapter 2

# Materials and Methods

## 2.1 Cell Culture

### 2.1.1 Cell lines

Cell lines used in this work are listed in Table 2.1. Cells were obtained either from repositories (ATCC, Korean Cell bank etc) or from collaborators. The identities of all cells were confirmed by correlation of somatic mutation data with those available in public databases (e.g. COSMIC cell line database).

### 2.1.2 Cell maintenance

Cell culture was carried out in a class II biological safety cabinet using sterile technique. Frozen cells were thawed rapidly at 37°C, centrifuged at 800g for five minutes, and the DMSO-containing medium aspirated. The cell pellet was resuspended in the appropriate growth medium, supplemented with 10% sterile-filtered Fetal Bovine Serum (FBS) (Sigma St. Louis, Missouri) and 1% Penicillin-Streptomycin (P/S) (Sigma), unless otherwise stated. Cells were cultured in a T75 tissue culture flask at 37°C in 5% $CO_2$ and passaged when 70-90% confluent. Cells were detached for subculture by incubation with 0.05% Tryspin-EDTA (ThermoFisher) at 37°C for 10

| Cell line | Source | Medium | *POLE* EDM | Microsatellite status |
|---|---|---|---|---|
| HCC2998 | Tomlinson Laboratory | RPMI (10% FBS and 1% P/S) | P286R | MSS |
| SNU-81 | Korean Cell bank | RPMI (10% FBS and 1% P/S) | P286R | MSS |
| HT115 | Tomlinson Laboratory | DMEM (10% FBS and 1% P/S) | V411L | MSS |
| HDC114 | Tomlinson Laboratory | IMDM (10% FBS and 1% P/S) | A456P | MSS |
| SNU-1040 | Korean Cell bank | RPMI (10% FBS and 1% P/S) | F367S | MSS |
| SW480 | Tomlinson Laboratory | L-15 (10% FBS and 1% P/S) | - | MSS |
| HTC116 | Tomlinson Laboratory | DMEM (10% FBS and 1% P/S) | - | MSI |
| HTC116 + MLH1 | Jonathan Grim Laboratory, University of Washington | DMEM (10% FBS and 1% P/S) | - | MSI |
| C32 | Tomlinson Laboratory | DMEM (10% FBS and 1% P/S) | - | MSS |
| HDC57 | Tomlinson Laboratory | DMEM (10% FBS and 1% P/S) | - | MSS |
| Colo320 | Tomlinson Laboratory | RPMI (10% FBS and 1% P/S) | - | MSS |
| HT29 | Tomlinson Laboratory | DMEM (10% FBS and 1% P/S) | - | MSS |
| T84 | Tomlinson Laboratory | DMEM (10% FBS and 1% P/S) | - | MSS |
| RKO | Tomlinson Laboratory | DMEM (10% FBS and 1% P/S) | - | MSI |
| GP2D | Tomlinson Laboratory | DMEM (10% FBS and 1% P/S) | - | MSI |
| LS180 | Tomlinson Laboratory | DMEM (10% FBS and 1% P/S) | - | MSS |

**Table 2.1:** Colorectal cell lines used in this thesis including source, medium and notable genetic features. RPMI, Roswell Park Memorial Institute; DMEM, Dulbeccos Modified Eagles Medium; IMDM, Iscove's Modified Dulbecco's Medium; L-15, Leibovitz-15; FBS, Fetal Bovine Serum; P/S, Penicillin/Streptomycin; EDM, Exonuclease domain mutation; MSS, Microsatellite stable; MSI, Microsatellite instability.

minutes. Trypsinisation was quenched by addition of growth medium containing 10% FBS. Cells for continued culture were plated at 77,000 cells/ml of medium in a T75 flask. All cell lines were tested for mycoplasma contamination every 2-3 months using the MycoAlert® mycoplasma detection kit (Lonza, Basel) as per the manufacturer's protocol.

For long-term storage, cells were harvested during log phase of growth, centrifuged at 800g then resuspended in 1ml freezing medium (10% FBS, 10% DMSO). Cells were transferred to a sterile cryovial and cooled to -80°C at 1°C per minute. Cryovials were stored in liquid nitrogen vapour between -140°C and -180°C.

#### 2.1.2.1 Counting cells

Cells were counted either manually using a FastRead102 disposable haemocytometer (Immune Systems, UK) or automatically with the Countess II automatic cell counter (LifeTech, Shenzhen). Calculation of cell concentration by manual counting used the following equation:

$$\text{cells/ml} = \frac{\text{Cell number}}{\text{Number of 4x4 grids counted}} \text{ x } 10^4$$

### 2.1.3 Transient transfection and siRNA knockdown

Cells for transfection were grown to 70-90% confluence in a 6-well plate. Transient transfection was performed using Lipofectamine® 2000 (ThermoFisher Scientific, UK) according to the manufacturer's protocol.

For plasmid and single stranded DNA, 10$\mu$l Lipofectamine® 2000 was diluted in 24$\mu$l of antibiotic and serum-free medium, then incubated at room temperature for 10 minutes to equilibrate. Up to 2500ng of plasmid and/or single-stranded DNA was diluted in antibiotic and serum-free medium then added to the diluted lipid. The solution was incubated at room temperature for five minutes to allow DNA/lipid complexes to form then added drop-wise to the cells. Culture medium was replaced 24 hours after transfection.

For gene knockdown, siRNA SMARTpool (Table 2.2) was transfected into cells using Lipofectamine® RNAiMAX (ThermoFisher Scientific, UK). 12$\mu$l RNAiMAX was diluted in 188$\mu$l antibiotic and serum-free medium in a 1.5ml eppendorf tube. In a separate tube, the appropriate concentration of siRNA was added to serum and antibiotic-free medium to a total volume of 200$\mu$l. RNAiMAX and siRNA were combined and incubated for five minutes at room temperature. RNAi-lipid complexes were added dropwise to cells at 60% confluence in a 6-well plate. Culture medium was replaced 24 hours after transfection.

Transfection efficiency was observed in live cells by Green fluorescent protein (pEGFP-C3 plasmid - Clontech) or SiGlo (Table 2.2) using the Nikon Eclipse TE2000U fluorescence inverted microscope (Nikon, Minato, Tokyo). Images were acquired with a Micropublisher 5.0 RTV digital camera, with Qimaging software.

| **siRNA** | **Product code** |
|---|---|
| SMARTpool ON-TARGETplus Human MLH1 siRNA | L-003906-00-0005 |
| SMARTpool ON-TARGETplus MSH6 siRNA | L-019287-00-0005 |
| ON-TARGETplus Non-targeting Pool (Scrambled) | D-001810-10-05 |
| siGLO Green Transfection Indicator | D-001630-01-05 |

**Table 2.2:** siRNA reagents and product codes for gene knockdown (Dharmacon).

### 2.1.4 Lentiviral production and transduction

#### 2.1.4.1 Generation of lentiviral particles

All lentiviral procedures were performed in a designated class II laminar flow biological safety cabinet. Waste from lentiviral particle production was disposed in clinical waste bins. HEK293T cells, below passage 15, were plated at a density of $7x10^6$ cells per 15cm dish (two dishes per lentivirus production) in IMDM medium supplemented with 10% FBS. On day two, when cells were 15% confluent and no latent infection was present, medium was replaced with IMDM + 10% FBS and 1% P/S and equilibrated to 37°C for two hours. The calcium chloride ($CaCl_2$) method was used to transfect cells. The following components required for lentiviral packaging were added to a 15ml Falcon tube: 32$\mu$g vector plasmid; 12.5$\mu$g pMDLg/pRREC (III generation) packaging plasmid; 9$\mu$l VSV-G envolope plasmid; 6.25$\mu$l pREV plasmid; 750$\mu$l 0.1XTE; 375$\mu$l endotoxin-free water and; 2.5M $CaCl_2$. 1250$\mu$l sterile 2X HEPES-buffered saline (HBS) was added drop-wise whilst vortexing. The solution was immediately added to the HEK293T cell culture dish and incubated for 16 hours at 37°C. The medium was replaced with 20ml of fresh IMDM medium + 10% FBS and 1% P/S with 10ng/$\mu$l doxycycline and 1mM Sodium Butyrate (NaBut) and incubated at 37°C for 24 hours. Lentiviral particles were harvested by filtering the medium through a 0.22$\mu$m filter and aliquoted in 2ml screw-top tubes. Lentiviral particles were stored at -80°C.

#### 2.1.4.2 Lentivirus titration

To ensure the correct multiplicity of infection (MOI) is achieved for optimum cell transduction, viral stocks were quantified using the Lenti-X p24 Rapid Titer Kit (Clontech) according to the manufacturer's instructions. A standard curve was generated by diluting 20$\mu$l of the enclosed p24 control (10ng/ml) with 980$\mu$l complete IMDM culture medium to create a 200pg/ml stock. The stock p24 solution was serial diluted in IMDM 2-fold six times.

Test viral samples were equilibrated to room temperature. In the meantime, 20$\mu$l lysis buffer was dispensed into the appropriate number of wells of p24 antibody-coated 8-well strips included in the kit. 200$\mu$l of standard curve dilution, test virus and blank medium were added to the wells in duplicate. The samples were incubated at 37°C for

60 minutes. The contents of the wells were aspirated, 100$\mu$l of biotin-conjugated anti-p24 antibody (included in the kit) was added to each well and incubated at 37°C for 60 minutes. The solution was aspirated, then 100$\mu$l of Streptavidin-HRP conjugate was added and incubated at room temperature for 30 minutes. The conjugate was aspirated and wells washed with 1X wash buffer. 100$\mu$l of Substrate Solution was immediately added to each well and incubated in the dark at room temperature for 30 minutes. 100$\mu$l stop solution was added and absorbance read at 450nm using the 96-well spectrophotometer plate reader (Molecular Devices, San Jose, California). The readings were normalised to the medium-only control and test virus quantified against the standard curve (Figure 2.1).

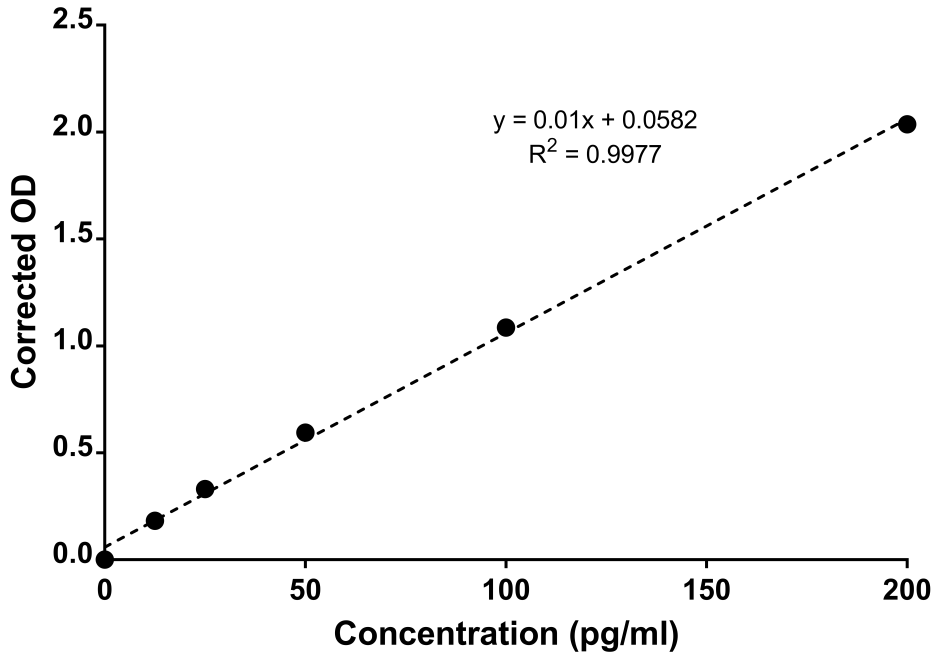



**Figure 2.1:** Standard curve of internal controls included with the Lenti-X p24 Rapid Titer Kit (Clontech). The final concentration of each sample (y) was calculated from the linear regression formula calculated from the standard curve.

#### 2.1.4.3 Transduction of cells

Mammalian cells were plated in 6cm dishes and cultured at 37°C until 30% confluent. Cells were transduced with 100ng of lentiviral particles in IMDM supplemented with 10% FBS, 1% P/S and 8ng/ml of Polybrene. Medium was changed 24 hours after

transduction and the cells passaged when 80% confluent.

### 2.1.5 Measuring cell viability and survival

#### 2.1.5.1 Absorbance and luminescence

Cell viability was measured using either the CellTiter 96® AQueous One Solution assay (Promega, Madison, Wisconsin) or the CellTiter-Glo® luminescent cell viability assay (Promega), according to the manufacturer's instructions. All experiments were performed in a 96-well tissue culture plate.

The CellTiter 96® AQueous One Solution assay uses the bioreduction of MTS tetrazolium compound by metabolically active cells into coloured formazan, thus it quantifies the total metabolic activity of a cell population rather than proliferation in the strict sense. For this, 20$\mu$l of reagent was added to each well of a 96-well tissue culture dish and incubated at 37°C for 1-2 hours. Absorbance was measured at 490nm using a 96-well spectrophotometer plate reader (Molecular Devices).

The CellTiter-Glo® Luminescent cell viability assay uses Luciferin catalysis by luciferase in the presence of magnesium, oxygen and ATP. The luminescent signal released is directly proportional to the number of viable cells. For this assay, 100$\mu$l of CellTiter-Glo reagent was added to 100$\mu$l cell culture in each well of an opaque walled 96-well plate and mixed for two minutes to induce cell lysis. The cells were incubated at room temperature for 10 minutes followed by luminescent signal quantification using the Luminoskan Ascent plate reader (Labsystems Diagnostics Oy, Finland).

#### 2.1.5.2 Clonogenic assay

Following preliminary optimisation experiments to determine plating efficiency, 50,000-150,000 experimental cells were plated in 10cm tissue culture dish containing 10ml of medium with 20% FBS and 1% P/S. 500 control cells were plated in triplicate in 10cm dishes in 10ml of medium with 20% FBS and 1% P/S to calculate plating efficiency. Cells were incubated for 24 hours to allow them to adhere and experimental cells were treated with an 40$\mu$M of 6-thioguanine. Experimental and control cells were incubated for 14-21 days, until colonies were visible to the naked eye. Cells were

washed with PBS then fixed and stained with 0.1% crystal violet in 20% methanol and colonies 20 cells or more in diameter were counted. Calculation of mutation rate was adapted from Glaab *et al.* (221):

$$\text{Mutation rate} = \frac{\text{Number of resistant colonies}}{\text{Number of cells plated x Plating efficiency}}$$

### 2.1.6 Clonal expansion

Clonal expansion is required to amplify cells derived from a single cell to establish an isogenic cell line for phenotypic characterisation or for DNA/RNA extraction.

For isogenic cell line generation, cells were plated at one cell/well in a 96-well tissue culture plate in 200$\mu$l of medium supplemented with 20% FBS and 1% P/S then cultured for 14 days. Cells were washed with PBS and detached with 30$\mu$l Trypsin-EDTA for 10 minutes.

If colonies were plated in a 10cm dish, cells were washed twice with PBS. One sterile Pyrex cloning cylinder (Corning) was coated in silicone grease (Sigma) and placed over a single colony to create a water-tight seal. Cells were detached with 30$\mu$l Trypsin-EDTA for 10 minutes.

Each colony was re-plated in a single well of a 48-well plate in 500$\mu$l of medium until 70-90% confluent. Cells were expanded until sufficient in number to freeze and/or extract DNA depending on the experiment.

### 2.1.7 Cell synchronisation

Cell lines were synchronised in the G1/S-phase cell cycle boundary by a double thymidine block, which relies on the addition of excess thymidine, causing replication stalling due to dNTP pool imbalance. Upon thymidine addition, cells stall at G1/S-phase or within S-phase if undergoing DNA replication. The second release and block allows cells stalled within S-phase to complete replication and stall at G1/S-phase upon a second exposure to thymidine.

Following optimisation, asynchronous cells were seeded at 1x$10^6$ cells in 10cm dishes in 10ml growth medium and cultured for eight hours at 37°C to adhere. Cells were treated with 2.5mM thymidine for 16 hours to block cells at G1/S-phase or in

S-phase if already undergoing genome duplication. Cells were then washed with PBS to release them from the thymidine block and incubated in fresh growth medium for eight hours. A second thymidine block (2.5mM thymidine) was then used for 16 hours arrests all cells at the G1/S-phase boundary.

### 2.1.8 Propidium iodide staining

Staining of DNA by propidium iodide permits experimental distinction between cells in G1 (2n), S (2n-4n) and G2 (4n) phase. Cells were trypsinised and resuspended at 1x$10^6$ cells/ml in PBS. To fix, 0.5ml of cell suspension was added to 1.5ml ice cold ethanol and incubated at 4°C for a minimum of two hours. Fixed cells were centrifuged at 700g for five minutes and resuspended in 0.2% Triton X-100 in PBS for five minutes to permeabilise. The cells were centrifuged at 700g for five minutes, washed in PBS and resuspended in 25$\mu$g/ml propidium iodide with 0.5mg/ml RNAse for 20 minutes. Cells were centrifuged at 100g in a benchtop centrifuge for 10 minutes, then resuspended in 1ml PBS ready for fluorescence activated cell sorting (FACS) analysis.

#### 2.1.8.1 FACS analysis

To distinguish cells in G1, S and G2-phase, propidium iodide-stained cells were analysed by fluorescence activated cell sorting (FACS) using the Sony SH800Z cell sorter (Sony Biotechnology, San Jose, California). Single cells only were analysed by adjusting the forward and side-scatter and the signal intensity at 488nm was quantified. Standard parameters used for FACS sorting are illustrated in Figure 2.2. FACS analysis was performed using FloJo® software and the stratification of cells into each phase of the cell cycle was performed automatically using unsynchronised cells as a reference to demarcate G1, S and G2-phase. The same parameters were used for each time-point.

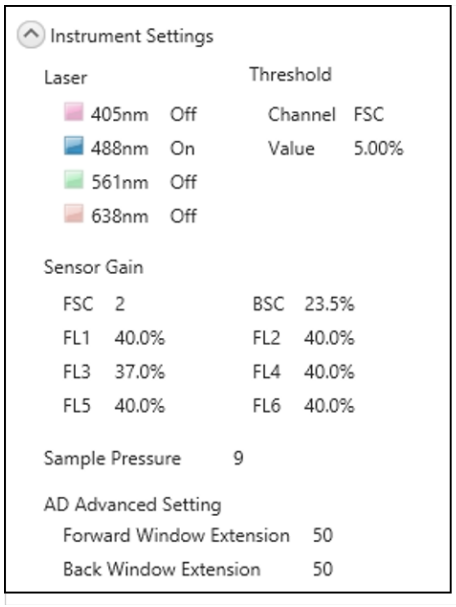



**Figure 2.2:** Standard FACS parameters for the analysis of propidium iodide-stained single cells.

### 2.1.9 DNA fibre immunofluorescence

The DNA fibre assay permits the analysis of newly synthesised DNA and was an adaptation of the protocol reported by Nieminuszczy *et al.* (222).

For this, 1x10$^6$ cells were plated in a 10cm dish in 10ml of medium and incubated at 37°C for 24 hours for cells to adhere and initiate the cell cycle. Nascent DNA at the replication fork was labelled stepwise with 25$\mu$M 5-Iodo-2'-deoxyuridine (IdU) for 20 minutes, then 250$\mu$M 5-Chloro-2'-deoxyuridine (CldU) for 20 minutes with incubation at 37°C in between addition of the thymine analogues. For analysis of DNA fork stalling, DNA was labelled for 20 minutes with 25$\mu$M IdU, with incubation at 37°C, then 2mM hydroxyurea (HU) and 250$\mu$M CldU was added and incubated at 37°C for two hours. Subsequently, cells were washed with ice-cold PBS to pause DNA replication and cells dissociated using Trypsin-EDTA. Detatched cells were resuspended in ice-cold PBS at 1x10$^6$ cells/ml. 2$\mu$l cell suspension was spotted onto a glass slide (at least three Superfrost® slides (VWR) per cell line) and air-dried at room temperature for five minutes before adding 7uL of lysis buffer (200 mM Tris-HCl, pH 7.5, 50 mM EDTA and 0.5% SDS). Lysis buffer was mixed gently with the cells using a pipette tip then incubated for two minutes at room temperature. The slide was tilted at a 15° angle for DNA fibres to spread down the length of the slide

and dried at room temperature.

DNA was fixed in 3:1 methanol:glacial acetic acid for 10 minutes, followed by denaturation in 2.5M hydrochloric acid for 80 minutes. Slides were washed for five minutes in PBS then incubated in blocking buffer (5% BSA in PBS) for 20 minutes in a humidified chamber. Slides were incubated at room temperature for two hours in a humidified chamber with 1° antibodies (1:25 mouse anti-BrdU for IdU (BD Biosciences, 347580) and 1:400 rat anti-BrdU for CldU (AbD serotec OBT0030)) in blocking buffer, then washed three times in PBS for five minutes each with gentle rocking. Slides were incubated in 2° antibody (1:400 Rabbit anti-mouse Alexa Fluor 568 and 1:400 Chicken anti-rat Alexa Fluor 488) for one hour at room temperature then washed three times in PBS for five minutes each, mounted with Prolong Gold® antifade mountant (ThermoFisher) and cured in the dark for 24 hours at room temperature. Fibres were analysed with an Olympus BX-51 microscope at 60X and images were taken with Genus Cytovision software (Leica).

A minimum of 450 fibres were analysed per cell line. Measurement of elongating DNA fibre tracts was performed using ImageJ software (223) using pixel length calibrated to 9.34 pixels/$\mu$m based on a 60X microscope objective.

### 2.1.10 Chromosome counts and Fluorescence in situ hybridization

Cytogenetics and Fluorescence in situ hybrisisation (FISH) were used to ascertain chromosome and gene copy number in human colorectal cancer isogenic cell lines. The protocol was adapted from Knight *et al.* (224). Firstly, 1x$10^6$ cells were seeded in a T75 flask and incubated at 37°C for 24 hours. Colecemid, a spindle fibre inhibitor, was then added at 30ng/ml for three hours to arrest cells in metaphase, after which, cells were detached by trypsin and centrifuged to pellet. Pelleted cells were then lysed in 10ml 75mM KCl at 37°C for 17 minutes, spun at 130g for 10 minutes and the cell pellet resuspended in fixative (3:1 methanol: glacial acetic acid) and incubated for 30 minutes at room temperature. The cells were centrifuged and resuspended in ice-cold fixative for 10 minutes. The cells were centrifuged one final time, resuspended in 200$\mu$l fixative and 30$\mu$l of cell suspension spotted on a clean slide then dried overnight. For

chromosome counts, slides were mounted in DAPI/Vectashield (Vector Laboratories). Chromosome number was quantified from a minimum of 25 metaphases per cell line.

For FISH, metaphase spreads were washed in 2X saline sodium citrate (SSC) at 70°C for 300 minutes, followed by one minute at room temperate in 0.1X SSC then denatured in 70mM sodium hydroxide for two minutes. Cells were washed in 0.2% SSC at 4°C for one minute, followed by dehydration through an ethanol series (70%, 90% and 100%) for two minutes each. 1020$\mu$l of BAC probe mix (listed in table 2.3) was applied to the slide, covered with a coverslip and incubated overnight at 37°C in a humidified chamber.

The slides were washed three times with 0.1 x SSC at 60°C. Probes were detected with 50$\mu$l of a solution containing anti-digoxigenin antibody conjugated with rhodamine (Sigma), avidin FITC or avidin Cy5 (Molecular Probes, Thermofisher), at 5$\mu$g/ml in 4 x SSC, 0.1% Tween-20 and 4% BSA. The slides were incubated at 37°C in a humidified chamber for 30 minutes. Finally, slides were washed three times in 4 x SSC, 0.1% Tween-20 at 42°C, followed by a wash in 4 x SSC.

Slides were mounted in DAPI/Vectashield mounting medium (Vector Laboratories), under a glass coverslip and analysed using an Olympus BX51 microscope. Images were collected using Cytovision Genus v7.1 software(Leica). A minimum of 50 metaphase or interphase cells were analysed per cell line.

| BAC | Start (hg19) | End (hg19) | Cytogenetic band | Symbol |
|---|---|---|---|---|
| RP11-25J3 | 75,892,543 | 76,047,496 | 12q21.2 | Chr12 Internal probe |
| RP11-148L11 | 133,038,915 | 133,243,361 | 12q24.33 | PolE |
| CTC221K181 | na | na | 12q24.3 | Telo 12q |

**Table 2.3:** Fluorescence in situ hybridization BAC probes.

## 2.2 DNA analysis

### 2.2.1 DNA extraction

#### 2.2.1.1 Formalin-fixed paraffin embedded (FFPE) tissue

DNA was extracted from Formalin-Fixed Paraffin-Embedded (FFPE) tissue using the HighPure FFPET DNA Isolation Kit (Roche, Basel) according to the manufacturer's protocol. To maximise the tumour cellularity of the samples for DNA extraction, areas of adenoma or carcinoma (as distinct from the tumour stroma) were identified on 4$\mu$m sections stained with haematoxylin and eosin (H&E) and used to guide manual dissection of tumour tissue from adjacent unstained 10$\mu$m sections using a 21 gauge needle.

Tumour tissue was re-suspended in 180$\mu$l of tissue lysis buffer, with the addition of 70$\mu$l proteinase K (at 22mg/ml) in a 2ml screw capped tube. The tube containing the sample was vortexed, spun briefly then incubated at 56°C with agitation until no visible tissue remained. The sample was then incubated at 90°C for 60 minutes, cooled gradually to room temperature then spun down briefly.

Tissue lysate was mixed with 200$\mu$l of DNA-binding buffer and incubated for 10 minutes, followed by the addition of 100$\mu$l of isoporopanol. The tissue lysis solution was added to a High Pure filter tube to bind the DNA and centrifuged at 8,000g for 60 seconds. The sample was washed once with wash buffer I and twice with wash buffer II to remove salts, proteins and cell debris. Genomic DNA was eluted in 30$\mu$l of Milli-Q (Millipore) water or elution buffer. Before DNA quantification by Nanodrop spectrophotometry (Section 2.2.2.1), the elute was centrifuged for two minutes at maximum speed to remove glass fibres then transferred to a clean 2ml screw-top tube.

#### 2.2.1.2 Fresh-frozen tissue and cultured cells

DNA was extracted from fresh-frozen tissue samples and cells in culture using the DNeasy blood and tissue kit (Qiagen, Hilden) according to the manufacturer's protocol. Snap-frozen tissue from euthanised mice was thawed on ice and up to 25mg

homogenised by micropestle in 180$\mu$l alkaline lysis buffer ATL (Qiagen) in a 1.5ml eppendorf tube. Following addition of 20$\mu$l proteinase K, tissue was digested by incubation at 56°C overnight. Samples were vortexed and 200$\mu$l Buffer AL and 200$\mu$l 100% ethanol added to the sample. The sample was mixed again by vortexing.

DNA extraction from cells in culture was done by a slightly modified version of this protocol. In this, up to 5x10$^6$ cells were dissociated by Trypsin-EDTA (Invitrogen) or manual scraping (Corning). Cells were pelleted by centrifugation at 800g for five minutes and the supernatant aspirated. Cells were resuspended in 200$\mu$l PBS, 200$\mu$l alkaline lysis buffer AL (Qiagen) and 20$\mu$l proteinase K and incubated for 10 minutes at 56°C.

Following fresh-frozen tissue digestion or cell lysis, the sample was added to a microspin column and centrifuged at 6,000g to bind DNA to the silica membrane. The column was washed once with AWI and once with AWII then DNA eluted in up to 200$\mu$l Tris-EDTA (TE) or Milli-Q water depending on downstream applications.

#### 2.2.1.3 Bacteria

Transformed plasmids for molecular cloning were purified using one of two methods depending on the number of bacteria in culture.

To screen for positively transformed clones, DNA was extracted from 4ml bacterial culture using the QIAprep Spin Miniprep Kit (Qiagen). Bacteria were pelleted at 6,800g for three minutes, the supernatant aspirated then cells resuspended in 250$\mu$l Buffer P1 with RNaseA. The solution was neutralised with buffer P2, inverted five times and incubated at room temperature for five minutes. The sample was adjusted to a high salt environment with the addition of buffer P3 to precipitate proteins, chromosomal DNA and cell debris for removal. Plasmid DNA was adsorbed onto a silica membrane in a high salt environment, washed and eluted in 50$\mu$l buffer EB (10mM Tris.Cl, pH 8.5).

For large 250ml bacterial cultures, plasmid DNA was purified using the plasmid Maxiprep kit (Qiagen). Instead of using silica membrane adsorption, DNA was bound to an anion-exchange column, washed, concentrated and desalted using isopropanol precipitation. DNA was then dissolved in sterile endotoxin-free water in a laminar flow hood for use in mammalian cell culture.

#### 2.2.1.4 Blood

Patient blood stored in vacutainers (with EDTA) at -80°C was equilibrated to room temperature and inverted several times to mix. DNA was extracted using the Maxwell® 16 Tissue DNA Purification Kit (Promega, Madison, Wisconsin) according to the manufacturer's protocol. 420$\mu$l of blood was added to well number one of the Maxwell cartridge and 230$\mu$l elution buffer was added to the blue elution tube. The magnetic plunger was placed in well seven of the Maxwell cartridge. Automated whole blood DNA purification by the Maxwell®16 AS2000 (Promega) in the standard elution volume (SEV) configuration was used for all samples.

#### 2.2.1.5 Hot sodium hydroxide and tris (HotSHOT) DNA extraction

DNA from mouse ear punches and cell pellets for genotyping PCR was extracted by alkaline lysis using the "HotSHOT" protocol as described by Truett *et al.* (225). 50$\mu$l of alkaline lysis buffer (25mM NaOH, 0.2mM EDTA, pH 12) was added to tissue or cells and incubated for 30 minutes at 95°C. The solution was neutralised with 50$\mu$l Tris-HCl (40mM , pH 5) and cooled to 4°C. Up to 2$\mu$l of product was used per PCR reaction.

### 2.2.2 Determination of DNA quantity and quality

#### 2.2.2.1 NanoDrop

Nucleic acid concentration and purity was quantified using the NanoDrop™ 2000 UV-visible spectrophotometer (ThermoFisher, Waltham, Massachusetts). The Nanodrop was calibrated with a blank control (nucleic acid diluent) then 1$\mu$l of DNA/RNA was loaded onto the optical pedestal. Nucleic acid concentration was calculated from the absorbance at 260nm using a modified Beer-Lambert equation and appropriate extinction coefficient for double-stranded DNA, single-stranded DNA or RNA. A $\frac{A260}{A280}$ ratio of 1.8 to 2.1 was taken as indicating acceptable sample purity.

#### 2.2.2.2 Qubit

The Qubit$^{TM}$ dsDNA High Sensitivity (HS) assay (ThermoFisher, UK) was used for high sensitivity quantification of genomic DNA. dsDNA high sensitivity reagent was diluted 1:200 in dsDNA HS Buffer and vortexed to mix. Either 1$\mu$l of DNA sample or 10$\mu$l of standard was added to individual 0.5ml Qubit assay tubes and normalised to 200$\mu$l with dsDNA HS reagent. Samples were vortexed and incubated for two minutes at room temperature. The samples were read on the Qubit 2.0 fluorometer and the DNA concentration automatically quantified from the calibration curve generated from readings of freshly prepared standards.

#### 2.2.2.3 TapeStation

The quality and quantity of DNA libraries for Next Generation Sequencing (NGS) were analysed using the TapeStation 2200 (Agilent, Santa Clara, California). Using a Screentape gel matrix, the TapeStation separates the DNA by molecular weight to assess DNA integrity and allow quantification of specific peaks associated with library size. Firstly, in separate tubes 1$\mu$l of ladder was mixed with 10$\mu$l sample buffer and 1$\mu$l of DNA library was mixed with 10$\mu$l sample buffer. Both samples were vortexed for one minute then spun down in a microcentrifuge. The tips, ScreenTape and DNA samples were loaded into the machine and run according to the manufacturer's instructions. Figure 2.3 illustrates an example output from the TapeStation 2200 software.

### 2.2.3 DNA purification from agarose gels

To remove unwanted PCR products (e.g. primer-dimers) or restriction digest products, DNA was purified from agarose gels using the QIAquick gel extraction kit (Qiagen) as per the manufacturer's protocol. DNA was electrophoresed on an agarose gel containing ethidium bromide for 20-40 minutes at 120 volts. The bands were visualised with a Dark Reader blue transilluminator (Clare Chemical Research, Colorado) and bands of the expected molecular weight were excised from the gel. Excised fragments were weighed and dissolved in buffer QG (100mg gel to 100$\mu$l buffer) at 50$^{o}$C

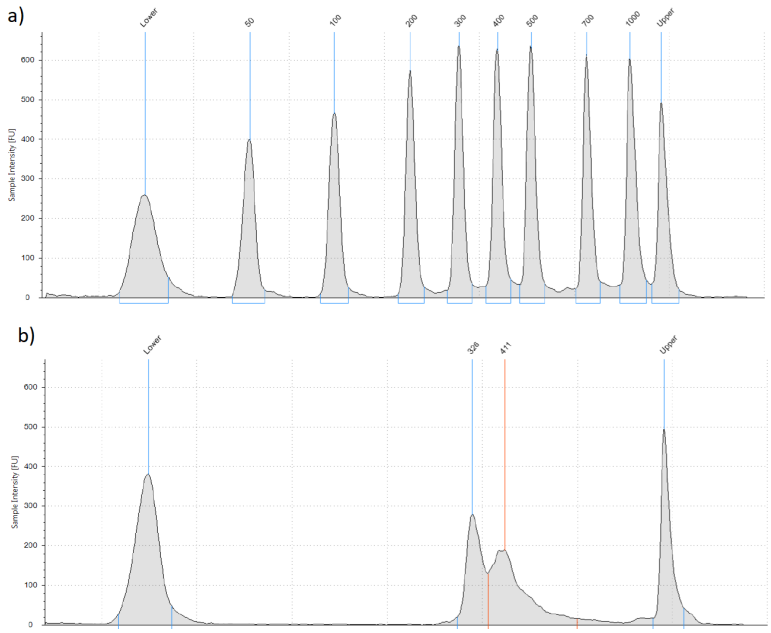



**Figure 2.3:** Example TapeStation output for the DNA ladder (a) and an NGS library (b).

for 10 minutes. One volume of isopropanol was added to the solution and vortexted, then the sample placed in a microspin column. DNA was purified by silica membrane adsorption in a high sodium solution, washed and eluted in up to 50$\mu$l Milli-Q water.

### 2.2.4 Polymerase chain reaction (PCR)

Primers for PCR amplification of genomic fragments were either in use in the host laboratory or were designed using Primer3 software (226). Primers were purchased from Life Technologies.

Prior to use, lyophilized primers were resuspended at 100$\mu$M in Milli-Q water then diluted to a working stock of 10$\mu$M in Milli-Q water. Individual PCR conditions varied according to primer tm and amplicon size and were optimised in all cases. In general, a standard PCR mix and reaction conditions for a 10$\mu$l reaction and amplification are set out in Tables 2.4 and 2.5 respectively.

The specific magnesium concentration, annealing temperature, extension time and additional reagents are specified in the primer tables in Appendix Section A.

PCR products were analysed alongside a 1kb Plus ladder (ThermoFisher), using agarose gel electrophoresis in Tris-Borate EDTA (TBE) running buffer. The gels

| Reagent | Working conc. |
|---|---|
| Buffer 10X (Bioline) | 1X |
| $MgCl_2$ | 1-2mM* |
| dNTPs (2.5mM) (bioline) | 100$\mu$M |
| Fw (10$\mu$M) | 0.4$\mu$M |
| Rv (10$\mu$M) | 0.4$\mu$M |
| Taq Bioline | 0.1U/$\mu$l |
| $ddH_2O$ | up to 10$\mu$l |
| DNA | 20ng* |

**Table 2.4:** Standard PCR master mix. * Denotes a variable parameter.

| | Cycles | Time | Temperature (°C) |
|---|---|---|---|
| **Initial denaturation** | 1 | 2 min | 95 |
| **Denaturation** | | 10-30sec | 95 |
| **Annealing** | 30-40* | 30 sec | 55-65* |
| **Elongation** | | 30-60 sec* | 72 |
| **Final elongation** | 1 | 10 min | 72 |

**Table 2.5:** Standard PCR reaction conditions. * Denotes a variable parameter.

were stained with ethidium bromide and were visualised using the ChemiDoc imager (BioRad, Hercules, California).

#### 2.2.4.1 Long-range PCR amplification

All long-range PCR reactions were performed with LA Taq DNA polymerase, containing 3'-5' exonuclease activity to proofread PCR errors (TAKARA BIO INC., Kusatsu, Shiga Prefecture) according to the manufacturer's protocol. Standard reagent master mix and PCR conditions are listed in Tables 2.6 and 2.7 respectively.

| Reagent | Working conc. |
|---|---|
| Buffer 10X | 1X |
| $MgCl_2$ | 1.5mM |
| dNTPs (2.5mM) | 100$\mu$M |
| Fw (10$\mu$M) | 0.5$\mu$M |
| Rv (10$\mu$M) | 0.5$\mu$M |
| TaKaRa LA Taq | 2.5U |
| $ddH_2O$ | - |
| DNA | 100ng |

**Table 2.6:** Standard long-range PCR master mix.

| | Cycles | Time | Temperature (°C) |
|---|---|---|---|
| Initial denaturation | 1 | 1 min | 95 |
| Denaturation | 30 | 30 sec | 95 |
| Annealing/Elongation | | 7 min | 68 |
| Final Elongation | 1 | 10 min | 72 |

**Table 2.7:** Standard long-range PCR reaction conditions.

#### 2.2.4.2 High-fidelity short-range PCR

Short-range PCR products from genomic DNA were amplified using FastStart High Fidelity (HiFi) Taq (Roche, Basel), which includes additional 3'-5' exonuclease activity to proofread replication errors. The reaction mix and amplification was performed according to the manufacturer's protocol. Standard reagent master mix and PCR conditions are listed in Tables 2.8 and 2.9 respectively.

| Reagent | Quantity ($\mu$l) |
|---|---|
| Buffer 10X (18mM $MgCl_2$ ) | 2.5 |
| dNTPs | 0.5 |
| Fw (10$\mu$M) | 1 |
| Rv (10$\mu$M) | 1 |
| FastStart HiFi (5U/$\mu$l) | 0.25 |
| dd$H_2O$ | 17.75 |
| DNA 25ng/$\mu$l | 2 |

**Table 2.8:** Standard High-Fidelity PCR master mix.

| | Cycles | Time | Temperature (°C) |
|---|---|---|---|
| Initial denaturation | 1 | 2 min | 95 |
| Denaturation | 30 | 30 sec | 95 |
| Annealing | | 30 sec | 60 |
| Elongation | | 30 sec | 72 |
| Final elongation | 1 | 7 min | 72 |

**Table 2.9:** Standard High-Fidelity PCR reaction conditions.

#### 2.2.4.3 Site-directed mutagenesis

The QuikChange II XL site-directed mutagenesis kit (Agilent, Santa Clara, California) was used insert specific point mutations into a plasmid according to the manufacturer's instructions. Two primers, complementary to the sense and antisense plasmid

DNA strands, containing the point mutation, were synthesised (Appendix Table A.3). Plasmids were mutagenised using the reaction conditions in Table 2.10. The reaction was placed on a thermal cycler and run according to the manufacturer's instructions (Table 2.11). The parental plasmid was digested at 37°C for one hour with 10U of *DpnI* and transformed into of XL10-Gold Ultracompetent cells (Section 2.4.4).

| **Reagent** | **Quantity ($\mu$l)** |
|---|---|
| Buffer 10X | 5 |
| dNTPs | 1 |
| Fw (125ng) | 1 |
| Rv (125ng) | 1 |
| QuikSolution | 1.5 |
| PfuUltra HF DNA polymerase (2.5 U/$\mu$l) | 1 |
| dd$H_2O$ | up to 50$\mu$l |
| DNA 50ng/$\mu$l | 1 |

**Table 2.10:** QuikChange II XL site-directed mutagenesis master mix.

| | **Cycles** | **Time** | **Temperature (°C)** |
|---|---|---|---|
| Initial denaturation | 1 | 2 min | 95 |
| Denaturation | | 20 sec | 95 |
| Annealing | 18 | 10 sec | 60 |
| Elongation | | 3 min | 68 |
| Final elongation | 1 | 5 min | 68 |

**Table 2.11:** QuikChange II XL site-directed mutagenesis reaction conditions.

### 2.2.5 Microsatellite instability testing

Microsatellite instability status was determined by analysis of mononucleotide and dinucleotide repeats using the modified Bethesda markers Bat25, Bat26, Bat40, D2S123, D5S346 and D17S250 and fluorescently labelled primers (listed in Table 2.12).

A standard triplex PCR mix consisted of 0.2-0.6ng primers, 5$\mu$l of Qiagen master mix (QMM) solution (Qiagen), 40ng of DNA topped up to 10$\mu$l with Milli-Q water. A standard PCR cycle (Section 2.2.4) with an extension of 90 seconds at 60°C and elongation of 45 seconds at 72°C was used to amplify dinucleotides and mononucleotides. Samples were sent to the University of Oxford Zoology Department sequencing facility. Dinucleotides and mononucleotide peaks were analysed using GeneMarker®

software (SoftGenetics) and the total frequency of peaks was compared with gDNA from the same patient. Microsatellite instability was defined as having >30% of unstable markers from at least three independent markers (Figure 2.4) (166).

| Microsatellite marker | Repeat type | Location | Forward primer (5'-3') | Reverse primer (5'-3') | Size (bp) |
|---|---|---|---|---|---|
| Bat25 | Mononucleotide | *c-kit*, chr4q12 | TCGCCTCCAAGAATGTAAGT | TCTGGATTTTAACTATGGCTC | 124 |
| Bat26 | Mononucleotide | *hMSH2*, chr2p16.3-p21 | TGACTACTTTTGACTTCAGCC | TTCTTCAGTATATGTCAATGAAAACA | 154 |
| Bat40 | Mononucleotide | *HSD3B2*, chr1p13.1 | CCTACACCACAACCCTGCTT | TGAGGTGGGAGGATAAATGG | 198 |
| D2S123 | Dinucleotide | *MSH2*, chr2p16 | AAACAGGATGCCTGCCTTTA | GGACTTTCCACCTATGGGAC | 211 |
| D5S346 | Dinucleotide | *APC*, chr5q21/22 | ACTCACTCTAGTGATAAATCGGG | AGCAGATAAGACAGTATTACTAGTT | 125 |
| D17S250 | Dinucleotide | *BRCA1*, chr17q11.2-q12 | GGAAGAATCAAATAGACAAT | GCTGGCCATATATATATTTAAACC | 152 |

**Table 2.12:** Primers used to determine microsatellite status. One of each primer pair was fluorescently labelled. Table adapted from Zhang *et al.* (227).

### 2.2.6 *Pla2g2a* genotyping by restriction length polymorphism (RFLP) analysis and Sanger sequencing

A 1bp insertion in *Pla2g2a*, thought to modify the polyp burden in $Apc^{\text{Min}}$ mice, abolishes a BamHI restriction site, making it possible to detect homozygous mutant, heterozygous mutant and wildtype genotypes from the size of digested PCR fragments run on an agarose gel.

Inbred C57BL/6J ($Pla2g2a^{\text{ex4fs/ex4fs}}$) and BALB/c ($Pla2g2a^{+/+}$) mice were sourced to use as negative and positive controls respectively for restriction digest. DNA was extracted from ear punches by HotSHOT and *Pla2g2a* genotyped using restriction length polymorphism as described in Gould *et al.* (228). A 476bp amplicon was synthesised by a standard PCR reaction using primers listed in Appendix Table A.4. 5$\mu$l of the PCR product was digested in 20U of BamHI restriction enzyme (NEB) and 1x CutSmart Buffer (NEB) for 2.5 hours then run on an agarose gel with ethidium bromide at 120 volts for 40 minutes. The bands were visualised on the ChemiDoc imager (BioRad) (Figure 2.5a). PCR products were Sanger sequenced for the 1bp thymine insertion (Section 2.2.7.1) and the peaks visualised using SnapGene® software (GSL Biotech; available at snapgene.com) (Figure 2.5b).

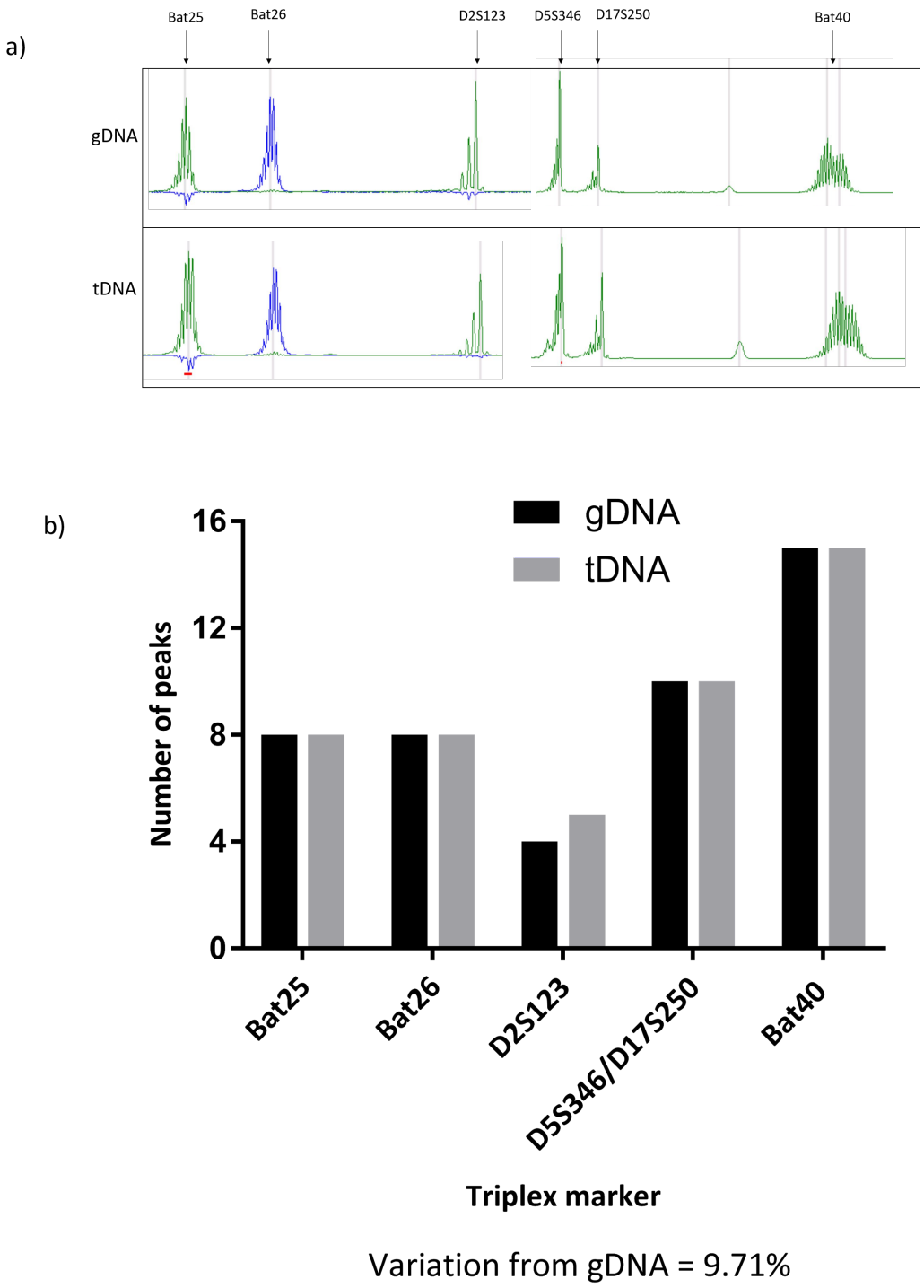



**Figure 2.4:** Microsatellite instability analysis. a) Mononucleotide and dinucleotide electropherograms for tDNA and gDNA visualised using GeneMarker® software (SoftGenetics). b) Number of peaks in tDNA and gDNA for each mononucleotide and dinucleotide marker. Cumulative percentage variation of tDNA fro gDNA is listed below the graph

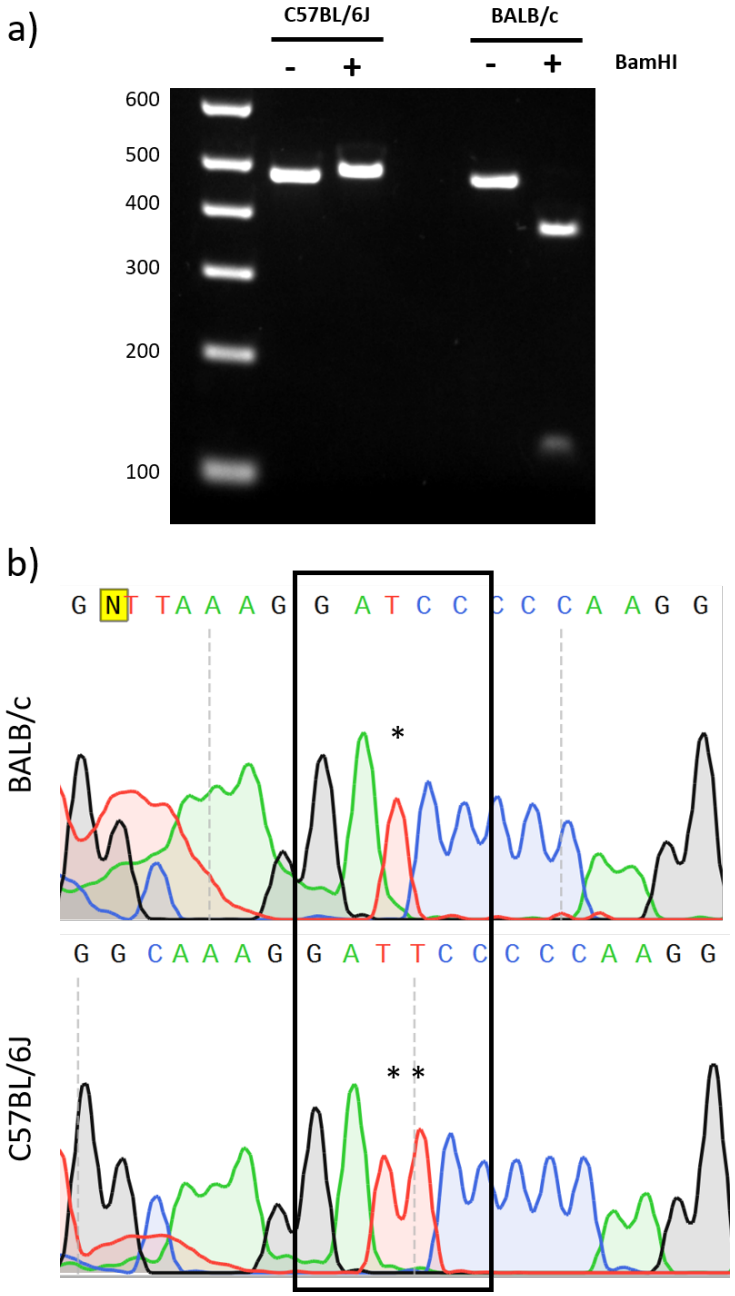



**Figure 2.5:** *Pla2g2a* genotyping using restriction fragment length polymorphism (RFLP) analysis and Sanger sequencing. a) Gel showing products from a BamHI restriction digest (RFLP analysis) with and without the restriction enzyme. In the presence of BamHI the wildtype *Pla2g2a* PCR product from BALB/c digest into 100bp and 400bp products. Homozygous *Pla2g2a*$^{ex4fs/ex4fs}$ PCR products from C57BL/6J mice do not digest due to disruption of the BamHI restriction site. In the absence of the enzyme, there is no evidence of PCR product degradation. b) Chromatogram of the 1bp *Pla2g2a* insertion site using primers from Appendix Table A.4. A homozygous 1bp thymine insertion mutation in exon 4 in *Pla2g2a* is present in C57BL/6J mice but not BALB/c mice. The star denotes the 1bp insertion site.

### 2.2.7 DNA/cDNA sequencing

#### 2.2.7.1 Sanger sequencing

PCR-amplified DNA for Sanger sequencing was prepared and analysed using the following method. First, to hydrolize excess primers and dNTPs, which can interfere with downstream sequencing, 2.5$\mu$l of PCR product was added to 1$\mu$l ExoSAP-IT$^{TM}$ (ThermoFisher) and incubated on a thermocycler at 37°C for 15 minutes, then 80°C for 15 minutes to inactivate the enzyme. Fluorescent-labelling of nucleotides was then performed by a 20$\mu$l PCR reaction using the reagents in Table 2.13 and PCR conditions in Table 2.14.

| **Reagent** | **Quantity ($\mu$l)** |
|---|---|
| Primer (10$\mu$M) | 0.5 |
| BigDye Terminator v3.1 (ThermoFisher) | 8 |
| dd$H_2O$ | 0.5 |
| PCR cleanup/Plasmid DNA | 1 |

**Table 2.13:** BigDye Terminator v3.1 master mix.

Excess dye terminators were removed by gel filtration (Qiagen DyEX 2.0 spin kit), the samples lyophilized by Speedvac Savant Concentrator (ThermoFisher) then run on the ABI 3730 DNA analyser (Applied Biosystems, Foster City, California) by the University of Oxford Zoology Department sequencing facility.

Mutations were called automatically using Mutation Surveyor from human genomic DNA sequencing (229) or manually using ApE (A plasmid Editor by M. Wayne Davis) or SnapGene® software (GSL Biotech; available at snapgene.com) when analysing plasmid and murine DNA sequencing data.

| | **Cycles** | **Time** | **Temperature (°C)** |
|---|---|---|---|
| Initial denaturation | 1 | 5 min | 96 |
| Denaturation | 35 | 30 sec | 96 |
| Annealing | | 45 sec | 50 |
| Elongation | | 4 min | 60 |
| Final elongation | 1 | 10 min | 60 |

**Table 2.14:** BigDye Terminator v3.1 PCR conditions.

## 2.3 *HPRT1* library preparation

DNA libraries from *HPRT1*-mutant colonies were prepared according to the following protocol. DNA from each 6-thioguanine-resistant colony and post-HAT-treated cells was quantified by Qubit (Section 2.2.2.2) and normalised to 25ng/$\mu$l in double-distilled $H_2O$. Each exon of the *HPRT1* gene was amplified using the FastStart High Fidelity (HiFi) Taq (Section 2.2.4.2) with the primer pairs listed in Appendix Table A.2.

Amplicons from each sample were run on a 2% agarose gel at 120 volts for 20 minutes and graded by band intensity fro 1-4 from strongest to weakest. Amplicons from each sample were pooled (grade 1 = 1$\mu$l, grade 4 = 4$\mu$l), the solution topped up to 100$\mu$l with dd$H_2O$ and vortexted to mix. One microlitre of pooled amplicons were indexed with Illumina barcodes using a standard PCR reaction mix listed in Table 2.4 in a volume of 25$\mu$l. Barcoding PCR conditions are listed in Table 2.15.

| | **Cycles** | **Time** | **Temperature (°C)** |
|---|---|---|---|
| Initial denaturation | 1 | 10 min | 95 |
| Denaturation | | 15 sec | 95 |
| Annealing | 12 | 30 sec | 60 |
| Elongation | | 60 sec | 72 |
| Final elongation | 1 | 3 min | 72 |

**Table 2.15:** MiSeq library barcoding PCR conditions.

Each barcoding reaction was run on an agarose gel at 120 volts for 40 minutes to confirm amplification. Subsequently, 10$\mu$l of each sample was pooled, gel purified and eluted in 30$\mu$l dd$H_2O$. To confirm library purify, the samples were analysed using the TapeStation 2200 and quantified using Qubit.

#### 2.3.0.1 Illumina MiSeq 3000

*HPRT1* DNA libraries were adjusted to 4nM in 20$\mu$l using the following equation:

$$\text{Concentration in nM} = \frac{\text{Concentration in ng}/\mu\text{l}}{\text{660g/mol x average library size in bp}} \text{ x } 10^6$$

Pooled libraries were re-quantified by Qubit and prepared for analysis on the Illumina MiSeq V2 500 cycles Nano kit. Sequencing was performed by Dr Annabelle Lewis, University of Oxford or by the Genomics Core, University of Birmingham.

### 2.3.1 TaqMan™ genotyping of point mutations

Point mutation screening of large numbers of samples was performed by competitive allele-specific PCR TaqMan™ (ThermoFisher). This uses sequence-specific forward and reverse primers and two quenched fluorescently-labelled (VIC or FAM) probes specific to the wildtype and mutant alleles respectively. Upon sequence amplification with AmpliTaq Gold™ DNA polymerase, the intrinsic 3'-5' exonuclease activity degrades the probe, cleaving the fluorescent dye away from the quencher, thus permitting fluorescent signal detection specific to the allele. The TaqMan reaction mix (in a 96-well optical plate) and thermal cycling conditions are listed in Tables 2.16 and 2.17 respectively.

| **Reagent** | **Volume ($\mu$l)** |
|---|---|
| 2X TaqMan genotyping master mix | 5 |
| 80X TaqMan Assay mix | 0.12 |
| gDNA (10ng/$\mu$l) | 2 |
| $H_2O$ | 2.88 |

**Table 2.16:** TaqMan genotyping master mix.

| | **Cycles** | **Time** | **Temperature (°C)** |
|---|---|---|---|
| Initial denaturation | 1 | 15 min | 95 |
| Denaturation | 40 | 10 sec | 95 |
| Annealing/Elongation | | 60 sec | 60 |

**Table 2.17:** TaqMan genotyping PCR conditions.

Post-PCR amplification, the relative VIC and FAM fluorescence was quantified on the ABI 7900HT Real-Time quantitative PCR (RT-qPCR) machine and sample genotype associated with the relative levels of each fluorescent dye. Two samples positive for the point mutation of interest, two wildtype samples and two no template reactions were included as controls.

## 2.4 DNA Cloning

### 2.4.1 Restriction digestion

Plasmids were digested at 37°C overnight using restriction enzymes (NEB, Ipswich, Massachusetts) as specified in each chapter. A standard restriction digest consisted of 1mg plasmid DNA or 1$\mu$l PCR product, 1X restriction digest buffer, and 1$\mu$l restriction enzyme and dd$H_2O$ to 10$\mu$l.

Assessment of the restriction digest was performed using gel electrophoresis at 120 volts for 40 minutes on a 1% agarose gel containing ethidium bromide and visualised using the ChemiDoc imager (BioRad).

### 2.4.2 DNA purification

Digested DNA or a PCR product was purified using the QIAquick PCR purification kit (QIAgen) according to the manufacturer's protocol. Five volumes of buffer PB was added to one volume of digested DNA or PCR product and vortexed to mix. The sample was placed in a QIAquick column and centrifuged, washed with buffer PE and dried by an additional centrifugation. DNA was eluted in 40$\mu$l of buffer EB or dd$H_2O$ into a 1.5ml eppendorf tube. DNA concentration was determined by Nanodrop spectrophotometry (2.2.2.1).

### 2.4.3 DNA ligation

DNA with sticky ends was ligated using T4 ligase (NEB) using the following reaction mix: 1X T4 ligase buffer (NEB), 1:2 vector DNA:insert DNA and nuclease-free water up to 20$\mu$l. The reaction was incubated overnight at 15°C in a waterbath and 1$\mu$l reaction mix was transformed into competent bacteria.

### 2.4.4 Bacterial transformation

Plasmids were cloned into XL10-Gold® (Stratagene, San Diego, California) Ultracompetent cells. 50$\mu$l of XL10-Gold cells were incubated with with 2$\mu$l $\beta$-Mercaptoethanol

in an ice-cold 14ml BD Falcon polypropylene tube for 10 minutes, then mixed with up to 50$\mu$l plasmid or 1$\mu$l of ligation reaction. The bacterial cells were incubated on ice for 30 minutes then heat shocked at 42°C for exactly 30 seconds in a waterbath. After one hour of recovery in 500 $\mu$l S.O.C. medium (ThermoFisher) at 37°C, the bacteria were spread at 1/10 and 9/10 dilution on LB agar plates with the appropriate selection antibiotic for the transformed plasmid. Plates were incubated at 37°C for 16 hours to permit single bacterial colony growth.

To amplify bacterial colonies, 4ml of LB broth with the appropriate antibiotic were inoculated with one colony and cultured at 37°C with shaking for 12-16 hours. Depending on the volume of DNA required, cells were either pelleted for Miniprep or 250ml of LB broth with antibiotic was inoculated with 1ml of bacterial culture and cultured for 12-16 hours at 37°C with shaking for Maxiprep plasmid purification (Section 2.2.1.3).

## 2.5 Analysis of RNA

### 2.5.1 RNA extraction

RNA was extracted from fresh frozen cell lines or mouse tissue using the RNeasy Plus Mini Kit (Qiagen) according to the manufacturer's instructions. Tissue was lysed by vortexing or mechanical disruption by micropestle in buffer RLT, containing guanidine-isothiocyanate to inactivate RNAses. The lysate was centrifuged at maximum speed to remove cell debris and the lysate transferred to a gDNA eliminator spin column. The flow through, mixed with ethanol was transferred to an RNeasy spin column to which the RNA bound, washed to remove contaminants and eluted in 50$\mu$l RNase-free water. The RNA elute was stored at -80°C.

RNA concentration and integrity was assessed using the NanoDrop$^{TM}$ 2000 (ThermoFisher) (Section 2.2.2.1) and gel electrophoresis respectively. Two crisp bands corresponding to 18S and 28S ribosomal RNA should be visible on the agarose gel with the 28S band approximately twice as intense as the 18S band (Figure 2.6).

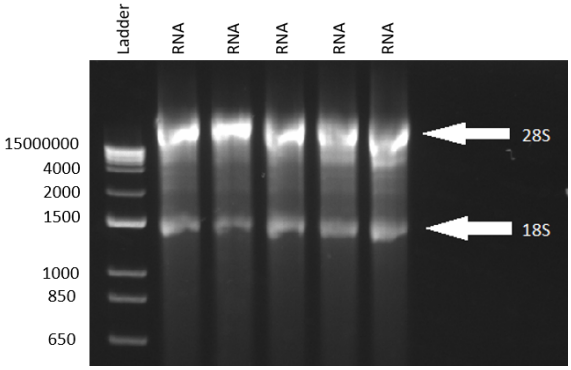



**Figure 2.6:** Bands on an agarose gel corresponding to 28S and 18S ribosomal RNA.

| **Reagent** | **Volume ($\mu$l)** |
|---|---|
| 10X RT Buffer | 2 |
| 25X dNTP Mix (100mM) | 0.8 |
| 10X RT Random primers | 2 |
| MultiScribe Reverse Transcriptase | 1 |
| Rnase Inhibitor | 1 |
| dd$H_2O$ | 3.2 |
| Dnase-treated RNA (10ng/$\mu$l | 10 |

**Table 2.18:** High Capacity cDNA reverse transcriptase kit master mix.

### 2.5.2 Reverse transcription of messenger RNA (mRNA) into complementary DNA (cDNA)

Prior to reverse transcription, RNA was treated with DNaseI (ThermoFisher) to remove genomic DNA which can confound downstream analysis. RNA was normalised to 1100ng/8$\mu$l in RNase-free $H_2O$ and added to 1$\mu$l 10X reaction buffer with $MgCl_2$ and 1$\mu$l DNaseI (ThermoFisher) in a 96-well PCR plate. The reaction was incubated at 37°C for 30 minutes to digest DNA, followed by the addition of 1$\mu$l 50mM EDTA and incubated at 65°C for 10 minutes to terminate the reaction. cDNA was synthesised using the High Capacity cDNA reverse transcriptase kit with RNase inhibitor (ThermoFisher) according to the manufacturer's instructions. The reverse transcription reaction mix and thermal cycling conditions are listed in Tables 2.18 and 2.19 respectively.

| Cycles | Time | Temperature (°C) |
|---|---|---|
| 1 | 10 min | 25 |
| 1 | 120 min | 37 |
| 1 | 5 min | 85 |

**Table 2.19:** High Capacity cDNA reverse transcriptase kit PCR conditions.

### 2.5.3 Real-Time quantitative PCR (RT-qPCR)

Real-Time quantitative PCR (RT-qPCR) measures the increase in fluorescent signal during PCR amplification, which is proportional to the total concentration of cDNA during each cycle. The PCR cycle at which the fluorescent signal rises above background levels ($C_t$) is used to quantify the starting concentration of mRNA (i.e. the higher the concentration of starting material, the lower the $C_t$ value). TaqMan gene expression assays were used for each assay (probes used in this study are listed in Table 2.20).

| Gene | ProbeID (ThermoFisher) |
|---|---|
| *hPOLE* | Hs00923952_m1 |
| *mPole* | Mm00448288_m1 |
| *hMSH6* | Hs00943000_m1 |
| *hMLH1* | Hs00179866_m1 |
| *hGAPDH* | Hs02758991_g1 |
| *mGAPDH* | Mm99999915_g1 |
| *hACTB* | Hs99999903_m1 |
| *mACTB* | Mm02619580_g1 |

**Table 2.20:** TaqMan Gene expression assay probes (20X).

A typical 20$\mu$l RT-qPCR reaction mix consisted of 10$\mu$l TaqMan Fast Universal PCR Master Mix (ThermoFisher), 50ng (0.5$\mu$l) cDNA, 1$\mu$l TaqMan Gene Expression Assay Mix (20x) and 8$\mu$l RNase-free $H_2O$ in a MicroAmp$^{TM}$ Fast Optical 96-Well Reaction Plate (ThermoFisher). Absolute fluorescence was quantified in real time on the ABI 7900HT cycler (Applied Biosystems) with automated $C_t$ threshold detecting using the conditions in Table 2.21.

Each sample was performed in duplicate (in triplcate where possible) and for each TaqMan gene expression assay probe, a no RT control and no template control were included to check for contaminating DNA and cDNA respectively.

Gene expression was calculated relative to an internal reference gene control (of-

| | Cycles | Time | Temperature (°C) |
|---|---|---|---|
| Initial denaturation | 1 | 20 sec | 95 |
| Denaturation | 40 | 1 sec | 95 |
| Annealing/Elongation | | 20 sec | 60 |

**Table 2.21:** TaqMan RT-qPCR thermal cycling conditions.

ten referred to as housekeeping genes) such as *GAPDH* or *β-actin* using the $\Delta\Delta C_t$ method, Livak *et al.* (230). Firstly, the $\Delta C_t$ value was calculated by subtracting the Ct of the internal reference gene from the $C_t$ of the gene of interest. The $\Delta\Delta$CT was calculated by subtracting the $\Delta C_t$ of the control (such as scrambled-transfected cells for siRNA gene knockdown or wildtype mice for relative gene expression). Gene expression relative to the control was quantified by using the equation $2^{-\Delta\Delta CT}$.

#### 2.5.3.1 Quantification of TaqMan probe efficiency

Reaction efficiency for pre-optimised TaqMan gene expression assay probes was quantified by serial dilution of cDNA from 50ng to 0.016ng (reducing 5-fold at each interval). A RT-qPCR reaction for each cDNA concentration was performed in triplicate using the gene expression probe of interest. $C_t$ values were plotted on the y-axis against cDNA concentration of the x-axis and a linear regression curve was plotted from the average $C_t$ value from each triplicate on a logarithmic scale. The slope of the standard curve was used to calculate the reaction efficiency using the following equation:

$$\text{Efficiency} = (10^{(1/\text{slope})}\text{-}1)$$

A reaction efficiency between 90-110% was deemed acceptable. An example standard curve for the *ACTB* (*β-actin*) probe is shown in Figure 2.7.

## 2.6 Analysis of protein

### 2.6.1 Protein purification and quantification

Adherent cells were washed with ice-cold PBS, dissociated from the flask or plate by manual scraping and resuspended in RIPA buffer (Tris-HCl 50 mM, NaCl 150mM,

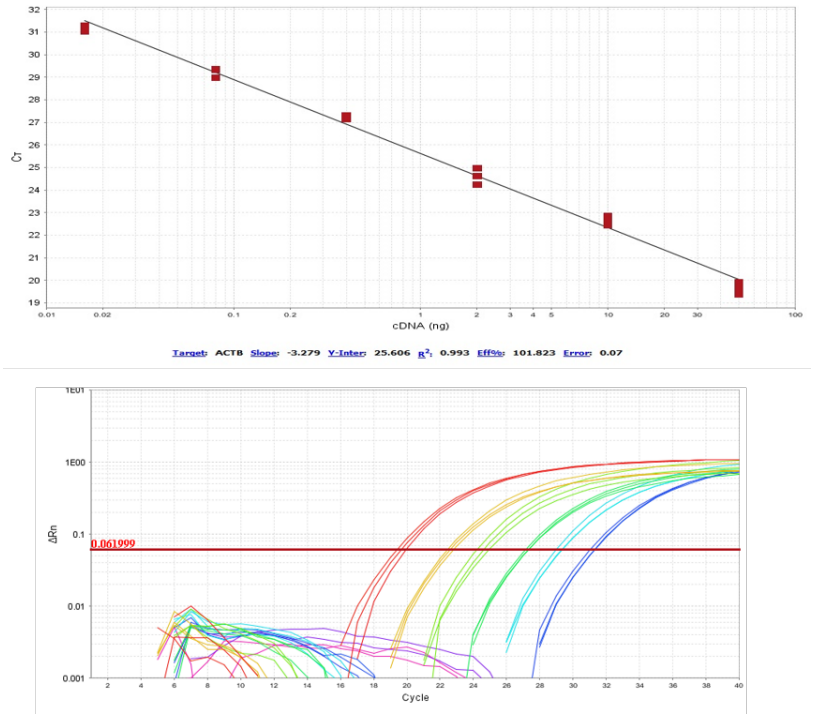



**Figure 2.7:** TaqMan *hACTB* reaction probe efficiency standard curve and amplification plot generated using QuantStudio Flex (ThermoFisher). The reaction efficiency is calculated to be 101.8%.

0.5% sodium deoxycholate, 1% Triton-X, 0:1% SDS, 5mM EDTA, pH 7.4) with 2X protease inhibitors (HALT$^{TM}$, ThermoFisher) and 10mM EDTA (ThermoFisher). Cells were lysed by incubating at 4°C for 30 minutes with regular vortexing. Cells were centrifuged for 20 minutes at 1,600g at 4°C to pellet cell debris and the supernatant containing protein lysate was aliquotted at 50$\mu$l in 1.5ml eppendorf tubes and stored at -80°C until use.

Protein concentration was quantified using the CB-X$^{TM}$ protein assay (G-Biosceinces®) following the manufacturer's instructions. The protein sample was diluted 1/10 with RIPA buffer in a total of 50$\mu$l, mixed with 1ml ice-cold CB-X buffer then centrifuged at 1,600g for five minutes. The supernatant was discarded and the pellet dissolved in 50$\mu$l CB-X buffer I and II. Finally 1ml of CB-X assay dye was added, vortexed then incubated for five minutes at room temperature. Absorbance was read at 595nm on the 96-well spectrophotometer plate reader (Molecular Devices) and the protein concentration ($\mu$g/$\mu$l) calculated from a standard curve (Figure 2.8).

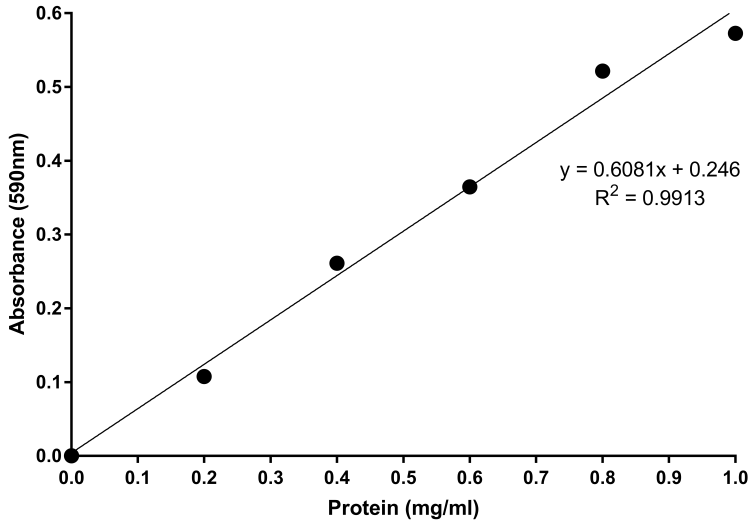



**Figure 2.8:** CB-X$^{TM}$ assay standard curve.

### 2.6.2 Western blotting

Protein lysate was mixed with 4x NuPAGE$^{TM}$ LDS sample buffer, 10X NuPAGE Sample Reducing Agent (ThermoFisher) and Milli-Q water then heated for 10 minutes at 70$^{o}$C to denature the protein. Denatured protein was run on a NuPAGE$^{TM}$ 4-12% Bis-Tris NuPAGE$^{TM}$ gel at 200 volts for 50 minutes alongside a Spectra$^{TM}$ Multicolor Broad Range Protein Ladder. 500$\mu$l antioxidant was added to the upper (cathode) buffer chamber to maintain proteins in a reduced state during electrophoresis. Protein was blotted onto a methanol-activated Polyvinylidene difluoride (PVDF) membrane (Immobilon P, Millipore, Burlington, Massachusetts) using semi-dry transfer at 20 volts for one hour.

Following protein transfer, the membrane was equilibrated in Tris-buffered saline, pH 7.6 (TBS) for five minutes, then blocked for one hour with blocking buffer (5% (w/v) milk (Marvel) in TBST (1x TBS, 0.1% Tween-20) with gentle rocking. The membrane was incubated with the appropriate primary antibody (Table 2.22) in blocking buffer for one hour at room temperature or overnight at 4$^{o}$C. Membranes were washed three times with TBST then incubated for one hour with the appropriate HRP-conjugated secondary antibody at 1:5000 dilution in blocking buffer, then washed with TBST. The membrane was incubated with Pierce$^{TM}$ Enhanced chemiluminescence (ECL) western blotting substrate (ThermoFisher) according to the

manufacturer's instructions. Chemiluminescence was detected using Amersham$^{TM}$ Hyperfilm$^{TM}$ (VWR) and developed using Xograph's Compact X4 processor.

| Primary Antibody | Manufacturer | Dilution | Secondary antibody | Manufacturer |
|---|---|---|---|---|
| hPOLE (93H3A) | ThermoFisher | 1:66 | Goat anti-mouse | DAKO |
| beta Actin (ab8227) | Abcam | 1:20,000 | Goat anti-rabbit | DAKO |
| MSH6 (44/MSH6 (RUO)) | BD Bioscience | 1:1500 | Goat anti-mouse | DAKO |
| MLH1 G168-728 (RUO) | BD Bioscience | 1:500 | Goat anti-mouse | DAKO |
| CHK1 (2G1D5) | CST | 1:1000 | Goat anti-mouse | DAKO |
| Phospho-CHK1 (Ser345) (133D3) | CST | 1:1000 | Goat anti-rabbit | DAKO |
| $\gamma$H2AX (Ser139) (20E3) | CST | 1:1000 | Goat anti-rabbit | DAKO |
| Apoptosis Western Blot Cocktail (ab136812) | Abcam | 1:250 | 100X HRP Conjugated 2$^{o}$ Antibody Cocktail | Abcam |
| HA-Tag (C29F4) | CST | 1:1000 | Goat anti-rabbit | DAKO |
| GFP (A-6455) | Invitrogen | 1:500 | Goat anti-rabbit | DAKO |

**Table 2.22:** Antibodies for western blotting.

### 2.6.3 Protein immunohistochemistry

Formalin-Fixed Paraffin-Embedded tissue sections (4$\mu$m) were de-waxed in xylene for 10 minutes, followed by rehydration through an ethanol series (70%, 90% and 100%) for two minutes each then distilled water for five minutes. Slides were placed in a pressure cooker with 10mM sodium citrate buffer (pH 6.0) for antigen retrieval. The slides were heated at full pressure for five minutes then gradually cooled for 20 minutes. Endogenous peroxidases within the tissue that could increase background staining were blocked in 3% hydrogen peroxide ($H_2O_2$) for 10 minutes.

Slides were washed twice with distilled water and once with TBST. Tissue was outlined with an ImmEdge Hydrophobic Barrier PAP Pen (Vector Laboratories, UK) and blocked with 5% normal goat serum in TBST for one hour within a humidified chamber. Blocking buffer was removed gently tapping the slide and the tissue was incubated with the appropriate primary antibody (Table 2.23) in 5% goat serum in TBST for two hours at room temperature followed by three washes in TBST with gentle rocking. Slides were incubated with a biotin-conjugated secondary antibody at 1:200 in TBST for one hour at room temperature.

Slides were washed three times with TBST for five minutes each, then incubated in ABC (Avidin/Biotinylated enzyme complex) (Vector Laboratories) for 30 minutes at room temperature then washed in dd$H_2O$. SIGMAFAST DAB (3, 3 - diaminobenzidine) solution (Sigma-Aldrich) was prepared and applied to the slide for 1-10 minutes and monitored microscopically for signal development. The reaction was

quenched in distilled water followed by staining of nuclear DNA with haemotoxylin for 15 seconds. Finally, tissue was dehydrated through an ethanol series and xylene for one minute each then mounted using DPX mounting media.

| Primary Antibody | Manufacturer | Dilution | Secondary Antibody |
|---|---|---|---|
| CD8 (SAB5500074) | Sigma | 1:100 | Goat anti-rabbit |
| CD45R RA3-6B2 (RUO) | BD Biosciences | 1:100 | Goat anti-rat |
| CD3 ab5690 | Abcam | 1:100 | Goat anti-rabbit |
| Ki67 (D3B5) | CST | 1:500 | Goat anti-rabbit |
| Cleaved Caspase-3 (AF835) | R&D Systems | 1:800 | Goat anti-rabbit |
| mCherry (ab167453) | Abcam | 1:200 | Goat anti-rabbit |

**Table 2.23:** Immunohistochemistry primary and secondary antibodies.

#### 2.6.3.1 Immunohistochemistry quantification

To assess proliferation in morphologically normal murine intestinal tissue samples, 4$\mu$m H&E-stained sections were viewed at 20X magnification on the Leica DM2500 optical microscope and imaged using the LasX software (Leica, Wetzlar). The percentage on Ki67-positive cells in 100 small intestine crypt-villus units was quantified from anonymised samples, using ImageJ software.

To assess apoptosis in morphologically normal murine intestinal tissue samples, 4$\mu$m H&E-stained sections were viewed at 20X magnification on the Leica DM2500 optical microscope and imaged using the LasX software (Leica). The total number of Cleaved Caspase-3 positive cells in 100 small intestine crypt-villus units were quantified from anonymised samples, using ImageJ software.

### 2.6.4 Visualisation of endogenous fluorescent proteins in mouse tissue

#### 2.6.4.1 Tissue embedding

Freshly dissected tissue was placed in the base of a cryomould and coated with Optimal Cutting Temperature (OCT) compound (VWR, Radnor, Pennsylvania). The cryomould was placed in Isopentane in a beaker and immersed in liquid nitrogen to

permit rapid, even freezing of the embedded tissue. Moulds were wrapped in aluminium foil to preserve fluorescence and stored at -80°C until sectioning.

#### 2.6.4.2 Fresh-frozen tissue sectioning

OCT-embedded murine tissue was equilibrated in the Leica CM3050 S research cryostat for 30 minutes at the optimum temperature for the specific tissue type (intestine and brain at -20°C and uterus at -15°C). Tissue was sectioned at 4$\mu$m, adhered to a Polysine slide (VWR) and stored at -80°C.

#### 2.6.4.3 Visualisation of fluorescent protein in fresh-frozen tissue

Prior to visualisation of fluorescent protein, slides were equilibrated to room temperature then washed in PBS to remove residual OCT compound. Slides were mounted with ProLong Diamond Antifade mountant with DAPI (ThermoFisher) and left to cure for 24 hours at room temperature in the dark. Fluorescent signal was detected using an Olympus BX-51 microscope and images were collected using Cytovision Genus v7.1 software (Leica).

## 2.7 Mouse husbandry and procedures

### 2.7.1 Animal husbandry

All mice were housed in The Wellcome Centre for Human Genetics, University of Oxford Functional Genetics Facility in individually ventilated cages with no more than six mice per cage. The environment was regulated with a 12 hour light dark cycle and mice were provided with water and normal chow with daily checks by animal technicians. All procedures were performed in accordance with UK Home Office regulations and the Animals (Scientific Procedures) Act 1986 under the project licence PB802BC1E and Personal licence I8E73D7F0.

### 2.7.2 Mouse lines

All but one strain used in the study has been previously reported (Table 2.24). Specific details of each mouse are included in Chapter four.

| Gene | Targeted construct | Nomenclature | MGI identifier | Reference |
|---|---|---|---|---|
| *Pgk*-Cre | transgene insertion 1 | $Tg(Pgk1\text{-}cre)^{1Lni}$ | MGI:2178050 | (231) |
| *Lgr5*-Cre | EGFP-IRES-creERT2 cassette inserted into exon 1 of *Lgr5* | $Lgr5^{tm1(cre/ERT2)Cle}$ | MGI:3764660 | (232) |
| *PR*-Cre | Cre recombinase inserted into exon 1 of the *PR* gene | $Pgr^{tm2(cre)Lyd}$ | MGI:3576366 | (71) |
| $Apc^{Min}$ | ENU-induced point mutation that alters nucleotide 2549 from a T to an A in *Apc* leading to a nonsense mutation at codon 850 | $Apc^{Min}$ | MGI:1856318 | (233) |

**Table 2.24:** Publicly available mouse strains used in this study. All had been backcrossed onto C57BL/6J for over ten generations.

Generation of the *Pole* L424V mouse and subsequent crosses will be described in Chapter Four (4).

### 2.7.3 Ear punches and genotyping

Ear punches, used for identification and DNA extraction were taken from pups after weaning (three weeks of age). No more than four ear punches were taken per animal. Genotyping of mice was performed by standard PCR using primers listed in Appendix Table A.4.

### 2.7.4 Breeding

Mice from six weeks of age were set up in breeding pairs or trios to obtain the desired genotype. Litter number was capped at six litters per mating pair or separated upon serial loss of litters for welfare purposes. Male breeders-only were used as experimental or control mice if required.

### 2.7.5 Induction of $ER^{T2}$ Cre-recombinase

Mouse models with the oestrogen receptor $ER^{T2}$ Cre recombinase require activation with tamoxifen. To activate $ER^{T2}$ Cre, 100$\mu$l tamoxifen solution (50mg tamoxifen, 0.5ml ethanol, 4.5ml corn oil) was administered by intraperitaneal injection every 24 hours over five consecutive days.

### 2.7.6 Necropsy

Mice were sacrificed at a pre-defined age, or when showing symptoms of distress, for example; weight loss; anaemia; hunching; anal/vaginal prolapse; distended abdomen and; visible external tumour. Mice were culled using the schedule 1 procedure, cervical dislocation followed by confirmation of permanent cessation of circulation.

#### 2.7.6.1 Gut and uterus preparation

The intestinal tract was dissected and divided into four sections, consisting of three equal lengths of the small intestine (proximal SB1, middle SB2 and distal SB3) and the colon and rectum. The intestine was flushed of debris with PBS and opened longitudinally onto filter paper with gut preparation apparatus (234). The intestines were pinned at either end to prevent dissociation from the filter paper and submerged in fixative.

The uterus was cut at the cervix and above each ovary to remove the uterus. The two uterine horns were separated and individually wrapped in tissue paper to prevent curving during fixation.

Tissue for downstream RNA and protein analysis were placed in cryovials, snap-frozen at necropsy in liquid nitrogen and stored at -80°C. All organs and tumours for histological analysis were fixed in 4% neutral buffered formalin (NBF) for 12-24 hours then stored in 70% ethanol until polyp counting 2.7.7 or processing 2.7.8.

### 2.7.7 Scoring of macroscopic phenotypes

Intestinal tissue was stained with methylene blue, which is absorbed by the intestinal mucosa, permitting visual analysis of colorectal polyps and adenocarcinomas. Formalin-fixed small and large intestine were submerged in methylene blue solution for three seconds, washed with PBS then examined under a stereoscopic dissecting microscope. Polyp burden was quantified, their diameter measured on their longest axis and categorised as follows: <1mm; 1-2mm; 2-3mm; or >3mm.

### 2.7.8 Tissue processing and embedding

Formalin-fixed intestinal tissue was rolled into swiss rolls with the epithelium on the interior. Uterine horns were cut laterally to expose the circular lumen. Other tissue was cut into cross sections. Tissue was placed in a histology cassette and processed in a Histomaster machine (Bavimed) using the following protocol:

| | |
|---|---|
| 70% ethanol | 45 minutes |
| 90% ethanol | 30 minutes |
| 100% ethanol | 30 minutes |
| 100% ethanol | 45 minutes |
| 100% ethanol | 45 minutes |
| Histoclear | 30 minutes x3 |
| Molton paraffin | 30 minutes |
| Molton paraffin | 45 minutes x2 |

Processed samples were embedded in hot paraffin wax using the paraffin embedding station (EG1150H, Leica) then cooled to -20°C until the wax was set.

### 2.7.9 Sectioning and histological staining of FFPE tissue

Formalin-Fixed Paraffin-Embedded tissue was sectioned using a microtome (RM2255, Leica) at either 10$\mu$m or 4$\mu$m for DNA extraction or histological staining/immunohistochemistry respectively. Tissue sections were adhered to a Superfrost® slide (VWR) then dried at 37°C. Tissue sections for DNA analysis were used immediately to reduce the risk of contamination.

For histological analysis, 4$\mu$m sections were baked at 60°C for one hour or overnight followed by five minutes in xylene, rehydration through an ethanol gradient (100%, 90% and 70%) for two minutes each then distilled $H_2O$ for two minutes. Slides were submerged in freshly filtered haemotoxylin for 30 seconds then washed briefly in running water. Slides were submerged in acid alcohol (70% ethanol, 1% 5M hydrochloric acid) to remove excess stain then washed in tap water for five minutes. Tissue was stained with eosin for two minutes, briefly dipped in distilled water, then dehydrated through an ethanol gradient for one minute each. Finally the slides were submerged in xylene for one minute then a coverslip mounted with DPX mounting medium.

## 2.8 Bioinformatics

All bioinformatic sequence analysis was aligned to Genome Reference Consortium Human Build 37 (GRCh37/hg19) genome assembly and performed using the a Unix terminal SmarTTY version 2.0 (http://smartty.sysprogs.com/).

### 2.8.1 BAM mapping and calling

Illumina adapters (Table 2.25) were removed from sequencing reads using Cutadapt version 1.13 (235). Paired reads were aligned with Burrows-Wheeler Aligner (BWA)-MEM and BAM files created using SAMtools. Variants were called using LoFreq software (236) with the relevant amplicon reference sequence(s) (the *HPRT1* sequencing reference sequence was chrX:133,584,044-133,644,829 - hg19).

| Illumina Adapter | Adapter sequence |
|---|---|
| 1 | AGATCGGAAGAGCACACGTCT |
| 2 | AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT |

**Table 2.25:** Illumina adapter sequences for *HPRT1* gene library preparation.

### 2.8.2 Variant annotation

Variant annotation was performed using the Ensembl VEP (Variant Effect Predictor) software (Table 2.26) by myself and Dr. Claire Palles, University of Birmingham, formerly University of Oxford. Non-default annotation databases are listed in Table 2.27.

### 2.8.3 Software

| Name | Use | Version | Citation |
|---|---|---|---|
| BWA-MEM | Genome alignment | 0.7.15 | (237) |
| SAMtools | Sort, index and generate BAM files | 1.2 | (237) |
| Platypus | Variant calling | 0.7.9.3 | (238) |
| LoFreq | Variant calling | 2.1.2 | (236) |
| VEP | VCF annotation | 87 | (239) |
| Cutadapt | Removing adapter sequences v | 1.13 | (235) |

**Table 2.26:** Software to process and analyse high throughput sequencing data.

| Description | Source | Date/Version |
|---|---|---|
| 1000G alternative allele frequency | www.1000genomes.org | 02/05/20 |
| ExAC alternative allele frequency | exac.broadinstitute.org | 0.3 |
| ESP6500 variant frequency | evs.gs.washington.edu/EVS | 01/07/2014 |
| UK10K variant frequency | www.uk10k.org | 22/07/2014 |
| COSMIC accession number | cancer.sanger.ac.uk/cancergenome/projects/cosmic | 10/02/2014 |
| Super duplication | genome.ucsc.edu/cgi-bin/hgTables | 17/11/2014 |
| Phylo-P score | genome.ucsc.edu/cgi-bin/hgTables | 26/06/2013 |
| Phastcons score | genome.ucsc.edu/cgi-bin/hgTables | 26/06/2013 |
| REPMASK class | genome.ucsc.edu/cgi-bin/hgTables | 27/11/2014 |
| HGMD PRO data | www.biobase-international.com/product/hgmd | 01/04/2014 |
| GERP score | mendel.stanford.edu/sidowlab/downloads/gerp | 01/05/2011 |
| ClinVar data | www.ncbi.nlm.nih.gov/clinvar | 03/02/2015 |

**Table 2.27:** VEP custom annotation tools.

## 2.9 Statistical analysis

Data were analysed using Student's $t$-test, Log-rank test, Chi-squared test, Fisher's exact test and two-tailed Mann-Whitney U test in Graphpad Prism version 7.04. The

one-way-ANOVA test was performed using R (RStudio, version 1.1.386). All $P$ values were two sided and $P \leq 0.05$ was considered statistically significant.

The Student's $t$-test was used to compare the means of two independent groups of samples, assuming a normal distribution. One-way and two-way ANOVA tests were used in Chapter four to compare the means two groups of data and perform multiple analyses of variance (MANOVA) in R. A two-way ANOVA was performed to test for an interaction between age or sex with with the independent and dependent variable (adjusted for multiple comparisons with Tukey's honestly significant difference (HSD) test). When the normality of the data was unknown a two-tail, non-parametric, Mann-Whitney U test was performed (DNA fibre analysis Chapter three).

The $\chi^2$ test was used to determine the fit between observed and expected data when analysing Mendelian ratios (Chapter four). The degrees of freedom were normally three. Fisher's exact test was used to compare categorical data to determine if there is a non-random association between two variables. Finally the Log-rank (Mantel-Cox) test was used for comparing survival data from two independent variables (two mouse genotypes in Chapter four).

# Chapter 3

# Functional characterisation of *POLE* exonuclease domain mutations in a cell-based system

## 3.1 Introduction

DNA Polymerase $\epsilon$ is the major leading strand replicative polymerase in eukaryotic cells. Its high nucleotide selectivity and intrinsic 3'-5' exonuclease activity are both critical to high fidelity DNA replication. Pol $\epsilon$ has been shown to play a role in cell cycle regulation and the repair of DNA damage (59, 72). As discussed in Chapter one, somatic *POLE* exonuclease domain mutations (EDMs) are associated with ultramutation (>100 mutations/Mb) in sporadic cancers of the colorectum, endometrium and brain, amongst others (1, 8, 157, 162). The postulate that this association reflects a causal relationship is supported by the demonstration that defects in Pol $\epsilon$ proofreading in *S. cerevisiae* and mice cause elevated mutation rates and tumour development in the latter (163, 240). Yet there are several important observations and discrepancies between the human cancers and existing experimental models that are currently

unexplained. For example, *POLE* exonuclease active site (D275/E277) mutations are rare in human cancer, possibly reflecting the observation that the mutation rate caused by cancer-associated *POLE* hotspot mutations (e.g. P286R and S459F) are considerably higher than those caused by loss of exonuclease activity alone (Figure 1.5) (163). While this suggests that there are additional mechanisms of mutagenesis beyond loss of proofreading, these remain elusive. Cancer-associated *POLE* exonuclease domain mutations are heterozygous and no evidence of second hits have been identified. Whilst this suggests a dominant negative effect on protein function this is as yet unproven in mammalian system.

Most *POLE* exonuclease domain mutant tumours are mismatch repair proficient although some cases have been reported, albeit infrequently (148). Loss of *POLE* exonuclease activity and mismatch repair is lethal or leads to growth defects in model systems due to a catastrophic increase in mutation rate (128, 129). The impact of cancer-associated *POLE* exonuclease domain mutations (e.g. P286R) with mismatch repair deficiency has not been studied in human cell lines.

### 3.1.1 Existing cellular models of *POLE* exonuclease domain mutations

Budding and fission yeast have been used extensively to study polymerase exonuclease domain mutations due to the ease of genetic manipulation and their relatively simple phenotypic characterisation (See Section 1.4.1 for more detail). However results from yeast studies are not always directly translatable to human cells due to the multitude of genetic and structural differences between species (indeed, the same applies to comparisons between budding and fission yeast). Recently, several studies have utilised higher eukaryotic cell models to study the phenotype and mechanism associated with *POLE* exonuclease domain mutations (139, 240, 241, 242, 243). I will discuss these models in the context of my project.

Isogenic cell models of *POLE* exonuclease domain and active site mutations were generated using the following methods: mutation knock-in using recombinant adeno-associated viruses (241); CRISPR-Cas9 nuclease genome editing (242, 243) or; purification of embryonic fibroblasts and embryonic stem cells from mouse models (139,

240) (Table 3.1).

The earliest quantification of mutation rate in *POLE* mutant mammalian cells was from an *Hprt1* forward mutation assay in mouse embryonic stem cells (mESCs) purified from heterozygous and homozygous *Pole* exonuclease active site mutant (exo-null) mice (139). Heterozygous *Pole* exo-null cells had a mutation rate approximately 4-fold above wildtype whilst in homozygous exo-null cells, the mutation rate was elevated 170-fold. All identified *Hprt1* mutations consisted of single base substitutions (139).

Mutations affecting the *POLE* exonuclease active site are rarely observed in human cancer. Mouse embryonic fibroblasts (MEFs) extracted from $Pole^{P286R/+}$ mice, harbouring the most common somatic cancer-associated *POLE* exonuclease domain mutation, had a significantly increased *Hprt1* mutation rate compared to wildtype MEFs. WGS of $Pole^{P286R/+}$ MEFs showed a mutation rate at least three orders of magnitude higher than wildtype MEFs (240). Consistent with human $POLE^{P286R}$ EDM tumours, mutations were predominantly base pair substitutions, with a high proportion of C→A, C→T and T→G mutations. However, there was less of a bias towards TCG→TTG and TTT→TGT mutations than in human *POLE* EDM cancers suggesting a difference in the mechanism of *POLE* exonuclease domain mutant-associated mutagenesis between mice and humans (1, 240).

Van Gool *et al.* introduced homozygous *Pole* exonuclease active-site (D275A/E277E) and cancer-associated exonuclease domain mutations (P286R, V411L and S297F) into mouse embryonic stem cells (mESCs), which significantly increased the *Hprt1* mutation rate compared to *Pole* wildtype cells (243). Consistent with the published literature, P286R was the strongest mutator with an 8-fold increase in mutation rate compared to wildtype, opposed to a 4-fold increase for D275A/E277E, V411L and S297F. However there was wide clonal variation in mutation rate particularly in S297F mutant cells, with approximately 100-fold difference in mutation rate between two clones (243). This study compares the mutation rate of homozygous cancer-associated mutations, however only heterozygous *POLE* exonuclease domain mutations have been detected in human cancer.

The mutation rate caused by heterozygous *POLE* exo-null variants was quantified in the microsatellite unstable (MLH1-deficient) human colorectal cancer cell line

HCT116 (241). The increase in *HPRT1* mutation rate in heterozygous *POLE* exo-null cells was approximately 4-fold compared to *POLE* wildtype cells. The number of single base pair substitutions in *HPRT1* was significantly higher than frameshifts in *POLE* exo-null cells. WGS validated the increased *POLE* exo-null mutation rate compared to wildtype cells, however the *POLE* exo- mutation rate was 5.8-fold higher than predicted by the *HPRT1* assay. When MLH1 activity was restored, the *POLE* exo-null mutation rate reduced to wildtype levels. From these data, the authors concluded that mismatch repair deficiency is a prerequisite for ultramutation in *POLE* exonuclease domain mutant cancers. It is important, however, to note that exonuclease active site mutations are seldom found in cancer (Section 1.5.3).

| **Cell line (Species)** | ***POLE* exonuclease domain mutations(s)** | **Limitations** | **Citation** |
|---|---|---|---|
| DT40 (Chicken) | Heterozygous and Homozygous D275A | Immortalised chicken cell lines | (242) |
| TK6 (Human B-cell) | Heterozygous and Homozygous D275A | Normal human B-cell line, exo-null *POLE* | (242) |
| HCT116 (Human) | Heterozygous D275A/E277A | *POLE* exo-null, mismatch repair deficient | (241) |
| MESCs (Murine) | Homozygous P286R, S297F, V411L, and D275A/E277A and Heterozygous S297F | Murine stem cells, homozygous cancer-associated *Pole* mutations | (243) |
| MEFs (Murine) | Heterozygous and Homozygous D275A/E277A | Murine cells, exo-null *Pole* | (139) |
| MEFs (Murine) | Heterozygous P286R | Murine cells | (240) |

**Table 3.1:** Published cell lines with *POLE* exonuclease domain mutations and their limitations.

These cell models have been used to test the sensitivity of *POLE* exonuclease domain mutant cells to chemotherapeutic agents and radiotherapy. Homozygous exonuclease domain mutant mouse embryonic stem cells were not differentially sensitive to commonly used chemotherapeutics, including: cisplatin (DNA crosslinking agent); doxorubicin (topoisomerase inhibitor); etoposide (topoisomerase inhibitor); methotrexate (dNTP synthesis inhibitor) and; paclitaxel (tubulin inhibitor), nor were they sensitive to ionizing radiation (243).

Nucleoside analogues were also tested based upon the hypothesis that increasing the mutation rate in ultramutator *POLE* exonuclease domain mutant cells could

push mutations over an "error threshold", reducing overall fitness, causing cell death (244, 245). Van Gool *et al.* and Tsuda *et al.* demonstrated that *POLE* exonuclease domain mutant mouse embryonic stem cells, chicken DT40 and human TK6 cells are sensitive to the nucleoside analogue cytarabine (242, 243). In addition, homozygous *POLE* exonuclease domain mutant mESCs were sensitive to the nucleoside analogue fludarabine compared to wildtype cells, although this difference was not statistically significant in heterozygous *Pole* S297F cells (243). Homozygous *POLE* exonuclease domain mutant mESCs were not differentially sensitive to the nucleoside analogues 5-fluorouracil (5-FU), gemcytabine, cladrabine or clofarabine (243).

Whilst these knock-in models have increased understanding of how these mutations affect untransformed cells and transformed cancer cells, they all suffer limitations which prevent their broader generalisation to cancers with spontaneous somatic *POLE* exonuclease domain mutations (Table 3.1).

### 3.1.2 CRISPR-Cas9

Zinc-finger nucleases (ZFNs), Transcription activator-like effector nucleases (TALENs) and most recently, Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Cas9 nuclease can all be used for genome editing, though the flexibility and economy of CRISPR-Cas9 has made it by far the most popular method in use. CRISPR-Cas9 endonucleases form part of a bacterial defence mechanism against invading phage and plasmid DNA (246). Its re-purposing as a genome editing tool involves the modification of *Streptococcus pyogenes* Cas9 (spCas9) to cause site-specific double-strand breaks (DSBs) at specific genomic loci by excising DNA immediately following a 5-NGG protospacer adjacent motif (PAM). Cas9 is guided by a single-stranded guide RNA (sgRNA) consisting of a 20 base pair homology sequence (protospacer) that binds to the target DNA region (247). The NGG PAM sequence occurs approximately every 8bp in the human genome, making it ideally suited for highly targeted genome editing (248). Cas9 endonuclease makes a DSB, activating intrinsic DSBR pathways. In proliferating human cells, 75% of DSBs are repaired by the error-prone, template-independent non-homologous end joining (NHEJ) pathway and the remaining 25% by template-dependent homology directed repair (HDR) (249). Us-

ing either a double stranded or single stranded oligonucleotide (dsODN or ssODN respectively) homologous to the target region, site-specific insertions, deletions or single base substitutions can be introduced via HDR (247).

The use of a 20bp sgRNA to target the endonuclease to the target region can cause off-target effects in areas of DNA with similar homology to the sgRNA. The D10A RuvC I domain mutant Cas9 nickase (Cas9n) causes a single-strand break which is preferentially repaired by high fidelity base-excision repair (250). A pair of offset sgRNAs complementary to opposite strands of the target site result in site-specific single strand breaks by Cas9n, mimicking a double strand break which can be repaired by the DSBR pathway but reducing off target effects by 50-100-fold (251, 252). Thus an important element in any experiment is to screen for off target effects.

### 3.1.3 Rationale for study and aims for chapter

The association of cancer-associated *POLE* exonuclease domain mutations with ultramutation has been observed in yeast, mice and knock-in mESCs, but a causal role of cancer-associated *POLE* mutations in ultramutation in human cells was unconfirmed. In addition, the mechanisms by which these mutations cause ultramutation and their impact on polymerase dynamics, DNA repair and concomitant MMR inactivation are currently unknown. My investigation aimed to study the factors contributing to the ultramutator phenotype and the molecular characteristics of human tumours in a clinically relevant human cancer cell model. Specific aims were as follows:

- Construct a $POLE^{\mathrm{P286R}}$ exonuclease domain mutant to wildtype revertant isogenic colorectal cancer cell line using CRISPR-Cas9n genome editing.
- Analyse CRISPR-derived isogenic cell lines for off-target effects associated with CRISPR editing.
- Quantify the mutation rate and spectrum of $POLE^{\mathrm{P286R}}$ and $POLE^{\mathrm{WT}}$ isogenic cell lines using the *HPRT1* forward mutation assay.
- Determine the impact of the $POLE^{\mathrm{P286R}}$ exonuclease domain mutations on DNA replication dynamics using whole-cell and single DNA molecular techniques.

- Determine whether knockdown of MMR proteins alters the viability or DNA damage response of *POLE*$^{P286R}$ mutant cell lines.
- Determine whether the *POLE*$^{P286R}$ mutation confers differential sensitivity to commonly used cytotoxics.

## 3.2 Methods

### 3.2.1 Design and generation of CRISPR-Cas9n mammalian expression vectors

The mammalian CRISPR-Cas9n expression plasmid, PX335, was kindly provided by Dr Ben Davies, Transgenics Core facility, WCHG (Figure 3.2a) (250). Key features of the plasmid include the CRISPR-Cas9 nickase gene, a guide RNA cloning site with hU6 promoter and a puromycin resistance cassette to select for positively transfected cells. I designed two offset single guide RNAs (sgRNAs) targeting the *POLE* c.857C→G variant using the open source Zhang lab CRISPR design tool (http://crispr.mit.edu/, accessed March 2014) (Figure 3.1a) (247). One sgRNA (sgRNA-1) was selected to overlap the c.857C→G mutation to minimise guide re-annealing once the mutation had been corrected. The other sgRNA (sgRNA-2) annealed 14 base pairs distal to sgRNA-1 (Figure 3.1).




**Figure 3.1:** Single guide RNA (sgRNA)-1 and sgRNA-2 sequences and binding sites at the *POLE* c.857C→G locus. The NGG PAM site and target number are labelled. The Cas9 nickase single strand break site is labelled with a blue arrow and the c.857C→G mutation is highlighted in red.

A single stranded oligonucleotide (ssODN), flanking the c.857C→G mutation by

ATTTTGTCCCCACAGGACCCTGTGGTTTTGGCATTTGACATTGAGACGACCAAACTGCCCCTCAAGTTTCC
TGATGCTGAGACAGACCAGATTATGATGATTTCCTACATGATCGATGGCCAGGTGAGCAGGTGGCTTCTG

**Table 3.2:** Single-stranded oligonucleotide (ssODN) repair template sequence.

70bp on either side containing the wildtype c.857C nucleotide, acted as the HDR repair template (Table 3.2). Because the cell lines are heterozygous for *POLE*$^{\text{P286R}}$, the wildtype allele may also act as an endogenous homology-directed repair template during double strand break repair.

Single gRNA-1 and sgRNA-2 were cloned into the two BbsI restriction sites of the PX335 plasmid (Figure 3.2b). Ligation of the sgRNA destroys the two BbsI restriction sites, therefore plasmids with integrated sgRNAs will not linearise upon a second BbsI restriction digest (Figure 3.2c). Selected plasmids were sequenced for integration of the correct sgRNA sequence (Figure 3.2d).

## 3.3 Results

Colorectal cancer cell lines with cancer-associated somatic *POLE* exonuclease domain mutations and a diploid karyotype were identified from a previous screen conducted in the host laboratory. Of 101 cell lines tested, five harboured heterozygous cancer-associated *POLE* exonuclease domain mutations (Table 3.3).

| Cell line | ***POLE* EDM** | **Somatic mutations (Exome)*** | **Age#** | **Sex#** | **% C:G→A:T*** | **MSI status#** | **Ploidy#** | **Publication** |
|---|---|---|---|---|---|---|---|---|
| HCC2998 | P286R | 14563 | NR | Male | 26.4 | MSS | 2.06 | NR |
| SNU-81 | P286R | 15469 | 53 | Male | 37.3 | MSS | 1.95 | (253) |
| HDC114 | A456P | NR | 62 | Male | NR | MSS | 2.045 | NR |
| HT115 | V411L | 8734 | NR | NR | 26.0 | MSS | 1.958 | NR |
| SNU-1040 | F367L | NR | 34 | Male | NR | MSS | 3.738 | (253) |

**Table 3.3:** *POLE* exonuclease domain mutant colorectal cancer cells lines. * Data from COSMIC cell lines database https://cancer.sanger.ac.uk/cell_lines (accessed 01/03/2019) (254), # Data from https://cellmodelpassports.sanger.ac.uk/passports/ (accessed 01/03/2019).

I decided to focus on the somatic *POLE*$^{\text{P286R}}$ mutated cell lines, HCC2998 and SNU-81, as this is the most prevalent somatic *POLE* exonuclease domain mutation found in cancer and is reliably associated with ultramutation, as discussed in Chapter

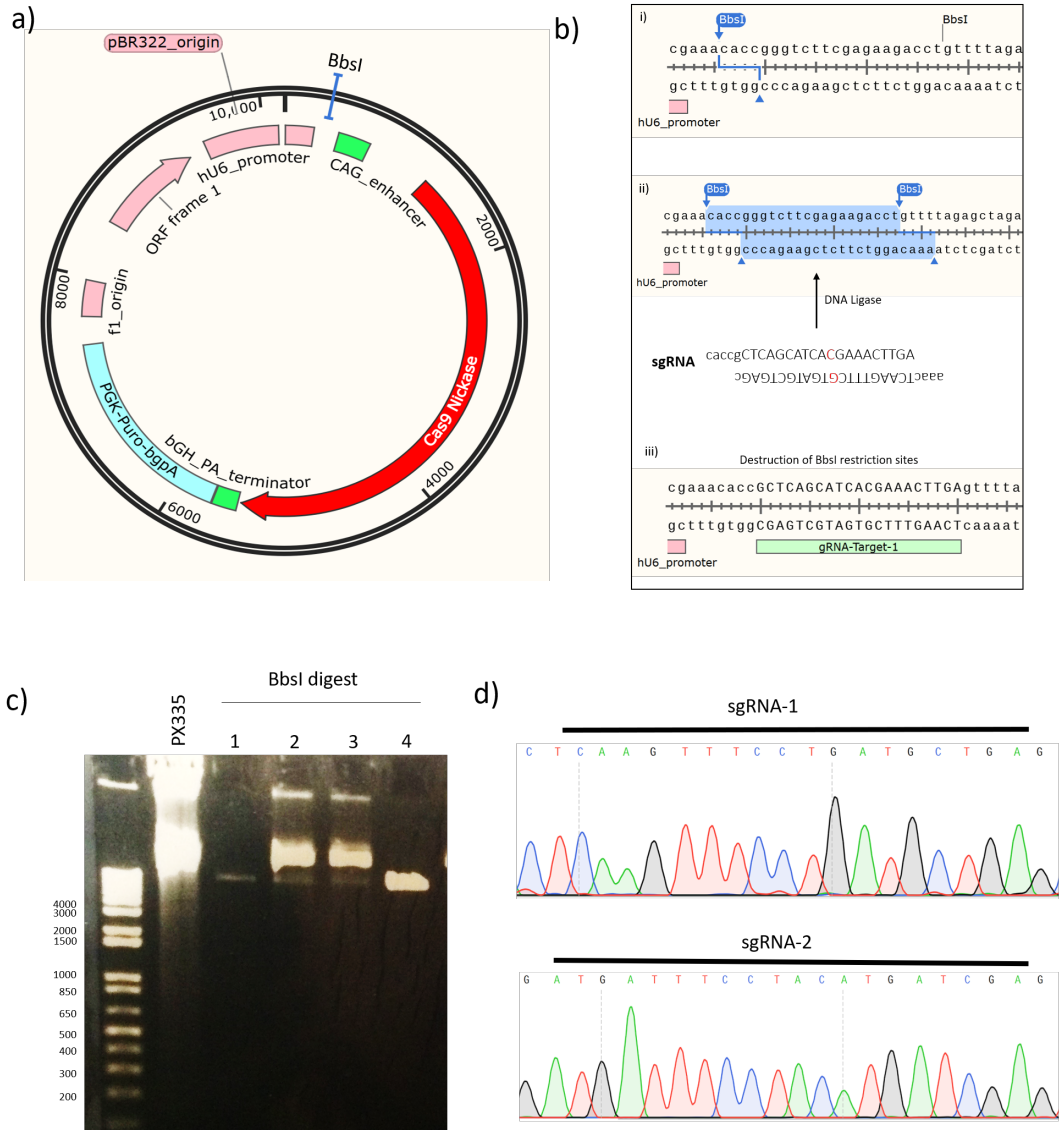



**Figure 3.2:** PX335 plasmid and cloning procedure. a) Commercial Cas9 nickase plasmid, PX335, kindly provided by Dr Ben Davis, Transgenics Core facility, WCHG, University of Oxford (250). The plasmid was adapted to include a puromycin resistance cassette to select for positively transfected cells. b) Schematic of sgRNA cloning procedure into the PX335 plasmid. i) Digest PX335 plasmid with BbsI restriction enzyme, to remove the sequence between the two labelled BbsI sites. ii) Ligate gRNA into the BbsI restriction sites next to the hU6 promoter. iii) The BbsI restriction sites are destroyed during sgRNA ligation. Ligated plasmids are transformed into competent cells to isolate single cloned plasmids. c) Purified plasmids are digested with BbsI and plasmids with integrated sgRNA (clones 2 and 3) will not digest. d) Chromatogram of integrated sgRNA-1 and sgRNA-2 in the PX335 plasmid, using sequencing primers listed in Appendix Table A.1. Figures generated using SnapGene software (from GSL Biotech; available at snapgene.com).

one. In both cases, the c.857C→G mutation causing the proline to argenine substitution at codon 286 was validated in-house by Sanger sequencing prior to its targeting.

To study the cellular and molecular phenotype conferred by the $POLE^{P286R}$ exonuclease domain mutation in clinically relevant isogenic cell models, I sought to correct the heterozygous c.857C→G (P286R) mutation to wildtype in HCC2998 and SNU-81 CRC cell lines. To reduce off-target effects of the Cas9 enzyme, I used the Cas9 double nickase (Cas9n) enzyme as described in the introduction. A 149bp ssODN repair template containing the wildtpye c.857C nucleotide was used as the homology-directed repair template.

### 3.3.1 Generation of isogenic *POLE* wildtype cell lines using CRISPR-Cas9n

$POLE^{P286R}$ mutant CRC cell lines, HCC2998 and SNU-81, were transiently co-transfected with the ssODN repair template and two PX335 plasmids (one containing sgRNA-1 and one containing sgRNA-2). A pEGFP-C3 control plasmid (Clontech) was used to visualise transfection efficiency, which was approximately 40% for HCC2998 and 30% for SNU-81, 24 hours following transfection (Figure 3.3). Transfected cells were selected with puromycin for 72 hours and live cells were plated at a limiting dilution in serum-rich medium to expand single cell clones. After 21 days, to permit sufficient growth, DNA was extracted from expanded clones using HotSHOT DNA extraction and exon 9 of *POLE* was screened for mutation correction by PCR amplification and Sanger sequencing with primers outside of the homology arm so as not to unintentionally amplify the repair template.

Sequencing of *POLE* exon 9 in HCC2998 and SNU-81 CRISPR-derived clones showed that 9.5% and 1.5% of clones respectively were corrected from $POLE^{P286R}$ to $POLE^{WT}$ (hereon referred to as Parental and CRISPR-WT respectively) (Table 3.4, Figure 3.4).

Several clones harboured insertions and deletions, consistent with the occurrence of error-prone NHEJ, however the majority of clones remained untargeted. The modest targeting frequency, 9.5% and 1.5% in HCC2998 and SNU-81 respectively may be due to a combination of limited transfection efficiency as demonstrated by green

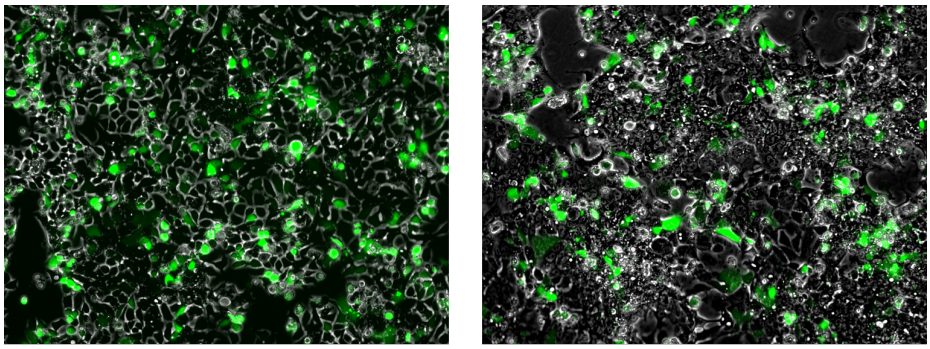



**Figure 3.3:** Green fluorescent protein expression as a visual representation of transfection efficiency. Images were acquired using the Nikon Eclipse TE2000U fluorescence inverted microscope on live cell cultures.

fluorescent protein (GFP) expression, the requirement for each cell to contain both sgRNA plasmids and the preference of NHEJ compared to homology-directed repair in cancer cells. If only one sgRNA plasmid is present in the cell, the single nick by Cas9n would be more likely to be repaired by high-fidelity base-excision repair.

| | **Number of clones (%)** | |
|---|---|---|
| **Cell line** | **HCC2998** | **SNU-81** |
| **Indels** | 8(19%) | 1(1.5%) |
| **c.857CG→CC (Corrected mutants)** | 4(9.5%) | 1(1.5%) |
| **Negative (mutant genotype)** | 30(71.5%) | 62(97%) |
| **Total clones screened** | 42 | 64 |

**Table 3.4:** Targeting efficiency of the CRISPR-Cas9 nickase in HCC2998 and SNU-81 cell lines.

### 3.3.2 Off-target effects

The -NGG PAM sequence occurs every 8bp throughout the human genome and CRISPR-Cas9 is thought to be able to tolerate up to five mismatches in the sgRNA protospacer (248). Although the use of the double Cas9-nickase and transient transfection is predicted to reduce the number of off-target mutations, given the potential for such events, I assessed predicted sites of sgRNA binding for evidence of mutagenesis.

Use of the Zhang-lab optimised CRISPR design tool (http://crispr.mit.edu) sug-

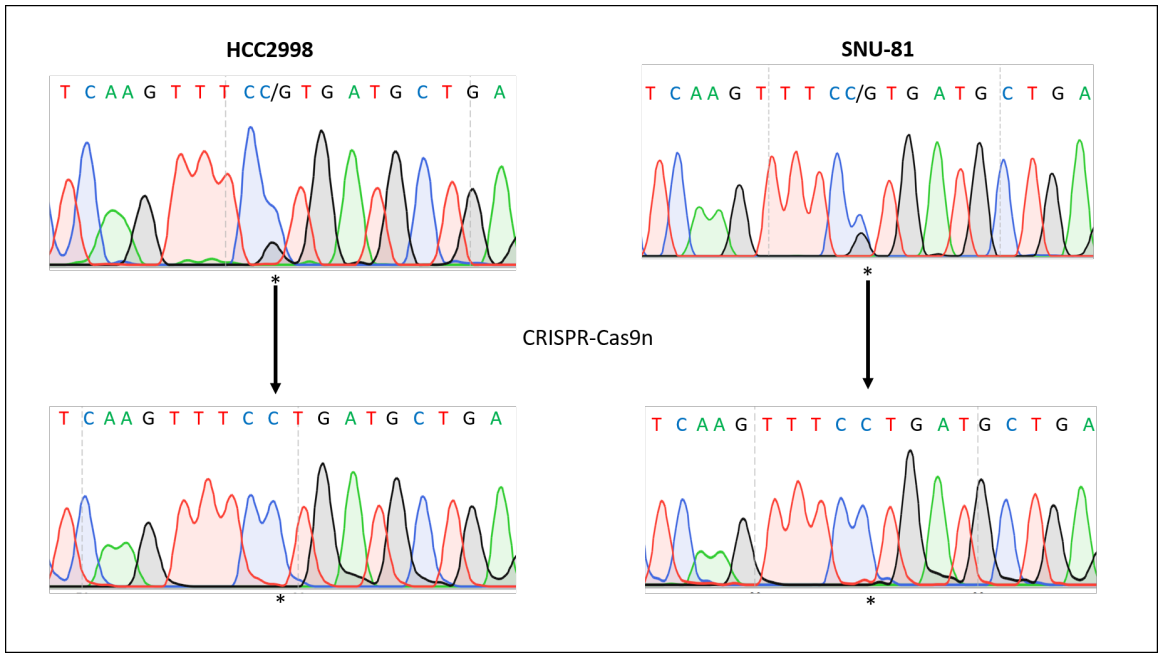



**Figure 3.4:** Chromatogram of the heterozygous *POLE* c.857C→G mutation and the corrected mutants before and after CRISPR-Cas9n editing respectively. The star denotes the corrected nucleotide.

gested that binding of both sgRNAs had no predicted off-target effects. Individually however, sgRNA-1 and sgRNA-2 had a predicted 119 and 277 off-target binding sites respectively with a minimum of three mismatched bases. I focused on potentially damaging off-target effects by pooling all hits within annotated genes. In total, 20 potential genic off-target effects were identified, 12 of which bound within exonic regions (Table 3.5).

I annotated each gene according to its function, intestinal expression and association with cancer. Three cancer-associated genes of interest were identified for their possible confounding of downstream results: *SYNE1*; *ARID2A*; and *BRCA2*. Sites of sgRNA binding in these genes were within exons. I therefore designed primers and performed Sanger sequencing of these regions to screen for mutations in CRISPR-WT isogenic cell lines. This analysis revealed no evidence of mutations in the off-target regions in any of these genes (data not shown).

| Position | Gene | Number of gRNA mis-matches | Gene position | Function | Cancer associated? | Expression in the intestine |
|---|---|---|---|---|---|---|
| chr6:152644636 | *SYNE1* | 4 | Exonic | Actin binding | Y | Yes |
| chr12:46246047 | *ARID2* | 4 | Exonic | Transcriptional regulation | Y | Low |
| chr13:32913611 | *BRCA2* | 4 | Exonic | DSB repair | Y | N/A |
| chr3:148457852 | *AGTR1* | 4 | Intronic | G-protein coupled receptor activity | N | Low |
| chr15:85657088 | *PDE8A* | 4 | Intronic | phosphodiesterase | N | Yes |
| chr22:36883717 | *FOXRED2* | 4 | Exonic | Oxidation Reduction | N | Low |
| chr17:56659091 | *TEX14* | 4 | Exonic | Cell division | N | NA |
| chr17:3962647 | *ZZEF1* | 4 | Intronic | Zinc finger nuclease | N | Yes |
| chr6:109321807 | *SESN1* | 4 | Exonic | DNA damage response | N | NA |
| chr20:18365773 | *DZANK1* | 4 | Intronic | Ankyrin repeats | N | Yes |
| chr5:56218903 | *MIER3* | 4 | Exonic | Transcription factor | N | Yes |
| chr1:113241352 | *MOV10* | 4 | Exonic | RNA helicase | N | NA |
| chr1:113241352 | *TNFSF11* | 4 | Exonic | Tumour necrosis factor | N | No |
| chr3:155304065 | *PLCH1* | 4 | Intronic | Signal transduction factor | N | Yes |
| chr2:102958846 | *IL1RL1* | 4 | Intronic | Immune response | N | Yes |
| chr6:46610096 | *CYP39A1* | 4 | Intronic | Lipid metabolism | N | NA |
| chr8:42147876 | *IKBKB* | 4 | Intronic | Phosphorylates | N | Yes |
| chr3:120047151 | *LRRC58* | 4 | Exonic | Leucine-rich repeat | N | Yes |
| chr6:90660768 | *BACH2* | 4 | Exonic | Transcriptional regulator | N | Yes |
| chr20:3995993 | *RNF24* | 4 | Exonic | Ion binding | N | Yes |

**Table 3.5:** Predicted off-target binding sites of sgRNA-1 and sgRNA-2 in within genic regions. Function and association with cancer were annotated using UCSC Gene Ontology (GO) Annotations (255). Gene expression data was collated from The Human Protein Atlas www.proteinatlas.org (256).

### 3.3.3 Large scale deletion in *POLE* isogenic cell lines

Interestingly, several CRISPR-derived clones from cells transfected without the ssODN repair template were corrected to $POLE^{\mathrm{WT}}$. It is possible, as mentioned above that the endogenous wildtype allele acted as the homology directed repair template. However, a SNP near the c.857C→G (P286R) mutation that was heterozygous in Parental cell lines was detected as homozygous in the Sanger sequence traces in CRISPR-WT cell lines. This suggested that a larger deletion within the $POLE^{\mathrm{P286R}}$ mutant allele may have occurred, rendering the CRISPR-WT isogenic cell lines hemizygous.

In collaboration with the Chromosome Dynamics Core facility at the WCHG I explored the possibility of a larger deletion by fluorescence in-situ hybridization (FISH) by using BAC probes targeting the chromosome 12 telomere, *POLE* and the chromosome 12 centromere to identify breaks and chromosomal rearrangements in and around *POLE* (Methods Table 2.3). First of all chromosome counts per cell in HCC2998 and SNU-81 Parental and CRISPR-WT cells were analysed from metaphase spreads. The majority of Parental HCC2998 cells were diploid (Figure 3.5), which is consistent with the reported ploidy in Table 3.1. The majority of HCC2998 CRISPR-WT clones were diploid, but a higher proportion of aneuploid cells (>50 chromosomes/cell) were found in CRISPR-WT clones 1 and 2 (72% and 44%

respectively) compared to Parental cell lines (29%), possibly as a result of CRISPR-Cas9n-induced breaks or a founder effect. In contrast, Parental SNU-81 cells were highly aneuploid and polyploid, with 44 to 151 chromosomes per cell (Figure 3.5). This is inconsistent with a ploidy of 1.95 previously reported (Table 3.1). The one CRISPR-WT clone remained highly aneuploid with 44 to 176 chromosomes per cell.

Cytogenetics, including FISH analysis, allows visualisation of large scale chromosomal gains, losses and rearrangements in single cells, which is not possible using short-read Sanger sequencing or even whole genome sequencing. Metaphase spreads from Parental cell lines and selected CRISPR-WT and uncorrected *POLE*$^{\text{P286R}}$ (CRISPR-MUT) CRISPR clones were analysed for binding of the three fluorescent probes on chromosome 12.

There was a reduction in *POLE* copy number in HCC2998 CRISPR-WT clones compared to the Parental and a CRISPR-MUT clone (Figure 3.6). Due to aneuploidy in SNU-81, a clear shift in *POLE* copy number was challenging to identify, although a small number of CRISPR-WT cells were found to contain only one copy of *POLE* compared to none of the Parental cells. Parental cells showed normal expression of the three BAC probes along chromosome 12 in 100% of cells analysed. The majority of CRISPR-WT clones contained at least one normal copy of chromosome 12 and at least one rearranged derivative with the absence of the *POLE* and the telomeric markers. The absence of *POLE* and telomeric probes in CRISPR-WT cell lines suggests a proximal deletion at chromosome 12q and the presence of a reduction in *POLE* fluorescent probe intensity in HCC2998 CRISPR-WT clone 2 is consistent with a deletion within the P286R mutant allele which affects probe binding (Figure 3.6). From these data I conclude the CRISPR-WT isogenic cell lines are hemizygous for *POLE*.

The *POLE* gene is located approximately 617.5kb proximal of the chromosome 12 telomere. Close telomeric proximity may increase the probability of large chromosomal breakage upon double strand breaks. The region distal to *POLE* contains 16 annotated genes which include: *GOLAG3*, a component of the gogi apparatus; *PGAM5*, a mitochondrial serine/threonine-protein phosphatase; *ANKLE2*, a mitotic regulator; *PXMP2*, Peroxisomal membrane 2 (pore-forming activity in the peroxisomal membrane); seven Zinc Finger Nucleases; and *CHFR* (Appendix Table B.1).

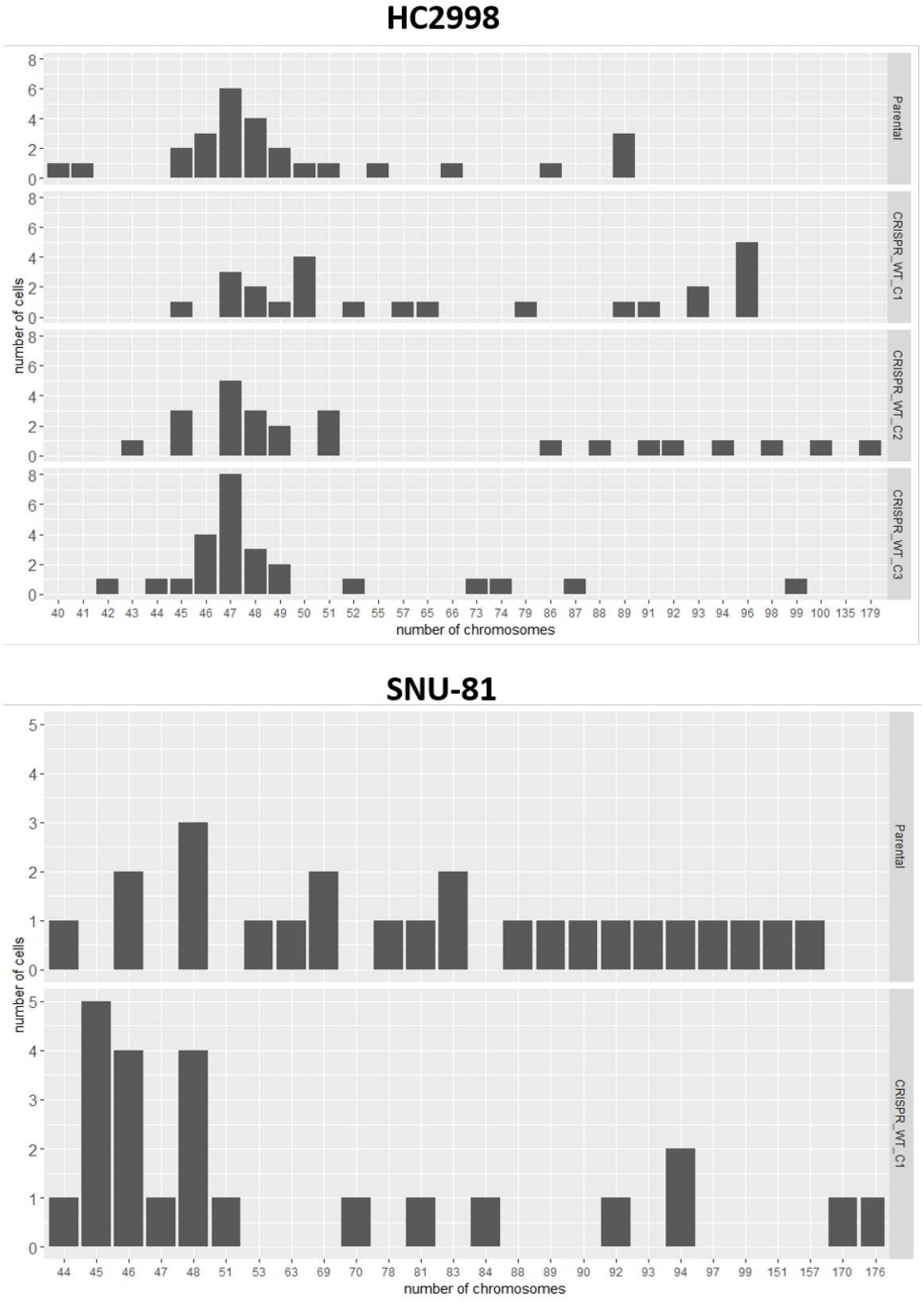



**Figure 3.5:** Chromosome counts of HCC2998 and SNU-81 Parental and CRISPR-WT clones from metaphase spreads.

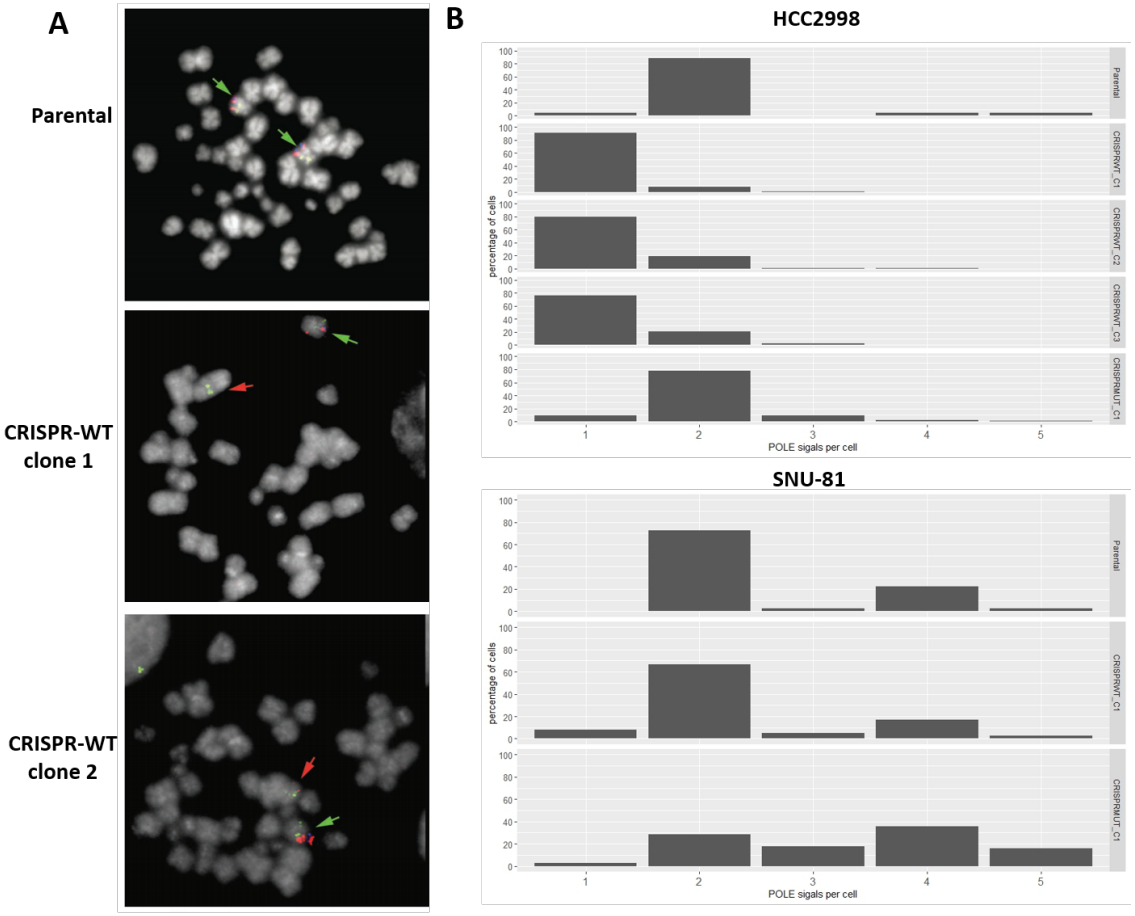



**Figure 3.6:** Cytogenetic analysis of isogenic cell lines A) Representative images of Fluorescent in-situ hybridization for *POLE* copy number analysis. *POLE* probe (RP11-148L11) localised in band 12q24.33 (hg19 chr12:133,038,915-133,243,361) is labelled red, Internal probe (RP11-25J3), localised in band 12q12 (hg19 chr12:75,892,543-76,047,496) is labelled green and telomeric Teloq probe (CTC221K18) localised in band 12q24.3 distal to *POLE* is labelled purple. Green arrows point to complete chromosome 12 containing all three probes and red arrows point to truncated/rearranged chromosome 12 lacking one or more probes. CRISPR-WT Clone 2 contains an truncated chromosome 12 with a weak red *POLE* signal, suggesting partial hybridisation of the probe due to a deletion within the CRISPR gRNA target region. b) *POLE* copy number in Parental, CRISPR-WT and CRISPR-MUT clones.

CHFR is involved maintenance of the cell cycle checkpoint into mitosis. Methylation-induced silencing has been reported in in 45% of colorectal cancer cell lines and 40% of primary colorectal in tumours and has been suggested to play a role in tumourigenesis (257). Furthermore, *Chfr*-knockout mice are mildly tumour-prone with 9% of developing lymphomas and epithelial tumours of the intestine, liver and lung by 40 weeks compared to 2% of controls (258). The potential implication of *CHFR* hemizygocity was taken into consideration in subsequent cell proliferation and cell cycle analysis.

### 3.3.4 POLE protein expression in isogenic cell lines

To ascertain whether loss of one *POLE* allele affected protein expression, I collected protein lysate from asynchronous *POLE* hemizygous CRISPR-WT and heterozygous Parental cells, then performed western blotting. Protein expression was quantified by pixel densitometry and normalised to the housekeeping protein $\beta$-actin based on two independent cell preparations. To permit interpretation of technical replicates, POLE protein levels in CRISPR-WT isogenic cell clones were normalised to the Parental cell line in each replicate.

POLE protein expression in CRISPR-WT isogenic cell lines was reduced relative to Parental cell lines (Figure 3.7), consistent with a reduction in *POLE* copy number. This reduction in protein level is taken into consideration in subsequent assays.

### 3.3.5 Mutation rate of Parental and CRISPR-WT isogenic cell lines

As discussed in the introduction, cancer-associated *POLE* exonuclease domain mutations are associated with tumour ultramutation and a characteristic 96-channel mutation signature (1, 123). It is unclear whether cancer-associated *POLE* exonuclease domain mutations are the sole driver of ultramutation in cancer. To determine this, I measured the spontaneous mutation rate in the Parental and CRISPR-WT isogenic colorectal cancer cell lines using the Hypoxanthine Phosphorybosyl Transferase (*HPRT1*) forward mutation assay.

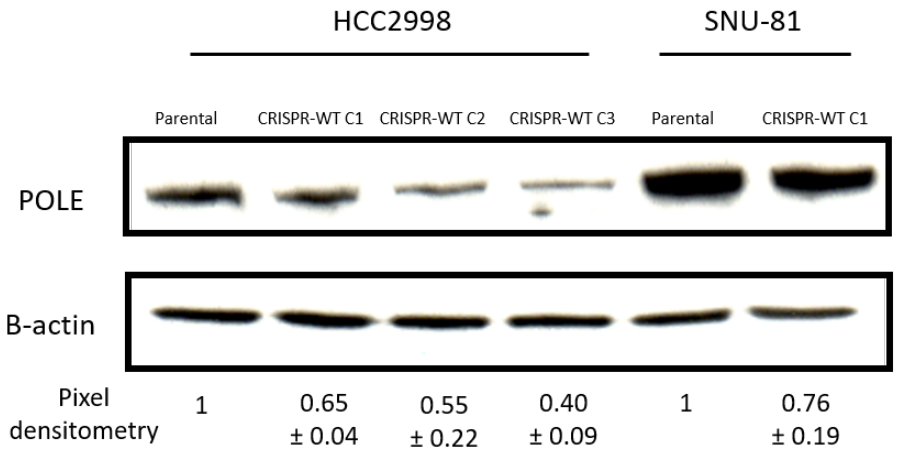



**Figure 3.7:** Representative western blot of POLE protein levels in Parental and CRISPR-WT isogenic cell lines. Pixel density for each sample was quantified using ImageJ and normalised to loading control $\beta$-actin and expressed as a proportion of Parental POLE expression (mean±SD from two replicate experiments).

HPRT1 is involved in purine nucleotide synthesis within the purine salvage pathway (Figure 3.8). HPRT1 metabolises the guanine analogue 6-thioguanine (6-TG) into toxic 6-thioguanine monophosphate, which is incorporated into DNA and RNA during genome replication and transcription. This acts as a replication blockade, resulting in replication fork stalling and cell death. Cells with loss-of-function (LOF) mutations in the *HPRT1* gene are resistant to 6-TG. The *HPRT1* gene is located on the X-chromosome, permitting detection of loss of function mutations in male cell lines heterogametic for sex chromosomes. Both HCC2998 and SNU-81 were derived from male patients with colorectal cancer (Table 3.1.).

The spontaneous mutation rate of *HPRT1* in *POLE* isogenic cell lines was determined using the method described in Glaab *et al.* (221) as follows. First, cells were exposed to Hypoxanthine-Aminopeterin-Thymidine (HAT) medium for five passages (15 days) to eliminate pre-existing *HPRT1* mutant cells. Aminopterin blocks the de-novo dNTP synthesis pathway, leaving the cells reliant on the purine salvage pathway. In cells with *HPRT1* loss of function mutation, both dNTP synthesis pathways are blocked, reducing dNTP pools, resulting in replication fork stalling and cell death. Secondly, cells were cultured in Hypoxanthine-Thymidine (HT) medium for two passages allowing them to recover, followed by five passages in normal growth medium to permit sufficient *HPRT1* mutation accumulation.

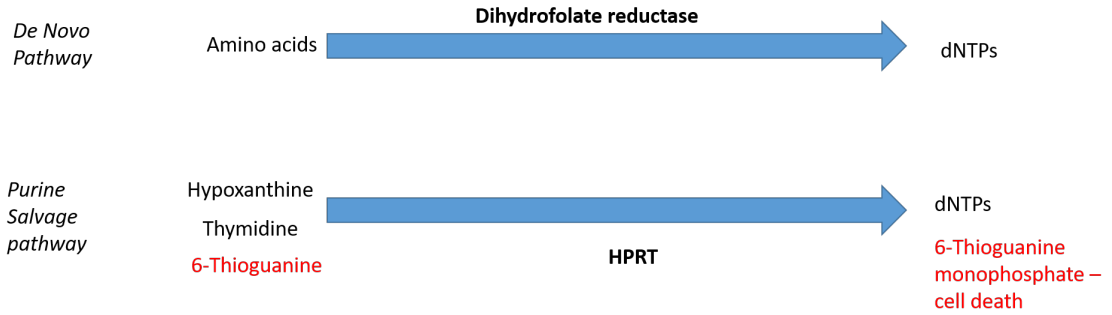



**Figure 3.8:** dNTP synthesis pathways and *HPRT1* assay mechanism. *de novo* synthesis of nucleotides is performed by dihydrofolate reductase (DHFR) and purine salvage is performed by Hypoxanthine Phosphoribosyltransferase 1 (HPRT1). To select for cells proficient for HPRT1 activity, cells are exposed to aminopterin which blocks DHFR selecting for *HPRT1*-proficient cells survive. To select for cells with *HPRT1* loss of function mutations, cells are exposed to 6-Thioguanine, which is metabolised by HPRT1 to create the toxic nucleoside analogue 6-thioguaninie monophosphate. Only cells with *HPRT1* loss of function, unable to metabolise 6-TG, will form viable colonies.

Finally, cells were plated at a density of 85,000 (HCC2998), and 150,000 (SNU-81 and HCC2998 CRISPR-WT clone 3) cells per 10cm dish, in ten dishes then cultured in normal growth medium, supplemented with 20% foetal bovine serum and 40$\mu$M 6-TG, for 21 days for *HPRT1* mutant colonies to form. In parallel, 500 cells per 10cm dish were plated in triplicate in normal growth medium, supplemented with 20% foetal bovine serum, to determine the plating efficiency (colony-forming ability) of each cell line. Colonies were fixed in methanol and stained with 0.1% crystal violet in PBS. Colonies >20 cells in diameter were scored and totalled (Figure 3.9).

To test this methodology, I performed a pilot *HPRT1* forward mutation assay on MLH1-deficient HCT116, MLH1-complemented HCT116 and MMR-proficient SW480 cell lines for which *HPRT1* mutation frequencies have previously been published in the assay setting (221).

The pilot study showed a 2-fold reduction in mutation rate in MLH1-complemented HCT116 compared to MLH1-deficient HCT116 (11.9 x $10^{-5}$ and 22.5 x $10^{-5}$ respectively) (Figure 3.10). The mutation rate of MLH1-complemented HCT116 reduced 16-fold compared to MLH1-deficient HCT116 in Glaab *et al.* (221). As expected, the MMR-proficient SW480 cell line exhibited a low mutation rate (0.26 x $10^{-5}$) with only one 6-TG-resistant cell out of one million cells plated (Figure 3.10).

Although the fold-difference in mutation rate in the pilot study differed from the

published literature, the results demonstrated the utility of this assay to capture the increase in mutation rate caused by clinically relevant perturbations to DNA-damage repair pathways and provided confidence for its use to examine Parental and CRISPR-WT isogenic cell lines.

I proceeded therefore to perform the *HPRT1* assay on the *POLE* isogenic cell lines. Plating efficiency ranged from 10-41%. Overall, SNU-81 Parental and CRISPR-WT clones had a lower plating efficiency than HCC298 Parental and CRISPR-WT clones ($P = 0.033$, Student's $t$-test) (Figure 3.9).

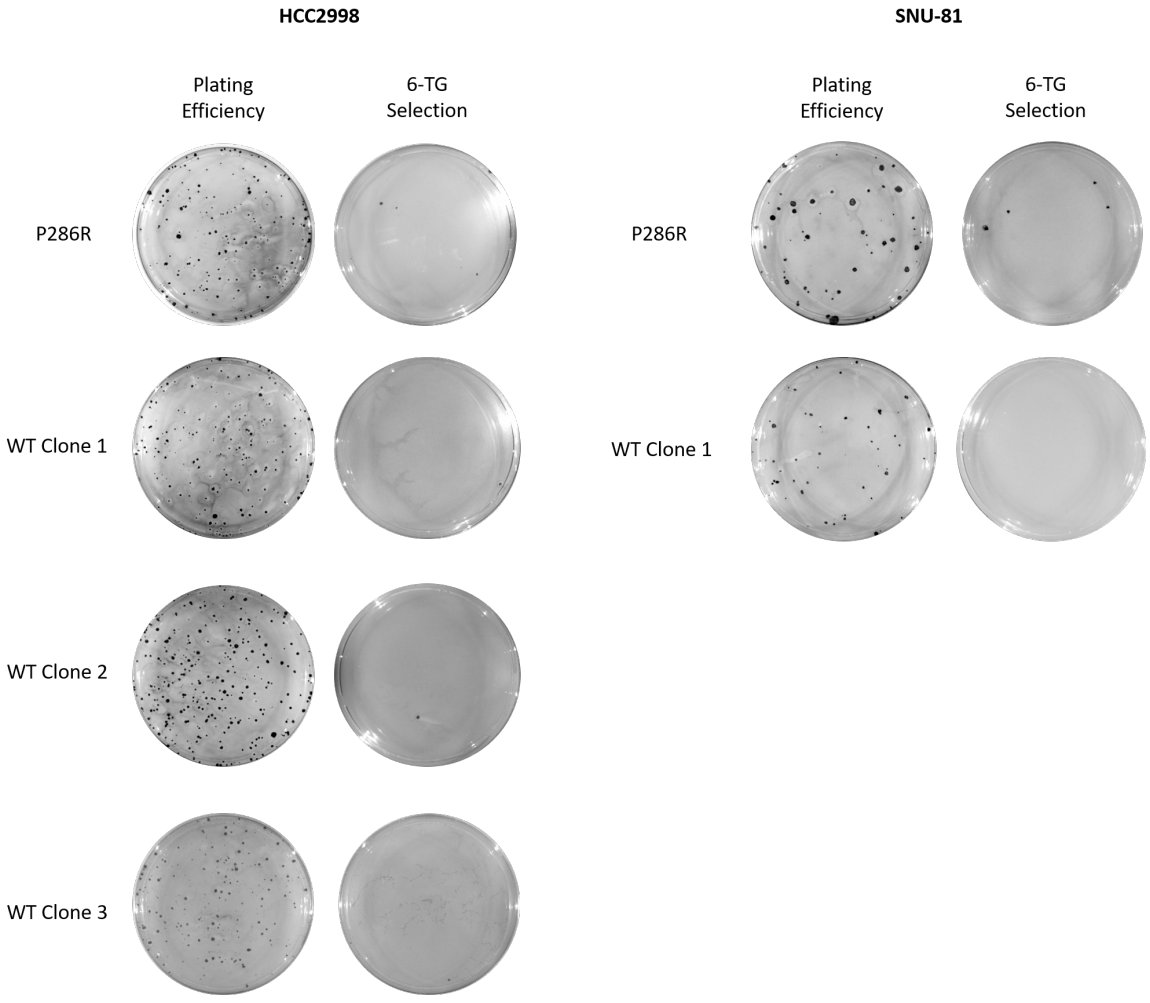



**Figure 3.9:** Representative images of plating efficiency and 6-TG selection plates for each cell line. Individual colonies over 20 cells in diameter were counted. At least two independent replicates were performed for each cell line.

In both the HCC2998 and SNU-81 cell lines, CRISPR-WT isogenic cell lines exhibited a striking reduction in mutation rate compared to the Parental cell line (Figure

3.10). These data confirmed the causal role of cancer-associated *POLE* exonuclease domain mutations in tumour ultramutation and suggest that any contributions of other secondary mutations to mutagenesis are at best modest.

Perhaps interestingly, SNU-81 Parental cell lines exhibited a 4-fold higher mutation rate than HCC2998 Parental cells. Speculatively, the increased mutation rate in SNU-81 cells may be due to an elevated $POLE^{P286R}$ copy number compared to the diploid HCC2998 cell line (Table 3.6).

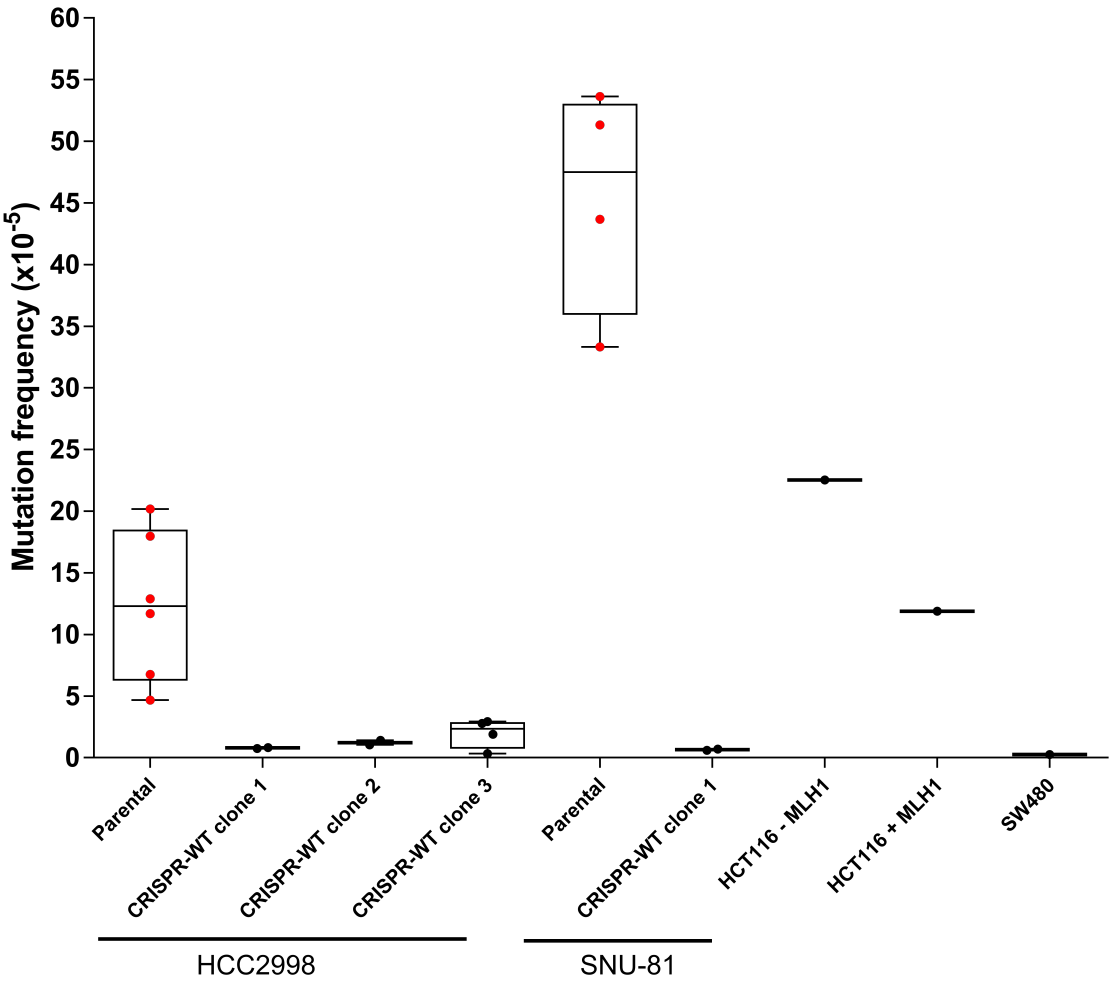



**Figure 3.10:** *HPRT1* mutation rate in HCC2998 and SNU-81 isogenic cell lines. Mutation rate was plotted for each replicate. Mean, interquartile range and minimum and maximum points are represented in the box plot. Mutation rate from the pilot study on HCT116 plus and minus *MLH1* and SW480 cells are included for comparison.

#### 3.3.5.1 Mutation spectrum of *HPRT1* mutant cell lines

Having confirmed the causal role of *POLE* exonuclease domain mutation in tumour ultramutaion, I next sought to determine whether it is also responsible for the characteristic mutation signature detected in human cancers carrying this variant (Signature 10) (Figure 1.6) (1, 158).

To do this, I first carried out an *in silico* analysis in collaboration with Bioinformatician Dr Roland Arnold to identify whether any of the single nucleotide variants in a trinucleotide context overrepresented in *POLE* Signature 10 could result in non-synonymous mutations within the *HPRT1* gene. Using an input of Signature 10 trinucleotide base pair substitutions (TCT>TAT, TCG>TTG and TTT>TGT) in 5'-3' and 3'-5' direction (Table 3.6), we designed a Python script to extract all possible *POLE* signature mutations from the *HPRT1* hg37 reference sequence.

| Trinucleotide | Ref | Alt |
|---|---|---|
| TCT | C | A |
| TCG | C | T |
| TTT | T | G |
| AGA | G | T |
| CGA | G | A |
| AAA | A | C |

**Table 3.6:** Single nucleotide variants in the Signature 10 trinucleotide context (TCT>TAT, TCG>TTG, TTT>TGT) in the 5'-3' and 3'-5' direction. Ref is the reference nucleotide and Alt is the mutant nucleotide.

I manually filtered variants from the *in silico* analysis to include only non-synonymous mutations, resulting in 51 possible missense and nonsense variants. Using Variant Effect Predictor (Methods Table 2.26), 40 mutations were predicted as deleterious or damaging by SIFT/PolyPhen scores (218, 219) and/or reported in a prior publication to be associated with Lesch-Nyhan syndrome, an *HPRT1*-deficiency disorder (discussed below) (http://www.lesch-nyhan.org, accessed 12/12/2018). After confirmation of possible loss of function variants in *HPRT1* as a result of Signature 10 mutations, I sequenced bulk HAT-treated cells and individual *HPRT1* mutant clones to determine the *HPRT1* mutation spectrum acquired in the mutation accumulation phase. A schematic of the protocol is depicted in Figure 3.11.

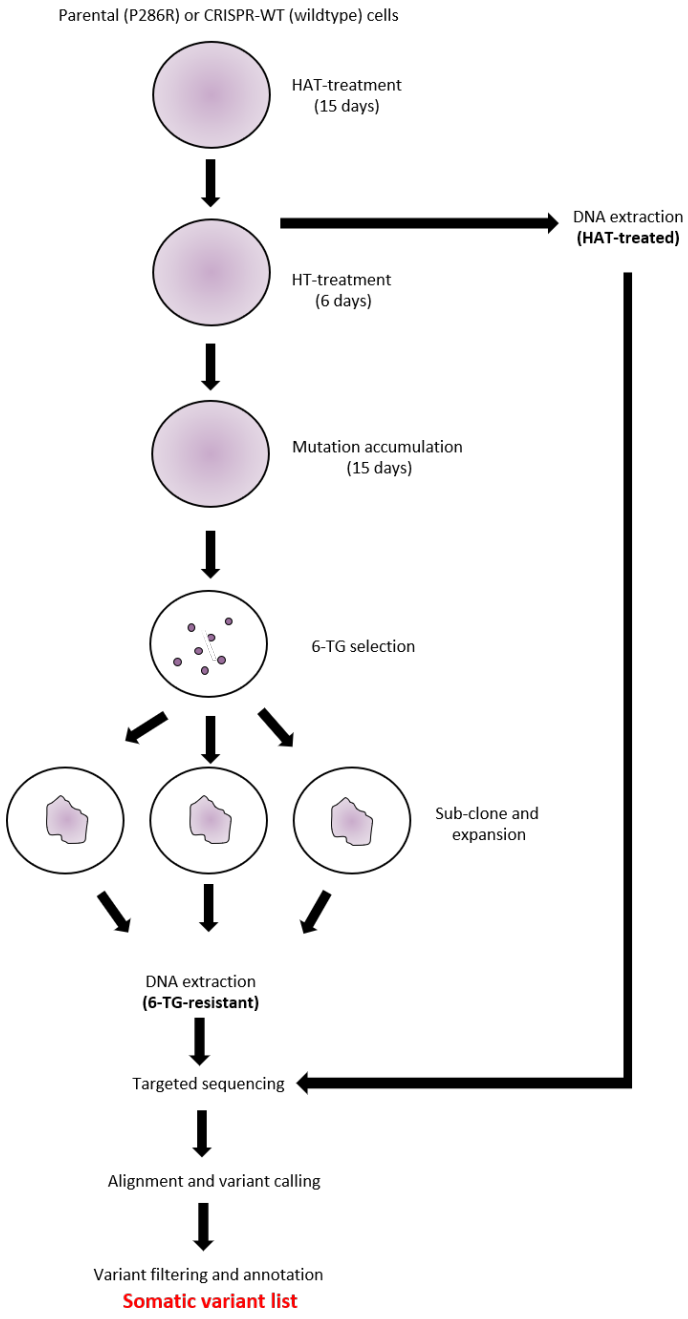



**Figure 3.11:** Schematic of the *HPRT1* sequencing protocol. Parental and CRISPR-WT cells underwent HAT-treatment to cleanse the population of pre-existing *HPRT1* mutant cells. Following HAT-treatment 2x$10^6$ cells were centrifuged and genomic DNA extracted for use as a control to for SNPs and pre-existing clonal *HPRT1* mutations. The remaining cells underwent HT-treatment, mutation accumulation and 6-TG selection as stipulated in Section 3.3.5. After 21 days of 6-TG selection, single 6-TG-resistant clones underwent clonal expansion for approximately 7-10 days followed by genomic DNA extraction. The *HPRT1* gene was sequenced in 6-TG-resistant cell clones and the HAT-treated bulk cell population followed by variant calling using LoFreq (236) and manual filtering.

To do this, I selected individual *HPRT1* mutant colonies (resistant to 6-TG), expanded them in medium with 40$\mu$M 6-TG and extracted genomic DNA for high throughput targeted sequencing. The rationale underpinning targeted DNA sequencing of gDNA rather than cDNA was to reduce the risk of missing truncating variants associated with nonsense-mediated decay, and to enable the detection of splice site variants and mutations downstream of nonsense mutations. Using Primer 3 (226), I designed custom primers to amplify the 9 exons of the *HPRT1* gene, with the primers for *HPRT1* exon 1 from Hinoi *et al.* (259) (Appendix Table A.2).

The primer design incorporated the requirement to include the Illumina adapters for paired-end sequencing - an important consideration for accurate sequence alignment, sample barcoding for multiplexing and the detection of indels (Table 3.7). *HPRT1* amplicon libraries were prepared using the protocol in Methods Section 2.3.

| | |
|---|---|
| Forward adapter | ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Reverse adapter | AGACGTGTGCTCTTCCGATCT |

**Table 3.7:** Illumina adapter sequences complementary to the adapter sequences on the Miseq Nano V2 flow cell, permitting clonal amplification of the DNA molecule into clonal clusters for sequencing.

PCR errors were minimised by using the Roche® FastStart High Fidelity (HiFi) Taq (Sigma) with 3'-5' exonuclease activity and a maximum of 30 cycles of PCR amplification, sufficient to enable amplicon visualisation. DNA bands were visualised on an agarose gel with ethidium bromide to confirm amplification and amplicon size (Figure 3.12). Amplicon intensity was visually graded on a scale of 1-4 from strongest to weakest and pooled according to grade (grade 1 = 1$\mu$l, grade 4 = 4$\mu$l). Each colony was individually barcoded, samples were pooled and gel extracted to remove primer dimers and surplus barcodes which may reduce sample read depth.

The quality of the final library was assessed using the TapeStation 2200 (Agilent). Results showed two major peaks at 326bp and 411bp which corresponds to amplicon size across the *HPRT1* library (Figure 3.13). There was no evidence of low molecular weight DNA, confirming the majority of small contaminating DNA molecules had been eliminated during library preparation (Figure 3.13).

Libraries were diluted to 4nM and the volume of library adjusted so that each amplicon was predicted to be sequenced to a minimum of 1000X read depth. Libraries

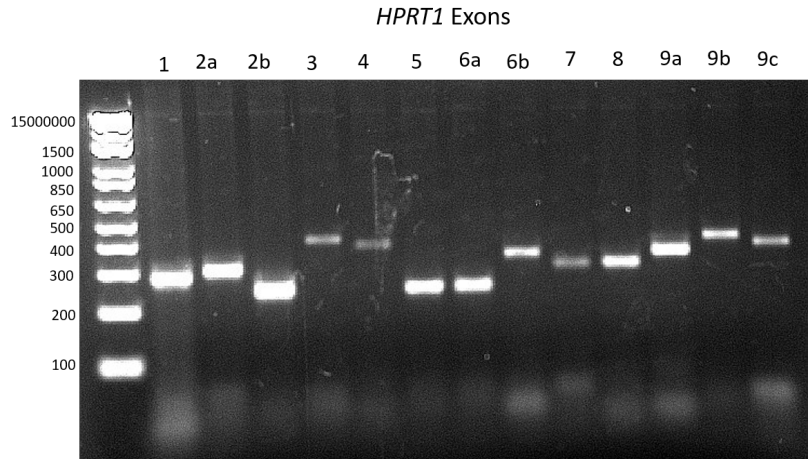



**Figure 3.12:** Agarose gel with ethidium bromide demonstrating successful amplification of the *HPRT1* exons for the MiSEQ NGS library. The first lane is a 1kb Plus ladder (ThermoFisher) to assess the size of each amplicon.

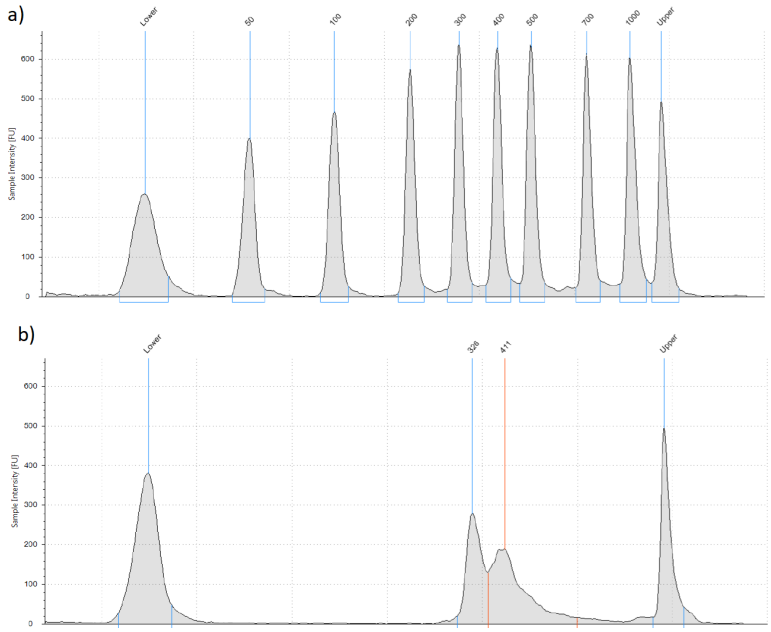



**Figure 3.13:** TapeStation plots for the *HPRT1* library. The upper plot is the standard ladder and the lower plot is the *HPRT1* DNA library. Two major peaks at 326bp and 411bp are visible with no evidence of contaminating primer dimers or adapter DNA.

were sequenced on a MiSeq V2 500 cycles Nano kit chip with a read capacity of $1x10^6$. All exons were successfully covered using visual assessment of raw reads with Integrative Genomics Viewer (IGV) (260). Illumina adapters were trimmed using Cutadapt (235) and raw sequencing fastq files were sorted and indexed into BAM files using SAMtools (237). I decided to use the variant caller LoFreq (236), for reasons of speed and sensitivity. The tool gives data on minor allele frequency (MAF), strand bias, read depth as well and homopolymer run length. Variants were annotated using Variant Effect Predictor (VEP) (Ensembl) (239).

Variants were manually filtered according to the following criteria:

- Removed variant if part of a homopolymer repeat >4 in length.
- Filtered to remove variants with a MAF of <0.3 (given a homogeneous cell population).
- Removed if observed in the HAT-treated bulk cell population (presumed single nucleotide polymorphisms or pre-existing *HPRT1* mutations).

Finally, to determine the trinucleotide context of each variant I used Galaxyproject.org (261), to add flanking base 3' and 5' of each called *HPRT1* variant using the "get flanks" tool. All filtered variants were validated by Sanger sequencing.

Twenty five individual 6-TG-resistant colonies, comprising thirteen HCC2998 Parental 6-TG-resistant, eight SNU-81 Parental 6-TG-resistant and four HCC2998 CRISPR-WT clone 1 6-TG-resistant colonies were analysed. The small number of 6-TG-resistant wildtype clones reflected the fact that, as expected given their low mutation rate, few colonies escaped the 6-TG selection. Each colony harboured at least one "predicted deleterious" mutation (SIFT/Polyphen) or stopgain/splice-site mutation, predicted to cause HPRT1 loss of function. Notably, nearly all of these variants (24/25) were single base-pair substitutions, with the single indel identified in the CRISPR-WT cell line.

Familial *HPRT1*-deficiency in humans causes Lesch Nyhan syndrome (LNS), an X-linked recessive disorder affecting males resulting in hyperuricaemia, mental retardation and self mutilation (262). The mutation spectrum of LNS is well characterized and comprises, in descending order of frequency, missense and nonsense mutations, splice site variants and indels (262). To aid confirmation of the pathogenicity

of *HPRT1* variants in my cell lines, I collated a list of 92 functional Lesch-Nyhan syndrome missense variants from lesch-nyhan.org (accessed 12/12/2018) (Appendix Table C.1). The trinucleotide context of each LNS variant was annotated using Galaxyproject.org "get flanks" tool (261).

Further analysis of the *HPRT1* variants detected in the 6-TG-resistant Parental and CRISPR-WT colonies revealed that 14/25 were recognised causal variants for human Lesch-Nyhan syndrome, confirming their pathogenicity. Interestingly, 4/13 of 6-TG-resistant HCC2998 Parental colonies and 3/8 6-TG-resistant SNU-81 Parental colonies harboured a F74L mutation at a T[C→A]T *POLE* signature trinucleotide hotspot. The overrepresentation of this variant provides functional evidence that this trinucleotide is a mutational hotspot for exonuclease domain mutant *Pol* $\epsilon$. None of the four 6-TG-resistant CRISPR-WT clones harboured *POLE* Signature 10 trinucleotide hotspot mutations.

As sequencing of only a few 6-TG-resistant CRISPR-WT colonies was possible, I utilised the cohort of pathogenic Lesch-Nyhan syndrome variants (See Appendix C.1) as $POLE^{\mathrm{WT}}$ controls to formally investigate the overrepresentation of Signature 10 mutations in 6-TG-resistant Parental cell lines. This revealed a clear and highly statistically significant bias for *POLE* Signature 10 mutations (TCT>TAT, TCG>TTG and TTT>TGT) in the 6-TG-resistant Parental cell lines compared to the Lesch-Nyhan syndrome control cohort (10/21 mutations vs. 6/92 mutations; $P<0.0001$, Fisher's Exact Test) (Table 3.8).

| | ***POLE* Signature 10 mutations** | **Other mutations** |
|---|---|---|
| **6-TG-resistant Parental colonies** | 10 | 11 |
| **Lesch-Nyhan syndrome variants** | 6 | 86 |

**Table 3.8:** Contingency table of *POLE* Signature 10 mutations (TCT>TAT, TCG>TTG and TTT>TGT) and all other mutations in 6-TG-resistant Parental clones and Lesch-Nyhan syndrome patients. There was a highly significant overrepresentation of *POLE* Signature 10 mutations in 6-TG-resistant *POLE* Parental cell lines compared to Lesch-Nyhan syndrome variants; $P<0.0001$, Fisher's Exact Test (For the full list of Lesch Nyhan syndrome variants, see Appendix Table C.1).

These data confirm that cancer-associated *POLE* exonuclease domain mutant Pol $\epsilon$ cause both tumour ultramutation and the characteristic mutation signature found in these cancers.

### 3.3.6 Impact of *POLE* exonuclease domain mutations on cell proliferation

Given the favourable prognosis of *POLE* exonuclease domain mutant cancers, I sought to determine whether this may be explained by an alteration in cell proliferation. Parental and CRISPR-WT HCC2998 and SNU-81 isogenic cell lines were plated at 1500 and 1000 cells per well respectively in a 96-well plate and their proliferation quantified over 120 hours using the CellTitre glo® cell viability assay (Figure 3.14a).

Cell doubling time was calculated by the following equation:

$$\text{Doubling time} = \frac{\text{T x ln(2)}}{\text{ln(Xe/Xb)}}$$

T = Incubation time (hrs), Xb = Reading at the beginning of incubation,
Xe = Reading at the end of incubation

Doubling time did not differ significantly between isogenic CRISPR-WT and Parental cell lines, with the exception of HCC2998 CRISPR-WT clone 3, which had the lowest relative POLE protein levels (Figure 3.14b). Doubling time in CRISPR-MUT (untargeted) clones did not significantly differ from the Parental cell line.

Although there was no reproducible difference in cell proliferation associated with the *POLE* mutation, there was significantly reduced proliferation in HCC2998 CRISPR-WT clone 3 and mildly reduced proliferation in HCC2998 CRISPR-WT clones 1 and 2. Possible reasons for this include, a proliferative advantage in cells containing the $POLE^{\text{P286R}}$ mutant allele or reduced cellular POLE protein levels in hemizygous CRISPR-WT cells. Further analysis, however, of the cell cycle and DNA replication dynamics was required. Furthermore, the significant reduction in cell proliferation in HCC2998 CRISPR-WT clone 3 lead me to exercise caution when using this cell line for further experiments.

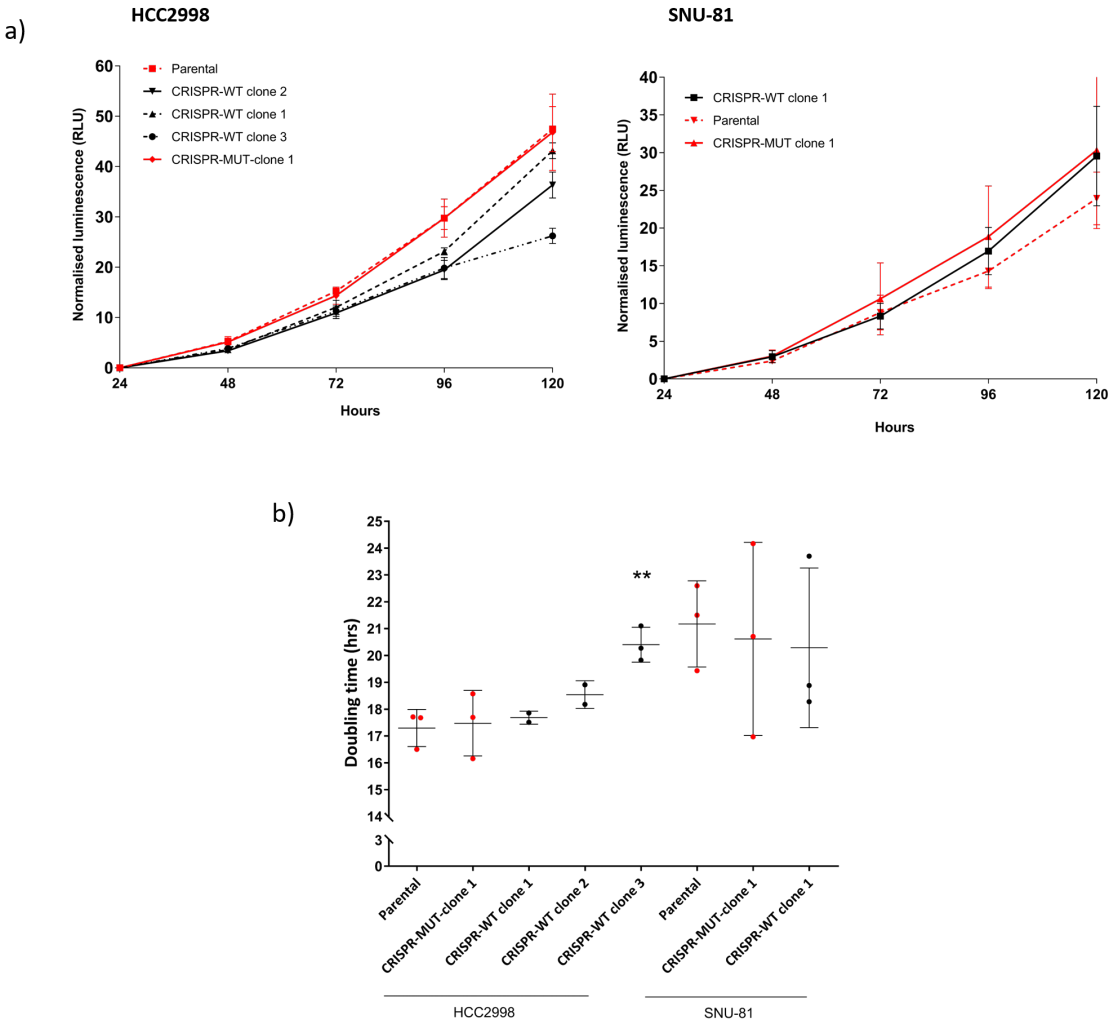



**Figure 3.14:** Cell proliferation curves and doubling time plot for HCC2998 and SNU-81 Parental, CRISPR-MUT (un-targeted) (red) and CRISPR-WT (black) isogenic cell lines. a) Normalised Luminescence readings from 24 to 120 hours. Data shown are mean±SEM. b) Box plot of the doubling time for each cell line as calculated by the above equation. Data shown are mean±SD from a minimum of two replicates. $^{**}P<0.005$; Student's $t$-test.

### 3.3.7 Impact of *POLE* exonuclease domain mutations on the cell cycle

Haploid fission yeast with a $POLE^{P286R}$ mutation have a delayed S-phase and increased RPA foci compared to $POLE^{exo-}$ and $POLE^{WT}$ strains. This observation is indicative of a DNA replication defect (Unpublished data, Kearsey lab, University of Oxford). Furthermore, there is evidence that Pol $\epsilon$ is involved in the S-phase checkpoint in budding yeast, through the binding of POLE to Mediator of the S-phase checkpoint, Mrc1 (orthologue of euakryotic Claspin) (72). It is plausible that a *POLE* exonuclease domain mutation may disrupt this protein interaction, resulting in cell cycle dysregulation. Mindful of this and the data from fission yeast, I next focused on the S-phase of the cell cycle in the isogenic cell lines to determine if there was any variation in S-phase progression in mammalian *POLE* EDM cells.

My initial attempts to synchronise cells at the G1/S boundary, by omitting foetal bovine serum (FBS) from the growth medium (serum starvation), proved unsuccessful. HCC2998 cells proved particularly sensitive to the removal of serum, becoming detached from the culture flask. As an alternative, I next used a double thymidine block (Methods Section 2.1.7).

After the successful pilot of this method, I proceeded to first test HCC2998 as it has a diploid genome, meaning that the stratification of cells by DNA content is more straightforward than in the aneuploid/polyploid SNU-81 cell line. HCC2998 Parental and CRISPR-WT cells were synchronised in G1/S-phase then released into fresh medium. Cells were sampled every hour for 10 hours following G1/S release, then again at 24 hours to ensure that cells were cycling normally and had not been adversely affected by the thymidine block.

At each time-point, the DNA content of single cells was analysed using propidium iodide (PI) staining and FACS sorting (Methods Section 2.1.8). FACS permits the parallel sorting of single cells and the distinction between cells in G1 (2n), S (2n-4n) and G2 (4n) phase of the cell cycle based on the PI fluorescence signal intensity. (Figure 3.15a).

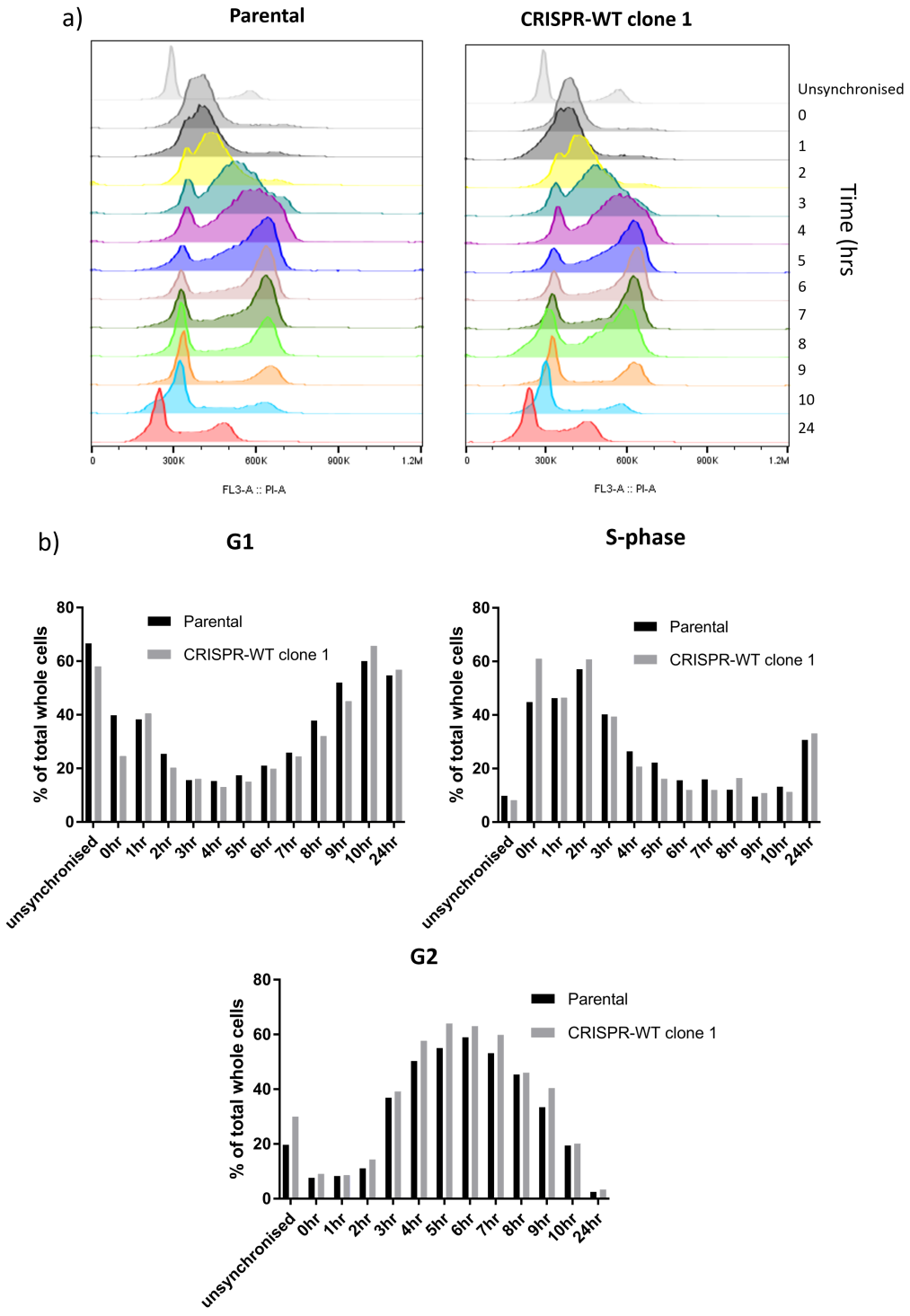



**Figure 3.15:** Cell cycle analysis of isogenic cell lines. a) Histograms were generated using FlowJo® showing the PI signal intensity which is relative to cellular DNA content in FACS-sorted single cells. b) Percentage of single cells in G1, S and G2 phases of the cell cycle over a 10 hour time period computed using FlowJo® software. The cellular DNA content corresponds to G1 (2n), S (2-4n) and G2 (4n) phases of the cell cycle.

This analysis confirmed similar S-phase progression between Parental and CRISPR-WT isogenic cell lines over the 10 hour time period (Figure 3.15b), confirming that neither the heterozygous $POLE^{P286R}$ exonuclease domain mutation, nor hemizygosity, confers an S-phase defect. The cell cycle resumed as normal after 24 hours.

### 3.3.8 DNA polymerase dynamics in *POLE* exonuclease domain mutant cell lines

In contrast to haploid yeast models, the $POLE^{P286R}$ mutation is heterozygous in both Parental HCC2998 and SNU-81 CRC cell lines. Consequently, quantification of cellular doubling time and S-phase progression in a bulk cell sample may be insufficiently sensitive to detect subtle variation in DNA replication caused by the $POLE^{P286R}$ mutation. To test for variation in DNA replication speed and DNA fork stalling, I next used a DNA fibre assay to analyse individual replication forks that have undergone DNA synthesis.

In this protocol, thymidine analogues are incorporated into the DNA by the replicative DNA polymerases *in vivo* and can be visually observed by fluorescently labelled antibodies. Dual labelling of replication forks allows the distinction between different replication fork structures as shown in Figure 3.16. This protocol is the same as that reported by Nieminuszczy *et al.* (222) and is detailed in Methods Section 2.1.9. HCC2998 Parental and CRISPR-WT clones 1 and 2 were pulse-labelled with the thymidine analogues 5-Iodo-2'-deoxyuridine (IdU), followed by 250$\mu$M 5-Chloro-2'-deoxyuridine (CldU) for 20 minutes each. Cells were lysed and DNA fibres spread on glass slides by gravity, fixed and integrated nucleoside analogues IdU and CldU were immunolabelled with Alexa Fluor 568 (red) and Alexa Fluor 488 (green) respectively.

Analysis of elongating replication fork length from a minimum of 450 dual-labelled DNA fibre tracts per cell line revealed small, but statistically significant, differences in DNA fibre length between both HCC2998 CRISPR-WT and Parental isogenic cell lines. Interestingly the difference in fork length was statistically significant between the two CRISPR-WT isogenic cell clones 1 and 2 ($P$ 0.0001-0.0004; two-tailed Mann-Whitney U test (Figure 3.17).

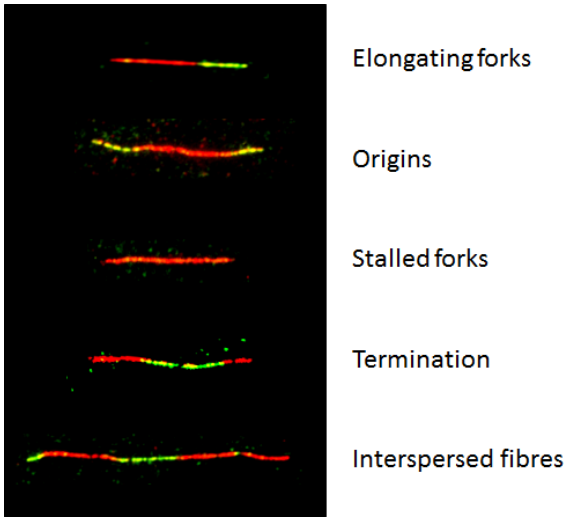



**Figure 3.16:** DNA replication fork structures generated from the DNA fibre assay. The red strand has been labelled with IdU for 20 minutes followed by the green strand, labelled with CldU for a further 20 minutes. After DNA labelling, replication is stalled by adding ice-cold Phosphate-buffered saline to cells.

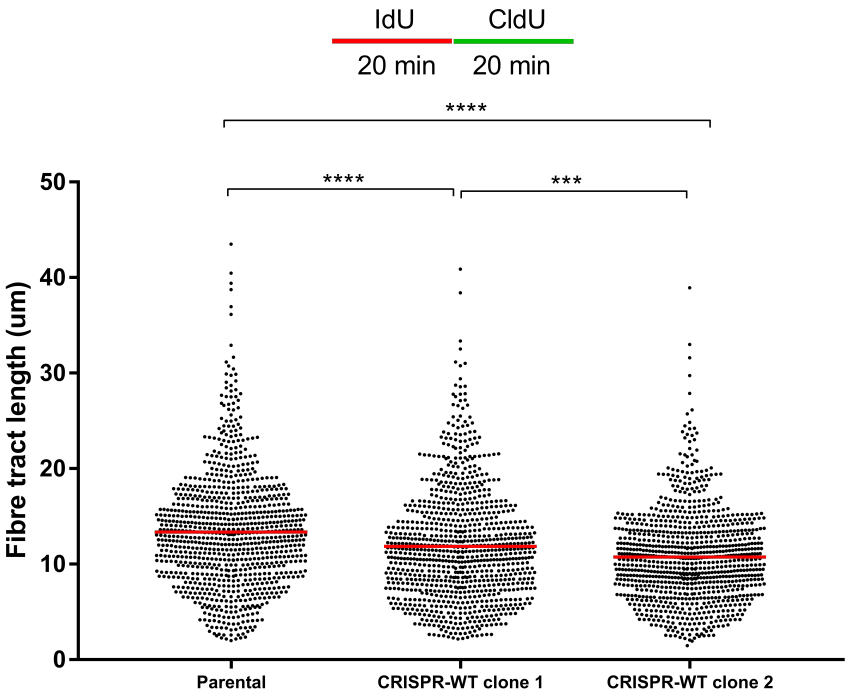



**Figure 3.17:** DNA fibre analysis in isogenic HCC2998 Parental and CRISPR-WT cell lines. DNA fibre-tract-length (IdU + CldU) was measured in elongating forks by fibre pixel length calibrated to 9.34$\mu$m/pixel (Cellular Imaging Core, WCHG, Oxford) using ImageJ (223). A minimum of 450 elongating forks were quantified per cell line. The red bar represents the mean. ***$P$<0.001, ****$P$<0.0001; two-tailed Mann-Whitney U test.

Unlike DNA combing, DNA fibre spreading is not done at a controlled speed, but by using gravity. Therefore I can conclude that although there is no substantial replication fork defect in Parental *POLE* exonuclease domain mutant cell lines, in the context of global cellular DNA replication, the subtle differences detected in replication fork velocity may contribute to the phenotypic characteristics observed in *POLE* exonuclease domain mutant cancers.

Regardless of the rate of DNA fork progression, dual labelling of DNA fibres permits distinction between different replication fork structures, such as: elongating fibres, which contain the first label followed by the second label; origins, which contain the first label in the middle followed by the second label on either end of the fibre; termination sites, which contain the first label at each end of the fibre followed by the second label colliding in the middle; and stalled replication forks, which contain the first label only (Figure 3.16).

I hypothesised that if *POLE* exonuclease domain mutations lead to increased polymerase stalling, this should be reflected in an increase in stalled fork structures, regardless of their length. To test this hypothesis, I counted 150 replication forks and calculated the proportion of each replication fork structure as listed in Figure 3.16. There was no statistically significant difference between the proportion of stalled forks ($P = 0.43$, Student's $t$-test) or other fork structures between Parental and CRISPR-WT cell lines.

To examine DNA replication fork progression under replication stress, I performed a pilot DNA fibre assay on Parental cells and CRISPR-WT clones 1 and 2 in the presence of the the deoxyribonucleotide synthesis inhibitor, hydroxyurea (HU). Cells were pulse-labelled with 25$\mu$M IdU for 20 minutes, followed by 250$\mu$M CldU and 2mM HU for two hours. DNA fibres were processed as per the protocol described previously. The frequency of elongating and stalled forks in each cell line was quantified from a minimum of 310 fibres. Upon hydroxyurea treatment, the percentage of stalled forks increased from 11% to 33% in Parental cells, 14% to 67% in CRISPR-WT clone 1 and 9% to 65% in CRISPR-WT clone 2, demonstrating that the experiment was successful in stalling DNA replication forks.

The proportion of stalled replication forks compared to elongating replication forks was reduced almost 2-fold in Parental cells (33%) compared to (67% and 65%)

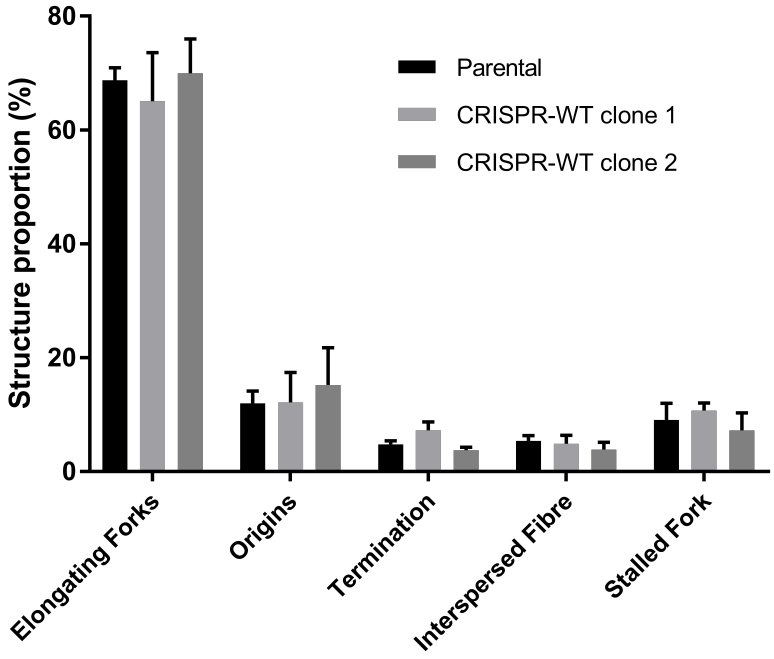



**Figure 3.18:** DNA replication fork structures in HCC2998 Parental and CRISPR-WT clone 1 and clone 2. Data are plotted as the % of each structure in a minimum of 150 individual fibres; mean + SD.

in CRISPR-WT clones 1 and 2 respective (Figure 3.19). These data suggest that *POLE* exonuclease domain mutant polymerases are able to replicate DNA, to a certain extent, upon dNTP pool depletion. Speculatively, the propensity to replicate past a mismatch may increase in *POLE* exonuclease domain mutant polymerases, hence reducing fork stalling upon replication stress. Due to time constraints I was unable to perform replicate experiments.

These data demonstrate that, under normal cell culture conditions, there is a subtle difference in replication fork speed, but no significant difference in the proportion of stalled forks between Parental and CRISPR-WT cells. Pilot data from hydroxyurea-treated cells, suggest that the *POLE* exonuclease domain mutant polymerase is more permissive than the wildtype polymerase under DNA replication stress.

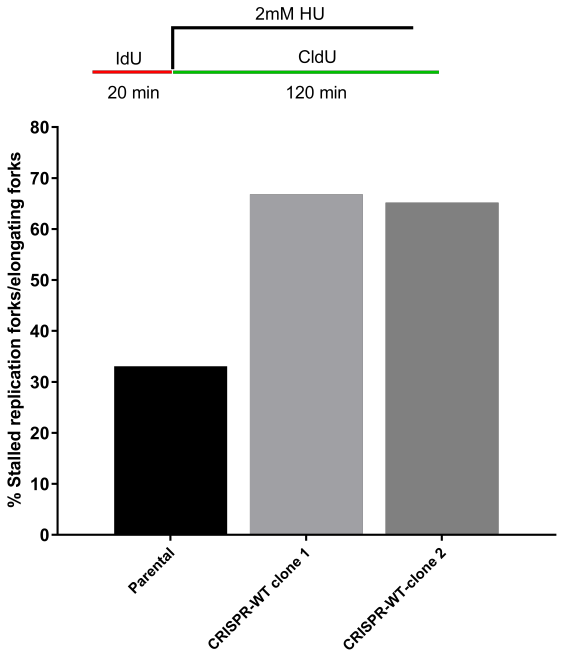



**Figure 3.19:** Proportion of stalled replication forks to elongating replication forks in HCC2998 Parental and CRISPR-WT clone 1 and clone 2 following hydroxyurea treatment. Data are plotted as the % of stalled forks in a minimum of 310 individual fibres.

### 3.3.9 DNA damage response in *POLE* exonuclease domain mutant cell lines

Given the ultramutator phenotype in *POLE* exonuclease domain mutant cell lines, I performed western blotting on protein lysates extracted from asynchronous Parental and CRISPR-WT cells lines for Phospho-CHK1 (Ser345), a component of the DNA damage checkpoint. When phosphorylated the protein is a marker of DNA damage response. Results showed no expression of Phospho-CHK1 in Parental or CRISPR-WT cell lines compared to the hydroxyurea-treated positive control. These data confirm that the DNA damage checkpoint is not activated in *POLE* exonuclease domain mutant cell lines (Figure 3.20).

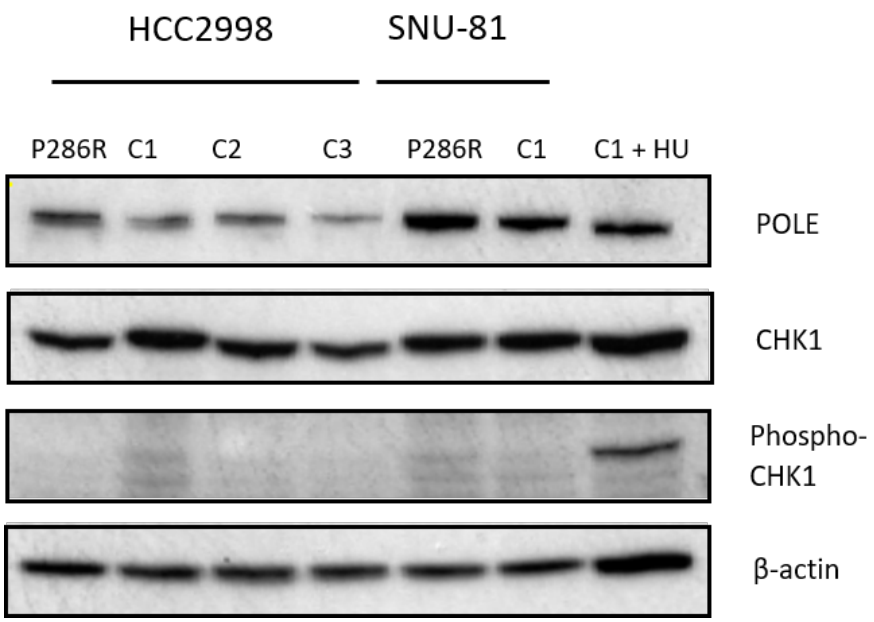



**Figure 3.20:** CHK1 and Phospho-CHK1 protein expression in Parental and CRISPR-WT isogenic cell lines in unstressed conditions. The positive control cell line underwent hydroxyurea treatment for two hours to to activate the DNA replication checkpoint.

### 3.3.10 Knockdown of MMR genes in *POLE* isogenic cell lines

#### 3.3.10.1 *MSH6* mutations in *POLE* mutant cell lines

Tumours with cancer-associated *POLE* exonuclease domain mutations have a strong tendency towards microsatellite stability (148, 205) despite an apparent enrichment of pathogenic missense and nonsense mutations in the mismatch repair gene *MSH6* (123, 144, 158). To further investigate this relationship, I sequenced a panel of five microsatellite stable (MSS), five *POLE* exonuclease domain mutant (including HCC2998 and SNU-81) and three MSI cell lines for exonic mutations in *MSH6*.

Consistent with the human tumour sequencing data, *POLE* exonuclease domain mutant cell lines were significantly enriched for missense and nonsense mutations in *MSH6* compared to $POLE^{WT}$ and MSI cells ($P = 0.0079$ and 0.0179 respectively, Fisher's exact text) (Table 3.9). Interestingly, none of the $POLE^{WT}$/MSS and MSI cell lines harboured non-synonymous *MSH6* mutations, despite GP2D being MSH6-deficient (Figure 3.22b). Further analysis of the *MSH6* mutations in *POLE* exonucle-

ase domain mutant cells revealed an enrichment of Signature 10-type mutations, with 2/7 TCT→TAT, 2/7 TCG→TTG and 1/7 TTT→TGT mutations. All mutations were C→A, C→T or T→G base pair substitutions, regardless of mutation context. These substitutions resulted in typical codon alterations associated with *POLE* mutations, including glutamic acid to stop codons (3/7) and arginine to glutamic acid (2/7) (Table 3.9).

| Cell line | ***POLE* EDM** | Microsatellite status | Non-synonymous ***MSH6* mutation (heterozygous)** |
|---|---|---|---|
| HCC2998 | P286R | MSS | K855N |
| SNU-81 | P286R | MSS | R922Q; E948X |
| HT115 | V411L | MSS | R911Q; E1322X |
| HDC114 | A456P | MSS | E641X |
| SNU-1040 | F367S | MSS | R243H |
| C32 | - | MSS | - |
| HDC57 | - | MSS | - |
| Colo320 | - | MSS | - |
| HT29 | - | MSS | - |
| T84 | - | MSS | - |
| RKO | - | MSI | - |
| GP2D | - | MSI | - |
| LS180 | - | MSI | - |

**Table 3.9:** Somatic *MSH6* mutations in *POLE* exonuclease domain mutant,*POLE* wildtype microsatellite stable (MSS) and microsatellite unstable (MSI) colorectal cancer cell lines identified by Sanger sequencing.

All three of the *POLE* exonuclease domain mutant cells with truncating heterozygous *MSH6* mutations are microsatellite stable, consistent with the observation that biallelic mutations alterations in MMR genes are required for mismatch repair deficiency and microsatellite instability to arise. However, the apparent abundance of heterozygous *MSH6* mutations without a second hit posed the question as to whether these mutations had a small, but important effect on mutation rate, whereas biallelic MMR gene activation might result in an increase in the mutation rate beyond that optimal for cell fitness. Indeed, as noted previously, combined loss of *POLE* exonuclease activity and MMR in budding yeast leads to a catastrophic increase in mutation rate, exceeding the error threshold, resulting in error-induced extinction (128).

To test this hypothesis, I used siRNA to transiently knock down MSH6 and MLH1 protein in the HCC2998 isogenic cell lines to test its effect on cell viability and DNA damage response.

#### 3.3.10.2 Cell viability of *MSH6* and *MLH1* knockdown *POLE* exonuclease domain mutant cell lines

The optimum concentration of siRNA for MSH6 and MLH1 protein knockdown was 25nM and 12.5nM siRNA respectively (Figure 3.21), using ON-TARGET *plus* siRNA SMARTpool (Dharmacon). Despite residual *MSH6* and *MLH1* mRNA expression at 48 hours, by 96 hours, western blotting detected no MLH1 or MSH6 protein (Figure 3.21). These results confirm successful knockdown of target mismatch repair proteins using siRNA.

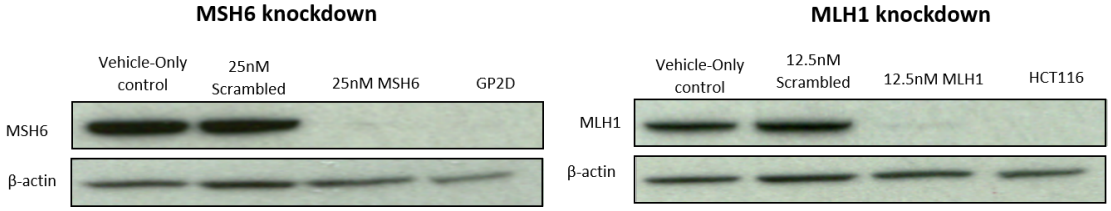



**Figure 3.21:** MLH1 and MSH6 protein expression 96 hours post-siRNA transfection. Vehicle-only control cells were transfected with Lipofectamine® RNAiMAX-only and MLH1 and MSH6-deficient cell lines HCT116 and GP2D respectively were used as controls for antibody specificity.

To determine if transient siRNA knockdown of MSH6 and MLH1 reduces cell fitness in HCC2998 Parental cell lines, cell viability was quantified using the CellTiter-Glo® luminescent cell viability assay (Promega) 96 hours post-transfection, relative to cells transfected with scrambled siRNA. Results demonstrated no significant difference in cell viability between Parental and CRISPR-WT clone 1 isogenic lines (Figure 3.22). These data confirm that transient knockdown of MMR genes does not lead to a reduction in cell viability in *POLE* exonuclease domain mutant cell lines.

As the limitations of transient siRNA knockdown meant that the duration of MMR loss may have been of insufficient length for cells to display a reduction in viability, I next examined markers of DNA damage response and apoptosis for the early evidence of replication stress that might occur prior to a global reduction in cell viability. I began by studying the well characterised marker of apoptosis, Cleaved Caspase-3 by western blotting. Results demonstrated that Cleaved Caspase-3 were not activated in response to MMR protein knockdown (Figure 3.23a). Next, DNA damage response was examined by western blotting of $\gamma$-H2AX, a marker of double strand breaks, in

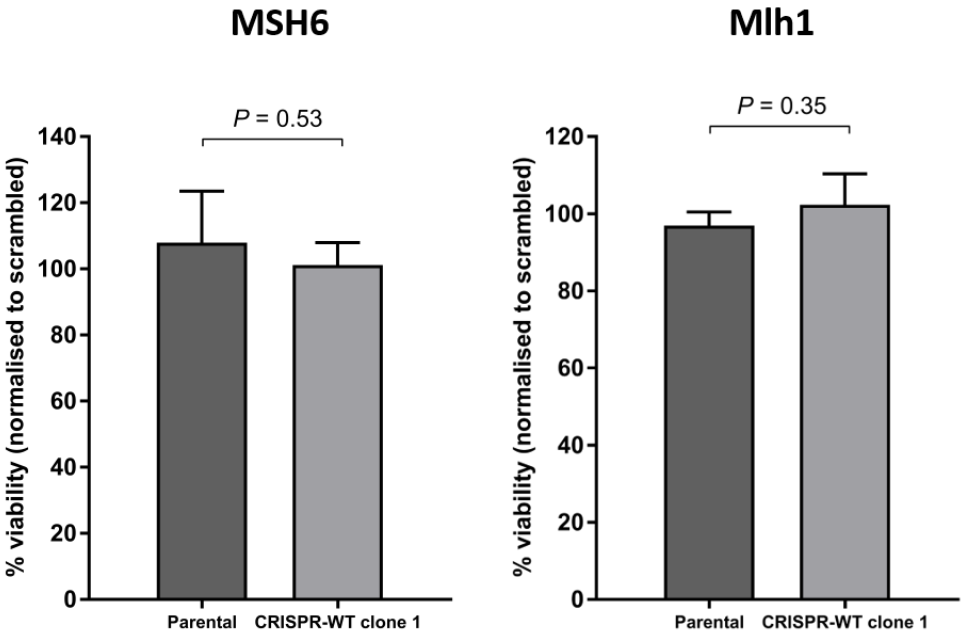



**Figure 3.22:** Cell viability of HCC2998 Parental and CRISPR-WT clone 1 isogenic cell lines 96 hours post MSH6 and MLH1 siRNA transfection. Cell viability was quantified using the CellTiter-Glo Luminescence cell viability assay (Promega). Data were normalised to cells transfected with Scrambled siRNA. Data shown are mean (n=3)+SD. $P$ = 0.35-0.53; Student's $t$-test.

MLH1 and MSH6 knockdown Parental *POLE* EDM cells. No expression of γ-H2AX was detected relative to the UV-irradiated positive control (Figure 3.23b).

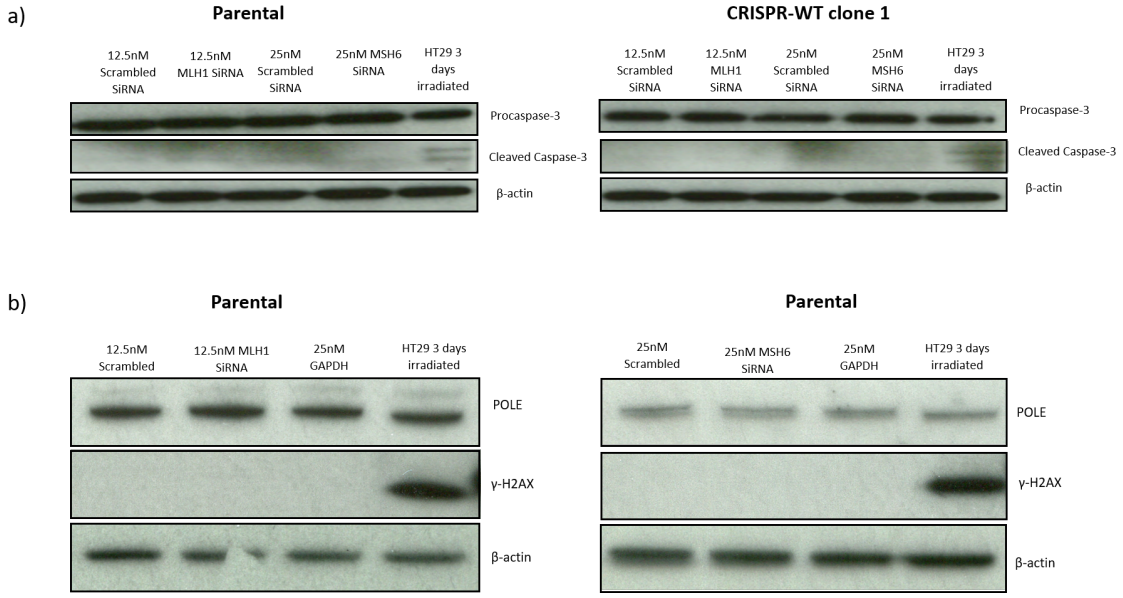



**Figure 3.23:** Expression of apoptotic and DNA damage response markers in Parental and CRISPR-WT isogenic cell lines. a) Western blot of apoptotic marker Cleaved Capspase-3 upon MMR knockdown. The positive control was UV irradiated for 30 minutes three days prior to protein lysate extraction. Neither MLH1 or MSH6 knockdown cells express markers of apoptosis. b) Western blot of DNA double-strand break marker γH2AX in MLH1 and MSH6 siRNA knockdown Parental *POLE* mutant cell lines. The positive control was UV irradiated for 30 minutes and protein extracted after three days.

Taken together, these data show that transient loss of MMR in cells with *POLE* exonuclease domain mutations does not result in a reduction in cell viability, increase apoptosis or replication stress, in contrast to the marked effects of this combination in yeast (128). It is possible that transient knockdown of MMR protein is insufficient for mammalian cells to exhibit a quantifiable reduction in cell viability and constitutive knockdown of MMR proteins is required.

### 3.3.11 Effect of *POLE*$^{P286R}$ exonuclease domain mutations on sensitivity to chemotherapeutic agents

As discussed previously, the common use of adjuvant chemotherapy and radiotherapy in colorectal and endometrial cancers means that analysis of clinical data are currently unable to determine whether the good prognosis of *POLE* exonuclease domain mutant tumours is independent of such therapies. Thus, determining whether *POLE* mutation alters the sensitivity to chemotherapy and radiotherapy is a critically unanswered question.

Although earlier studies using a panel of *POLE* exonuclease domain mutant and wildtype CRC cell lines in our laboratory suggested that *POLE* exonuclease domain mutant cells have an increased sensitivity to 5-fluorouracil (5-FU) (Figure 3.24), interpretation of these data is complicated by the heterogeneous genetic background of the cell lines. Furthermore, as noted previously, studies of murine ES cells and human B-cell lines have suggested that *POLE* exonuclease domain mutations confer sensitivity to the nucleoside analogues cytarabine and fludarabine (242, 243).

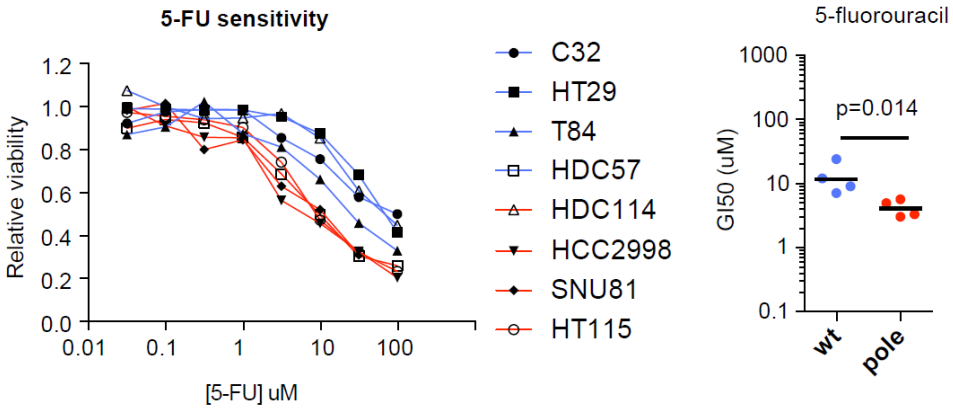



**Figure 3.24:** 5-fluorouracil sensitivity in four *POLE* exonuclease domain mutant (red line) and four *POLE* wildtype (blue line) CRC cell lines previously tested in the Tomlinson laboratory. GI50 was calculated using GraphPad Prism version 7.00, GraphPad Software, La Jolla California USA, www.graphpad.com. There was a significant difference in 5-FU sensitivity between *POLE* exonuclease domain mutant and wildtype cell lines; $P = 0.014$ (Student's $t$-test).

To test the effect of cancer-associated *POLE* exonuclease domain mutations on

sensitivity to cytotoxics in clinical use, and nucleoside analogues in a clinically relevant model system, I used the isogenic HCC2998 cell lines to estimate their response to nucleoside analogues 5-FU, fludarabine and the commonly used cytotoxic camptothecin, a topoisomerase inhibitor, using a cell viability assay. HCC2998 was selected over SNU-81 as multiple CRISPR-WT clones were available to rule out clonal variations in sensitivity.

Preliminary optimisation experiments focussed on calculating the GI50 (the concentration at which cell proliferation is inhibited by 50%) of each HCC2998 isogenic cell line using Prism by fitting a nonlinear regression model with variable slope (4 parameters). Data demonstrated no evidence of differential sensitivity to 5-FU, camptothecin or fludarabine between Parental and CRISPR-WT cell lines, despite variability between clones (Figure 3.25).

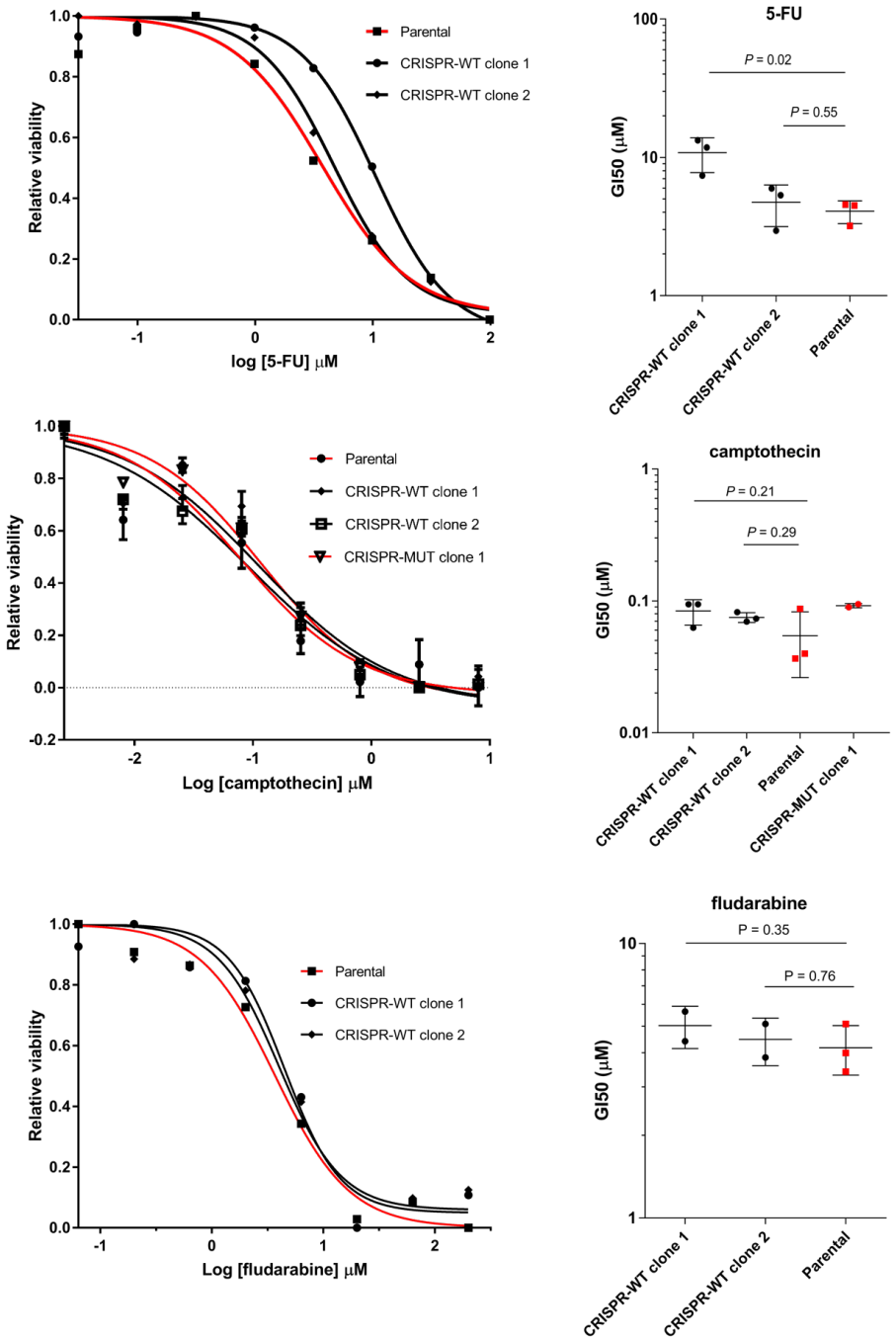



**Figure 3.25:** Dose-response curves of HCC2998 isogenic cell lines exposed to 5-FU, camptothecin and fludarabine. Data were normalised to the vehicle-only control. GI50 was calculated using the dose-response - Variable slope application in GraphPad Prism version 7.00, GraphPad Software, La Jolla California USA, www.graphpad.com. There was no significant difference in 5-FU, camptothecin and fludarabine sensitivity between Parental and CRISPR-WT cell lines, $P$ values are labelled on each graph (Student's $t$-test).

To test HCC2998 isogenic cell line sensitivity to the nucleoside analogue cytarabine, a pilot cell viability assay was performed. Results showed a plateau in relative cell viability from 0.5$\mu$M to 500$\mu$M of cytarabine, with a minimum relative cell viability of 0.59. These data suggest that both Parental and CRISPR-WT clone 1 cells are resistant to cytarabine (Figure 3.26).

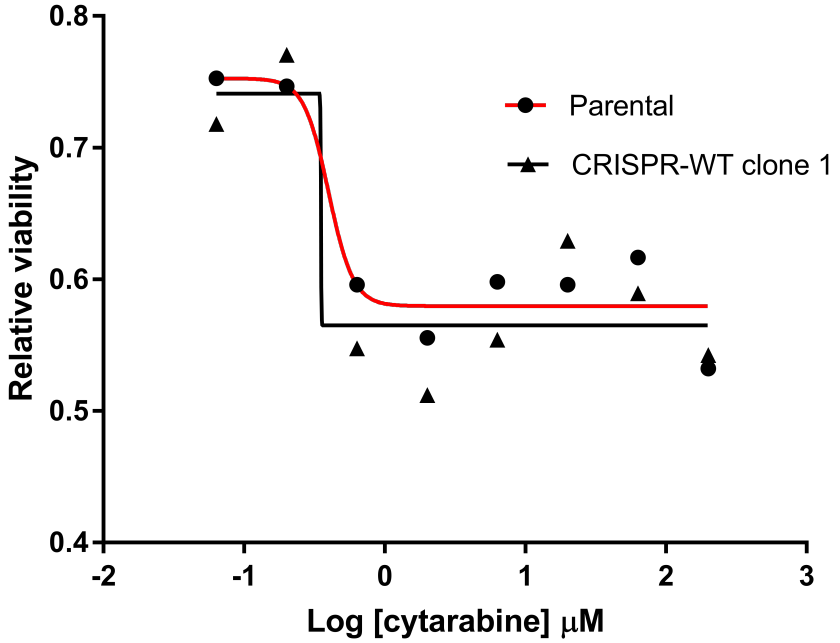



**Figure 3.26:** Pilot dose-response curve of HCC2998 isogenic cell lines exposed to cytarabine (concentration range of 0.15$\mu$M to 500$\mu$M). Data were normalised to the vehicle-only control. The curve was fitted by variable slope application in GraphPad Prism version 7.00, GraphPad Software, La Jolla California USA, www.graphpad.com. The GI50 was not able to be calculated from these data.

### 3.3.12 Overexpression of *POLE* in diploid MSS colorectal cancer cell lines

Long-term culture of the $POLE^{P286R}$ exonuclease domain mutant cell lines, HCC2998 and SNU-81, may have led to adaptation involving secondary mutations, altering the DNA damage response and confounding my data. To study the functional consequence of introducing the $POLE^{P286R}$ mutation into *POLE* wildtype colorectal cancer cells and to permit comparison with isogenic D275A/E277A exonuclease-deficient

mutants, I overexpressed *POLE* cDNA in the diploid, microsatellite stable CRC cell line SW480.

#### 3.3.12.1 Constitutive overexpression of exonuclease domain mutant POLE in CRC cells

To enable constitutive expression for mutagenesis and DNA damage response analysis, C-terminal HA-tagged $POLE^{P286R}$ (P286R), $POLE^{D275A/E277A}$ (exo-null) (experimental) and wildtype *POLE* (control) cDNA was overexpressed in the CRC cell line SW480. The HA tag is a highly immunoreactive epitope and at 14 amino acids long is predicted less likely to interfere with protein folding or binding.

The full-length 7Kb *POLE* cDNA was amplified by long-range PCR (Primers shown in Figure 3.27a and listed in Appendix Table A.3), from a baculovirus expression vector (pFASTBAC) (kindly provided by Sue Cotterill, St George's University of London) and cloned into the the 710 lentiviral expression vector (Figure 3.27b). The 710 plasmid-integrated *POLE* cDNA was sequenced for mutations using primers listed in Appendix Table A.3 and no mutations were detected.

Site-directed mutagenesis was used to introduce the P286R (c.857C→G) and D275A/E277A exo-null (c.824A→C/830A→C) mutations to the *POLE* cDNA within the initial pFASTBAC plasmid (Figure 3.27c, Methods Section 2.2.4.3). The 877bp fragment of *POLE* cDNA containing the P286R and D275A/E277A mutations was excised from each pFASTBAC plasmid using restriction enzymes NsiI and BstEII.

The 710 plasmid was digested using the same restriction enzymes to remove the corresponding *POLE* cDNA fragment. The mutant *POLE* cassettes were ligated into the 710 plasmids to create full length P286R and exo-null *POLE* cDNA vectors ready to generate lentiviral particles. The entire *POLE* cDNA was Sanger sequenced and no additional mutations were identified (Primers in Appendix Table A.3).

Lentiviral particles for the three *POLE* variant vector plasmids were packaged in HEK293T cells following the protocol in Methods Section 2.1.4.1. The concentration of purified lentivirus was quantified by p24 ELISA (Methods Section 2.1.4.2). I transduced SW480 cells and the efficiency was estimated using a GFP control plasmid (Figure 3.28a). Seven days post-transduction, cells were harvested from two inde-

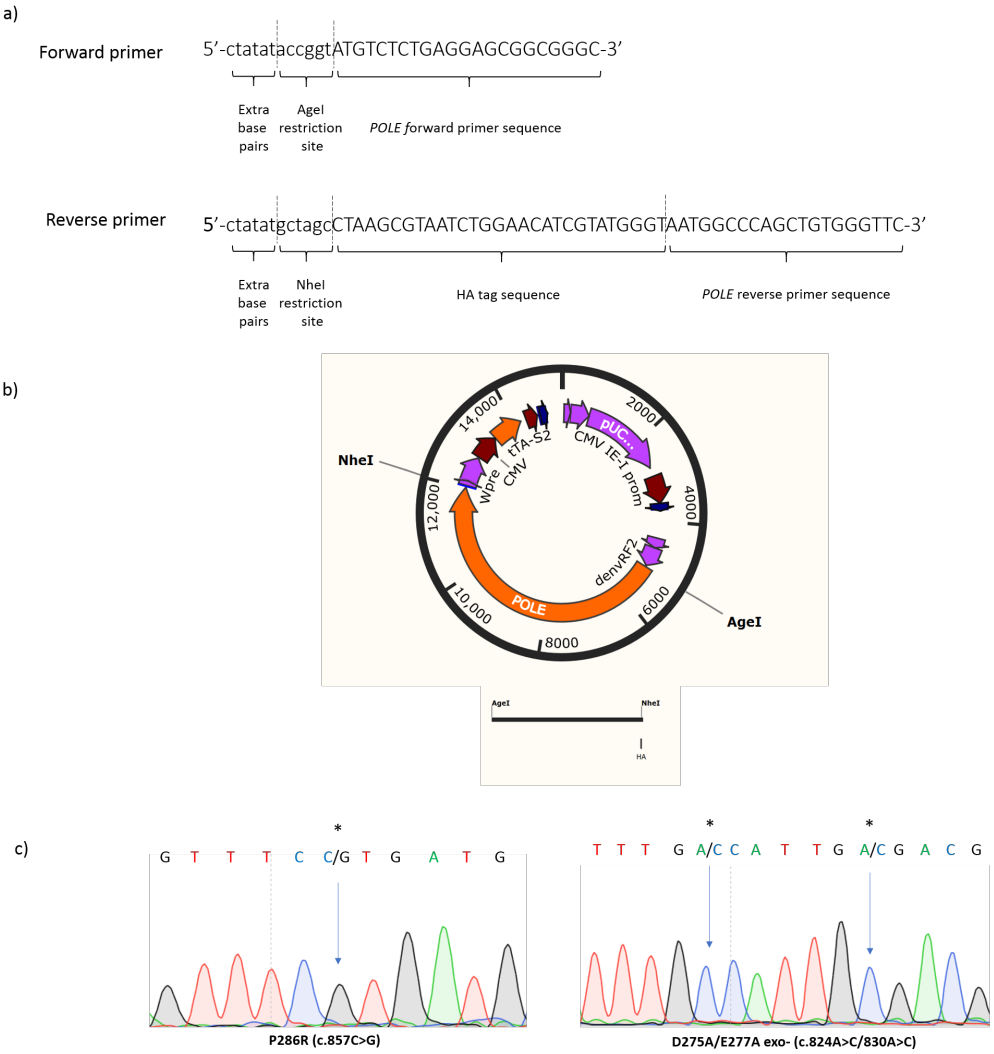



**Figure 3.27:** Amplification, cloning and site-directed mutagenesis of *POLE* cDNA. a) Annotated primer sequences to amplify full-length *POLE* cDNA. Primers include the AgeI and NheI restriction sites to permit 710 vector cloning and the C-terminal HA tag for protein immunodetection (Appendix Table A.3). b) Map of the 8318bp lentiviral expression plasmid 710. Key components are labelled as follows; CMV, cytomegalovirus promoter; TetO7, Tetracycline operator; tTA-S2, tTA-S2, tetR plus C-terminus of VP16; RRE, Rev Response Element; Wpre, Woodchuck hepatitis virus post-transcriptional regulatory element. NheI and AgeI are restriction sites used in the cloning procedure. SnapGene software (from GSL Biotech; available at snapgene.com). c) Chromatograms of *POLE* P286R (c.857C→G) (left) and D275A/E277A exo-null (c.824A→C/830A→C) (right) mutations introduced by site-directed *POLE* mutagenesis. The construct was sequenced directly from the plasmid using primers POLE 3F and POLE 3R (Appendix Table A.3).

pendent transduction wells and POLE-HA expression analysed by HA-tag western blotting (Figure 3.28b). HA-tagged POLE was expressed in transduced SW480 cells, confirming successful transduction of the 710 plasmid. As the aim was to generate constitutively expressing cells, I single-cell cloned the P286R, exo-null and wildtype POLE-HA transduced cells to select for colonies that had successfully integrated the plasmid into the genome.

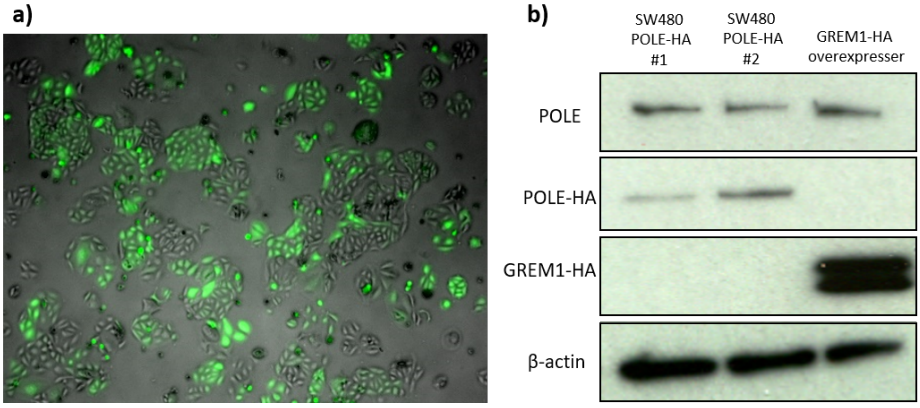



**Figure 3.28:** GFP transfection efficiency and POLE-HA expression in SW480 cells. a) Representative overlaid fluorescence and phase contrast microscopy image of *POLE*-transduced cells co-transfected with GFP plasmid pEGFP-C3 (FITC filter). b) Western blot of POLE, POLE-HA, GREM1-HA and $\beta$-ACTIN. Lane one and lane two are two lysate preparations of POLE-HA wildtype transduced SW480 cells seven days after transduction. These cells show positive expression for HA-tagged POLE protein. Lane three is a positive control for HA antibody specificity - lysate from a cell line with stable expression of HA-tagged GREM1 protein.

Following clonal expansion of POLE-HA P286R, exo-null and wildtype transduced SW480 cells, genomic DNA was extracted to test for integration of the viral plasmid into the nuclear genome (nuclear DNA was not extracted from POLE-HA wildtype transduced clones). The region spanning the HA tag was PCR amplified using primers "Full *POLE* cDNA R" and "*POLE* Fw 9" (Appendix Table A.3). Amplified construct was present in 9/21 and 6/12 POLE-HA P286R an POLE-HA exo-null clones respectively (Figure 3.29a). Western blotting for the HA-tag was performed to detect POLE-HA protein in five POLE-HA exonuclease domain mutant clones (POLE-HA P286R clones 8, 15 and 26 and POLE-HA exo-null clones 8 and 15) and 17 previously untested POLE-HA wildtype clones. Three POLE-HA P286R, two exo-null clones and 3/17 POLE-HA wildtype clones (clones 6, 8 and 16) demonstrated positive HA-tagged POLE expression at the expected molecular weight of 262kDa (Figure

3.29b).

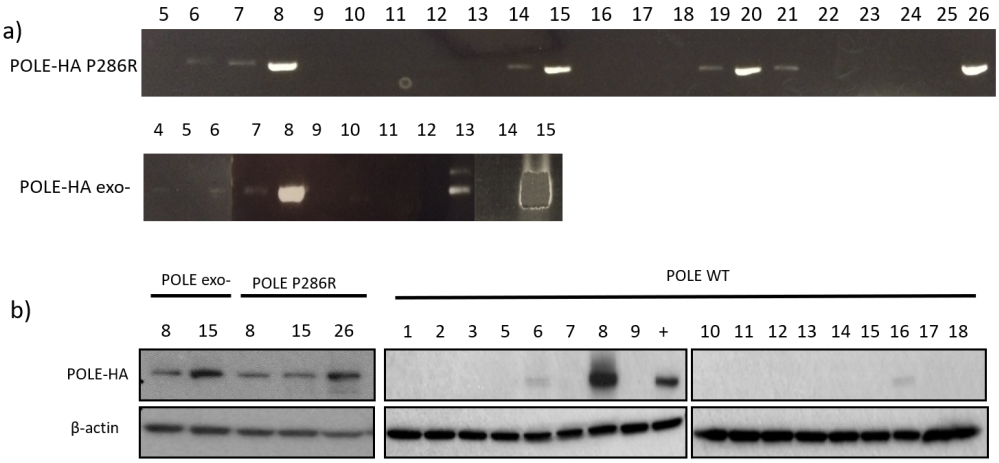



**Figure 3.29:** Screening of SW480 clones transduced with POLE-HA lentivirus. Clone numbers are labelled above each corresponding lane a) PCR amplification of genomic DNA using primers "Full POLE cDNA R" and "POLE Fw 9" (Appendix Table A.3) which amplify HA-tagged *POLE*. This enables detection of cells with integrated expression vectors b) Western blot of lentiviral transduced clone lysates using the HA antibody to identify clones with stable POLE-HA protein expression.

The 710 expression plasmid contains a Tet-off system that switches off transcription of the integrated gene in the presence of the tetracycline analogue, doxycycline (Figure 3.30a). To test if the Tet-off system was functional, POLE-HA P286R clone 8 was cultured with and without 10$\mu$g/ml of doxycycline for seven days, replacing doxycycline every 72 hours. Protein was extracted and western blotting was performed to detect POLE-HA expression. Figure 3.30b shows no expression of HA-tagged POLE in doxycycline-treated cells compared to untreated cells demonstrating that the Tet-off system is functional. There was no overall reduction in POLE expression levels after doxycycline treatment (Figure 3.30b).

I next sought to determine the spontaneous mutation rate in POLE-HA P286R and exo-null overexpressing cell lines using the *HPRT1* forward mutation assay as previously described in Section 3.3.5. SW480 cells have an exceptionally low mutation rate (0.26 x $10^{-5}$) as calculated in Section 3.3.5. SW480 P286R clone 8, SW480 exo-null clone 8 and SW480 wildtype clone 6 (control) were first purged of pre-existing *HPRT1* mutant cells using HAT and HT medium followed by culture in normal medium for five passages (15 days) to accumulate mutations. Cells were plated at a density of 1x$10^5$ cells per 10cm dish in 20 dishes and cultured in normal growth

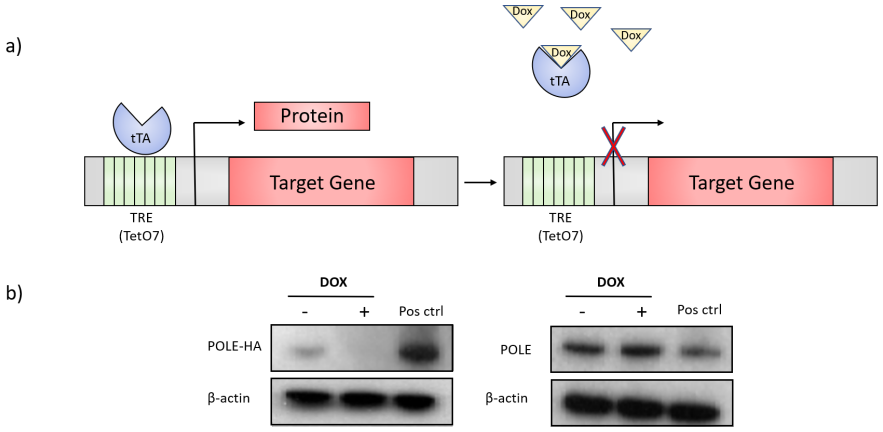



**Figure 3.30:** Tetracycline off system. a) Schematic of the Tet-off system. Briefly, the Tetracycline-response element (TRE), comprised of seven 19bp repeats called the tet operator (labelled TetO7 on the 710 plasmid in Figure 3.27) which, in the presence of the active tetracycline-controlled transactivator (labelled tTA-S2 in the 710 plasmid in Figure 3.27) promotes expression of the target gene. In the presence of doxycycline, tTA-S2 is inactivated which prevents binding to the TRE and inhibits gene expression. b) POLE-HA and global POLE protein expression in cells incubated with and without the addition of 10$\mu$g/ml doxycycline (Dox) for seven days.

medium plus 40$\mu$M 6-TG for 12 days. In parallel, 500 cells per 10cm dish were plated in triplicate to measure plating efficiency. Colonies were fixed in methanol and stained with 0.1% crystal violet in PBS.

No colonies were visible on the 6-TG-treated P286R, exo-null or wildtype SW480 plates, demonstrating that no cells had acquired pathogenic mutations in the *HPRT1* gene (Figure 3.31). As no viable cells were present, I was unable to calculate a mutation rate from this experiment. A possible reason why there were no viable colonies on the POLE-HA P286R and exo-null 6-TG-treated plates is that there was selection against the presence of the viral plasmid in the genome, for example owing to too high a mutation rate, particularly if multiple copies of the mutant construct were integrated into the genome. To test this hypothesis, the mutation-accumulation steps of the experiment were repeated and protein lysates collected at multiple passages which underwent western blotting for HA-tagged POLE.

There was no evidence for loss of POLE-HA protein expression in POLE-HA wild-type, exo-null or P286R transduced clones through serial passages, arguing against negative selection of the POLE-HA allele but not ruling out a reduction in cell viability (Figure 3.31b).

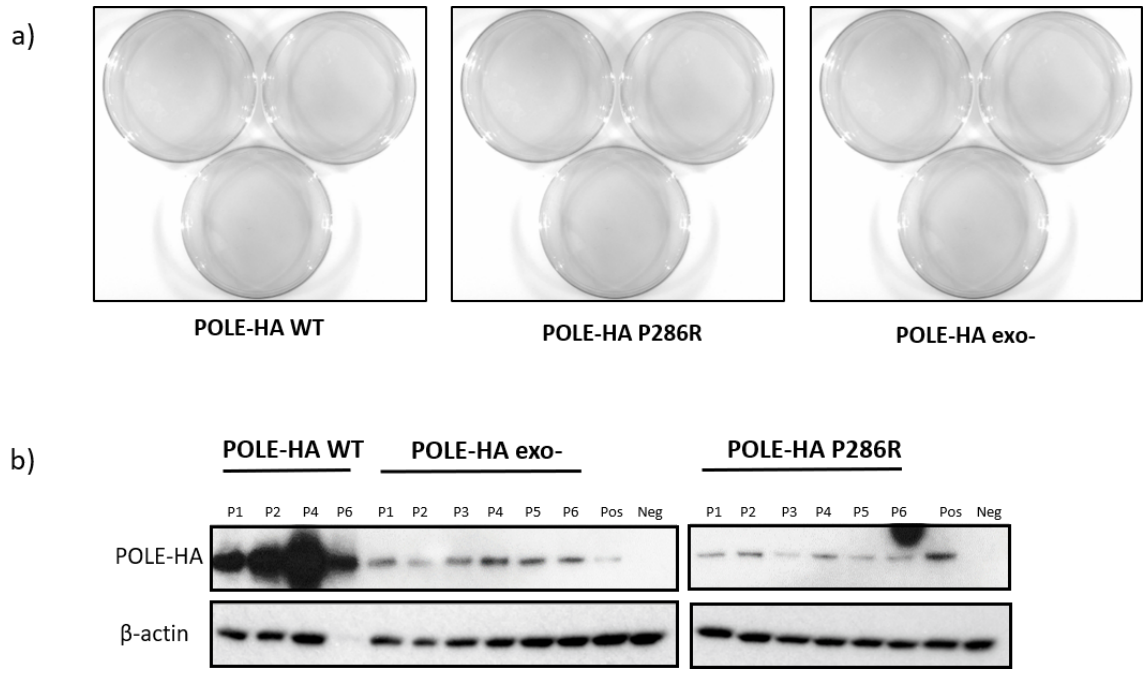



**Figure 3.31:** *HPRT1* forward mutation assay results from POLE-HA P286R, exo-null and wildtype overexpressing cell lines. a) Representative images of 10cm dishes containing 1 x $10^5$ cells following treatment with 40$\mu$M 6-TG for 14 days. No viable colonies are visible when stained with crystal violet. b) Western blotting of POLE-HA protein expression over *HPRT1* mutation accumulation passages. Each lane represents a mutation accumulation passage number of the individual clone (i.e. P2 is passage 2).

The plating efficiencies of exo-null and P286R POLE-HA clones were higher than the POLE-HA wildtype clone (Figure 3.32). Colony morphology also varied between cell lines, with the exo-null clone producing visibly smaller colonies, with more small inviable colonies below 20 cells in diameter than the POLE-HA P286R and wildtype clones. Further studies are required to determine if the difference in colony morphology is due to clonal differences or an effect of *POLE* genotype.

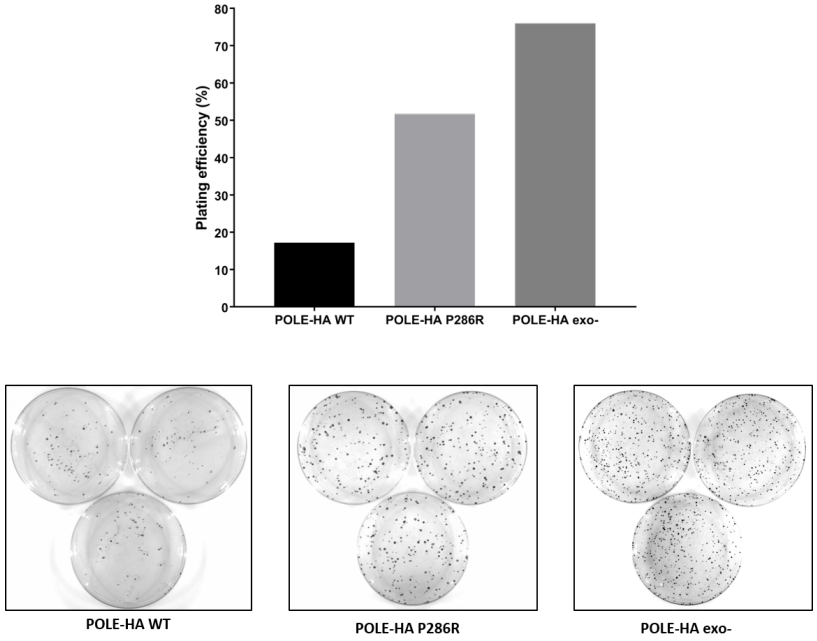



**Figure 3.32:** Plating efficiency (PE) of SW480 POLE-HA P286R, exo-null and wildtype clones. The plating efficiency for each cell line was calculated using the following equation: % PE = (total number of colonies/number of cells plated) x 100.

## 3.4 Discussion

This work is the first to characterise clinically relevant *POLE* exonuclease domain mutations in an isogenic human cell model. My analysis differs significantly from previous studies of *POLE* exonuclease domain knock-in cell models, as I have used patient-derived colorectal cancer cell lines with pre-existing $POLE^{\mathrm{P286R}}$ mutations to generate $POLE^{\mathrm{WT}}$ isogenic derivatives using CRISPR-Cas9n genome editing.

By contrast with wildtype counterparts, I have shown that it is the *POLE* mutation that determines the key molecular characteristics of *POLE* exonuclease do-

main mutant cancers, including ultramutation, a distinct mutation signature and microsatellite stability/low number of indels. By more faithfully recapitulating the status of *POLE* mutant cancers, this model may permit more informative analyses of the underlying mechanisms and consequences of these mutations.

### 3.4.1 Generation and genetic characterisation of *POLE* isogenic cell lines

CRISPR-Cas9 genome editing has been described as "affordable and easy to engineer" (250). Whilst this true with regards to the methodology, requiring a Cas9-expressing plasmid, a single guide RNA and a repair template, this study has demonstrated that the engineering of cells as well as careful post-CRISPR genetic validation of cells can prove challenging.

Sanger sequencing of CRISPR-derived clones suggested successful correction of the $POLE^{P286R}$ mutation to wildtype in HCC2998 and SNU-81 colorectal cancer cell lines. However, fluorescence in-situ hybridisation of *POLE*, the Chr12 telomere and the Chr12 centromere revealed a deletion of the $POLE^{P286R}$ allele which was not detectable by Sanger sequencing, thus generating hemizygous $POLE^{WT}$ isogenic cell lines. The close proximity of the *POLE* gene to the end of chromosome 12 (617.5kb) may have resulted in deletion of the distal portion of the chromosome, including *POLE*, during CRISPR editing, although positive selection of pre-existing hemizygous cells during single-cell cloning is also a possibility. Consequently, the possible confounding by *POLE* hemizygosity was taken into consideration throughout my analysis of these cell lines, although it should be noted that humans with heterozygous germline truncating mutations in *POLE* have been identified and have no observable phenotype (unpublished data, Tomlinson laboratory).

Whilst the results that I have generated clearly demonstrate the utility of these cell models, it is of interest to repeat these analyses in isogenic cell lines with two copies of the *POLE* widltype allele when generated.

Since the initial publication of the use of CRISPR-Cas9 in genome editing in 2013 (250), there has been much improvement in the specificity and efficiency of the technique. To increase targeting efficiency, use of 5' and 3' phosphorothioate modified

ssODNs decrease exonucleolytic degradation at exposed oligo termini (263). Alternatively, a plasmid-integrated dsDNA repair template would permit integration of a fluorescent marker into the genome to select for successfully edited cells and track non-specific oligonucleotide integration. There is growing evidence that synchronising cells in metaphase, with nocodozole, or S-phase, using a thymidine block, can enhance CRISPR-mediated homology-directed repair (33% compared to 10% in unsynchronised cells) (251, 264). Recently, small molecules have been used to increase homology-directed repair, either by inhibiting proteins involved in the NHEJ pathway, such as Ligase IV by SCR7 or L755507 (265, 266), or by stimulating homology-directed repair pathway proteins such as RAD51 by RS-1 (267).

Nevertheless, as data from my isogenic cell lines have demonstrated, screening for off-target effects of CRISPR genome editing and/or clonal selection using both sequencing and cytogenetic techniques is crucial to confirm the genomic sequence and structural integrity of isogenic cell lines.

### 3.4.2 Causal role of *POLE* exonuclease domain mutation in ultramutation and mutation Signature 10

Using the isogenic lines, I have provided the first evidence in a human cell line that clinically relevant $POLE^{P286R}$ mutations have a direct causal role in tumour ultramutation and the characteristic mutation Signature 10.

Although mutation rate and spectrum results from *POLE* isogenic cell from the *HPRT1* assay were concordant with the published literature on *POLE* exonuclease domain mutant cancers (1, 161, 162), it was not possible to measure the global mutation rate and spectrum of the lines as the *HPRT1* assay surveys a single locus in the human genome. Hodel *et al.* reported a 5.8 fold difference in mutation rate between the *HPRT1* assay and whole exome sequencing of HCT116 cells. Mutation rate can vary throughout the genome due to variations in early and late replicating regions, chromatin structure and methylation patterns. Whole genome sequencing of isogenic Parental and CRISPR-WT cell lines will allow greater statistical power to compare the mutation rate and spectrum, particularly in *POLE* exonuclease domain wildtype cell lines with a relatively low mutation rate (241).

The mutation rate varied by several orders of magnitude between HCC2998 and SNU-81 Parental cell lines, potentially resulting from different genetic mutations between cell lines or variations in *HPRT1* gene copy number. Both Parental cell lines, however, demonstrated an ultramutator phenotype compared with CRISPR-WT clones. Importantly, this difference was demonstrated not to simply be the result of secondary mutations and genome aberrations in cancer cells, which could potentially act as secondary drivers of mutagenesis.

### 3.4.3 *POLE* exonuclease domain mutations do not cause a substantial replication defect in CRC cell lines

Studies of *POLE* exonuclease domain mutations in cell-free assays and in yeast have demonstrated that $POLE^{P286R}$ confers a mutation rate that exceeds *POLE* exonuclease active site mutants by several orders of magnitude (161, 162, 163), suggesting a defect beyond loss of proofreading.

Despite the crucial role of Pol $\epsilon$ in replicating the leading strand (94), data from cell proliferation, cell cycle and DNA fibre analysis in isogenic CRC cell lines demonstrate that *POLE* exonuclease domain mutations do not cause a substantial replication defect under normal growth conditions, although subtle variation in fork speed was observed in DNA fibre analysis. Preliminary DNA fibre data from hydroxyurea-treated cells suggests that exonuclease domain mutant polymerases may be more tolerant to a reduction in dNTP levels, or suffer less DNA damage than wildtype polymerases, which requires further investigation.

Haploid fission yeast with $POLE^{P286R}$ mutations have delayed S-phase progression (unpublished, Kearsey lab, University of Oxford). It is possible that a replication defect in heterozygous *POLE* exonuclease domain mutant human cell lines are below the limit of detection for the assays used under normal conditions, as the wildtype polymerase can mask variations in replication speed. Generation of a hemizygous or homozygous $POLE^{P286R}$ mutant cell line will permit investigation of the mutant polymerase *in vivo* without interference from the wildtype polymerase. To assess DNA replication, cell-free assays of the human $POLE^{P286R}$ holoenzyme will determine its intrinsic polymerisation rate.

### 3.4.4 Relationship of *POLE* exonuclease domain mutation with MMR loss

Perhaps surprisingly, given that most *POLE* exonuclease domain mutant cancers are microsatellite stable, Shlien and collegues have hypothesised that mismatch repair deficiency is essential for *POLE* mutagenesis (164, 241). However the effect of *POLE* on ultramutation that I have demonstrated in microsatellite stable, MMR-proficient HCC2998 and SNU-81 cell lines show no evidence of a requirement for mismatch repair deficiency for ultramutation, which is inconsistent with this hypothesis. My results add to the evidence that microsatellite unstable cancer with combined *POLE* mutations and MMR-deficiency are likely a distinct cancer subset that have a unique mutation signature and prognosis and the mechanism of mutagenesis should be treated as distinct from classic MSS *POLE* exonuclease domain mutant cancer (171, 210).

I hypothesised that mismatch repair deficiency in *POLE* exonuclease domain mutant cells would increase the mutation burden to a lethal threshold, resulting in cell death. Transient knockdown of MSH6 and MLH1 did not result in reduced cell viability, nor did it lead to an elevated DNA damage response in *POLE* exonuclease domain mutant cells. Haploid budding yeast with dual loss of polymerase exonuclease activity and MMR are either unviable or have a severe growth defect (128, 129). In comparison, diploid human cell lines have a longer doubling time, thus may require constitutive knockdown of mismatch repair genes, using CRISPR-Cas9 or shRNA, to observe a reduction in cell viability, or increase in DNA damage response.

### 3.4.5 Nucleoside analogues may not be a suitable treatment for heterozygous *POLE* exonuclease domain mutant cancers

Using the isogenic cell lines, I demonstrated no differential sensitivity to the commonly used clinical cytotoxic camptothecin or nucleoside analogues, 5-FU and fludarabine, between *POLE* exonuclease domain mutant and *POLE* wildtype isogenic cancer cell

lines. Crucially, my data demonstrates the importance of using genetically homogeneous isogenic cell lines for comparative cell viability assays based on a single gene mutation, as previous data from the Tomlinson laboratory in heterogeneous CRC cell lines suggested that *POLE* exonuclease domain mutant cells had significantly increased sensitivity to 5-FU.

Van Gool *et al.* proposed the nucleoside analogues fludarabine and cytarabine as potential treatment options for *POLE* exonuclease domain cancers (243). These data are supported by previous studies in *POLE* exonuclease active site mutant human TK6 cells which are acutely sensitive to cytarabine (242). The authors suggest that the mutant polymerase is less able to extend from the misincorporated nucleoside analogue, causing replication fork stalling, fork collapse and cell death (242). However, both studies are limited by their use of murine embryonic stem cells (243) and homozygous cancer-associated and exonuclease active site mutant *POLE* (242, 243), which are rarely if not at all observed in ultramutated cancer (Table 3.1).

Data from isogenic *POLE* exonuclease domain mutant cell lines show no differential sensitivity to fludarabine and they were resistant to cytarabine. To put my results into context, exposure to 500$\mu$M cytarabine reduced Parental and CRISPR-WT clone 1 viability by 47% and 46% respectively, whilst 0.14$\mu$M cytarabine killed 50% of heterozygous $POLE^{S297F}$ murine embryonic stem cells (243). Further cell viability assays using cytarabine in SNU-81 isogenic cell lines and the heterogeneous *POLE* exonuclease domain mutant cell panel are required to determine if the cytarabine is a potential treatment option in heterozygous *POLE* exonuclease domain mutant cancers.

In relation to nucleoside analogues, studies in *S. cerevisiae* show that modifications in dNTP pool levels in error-prone *Pole* $\epsilon$ and $\delta$ result in a lethal increase in mutagenesis at high dNTP concentrations, or reverse the mutator phenotype at low dNTP concentrations (133, 268). The result of dNTP pool modification in *POLE* exonuclease domain mutant human cells is an interesting avenue for further investigation into potential therapies for patients with *POLE* exonuclease domain mutant cancers.

### 3.4.6 Summary

Clinically relevant *POLE*$^{P286R}$ mutations have a direct causal role in tumour ultramutation and the *POLE* exonuclease domain mutation-associated mutation Signature 10. Intriguingly, despite Pol epsilon's crucial role in replicating half of all genomic DNA, global DNA replication is unpurturbed and the DNA damage response is not activated in *POLE* exonuclease domain mutant cells. Preliminary data, however, suggests that exonuclease domain mutant Pol $\epsilon$ may be more tolerant to DNA replication stress than wildtype polymerases.

Multiple missense and nonsense mutations in mismatch repair genes such as *MSH6* are present in *POLE* exonuclease domain mutant tumours, however transient knockdown of MMR proteins in *POLE* exonuclease domain mutant cells does not lead to a reduction in cell viability, nor does it increase DNA damage response. It will be interesting to determine if chemically-induced replication stress or constitutive mismatch repair deficiency affect the DNA replication dynamics, DNA damage response and cell viability in isogenic *POLE* exonuclease domain mutant cell lines and if this could reveal potential therapeutic targets.

# Chapter 4

# Characterisation of the *Pole*$^{L424V}$ mutation in mouse models

## 4.1 Introduction

Dominantly inherited germline *POLE* exonuclease domain mutations predispose carriers to colonic adenomas and polyps, in addition to cancers of the intestine, endometrium, breast, duodenum, ovary and brain (Table 1.4) (144). At least eight germline hotspot mutations in *POLE* have been reported to date; these and their clinical associations are shown in Table 1.4.

The most common germline variant is *POLE*$^{L424V}$, which has been reported in over 20 families and as two *de novo* germline variants (123). Carriers of *POLE*$^{L424V}$ are highly predisposed to develop pre-malignant colorectal adenomas/polyps and invasive colorectal carcinoma (92% and 61% respectively), glioblastoma and endometrial cancer and have an early age of disease onset (mean 39 years, range 15-60 years). The reason for the tissue specificity effects caused by these germline *POLE* variants and the early development of colorectal polyps, adenomas and cancer is currently unclear.

In this Chapter, I sought to generate and phenotypically characterise *Pole*$^{L424V}$ mutant mice using both constitutive and tissue-specific conditional mouse models.

### 4.1.1 Existing *Pole* exonuclease domain mutant mice

To date, two mouse models with alterations within the *Pole* exonuclease domain have been published (139, 240). In the first, the two codons of the exonuclease active site residues D272 and E274 (corresponding to D275 and E277 in humans), were substituted to alanine, resulting in complete loss of *Pole* exonuclease activity, hereafter referred to as exo-null (125, 139).

Homozygous *Pole* exo-null mice developed tumours, consisting primarily of intestinal adenomas and carcinomas (45%), histocytic sarcomas (36%) and nodal and follicular lymphomas (24% and 15% respectively) (139). With the exception of intestinal adenomas and carcinomas, the tumour spectrum in exo-null mice differs markedly from that observed in human carriers of germline *POLE* exonuclease domain mutations, in which lymphomas and sarcomas are uncommon. Heterozygous exo-null mice display no macroscopic phenotype, despite having an elevated cellular mutation rate (further detail in Section 1.4.2). The phenotypic discordance of exo-null mice with human germline *POLE* exonuclease domain mutant carriers and the fact that *POLE* exonuclease domain mutations are seldom found in human cancers, means that the *Pole* exo-null mouse is of limited clinical relevance as an experimental model.

The second of these, a conditional heterozygous Pole$^{P286R/+}$ allele, modelling the most common somatic *POLE* exonuclease domain mutation in human cancer was reported in 2018 (240). When crossed with the ubiquitous *Ddx4*-Cre promoter (269), heterozygous mutants showed a striking reduction in survival compared to wildtype controls, with a median survival of 138 days (range 72-293 days). *Pole*$^{P286R/+}$ mice developed multiple invasive tumours (average 2.1/animal) across a diverse range of tissue types including the lung, thymus, mammary glands, uterus, skin and intestine (240).

Analysis of *Pole*$^{P286R/+}$ mouse embryonic fibroblasts revealed an elevated mutation rate using the *Hprt1* forward mutation assay, and tumour WGS revealed hyper- and ultramutation ranging from 10-100 mutations/Mb, similar to that detected in human tumours with somatic *POLE* exonuclease domain mutations. Notably, these tumours had a high frequency of TCT→TAT, C→T and TTT→TGT mutations, recapitulating the mutation signature observed in human tumours with somatic *POLE* exonuclease domain mutations.

Interestingly, hemizygous $Pole^{P286R}$ mice exhibited embryonic lethality (240), a similar observation to the embryonic lethality in homozygous *Pole* exo-null mutants crossed with homozgous *Mlh1*-deficient or *Msh2*-deficient alleles (139). These data support the hypothesis that only a single copy of a *POLE* exonuclease domain mutation is capable of causing an increase in the mutation rate sufficient to initiate tumourigenesis and are consistent with the notion that homozygous *POLE* exonuclease domain mutations, or loss of heterozygosity may be lethal to the cell (144, 158, 162).

### 4.1.2 *Adenomatous Polyposis Coli (Apc)* mutant mouse models

The Adenomatous Polyposis Coli (*APC*) gene encodes a 312kDa protein involved in a broad range of cellular functions, including cell adhesion, apoptosis, mitotic chromosomal segregation and $\beta$-catenin regulation in the Wnt signalling pathway.

The Wnt signalling pathway, depicted in Figure 4.1, is responsible for regulating stem cell proliferation and differentiation in the mammary glands, stomach, kidney, ovary and intestine (270, 271).

In the absence of Wnt binding to its ligands, Frizzled and LRP6, $\beta$-catenin binds to a "destruction complex" comprising GSK3$\beta$, Axin and APC, resulting in $\beta$-catenin phosphorylation, ubiquitination and proteasome degradation. In the presence of Wnt, GSK3$\beta$ of the "destruction complex" is inhibited by GBP, increasing $\beta$-catenin levels, resulting in the activation of genes involved in cellular proliferation, apoptosis and cell cycle progression such as *Lgr5*, *Axin*, *CCND1* (Cyclin-D1) and *MYK* (272). APC loss of function disrupts the destruction complex, resulting in the failure to degrade $\beta$-catenin, mimicking constitutive Wnt pathway signalling, leading to increased cellular proliferation, invasion and ultimately metastasis (270, 273).

APC is a classic tumour suppressor gene, with biallelic loss of function mutations occurring in >80% somatic CRC cases. Germline *APC* mutations cause familial adenomatous polyposis (FAP), a dominant inherited predisposition to colorectal cancer (270, 274). Somatic mutations tend to occur between codons 1284-1580 (mutation cluster region), with mutation hotspots at amino acids 1061, 1309 and 1462 (274, 275). Interestingly, while *APC* is frequently mutated in sporadic and germline *POLE* ex-

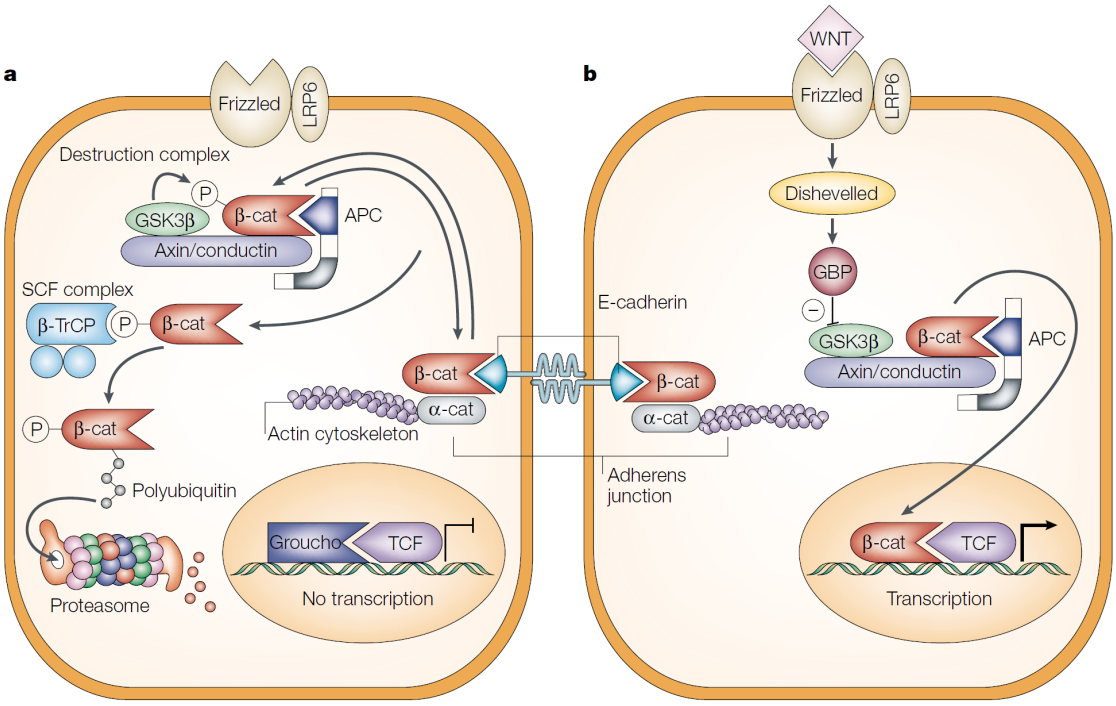



**Figure 4.1:** Wnt signalling pathway schematic from Fodde *et al.* (270) a) In the absence of Wnt protein, free $\beta$-catenin is recruited to a "destruction complex" with GSK3$\beta$, Axin and APC resulting in its phosphorylation, ubiquitination and proteasome degradation. This inhibits the transcription of genes involved in cellular proliferation and adherence. b) In the presence of Wnt, Wnt binds in a complex with the receptor protein Frizzled. This in complex with Lrp6 activates Dishevelled and GBP inhibiting GSK3$\beta$ and inactivating the $\beta$-catenin destruction complex. Increased cytoplasmic $\beta$-catenin can enter the nucleus and activate the transcription of cellular proliferation genes in complex with TCF.

onuclease domain mutant tumours (10), the specific *APC* amino acid substitutions differ from those commonly found in sporadic colorectal cancer, with recurrent truncating mutations at codons 1338 and 1114 (123). The latter is noteworthy, as this variant only partially deletes the $\beta$-catenin binding domain, raising the possibility that residual $\beta$-catenin binding may result in incomplete Wnt pathway activation in *POLE* exonuclease domain mutant cancers compared to traditional *APC* hotspot mutations (Figure 4.2).

Several mouse models have been generated to recreate the *APC* mutations and consequent polyposis observed in human cancer. The most widely studied of these models is *Apc*$^{\text{Min}}$ (Multiple intestinal neoplasia). This was generated through an *N*-ethyl-*N*-nitrosourea (ENU) random mutagenesis screen (233) and carries a trun-

cating *Apc* mutation at codon 850, similar to germline *APC* mutations in human patients with FAP. Mechanistically, this alteration deletes the $\beta$-catenin binding domain, resulting in constitutive $\beta$-catenin activation (Figure 4.2). Similar to patients with FAP, *Apc*$^{\text{Min}}$ mice develop >100 polyps in the small intestine and have a risk of developing mammary and endometrial adenocarcinomas (276).

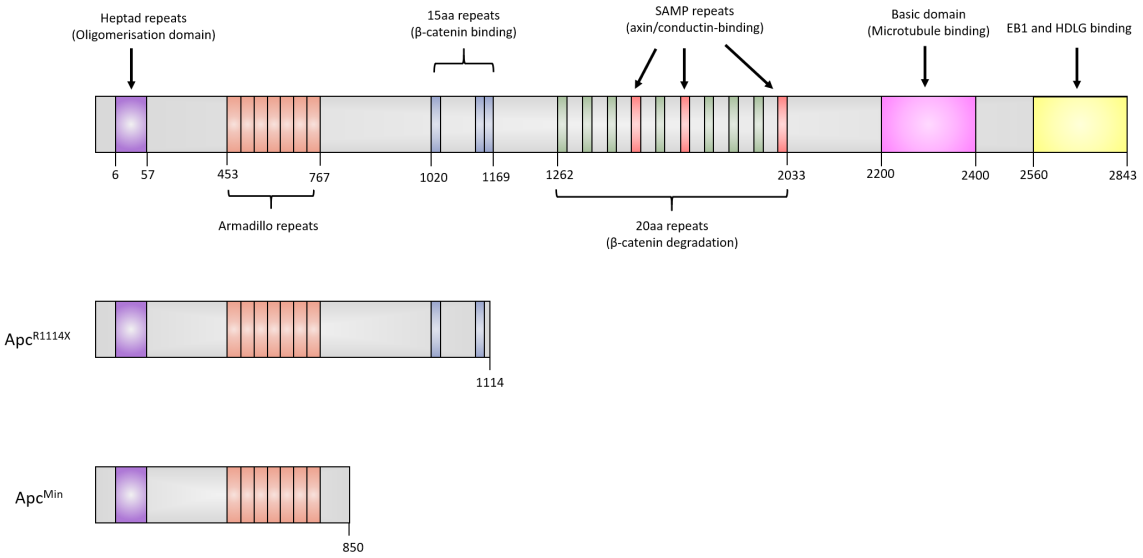



**Figure 4.2:** Human APC protein domains. APC protein domains are highly conserved in the murine Apc protein (90.2% sequence identity, UniProt amino acid alignment). The remaining protein domains for human *APC* R1114X, (mutation hotspot in *POLE* mutant colorectal and endometrial cancer) and Apc$^{\text{Min}}$ mouse (truncated at codon 850) are depicted below. Adapted from Sieber *et al.* (273).

## 4.2 Tissue-specific recombination by Cre-loxP

Cre-LoxP recombination permits the generation of tissue-specific and inducible genetic alterations in mice. The enzyme Cre (cyclisation)-recombinase and LoxP (Locus of x(cross)-over in P1) sites are derived from bacteriophage P1, where genomic recombination is a component of the viral lifecycle. Cre-recombinase is a 38kDa enzyme that catalyses recombination between two LoxP recognition sites, comprised of two 13bp palindromic recognition sites separated by an 8bp asymmetric spacer region.

The orientation of two LoxP sites determines the specific genomic rearrangement induced by Cre recombinase: two LoxP sites in the same orientation on the same

DNA strand will result in deletion of the enclosed DNA sequence; two inverted LoxP sites on the same DNA strand will result in an inversion within the enclosed region; two LoxP sites in the same orientation on different DNA strands, will result in a translocation. A gene segment flanked by LoxP sites is classed as a "floxed" allele.

Spacial control of Cre expression is possible when under a tissue-specific promoter. This can be from ubiquitous expression under the 3-phosphoglycerate kinase (Pgk) promoter which is expressed from the diploid phase of oogenesis, to tissue-specific expression when under the Progesterone Receptor promoter which is active in the uterus, ovary, oviduct, pituitary and mammary glands.

Temporal control of recombination is possible using inducible Cre-recombinases. One of the most prevalent examples is the tamoxifen-inducible Cre recombinase oestrogen receptor T2 fusion protein (CreER$^{T2}$), which is unable to bind to oestrogen under endogenous concentrations, but can bind to the synthetic oestrogen analogue, tamoxifen, at a lower concentration. In the absence of tamoxifen, CreER$^{T2}$ remains in the cytoplasm. Binding of tamoxifen to ER$^{T2}$ enables Cre translocation into the nucleus to initiate recombination.

It is possible for both spacial and temporal control of Cre recombinase expression using an inducible Cre recombinase under a tissue-specific promoter. One example is the *Lgr5-EGFP-IRES-CreER*$^{T2}$ allele, from which CreERT2 is expressed in the stem cells of the intestine (and other tissues) and translocates to the nucleus upon exposure to tamoxifen.

## 4.3 Aims of chapter

I hypothesised that expression of the mutation corresponding to the most common germline *POLE* germline exonuclease domain mutation, L424V, in the mouse would recapitulate the tumour phenotype observed in patients and that study of this model will help determine the mechanism underpinning this. I sought to do this by a combination of constitutive (germline) expression of the mutant allele and its conditional expression in clinically relevant tissues of the intestine and uterus.

Specific aims included:

- Generation of a *Pole* L424V-floxed (*Pole*$^{fl(L424V)/+}$) knock-in mouse model by

homologous recombination.

- Generation of a constitutive mouse model by breeding *Pole*$^{fl(L424V)/+}$ mice with *PgkCre* mice.
- Generation of a conditional intestinal-specific mouse model by breeding *Pole*$^{fl(L424V)/+}$ mice with inducible *Lgr5-EGFP-IRES-CreER*$^{T2}$ (*Lgr5CreERT2*) mice and assess the intestinal phenotype.
- Generation of a uterine-specific mouse model by breeding *Pole*$^{fl(L424V)/+}$ mice with *PRCre* (Progesterone-Receptor Cre) mice and assess the phenotype in the uterus and mammary glands.
- To cross constitutive (germline) *Pole*$^{L424V/+}$ mice with *Apc*$^{Min}$ mice to determine if *Pole*$^{L424V/+}$ modifies the *Apc*$^{Min}$ intestinal polyposis phenotype.

## 4.4 Results

### 4.4.1 Generation of a conditional *Pole* L424V knock-in mouse model

Murine and human *Pole* and *POLE* genes respectively are well conserved with both containing 49 exons at 90.6% sequence identity (ClustalO alignment (277)) (Table 4.3a). Codon L424, located within exon 13 is identical in mice and humans. The construct to knock-in the inducible *Pole* L424V point mutation (c.1270C→G) (Figure 4.3b) was designed by Professor Ian Tomlinson, University of Birmingham and synthesised by Ozgene ltd.

Expression of the *Pole*$^{L424V}$ knock-in mutant allele requires two independent Cre recombination events. To ensure these are discrete, the construct includes two conventional LoxP sites and two mutant LoxP2272 sites which are not cross-compatible. Prior to recombination, the floxed *Pole* construct expresses the wildtype *Pole* sequence as wildtype exon 13 is orientated 5'-3' in the direction of transcription (Figure

4.3b). The mCherry cassette, *PGK* promoter and antisense L424V mutant exon 13 are predicted to be removed by splicing.

a)

| | Chromosome | Protein size | Total Exons | L424V Exon | Sequence Identity | *Pole* L424V (CTC>GTC) amino acid context |
|---|---|---|---|---|---|---|
| **Human** | 12 | 2286 (262kDa) | 49 | 13 | 90.6% | 420 430<br>WVKRDSYLPVGSHNLKAAAKAKLGYDPVE<br>WVKRDSYLPVGSHNLKAAAKAKLGYDPVE<br>********** ** ********** |
| **Mouse** | 5 | 2283 (262kDa) | 49 | 13 | | |

b)

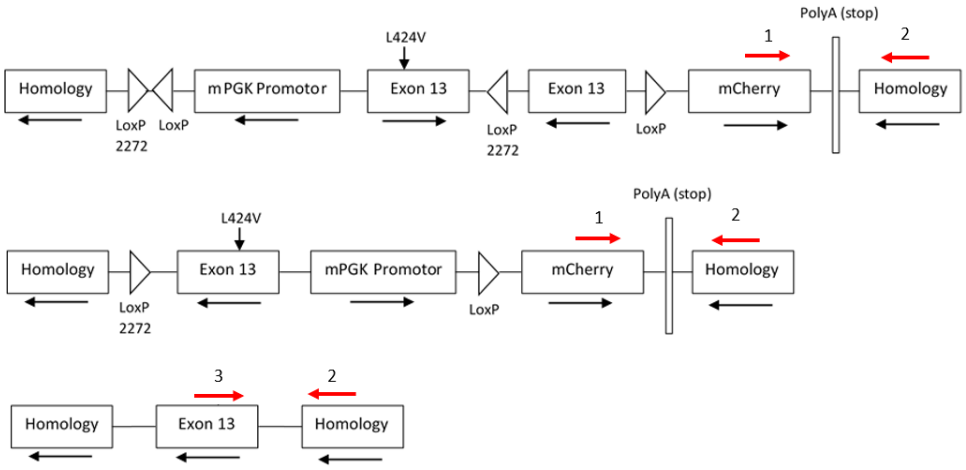



**Figure 4.3:** Conservation of *POLE* and the *Pole*$^{fl(L424V)}$ knock-in targeting construct. a) Human and murine POLE protein information and sequence identity in the context of the L424V mutation. The amino acids are perfectly conserved around the *POLE*$^{L424V}$ mutation. b) The top diagram represents the targeting construct to generate the L424V mouse. Areas of homology represent the sequences required for accurate homologous recombination into the mouse genome. The second diagram down represents the recombined construct following Cre recombination. The L424V mutant exon is in the coding direction and the mCherry cassette is expressed via the *PGK* promoter. Finally the third diagram is the wildtype *Pole* allele. The black arrows represent the coding orientation of the exon, mCherry gene and PGK promoter. The red arrows represent the primers used for genotyping the mutant and the wildtype alleles.

In the presence of Cre, the first recombination event is predicted to be an inversion of the sequence between the two LoxP2272 sequences. This will orientate L424V mutant exon 13 in the direction of transcription and orientate the *PGK* promoter to drive mCherry gene expression. This inversion will also flip the enclosed LoxP site to the same orientation as the second LoxP site, enabling Cre-mediated deletion of the wildtype exon 13. Thus, the fully recombined construct should express both the *Pole* L424V mutant allele under the endogenous promoter and the mCherry reporter gene

under the constitutive *PGK* promoter.

Blastocyst injection and initial chimeras were generated in the laboratory of Dr David Adams at The Wellcome Sanger Institute, University of Cambridge and were on a C57BL/6N background. I took over this project upon receipt of the $Pole^{fl(L424V)/+}$ mice to the facility at which point, no breeding or phenotyping had been performed.

Full nomenclatures and a description of all crosses used throughout this chapter are listed in Appendix Table D.1

For routine genotyping of heterozygous, unrecombined *POLE* L424V mutant mice (hereon referred to as $Pole^{fl(L424V)/+}$), I designed a multiplex PCR (primers listed in Appendix Table A.4) covering a common genomic sequence (primer 2), mCherry (primer 1) and exon 13 (primer3). The results of a representative reaction are shown in Figure 4.4 and demonstrate a 364bp band corresponding to the wildtype allele and a 668bp band corresponding to the $Pole^{fl(L424V)}$ allele.

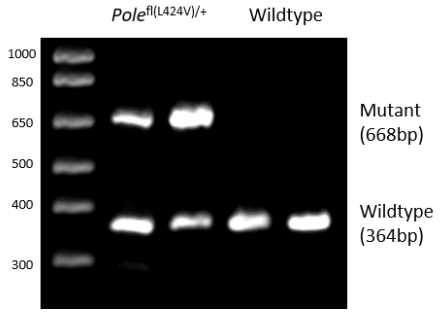



**Figure 4.4:** Multiplex PCR from DNA extracted from two heterozygous $Pole^{fl(L424V)/+}$ (lanes 1 and 2) and two wildtype (lanes 3 and 4) mice. The gel shows bands from the mutant (668bp) and wildtype (364bp) *Pole* alleles amplified using primers 1,2 and 3.

Heterozygous $Pole^{fl(L424V)/+}$ mice were crossed and offspring genotyped, however after eight independent crosses (21 litters), no homozygous $Pole^{fl(L424V)/fl(L424V)}$ mice were identified ($P = 8 \times 10^{-8}$, $\chi^2$ test) (Table 4.1).

Due to the complex nature of the unrecombined $Pole^{fl(L424V)/+}$ allele, which contained an mCherry cassette and *PGK* promoter in introns 12 and 13 respectively, it was plausible that these components could disrupt expression of *Pole* and surrounding genes to an extent that could prove lethal in homozygosity. To investigate this, fresh-frozen small intestine, uterus and kidney tissue were harvested from from five $Pole^{fl(L424V)/+}$ and five wildtype mice. mRNA was purified, reverse transcribed

| Genotype | Male | Female | Total |
|---|---|---|---|
| **fl(L424V)/fl(L424V)** | 0 | 0 | 0 |
| **fl(L424V)/+** | 30 | 28 | 58 |
| **Wildtype +/+** | 19 | 17 | 36 |
| **Total** | 49 | 45 | 94 |

**Table 4.1:** Contingency table of *Pole*$^{fl(L424V)/+}$ x *Pole*$^{fl(L424V)/+}$ offspring. No homozygous *Pole*$^{fl(L424V)/fl(L424V)}$ offspring were generated, suggesting embryonic lethality of homozygotes ($P$ = 8 x $10^{-8}$, $\chi^2$ test). Mice from from 21 litters from 8 breeding pairs were analysed (n = 94).

into cDNA and *Pole* gene expression quantified using RT-qPCR. Using a modified $\delta\delta$Ct method incorporating the use of two housekeeping genes *Actb* and *Gapdh* (278), relative *Pole* gene expression was determined.

This analysis revealed a significant increase in *Pole* mRNA expression in the small intestine and kidney of *Pole*$^{fl(L424V)/+}$ mice, though interestingly no difference was observed in the uterus (Figure 4.6).

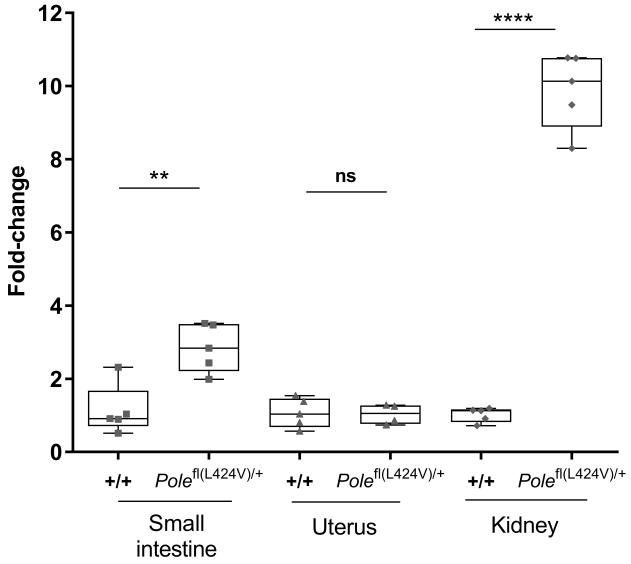



**Figure 4.5:** Fold-change in *Pole* mRNA expression in the small intestine, uterus and kidney from five *Pole*$^{fl(L424V)/+}$ mice and five wildtype (+/+) mice. TaqMan probe Mm00448288_m1 was used to measure *Pole* cDNA levels relative to housekeeping genes *Actb* (Mm02619580_g1) and *Gapdh* (Mm99999915_g1) (Probes listed in Methods Table 2.20). **$P$<0.01, ****$P$<0.0001, ns - $P$= 0.89 (Student's $t$-test).

These data demonstrate that the *Pole*$^{fl(L424V)}$ allele is associated with increased *Pole* mRNA expression in heterozygous *Pole*$^{fl(L424V)/+}$ mice and may be associated

with embryonic lethality of homozygotes.

Due to the absence of homozygous *POLE* exonuclease domain mutations in human cancer, the mechanism or timing of embryonic lethality in homozygous *Pole*$^{fl(L424V)/fl(L424V)}$ mice was not investigated. I speculate, however, that the internal PGK promoter may upregulate *Pole* mRNA expression in organs with a high cellular turnover such as the intestine and kidney which may be toxic in homozygous *Pole*$^{fl(L424V)/fl(L424V)}$ embryos. Because *Pole* mRNA levels are upregulated in the clinically relevant tissue of the intestine, Pole$^{fl(L424V)/+}$ mice will be included as a control in further experiments.

### 4.4.2 Generation of a constitutive *Pole* L424V knock-in mouse model

To study the effects of germline Pole$^{L424V}$ mutations, I crossed *Pole*$^{fl(L424V)/+}$ females with Tg(Pgk1-cre)1Lni males (MGI:2178050) (hereon referred to as *PgkCre*) (231), in which the ubiquitously expressed *Phosphoglycerate kinase-1* (Pgk) promoter drives Cre recombinase in all tissues from the diploid phase of oogenesis. Because the *PgkCre* gene is under maternal control, offspring that do not inherit Cre recombinase from the mother typically undergo LoxP recombination, likely due to residual Cre protein within the cytoplasm of the oocyte (231). Offspring from these crosses included *PgkCre*,*Pole*$^{fl(L424V)/+}$, *Pole*$^{fl(L424V)/+}$ (control) and *PgkCre* (control) animals.

#### 4.4.2.1 Confirmation of *Pole* recombination in *PgkCre*,*Pole*$^{fl(L424V)/+}$ mice

Recombination of the floxed *Pole* allele in *PgkCre*,*Pole*$^{fl(L424V)/+}$ mice was confirmed by split PCR, as shown in Figure 4.6a. PCR number 1 involved primers 4 and 5, which amplify the floxed (unrecombined) *Pole* allele at 582bp. PCR number 2, involving primers 4 and 6, amplify recombined *Pole* alleles at 196bp (Primers listed in Appendix Table A.4). Neither PCR should amplify the wildtype allele. The results from the two PCR amplifications show clear recombination in *PgkCre*,*Pole*$^{fl(L424V)/+}$ mice (Figure 4.6a).

Expression of the L424V mutation in the small intestine of *PgkCre*,*Pole*$^{fl(L424V)/+}$ mice was confirmed by Sanger sequencing of exon 13 from extracted cDNA (Ap-

pendix Table A.4). The chromatogram in Figure 4.6b confirmed expression of the c.1270C→G (L424V) mutation, albeit at a lower amplitude than the wildtype allele, suggesting incomplete recombination or under-expression of the mutant *Pole* allele compared to the wildtype allele.

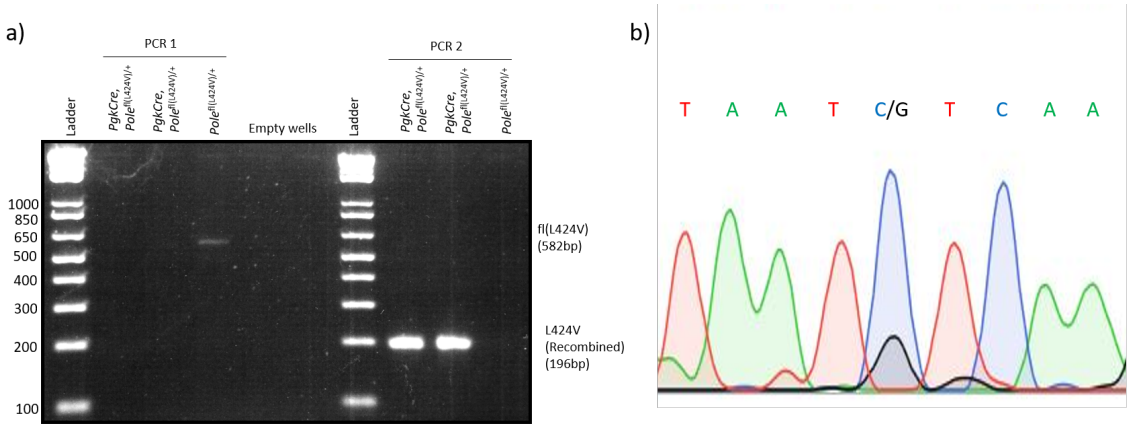



**Figure 4.6:** *Pole* recombination PCR and *Pole* L424V expression chromatogram. a) Recombination PCR from two *PgkCre*,*Pole*$^{fl(L424V)/+}$ mice and one *Pole*$^{fl(L424V)/+}$ mouse. b) Chromatogram of the heterozygous c.1270C→G mutation from cDNA extracted from in intestine of a *PgkCre*,*Pole*$^{fl(L424V)/+}$ mouse. The c.1270C→G mutation is at a lower amplitude than the wildtype allele. cDNA was sequenced using primers 11F and 15R listed in Appendix Table A.4.

In addition to expression of the *Pole* L424V allele, upon recombination, the internal PGK promoter is positioned to drive expression of the mCherry fluorescent protein included for purposes of lineage tracing and cell sorting (Figure 4.3b). Unfortunately, microscopic visualisation of fresh-frozen brain and intestine sections from *PgkCre*,*Pole*$^{fl(L424V)/+}$ mice failed to demonstrate mCherry fluorescent signal (Figure 4.7a), suggesting that the mCherry fluorescent signal may have been too weak to visualise or that the protein had degraded during tissue processing.

This result was confirmed by immunohistochemistry using anti-mCherry antibody and a Red fluorescent protein (RFP)-transfected cell line as a positive control. While the RFP cell line demonstrated obvious mCherry expression, the *PgkCre*,*Pole*$^{fl(L424V)/+}$ mice showed no mCherry expression (Figure 4.7b). These results indicate that either the mCherry gene is not expressed after recombination of the *Pole*$^{fl(L424V)/+}$ knock-in allele or that the level of expression is below the limit of detection for this assay. The reason for this is unclear.

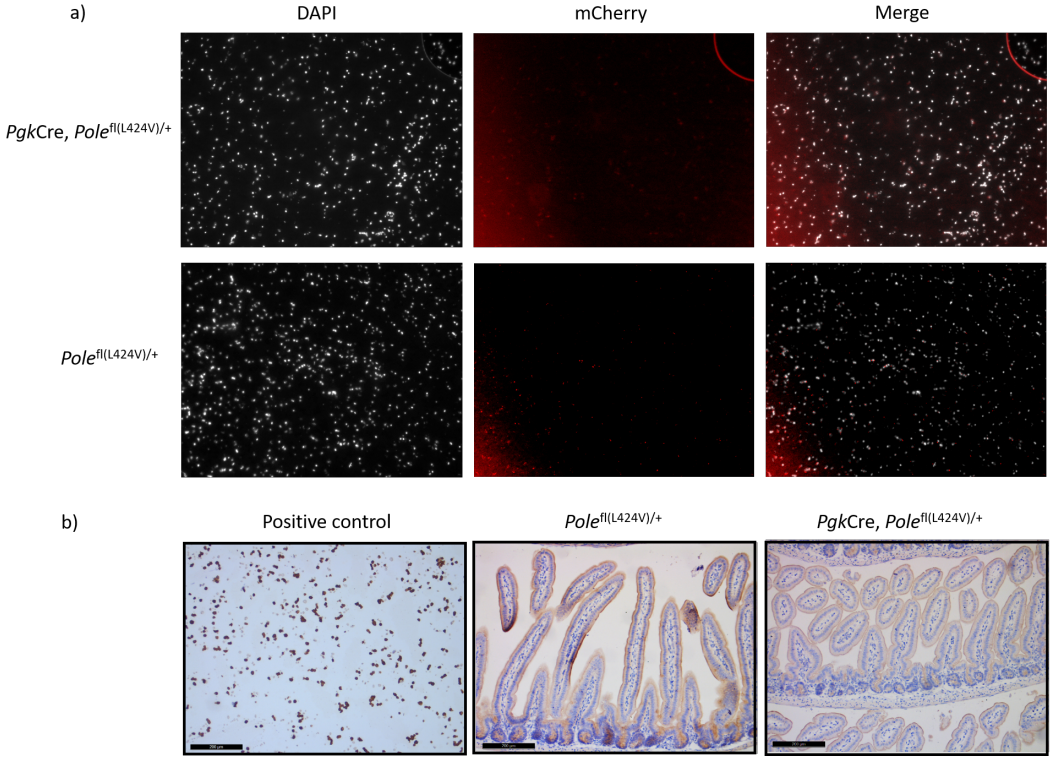



**Figure 4.7:** Analysis of mCherry protein expression. a) Representative images of DAPI-stained DNA and mCherry protein expression in fresh-frozen brain tissue from *Pole*$^{fl(L424V)/+}$ and *PgkCre*,*Pole*$^{fl(L424V)/+}$ mice. Images were acquired using the Olympus-bx-51 microscope and Cytovision Genus v7.1 software at 40X. b) Representative images of mCherry immunohistochemistry from the small intestine of *Pole*$^{fl(L424V)/+}$ and *PgkCre*,*Pole*$^{fl(L424V)/+}$ mice. Positive controls are RFP-expressing cell lines. Both assays were negative for mCherry protein expression.

#### 4.4.2.2 Phenotypic analysis of *PgkCre,Pole*$^{fl(L424V)/+}$ mice

To investigate the role of *Pole* L424V mutations in tumourigenesis, I aged a cohort of four *PgkCre,Pole*$^{fl(L424V)/+}$ (experimental), three *Pole*$^{fl(L424V)/+}$ (control) and five *PgkCre* (control) mice to 52 weeks of age for phenotypic analysis. These animals showed no gross morphological abnormalities at necropsy (Appendix Table E.1). Clinically relevant tissues (intestine, uterus (when possible), brain and lung) were dissected and were normal on microsopic analysis (Appendix Table E.2, Figure E.1a).

A cohort of two *PgkCre,Pole*$^{fl(L424V)/+}$ and two *Pole*$^{fl(L424V)/+}$ mice were aged to 81 weeks (Appendix Table E.1). One *PgkCre,Pole*$^{fl(L424V)/+}$ mouse presented with gastric and neck tumours at necropsy, which were characterised as lymphomas upon microscopic analysis (Figure 4.8). All other mice were normal at necropsy and microscopy (Appendix Tables E.1 E.2, Figure E.1b).

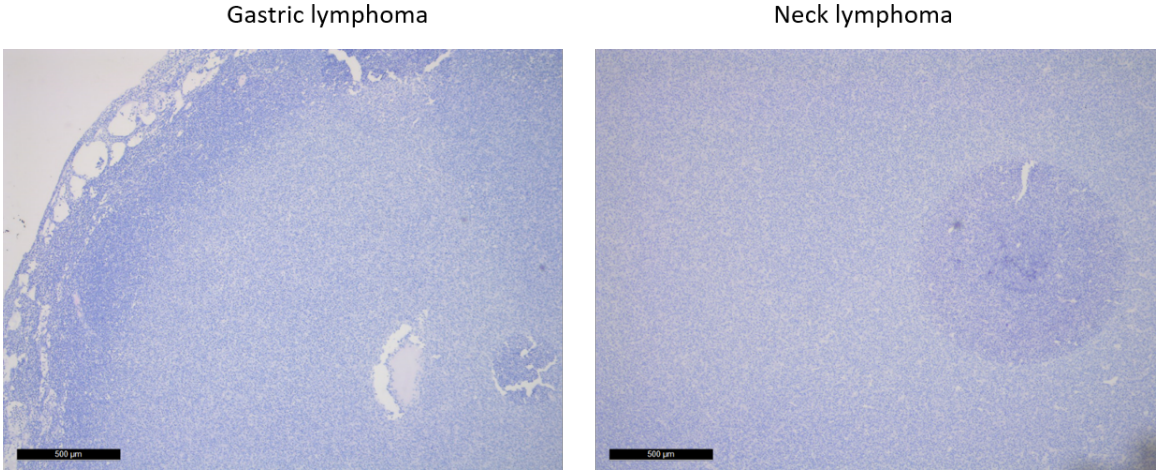



**Figure 4.8:** Gastric (Stomach) and neck lymphomas from a *PgkCre,Pole*$^{fl(L424V)/+}$ mouse at 81.7 weeks. Scale bars represent 500$\mu$m in both images.

At 17.6 weeks one *PgkCre,Pole*$^{fl(L424V)/+}$ mouse developed cachexia and became inactive. An intrathoracic tumour was identified at necropsy. Microscopic analysis confirmed the tumour as a lymphoma (Figure 4.9). Immunohistochemistry showed the lymphoma to be CD3 positive and CD45 negative, which suggest that the tumour is a T-cell lymphoblastic lymphoma (Figure 4.9).

Lymphomas were observed in 2/7 (29%) experimental *PgkCre,Pole*$^{fl(L424V)/+}$ mice and 0/10 control mice ($P = 0.157$, Log-rank test). Lymphomas were also observed in 35% and 55% of homozygous *Pole* exo-null and heterozygous *Pole* P286R mice

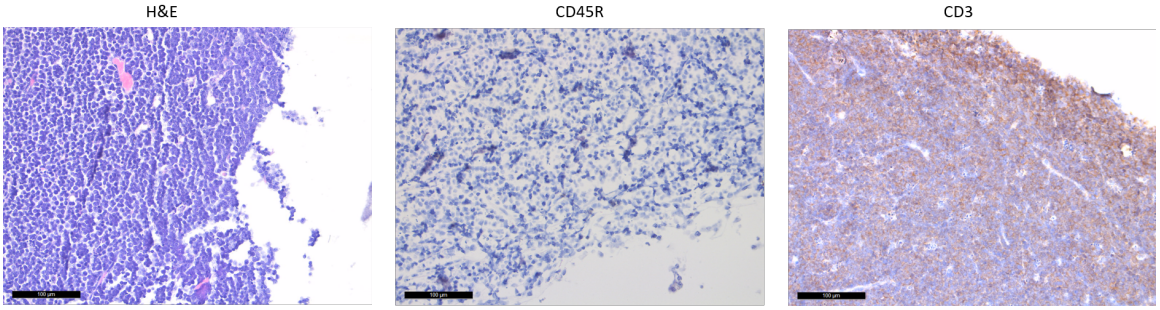



**Figure 4.9:** Representative H&E-staned intrathoracic tumour sections from a 17.6 week *PgkCre*,*Pole*$^{fl(L424V)/+}$ mouse. B and T cell expression was assessed by CD45R and CD3 immunohistochemistry respectively. The tumour was CD45R negative and CD3 positive. Images were taken at 20X and scale bars represent 100$\mu$m.

respectively (139, 240). These data suggest that lymphomas in the experimental cohort may be a result of the *Pole* L424V mutation, however the small cohort meant that I was underpowered to confirm this result.

### 4.4.3 Molecular analysis of *PgkCre*,*Pole*$^{fl(L424V)/+}$ mice

As *PgkCre*,*Pole*$^{fl(L424V)/+}$ mice had not developed macroscopic or microscopic polyps of the intestine by 81 weeks, I next examined molecular markers of proliferation (Ki67) and apoptosis (Cleaved Caspase-3) by immunohistochemistry that might be upregulated prior to the onset of tumourigenesis. Immunohistochemistry was performed on morphologically normal small intestinal tissue sections from three *PgkCre*,*Pole*$^{fl(L424V)/+}$, three *Pole*$^{fl(L424V)/+}$ and three *PgkCre* mice at 52 weeks.

Cleaved Caspase-3 expression was significantly increased in the small intestine of *PgkCre*, *Pole*$^{fl(L424V)/+}$ mice relative to *Pole*$^{fl(L424V)/+}$ mice ($P = 0.047$, Student's $t$-test). Increased expression, however, was not significant between *PgkCre*,*Pole*$^{fl(L424V)/+}$ and *PgkCre* mice (Figure 4.10). There was no significant difference in intestinal Ki67 levels, between *PgkCre*,*Pole*$^{fl(L424V)/+}$, *Pole*$^{fl(L424V)/+}$ and *PgkCre* mice (Figure 4.10).

These data suggest that constitutive expression of the *Pole* L424V mutation increases apoptosis in the small intestine, potentially due to an increase in mutation rate, resulting in an accumulation of pathogenic mutations in essential genes.

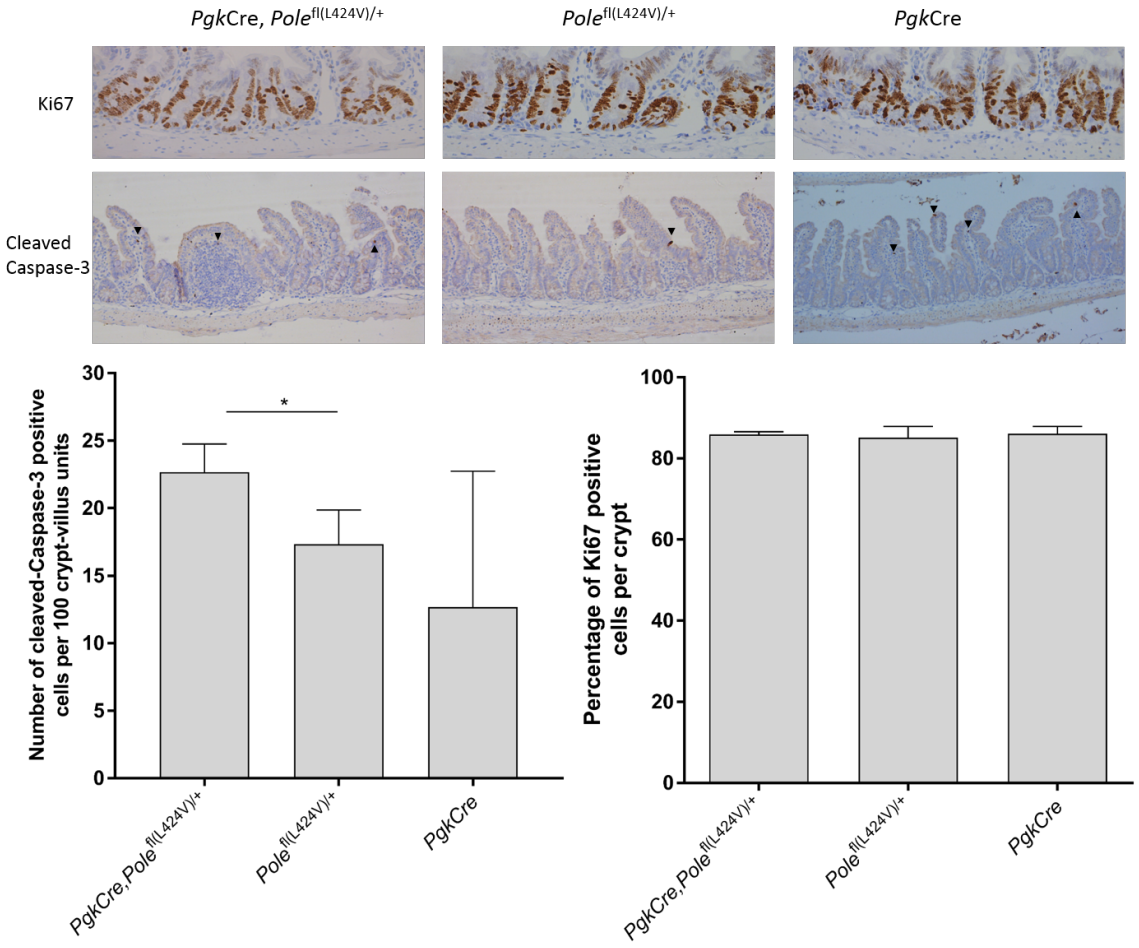



**Figure 4.10:** Molecular analysis of apoptosis and proliferation in morphologically normal 52 week old small intestinal tissue by Cleaved Caspsase-3 and Ki67 immunohistochemistry respectively. Cleaved Caspase-3 levels were quantified from 100 crypt-villus units across SB1, SB2 and SB3. Ki67 levels were assessed by the percentage of Ki67 positive cells across 100 crypt-villus units. Three mice of each genotype (*PgkCre*,$Pole^{\mathrm{fl(L424V)/+}}$, $Pole^{\mathrm{fl(L424V)/+}}$ and *PgkCre*) were assessed. $^{*}P<0.05$, Student's $t$ test, mean + SD (n=3).

#### 4.4.3.1 Generation of a germline *Pole* L424V knock-in mouse

While the *Pgk* promoter is expressed from early embryogenesis, the possibility of mosaic Cre expression and recombination within tissues could not be excluded. To ensure ubiquitous expression of the *Pole* L424V mutation, I generated a germline *Pole*$^{\mathrm{L424V/+}}$ mouse by crossing a *PgkCre*,*Pole*$^{\mathrm{fl(L424V)/+}}$ female with a wildtype male (Generation 1). Because offspring from female *PgkCre* carriers express residual Cre during early embryogenesis in the absence of the inherited *PgkCre* allele, *Pole* L424V/+ genotype offspring should carry the fully recombined allele (hereon referred to as *Pole*$^{\mathrm{L424V/+}}$) in all cells (Generation 2). This prediction was confirmed by recombination-specific PCR discussed in Section 4.4.2.1. To ensure this was also the case in the germline, *Pole*$^{\mathrm{L424V/+}}$ and wildtype mice from Generation 2 were crossed and *Pole*$^{\mathrm{L424V/+}}$ offspring tested for recombination. This confirmed the generation of mice with germline inheritance of the recombined *Pole*$^{\mathrm{L424V/+}}$ mutant allele.

As discussed in Section 4.4.1, the *Pole*$^{\mathrm{fl(L424V)}}$ allele is embryonic lethal in homozygotes. To determine if germline *Pole*$^{\mathrm{L424V/+}}$ mice can produce homozygous offspring, *Pole*$^{\mathrm{L424V/+}}$ males were crossed with *Pole*$^{\mathrm{L424V/+}}$ females. No homozygous offspring from seven litters (two matings, 26 offspring) were detected, confirming that the homozygous germline *Pole*$^{\mathrm{L424V/L424V}}$ allele is embryonic lethal ($P = 0.0036$, $\chi^2$ test).

| **Genotype** | **Total** |
|---|---|
| **L424V/L424V** | 0 |
| **L424V/+** | 14 |
| **+/+** | 12 |
| **Total** | 26 |

**Table 4.2:** Contingency table of offspring from *Pole*$^{\mathrm{L424V/+}}$ x *Pole*$^{\mathrm{L424V/+}}$ matings. No homozygous *Pole*$^{\mathrm{L424V/L424V}}$ offspring were generated, confirming embryonic lethality of homozygotes ($P = 0.0036$, $\chi^2$ test).

To enable survival and phenotypic analysis of *Pole*$^{\mathrm{(L424V)/+}}$ mice, a cohort of 27 *Pole*$^{\mathrm{L424V/+}}$ and 20 wildtype mice are ageing and will be sacrificed upon signs of distress. No mice have become moribund thus far with the oldest mouse currently at 73 weeks of age.

### 4.4.4 Generation of an intestinal-specific *Pole* L424V knock-in mouse model

To further investigate the role of *Pole* L424V in intestinal tumourigenesis, I crossed *Pole*$^{fl(L424V)/+}$ mice with the inducible Cre recombinase, Lgr5$^{tm1(cre/ERT2)Cle}$ (MGI:3764660) (hereon referred to as *Lgr5CreERT2*) (Figure 4.11). *Lgr5CreERT2* is located in the endogenous locus of *Lgr5*, a Wnt target gene expressed in the stem cells of the small intestine and colonic crypts (and stem cells of the stomach, mammary glands, and hair follicle stem cells) (232). Offspring were generated at expected Mendelian ratios.

Experimental *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$ and control *Lgr5CreERT2* and *Pole*$^{fl(L424V)/+}$ offspring recombination was induced by intraperitoneal injection of tamoxifen at six weeks of age. The *Lgr5CreERT2* allele also results in EGFP expression (Figure 4.11). Immunohistochemistry for EGFP protein in FFPE intestinal tissue confirmed *Lgr5CreERT2* expression, thus recombination of the target allele in the intestinal crypt of 26 week old *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$ mice but not in *Pole*$^{fl(L424V)/+}$ controls (Figure 4.11b).

Expression of *Lgr5CreERT2* was mosaic, consistent with the previous literature (279); an important consideration for the phenotypic characterisation of these mice.

#### 4.4.4.1 Phenotyping of *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$ mice

A cohort of five experimental *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$ and five control *Lgr5CreERT2* mice were sacrificed at 26 weeks for phenotypic analysis (Appendix Table F.1). These animals showed no signs of intestinal tumourigenesis (e.g. pale feet, hunching) prior to culling and no gross morphological abnormalities at necropsy (Appendix Table F.1). Further analysis of these animals, by methylene blue staining of the intestine, revealed no polyps (Appendix Figure F.1a) and microscopic analysis of H&E-stained intestinal sections also demonstrate normal appearances with no signs of early microadenomas (Appendix Table F.2, Figure F.1b)

A further cohort three mice of each genotype were aged to 52 weeks. Necropsy of one *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$ mouse revealed splenomegaly, shown in Figure 4.12, but no other macroscopic abnormality; the remaining mice were macroscopically normal at sacrifice (Appendix Table F.1).

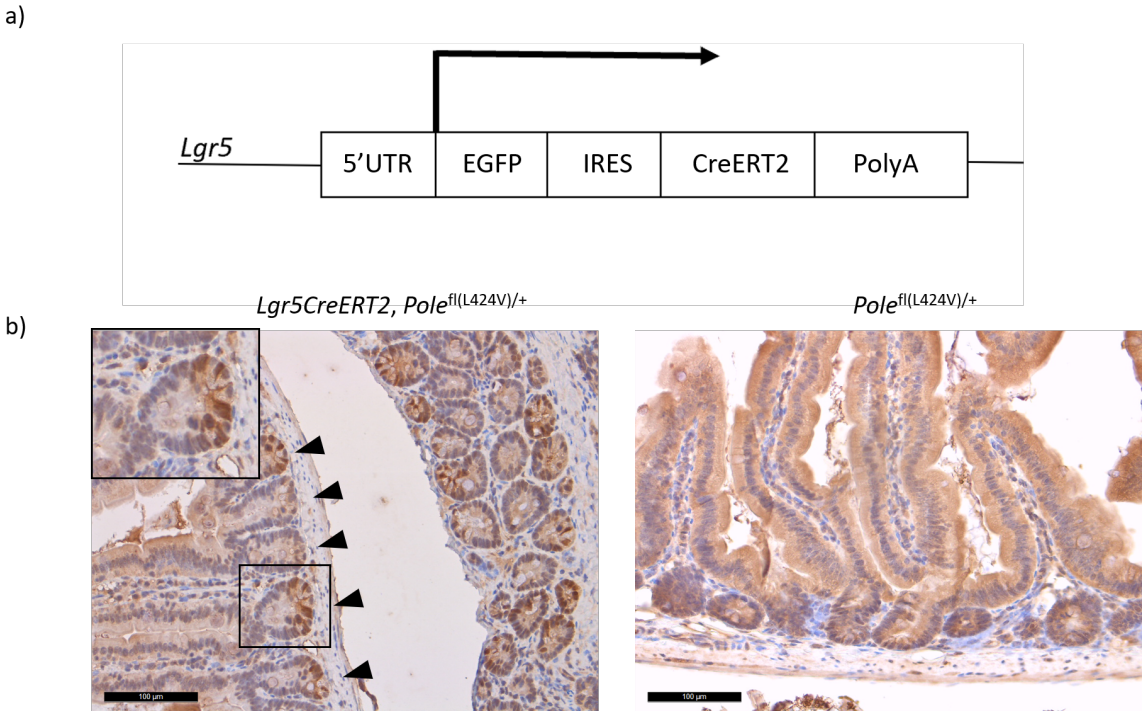



**Figure 4.11:** *Lgr5CreERT2* construct and Lgr5-EGFP protein expression in intestinal stem cells a) Schematic of the *EGFP-IRES-creERT2* (*Lgr5CreERT2*) cassette inserted into codon one of the *Lgr5* gene (232). Translocation of the Cre recombinase to the nucleus for allele recombination is tamoxifen-inducible. b) EGFP immunohistochemistry in experimental *Lgr5CreERT2*,*Pole*$^{fl(L424V)/+}$ and control *Pole*$^{fl(L424V)/+}$ mice (antibody is listed in Table 2.23). EGFP is expressed in the intestinal stem cells of *Lgr5CreERT2*,*Pole*$^{fl(L424V)/+}$ mice as shown by the black arrows. Scale bars represent 100$\mu$m.

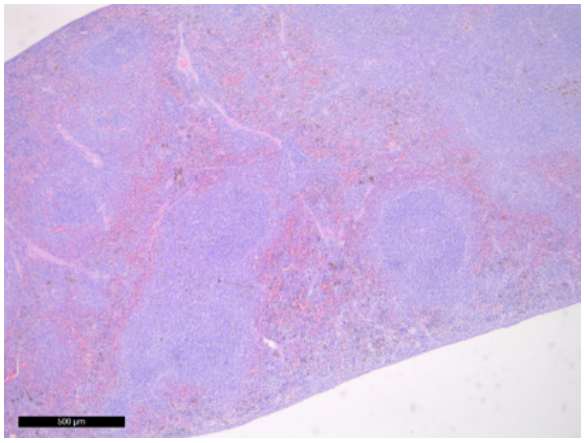


**Figure 4.12:** Microscopic analysis of a H&E-stained spleen section from a 52 week old *Lgr5CreERT2*,*Pole*$^{fl(L424V)/+}$ mouse with splenomegaly. Scale bar represents 500$\mu$m.

Macroscopic and micoscopic analysis of the small intestine and colon from all of these animals showed no polyps or microadenomas respectively (Appendix Table F.2, Figure F.1c).

Given the possibility that a phenotype associated with the L424V mutation may only be evident at a more advanced age, I aged a further five experimental *Lgr5CreERT2*, *Pole*$^{fl(L424V)/+}$, three control *Lgr5CreERT2* and five control *Pole*$^{fl(L424V)/+}$ mice to 85-90 weeks. All *Lgr5CreERT2*,*Pole*$^{fl(L424V)/+}$ were healthy prior to sacrifice and displayed normal macroscopic appearances at necropsy (Appendix Table F.1). Analysis of these animals by methylene blue staining of the intestines revealed no polyps and microscopic analysis of the small intestine and colon were normal (Appendix Table F.1, Figure F.1d).

A control *Lgr5CreERT2* mouse had a large liver tumour which had prominent lymphocytic infiltrate, another control *Lgr5CreERT2* mouse died aged 18 weeks with an inguinal tumour (Figures not shown) and one control *Pole*$^{fl(L424V)/+}$ mouse had splenomegaly at 90.6 weeks, although no cause was identified at necropsy or microscopy.

### 4.4.5 Generation and phenotyping of a uterine-specific *Pole* L424V mouse model

Endometrial cancer is a prominent phenotype in female carriers of germline *POLE*$^{L424V}$ mutations in humans. To assess the impact of *Pole* L424V in the uterus, I used Progesterone Receptor Cre (Pgr$^{tm2(cre)Lyd}$) (MGI:3576366) (hereon referred to as *PRCre*) which is in the endogenous locus, to initiate recombination (280). In addition to uterine expression, the Progesterone Receptor promoter expresses Cre in other tissues under progesterone regulation including the mammary glands, the ovary and the pituitary (280). Offspring were generated at expected Mendelian ratios.

A cohort of five experimental *PRCre*,*Pole*$^{fl(L424V)/+}$ and five control *PRCre* mice were sacrificed at 24 weeks for phenotypic analysis (Appendix Table G.1). All mice were seemingly healthy on sacrifice and showed no gross macroscopic morphological abnormalities at necropsy. Micospcopy of H&E-stained sections from the uterus and mammary glands were also normal (Appendix Table G.2, Figure G.1).

Additional cohorts of these mice (three of each genotype) were aged to 52 weeks and sacrificed. One out of three experimental mice was normal at necropsy, while one had a moderately enlarged uterus, demonstrated to be due to normal post-menopausal fluid retention (hydrometrium) on histological analysis (Figure 4.13), the other had a mammary tumour, which was shown to be normal upon histological analysis (Figure not shown). All control mice were both macroscopically and microscopically normal (Figure 4.13), with the exception of one *PRCre* mouse with mild splenomegaly but normal uterus and mammary glands (Figure not shown).

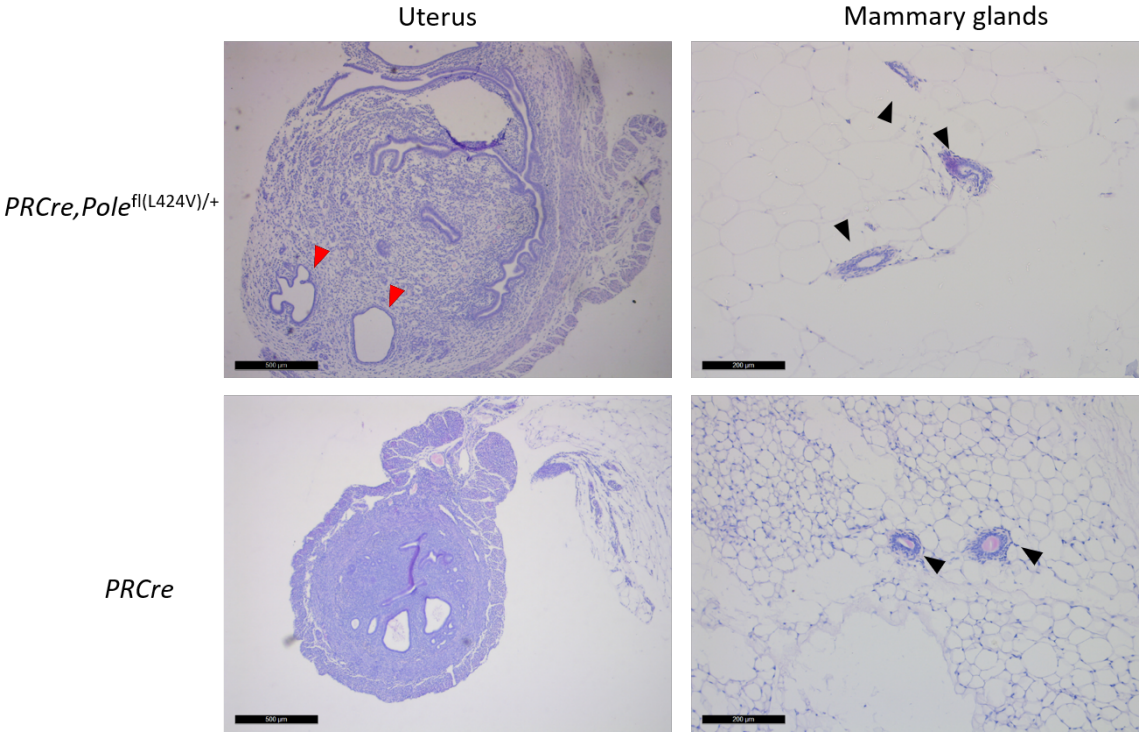



**Figure 4.13:** Representative images of the uterus and mammary glands from one experimental *PRCre*,*Pole*$^{fl(L424V)/+}$ mouse with hydrometrium and one control *PRCre* mouse at 52 weeks. Individual mammary glands are highlighted with black arrows and fluid-filled glands in the uterus are highlighted with red arrows. Scale bars represent 500$\mu$m for the uterus and 200$\mu$m for mammary glands.

Longer-term surival analysis of these animals is presently ongoing. To date, 11 *PRCre*,*Pole*$^{fl(L424V)/+}$ and nine *PRCre* mice have been aged to 77 - 82 weeks (Appendix Table G.1). One experimental *PRCre*,*Pole*$^{fl(L424V)/+}$ was sacrificed for cachexia; necropsy revealed a visible lung tumour which was shown to be a lymphoma on histological analysis (Figure 4.14C). The remaining animals were all seemingly healthy at sacrifice; the most common abnormal finding at necropsy in these

was splenomegaly, evident in 4/11 and 1/9 *PRCre*,*Pole*$^{fl(L424V)/+}$ and *PRCre* mice respectively ($P = 0.32$, Fisher's exact test) (Appendix Table G.1).

Microscopic analysis of H&E-stained uterus, mammary gland and tumour sections showed lymphocytic infiltrate and lymphomas in 2/11 *PRCre*,*Pole*$^{fl(L424V)/+}$ and 1/9 *PRCre* mice ($P = 1$, Fisher's exact test). One *PRCre*,*Pole*$^{fl(L424V)/+}$ mouse had three lymphomas in total in dissected tissues (uterus, mammary gland and lung) (Figure 4.14 and Appendix Table G.2).

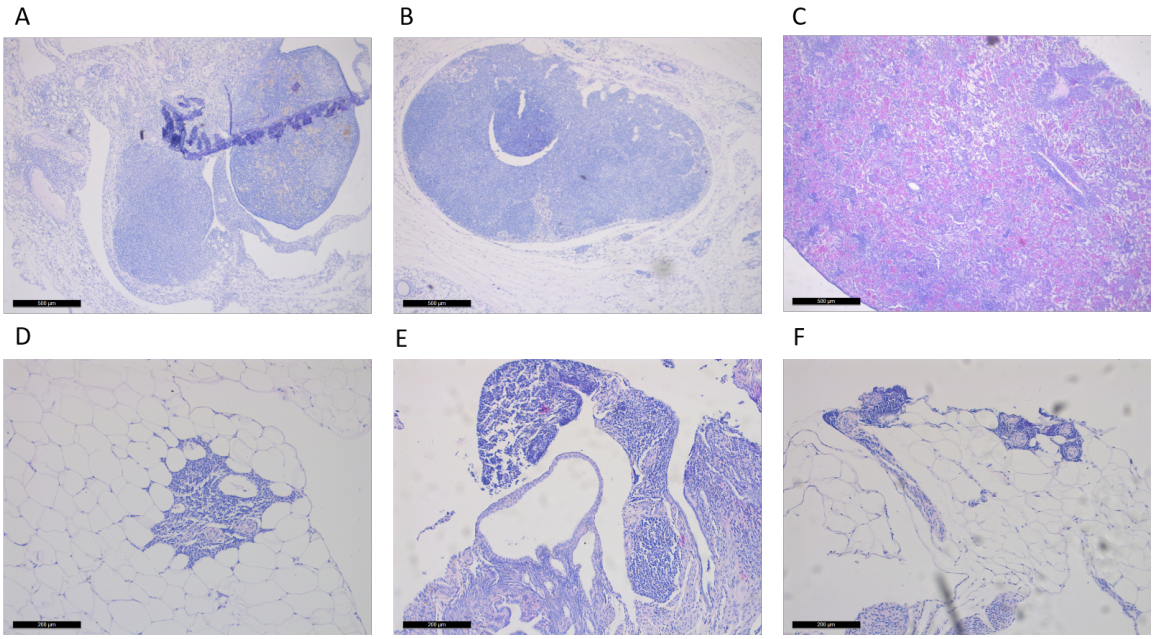



**Figure 4.14:** Lymphomas and organs with lymphocytic infiltrate in *PRCre*,*Pole*$^{fl(L424V)/+}$ and *PRCre* mice. A) Lymphoma within the fat surrounding the uterus, genotype *PRCre*,*Pole*$^{fl(L424V)/+}$. B) Lymphoma within the mammary tissue, genotype *PRCre*,*Pole*$^{fl(L424V)/+}$. C) Lung tumour with lymphocytic infiltrate, genotype *PRCre*,*Pole*$^{fl(L424V)/+}$ D) Lymphocytic infiltrate in the fat surrounding the uterus, genotype *PRCre*,*Pole*$^{fl(L424V)/+}$ E) Lymphoid tissue within the uterus, genotype *PRCre*. F) Lymphocyte cells in the fat surrounding the uterus, genotype *PRCre*. Scale bars in figures A,B and C represent 500$\mu$m and scale bars in figures D,E and F represent 200$\mu$m.

Uterine hyperplasia with glandular crowding and atypical cellular nuclei was observed upon microscopic analysis in one experimental and one control mouse ($P = 1$, Fisher's exact test) (Figure 4.15, Table G.2). Finally, one *PRCre*,*Pole*$^{fl(L424V)/+}$ mouse was suspected to have an ovarian haemorrhagic lesion upon microscopic analysis, however confirmation is required (Appendix Table G.2, Figure not shown).

Overall, there was no statistical difference in gross morphological and microscopic

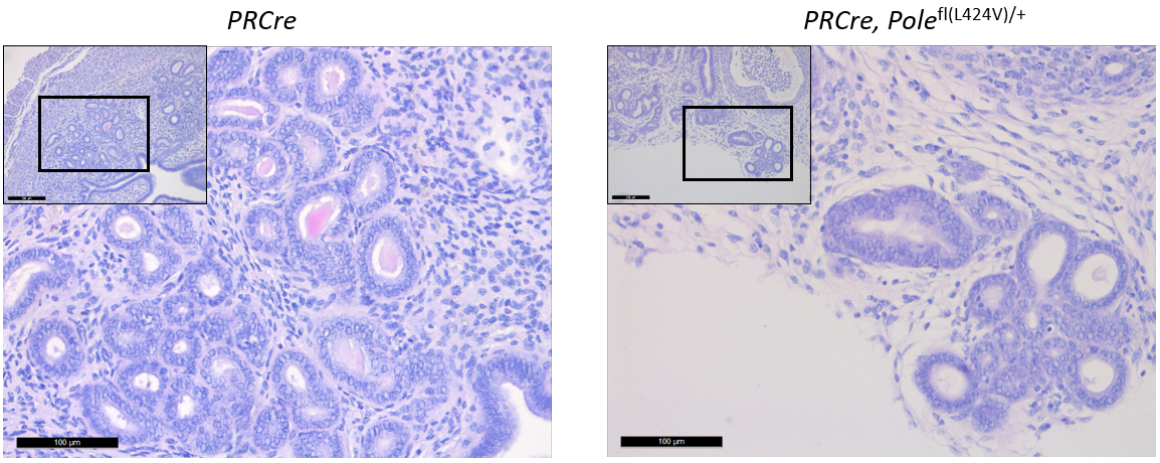



**Figure 4.15:** Uterine hyperplasia from one experimental *PRCre*,*Pole*$^{fl(L424V)/+}$ and one control *PRCre* mouse. Glands within the uterus are crowded with nuclear atypia. Scale bars represent 200$\mu$m in the small image and 100$\mu$m in the large image

phenotype between experimental *PRCre*,*Pole*$^{fl(L424V)/+}$ and control *PRCre* mice.

### 4.4.6 Investigation of *Pole*$^{L424V/+}$ in an *Apc*-mutant background

Loss of function mutations in the tumour suppressor gene *APC* drive colorectal carcinogenesis and are common in both somatic and germline *POLE* exonuclease domain mutant colorectal cancer (123). The ENU mutagen-generated *Apc*$^{Min}$ allele (MGI:1856318) contains a nonsense mutation at codon 850 within the $\beta$-catenin binding domain of *Apc* (Figure 4.2). *Apc*$^{Min}$ mice develop multiple intestinal polyps, mammary gland tumours and occasionally uterine tumours with complete or partial loss of heterozygosity evident in polyps and tumours (281).

As the lack of intestinal phenotype in heterozygous *Pole* L424V mutant mice suggested the effects of this mutation may be too subtle to initiate cause gross tumourigenesis during the timescales of my previous experiments, I crossed germline *Pole*$^{L424V/+}$ mice with *Apc*$^{Min}$ mice to determine if a phenotype associated with the *Pole*$^{L424V/+}$ allele was unmasked by an accompanying *Apc*$^{Min}$ mutation.

#### 4.4.6.1 Generation of compound mutant *Apc*$^{\mathbf{Min}}$,*Pole*$^{\mathbf{L424V/+}}$ mice

Male *Apc*$^{\mathrm{Min}}$ mice were crossed with female *Pole*$^{\mathrm{L424V/+}}$ mice to generate compound mutant *Apc*$^{\mathrm{Min}}$,*Pole*$^{\mathrm{L424V/+}}$ offspring. All *Apc*$^{\mathrm{Min}}$ mice used in the study were heterozygous, as homozygotes die in utero (233). Offspring were generated at expected Mendelian ratios.

Experimental *Apc*$^{\mathrm{Min}}$,*Pole*$^{\mathrm{L424V/+}}$ mice were aged and sacrificed on signs of a phenotype such as pale feet (anaemia), cachexia or a visible external tumour. Formalin-fixed intestine, spleen, uterus, lung and if present, visible tumours and fresh-frozen small intestinal polyps (if present), ear and uterus were collected from each mouse. To ensure matching time-points for quantification of polyp number, sex-matched (if possible) *Apc*$^{\mathrm{Min}}$ littermate controls were culled at the same time as each experimental mouse. The median survival of 16 experimental *Apc*$^{\mathrm{Min}}$,*Pole*$^{\mathrm{L424V/+}}$ mice and 12 control *Apc*$^{\mathrm{Min}}$ mice generated from 13 breeding pairs was 132 (104-187) days and 142 (88-187) days respectively.

#### 4.4.6.2 Macroscopic analysis of L424V/+ *Apc*$^{\mathbf{Min}}$ compound mice

As anticipated, the predominant phenotype observed in all *Apc*$^{\mathrm{Min}}$ crosses was intestinal polyposis in the small intestine and to a lesser extent, the colon. Intestines were processed as described in Methods Section 2.7.6.1 and stained with methylene blue. Stained intestines were visualised under a dissection microscope and polyp burden and diameter were measured. Polyps were categorised by size as <1mm, 1-2mm, 2-3mm and >3mm maximum diameter.

11/13 of experimental and 2/4 control mice with weight loss and signs of distress during the study were found to have large colonic polyps, in some cases blocking the intestinal lumen (Figure 4.16c). Formal comparison of colonic polyp burden revealed a statistically significant increase in *Apc*$^{\mathrm{Min}}$,*Pole*$^{\mathrm{L424V/+}}$ compound mutant mice compared to *Apc*$^{\mathrm{Min}}$, adjusting for age and sex ($P = 0.045$, One-way ANOVA), (Figure 4.16a). Interestingly, *Apc*$^{\mathrm{Min}}$,*Pole*$^{\mathrm{L424V/+}}$ compound mutant mice also demonstrated a tendency towards larger colonic polyps than *Apc*$^{\mathrm{Min}}$, demonstrated in Figure 4.16b. Although this result was not statistically significant ($P = 0.244$, One-way ANOVA), it raises the possibility of more rapid polyp growth in this group.

Unfortunately, the small number of colonic polyps in $Apc^{Min}$ mice meant that I was underpowered to confirm this and more animals are required to validate these findings.

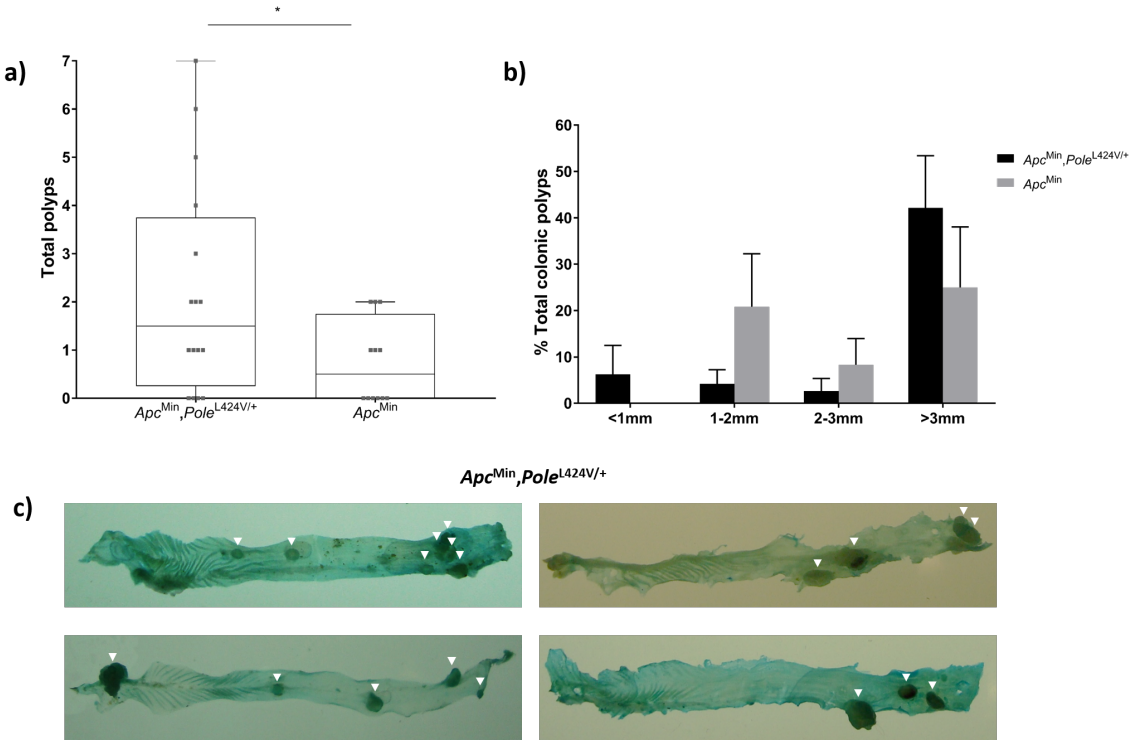



**Figure 4.16:** The heterozygous L424V/+ mutation significantly increases colonic polyp burden in $Apc^{Min}$ mice. a) Box and whisker plot of colonic polyp burden per mouse. The box shows the median value, and the 25th and 75th percentile. The whiskers cover the minimum to the maximum values. b) Bar chart showing the proportion of polyps in four size categories from the colon (Mean + SEM). c) Representative images of methylene blue stained colon from $Apc^{Min}$,$Pole^{L424V/+}$ compound mice. Colonic polyps are highlighted with white arrows.

In contrast to the colonic analyses, examination of the small intestine revealed no statistically significant difference in polyp burden between $Apc^{Min}$,$Pole^{L424V/+}$ compound mutant and $Apc^{Min}$ mice, adjusting for age and sex ($P = 0.81$, One-way ANOVA) (Figure 4.17a). Similarly, the $Pole^{L424V/+}$ allele did not appear to alter the location or size of small intestinal polyps (Figure 4.17c and 4.17d).

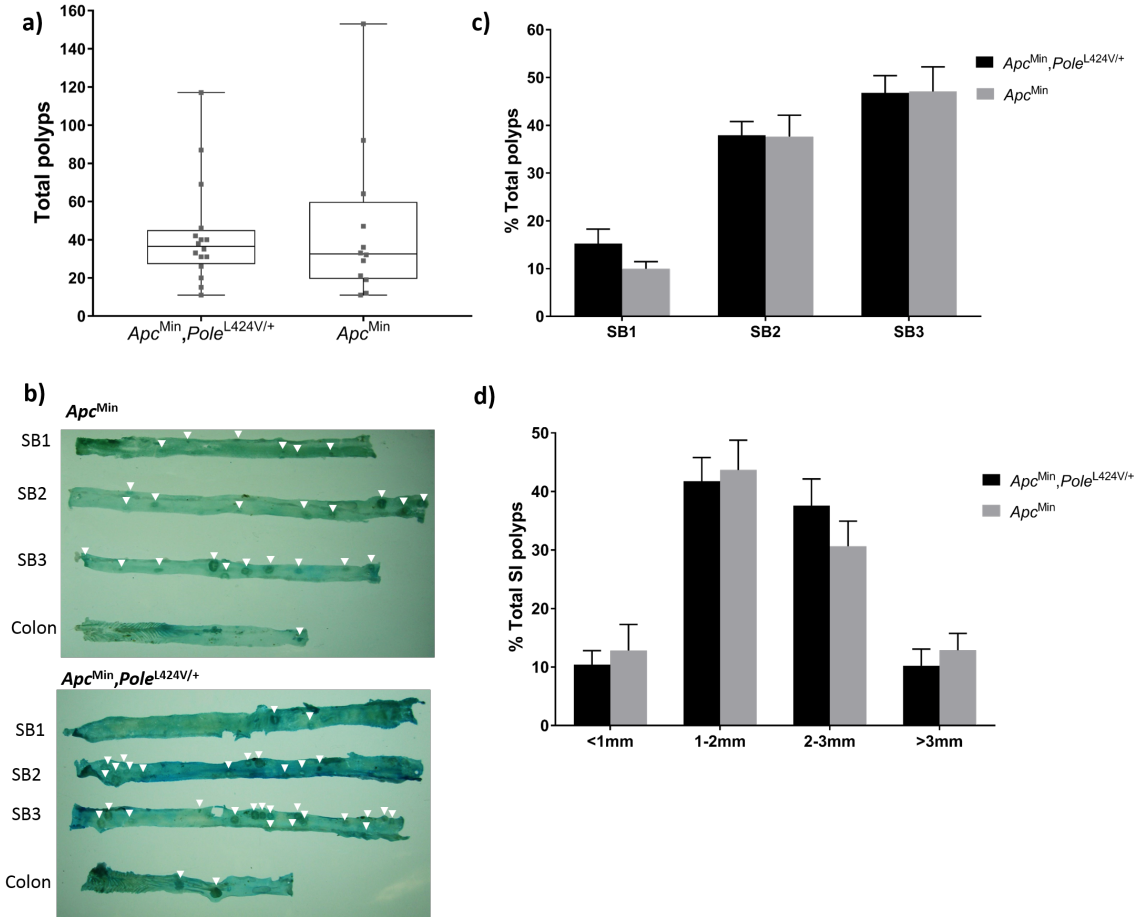



**Figure 4.17:** The heterozygous L424V/+ mutation does not significantly increase small intestinal polyp burden or size in $Apc^{Min}$ mice. a) Box and whisker plot of polyp burden in the small intestine. The box shows the median value, and the 25th and 75th percentile. The whiskers cover the minimum to the maximum values. b) Representative images of methylene blue stained small intestine and colon from experimental $Apc^{Min}$,$Pole^{L424V/+}$ and control $Apc^{Min}$ mice. Polyps are highlighted with white arrows. c) Bar chart showing the proportion of total polyps found in SB1, SB2 and SB3 (Mean + SEM). d) Bar chart showing the proportion of polyps in four size categories from the small intestine (Mean + SEM). Raw polyp counts are listed in Appendix Table I.1.

#### 4.4.6.3 Modifiers of Min

The wide variation in polyp burden within each genotype (11-153 and 11-117 for $Apc^{Min}$,$Pole^{L424V/+}$ and $Apc^{Min}$ respectively), raised the possibility that additional genetic or environmental factors might be influencing tumour multiplicity, thus complicating the analyses.

One such genetic factor in the context of $Apc^{Min}$ mice is Modifier of Min-1 (Mom1). Mom1 is a genetic locus mapping to a 15 centimorgan region on distal chromosome 4. Specific strains have been characterised as either $Mom1^{s}$ (mutant) or $Mom1^{r}$ (wild-type) (282), the former being associated with a higher number of $Apc^{Min}$-associated intestinal adenomas. A convincing candidate gene within this region is Phospholipase A2, membrane associated (*Pla2g2a*), expressed in the paneth cells of the intestinal crypts (228, 283). A homozygous 1bp frameshift mutation in exon 4 of *Pla2g2a*, resulting in loss of function, occurs in some inbred mouse strains (283) and has been shown to perfectly correlate with Mom1 status (228).

Although pure C57BL/6J mice are $Pla2g2a^{ex4fs/ex4fs}$ (228), founder $Pole^{fl(L424V)/+}$ mice were derived from a C57BL/6N colony, and for reasons of time, had not been backcrossed onto C57BL/6J. To determine if variation in *Pla2g2a* mutation status might explain the variation in polyp frequency across $Apc^{Min}$ experimental and control mice, I genotyped *Pla2g2a* using restriction fragment length polymorphism (RFLP) analysis and Sanger sequencing (Methods Section 2.2.6). This confirmed that all mice were homozygous $Pla2g2a^{ex4fs/ex4fs}$ (Figure 4.18), thus excluding Mom1/*Pla2g2a* polymorphism as a cause of the variability.

## 4.5 Discussion

The principle result of this chapter was that mouse models of the *Pole* L424V exonuclease domain mutations do not recapitulate the highly penetrant intestinal polyposis and endometrial cancer phenotype observed in human carriers of the germline $POLE^{L424V}$ mutation.

Both homozygous unrecombined ($Pole^{fl(L424V)}$) and germline recombined ($Pole^{L424V}$) alleles were embryonic lethal, possibly due to *Pole* gene dysregulation. Constitutive

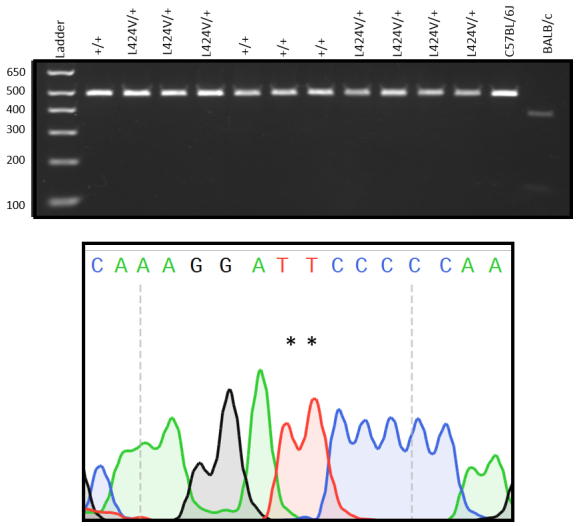



**Figure 4.18:** *Pla2g2a* genotyping of *Apc*$^{Min}$,*Pole*$^{L424V/+}$ and *Apc*$^{Min}$ using restriction fragment length polymorphism analysis (upper gel) and Sanger sequencing (lower chromatogram). +/+ denotes *Apc*$^{Min}$ mice and L424V/+ denotes *Apc*$^{Min}$,*Pole*$^{L424V/+}$ mice. C57BL/6J is a positive control for a homozygous 1bp *Pla2g2a* insertion and BALB/c is a positive control for wildtype *Pla2g2a*. All samples tested were homozygous *Pla2g2a*$^{ex4fs/ex4fs}$.

heterozygous *Pole*$^{L424V/+}$ mice are viable and express the mutant allele. Phenotypic analysis of constitutive knock-in, uterine-specific and inducible intestinal heterozygous *Pole* L424V models revealed no gross morphological or microscopic abnormalities, however, increased protein expression of Cleaved Caspase-3, an apoptotic marker, was observed in the small intestine of constitutive knock-in models.

When crossed with *Apc*$^{Min}$, *Pole*$^{L424V/+}$ drove a statistically significant increase in colonic polyp burden compared to *Apc*$^{Min}$ controls, independent of the genetic Modifyer of Min, Mom1. Compound mutant mice had a tendency towards larger polyps, suggesting more rapid polyp growth in *Pole*$^{L424V/+}$ mutants. Small intestinal polyp burden and size was unaffected.

### 4.5.1 *Pole* fl(L424V) is homozygous lethal in mice

The *Pole* fl(L424V) construct used in this study relies on two independent LoxP recombination events to delete wildtype exon 13 and express L424V mutant exon 13. The construct also contains an mCherry cassette under the *PGK* promoter upon recombination. The demonstration that unrecombined *Pole* fl(L424V) is embryonic

lethal in homozygosity, along with increased *Pole* mRNA expression in the intestine and kidney raises the possibility that the allele may act as a null or a hypermorph, resulting in loss of function or toxicity respectively, potentially as a consequence of defective gene splicing, intrinsic PGK promoter activity or increased $\beta$ hairpin loop formation around the GC-rich region of the L424 codon (described in Section 1.5). Determination of this was not performed due to consideration of time and effort.

Interestingly, I found that the germline recombined *Pole* L424V allele was also embryonic lethal in homozygocity. I decided not to explore the timing of lethality in pregnant females as this was not directly relevant to my investigation of the contribution of *Pole* L424V to cancer. However, the perturbation of *Pole* mRNA underscored the importance of inclusion of *Pole*$^{fl(L424V)/+}$ mutants as controls in experiments.

### 4.5.2 *Pole* L424V/+ alone does not induce tumourigenesis in mice

I generated three crosses of *Pole*$^{fl(L424V)/+}$ mice to study the effect of this mutation on tumourigenesis: (i) a cross with *PgkCre* that is ubiquitously expressed from early embryogenesis; (ii) an intestinal-specific cross with the tamoxifen-inducible *Lgr5CreERT2* and; (iii) a uterine-specific cross with *PRCre*. In all three crosses, there was no evidence of tumourigenesis or pre-cancerous lesions associated with the *Pole*$^{fl(L424V)/+}$ allele up to 80-90 weeks of age. This contrasts with patients carrying germline *POLE*$^{L424V}$ mutations who develop tumours at a mean age of 39 years (range 14-60 years), corresponding to a biological age of approximately 40-56 weeks in mice (284) (Table 1.4).

Independent of age or specific Cre recombinase, 4/30 experimental Cre,*Pole*$^{fl(L424V)/+}$ and 2/31 control mice developed lymphomas, although this was not associated with the *Pole* L424V mutation ($P = 0.42$, Fisher's exact test). As the C57BL/6J inbred strain is known to be prone to lymphoma and leukaemia development, it is plausible that this just represented a strain-dependent phenomenon (285). An activation of Cleaved Caspase-3 expression in the small intestine of *PgkCre*,*Pole*$^{fl(L424V)/+}$ mice may reflect an increase in cellular mutation rate resulting in damaging mutations in essential genes, however, this requires further investigation.

There are several possible explanations as to why these mice have no gross morphological or microscopic phenotype. Firstly, the embryonic lethality of mice homozygous for the unrecombined and recombined allele raises the possibility that the targeted allele may not be physiologically expressed, or expressed to a lower degree than physiological levels. Although analysis of cDNA confirmed successful recombination of the *Pole* L424V allele, expression of the mutant appeared to be lower than the wildtype, consistent with the latter hypothesis. Determination of whether this reflects mosaic recombination in *PgkCre*,*Pole*$^{fl(L424V)/+}$ mice was not possible in the timeframe of this project, but would be of interest to define. Similarly, *Lgr5CreERT2* is not expressed in all intestinal stem cells and may have delayed to onset of a visible phenotype.

Finally, although the L424V mutation leads to tumourigenesis in humans, this may not be the case in mice. For example, biallelic loss of the DNA glycosylase *MUTYH* and *NTHL1* in the human germline (MUTYH-associated polyposis and NTHL1-associated polyposis) predisposes carriers to colorectal adenomas and cancer. Homozygous *Myh* and *Nthl1* knockout mice, however, do not develop tumours despite an increased mutation rate in embryonic stem cells (286, 287, 288). Indeed, heterozygous *Pole* exonuclease-deficient mice also appear phenotypically normal, even though they display a mutator phenotype compared to control animals and homozygosity is required to reveal a tumour phenotype (139). It is plausible therefore that L424V mutation, which has a mutation rate similar to that of exonuclease active site mutations in budding yeast (197), mice may only develop tumours with homozygous L424V mutations.

### 4.5.3 *Pole* L424V increases colonic polyposis in *Apc*$^{Min}$ mice

Interestingly, compound mutant *Apc*$^{Min}$,*Pole*$^{L424V/+}$ mice display a propensity to develop multiple large adenomas (>3mm) in the colon resulting in bowel obstruction, prolapse and cachexia in several cases. Colonic adenomas are uncommon in *Apc*$^{Min}$ mice, which typically develop 20 to >100 intestinal polyps, the vast majority of which are proximal small intestine (233). In contrast, the addition of a germline *Pole* L424V/+ mutation did not affect small intestinal polyp burden or morphology in compound mutant mice.

The initiation of intestinal polyposis in $Apc^{Min}$ mutant mice is typically by loss of heterozygosity or a pathogenic mutation in the wildtype allele leading to partial or total loss of function. Using $Apc^{Min}$ mice it has been shown that although the number of colonic microscopic aberrant crypt foci, precursors to colonic polyps, exceeds the number small intestinal polyps, most do not go on to develop into macroscopic polyps (289). This suggests that there is a rate-limiting step in the colonic crypts that inhibits tumour progression (290).

Several *Apc* compound mutant mice such as $Apc^{Min}$ $BubR1^{+/-}$ and $Apc^{\Delta 716}$ $Cdx2^{+/-}$ show a clear propensity to develop colonic rather than small intestinal polyps (291, 292). Both *BubR1* and *Cdx2* genes regulate chromosomal segregation during mitosis; the chromosomal instability caused by their loss in these mice is thought to accelerate *Apc* loss of heterozygosity, resulting in progression of colon aberrant crypt foci to adenomas. An increase in the mutation rate, due to the *Pole* L424V mutation in compound mutant $Apc^{Min}Pole^{L424V/+}$ mice, is predicted to increase the probability of pathogenic point mutations in wildtype *Apc*. This in turn would drive colonic aberrant crypt foci progression to polyps, although perhaps to a more moderate degree than in $Apc^{Min}$ $BubR1^{+/-}$ compound mutant mice.

# Chapter 5

# When do somatic *POLE* exonuclease domain mutations occur in colorectal cancer?

## 5.1 Introduction

### 5.1.1 Timing of *POLE* mutations in cancer

Determination of the timing of somatic mutations in cancer is important for the analysis of tumour evolution, clonal selection and may also carry therapeutic implications (293, 294). The association of germline *POLE* mutations with colorectal cancer, with no evidence of a second hit (144), suggests that heterozygous exonuclease domain mutations are true drivers of tumourigenesis.

Studies of relative variant allele frequency of somatic mutations in cancer has suggested that *POLE* mutations are present in the earliest persisting clone in sporadic glioblastomas (164, 181), while pan-cancer studies have shown that pathogenic cancer-associated *POLE* exonuclease domain mutations are typically clonal, with sub-clonal

*POLE* EDMs predicted to be non-functional passengers (171, 186, 210). Acquisition of a pathogenic *POLE* exonuclease domain mutation precedes an explosive increase in mutations (171, 186), with the majority of subclonal mutations exhibiting a strong *POLE*-associated mutation signature (Signature 10, Alexandrov *et al.* (1)) (156, 158, 162, 181, 196).

More importantly, cancer driver genes in colorectal and endometrial cancers often display evidence of the *POLE* mutational bias, resulting in non-hotspot amino acid changes (Section 1.5.3) that are not observed in *POLE* itself (10, 186). This pattern suggests that somatic *POLE* mutation is an early or initiating event in sporadic cancers (162).

Mathematical models have predicted that a mutator phenotype gives cells a proliferative advantage due to the rapid acquisition of secondary driver mutations and the ability to adapt in a changing tumour microenvironment (295, 296). Analysis of glioblastomas from patients with constitutional biallelic mismatch repair deficiency revealed that the acquisition of pathogenic *POLE* EDMs was associated with rapid tumour growth (164). Retrospective analyses showed that patients with *POLE* exonuclease domain mutations were diagnosed at an earlier age than patients without *POLE* mutations (166, 167), consistent with accelerated growth of *POLE*-mutant tumours. Finally, an extreme case study of a patient diagnosed with multiple intestinal adenomas and carcinomas was shown to harbour a de-novo $POLE^{V411L}$ germline mutation with no additional congenital defects (155). All of these data suggest that *POLE* exonuclease domain mutations may accelerate the acquisition of additional events for carcinogenesis, at least in early disease stages.

Although studies using *POLE* exonuclease domain mutation clonality and signature 10 mutation preponderance are used to support the notion that somatic *POLE* mutations are an early event in cancer, direct evidence that *POLE* mutations occur in pre-cancers has been lacking. Recent published work, however, part of which is reported below, have shown that pathogenic *POLE* mutations occur in the precursors of both endometrial and colorectal cancers (endometrial epithelial neoplasia and colorectal adenomas respectively) (186).

### 5.1.2 Immune response, prognosis and therapies

As mentioned in Chapter one, tumours with *POLE* exonuclease domain mutations have a prominent T-cell infiltrate (166, 175, 178, 179, 181, 187, 205) and activation of immune checkpoints (166, 205). This is thought to relate to the increased number of neoantigens these tumours are predicted to harbour, compared to tumours with a lower mutation burden. The high mutation rate in *POLE*-mutant cancers may also lead to positive selection of cells with up-regulation of immunosupressive checkpoints PD-1 and CTLA-4, which help cells evade immune destruction (186, 205).

If *POLE* exonuclease domain mutations are indeed the initial tumour driver, patients with *POLE* mutations may be good candidates for personalised management which accounts for the characteristics of these tumours, or for therapies which target any vulnerability the mutant polymerase causes, reducing unnecessary treatments and side effects.

The association of *POLE* exonuclease domain mutations with a good prognosis in endometrial cancer and potentially colorectal and brain cancer is one of the most promising characteristics that can be exploited for therapeutic benefit. Despite often being high grade, endometrial cancers with *POLE* EDMs have almost a 100% recurrence-free survival (157, 180, 207). These data suggest that these patients may benefit from receiving either no, or less aggressive, post-operative chemotherapy and radiotherapy compared to patients with high grade tumours without *POLE* mutations (179, 180, 188).

It has been suggested that the association between *POLE* exonuclease domain mutations and a good prognosis may be due to high tumour immunogenicity, which also makes cancer harbouring these mutations good candidates for immune checkpoint inhibition in the setting of recurrent disease (190, 205). It is, however, currently unclear whether such enhanced immunogenicicty is a feature of pre-cancerous colorectal adenomas harbouring cancer-associated *POLE* exonuclease domain mutations.

### 5.1.3 Aims of this chapter

- To determine if *POLE* exonuclease domain mutations occur earlier than the transition to a carcinoma by quantifying the frequency of cancer-associated

*POLE* EDMs in early intestinal adenomas.

- To quantify the density of immune infiltrate in adenomas with and without *POLE* exonuclease domain mutations.
- To compare the mutation frequency and spectrum of early adenomas with advanced colorectal tumours which have somatic *POLE* exonuclease domain mutations.

## 5.2 Methods

The frequency of *POLE* exonuclease domain mutations in colorectal cancer is 0.5-3%, compared to 7-12% in endometrial cancer (123). To ensure sufficient probability of detecting a *POLE* exonuclease domain mutant pre-cancer, 389 FFPE intestinal adenomas were collated from 261 patients recruited through the COloRectal Gene Identification (CORGI) study (Tomlinson Group, University of Birmingham, formerly University of Oxford). All patients recruited to CORGI had been been screened for germline mutations in known colorectal cancer predisposition genes associated with Familial Adenomatous Polyposis (*APC*), *MUTYH*-associated polyposis and Lynch syndrome (MMR genes).

Areas of dysplasia were marked on 4$\mu$m H&E stained sections and DNA extraction performed from adjacent unstained tissue (Methods Section 2.2.1.1). DNA was screened for the somatic cancer-associated *POLE* exonuclease domain mutations P286R, V411L and S459F (which account for approximately 80% of somatic cancer-associated *POLE* variants in CRC (123, 162, 166)) by Sanger sequencing of exons 9, 13 and 14 (Primers listed in Appendix Table A.5) and/or by TaqMan genotyping.

## 5.3 Results

Some of the results presented in this chapter have been published in Temko *et al.* (186).

### 5.3.1 Selection of Adenomas to screen for *POLE* proofreading domain mutations

A total of seven somatic *POLE* exonuclease domain mutations were identified in the cohort of 389 intestinal adenomas, 5/6 of which were variants of unknown significance (VUS). Two VUS, A456V and A456T, were identified at codon 456, a *POLE* mutation hotspot (Figure 5.1 and 1.4). The two A456 variants are not listed in the TCGA database, so were not taken forward for further analysis (14).

Two adenomas harboured a T457M variant, one in the heterozygous and the other in the homozygous state. $POLE^{\mathrm{T457M}}$ is reported in the TCGA database (14) in a uterine carcinosarcoma (176), adrenocortical carcinoma (14) and in two adenocarcinomas (194). The uterine carcinosarcoma was ultramutated with 5499 somatic coding variants. The adrenocortical carcinoma, however, harboured 519 mutations. The pathogenicity of the T457M variant is therefore uncertain, particularly as the mutation was homozygous in one patient, a state discordant with all confirmed pathogenic *POLE* exonuclease domain mutations detected to date, which are universally heterozygous.

One adenoma contained a cancer-associated *POLE* V411L mutation (Figure 5.2). The total number of confirmed pathogenic *POLE* exonuclease domain mutations was 0.26% adenomas (0.38% of patients), which is slightly lower than the previously reported frequency in established colorectal cancers of 0.5-3%. Similar to *POLE* exonuclease domain mutant colorectal carcinomas, the $POLE^{\mathrm{V411L}}$ mutant adenoma had unremarkable histology (Figure 5.2).

### 5.3.2 Immune infiltrate quantification

I hypothesised that the early acquisition of *POLE* exonuclease domain mutations would result in a rapid increase in mutation rate, resulting in an increased number of neoantigens and an enhanced immune response. Immunohistochemistry for the T-cell marker, $CD8^{+}$ was performed on the *POLE* exonuclease domain mutant ($POLE^{\mathrm{V411L}}$) adenoma and five *POLE* wildtype ($POLE^{\mathrm{WT}}$) adenomas. The $POLE^{\mathrm{V411L}}$ adenoma exhibited a dense $CD8^{+}$ cellular infiltrate compared with the five $POLE^{\mathrm{WT}}$ adenomas

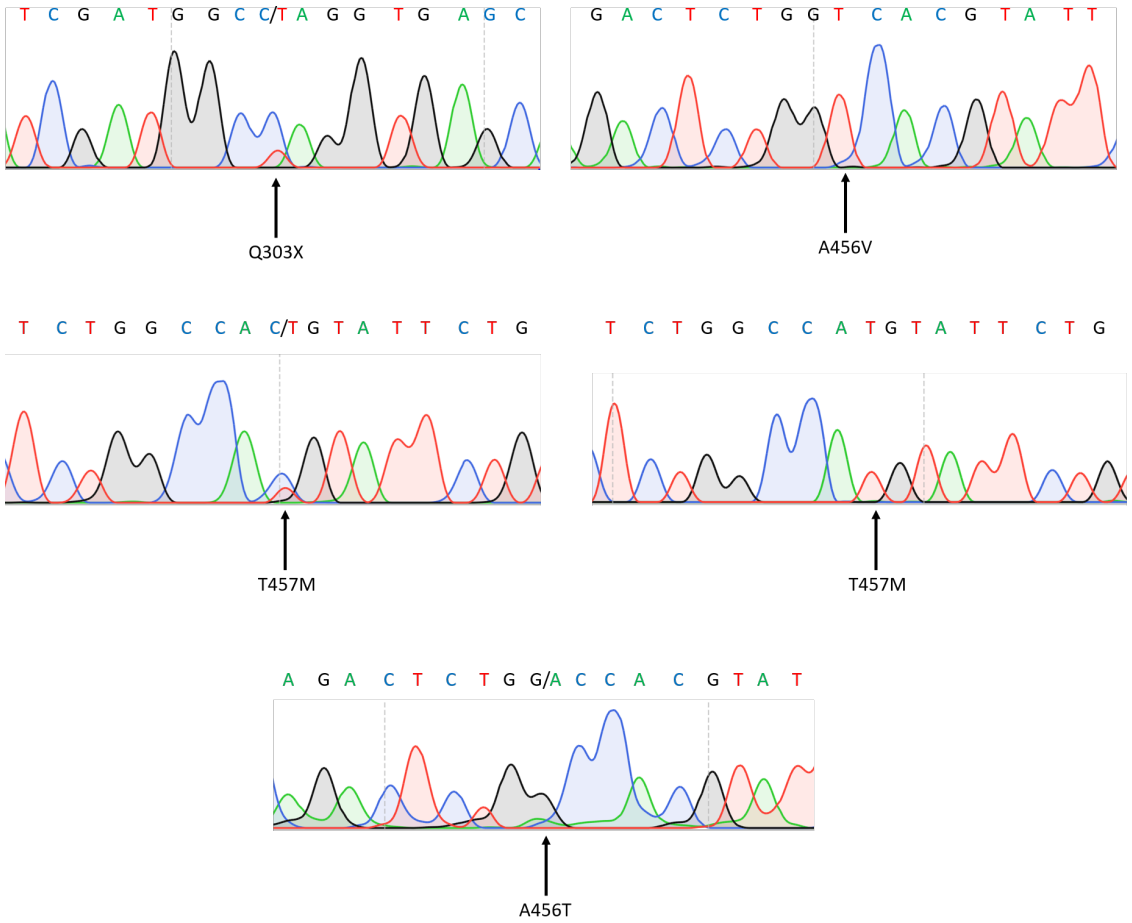



**Figure 5.1:** Chromatograms of somatic *POLE* exonuclease domain variants of unknown significance in intestinal adenomas.

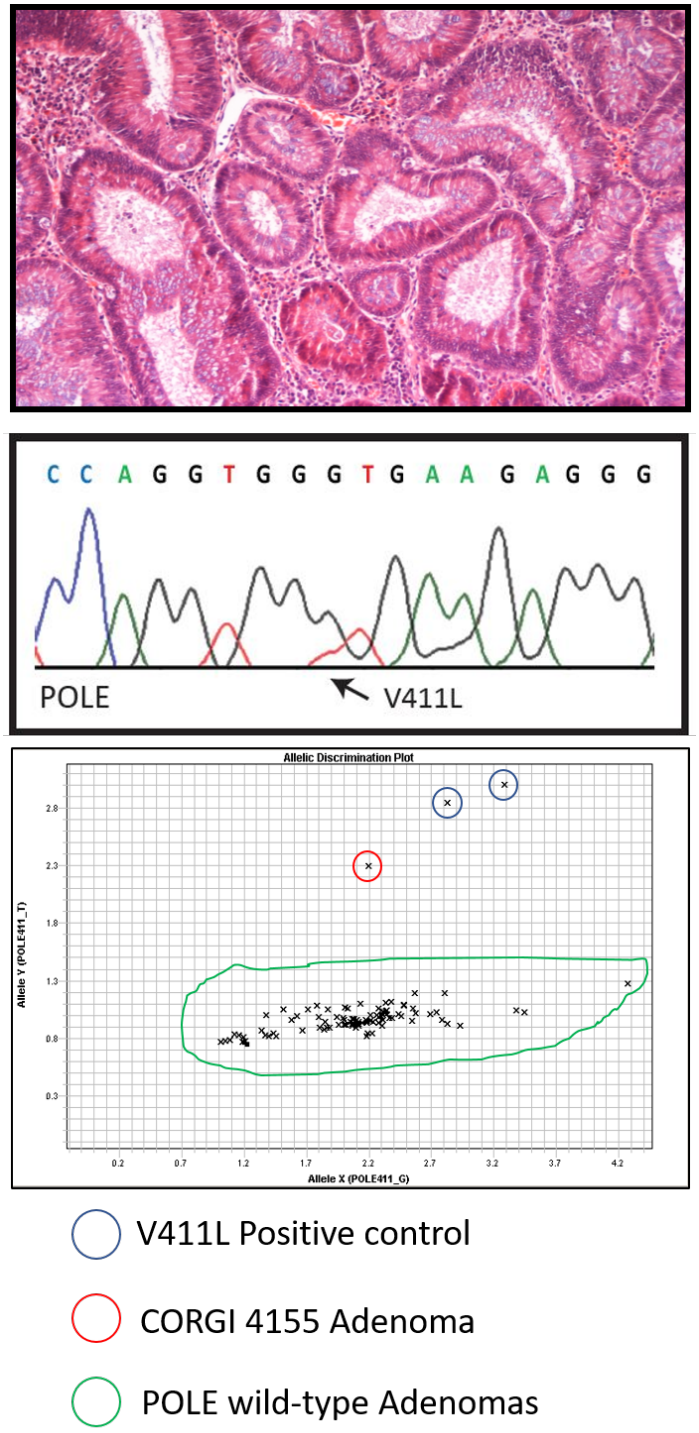



**Figure 5.2:** *POLE*$^{V411L}$ mutant adenoma H&E, chromatogram and TaqMan plot. The results show a heterozygous G→T missense variant resulting in a cancer-associated *POLE*$^{V411L}$ exonuclease domain mutation (adapted from Temko *et al.* (186)).

(154.9 versus median 34.0 CD8$^+$ cells per highpower field (HPF)), though statistical testing was not possible (Figure 5.3).

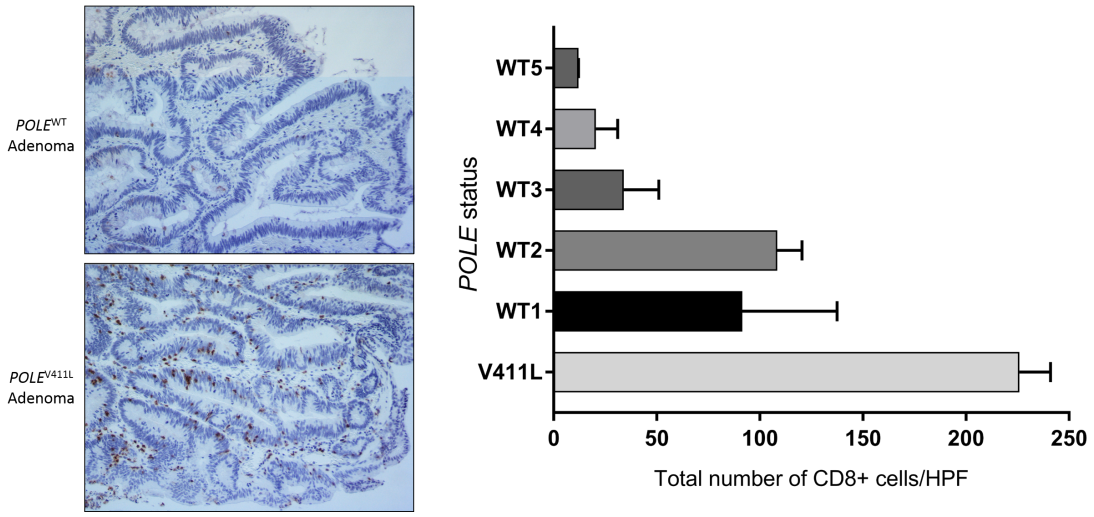



**Figure 5.3:** CD8$^+$ immunohistochemistry on colorectal adenoma sections a) Representative images of CD8$^+$ immunohistochemistry from *POLE*$^{V411L}$ and *POLE*$^{WT}$ adenoma sections b) Mean and standard deviation of CD8$^+$ density (mean of two HPFs) for *POLE*$^{V411L}$ and *POLE*$^{WT}$ adenomas.

## 5.4 Microsatellite instability testing

Microsatellite unstable colorectal and endometrial tumours are hypermutated and display increased immune infiltrate (297). To exclude the possibility of microsatellite instability contributing to the increased CD8$^+$ immune infiltrate in the *POLE*$^{V411L}$ adenoma, I tested mononucleotide and dinucleotide microsatellite markers in tumour (tDNA) and genomic DNA (gDNA) using PCR analysis (Methods Section 2.2.5). No sample exceeded 30% variation between tDNA and gDNA microsatellite peak frequency (Table 5.4), therefore all samples were confirmed microsatellite stable. These data exclude microsatellite instability as a contributing factor for the increased CD8$^+$ immune cell infiltrate in the *POLE*$^{V411L}$ mutant adenoma.

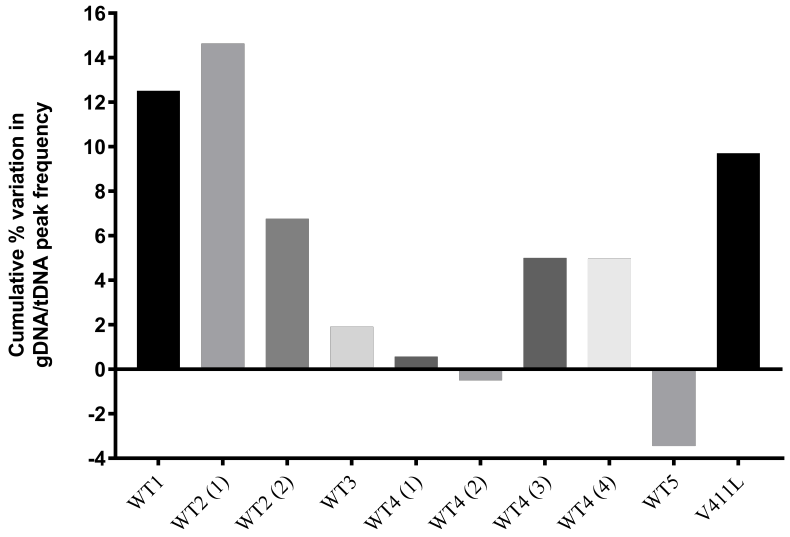



**Figure 5.4:** Cumulative % variation in mononucleotide and dinucleotide microsatellite marker peak frequency between gDNA and tDNA from in $POLE^{WT}$ and $POLE^{V411L}$ adenomas. (Samples with +/-30% variation are classed as microsatellite unstable.). Numbers in brackets are independent adenomas from the same patient.

### 5.4.1 Next generation sequencing of *POLE* EDM adenomas

I next wished to determine whether colorectal adenomas with *POLE* exonuclease domain mutations also display similar ultramutation and other molecular characteristics to *POLE*-mutant colorectal carcinomas. The $POLE^{V411L}$ adenoma and a $POLE^{WT}$ control were sequenced using a colorectal cancer molecular inversion probe (MIP) capture panel consisting of 30 commonly mutated genes and recurrent fusion events (Appendix Table J.1) in collaboration with Dr Claire Palles, University of Birmingham who made the libraries and sequenced the DNA. MIP capture panel sequencing comprises short reads and requires the variant to be on both the forward and reverse DNA strands, making the technique suitable for low concentration, poor quality FFPE DNA, thus reducing the probability of miscalling cytosine deamination as thymine or other DNA lesions that occur during tissue processing.

Adenoma and associated germline DNA samples were sequenced on the Illumina NextSeq platform and the data processed by the Palles laboratory. Variants were annotated using VEP (Ensembl) (239) and were excluded if: a germline variant; within a homopolymer run >4 in length; had significant strand bias; a confirmed single nucleotide polymorphism; had a read depth <20 and; had a variant allele

frequency <5%, thus excluding PCR and sequencing errors.

Post-filtering, 57 somatic variants were called for the $POLE^{V411L}$ exonuclease domain mutant adenoma, which was nearly double that of the $POLE^{WT}$ adenoma at 29 somatic variants. The variant allele frequency of $POLE^{V411L}$ was similar to predicted pathogenic variants in major driver genes such as *APC*, *TP53* and *PTEN*, suggesting that $POLE^{V411L}$ also occured early in tumourigenesis (Figure 5.5). The majority of somatic variants (89%) had a VAF less than $POLE^{V411L}$, again consistent with the hypothesis that *POLE* exonuclease domain mutations are an early event in colorectal cancer development (Figure 5.5).

*POLE* exonuclease domain mutant tumours have a distinct mutation signature (1). I compared the 6-channel mutation signature between the $POLE^{V411L}$ and $POLE^{WT}$ adenomas, as the number of genes sequenced was not sufficient to extract the 96 channel mutation signature reported by Alexandrov *et al.* (1).

Interestingly, the $POLE^{V411L}$ adenoma had a high frequency of T→G mutations (a component of the *POLE* COSMIC signature) compared to the $POLE^{WT}$ adenoma (Figure 5.6). Unexpectedly, the hotspot mutations C→A and C→T were not increased compared to the $POLE^{WT}$ adenoma.

In the $POLE^{V411L}$ adenoma, amino acid substitutions R→X, E→X and R→Q, known to be enriched in *POLE* exonuclease domain mutant cancers (123), were present in major colorectal cancer driver genes, *APC*, *TP53* and *TCF7L2* (4/57). These amino acid changes were not observed in the $POLE^{WT}$ adenoma (0/29), however the difference was not significant ($P = 0.295$, Fisher's exact test).

Although the breadth of the DNA sequenced by the MIP panel was sufficient to detect an increased number of mutations in the $POLE^{V411L}$ adenoma, it was inadequate to derive a *POLE*-associated mutation signature. Whole exome or whole genome sequencing in addition to more *POLE* exonuclease domain mutant colorectal adenomas will be required to provide sufficient power to detect ultramutation and mutation signatures with confidence.

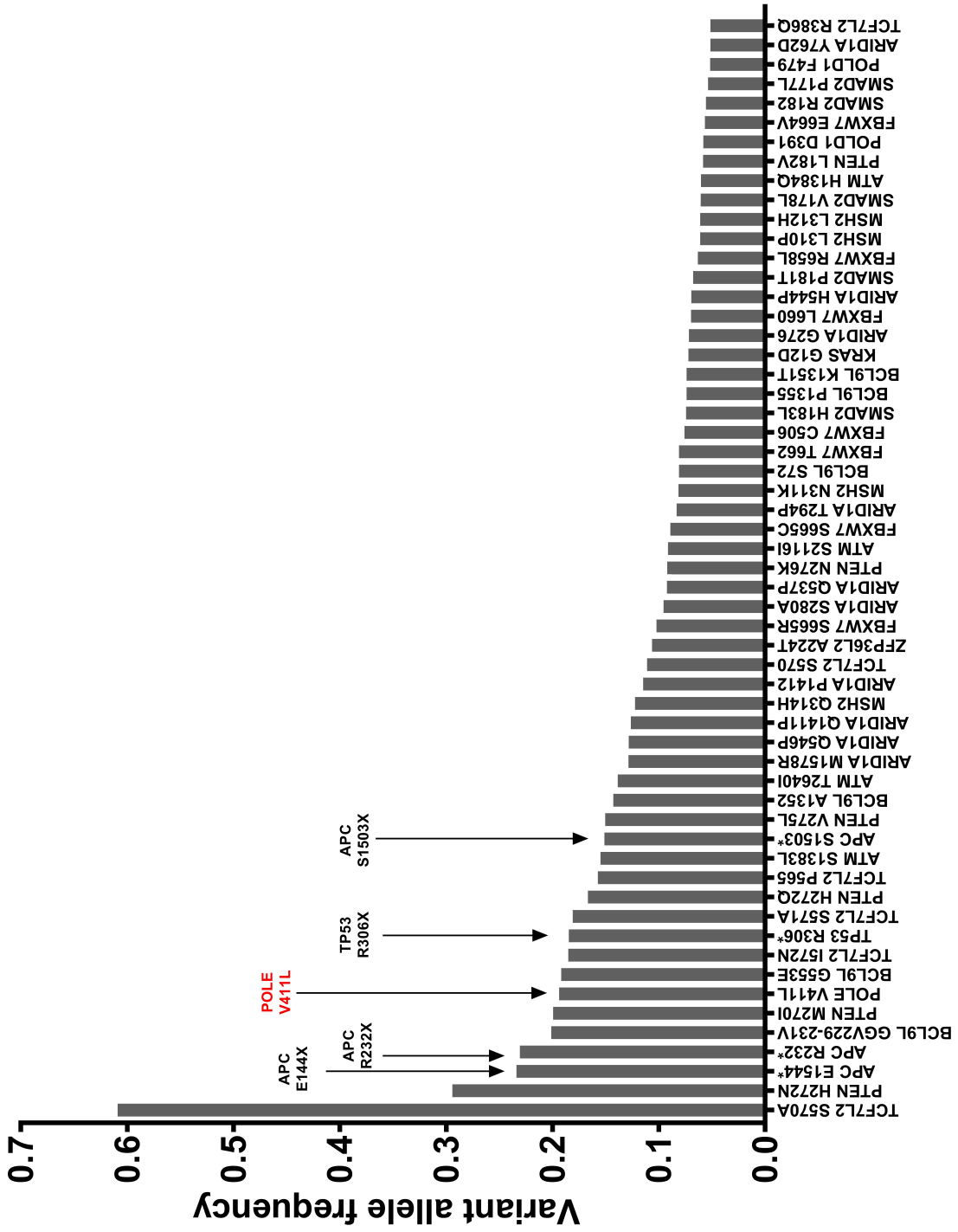



**Figure 5.5:** Variant allele frequency (VAF) of genic mutations in the $POLE^{V411L}$ exonuclease domain mutant colorectal adenoma. The $POLE^{V411L}$ variant has a VAF of 0.194. Colorectal cancer driver genes with somatic nonsense mutations are highlighted.

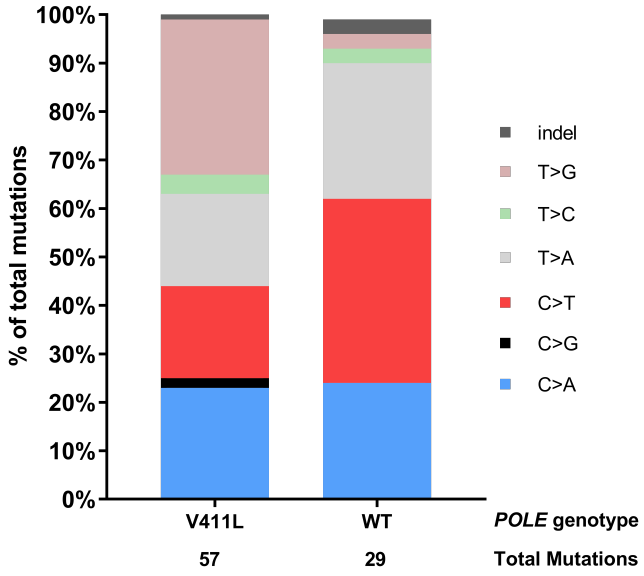



**Figure 5.6:** Mutation spectrum of $POLE^{V411L}$ exonuclease domain mutant and $POLE^{WT}$ colorectal adenomas.

## 5.5 Discussion

### 5.5.1 *POLE* exonuclease domain mutations are an early event in colorectal cancer development

Previous studies have shown that somatic *POLE* exonuclease domain mutations are frequently clonal in tumour DNA (164, 171, 181, 210), suggesting that they are an early event in tumour developent and may be the initial driver of tumourigenesis. Until recently, there has been no direct evidence that *POLE* exonuclease domain mutations occur in pre-cancerous lesions. Using Sanger sequencing and TaqMan genotyping, I identified the cancer-associated $POLE^{V411L}$ exonuclease domain mutation in a non-invasive colorectal adenoma, as part of a collaboration in which colleagues also identified cancer-associated *POLE* exonuclease domain mutations in spatially distinct endometrial cancer precursor lesions. These results demonstrate that *POLE* exonuclease domain mutations are an early event in colorectal and endometrial cancer (186).

The frequency of cancer-associated *POLE* exonuclease domain mutations from

adenomas in the CORGI patient cohort was 0.38%, compared to a frequency of 0.5-3% in CRC. Possible reasons for the lower *POLE* mutation frequency in screened adenomas include: i) targeted sequencing of specific exons and nucleotide variants rather than screening of the whole exonuclease domain; ii) experimental cohort bias from CORGI patients who are recruited due to a family history or early onset CRC and; iii) *POLE* exonuclease domain mutant adenomas may transition from adenoma to carcinoma faster, making the probability of pre-cancer detection less likely.

To improve variant detection, I would use targeted deep sequencing of the entire *POLE* exonuclease domain. The probability of detecting variants in tumours with a low cancer cell fraction as a consequence of dense immune infiltrate would then be maximised.

The early acquisition of cancer-associated somatic *POLE* exonuclease domain mutations and the rapid acquisition of mutations that they are likely to cause, may explain the association of *POLE* mutations with a young age of onset (<50 years). If *POLE* exonuclease domain mutations initiate tumourigenesis, this may explain why they are generally observed in organs with a rapid tissue turnover, such as the endometrium, intestine and the brain. Malignant precursor lesions from other tumours will need to be analysed to determine if this is the case.

### 5.5.2 *POLE* exonuclease domain mutant adenomas have an increased T-cell immune infiltrate

Colorectal and endometrial tumours with cancer-associated *POLE* exonuclease domain mutations both have an increased lymphocytic infiltrate, analogous to that of microsatellite unstable tumours (166, 205). Using $CD8^+$ immunohistochemistry as a marker of immune infiltrate, I found that the $POLE^{V411L}$ mutant adenoma had an increased T-cell infiltrate compared to $POLE^{WT}$ adenomas.

The results are concordant with the increased lymphocytic infiltrate observed in *POLE* exonuclease domain mutant colorectal cancer, suggesting that early adenomas have already acquired a sufficient number of somatic mutations to increase their neoantigen burden. Only one adenoma sample, however, has been analysed and I could not make a statistical comparison. To achieve this, additional samples and

more detailed histological and transcriptomic analysis of T-cell infiltrate and immune checkpoint expression will be required. However, it is interesting to speculate that $CD8^+$ may prove a useful biomarker for *POLE* exonuclease domain mutations in precursor lesions of the bowel and endometrium in the future.

### 5.5.3 The *POLE*$^{V411L}$ adenoma shows an amino acid substitution bias

The *POLE*$^{V411L}$ mutant adenoma had an enrichment of R→X, E→X and R→Q mutations in cancer driver genes: *APC* E1544X, R232X; *TP53* R306X; and *TCF7L2* R386Q. These amino acid changes have shown to be enriched in *POLE* exonuclease domain mutant cancers as a result of the *POLE* mutation Signature 10. Although this suggests that this adenoma bears the signs of a *POLE* mutation signature, the MIP panel focusses on CRC driver genes. The number of mutations called, therefore, is not sufficient to extract a full mutation signature. Whole exome or whole genome sequencing will be required to extract a comprehensive mutation signature and to permit estimation of the timing of mutations based on variant allele frequency.

Results from my study warrant further investigation with a larger sample size. They also suggest that the traditional adenoma-carcinoma sequence may not apply in *POLE* exonuclease domain mutant cancers. For example, *TP53* mutations are usually a late event in CRC development, but these data suggest that it may be early event in *POLE* exonuclease domain mutant cancer.

# Chapter 6

# Discussion and future work

## 6.1 Discussion and future work

As described in Chapter one, somatic and germline *POLE* exonuclease domain mutations define a subset of colorectal and endometrial cancers with distinct clinicopathological and molecular characteristics (123). The most striking of these comprise their exceptional mutational burden, termed "ultramutation" and a distinct mutation signature (1). These data, combined with correlative genomic data from advanced colorectal cancer (166), endometrial cancer (162) and brain cancer (181), have led to speculation that somatic *POLE* exonuclease domain mutations may be an early or even an initiating mutation in human cancer.

Until now, however, confirmation of the causative role of cancer-associated *POLE* exonuclease domain mutations in ultramutation in human cells and evidence of their early occurrence was lacking. Furthermore, whilst phenotypic characterisation of cancer-associated *POLE* mutations in mouse models was recently published by Li *et al.* (240), functional analysis of the clinically relevant tissues of the intestine and uterus was not reported.

In the work presented in this thesis, I have sought to address these knowledge gaps by: the generation and functional characterisation of an isogenic colorectal cancer cell model of the most common somatic *POLE* exonuclease domain mutation, $POLE^{\mathrm{P286R}}$

(Chapter 3); the generation and phenotypic characterisation of intestinal-specific, uterine-specific and constitutive mouse models of the most comomon germline *POLE* exonuclease domain mutation, $Pole^{L424V}$ (Chapter 4); the sequencing and molecular analysis of a large cohort of colorectal pre-cancers (adenomas) (Chapter 5).

My results demonstrate that:

1. Somatic *POLE* exonuclease domain mutations cause ultramutation, independent of secondary cancer driver mutations.

2. Somatic *POLE* exonuclease domain mutations cause the specific mutation spectrum reported in Alexandrov *et al.*, (Signature 10) (1).

3. Germline $Pole^{L424V}$ mutations *in vivo* cause no gross macroscopic or histolopathological abnormalities in the intestine or uterus, although the germline $Pole^{L424V}$ mutation increases the colonic polyp burden in $Apc^{Min}$ mice.

4. Somatic *POLE* exonuclease domain mutations are an early or initiating event in colorectal and endometrial cancer.

### 6.1.1 Somatic *POLE* exonuclease domain mutations causes ultramutation and mutational bias in cancer

Perhaps the most notable finding from my work was the confirmation that cancer-associated *POLE* exonuclease domain mutations are causal for tumour ultramutation and a shift in mutational bias, in keeping with the characteristic mutation signature found in these cancers (1). *POLE* encodes the catalytic subunit of Pol $\epsilon$, the major leading strand DNA replicase, but I found no evidence that somatic *POLE* exonuclease domain mutations cause stalling of replication forks, induction of the DNA damage response or a cell cycle defect. There was some evidence, however, that polymerase exonuclease domain mutations may subtly affect its intracellular DNA replication rate. Furthermore, the alteration in mutation rate and spectrum in the isogenic *POLE* exonuclease domain mutant cells occurred despite apparently proficient mismatch repair. Collectively, these data suggest that in mammalian cells, the

ultramutator phenotype and mutation spectrum is an intrinsic property of the mutant polymerase protein.

Repair of DNA mismatches by Pol $\epsilon$ relies on the balance between the polymerase and exonuclease activity of the enzyme (298). Under normal conditions, wildtype Pol $\epsilon$ is balanced towards polymerisation, however, upon incorporating a DNA mismatch, the balance is thought to shift towards exonuclease activity, enabling the polymerase to excise the mismatched base (298).

Structural data from the catalytic core of *S. cerevisiae pol2-P301R* (analogous to human *POLE*-P286R, the most common somatic *POLE* exonuclease domain mutation in cancer), demonstrated that the mutant arginine residue perturbs single-stranded DNA binding in the exonuclease active site (299). The steric hindrance is predicted to significantly reduce exonuclease activity by shifting the balance towards increased polymerase activity, thus increasing the propensity of the mutant enzyme to polymerise past mismatches resulting in a mutation rate increase (299). The prediction is supported by cell-free assays of the *S. cerevisiae* Pol $\epsilon$ P286R mutant holoenzyme, which extends DNA more readily from a mismatched base than the wildtype enzyme and replicates DNA at a faster rate than the wildtype polymerase, thus increasing the mutation rate (300).

Although this is a plausible mechanism for *POLE* exonuclease domain mutant-induced ultramutation, I did not detect a substantial increase in DNA fibre length in mammalian $POLE^{\mathrm{P286R}}$ mutant cancer cells compared to wildtype isogenic derivatives. Possible reasons for this discrepancy include: an additional factor in cells not present *in vitro*, such as the replicative helicase or PCNA regulating DNA replication fork speed; the DNA fibre assay is not sufficiently sensitive to detect such an increase in DNA polymerase replication velocity, particularly in a heterozygous mutant cell line; or variation in the mechanism of *POLE* exonuclease domain mutation-induced mutagenesis between budding yeast and humans. For example, $POLE^{\mathrm{V411L}}$ is associated with ultramutation in human tumours (123) but does not exhibit an increased mutation rate in budding yeast (163).

To determine the precise mechanism of cancer-associated *POLE* exonuclease domain mutation-induced mutagenesis in mammalian cells, accurate quantification of DNA replication velocity by exonuclease domain mutant Pol $\epsilon$ is required. DNA

combing, although similar to DNA fibre analysis, elongates labelled DNA strands on a glass slide at a constant rate, permitting accurate quantification of intracellular DNA replication speed.

### 6.1.2 Germline *Pole*$^{L424V/+}$ mutation increases in colonic polyp burden in tumour-prone *Apc*$^{Min}$ mice

An intriguing finding from Chapter four was the difference in phenotype between murine and human carriers of *POLE* L424V mutations. Patients with germline *POLE*$^{L424V}$ mutations are genetically predisposed to develop intestinal polyps (92% of cases) and colorectal and endometrial cancer (61% and 2% respectively) at a mean age of 39 years (144). In contrast, heterozygous *Pole*$^{L424V/+}$ knock-in mice of a similar biological age have no gross morphological or histopathological abnormalities of the intestine and uterus. Indeed, it was only when crossed with *Apc*$^{Min}$ that a significant increase in colonic polyp burden, associated with the germline Pole$^{L424V/+}$ mutation, was observed.

Why is *Pole*$^{L424V}$ alone unable to initiate tumourigenesis in mice when it is highly penetrant in humans? As discussed in Chapter four, loss of base-excision repair proteins *MUTYH* or *NTHL1* in humans cause autosomal recessive colorectal cancer syndromes, MUTYH-associated polyposis (MAP) and NTHL1-associated polyposis (NAP) respectively (23, 30). *Myh* and *Nthl*-knockout mice, however, are indistinguishable from wildtype controls (286, 288).

It was postulated by Sieber *et al.* that in *Myh*-deficient mice, the intrinsic increase in mutation rate was not sufficient to induce intestinal polyposis (29). In humans, the vast majority MAP colonic adenomas and carcinomas harbour truncating mutations in *APC*, caused by an accumulation of 8-oxoG-induced C→T transversions (301). In humans, germline *POLE*$^{L424V}$ exonuclease domain mutant colorectal and endometrial cancers have a preponderance of nonsense mutations in cancer driver genes such as *TP53*, *FBXW7* and *APC* (123, 144). Together with the observation that mice harbouring germline *Pole*$^{L424V/+}$ mutations significantly increase colonic polyp burden in *Apc*$^{Min}$ mice, suggest that alleles that confer a modest mutator phenotype, such as *Myh*$^{-/-}$ and *Pole*$^{L424V}$ may require additional mutations in cancer driver genes

(mini-drivers (9)) to initiate tumourigenesis.

*Myh*-deficiency increases cystic crypt burden in the small intestine and significantly increases colonic aberrant crypt foci (ACF), precursors to colonic polyps, in $Apc^{\mathrm{Min}}$ mice (29). The limited lifespan of $Apc^{\mathrm{Min}}$ mice (typically caused by blood loss secondary to small intestinal polyposis) means that small intestinal and colonic polyps may not have sufficient time to develop under a modest mutator phenotype. Quantification of cystic crypts and colonic ACFs in $Apc^{\mathrm{Min}}$,$Pole^{\mathrm{L424V/+}}$ compound mutant mice may be a more sensitive measure of a histopathological phenotype (discussed in Section 6.2).

Human germline and somatic *POLE*-mutant cancers tend to harbour nonsense mutations at predicted functionally sub-optimal codon positions in cancer driver genes (10). It is plausible that these additional driver mutations in *APC* (and other driver genes) may be a pre-requisite for *POLE* exonuclease domain-induced tumourigenesis. Furthermore, functionally sub-optimal nonsense mutations in *APC* that permit residual Wnt signalling may be a conditional factor for germline *POLE* exonuclease domain mutation-induced tumourigenesis in the intestine and possibly the uterus. It is in *POLE* exonuclease domain mutant endometrial tumours that *APC* nonsense mutations tend to occur over classic $\beta$-catenin gain of function mutations (302, 303). Such mutational bias towards predicted functionally sub-optimal *APC* mutations in *POLE* exonuclease domain mutant cancer is particularly pertinent, as constitutively high $\beta$-catenin levels in melanoma has been demonstrated to inhibit T-cell priming, reducing tumour immunogenicity (304, 305). Residual Wnt signalling *POLE* exonuclease domain tumours, therefore, may be a factor in their high immunogenicity and good prognosis (166, 180).

It will be of particular interest to perform whole genome sequencing of colonic polyps from $Apc^{\mathrm{Min}}$,$Pole^{\mathrm{L424V/+}}$ compound mutant mice to determine if $POLE^{\mathrm{L424V}}$ confers a mutator phenotype and to analyse second hits in *Apc*. Moreover, analysis of Wnt signalling by *Apc* and $\beta$-catenin immunohistochemistry, western blotting and RNAseq to quantify if residual Apc and $\beta$-catenin expression is present in $Pole^{\mathrm{L424V}}$ exonuclease domain mutant and wildtype mice in an *Apc*-mutant background would be informative. Finally, the introduction of specific *Apc* mutations in intestinal mouse organoids using CRISPR-Cas9 would permit analysis of Wnt pathway inactivation

associated with specific *Apc* mutations observed in $Pole^{L424V}$ exonuclease domain mutant polyps.

### 6.1.3 Therapeutic implications of *POLE* exonuclease domain mutations

Data from this thesis has confirmed that somatic *POLE* exonuclease domain mutations occur in colorectal pre-cancers and exhibit an increased T-cell infiltrate and mutation rate, concordant with that observed in *POLE* exonuclease domain mutant colorectal carcinomas (166).

Clinical trials of immune checkpoint inhibitors PD-L1 (Avelumab) and PD-L (Pembrolizumab) ((NCT02912572, NCT02899793, NCT02658279 and 2017-000370-10 (EudraCT)) are currently ongoing in patients with recurrent or metastatic ultramutated *POLE* exonuclease domain mutant endometrial cancer, colorectal cancer and glioblastomas. Whether the apparently increased T-cell infiltrate in *POLE* exonuclease domain mutant adenomas will prove useful in their stratification requires future studies. Finally, it is of interest to determine T-cell infiltrate in colorectal adenomas from patients with germline *POLE* and *POLD1* exonuclease domain mutations, as this could assist in determining suitable treatment options for these patients.

Preliminary compound sensitivity assays in *POLE* exonuclease domain mutant isogenic colorectal cancer cell lines demonstrated their utility as clinically relevant models for preclinical testing of cytotoxics and radiotherapy to target ultramutated *POLE* exonuclease domain mutant cells (discussed in Future Work 6.2).

## 6.2 Ongoing and future work

Several questions have arisen throughout this thesis which require further investigation. Below is a summary of ongoing and future experiments that aim to address these.

1. **Quantification of the mutation rate and the mutation spectrum in *POLE* exonuclease domain mutant isogenic colorectal cancer cell lines**

**by whole genome sequencing**

The *HPRT1* mutation rate in $POLE^{P286R}$ mutant colorectal cancer cell lines was significantly elevated compared to $POLE^{WT}$ isogenic derivatives. As discussed in Section 3.4.2, reporter genes underestimate global genomic mutation rates by several fold compared to whole exome sequencing (241). Whole genome sequencing of *POLE* exonuclease domain mutant isogenic cell lines will permit precise quantification of the global mutation rate and enable the analysis of mutations in a strand-specific, replication timing and methylation-specific context.

Hodel *et al.* utilised whole exome sequencing to quantify the mutation rate in colorectal HCT116 cells from a bulk cell sample, however, this technique was unable to detect newly acquired mutations in single cells and small sub-clones. This technique was not, therefore, an accurate quantification of mutation rate at the given time point (241). As an alternative, I have devised a method, utilising clonal expansion of cell lines, to quantify the absolute mutation rate at a specified time point and this will be performed by members of the Tomlinson laboratory, University of Birmingham. The method is outlined in the schematic in Figure 6.1.

A single mammalian cell will be cultured until the first cell division, then one daughter cell sub-cloned to sequence DNA at $P_o$. Similar to Hodel *et al.* (241), the remaining single cell will be cultured for 70 cell doublings ($P_x$), split into single cells to undergo clonal expansion. DNA will be extracted from each expanded clone and whole genome sequenced at low-depth to detect clonal mutations. Clonal mutations from $P_o$ will be removed from variants called from expanded clones as they are likely to be pre-existing or germline mutations. The remaining clonal mutations (VAF>3) will be analysed for mutation rate and spectrum.

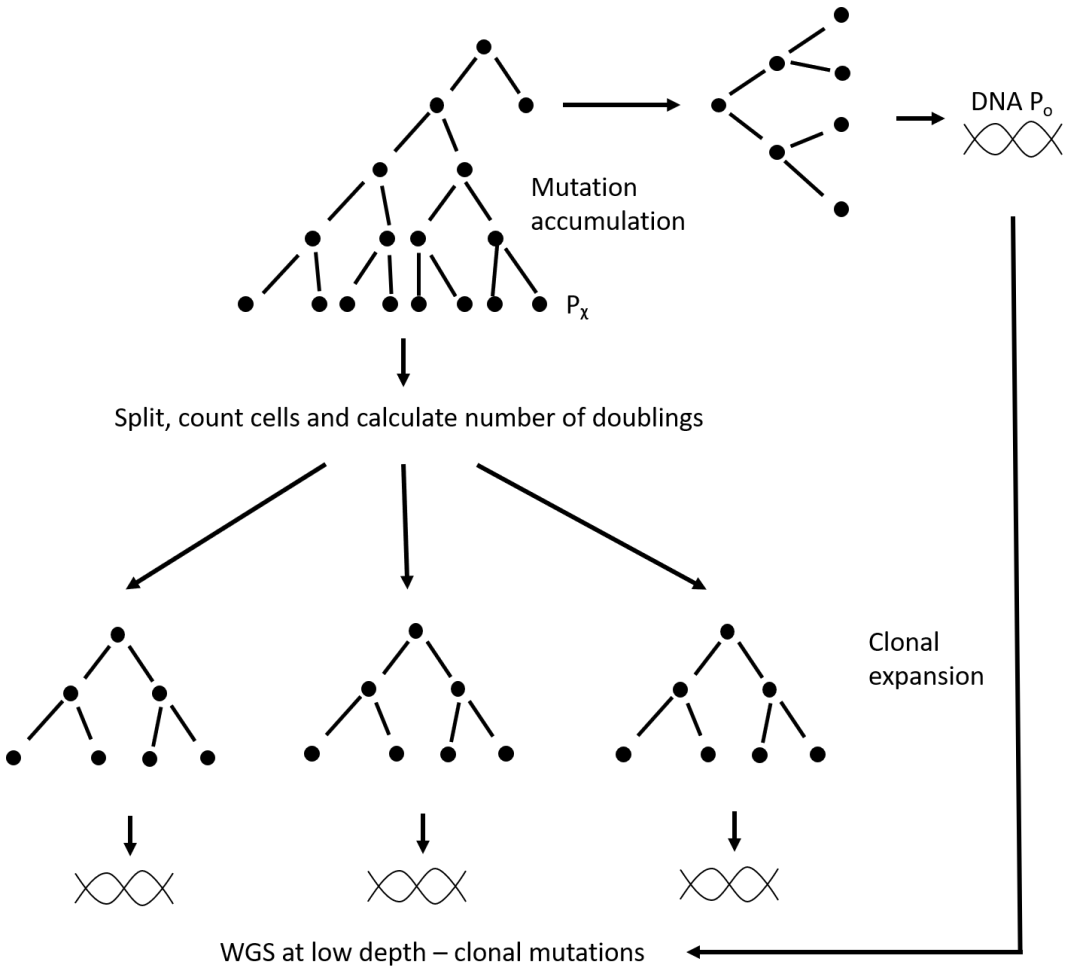



**Figure 6.1:** Whole genome sequencing mutation rate analysis schematic in *POLE* exonuclease domain mutant isogenic cell lines. A cell is seeded to undergo one round of DNA replication. The daughter cell is harvested and re-seeded on a separate plate for clonal expansion and DNA extraction ($P_o$). The remaining cell is cultured for 70 cell doublings followed by harvesting and re-plating as single cells for clonal expansion. DNA is extracted from expanded cells then whole genome sequenced at a low depth (30X). DNA is mapped and variants called followed by variant filtering to include clonal mutations (VAF >0.3). Variants detected in $P_o$ (SNPs and pre-existing somatic mutations) are subtracted from variants detected in $P_x$.

2. **Lethality or reduced viability in *POLE*$^{P286R}$ and constitutive MMR-deficient cells**

   Previous studies have suggested that mismatch repair deficiency is a pre-requisite for cancer-associated *POLE* exonuclease domain mutation-associated ultramutation (164, 241). Mismatch repair deficiency and *POLE* exonuclease domain mutations, however, rarely co-occur in tumours. Experiments presented in

Chapter three confirmed that transient knockdown of the MMR proteins, MLH1 and MSH6, does not decrease the cell viability or increase double-stand breaks in $POLE^{P286R}$ mutant colorectal cancer cell lines; however it is possible that the time-frame of knockdown was insufficient to compromise cell viability.

To determine if MMR deficiency reduces viability or increases DNA damage response in combination with $POLE^{P286R}$ mutations in human tumour cells over a longer period of time, it will be informative to knock out *MLH1* or *MSH6* from *POLE* exonuclease domain mutant and wildtype isogenic cell lines using CRISPR-Cas9. Mismatch repair knockout cell lines will be cultured for 70 cell doublings with cell viability and DNA damage response assessed periodically, using the assays described in Chapter three.

It will be of interest to correlate specific novel mutations with cell viability to detect potential antimutator mutations, which were observed in *POLE* exonuclease-deficient budding yeast with MSH2 and MSH6 deficiency (128, 129). Using the protocol devised in Figure 6.1, the mutation rate and spectrum will be quantified in MMRD *POLE* exonuclease domain mutant cell lines.

3. **Quantification of small intestinal cystic crypt and colonic abberant crypt foci burden in $Apc^{Min}Pole^{L424V/+}$ compound mutant mice**

As discussed in Section 6.1.2, quantification of cystic crypts and colonic ACF may be a more sensitive measure of a histopathological phenotype in $Apc^{Min}$, $Pole^{L424V/+}$ compound mutant mice. Using a protocol adapted from Yamada *et al.*, the mucosal surface of the small intestinal and colon will be observed microscopically and cystic crypts and ACFs quantified respectively (289).

The small intestine and colon from experimental $Apc^{Min}Pole^{L424V/+}$ and control $Pole^{L424V/+}$ mice will be divided into longitudinal sections and embedded crypt/villi-side down in paraffin. Serial *en face* sections at 4$\mu$m will be cut across the mucosal surface and stained with H&E to quantify microscopic lesions (Figure 6.2).

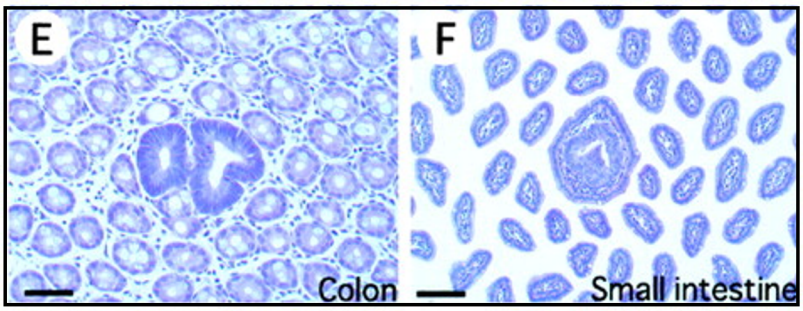



**Figure 6.2:** H&E-stained FFPE tissue sections of the small intestine and colonic mucosal surface showing small intestinal cystic crypts and colonic aberrant crypt foci in $Apc^{\mathrm{Min}}$ mice. Scale bar represents 50$\mu$m. Figure adapted from Yamada *et al.* (289).

As mentioned above, the limited lifespan of $Apc^{\mathrm{Min}}$ mice may not permit sufficient time for growth of macroscopic polyps in $Apc^{\mathrm{Min}}$, $Pole^{\mathrm{L424V/+}}$ compound mutant mice. In addition to quantifying microscopic precursor lesions, crossing $Pole^{\mathrm{L424V/+}}$ mice with an *Apc* mutant mouse that has an attenuated intestinal polyposis phenotype such as $Apc^{1638\mathrm{N}}$ (306), may reveal a gross morphological consequence of a modest mutator phenotype.

4. **Quantification of tissue-specific mutation rate in germline $\boldsymbol{Pole^{\mathrm{L424V/+}}}$-derived organoids**

   Whilst $Pole^{\mathrm{L424V/+}}$ mice do not display evidence of tumour predisposition, it is possible that the *Pole* L424V mutation alone also increases the basal mutation rate by an amount insufficient to cause tumourigenesis in the mouse.

   Determining the tissue-specific mutation rate using organoids derived from germline $Pole^{\mathrm{L424V/+}}$ mice will address this and may explain why tumours from patients carrying germline *POLE* exonuclease domain mutations (and other DNA repair defects) tend to occur in the intestine, uterus and brain.

   In collaboration with Dr Maria Pinna, Tomlinson group, University of Birmingham, using published protocols (307, 308, 309, 310, 311), we have successfully generated organoids from the small intestine and colon, uterus, liver, skin, and oesophagus from germline $Pole^{\mathrm{L424V}}$ and wildtype age-matched mice. Sequencing of organoids will be adapted from the protocol published in Jager *et al.* (312) (Figure 6.3) to obtain the mutation rate.

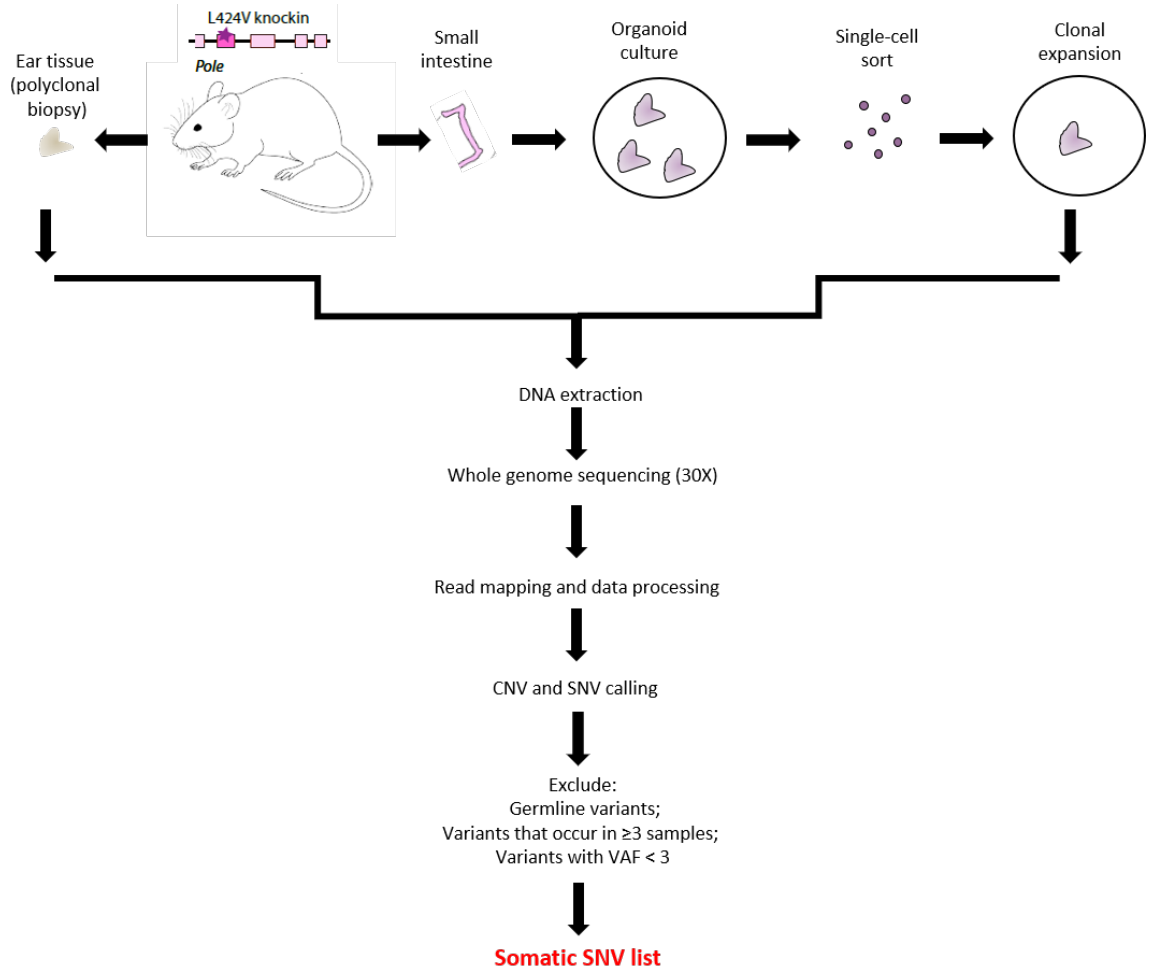



**Figure 6.3:** A schematic adapted from Jager *et al.* to quantify mutation rate and spectrum in germline *Pole*$^{L424V/+}$ mice using organoid culture and whole genome sequencing (312). Organoids are cultured from age-matched germline *Pole*$^{L424V}$ and wildtype littermates. An ear clip is taken at necropsy to use as a germline DNA control. Organoids are split into single cells to undergo clonal expansion followed by DNA extraction. Both organoid and control DNA will undergo whole genome sequencing to extract somatic variants that were generated during the lifespan of the mouse.

Organoid cultures will be dissociated into single stem cells by FACS sorting and plated at a limiting dilution to undergo clonal expansion. Clonal organoids will be cultured for approximately six weeks, splitting 1:4 every week (1:2 if slow-growing). DNA from clonal organoid cultures, sourced from from at least six wells in a 24 well plate and normal ear tissue as a control, will be extracted to undergo whole genome sequencing at 30X, according to standard protocols. In collaboration with bioinformaticians in the Tomlinson laboratory, the data will undergo bioinformatic analysis (Figure 6.3). Firstly, copy number variation

analysis will be performed as if the ploidy deviates from two, this could affect SNV calling. Single nucleotide variants will be mapped and called using GATK according to the parameters set out in Jager *et al.* (312). Germline variants identified in the polyclonal ear biopsy and variants that occur in three or more samples which could represent unknown SNPs or sequencing mutation hotspots will be excluded. Variants with a VAF <3 will be excluded to remove sub-clonal variants. These data will be used to determine the lifetime mutation rate and spectrum of $POLE^{L424V/+}$ mice in various tissue types. For comparison between litters, mutation rate will be adjusted for age at the time of necropsy.

5. **High throughput compound screen in *POLE* exonuclease domain mutant isogenic cell lines**

   The *POLE* exonuclease domain mutant and wildtype isogenic cell lines are useful pre-clinical models. It will be informative to test the viability of these cells when exposed to a wide range of cytotoxics, using a high throughput compound screen to identify potential cytotoxics that specifically target *POLE* exonuclease domain mutant cancer cells. The screen will determine if the favourable outcome of patients with *POLE* exonuclease domain mutant cancers is independent of neoadjuvent chemotherapy and radiotherapy.

   In collaboration with Dr Farhat Kanhim, University of Birmingham, a high throughput compound screen will be performed on $POLE^{P286R}$ and $POLE^{WT}$ isogenic colorectal cancer cell lines, to identify compounds that result in a differential reduction in cell viability.

## 6.3 Concluding statement

In this thesis I have shown that $POLE^{P286R}$ exonuclease domain mutations cause ultramutation and the distinct mutation signature observed in *POLE* exonuclease domain mutant cancers. Although *POLE* exonuclease domain mutations are found in colorectal and endometrial pre-cancers, data from $Pole^{L424V/+}$ exonuclease domain mutant mice suggest that functionally sub-optimal mutations in secondary cancer-driver genes may be a pre-requisite for *POLE* exonuclease domain mutant tumour

initiation. Future work will seek to better understand the mechanism of ultramutation and the specific mutation signature in *POLE* exonuclease domain mutant tumours. This will aid detection and therapeutic targeting of tumours carrying *POLE* exonuclease domain mutations, with the hope of improving patient prognosis.

# Appendix A

# Primer sequences and PCR conditions

| Primer name | Primer sequence 5'-3' | product size (bp) |
|---|---|---|
| **Amplification of cloned sgRNAs** | | |
| PX335 cloning plasmid F | GGCCTATTTCCCATGATTCCT | |
| PX335 cloning plasmid R | GGGCCATTTACCGTAAGTTATGT | |
| Used to Sanger sequence directly from the plasmid | | |
| ***POLE* exon 9 amplification** | | |
| *POLE* Exon 9F Blood/tissue | CTAACAGTGGGGCAGATGCT | |
| *POLE* Exon 9R Blood/tissue | GGTGTTCAGGGAGGCCTAAT | 258 |
| $MgCl_2$, 1.5mM; Annealing temperature, 65°C; Ext. time, 30 seconds | | |
| **Primers for CRISPR off-target effects** | | |
| *ARID2* F | TGAAATGCATGTGGGAAGTCT | 568 |
| *ARID2* R | GGTTGTAGAATGTGGTGAGACA | |
| *BRCA2* F | TCTATTGAGACTGTGGTGCCA | 596 |
| *BRCA2* R | ACAGTTTGTGGGTATGCATTTG | |
| *SYNE1* F | AGGAGATCAAGACCACGGTG | 493 |
| *SYNE1* R | AAGGCCACTGAAATGATTGATC | |
| $MgCl_2$, 1.5mM ; Annealing temperature, 60°C °C; Ext. time, 30 seconds | | |

**Table A.1:** Primers for the generation and genetic characterisation of CRISPR-Cas9n constructs and cell lines. Standard PCR conditions were used in all cases (Methods Section 2.2.4).

| Primer name | Primer sequence 5'-3' | Product size (bp) |
|---|---|---|
| HPRT_1F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTCCTCCTGAGCAGTCAGC | 283 |
| HPRT_1R | AGACGTGTGCTCTTCCGATCTCGTGACGTAAAGCCGAACC | |
| HPRT_2aF | ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGGCCTGTTGTTTTCTTACAT | 304 |
| HPRT_2aR | AGACGTGTGCTCTTCCGATCTCCCTTTCCAAATCCTCAGCA | |
| HPRT_2bF | ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGTGATGATGAACCAGGTTATGAC | 243 |
| HPRT_2bR | AGACGTGTGCTCTTCCGATCTCAAATAGCAAGTACTCAGAACAGC | |
| HPRT_3F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGGTTGGTGTGGAAGTTTAATGA | 408 |
| HPRT_3R | AGACGTGTGCTCTTCCGATCTTGGTTTGCAGAGATTCAAAGAAA | |
| HPRT_4F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCTAGCTAACTTCTCAAATCTTCT | 383 |
| HPRT_4R | AGACGTGTGCTCTTCCGATCTCCTAGACTGCTTCCAAGGGT | |
| HPRT_5F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGGGTTGTTATGATGTGATTTGACT | 245 |
| HPRT_5R | AGACGTGTGCTCTTCCGATCTGGAATTTCTCTCCCTGGCTT | |
| HPRT_6aF | ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTGTGCTGAAAAGAATGACAGT | 244 |
| HPRT_6aR | AGACGTGTGCTCTTCCGATCTTACCTTGCGACCTTGACCAT | |
| HPRT_6bF | ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGAAAGGATATAATTGACACTGGCA | 343 |
| HPRT_6bR | AGACGTGTGCTCTTCCGATCTGGCAAGGAAGTGACTGTAATTATGA | |
| HPRT_7F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCACGGATGAAATGAAACAGT | 304 |
| HPRT_7R | AGACGTGTGCTCTTCCGATCTGCCTCTCTACAAATATTCTCTAAGC | |
| HPRT_8F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGAGAATATTTGTAGAGAGGCACA | 309 |
| HPRT_8R | AGACGTGTGCTCTTCCGATCTTGAGGTGCTGGAAGGAGAAA | |
| HPRT_9aF | ACACTCTTTCCCTACACGACGCTCTTCCGATCTACACTCCCCAAAAGTTACTGATC | 347 |
| HPRT_9aR | AGACGTGTGCTCTTCCGATCTAGATGGCCACAGAACTAGAACA | |
| HPRT_9bF | ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGTTCTAGTTCTGTGGCCATCT | 406 |
| HPRT_9bR | AGACGTGTGCTCTTCCGATCTGCTTCTTACTTTTCTAACACACGG | |
| HPRT_9cF | ACACTCTTTCCCTACACGACGCTCTTCCGATCTACCACCGTGTGTTAGAAAAGT | 381 |
| HPRT_9cR | AGACGTGTGCTCTTCCGATCTACAAGGAAAATGGAACCAACAGA | |

**Table A.2:** *HPRT1* exon sequencing primers including Illumina next generation sequencing adapters. All PCR reactions were performed using Roche® FastStart High Fidelity (HiFi) Taq (Sigma) (Methods Section 2.2.4.2), 1.8mM $MgCl_2$, 30 seconds extension time, $60^0$C annealing temperature.

| | |
|---|---|
| **_POLE_ full-length cDNA overexpression primers construct** | |
| Full *POLE* cDNA F | CTATATACCGGTATGTCTCTGAGGAGCGGCGGGC |
| Full *POLE* cDNA R | CTATATGCTAGCCTAAGCGTAATCTGGAACATCGTATGGGTAATGGCCCAGCTGTGGGTTC |
| Long-range PCR reaction (Methods Section (2.2.4.1): 7 minute annealing/extension time, 1.5mM $MgCl_2$. | |
| **Site-directed mutagenesis primers for generation of _POLE_ exonuclease domain mutant overexpression constructs** | |
| SDM P286R 857C→G F | CTGCCCCTCAAGTTTCGTGATGCTGAGACAGAC |
| SDM P286R 857C→G R | GTCTGTCTCAGCATCACGAAACTTGAGGGGCAG |
| SDM D275A/D277A EXONULL F | GTGGTTTTGGCATTTGCCATTGCGACGACCAAACTGCCC |
| SDM D275A/D277A EXONULL R | GGGCAGTTTGGTCGTCGCAATGGCAAATGCCAAAACCAC |
| Reaction performed according to protocol in Methods Section 2.2.4.3 | |
| **_POLE_ cDNA sequencing primers** | |
| *POLE* R 2 | GCGCATGTCCACAATGTTG |
| *POLE* F 3 | ACCTCTAAGAAGATAGCTGACCA |
| *POLE* R 3 | GGCCGCCTTGAGATTATGAC |
| *POLE* F 4 | GTGGGTGAAGAGGGACAGTT |
| *POLE* R 4 | CCCTTTGAACTCGTAACGCC |
| *POLE* F 5 | GCGGGAAAACTCCTTCTACG |
| *POLE* R 5 | TCTGCTCCCCGTAATCTTCC |
| *POLE* F 6 | AACATGCCTGACTCTGAGCT |
| *POLE* R 6 | CTGGTCTCGCTGATCTGCA |
| *POLE* F 7 | CTTGCGAAGAACTGCCCG |
| *POLE* R 7 | CTGCTCATCTCGAAGGCCT |
| *POLE* F 8 | TCAACTATGGGGTCCTGGAC |
| *POLE* R 8 | AGTCTTGCAGTCCACAGTGA |
| *POLE* F 9 | CTCGATGCTGGGAATTTCTTCT |
| *POLE* R 9 | TGTAGAACTTCCACCAGCGT |
| Used to directly sequence from the plasmid | |

**Table A.3:** Sequencing and site-directed mutagenesis primers for the generation of *POLE* overexpression constructs.

| Primer name | Primer sequence 5'-3' | Product and size (bp) |
|---|---|---|
| | **Multiplex PCR - *Pole*$^{fl(L424V)}$ genotyping** | |
| Pole2 | TTTGAGAGGCACAGAGAGGC | |
| Pole1 | GCCCCGTAATGCAGAAGAAG | floxed (668) |
| Pole3 | GCTCTACAGGGTCATAGCCA | wildtype (364) |
| | $MgCl_2$, 1.5mM; Annealing temperature, 60°C; Ext. time, 45 seconds | |
| | **Independent PCRs (same forward primer) - Recombination PCR *Pole* L424V** | |
| RecombinedF | CCATGACGTCGGATCCAGAT | |
| RecombinedR | TTTCCCAAGGCAGTCTGGAG | Unrecombined PCR 1 (582) |
| Exon13R | TGGCTATGACCCTGTAGAGC | Recombined PCR 2 (196) |
| | $MgCl_2$, 1.5mM; Annealing temperature, 55°C; Ext. time, 30 seconds | |
| | ***Pole* cDNA sequencing primers** | |
| *Pole*$^{L424Vfl}$ Exon11F | TTATGGTCACCTACAATGGGG | |
| *Pole*$^{L424Vfl}$ Exon15R | CACCAGACTCTAGTGCCTCC | *Pole* exon 13 cDNA sequencing (666) |
| | $MgCl_2$,1.5mM ; Annealing temperature,60°C ; Ext. time, 30 seconds | |
| | ***Cre* genotyping** | |
| *Cre* F | CGGTCGATGCAACGAGTGATGAGG | |
| *Cre* R | CCAGAGACGGAAATCCATCGCTCG | Cre *600* |
| | $MgCl_2$, 1.5mM; Annealing temperature, 55°C; Ext. time, 60 seconds; Q-solution, 0.4X | |
| | **Multiplex PCR - *Apc*$^{Min}$ genotyping** | |
| *Apc*$^{Min}$ MT | TGAGAAAGACAGAAGTTA | |
| *Apc*$^{Min}$ 9 | GCCATCCCTTCACGTTAG | wildtype (618) |
| *Apc*$^{Min}$ 15 | TTCCACTTTGGCATAAGGC | mutant (327) |
| | $MgCl_2$, 1.5mM; Annealing temperature, 55 °C; Ext. time, 60 seconds | |
| | ***Pla2g2a* genotyping** | |
| Pla2g2a F | GTCCAAGGGAACATTGCG | |
| Pla2g2a R | AGAACAGGTGATTTGGCCC | Pla2g2a (476) |
| | $MgCl_2$, 1mM; Annealing temperature, 55°C; Ext. time, 60 seconds | |

**Table A.4:** Mouse genotyping and sequencing primers. All reactions use a standard PCR reaction using Taq (Bioline).

| Primer name | Primer sequence 5'-3' | product size (bp) |
|---|---|---|
| *POLE* Exon 9F FFPE | TGCTTATTTTGTCCCCACAG | |
| *POLE* Exon 9R FFPE | TACTTCCCAGAAGCCACCTG | 153 |
| *POLE* Exon 13F FFPE | TCTGTTCTCATTCTCCTTCCAG | |
| *POLE* Exon 13R FFPE | CGGGATGTGGCTTACGTG | 174 |
| *POLE* Exon 14F FFPE | TCTGGCGTTCTCTCCTCAG | |
| *POLE* Exon 14R FFPE | CGACAGGACAGATAATGCTCA | 155 |
| Standard PCR: $MgCl_2$, 1.5mM; Annealing temperature, 55 °C; Ext. time, 30 seconds | | |

**Table A.5:** FFPE DNA *POLE* sequencing primers.

# Appendix B

# Genes distal to *POLE* on chromosome 12q

| **Genes distal to *POLE* on chr12q** |
|---|
| *PXMP2* |
| *PGAM5* |
| *ANKLE2* |
| *COLGA3* |
| *CHFR* |
| *ZNF605* |
| *ZNF26* |
| *ZNF84* |
| *ZNF140* |
| *ZNF891* |
| *ZNF269* |
| *ZNF10* |

**Table B.1:** Genes distal to *POLE* on chromosome 12q.

# Appendix C

# Comprehensive list of Lesch Nyhan syndrome variants

**Table C.1:** A comprehensive list of published Lesch-Nyhan syndrome variants collated from http://www.lesch-nyhan.org/en, accessed 12/12/2018.

| Mutation (HGVS) | Trinucleotide Context | Result |
|---|---|---|
| c.1A >T | ATA >TAT | Met >Leu |
| c.2T >C | ATG >ACG | Met >Thr |
| c.3G >A | CCA >CTA | Met >Ile |
| c.20G >A | GCC >GTC | Gly >Asp |
| c.23T >G | GTC >GGC | Val >Gly |
| c.29T >G | ATT >AGT | Ile >Ser |
| c.29T >A | ATT >AAT | Ile >Asn |
| c.35A >C | ATC >AGC | Asp >Ala |
| c.46G >A | CCT >CTT | Gly >Ser |
| c.46G >C | CCT >CGT | Gly >Ala |
| c.47G >A | ACC >ATC | Gly >Asp |
| c.47G >T | ACC >AAC | Gly >Val |
| c.50A >G | ATA >ACA | Tyr >Cys |

Continued on next page

Table C.1 – continued from previous page

| **Mutation (HGVS)** | **Trinucleotide Context** | **Result** |
|---|---|---|
| c.59A >T | ATC >AAC | Asp >Val |
| c.65T >G | TTT >TGT | Phe >Cys |
| c.68G >T | GCA >GAA | Cys >Phe |
| c.69C >G | GCA >GGA | Cys >Trp |
| c.73C >A | ACC >AAC | Pro >Thr |
| c.74C >G | CCT >CGT | Pro >Arg |
| c.84T >A | ATG >AAG | Tyr >stop |
| c.85G >C | GCA >GGA | Ala >Pro |
| c.93T >G | ATT >AGT | Asp >Glu |
| c.95T >C | TTG >TCG | Leu >Ser |
| c.96G >C | CCA >CGA | Leu >Phe |
| c.112C >T | TCC >TTC | Pro >Ser |
| c.113C >T | CCT >CTT | Pro >Leu |
| c.115C >G | TCA >TGA | His >Asp |
| c.118G >A | CCA >CTA | Gly >Arg |
| c.122T >C | CTA >CCA | Leu >Pro |
| c.124A >T | ATT >AAT | Ile >Phe |
| c.125T >C | ATT >ACT | Ile >Thr |
| c.130G >T | TCC >TAC | Asp >Try |
| c.130G >C | TCC >AGC | Asp >His |
| c.131A >T | GTC >GAC | Asp >Val |
| c.131A >G | GTC >GCC | Asp >Gly |
| c.133A >G | CTG >CCG | Arg >Gly |

Continued on next page

Table C.1 – continued from previous page

| **Mutation (HGVS)** | **Trinucleotide Context** | **Result** |
|---|---|---|
| c.134G >A | CCT >CTT | Arg >Lys |
| c.140A >T | TTC >TAC | Glu >Val |
| c.140A >G | TTC >TCT | Glu >Gly |
| c.142C >T | ACG >ATG | Arg >Cys |
| c.143G >A | ACG >ATG | Arg >His |
| c.145C >T | TCT >TTT | Leu >Phe |
| c.146T >G | CTT >AGT | Leu >Arg |
| c.148G >A | GCA >GTA | Ala >Thr |
| c.148G >C | GCA >GGA | Ala >Pro |
| c.149C >T | GCT >GTT | Ala >Val |
| c.149C >A | GCT >GAT | Ala >Asp |
| c.151C >G | TCG >TGG | Arg >Gly |
| c.151C >T | TCG >TTG | Arg >stop |
| c.152G >A | GTG >GAG | Arg >Gln |
| c.152G >C | TCG >TGG | Arg >Pro |
| c.157G >A | ACA >ATA | Val >Met |
| c.158T >C | GTG >GCG | Val >Ala |
| c.158T >G | GTG >GGG | Val >Gly |
| c.160A >C | ATC >AGC | Met >Leu |
| c.169A >G | ATC >ACC | Met >Val |
| c.170T >C | ATG >ACG | Met >Thr |
| c.173G >A | TCC >TTC | Gly >Arg |
| c.179A >G | ATG >ACG | His >Arg |

Continued on next page

Table C.1 – continued from previous page

| **Mutation (HGVS)** | **Trinucleotide Context** | **Result** |
|---|---|---|
| c.190G >C | GCT >GGT | Ala >Pro |
| c.191C >A | GCC >GAC | Ala >Asp |
| c.193C >T | CCT >CTT | Leu >Phe |
| c.194T >C | CTC >CCC | Leu >Pro |
| c.196T >C | CTG >CCG | Cys >Arg |
| c.200T >C | GTG >CCG | Val >Ala |
| c.203T >C | CTC >CCC | Leu >Pro |
| c.203T >G | CTC >CCC | Leu >Arg |
| c.208G >A | CCC >CTC | Gly >Arg |
| c.208G >T | CCC >CAC | Gly >Trp |
| c.208G >C | CCC >CGC | Gly >Arg |
| c.209G >A | CCC >CTC | Gly >Glu |
| c.211G >C | CCC >CGC | Gly >Arg |
| c.212G >A | GCC >GTC | Gly >Asp |
| c.212G >T | GCC >GAC | Gly >Val |
| c.215A >G | ATA >ACA | Tyr >Cys |
| c.220T >G | ATT >AGT | Phe >Val |
| c.221T >G | TTC >TGC | Phe >Cys |
| c.222C >A | TCT >TAT | Phe >Leu |
| c.222C >G | TCT >TGT | Phe >Leu |
| c.232C >G | CCT >CGT | Leu >Val |
| c.233T >A | CTG >CAG | Leu >Gln |
| c.239A >T | ATC >AAC | Asp >Val |

Continued on next page

Table C.1 – continued from previous page

| **Mutation (HGVS)** | **Trinucleotide Context** | **Result** |
|---|---|---|
| c.293A >G | ATC >ACC | Asp >Gly |
| c.295T >G | TTT >TGT | Phe >Val |
| c.310A >G | CTC >CCC | Ser >Arg |
| c.325C >T | CCA >CTA | Gln >stop |
| c.329C >T | TCA >TTA | Ser >Leu |
| c.368C >G | TCA >TGA | Ser >stop |
| c.370A >C | GTT >GGT | Thr >Pro |
| c.371C >G | ACT >AGT | Thr >Ser |
| c.374T >C | TTA >TCA | Leu >Ser |
| c.379G >T | GTC >GAC | Gly >Val |
| c.389T >A | GTC >GAC | Val >Asp |
| c.392T >C | TTG >TCG | Leu >Ser |
| c.392T >G | TTG >TGG | Leu >Trp |
| c.395T >G | ATT >AGT | Ile >Ser |
| c.395T >C | ATT >ACT | Ile >Thr |
| c.396T >G | TTG >TGG | Ile >Met |
| c.397G >A | ACA >ATA | Val >Met |
| c.404A >T | ATC >AAC | Asp >Val |
| c.404A >G | ATC >ACC | Asp >Gly |
| c.407T >A | ATA >AAA | Ile >Lys |
| c.407T >C | ATA >ACA | Ile >Thr |
| c.410T >C | ATT >ACT | Ile >Thr |
| c.415A >C | GTG >GGG | Thr >Pro |

Continued on next page

Table C.1 – continued from previous page

| **Mutation (HGVS)** | **Trinucleotide Context** | **Result** |
|---|---|---|
| c.418G >C | CCA >CGA | Gly >Arg |
| c.419G >A | GCC >GTC | Gly >Asp |
| c.430C >T | GCA >GTA | Gln >stop |
| c.439C >T | GCT >GTA | Leu >Phe |
| c.440T >C | CTT >CCT | Leu >Pro |
| c.449T >G | GTC >GGC | Val >Gly |
| c.454C >T | GCA >GTA | Gln >stop |
| c.463C >T | TCC >TTC | Pro >Ser |
| c.472G >T | ACC >AAC | Val >Phe |
| c.473T >G | GTC >GGC | Val >Gly |
| c.475A >G | TTG >TCG | Lys >Glu |
| c.476A >C | CTT >CGT | Lys >Thr |
| c.481G >A | GCG >GTG | Ala >Thr |
| c.481G >T | GCG >GAG | Ala >Ser |
| c.482C >A | GCA >GAA | Ala >Glu |
| c.484A >C | GCT >GGT | Ser >Arg |
| c.485G >A | GCT >GTT | Ser >Asn |
| c.485G >C | TTG >TGG | Ser >Thr |
| c.486C >G | GCT >GGT | Ser >Arg |
| c.488T >C | TTG >TCG | Leu >Ser |
| c.500G >T | CCT >CAT | Arg >Met |
| c.500G >C | CCT >CGT | Arg >Thr |
| c.503C >T | ACC >ATC | Thr >Ile |

Continued on next page

Table C.1 – continued from previous page

| **Mutation (HGVS)** | **Trinucleotide Context** | **Result** |
|---|---|---|
| c.508C >T | ACG >ATG | Arg >stop |
| c.526C >T | GCC >GTC | Pro >Ser |
| c.527C >T | CCA >CTA | Pro >Leu |
| c.529G >T | TCT >TAT | Asp >Tyr |
| c.530A >T | GTC >GAC | Asp >Val |
| c.539G >A | TCC >TTC | Gly >Glu |
| c.541T >C | ATT >ACT | Phe >Leu |
| c.548T >C | ATT >ACT | Ile >Thr |
| c.550C >A | TCC >TAC | Pro >Thr |
| c.554A >T | GTC >GAC | Asp >Val |
| c.554A >G | GTC >GCC | Asp >Gly |
| c.558G >C | ACT >AGT | Lys >Asn |
| c.563T >C | GTT >GCT | Val >Ala |
| c.565G >T | ACA >AAA | Val >Leu |
| c.568G >C | CCT >CGT | Gly >Arg |
| c.569G >A | TCC >TTC | Gly >Glu |
| c.574G >A | GCA >GTA | Ala >Thr |
| c.575C >T | GCC >GTC | Ala >Val |
| c.577C >T | CCT >CTT | Leu >Phe |
| c.579T >G | TTG >TGG | Leu >Arg |
| c.580G >A | TCA >TTA | Asp >Asn |
| c.580G >T | TCA >TAA | Asp >Tyr |
| c.580G >C | TCA >TGA | Asp >His |

Continued on next page

Table C.1 – continued from previous page

| **Mutation (HGVS)** | **Trinucleotide Context** | **Result** |
|---|---|---|
| c.582C >G | ACT >AGT | Asp >Glu |
| c.584A >C | ATA >AGA | Tyr >Ser |
| c.584A >G | ATA >ACA | Tyr >Cys |
| c.586A >G | ATA >ACA | Asn >Asp |
| c.590A >T | TTC >TAC | Glu >Val |
| c.595T >G | CTT >CGT | Phe >Val |
| c.596T >G | TTC >TGC | Phe >Cys |
| c.597C >G | TCA >TTA | Phe >Leu |
| c.599G >C | TCA >TGA | Arg >Thr |
| c.601G >A | TCA >TTA | Asp >Asn |
| c.601G >T | TCC >TAC | Asp >Tyr |
| c.602A >T | ATC >AAC | Asp >Val |
| c.602A >G | ATC >ACC | Asp >Gly |
| c.606G >T | TCA >TAA | Leu >Phe |
| c.610C >T | TCA >TTA | His >Tyr |
| c.610C >G | TCA >TGA | His >Asp |
| c.611A >T | TGT >TAT | His >Leu |
| c.617G >A | ACA >ATA | Cys >Tyr |
| c.618T >A | GTG >GAG | Cys >stop |
| c.634G >T | CCA >CAA | Gly >stop |
| c.635G >A | TCC >TTC | Gly >Glu |

# Appendix D

# Nomenclature of *Pole*$^{\mathbf{L424V}}$ mutant mice

| Genotype | Explanation |
|---|---|
| *Pole*$^{\text{fl(L424V)/+}}$ | Heterozygous floxed *POLE* L424V allele |
| *Pole*$^{\text{L424V/+}}$ | Heterozygous recombined germline *POLE* L424V knockiin allele |
| *PgkCre*,*Pole*$^{\text{fl(L424V)/+}}$ | Heterozygous floxed *POLE* L424V allele crossed with *Pgk*-Cre (Tg(Pgk1-cre)1Lni (MGI:2178050)) (231) |
| *Lgr5ERT2Cre*,*Pole*$^{\text{fl(L424V)/+}}$ | Heterozygous floxed *POLE* L424V allele crossed with *Lgr5*$^{\text{tm1(cre/ERT2)Cle}}$ (MGI:3764660) (232) |
| *PRCre*,Pole$^{\text{fl(L424V)/+}}$ | Heterozygous floxed *POLE* L424V allele crossed with Pgr$^{\text{tm2(cre)Lyd}}$ (MGI:3576366) (280) |
| *Apc*$^{\text{Min}}$,*Pole*$^{\text{L424V/+}}$ | Heterozygous recombined germline *POLE* L424V knock-in allele crossed with *Apc*$^{\text{Min}}$ (MGI:1856318) (233)) |

**Table D.1:** Nomenclature of *Pole*-mutant mice used in this study.

# Appendix E

# Gross morphological and histological analysis of constitutive *PgkCre*, *Pole*$^{\text{fl(L424V)/+}}$ mice

| Mouse ID | Genotype | Age (Weeks) | Reason for sacrifice | Necropsy results |
|---|---|---|---|---|
| TLPW2.1c | *PgkCre*,*Pole*$^{\text{fl(L424V)/+}}$ | 52 | Time-point | Normal upon necropsy |
| TLPW2.2d | *PgkCre*,*Pole*$^{\text{fl(L424V)/+}}$ | 52.4 | Time-point | Normal upon necropsy |
| TLPW3.1c | *PgkCre*,*Pole*$^{\text{fl(L424V)/+}}$ | 53 | Time-point | Normal upon necropsy |
| TLPW2.1a | *PgkCre* | 52 | Time-point | Normal upon necropsy |
| TLPW2.1b | *PgkCre* | 52 | Time-point | Normal upon necropsy |
| TLPW2.1d | *PgkCre* | 52 | Time-point | Normal upon necropsy |
| TLPW2.1f | *PgkCre* | 52 | Time-point | Normal upon necropsy |
| TLPW2.2f | *PgkCre* | 52.4 | Time-point | Normal upon necropsy |
| TLPW2.1e | *Pole*$^{\text{fl(L424V)/+}}$ | 52 | Time-point | Normal upon necropsy |
| TLPW2.1g | *Pole*$^{\text{fl(L424V)/+}}$ | 52 | Time-point | Normal upon necropsy |
| TLPW2.2e | *Pole*$^{\text{fl(L424V)/+}}$ | 52.4 | Time-point | Normal upon necropsy |
| TLPW2.3d | *PgkCre*,*Pole*$^{\text{fl(L424V)/+}}$ | 81.4 | Mass on neck | Stomach and Neck tumours |
| TLPW3.1h | *PgkCre*,*Pole*$^{\text{fl(L424V)/+}}$ | 81.7 | Time-point | Normal upon necropsy |
| TLPW2.3c | *Pole*$^{\text{fl(L424V)/+}}$ | 81.4 | Time-point | Normal upon necropsy |
| TLPW3.1g | *Pole*$^{\text{fl(L424V)/+}}$ | 81.7 | Time-point | Normal upon necropsy |
| TLPW2.2a | *PgkCre*,*Pole*$^{\text{fl(L424V)/+}}$ | 17.6 | Cachexia | Intrathoracic tumour |

**Table E.1:** Reason for sacrifice, age of sacrifice and results from necropsy from *PgkCre*,*Pole*$^{\text{fl(L424V)/+}}$ (experimental), *PgkCre* (control), and *Pole*$^{\text{fl(L424V)/+}}$ (control) mice.

| Mouse ID | Genotype | Age (Weeks) | Tissue | Histopathology results |
|---|---|---|---|---|
| TLPW2.2d | *PgkCre*,$Pole^{fl(L424V)/+}$ | 52.4 | Intestine | Normal |
| TLPW2.2d | *PgkCre*,$Pole^{fl(L424V)/+}$ | 52.4 | Lung | Normal |
| TLPW2.2d | *PgkCre*,$Pole^{fl(L424V)/+}$ | 52.4 | Brain | Normal |
| TLPW3.1c | *PgkCre*,$Pole^{fl(L424V)/+}$ | 53 | Intestine | Normal |
| TLPW3.1c | *PgkCre*,$Pole^{fl(L424V)/+}$ | 53 | Lung | Normal |
| TLPW3.1c | *PgkCre*,$Pole^{fl(L424V)/+}$ | 53 | Brain | Normal |
| TLPW2.1c | *PgkCre*,$Pole^{fl(L424V)/+}$ | 52 | Intestine | Normal |
| TLPW2.1c | *PgkCre*,$Pole^{fl(L424V)/+}$ | 52 | Lung | Normal |
| TLPW2.1c | *PgkCre*,$Pole^{fl(L424V)/+}$ | 52 | Brain | Normal |
| TLPW2.1a | *PgkCre* | 52 | Intestine | Normal |
| TLPW2.1a | *PgkCre* | 52 | Lung | Normal |
| TLPW2.1a | *PgkCre* | 52 | Brain | Normal |
| TLPW2.1b | *PgkCre* | 52 | Intestine | Normal |
| TLPW2.1b | *PgkCre* | 52 | Lung | Normal |
| TLPW2.1b | *PgkCre* | 52 | Brain | Normal |
| TLPW2.1d | *PgkCre* | 52 | Intestine | Normal |
| TLPW2.1d | *PgkCre* | 52 | Lung | Normal |
| TLPW2.1d | *PgkCre* | 52 | Brain | Normal |
| TLPW2.1f | *PgkCre* | 52 | Intestine | Normal |
| TLPW2.1f | *PgkCre* | 52 | Lung | Normal |
| TLPW2.1f | *PgkCre* | 52 | Brain | Normal |
| TLPW2.1f | *PgkCre* | 52 | Uterus | Normal |
| TLPW2.2f | *PgkCre* | 52.4 | Intestine | Normal |
| TLPW2.2f | *PgkCre* | 52.4 | Lung | Normal |
| TLPW2.2f | *PgkCre* | 52.4 | Uterus | Normal |
| TLPW2.2f | *PgkCre* | 52.4 | Brain | Normal |
| TLPW2.1e | $Pole^{fl(L424V)/+}$ | 52 | Intestine | Normal |
| TLPW2.1e | $Pole^{fl(L424V)/+}$ | 52 | Lung | Normal |
| TLPW2.1e | $Pole^{fl(L424V)/+}$ | 52 | Uterus | Normal |
| TLPW2.1e | $Pole^{fl(L424V)/+}$ | 52 | Brain | Normal |
| TLPW2.1g | $Pole^{fl(L424V)/+}$ | 52 | Intestine | Normal |
| TLPW2.1g | $Pole^{fl(L424V)/+}$ | 52 | Lung | Normal |
| TLPW2.1g | $Pole^{fl(L424V)/+}$ | 52 | Uterus | Normal |
| TLPW2.1g | $Pole^{fl(L424V)/+}$ | 52 | Brain | Normal |
| TLPW2.2e | $Pole^{fl(L424V)/+}$ | 52.4 | Intestine | Normal |
| TLPW2.2e | $Pole^{fl(L424V)/+}$ | 52.4 | Lung | Normal |
| TLPW2.2e | $Pole^{fl(L424V)/+}$ | 52.4 | Uterus | Normal |
| TLPW2.2e | $Pole^{fl(L424V)/+}$ | 52.4 | Brain | Normal |
| TLPW3.1h | *PgkCre*,$Pole^{fl(L424V)/+}$ | 84.7 | Intestine | Normal |
| TLPW3.1h | *PgkCre*,$Pole^{fl(L424V)/+}$ | 84.7 | Uterus | Normal |
| TLPW3.1h | *PgkCre*,$Pole^{fl(L424V)/+}$ | 84.7 | Brain | Normal |
| TLPW3.1h | *PgkCre*,$Pole^{fl(L424V)/+}$ | 84.7 | Lung | Normal |
| TLPW2.3d | *PgkCre*,$Pole^{fl(L424V)/+}$ | 81.4 | Neck tumour | lymphoma |
| TLPW2.3d | *PgkCre*,$Pole^{fl(L424V)/+}$ | 81.4 | Abdominal tumour | lymphoma |
| TLPW2.3c | $Pole^{fl(L424V)/+}$ | 84.4 | Intestine | Normal |
| TLPW2.3c | $Pole^{fl(L424V)/+}$ | 84.4 | Lung | Normal |
| TLPW2.3c | $Pole^{fl(L424V)/+}$ | 84.4 | Uterus | Normal |
| TLPW2.3c | $Pole^{fl(L424V)/+}$ | 84.4 | Brain | Normal |
| TLPW3.1g | $Pole^{fl(L424V)/+}$ | 84.7 | Intestine | Normal |
| TLPW3.1g | $Pole^{fl(L424V)/+}$ | 84.7 | Lung | Normal |
| TLPW3.1g | $Pole^{fl(L424V)/+}$ | 84.7 | Uterus | Normal |
| TLPW3.1g | $Pole^{fl(L424V)/+}$ | 84.7 | Brain | Normal |
| TLPW2.2a | *PgkCre*,$Pole^{fl(L424V)/+}$ | 17.6 | Intrathoracic tumour | T-cell lymphoblastic lymphoma |

**Table E.2:** *PgkCre*,$Pole^{fl(L424V)/+}$, $Pole^{fl(L424V)/+}$ and *PgkCre* mouse histopathology results from intestine, uterus, brain, lung and tumour tissue sections.

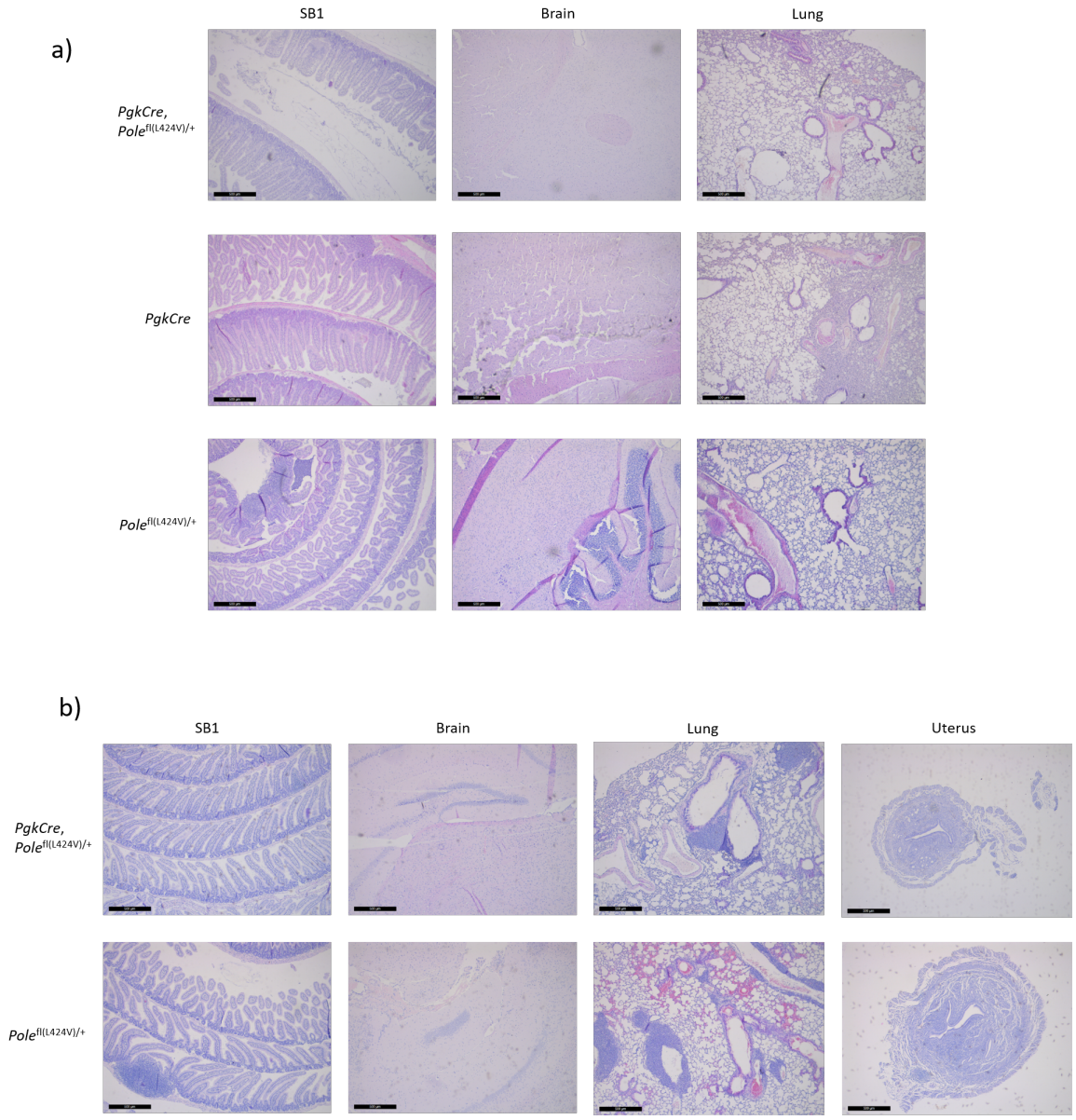



**Figure E.1:** Representative images of H&E stained FFPE tissue sections from experimental (*PgkCre*,*Pole*$^{fl(L424V)/+}$) and control (*Pole*$^{fl(L424V)/+}$ and *PgkCre*) mice. a) Representative images of the small intestine (SB1), brain and lung from experimental and control animals sacrificed at 52 weeks. b) Representative images of the small intestine (SB1), brain lung and uterus from experimental and control animals sacrificed at 81 weeks. Scale bars represent 500$\mu$m in all images.

# Appendix F

# Gross morphological and histological analysis of inducible *Lgr5CreERT2*, *Pole*$^{\text{fl(L424V)/+}}$ mice

| Mouse ID | Genotype | Age (Weeks) | Reason for sacrifice | Necropsy results |
|---|---|---|---|---|
| THNW3.4a | *Lgr5CreERT2,Pole*$^{\text{fl(L424V)/+}}$ | 26.4 | Time-point | Normal upon necropsy |
| THNW3.4b | *Lgr5CreERT2,Pole*$^{\text{fl(L424V)/+}}$ | 26.4 | Time-point | Normal upon necropsy |
| THNW3.4c | *Lgr5CreERT2,Pole*$^{\text{fl(L424V)/+}}$ | 26.4 | Time-point | Normal upon necropsy |
| THNW3.4d | *Lgr5CreERT2,Pole*$^{\text{fl(L424V)/+}}$ | 26.4 | Time-point | Normal upon necropsy |
| THNW3.4j | *Lgr5CreERT2,Pole*$^{\text{fl(L424V)/+}}$ | 26.4 | Time-point | Normal upon necropsy |
| THNW1.7a | *Lgr5CreERT2* | 26.7 | Time-point | Normal upon necropsy |
| THNW1.7e | *Lgr5CreERT2* | 26.7 | Time-point | Normal upon necropsy |
| THNW1.7f | *Lgr5CreERT2* | 26.7 | Time-point | Normal upon necropsy |
| THNW3.4e | *Lgr5CreERT2* | 26.4 | Time-point | Normal upon necropsy |
| THNW3.4g | *Lgr5CreERT2* | 26.4 | Time-point | Normal upon necropsy |
| THNW1.1e | *Lgr5CreERT2,Pole*$^{\text{fl(L424V)/+}}$ | 51.9 | Time-point | Normal upon necropsy |
| THNW2.2e | *Lgr5CreERT2,Pole*$^{\text{fl(L424V)/+}}$ | 52 | Time-point | Splenomegaly |
| THNW2.2f | *Lgr5CreERT2,Pole*$^{\text{fl(L424V)/+}}$ | 52 | Time-point | Normal upon necropsy |
| THNW3.2b | *Lgr5CreERT2* | 52 | Time-point | Normal upon necropsy |
| THNW3.2e | *Lgr5CreERT2* | 52 | Time-point | Normal upon necropsy |
| THNW1.8b | *Lgr5CreERT2* | 52.3 | Time-point | Normal upon necropsy |
| THNW1.1a | *Pole*$^{\text{fl(L424V)/+}}$ | 51.9 | Time-point | Normal upon necropsy |
| THNW1.1b | *Pole*$^{\text{fl(L424V)/+}}$ | 51.9 | Time-point | Normal upon necropsy |
| THNW3.2g | *Pole*$^{\text{fl(L424V)/+}}$ | 52 | Time-point | Normal upon necropsy |
| THNW1.2b | *Lgr5CreERT2,Pole*$^{\text{fl(L424V)/+}}$ | 87.1 | Time-point | Normal upon necropsy |
| THNW1.2d | *Lgr5CreERT2,Pole*$^{\text{fl(L424V)/+}}$ | 85.1 | Time-point | Normal upon necropsy |
| THNW1.2e | *Lgr5CreERT2,Pole*$^{\text{fl(L424V)/+}}$ | 87.1 | Time-point | Normal upon necropsy |
| THNW2.1f | *Lgr5CreERT2,Pole*$^{\text{fl(L424V)/+}}$ | 86.4 | Time-point | Normal upon necropsy |
| THNW2.2i | *Lgr5CreERT2,Pole*$^{\text{fl(L424V)/+}}$ | 89.3 | Time-point | Normal upon necropsy |
| THNW2.2f | *Lgr5CreERT2* | 86.1 | Time-point | Normal upon necropsy |
| THNW2.2b | *Lgr5CreERT2* | 89.3 | Time-point | Normal upon necropsy |
| THNW2.2c | *Lgr5CreERT2* | 89.3 | Time-point | Liver tumour, all else normal |
| THNW3.1h | *Pole*$^{\text{fl(L424V)/+}}$ | 90.6 | Time-point | Splenomegaly, all else normal |
| THNW3.2a | *Pole*$^{\text{fl(L424V)/+}}$ | 85.3 | Time-point | Normal upon necropsy |
| THNW3.2d | *Pole*$^{\text{fl(L424V)/+}}$ | 85.3 | Time-point | Normal upon necropsy |
| THNW1.3b | *Pole*$^{\text{fl(L424V)/+}}$ | 91.1 | Time-point | Normal upon necropsy |
| THNW3.1g | *Pole*$^{\text{fl(L424V)/+}}$ | 88.3 | Time-point | Normal upon necropsy |
| THNW3.4f | *Lgr5CreERT2* | 18.4 | Cachexia | External tumour left inguinal region. Appears cystic on dissection |

**Table F.1:** Reason for sacrifice, age of sacrifice and results from necropsy from *Lgr5CreERT2*, *Pole*$^{\text{fl(L424V)/+}}$, *Lgr5*-Cre and *Pole*$^{\text{fl(L424V)/+}}$ mice.

| Mouse ID | Genotype | Age (Weeks) | Tissue | Histopathology results |
|---|---|---|---|---|
| THNW3.4a | *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$ | 26.4 | Intestine | Normal |
| THNW3.4b | *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$ | 26.4 | Intestine | Normal |
| THNW3.4c | *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$ | 26.4 | Intestine | Normal |
| THNW3.4d | *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$ | 26.4 | Intestine | Normal |
| THNW3.4j | *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$ | 26.4 | Intestine | Normal |
| THNW1.7a | *Lgr5CreERT2* | 26.7 | Intestine | Normal |
| THNW1.7e | *Lgr5CreERT2* | 26.7 | Intestine | Normal |
| THNW1.7f | *Lgr5CreERT2* | 26.7 | Intestine | Normal |
| THNW3.4e | *Lgr5CreERT2* | 26.4 | Intestine | Normal |
| THNW3.4g | *Lgr5CreERT2* | 26.4 | Intestine | Normal |
| THNW1.1e | *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$ | 51.9 | Intestine | Normal |
| THNW2.2e | *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$ | 52 | Intestine | Normal |
| THNW2.2f | *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$ | 52 | Intestine | Normal |
| THNW3.2b | *Lgr5CreERT2* | 52 | Intestine | Normal |
| THNW3.2e | *Lgr5CreERT2* | 52 | Intestine | Normal |
| THNW1.8b | *Lgr5CreERT2* | 52.3 | Intestine | Normal |
| THNW1.1a | *Pole*$^{fl(L424V)/+}$ | 51.9 | Intestine | Normal |
| THNW1.1b | *Pole*$^{fl(L424V)/+}$ | 51.9 | Intestine | Normal |
| THNW3.2g | *Pole*$^{fl(L424V)/+}$ | 52 | Intestine | Normal |
| THNW1.2b | *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$ | 87.1 | Intestine | Normal |
| THNW1.2d | *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$ | 85.1 | Intestine | Normal |
| THNW1.2e | *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$ | 87.1 | Intestine | Normal |
| THNW2.2i | *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$ | 89.3 | Intestine | Normal |
| THNW2.1f | *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$ | 86.4 | Intestine | Normal |
| THNW1.2f | *Lgr5CreERT2* | 86.1 | Intestine | Normal |
| THNW2.2c | *Lgr5CreERT2* | 89.3 | Liver tumour | Lymphocytic infiltrate |
| THNW2.2c | *Lgr5CreERT2* | 89.3 | Intestine | Normal |
| THNW2.2b | *Lgr5CreERT2* | 89.3 | Intestine | Normal |
| THNW3.2a | *Pole*$^{fl(L424V)/+}$ | 85.3 | Intestine | Normal |
| THNW3.2d | *Pole*$^{fl(L424V)/+}$ | 85.3 | Intestine | Normal |
| THNW3.1g | *Pole*$^{fl(L424V)/+}$ | 88.3 | Intestine | Normal |
| THNW3.1h | *Pole*$^{fl(L424V)/+}$ | 90.6 | Intestine | Normal |
| THNW1.3b | *Pole*$^{fl(L424V)/+}$ | 91.1 | Intestine | Normal |
| THNW3.4f | *Lgr5CreERT2* | 18.4 | Abdominal tumour | Normal |
| THNW3.4f | *Lgr5CreERT2* | 18.4 | Intestine | Normal |

**Table F.2:** Intestine and tumour histopathology results from *Lgr5CreERT2,Pole*$^{fl(L424V)/+}$, *Lgr5CreERT2* and *Pole*$^{fl(L424V)/+}$ mice.

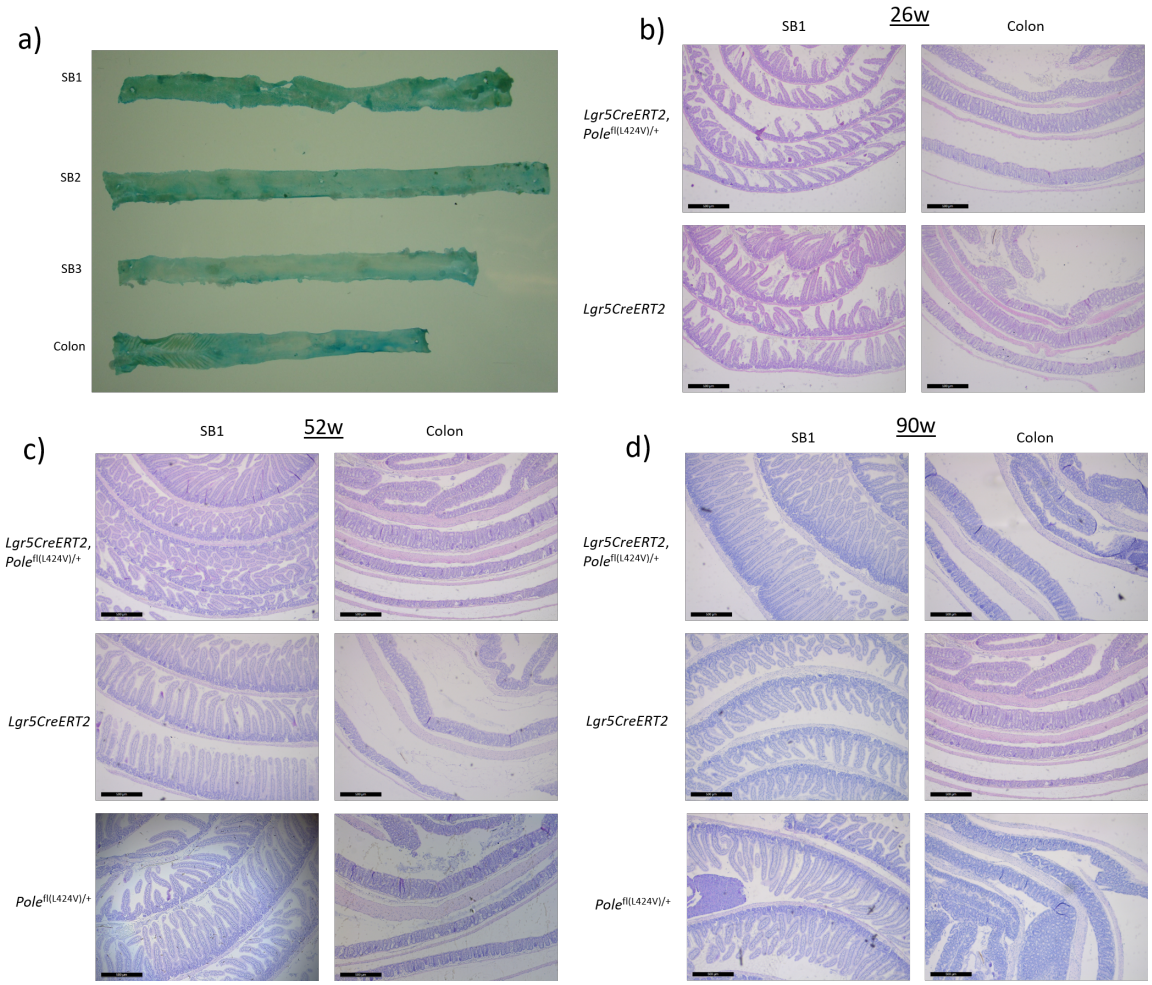



**Figure F.1:** Representative images of methylene blue stained intestine and H&E stained intestinal FFPE tissue sections from experimental *Lgr5CreERT2*,*Pole*$^{\mathrm{fl(L424V)/+}}$ and control *Lgr5CreERT2* and *Pole*$^{\mathrm{fl(L424V)/+}}$ mice. a) Methylene blue stained small intestine and colon. b) Experimental and control H&E-stained intestinal sections from 26 week old mice. c) Experimental and control H&E-stained intestinal sections from 52 week old mice. d) Experimental and control H&E-stained intestinal sections from 85-90 week old mice. No intestinal polyps or microadenomas were observed. Scale bars represent 500$\mu$m.

# Appendix G

# Gross morphological and histological analysis of *PRCre,Pole*$^{\text{fl(L424V)/+}}$ mice

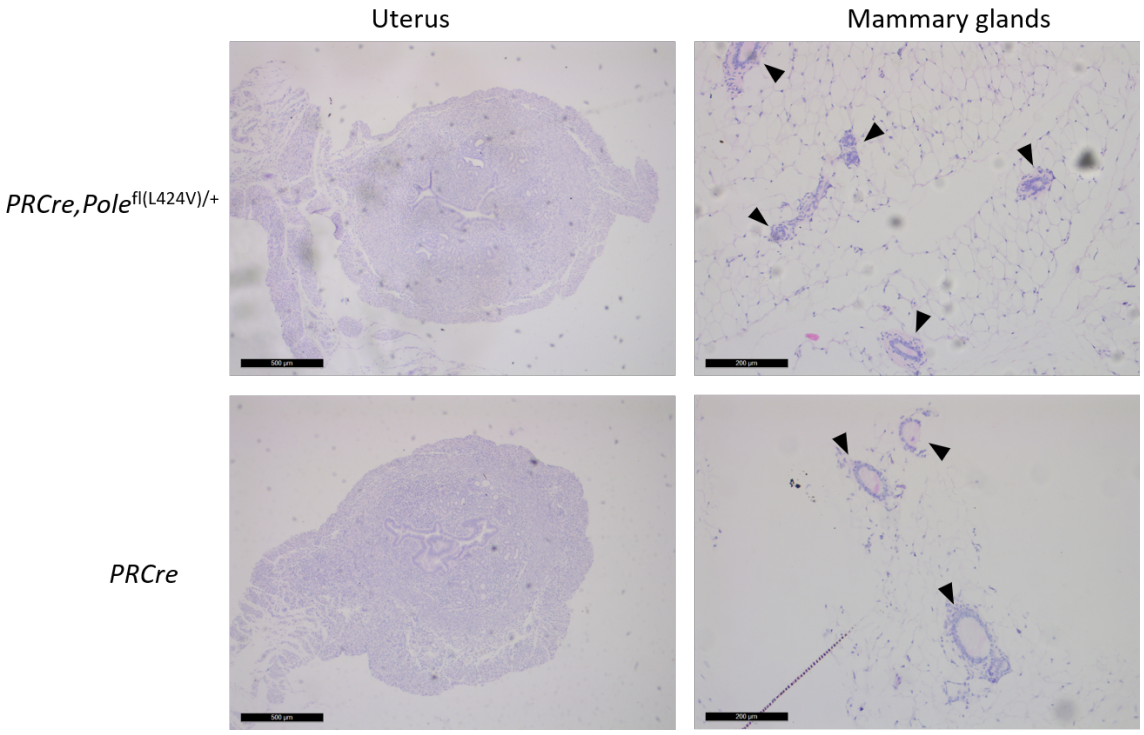



**Figure G.1:** Representative images of H&E stained uterus and mammary gland sections from experimental *PRCre,Pole*$^{\text{fl(L424V)/+}}$ and control *PRCre* mice at at 24 weeks. All tissue was normal on analysis. Scale bars represent 500$\mu$m for the uterus and 200$\mu$m for mammary glands.

| Mouse ID | Genotype | Age (Weeks) | Reason for sacrifice | Necropsy results |
|---|---|---|---|---|
| TPBW5.1e | *PRCre,Pole*$^{fl(L424V)/+}$ | 24 | Time-point | Normal upon necropsy |
| TPBW5.1f | *PRCre,Pole*$^{fl(L424V)/+}$ | 24 | Time-point | Normal upon necropsy |
| TPBW5.1h | *PRCre,Pole*$^{fl(L424V)/+}$ | 24 | Time-point | Normal upon necropsy |
| TPBW3.5g | *PRCre,Pole*$^{fl(L424V)/+}$ | 24 | Time-point | Normal upon necropsy |
| TPBW1.8d | *PRCre,Pole*$^{fl(L424V)/+}$ | 24 | Time-point | Normal upon necropsy |
| TPBW5.1d | *PRCre* | 24 | Time-point | Normal upon necropsy |
| TPBW5.1g | *PRCre* | 24 | Time-point | Normal upon necropsy |
| TPBW5.1i | *PRCre* | 24 | Time-point | Normal upon necropsy |
| TPBW3.5h | *PRCre* | 24 | Time-point | Normal upon necropsy |
| TPBW1.8c | *PRCre* | 24 | Time-point | Normal upon necropsy |
| TPBW2.1c | *PRCre,Pole*$^{fl(L424V)/+}$ | 52 | Time-point | Mammary tumour |
| TPBW2.1e | *PRCre,Pole*$^{fl(L424V)/+}$ | 52 | Time-point | Enlarged and bumpy uterus |
| TPBW2.1g | *PRCre,Pole*$^{fl(L424V)/+}$ | 52 | Time-point | Normal upon necropsy |
| TPBW2.1d | *PRCre* | 52 | Time-point | Normal upon necropsy |
| TPBW1.1e | *PRCre* | 52.1 | Time-point | Normal upon necropsy |
| TPBW1.1f | *PRCre* | 52.1 | Time-point | Splenomegaly |
| TPBW6.2f | *PRCre,Pole*$^{fl(L424V)/+}$ | 77.1 | Time-point | Normal upon necropsy |
| TPBW1.6f | *PRCre,Pole*$^{fl(L424V)/+}$ | 79.4 | Time-point | Normal upon necropsy |
| TPBW1.6g | *PRCre,Pole*$^{fl(L424V)/+}$ | 79.4 | Time-point | Splenomegaly |
| TPBW3.3j | *PRCre,Pole*$^{fl(L424V)/+}$ | 81.9 | Time-point | Normal upon necropsy |
| TPBW3.3m | *PRCre,Pole*$^{fl(L424V)/+}$ | 81.9 | Time-point | Normal upon necropsy |
| TPBW3.3k | *PRCre,Pole*$^{fl(L424V)/+}$ | 81.9 | Time-point | Normal upon necropsy |
| TPBW3.3l | *PRCre,Pole*$^{fl(L424V)/+}$ | 81.9 | Time-point | Normal upon necropsy |
| TPBW3.1a | *PRCre,Pole*$^{fl(L424V)/+}$ | 77.7 | Cachexia | Splenomegaly and liver tumour |
| TPBW3.1d | *PRCre,Pole*$^{fl(L424V)/+}$ | 78.7 | Time-point | Enlarged uterus and Splenomegaly |
| TPBW3.1e | *PRCre,Pole*$^{fl(L424V)/+}$ | 78.7 | Time-point | Normal upon necropsy |
| TPBW3.1f | *PRCre,Pole*$^{fl(L424V)/+}$ | 78.7 | Time-point | Enlarged uterus and Splenomegaly |
| TPBW6.2g | *PRCre* | 77.1 | Time-point | Normal upon necropsy |
| TPBW6.2e | *PRCre* | 77.1 | Time-point | Normal upon necropsy |
| TPBW6.1f | *PRCre* | 80.4 | Time-point | Normal upon necropsy |
| TPBW1.4a | *PRCre* | 81.9 | Time-point | Normal upon necropsy |
| TPBW3.3i | *PRCre* | 81.9 | Time-point | Normal upon necropsy |
| TPBW3.1b | *PRCre* | 78.7 | Time-point | Splenomegaly and Mottled spleen |
| TPBW3.1c | *PRCre* | 78.7 | Time-point | Normal upon necropsy |
| TPBW3.1g | *PRCre* | 78.7 | Time-point | Normal upon necropsy |
| TPBW1.3b | *PRCre* | 79.3 | Time-point | Normal upon necropsy |

**Table G.1:** Reason for sacrifice, age of sacrifice and results from necropsy from *PRCre,Pole*$^{fl(L424V)/+}$ and *PRCre* mice

| Mouse ID | Genotype | Age (Weeks) | Tissue | Histopathology results |
|---|---|---|---|---|
| TPBW5.1e | *PRCre,Pole*$^{fl(L424V)/+}$ | 24 | Uterus & Ovary | Normal |
| TPBW5.1e | *PRCre,Pole*$^{fl(L424V)/+}$ | 24 | Mammary gland | Normal |
| TPBW5.1f | *PRCre,Pole*$^{fl(L424V)/+}$ | 24 | Uterus & Ovary | Normal |
| TPBW5.1f | *PRCre,Pole*$^{fl(L424V)/+}$ | 24 | Mammary gland | Normal |
| TPBW5.1h | *PRCre,Pole*$^{fl(L424V)/+}$ | 24 | Uterus & Ovary | Normal |
| TPBW5.1h | *PRCre,Pole*$^{fl(L424V)/+}$ | 24 | Mammary gland | Normal |
| TPBW5.1h | *PRCre,Pole*$^{fl(L424V)/+}$ | 24 | sternum lump | Normal |
| TPBW3.5g | *PRCre,Pole*$^{fl(L424V)/+}$ | 24 | Uterus & Ovary | Normal |
| TPBW3.5g | *PRCre,Pole*$^{fl(L424V)/+}$ | 24 | Mammary gland | Normal |
| TPBW1.8d | *PRCre,Pole*$^{fl(L424V)/+}$ | 24 | Uterus & Ovary | Normal |
| TPBW1.8d | *PRCre,Pole*$^{fl(L424V)/+}$ | 24 | Mammary gland | Normal |
| TPBW5.1d | *PRCre* | 24 | Uterus & Ovary | Normal |
| TPBW5.1d | *PRCre* | 24 | Mammary gland | Normal |
| TPBW5.1g | *PRCre* | 24 | Uterus & Ovary | Normal |
| TPBW5.1g | *PRCre* | 24 | Mammary gland | Normal |
| TPBW5.1i | *PRCre* | 24 | Uterus & Ovary | Normal |
| TPBW5.1i | *PRCre* | 24 | Mammary gland | Normal |
| TPBW3.5h | *PRCre* | 24 | Uterus & Ovary | Normal |
| TPBW3.5h | *PRCre* | 24 | Mammary gland | Normal |
| TPBW1.8c | *PRCre* | 24 | Uterus & Ovary | Normal |
| TPBW1.8c | *PRCre* | 24 | Mammary gland | Normal |
| TPBW2.1c | *PRCre,Pole*$^{fl(L424V)/+}$ | 52 | Uterus & Ovary | Normal |
| TPBW2.1c | *PRCre,Pole*$^{fl(L424V)/+}$ | 52 | Mammary gland lump | Normal |
| TPBW2.1e | *PRCre,Pole*$^{fl(L424V)/+}$ | 52 | Uterus & Ovary | Hydrouterus |
| TPBW2.1e | *PRCre,Pole*$^{fl(L424V)/+}$ | 52 | Mammary gland | Normal |
| TPBW2.1g | *PRCre,Pole*$^{fl(L424V)/+}$ | 52 | Uterus & Ovary | Normal |
| TPBW2.1g | *PRCre,Pole*$^{fl(L424V)/+}$ | 52 | Mammary gland | Normal |
| TPBW2.1d | *PRCre* | 52 | Uterus & Ovary | Normal |
| TPBW1.1f | *PRCre* | 52.1 | Uterus & Ovary | Normal |
| TPBW1.1f | *PRCre* | 52.1 | Mammary gland | Normal |
| TPBW6.2f | *PRCre,Pole*$^{fl(L424V)/+}$ | 77.1 | Uterus & Ovary | Normal |
| TPBW6.2f | *PRCre,Pole*$^{fl(L424V)/+}$ | 77.1 | Mammary gland | Normal |
| TPBW1.6f | *PRCre,Pole*$^{fl(L424V)/+}$ | 79.4 | Uterus & Ovary | Normal |
| TPBW1.6f | *PRCre,Pole*$^{fl(L424V)/+}$ | 79.4 | Mammary gland | Normal |
| TPBW1.6g | *PRCre,Pole*$^{fl(L424V)/+}$ | 79.4 | Uterus & Ovary | Normal |
| TPBW1.6g | *PRCre,Pole*$^{fl(L424V)/+}$ | 79.4 | Mammary gland | Normal |
| TPBW1.6g | *PRCre,Pole*$^{fl(L424V)/+}$ | 79.4 | Spleen | Normal |
| TPBW3.3j | *PRCre,Pole*$^{fl(L424V)/+}$ | 81.9 | Uterus & Ovary | Normal |
| TPBW3.3j | *PRCre,Pole*$^{fl(L424V)/+}$ | 81.9 | Mammary gland | Normal |
| TPBW3.3m | *PRCre,Pole*$^{fl(L424V)/+}$ | 81.9 | Uterus & Ovary | Normal |
| TPBW3.3m | *PRCre,Pole*$^{fl(L424V)/+}$ | 82.9 | Mammary gland | Normal |
| TPBW3.3k | *PRCre,Pole*$^{fl(L424V)/+}$ | 81.9 | Uterus & Ovary | Ovarian haemorragic lesion |
| TPBW3.3k | *PRCre,Pole*$^{fl(L424V)/+}$ | 82.9 | Mammary gland | Normal |
| TPBW3.3l | *PRCre,Pole*$^{fl(L424V)/+}$ | 81.9 | Uterus & Ovary | Normal |
| TPBW3.1a | *PRCre,Pole*$^{fl(L424V)/+}$ | 77.7 | Uterus & Ovary | Lymphoma |
| TPBW3.1a | *PRCre,Pole*$^{fl(L424V)/+}$ | 77.7 | Mammary gland | Lymphoma in Mammary gland |
| TPBW3.1a | *PRCre,Pole*$^{fl(L424V)/+}$ | 77.7 | Lung tumour | high lymphocytic infiltrate |
| TPBW3.1a | *PRCre,Pole*$^{fl(L424V)/+}$ | 77.7 | Lung | high lymphocytic infiltrate |
| TPBW3.1d | *PRCre,Pole*$^{fl(L424V)/+}$ | 78.7 | Uterus & Ovary | Hydrouterus |
| TPBW3.1d | *PRCre,Pole*$^{fl(L424V)/+}$ | 78.7 | Mammary gland | Normal |
| TPBW3.1e | *PRCre,Pole*$^{fl(L424V)/+}$ | 78.7 | Uterus & Ovary | Normal |
| TPBW3.1e | *PRCre,Pole*$^{fl(L424V)/+}$ | 78.7 | Mammary gland | Normal |
| TPBW3.1f | *PRCre,Pole*$^{fl(L424V)/+}$ | 78.7 | Uterus & Ovary | lymphoid tissue in fat, uterine atypical hyperplasia |
| TPBW3.1f | *PRCre,Pole*$^{fl(L424V)/+}$ | 78.7 | Mammary gland | Normal |
| TPBW3.1f | *PRCre,Pole*$^{fl(L424V)/+}$ | 78.7 | Spleen | Normal |
| TPBW6.2g | *PRCre* | 77.1 | Uterus & Ovary | Normal |
| TPBW6.2g | *PRCre* | 77.1 | Mammary gland | Normal |
| TPBW1.4a | *PRCre* | 81.9 | Uterus & Ovary | Normal |
| TPBW1.4a | *PRCre* | 81.9 | Mammary gland | Normal |
| TPBW3.1b | *PRCre* | 78.7 | Uterus & Ovary | Hyperplasia of the uterus |
| TPBW3.1b | *PRCre* | 78.7 | Mammary gland | Normal |
| TPBW3.1b | *PRCre* | 78.7 | Mottled spleen | Ask for advice |
| TPBW3.1c | *PRCre* | 78.7 | Uterus & Ovary | Normal |
| TPBW3.1c | *PRCre* | 78.7 | Mammary gland | Normal |
| TPBW3.1g | *PRCre* | 78.7 | Uterus & Ovary | Lymphoid tissue |
| TPBW3.1g | *PRCre* | 78.7 | Mammary gland | Normal |
| TPBW1.3b | *PRCre* | 79.3 | Uterus & Ovary | Normal |
| TPBW1.3b | *PRCre* | 79.3 | Mammary gland | Normal |

**Table G.2:** Histopathology results from *PRCre,Pole*$^{fl(L424V)/+}$ and *PRCre* mice.

# Appendix H

# Gross morphological analysis of compound mutant $\boldsymbol{Apc}^{\mathrm{Min}}$,$\boldsymbol{Pole}^{\mathrm{L424V/+}}$ mice

| Mouse ID | Genotype | Age (Days) | Sex | Litter number | Reason for sacrifice | Necropsy results |
|---|---|---|---|---|---|---|
| TIWW1.1b | $Apc^{Min}$,$Pole^{L424V/+}$ | 125 | Male | 1 | Anaemia, Cachexia | Intestinal polyposis, Splenomegaly |
| TIWW1.1c | $Apc^{Min}$ | 126 | Female | 1 | Control to TIWW1.1b | Intestinal polyposis, Splenomegaly |
| TIWW1.3a | $Apc^{Min}$,$Pole^{L424V/+}$ | 139 | Male | 2 | Anaemia, Cachexia | Intestinal polyposis, Splenomegaly |
| TIWW1.3c | $Apc^{Min}$ | 142 | Male | 2 | Control to TIWW1.3a | Intestinal polyposis, Splenomegaly |
| TIWW1.3e | $Apc^{Min}$ | 142 | Female | 2 | Control to TIWW1.3a | Intestinal polyposis, Splenomegaly |
| TIWW1.4f | $Apc^{Min}$,$Pole^{L424V/+}$ | 106 | Female | 3 | Anaemia, Cachexia | Intestinal polyposis, Splenomegaly |
| TIWW1.4b | $Apc^{Min}$,$Pole^{L424V/+}$ | 111 | Female | 3 | Control to TIWW1.4f | Intestinal polyposis, Splenomegaly |
| TIWW1.4e | $Apc^{Min}$,$Pole^{L424V/+}$ | 111 | Female | 3 | Control to TIWW1.4f | Intestinal polyposis, Splenomegaly |
| TIWW1.5f | $Apc^{Min}$ | 175 | Male | 4 | Anaemia, Cachexia, Rectal bleeding | Intestinal polyposis, Splenomegaly |
| TIWW1.5g | $Apc^{Min}$,$Pole^{L424V/+}$ | 175 | Female | 4 | Littermate to TIWW1.5f (No ApcMin within litter) | Intestinal polyposis, Splenomegaly |
| TIWW4.1a | $Apc^{Min}$ | 115 | Male | 5 | Control to TIWW4.1b | Intestinal polyposis, Splenomegaly |
| TIWW4.1b | $Apc^{Min}$,$Pole^{L424V/+}$ | 115 | Female | 5 | Anaemia, Cachexia | Intestinal polyposis, Splenomegaly |
| TIWW4.2a | $Apc^{Min}$ | 132 | Male | 6 | Anaemia, Cachexia | Large Intestinal polyposis, Splenomegaly |
| TIWW4.2c | $Apc^{Min}$,$Pole^{L424V/+}$ | 132 | Female | 6 | Littermate to TIWW4.2a (no controls left in litter) | Intestinal polyposis, Splenomegaly |
| TIWW4.3e | $Apc^{Min}$ | 143 | Female | 7 | Control to TIWW5.1b | Intestinal polyposis, Splenomegaly |
| TIWW5.1b | $Apc^{Min}$,$Pole^{L424V/+}$ | 143 | Female | 8 | Cachexia, Solid mass above the left hind leg, Rectal bleeding | Tumour (potentially skin or mammary origin), Large polyp in colon |
| TIWW5.2b | $Apc^{Min}$,$Pole^{L424V/+}$ | 126 | Female | 9 | Anal prolapse | Large bloody polyp in colon |
| TIWW5.2g | $Apc^{Min}$,$Pole^{L424V/+}$ | 173 | Female | 9 | Anaemia, Cachexia, Anal prolapse | Intestinal polyposis - mostly in the small bowel and Splenomegaly |
| TIWW5.2d | $Apc^{Min}$,$Pole^{L424V/+}$ | 182 | Female | 9 | Anaemia, Cachexia, Anal bleeding | Intestinal polyposis, Splenomegaly |
| TIWW5.2f | $Apc^{Min}$,$Pole^{L424V/+}$ | 182 | Female | 9 | Anaemia, Cachexia, Anal bleeding | Intestinal polyposis, Mammary tumour, Splenomegaly |
| TIWW5.3c | $Apc^{Min}$,$Pole^{L424V/+}$ | 104 | Male | 10 | Anaemia, Cachexia, Anal bleeding | Intestinal polyposis (large mass in SB1 and multiple colonic polyps), Splenomegaly |
| TIWW5.3e | $Apc^{Min}$ | 104 | Male | 10 | Control to TIWW5.3c, rectal bleeding | Intestinal polyposis, Splenomegaly |
| TIWW5.4b | $Apc^{Min}$ | 151 | Male | 11 | Anaemia, Cachexia | Intestinal polyposis, Splenomegaly |
| TIWW5.4e | $Apc^{Min}$,$Pole^{L424V/+}$ | 157 | Female | 11 | Anaemia, Cachexia | Intestinal polyposis, Splenomegaly |
| TIWW6.1c | $Apc^{Min}$ | 187 | Male | 12 | Control to TIWW6.1d | Intestinal polyposis, Splenomegaly |
| TIWW6.1d | $Apc^{Min}$,$Pole^{L424V/+}$ | 187 | Female | 12 | Anaemia, Cachexia, Rectal bleeding | Intestinal polyposis (Multiple colonic polyps), Splenomegaly |
| TIWW6.1f | $Apc^{Min}$ | 187 | Female | 12 | Control to TIWW6.1d | Intestinal polyposis, Splenomegaly |
| TIWW7.1f | $Apc^{Min}$ | 88 | Female | 13 | Anaemia, Cachexia | Intestinal polyposis (small intestine covered in polyps), Splenomegaly |

**Table H.1:** Gross morphological analysis of $Apc^{Min}$,$Pole^{L424V/+}$ and $Apc^{Min}$ mice. Mice were culled when displaying a phenotype such as hunching, cachexia, anal prolapse or severe rectal bleeding or a visible external tumour. The abdomen and chest cavity were examined for abnormalities, specifically focussing on the intestine, uterus, mammary tissue and lung.

# Appendix I

# Raw intestinal polyp counts from compound mutant $\boldsymbol{Apc}^{\mathrm{Min}}$,$\boldsymbol{Pole}^{\mathrm{L424V}/+}$ mice

| | | | | SB1 | | | | SB2 | | | | SB3 | | | | Colon | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Mouse ID** | **Genotype** | **Sex** | **Age** | **<1mm** | **1-2mm** | **2-3mm** | **>3mm** | **<1mm** | **1-2mm** | **2-3mm** | **>3mm** | **<1mm** | **1-2mm** | **2-3mm** | **>3mm** | **<1mm** | **1-2mm** | **2-3mm** | **>3mm** |
| TIWW5.3c | $Apc^{Min}$,$Pole^{L424V/+}$ | Male | 104 | 1 | 1 | 3 | 3 | 0 | 4 | 4 | 0 | 3 | 14 | 5 | 1 | 0 | 3 | 3 | 1 |
| TIWW1.4f | $Apc^{Min}$,$Pole^{L424V/+}$ | Female | 106 | 2 | 1 | 4 | 3 | 7 | 38 | 10 | 0 | 7 | 34 | 8 | 0 | 0 | 0 | 0 | 3 |
| TIWW1.4b | $Apc^{Min}$,$Pole^{L424V/+}$ | Female | 111 | 0 | 0 | 0 | 0 | 2 | 6 | 3 | 0 | 7 | 8 | 0 | 0 | 0 | 0 | 0 | 0 |
| TIWW1.4e | $Apc^{Min}$,$Pole^{L424V/+}$ | Female | 111 | 1 | 1 | 2 | 0 | 0 | 3 | 6 | 3 | 0 | 7 | 8 | 0 | 0 | 0 | 0 | 0 |
| TIWW4.1b | $Apc^{Min}$,$Pole^{L424V/+}$ | Female | 115 | 0 | 2 | 2 | 0 | 3 | 5 | 0 | 0 | 0 | 13 | 7 | 1 | 0 | 0 | 0 | 0 |
| TIWW1.1b | $Apc^{Min}$,$Pole^{L424V/+}$ | Male | 125 | 0 | 0 | 7 | 2 | 6 | 13 | 12 | 0 | 0 | 11 | 12 | 5 | 0 | 0 | 0 | 1 |
| TIWW5.2b | $Apc^{Min}$,$Pole^{L424V/+}$ | Female | 126 | 0 | 0 | 2 | 0 | 5 | 12 | 3 | 0 | 2 | 13 | 3 | 1 | 0 | 0 | 0 | 1 |
| TIWW4.2c | $Apc^{Min}$,$Pole^{L424V/+}$ | Female | 132 | 0 | 1 | 3 | 0 | 0 | 1 | 5 | 1 | 4 | 11 | 3 | 0 | 2 | 0 | 0 | 0 |
| TIWW1.3a | $Apc^{Min}$,$Pole^{L424V/+}$ | Male | 139 | 0 | 2 | 0 | 0 | 0 | 6 | 8 | 2 | 1 | 9 | 8 | 0 | 0 | 0 | 0 | 2 |
| TIWW5.1b | $Apc^{Min}$,$Pole^{L424V/+}$ | Female | 143 | 0 | 0 | 4 | 0 | 0 | 2 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 |
| TIWW5.4e | $Apc^{Min}$,$Pole^{L424V/+}$ | Female | 157 | 0 | 0 | 2 | 1 | 6 | 4 | 1 | 2 | 3 | 10 | 10 | 1 | 0 | 0 | 0 | 0 |
| TIWW5.2g | $Apc^{Min}$,$Pole^{L424V/+}$ | Female | 173 | 1 | 2 | 6 | 7 | 0 | 4 | 2 | 1 | 0 | 6 | 8 | 2 | 0 | 0 | 0 | 0 |
| TIWW1.5g | $Apc^{Min}$,$Pole^{L424V/+}$ | Female | 175 | 0 | 0 | 2 | 1 | 0 | 7 | 2 | 1 | 0 | 13 | 6 | 2 | 0 | 0 | 0 | 0 |
| TIWW5.2d | $Apc^{Min}$,$Pole^{L424V/+}$ | Female | 182 | 0 | 0 | 0 | 2 | 0 | 1 | 1 | 3 | 2 | 1 | 1 | 0 | 0 | 1 | 0 | 2 |
| TIWW5.2f | $Apc^{Min}$,$Pole^{L424V/+}$ | Female | 182 | 0 | 0 | 2 | 1 | 0 | 1 | 6 | 1 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 3 |
| TIWW6.1d | $Apc^{Min}$,$Pole^{L424V/+}$ | Female | 187 | 0 | 1 | 7 | 2 | 4 | 14 | 14 | 6 | 4 | 13 | 16 | 0 | 0 | 0 | 0 | 3 |
| TIWW7.1f | $Apc^{Min}$ | Female | 88 | 0 | 5 | 12 | 2 | 1 | 43 | 16 | 3 | 6 | 56 | 7 | 1 | 0 | 1 | 0 | 0 |
| TIWW5.3e | $Apc^{Min}$ | Male | 104 | 0 | 0 | 2 | 0 | 0 | 2 | 1 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 1 | 2 |
| TIWW4.1a | $Apc^{Min}$ | Male | 115 | 1 | 0 | 0 | 0 | 5 | 2 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| TIWW1.1c | $Apc^{Min}$ | Female | 126 | 0 | 0 | 0 | 0 | 0 | 2 | 3 | 3 | 7 | 8 | 8 | 5 | 0 | 0 | 0 | 0 |
| TIWW4.2a | $Apc^{Min}$ | Male | 132 | 0 | 2 | 5 | 3 | 0 | 14 | 9 | 3 | 3 | 40 | 12 | 1 | 0 | 0 | 0 | 0 |
| TIWW1.3c | $Apc^{Min}$ | Male | 142 | 0 | 0 | 2 | 1 | 0 | 0 | 6 | 1 | 0 | 3 | 4 | 4 | 0 | 0 | 0 | 0 |
| TIWW1.3e | $Apc^{Min}$ | Female | 142 | 1 | 1 | 1 | 0 | 4 | 5 | 4 | 2 | 2 | 7 | 2 | 2 | 0 | 0 | 0 | 1 |
| TIWW4.3e | $Apc^{Min}$ | Female | 143 | 0 | 1 | 3 | 0 | 0 | 11 | 7 | 1 | 8 | 20 | 13 | 0 | 0 | 0 | 0 | 0 |
| TIWW5.4b | $Apc^{Min}$ | Male | 151 | 0 | 0 | 3 | 2 | 0 | 4 | 7 | 3 | 0 | 3 | 7 | 2 | 0 | 0 | 1 | 2 |
| TIWW1.5f | $Apc^{Min}$ | Male | 175 | 0 | 0 | 2 | 0 | 2 | 5 | 1 | 0 | 2 | 10 | 5 | 2 | 0 | 0 | 0 | 0 |
| TIWW6.1c | $Apc^{Min}$ | Male | 187 | 0 | 0 | 1 | 0 | 0 | 6 | 2 | 3 | 1 | 0 | 3 | 2 | 0 | 1 | 0 | 0 |
| TIWW6.1f | $Apc^{Min}$ | Female | 187 | 1 | 3 | 1 | 1 | 4 | 8 | 4 | 0 | 3 | 10 | 6 | 4 | 0 | 1 | 0 | 0 |

**Table I.1:** Raw intestinal polyp counts from $Apc^{Min}$,$Pole^{L424V/+}$ mutant compound and $Apc^{Min}$ mice. Polyps are listed according to intestinal location in the small intestine (SB1, SB2, SB3) and colon grouped by size (diameter in mm).

| | gRNA/tRNA peak ratio | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Triplex Probe** | **WT1** | **WT2 (Adenoma 1)** | **WT2 (Adenoma 2)** | **WT3** | **WT4 (Adenoma 1)** | **WT4 (Adenoma 2)** | **WT4 (Adenoma 3)** | **WT4 (Adenoma 4)** | **WT5** | **V411L** |
| **Bat25** | 0.90 | 1.0 | 1.0 | 1.0 | 0.90 | 0.90 | 0.90 | 0.82 | 1.0 | 1.0 |
| **Bat26** | 0.89 | 1.14 | 1.0 | 1.0.90 | 1.0 | 1.0 | 1.0 | 1.0 | 0.90 | 1.0 |
| **D2S123** | 0.77 | 0.29 | 1.0 | 0.91 | 1.0 | 1.0 | 0.35 | 1.0 | 1.27 | 0.80 |
| **D5S346/D17S250** | 0.88 | 0.92 | 0.73 | 0.93 | 1.0 | 1.25 | 1.50 | 1.0 | 1.0 | 0.71 |
| **Bat40** | 0.94 | 0.92 | 0.93 | 1.17 | 1.07 | 0.88 | 1.0 | 0.93 | 1.0 | 1.0 |
| **cumulative % variation from gDNA** | 12.51% | 14.63% | 6.76% | 1.91% | 0.57% | -0.50% | 5.00% | 4.97% | -3.45% | 9.71% |

**Table I.2:** Ratio of mononucleotide and dinucleotide marker peaks in gDNA/tDNA for adenomas from $POLE^{WT}$ and $POLE^{V411L}$ adenomas. Listed in the final row is % variation in peak frequency of gDNA vs tDNA (Samples with +/-30% variation is classed as microsatellite unstable).

# Appendix J

# Molecular Inversion Probe panel genes and target regions

| Genes | | | Gene fusions |
|---|---|---|---|
| *ACVR2A* | *MSH6* | *CTNNB1* | *PTPRK* (e1)-*RSPO3* (e2) |
| *APC* | *NRAS* | *ELF3* | *PTP*RK (e7)-*RSPO3* (e2) |
| *ARID1A* | *PIK3CA* | *FBXW7* | *EIF3E* (e1)-*RSPO2* |
| *ATM* | *POLD1* | *GNAS* | |
| *B2M* | *POLE* | *KRAS* | |
| *BCL9L* | *PTEN* | *MLH1* | |
| *BMPR2* | *RNF43* | *MSH2* | |
| *BRAF* | *RPL22* | *SMAD2* | |
| *TCF7L2* | *TP53* | *SMAD4* | |
| *TGIF1* | *ZFP36L2* | *SOX9* | |

**Table J.1:** Genes and gene fusion events covered by the Molecular Inversion Probe (MIP) panel covering 63.1kb. 30 genes, either full coding region (with 10bp flanking region) or hotspots or domains. In addition, 300 MIPs targeting 1800 TTTCTTT sites. These are the most common signature mutation seen when there is impaired POLE/POLD1 proofreading due to a functional mutation in the exonuclease domain. *APC* codons 1-1600; *POLD1*: chr19:50905938-50910304, *POLE*: chr12:133249809-133253238, *GNAS* codon 200, *BRAF*: exons 11 and 15 (chr7: 140453074-140453193, chr7: 140481375_140481493, *CTNNB1* hotspots: chr3: 41266444-41266698, chr3: 41267150-41267352, chr3: 41274831-41274935, chr3: 41275019-41275358, chr3: 41277214-41277334, chr3: 41266016-41266244, *PIK3CA* hotspots: chr3: 178916613178916965, chr3: 178935997-178936122, chr3: 178951881-178952152.

# References


[1] L. B. Alexandrov, S. Nik-Zainal, D. C. Wedge, S. A. J. R. Aparicio, S. Behjati, A. V. Biankin, G. R. Bignell, N. Bolli, A. Borg, A.-L. Brresen-Dale, S. Boyault, B. Burkhardt, A. P. Butler, C. Caldas, H. R. Davies, C. Desmedt, R. Eils, J. E. Eyfjrd, J. A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilicic, S. Imbeaud, M. Imielinski, M. Imielinsk, N. Jger, D. T. W. Jones, D. Jones, S. Knappskog, M. Kool, S. R. Lakhani, C. Lpez-Otn, S. Martin, N. C. Munshi, H. Nakamura, P. A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J. V. Pearson, X. S. Puente, K. Raine, M. Ramakrishna, A. L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T. N. Schumacher, P. N. Span, J. W. Teague, Y. Totoki, A. N. J. Tutt, R. Valds-Mas, M. M. van Buuren, L. van 't Veer, A. Vincent-Salomon, N. Waddell, L. R. Yates, J. Zucman-Rossi, P. A. Futreal, U. McDermott, P. Lichter, M. Meyerson, S. M. Grimmond, R. Siebert, E. Campo, T. Shibata, S. M. Pfister, P. J. Campbell, and M. R. Stratton, "Signatures of mutational processes in human cancer," *Nature*, vol. 500, no. 7463, pp. 415–421, 2013. xiii, 5, 29, 32, 76, 78, 92, 97, 129, 166, 174, 179, 180

[2] D. Hanahan and R. A. Weinberg, "The hallmarks of cancer," *Cell*, vol. 100, no. 1, pp. 57–70, 2000. 2

[3] D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: The next generation," *Cell*, vol. 144, no. 5, pp. 646–674, 2011. 2

[4] E. R. Fearon, "A genetic model for colorectal tumorigenesis," *Annu Rev Pathol*, vol. 61, no. 5, pp. 759–767, 2011. 2, 8

[5] E. F. Fearon and B. Vogelstein, "A genetic model for colorectal tumorigenesis," *Cell*, vol. 61, no. 5, pp. 759–767, 1990. 2, 3, 4

[6] L. D. Wood, D. W. Parsons, S. Jones, J. Lin, T. Sjblom, R. J. Leary, D. Shen, S. M. Boca, T. Barber, J. Ptak, N. Silliman, S. Szabo, Z. Dezso, V. Ustyanksky, T. Nikolskaya, Y. Nikolsky, R. Karchin, P. a. Wilson, J. S. Kaminker, Z. Zhang, R. Croshaw, J. Willis, D. Dawson, M. Shipitsin, J. K. V. Willson, S. Sukumar, K. Polyak, B. H. Park, C. L. Pethiyagoda, P. V. K. Pant, D. G. Ballinger, A. B. Sparks, J. Hartigan, D. R. Smith, E. Suh, N. Papadopoulos, P. Buckhaults, S. D. Markowitz, G. Parmigiani, K. W. Kinzler, V. E. Velculescu, and B. Vogelstein, "The genomic landscapes of human breast and colorectal cancers," *Science (New York, N.Y.)*, vol. 318, no. 5853, pp. 1108–1113, 2007. 2

[7] W. F. Bodmer, "Cancer genetics: colorectal cancer as a model," *Journal of Human Genetics*, vol. 51, no. 5, pp. 391–396, 2006. 2

[8] T. C. G. A. Network, "Comprehensive molecular characterization of human colon and rectal cancer," *Nature*, vol. 487, no. 7407, pp. 330–337, 2012. 3, 9, 20, 22, 23, 29, 30, 31, 76

[9] F. Castro-Giner, P. Ratcliffe, and I. Tomlinson, "The mini-driver model of polygenic cancer evolution," *Nature Reviews Cancer*, vol. 15, no. 11, pp. 680–685, 2015. 3, 183

[10] E. Heitzer and I. Tomlinson, "Replicative dna polymerase mutations in cancer," *Current Opinion in Genetics  Development*, vol. 24, pp. 107–113, 2014. 3, 30, 137, 166, 183

[11] W. Bodmer, "Genetic instability is not a requirement for tumor development," *Cancer Research*, vol. 68, no. 10, pp. 3558–3561, 2008. 4

[12] C. Tomasetti and B. Vogelstein, "Variation in cancer risk among tissues can be explained by the number of stem cell divisions," *Science*, vol. 347, no. 6217, pp. 78–81, 2015. 4

[13] L. A. Loeb, C. F. Springgate, and N. Battula, "Errors in dna replication as a basis of malignant changes 1," *Cancer Research*, vol. 34, no. 9, pp. 2311–2321, 1974. 4

[14] E. Cerami, J. Gao, U. Dogrusoz, B. E. Gross, S. Onur, E. Larsson, Y. Antipin, B. Reva, A. P. Goldberg, and C. Sander, "The cbio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data," *Cancer discovery*, vol. 2, no. 5, pp. 401–404, 2014. 4, 22, 23, 25, 32, 169

[15] T. Lindahl and R. D. Wood, "Quality control by dna repair," *Science*, vol. 286, no. 5446, pp. 1897–1905, 1999. 5

[16] S. S. Lange, K.-i. Takata, and R. D. Wood, "Dna polymerases and cancer," *Nature Reviews Cancer*, vol. 11, no. 2, pp. 96–110, 2011. 5

[17] R. D. Kolodner, C. D. Putnam, and K. Myung, "Maintenance of genome stability in saccharomyces cerevisiae," *Science*, vol. 297, no. 5581, pp. 552–557, 2002. 5

[18] C. Kandoth, M. D. McLellan, F. Vandin, K. Ye, B. Niu, C. Lu, M. Xie, Q. Zhang, J. F. McMichael, M. A. Wyczalkowski, M. D. Leiserson, C. A. Miller, J. S. Welch, M. J. Walter, M. C. Wendl, T. J. Ley, R. K. Wilson, B. J. Raphael, and L. Ding, "Mutational landscape and significance across 12 major cancer types," *Nature*, vol. 502, no. 7471, pp. 333–339, 2013. 5

[19] J. E. Cleaver, "Defective repair replication of dna in xeroderma pigmentosum," *Nature*, vol. 218, no. 5142, pp. 652–656, 1968. 5

[20] D. Stoppa-Lyonnet, "The biological effects and clinical implications of brca mutations: where do we go from here?," *European Journal Of Human Genetics*, vol. 24, pp. S3–S9, 2016. 5

[21] S. Negrini, V. G. Gorgoulis, and T. D. Halazonetis, "Genomic instability  an evolving hallmark of cancer," *Nature Reviews Molecular Cell Biology*, vol. 11, no. 3, pp. 220–228, 2010. 6

[22] N. Romero-Laorden and E. Castro, "Inherited mutations in dna repair genes and cancer risk," *Current Problems in Cancer*, vol. 41, no. 4, pp. 251–264, 2017. 6

[23] N. Al-Tassan, N. H. Chmiel, J. Maynard, N. Fleming, A. L. Livingston, G. T. Williams, A. K. Hodges, D. R. Davies, S. S. David, J. R. Sampson, and J. P. Cheadle, "Inherited variants of myh associated with somatic g:c - t:a mutations in colorectal tumors," *Nature Genetics*, vol. 30, no. 2, pp. 227–232, 2002. 6, 7, 182

[24] S. Jones, P. Emmerson, J. Maynard, J. M. Best, S. Jordan, G. T. Williams, J. R. Sampson, and J. P. Cheadle, "Biallelic germline mutations in myh predispose to multiple colorectal adenoma and somatic g:c - t:a mutations," *Human Molecular Genetics*, vol. 11, no. 23, pp. 2961–2967, 2002. 6, 7

[25] S. S. Wallace, "Base excision repair: a critical player in many games," *DNA Repair (Amst)*, vol. 19, pp. 14–26, 2014. 6

[26] T. Lindahl, "Instability and decay of the primary structure of dna," *Nature*, vol. 362, no. 6422, pp. 709–715, 1993. 6, 31

[27] S. S. David, V. L. O'Shea, and S. Kundu, "Base-excision repair of oxidative dna damage," *Nature*, vol. 447, no. 7147, pp. 941–950, 2007. 7

[28] J. R. Sampson and N. Jones, "Mutyh-associated polyposis," *Best Practice Research Clinical Gastroenterology*, vol. 23, no. 2, pp. 209–218, 2009. 7

[29] O. M. Sieber, K. M. Howarth, C. Thirlwell, A. Rowan, N. Mandir, R. A. Goodlad, A. Gilkar, B. Spencer-Dene, G. Stamp, V. Johnson, A. Silver, H. Yang, J. H. Miller, M. Ilyas, and I. P. Tomlinson, "Myh deficiency enhances intestinal tumorigenesis in multiple intestinal neoplasia (apc min/+) mice," *Cancer Research*, vol. 64, no. 24, pp. 8876–8881, 2004. 7, 18, 182, 183

[30] R. D. Weren, M. J. Ligtenberg, C. M. Kets, R. M. de Voer, E. T. Verwiel, L. Spruijt, W. A. van Zelst-Stams, M. C. Jongmans, C. Gilissen, J. Y. Hehir-Kwa, A. Hoischen, J. Shendure, E. A. Boyle, E. J. Kamping, I. D. Nagtegaal, B. B. Tops, F. M. Nagengast, A. Geurts van Kessel, J. H. van Krieken, R. P. Kuiper, and N. Hoogerbrugge, "A germline homozygous mutation in the base-excision repair gene nthl1 causes adenomatous polyposis and colorectal cancer," *Nature Genetics*, vol. 47, no. 6, pp. 668–671, 2015. 7, 182

[31] M. K. Chan, M. T. Ocampo-Hafalla, V. Vartanian, P. Jaruga, G. Kirkali, K. L. Koenig, S. Brown, R. S. Lloyd, M. Dizdaroglu, and G. W. Teebor, "Targeted deletion of the genes encoding nth1 and neil1 dna n-glycosylases reveals the existence of novel carcinogenic oxidative damage to dna," *DNA Repair*, vol. 8, no. 7, pp. 786–794, 2009. 7

[32] H. T. Tran, D. A. Gordenin, and M. A. Resnick, "The 3 to 5 exonucleases of dna polymerases delta and epsilon and the 5 to 3 exonuclease exo1 have major roles in postreplication mutation avoidance in saccharomyces cerevisiae," *Molecular and Cellular Biology*, vol. 19, no. 3, pp. 2000–2007, 1999. 8, 11

[33] T. A. Kunkel and D. A. Erie, "Eukaryotic mismatch repair in relation to dna replication," *Annu Rev Genet*, vol. 49, pp. 291–313, 2015. 8

[34] H. T. Tran, J. D. Keen, M. Kricker, M. a. Resnick, and D. a. Gordenin, "Hypermutability of homonucleotide runs in mismatch repair and dna polymerase proofreading yeast mutants," *Molecular and Cellular Biology*, vol. 17, no. 5, pp. 2859–2865, 1997. 8

[35] R. Fishel, M. K. Lescoe, M. R. S. Rao, N. G. Copeland, N. A. Jenkins, J. Garber, M. Kane, and R. Kolodner, "The human mutator gene homolog msh2 and its association with hereditary nonpolyposis colon cancer," *Cell*, vol. 75, no. 5, pp. 1027–1038, 1993. 8

[36] F. S. Leach, N. C. Nicolaides, N. Papadopoulos, B. Liu, J. Jen, R. Parsons, P. Peltomaki, P. Sistonen, L. A. Aaltonen, M. Nystrom-Lahti, and et al., "Mutations of a muts homolog in hereditary nonpolyposis colorectal cancer," *Cell*, vol. 75, no. 6, pp. 1215–1225, 1993. 8

[37] C. E. Bronner, S. M. Baker, P. T. Morrison, G. Warren, L. G. Smith, M. K. Lescoe, M. Kane, C. Earabino, J. Lipford, A. Lindblom, and et al., "Mutation in the dna mismatch repair gene homologue hmlh1 is associated with hereditary non-polyposis colon cancer," *Nature*, vol. 368, no. 6468, pp. 258–261, 1994. 8

[38] N. Papadopoulos, N. C. Nicolaides, Y. F. Wei, S. M. Ruben, K. C. Carter, C. A. Rosen, W. A. Haseltine, R. D. Fleischmann, C. M. Fraser, M. D. Adams, and et al., "Mutation of a mutl homolog in hereditary colon cancer," *Science*, vol. 263, no. 5153, pp. 1625–1629, 1994. 8

[39] W. Edelmann, K. Yang, A. Umar, J. Heyer, K. Lau, K. H. Fan, W. Liedtke, P. E. Cohen, M. F. Kane, J. R. Lipford, N. J. Yu, G. F. Crouse, J. W. Pollard, T. Kunkel, M. Lipkin, R. Kolodner, and R. Kucherlapati, "Mutation in the mismatch repair gene msh6 causes cancer susceptibility," *Cell*, vol. 91, no. 4, pp. 467–477, 1997. 8

[40] J. G. Herman, A. Umar, K. Polyak, J. R. Graff, N. Ahuja, J.-P. J. Issa, S. Markowitz, J. K. V. Willson, S. R. Hamilton, K. W. Kinzler, M. F. Kane, R. D. Kolodner, B. Vogelstein, T. A. Kunkel, and S. B. Baylin, "Incidence and functional consequences of hmlh1 promoter hypermethylation in colorectal carcinoma," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 12, pp. 6870–6875, 1998. 8

[41] D. J. Weisenberger, K. D. Siegmund, M. Campan, J. Young, T. I. Long, M. A. Faasse, G. H. Kang, M. Widschwendter, D. Weener, D. Buchanan, H. Koh, L. Simms, M. Barker, B. Leggett, J. Levine, M. Kim, A. J. French, S. N. Thibodeau, J. Jass, R. Haile, and P. W. Laird, "Cpg island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with braf mutation in colorectal cancer," *Nature Genetics*, vol. 38, no. 7, pp. 787–793, 2006. 8

[42] A. de la Chapelle and H. Hampel, "Clinical relevance of microsatellite instability in colorectal cancer," *Journal of Clinical Oncology*, vol. 28, no. 20, pp. 3380–3387, 2010. 9

[43] R. Okazaki, T. Okazaki, K. Sakabe, K. Sugimoto, and A. Sugino, "Mechanism of dna chain growth. i. possible discontinuity and unusual secondary structure of newly synthesized chains," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 59, no. 2, pp. 598–605, 1968. 9

[44] A. Costa, I. V. Hood, and J. M. Berger, "Mechanisms for initiating cellular dna replication," *Annual review of biochemistry*, vol. 82, pp. 25–54, 2013. 9

[45] P. M. Burgers and T. A. Kunkel, "Eukaryotic dna replication fork," *Annu Rev Biochem*, vol. 86, no. 1, pp. 417–438, 2017. 9, 10, 12

[46] S. Muramatsu, K. Hirai, Y. S. Tak, Y. Kamimura, and H. Araki, "Cdk-dependent complex formation between replication proteins dpb11, sld2, pol epsilon, and gins in budding yeast," *Genes and Development*, vol. 24, no. 6, pp. 602–612, 2010. 9

[47] S. E. Moyer, P. W. Lewis, and M. R. Botchan, "Isolation of the cdc45/mcm2-7/gins (cmg) complex, a candidate for the eukaryotic dna replication fork helicase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 27, pp. 10236–10241, 2006. 10

[48] A. Morrison, H. Araki, A. B. Clark, R. K. Hamatake, and A. Sugino, "A third essential dna polymerase in s. cerevisiae," *Cell*, vol. 62, no. 6, pp. 1143–1151, 1990. 10

[49] M. Budd and J. L. Campbell, "Temperature-sensitive mutations in the yeast dna polymerase i gene," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 84, no. 9, pp. 2838–2842, 1987. 10

[50] M. E. Budd and J. L. Campbell, "Dna polymerases delta and epsilon are required for chromosomal replication in saccharomyces cerevisiae," *Molecular Cell Biology*, vol. 13, no. 1, pp. 496–505, 1993. 10

[51] S. Francesconi, H. Park, and T. S. Wang, "Fission yeast with dna polymerase delta temperature-sensitive alleles exhibits cell division cycle phenotype," *Nucleic Acids Research*, vol. 21, no. 16, pp. 3821–3828, 1993. 10

[52] L. Pellegrini, "The pol alpha-primase complex," *Subcell Biochem*, vol. 62, pp. 157–169, 2012. 10

[53] A. Bernad, L. Blanco, J. M. Lazaro, G. Martin, and M. Salas, "A conserved 3' - 5' exonuclease active site in prokaryotic and eukaryotic dna polymerases," *Cell*, vol. 59, no. 1, pp. 219–228, 1989. 10, 14

[54] M. Simon, L. Giot, and G. Faye, "The 3' to 5' exonuclease activity located in the dna polymerase delta subunit of saccharomyces cerevisiae is required for accurate replication," *The EMBO journal*, vol. 10, no. 8, pp. 2165–2170, 1991. 10, 17

[55] O. Chilkova, B. H. Jonsson, and E. Johansson, "The quaternary structure of dna polymerase epsilon from saccharomyces cerevisiae," *Journal of Biological Chemistry*, vol. 278, no. 16, pp. 14082–14086, 2003. 10

[56] M. J. Prindle, M. W. Schmitt, F. Parmeggiani, and L. a. Loeb, "A substitution in the fingers domain of dna polymerase delta reduces fidelity by altering nucleotide discrimination in the catalytic site," *The Journal of Biological Chemistry*, vol. 288, no. 8, pp. 5572–5580, 2013. 10

[57] Z. F. Pursell and T. A. Kunkel, "Dna polymerase epsilon: a polymerase of unusual size (and complexity)," *Progress in nucleic acid research and molecular biology*, vol. 82, pp. 101–145, 2008. 10

[58] T. Kesti, K. Flick, S. Kernen, J. E. Syvoja, and C. Wittenberg, "Dna polymerase epsilon catalytic domains are dispensable for dna replication, dna repair, and cell viability," *Molecular Cell*, vol. 3, no. 5, pp. 679–685, 1999. 10

[59] R. Dua, D. L. Levy, and J. L. Campbell, "Analysis of the essential functions of the c-terminal protein/protein interaction domain of saccharomyces cerevisiae pol epsilon and its unexpected ability to support growth in the absence of the dna polymerase domain," *Journal of Biological Chemistry*, vol. 274, no. 32, pp. 22283–22288, 1999. 10, 11, 76

[60] K. E. Zahn and S. Doubli, "Look ma, no pcna: how dna polymerase epsilon synthesizes long stretches of dna without a processivity factor," *Nature Structural & Molecular Biology*, vol. 21, no. 1, pp. 12–14, 2014. 10, 14

[61] M. Hogg, P. Osterman, G. O. Bylund, R. A. Ganai, E. B. Lundstrom, A. E. Sauer-Eriksson, and E. Johansson, "Structural basis for processive dna synthesis by yeast dna polymerase epsilon," *Nature Structural & Molecular Biology*, vol. 21, no. 1, pp. 49–55, 2014. 11, 14

[62] C. Nishida, P. Reinhard, and S. Linn, "Dna repair synthesis in human fibroblasts requires dna polymerase delta," *Journal of Biological Chemistry*, vol. 263, no. 1, pp. 501–510, 1988. 11

[63] A. R. Lehmann, "Dna polymerases and repair synthesis in ner in human cells," *DNA Repair (Amst)*, vol. 10, no. 7, pp. 730–733, 2011. 11

[64] A. K. Rytkonen, M. Vaara, T. Nethanel, G. Kaufmann, R. Sormunen, E. Laara, H. P. Nasheuer, A. Rahmeh, M. Y. Lee, J. E. Syvaoja, and H. Pospiech, "Distinctive activities of dna polymerases during human dna replication," *Febs j*, vol. 273, no. 13, pp. 2984–3001, 2006. 11

[65] A. Blank, B. Kim, and L. A. Loeb, "Dna polymerase delta is required for base excision repair of dna methylation damage in saccharomyces cerevisiae," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 19, pp. 9047–9051, 1994. 11

[66] M. Stucki, B. Pascucci, E. Parlanti, P. Fortini, S. H. Wilson, U. Hubscher, and E. Dogliotti, "Mammalian base excision repair by dna polymerases delta and epsilon," *Oncogene*, vol. 17, no. 7, pp. 835–843, 1998. 11, 30

[67] E. Parlanti, B. Pascucci, G. Terrados, L. Blanco, and E. Dogliotti, "Aphidicolin-resistant and -sensitive base excision repair in wild-type and dna polymerase beta-defective mouse cells," *DNA Repair*, vol. 3, no. 7, pp. 703–710, 2004. 11

[68] N. Bowen, C. E. Smith, A. Srivatsan, S. Willcox, J. D. Griffith, and R. D. Kolodner, "Reconstitution of long and short patch mismatch repair reactions using saccharomyces cerevisiae proteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 46, pp. 18472–18477, 2013. 11

[69] Y. Zhang, F. Yuan, S. R. Presnell, K. Tian, Y. Gao, A. E. Tomkinson, L. Gu, and G. M. Li, "Reconstitution of 5'-directed human mismatch repair in a purified system," *Cell*, vol. 122, no. 5, pp. 693–705, 2005. 11

[70] L. Costantino, S. K. Sotiriou, J. K. Rantala, S. Magin, E. Mladenov, T. Helleday, J. E. Haber, G. Iliakis, O. P. Kallioniemi, and T. D. Halazonetis, "Break-induced replication repair of damaged forks induces genomic duplications in human cells," *Science*, vol. 343, no. 6166, pp. 88–91, 2014. 11, 12

[71] J. R. Lydeard, S. Jain, M. Yamaguchi, and J. E. Haber, "Break-induced replication and telomerase-independent telomere maintenance require pol32," *Nature*, vol. 448, no. 7155, pp. 820–823, 2007. 11, 70

[72] H. Lou, M. Komata, Y. Katou, Z. Guan, C. C. Reis, M. Budd, K. Shirahige, and J. L. Campbell, "Mrc1 and dna polymerase epsilon function together in linking dna replication and the s phase checkpoint," *Molecular Cell*, vol. 32, no. 1, pp. 106–117, 2009. 11, 76, 105

[73] T. A. Navas, Z. Zhou, and S. J. Elledge, "Dna polymerase epsilon links the dna replication machinery to the s phase checkpoint," *Cell*, vol. 80, no. 1, pp. 29–39, 1995. 11

[74] O. Chilkova, P. Stenlund, I. Isoz, C. M. Stith, P. Grabowski, E. B. Lundstrom, P. M. Burgers, and E. Johansson, "The eukaryotic leading and lagging strand dna polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of pcna," *Nucleic Acids Research*, vol. 35, no. 19, pp. 6588–6597, 2007. 11, 14

[75] A. G. Baranovskiy, N. D. Babayeva, V. G. Liston, I. B. Rogozin, E. V. Koonin, Y. I. Pavlov, D. G. Vassylyev, and T. H. Tahirov, "X-ray structure of the complex of regulatory subunits of human dna polymerase delta," *Cell Cycle*, vol. 7, no. 19, pp. 3026–3036, 2008. 11

[76] V. N. Podust, L. S. Chang, R. Ott, G. L. Dianov, and E. Fanning, "Reconstitution of human dna polymerase delta using recombinant baculoviruses: the p12 subunit potentiates dna polymerizing activity of the four-subunit enzyme," *Journal of Biological Chemistry*, vol. 277, no. 6, pp. 3894–3901, 2002. 11

[77] Y. Zhou, X. Meng, S. Zhang, E. Y. Lee, and M. Y. Lee, "Characterization of human dna polymerase delta and its subassemblies reconstituted by expression in the multibac system," *PLoS One*, vol. 7, no. 6, p. e39156, 2012. 11, 22

[78] M. Y. W. T. Lee, S. Zhang, S. Hua, S. Lin, J. Chea, X. Wang, C. Leroy, A. Wong, Z. Zhang, and E. Y. C. Lee, "Regulation of human dna polymerase delta in the cellular responses to dna damage," *Environmental and Molecular Mutagenesis*, vol. 53, no. 9, pp. 683–698, 2012. 11

[79] X. Meng, Y. Zhou, S. Zhang, E. Y. C. Lee, D. N. Frick, and M. Y. W. T. Lee, "Dna damage alters dna polymerase delta to a form that exhibits increased discrimination against modified template bases and mismatched primers," *Nucleic Acids Research*, vol. 37, no. 2, pp. 647–657, 2009. 11

[80] A. G. Baranovskiy, A. G. Lada, H. M. Siebler, Y. Zhang, Y. I. Pavlov, and T. H. Tahirov, "Dna polymerase delta and zeta switch by sharing accessory subunits of dna polymerase delta," *Journal of Biological Chemistry*, vol. 287, no. 21, pp. 17281–17287, 2012. 11

[81] M. G. Dunlop, S. E. Dobbins, S. M. Farrington, A. M. Jones, C. Palles, N. Whiffin, A. Tenesa, S. Spain, P. Broderick, L.-Y. Ooi, E. Domingo, C. Smillie, M. Henrion, M. Frampton, L. Martin, G. Grimes, M. Gorman, C. Semple, Y. P. Ma, E. Barclay, J. Prendergast, J.-B. Cazier, B. Olver, S. Penegar, S. Lubbe, I. Chander, L. G. Carvajal-Carmona, S. Ballereau, A. Lloyd, J. Vijayakrishnan, L. Zgaga, I. Rudan, E. Theodoratou, J. M. Starr, I. Deary, I. Kirac, D. Kovacevi, L. a. Aaltonen, L. Renkonen-Sinisalo, J.-P. Mecklin, K. Matsuda, Y. Nakamura, Y. Okada, S. Gallinger, D. J. Duggan, D. Conti, P. Newcomb, J. Hopper, M. a. Jenkins, F. Schumacher, G. Casey, D. Easton, M. Shah, P. Pharoah, A. Lindblom, T. Liu, C. G. Smith, H. West, J. P. Cheadle, R. Midgley, D. J. Kerr, H. Campbell, I. P. Tomlinson, and R. S. Houlston, "Common variation near cdkn1a, pold3 and shroom2 influences colorectal cancer risk," *Nature Genetics*, vol. 44, no. 7, pp. 770–776, 2012. 12

[82] E. Tumini, S. Barroso, C. P. Calero, and A. Aguilera, "Roles of human pold1 and pold3 in genome stability," *Scientific Reports*, vol. 6, p. 38873, 2016. 12

[83] Q. M. Huang, S. Tomida, Y. Masuda, C. Arima, K. Cao, T.-a. Kasahara, H. Osada, Y. Yatabe, T. Akashi, K. Kamiya, T. Takahashi, and M. Suzuki, "Regulation of dna polymerase pold4 influences genomic instability in lung cancer," *Cancer Research*, vol. 70, no. 21, pp. 8407–8416, 2010. 12

[84] Y. Li, H. Asahara, V. S. Patel, S. Zhou, and S. Linn, "Purification, cdna cloning, and gene mapping of the small subunit of human dna polymerase," *Journal of Biological Chemistry*, vol. 272, no. 51, pp. 32337–32344, 1997. 12

[85] M. Garbacz, H. Araki, K. Flis, A. Bebenek, A. Zawada, P. Jonczyk, K. Makiela-Dzbenska, and I. J. Fijalkowska, "Fidelity consequences of the impaired interaction between dna polymerase epsilon and the gins complex," *DNA Repair*, vol. 29, pp. 23–35., 2015. 12

[86] S. Sengupta, F. van Deursen, G. de Piccoli, and K. Labib, "Dpb2 integrates the leading-strand dna polymerase into the eukaryotic replisome," *Current Biology*, vol. 23, no. 7, pp. 543–552, 2013. 12

[87] L. D. Langston, D. Zhang, O. Yurieva, R. E. Georgescu, J. Finkelstein, N. Y. Yao, C. Indiani, and M. E. O'Donnell, "Cmg helicase and dna polymerase epsilon form a functional 15-subunit holoenzyme for eukaryotic leading-strand dna replication," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 43, pp. 15390–15395, 2014. 12

[88] R. Dua, D. L. Levy, and J. L. Campbell, "Role of the putative zinc finger domain of saccharomyces cerevisiae dna polymerase epsilon in dna replication and the s/m checkpoint pathway," *Journal of Biological Chemistry*, vol. 273, no. 45, pp. 30046–30055, 1998. 12

[89] R. Dua, D. L. Levy, C. M. Li, P. M. Snow, and J. L. Campbell, "In vivo reconstitution of saccharomyces cerevisiae dna polymerase epsilon in insect cells. purification and characterization," *Journal of Biological Chemistry*, vol. 277, no. 10, pp. 7889–7896, 2002. 12

[90] A. Aksenova, K. Volkov, J. Maceluch, Z. F. Pursell, I. B. Rogozin, T. a. Kunkel, Y. I. Pavlov, and E. Johansson, "Mismatch repair-independent increase in spontaneous mutagenesis in yeast lacking non-essential subunits of dna polymerase epsilon," *PLoS Genetics*, vol. 6, no. 11, p. e1001209, 2010. 12

[91] R. Sahashi, R. Matsuda, O. Suyari, M. Kawai, H. Yoshida, S. Cotterill, and M. Yamaguchi, "Functional analysis of drosophila dna polymerase epsilon p58 subunit," *American journal of cancer research*, vol. 3, no. 5, pp. 478–489, 2013. 12

[92] R. Bellelli, V. Borel, C. Logan, J. Svendsen, D. E. Cox, E. Nye, K. Metcalfe, S. M. OConnell, G. Stamp, H. R. Flynn, A. P. Snijders, F. Lassailly, A. Jackson, and S. J. Boulton, "Pol epsilon instability drives replication stress, abnormal development, and tumorigenesis," *Molecular Cell*, vol. 70, no. 4, pp. 707–721.e7, 2018. 12

[93] A. Morrison and A. Sugino, "The 3 - 5 exonucleases of both dna polymerases delta and epslion participate in correcting errors of dna replication in saccharomyces cerevisiae," *Molecular and General Genetics MGG*, vol. 242, no. 3, pp. 289–296, 1994. 12, 17, 18

[94] Z. F. Pursell, I. Isoz, E. B. Lundstrm, E. Johansson, and T. a. Kunkel, "Regulation of b family dna polymerase fidelity by a conserved active site residue: Characterization of m644w, m644l and m644f mutants of yeast dna polymerase epsilon," *Nucleic Acids Research*, vol. 35, no. 9, pp. 3076–3086, 2007. 12, 25, 130

[95] S. A. N. McElhinny, C. M. Stith, P. M. J. Burgers, and T. A. Kunkel, "Inefficient proofreading and biased error rates during inaccurate dna synthesis by a mutant derivative of saccharomyces cerevisiae dna polymerase delta," *Journal of Biological Chemistry*, vol. 282, no. 4, pp. 2324–2332, 2007. 12

[96] Z. F. Pursell, I. Isoz, E.-b. Lundstrm, E. Johansson, and A. Thomas, "Yeast dna polymerase epsilon participates in leading-strand dna replication," *Science*, vol. 317, no. 5834, pp. 127–130, 2007. 12

[97] N. McElhinny, S. A. Gordenin, D. A., C. M. Stith, P. M. J. Burgers, and T. A. Kunkel, "Division of labour at the eukaryotic replication fork," *Molecular Cell*, vol. 30, no. 2, pp. 137–144, 2008. 12

[98] S. A. N. McElhinny, D. Kumar, A. B. Clark, D. L. Watt, B. E. Watts, E.-B. Lundstrm, E. Johansson, A. Chabes, and T. A. Kunkel, "Genome instability due to ribonucleotide incorporation into dna," *Nature Chemical Biology*, vol. 6, no. 10, p. 774, 2010. 13

[99] I. Miyabe, T. a. Kunkel, and A. M. Carr, "The major roles of dna polymerases epsilon and delta at the eukaryotic replication fork are evolutionarily conserved," *PLoS Genetics*, vol. 7, no. 12, p. e1002407, 2011. 13

[100] A. A. Larrea, S. A. Lujan, S. A. N. McElhinny, P. A. Mieczkowski, M. A. Resnick, D. A. Gordenin, and T. A. Kunkel, "Genome-wide model for the normal eukaryotic dna replication fork," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 41, pp. 17674–17679, 2010. 13

[101] Y. Daigaku, A. Keszthelyi, C. A. Mller, I. Miyabe, T. Brooks, R. Retkute, M. Hubank, C. A. Nieduszynski, and A. M. Carr, "A global profile of replicative polymerase usage," *Nature Structural  Molecular Biology*, vol. 22, no. 3, pp. 192–198, 2015. 13

[102] M. A. Garbacz, S. A. Lujan, A. B. Burkholder, P. B. Cox, Q. Wu, Z. X. Zhou, J. E. Haber, and T. A. Kunkel, "Evidence that dna polymerase delta contributes to initiating leading strand dna replication in saccharomyces cerevisiae," *Nature Communications*, vol. 9, no. 1, p. 858, 2018. 13

[103] M. A. M. Reijns, H. Kemp, J. Ding, S. M. de Proc, A. P. Jackson, and M. S. Taylor, "Lagging-strand replication shapes the mutational landscape of the genome," *Nature*, vol. 518, no. 7540, pp. 502–506, 2015. 13

[104] A. R. Clausen, S. A. Lujan, A. B. Burkholder, C. D. Orebaugh, J. S. Williams, M. F. Clausen, E. P. Malc, P. A. Mieczkowski, D. C. Fargo, D. J. Smith, and T. A. Kunkel, "Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation," *Nature Structural & Molecular Biology*, vol. 22, no. 3, pp. 185–191, 2015. 13

[105] I. Miyabe, K. Mizuno, A. Keszthelyi, Y. Daigaku, M. Skouteri, S. Mohebi, T. A. Kunkel, J. M. Murray, and A. M. Carr, "Polymerase delta replicates both strands after homologous recombination-dependent fork restart," *Nature Structural & Molecular Biology*, vol. 22, no. 11, pp. 932–938, 2015. 13

[106] R. E. Johnson, R. Klassen, L. Prakash, and S. Prakash, "A major role of dna polymerase delta in replication of both the leading and lagging dna strands," *Molecular Cell*, vol. 59, no. 2, pp. 163–175, 2015. 13

[107] C. L. Flood, G. P. Rodriguez, G. Bao, A. H. Shockley, Y. W. Kow, and G. F. Crouse, "Replicative dna polymerase delta but not epsilon proofreads errors in cis and in trans," *PLOS Genetics*, vol. 11, no. 3, p. e1005049, 2015. 13

[108] R. E. Georgescu, L. Langston, N. Y. Yao, O. Yurieva, D. Zhang, J. Finkelstein, T. Agarwal, and M. E. O'Donnell, "Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork," *Nature Structural & Molecular Biology*, vol. 21, no. 8, pp. 664–670, 2014. 13

[109] R. E. Georgescu, G. D. Schauer, N. Y. Yao, L. D. Langston, O. Yurieva, D. Zhang, J. Finkelstein, and M. E. O'Donnell, "Reconstitution of a eukaryotic replisome reveals suppression mechanisms that define leading/lagging strand operation," *Elife*, vol. 4, p. e04988, 2015. 13

[110] J. Sun, Y. Shi, R. E. Georgescu, Z. Yuan, B. T. Chait, H. Li, and M. E. O'Donnell, "The architecture of a eukaryotic replisome," *Nature Structural & Molecular Biology*, vol. 22, no. 12, pp. 976–982, 2015. 13

[111] G. Chui and S. Linn, "Further characterization of hela dna polymerase epsilon," *Journal of Biological Chemistry*, vol. 270, no. 14, pp. 7799–7808, 1995. 14

[112] A. K. Showalter and M.-D. Tsai, "A reexamination of the nucleotide incorporation fidelity of dna polymerases," *Biochemistry*, vol. 41, no. 34, pp. 10571–10576, 2002. 14

[113] T. A. Kunkel and P. M. Burgers, "Dividing the workload at a eukaryotic replication fork," *Trends in Cell Biology*, vol. 18, no. 11, pp. 521–7, 2008. 14

[114] S. D. McCulloch and T. a. Kunkel, "The fidelity of dna synthesis by eukaryotic replicative and translesion synthesis polymerases," *Cell Research*, vol. 18, no. 1, pp. 148–161, 2008. 14

[115] J. a. St Charles, S. E. Liberti, J. S. Williams, S. a. Lujan, and T. a. Kunkel, "Quantifying the contributions of base selectivity, proofreading and mismatch repair to nuclear dna replication in saccharomyces cerevisiae," *DNA Repair*, vol. 31, pp. 41–51, 2015. 14

[116] I. V. Shevelev and U. Hbscher, "The 3'- 5' exonucleases," *Nature Reviews Molecular cell biology*, vol. 3, no. 5, pp. 364–376, 2002. 14

[117] S. Briggs and I. Tomlinson, "Germline and somatic polymerase epsilon and delta mutations define a new class of hypermutated colorectal and endometrial cancers," *The Journal of Pathology*, vol. 230, no. 2, pp. 148–153, 2013. 14, 19, 27, 28

[118] Y. Shamoo and T. A. Steitz, "Building a replisome from interacting pieces: sliding clamp complexed to a peptide from dna polymerase and a polymerase editing complex," *Cell*, vol. 99, no. 2, pp. 155–166, 1999. 14

[119] E. E. Henninger and Z. F. Pursell, "Dna polymerase epsilon and its roles in genome stability," *IUBMB Life*, vol. 66, no. 5, pp. 339–351, 2014. 14

[120] A. Del Prado, E. Franco-Echevarria, B. Gonzalez, L. Blanco, M. Salas, and M. de Vega, "Noncatalytic aspartate at the exonuclease domain of proofreading dna polymerases regulates both degradative and synthetic activities," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 13, pp. E2921–E2929, 2018. 14

[121] R. a. Ganai, G. O. Bylund, and E. Johansson, "Switching between polymerase and exonuclease sites in dna polymerase epsilon," *Nucleic Acids Research*, vol. 4, no. 2, pp. 932–942, 2014. 14

[122] V. A. Simossis, J. Kleinjung, and J. Heringa, "Homology-extended sequence alignment," *Nucleic Acids Research*, vol. 33, no. 3, pp. 816–824, 2005. 15

[123] E. Rayner, I. C. Van Gool, C. Palles, S. E. Kearsey, T. Bosse, I. Tomlinson, and D. N. Church, "A panoply of errors: polymerase proofreading domain mutations in cancer," *Nature Reviews Cancer*, vol. 16, no. 2, pp. 71–81, 2016. 16, 20, 23, 24, 26, 31, 32, 92, 112, 134, 137, 156, 168, 174, 179, 181, 182

[124] Y. I. Pavlov, C. Frahm, S. a. N. McElhinny, a. Niimi, M. Suzuki, and T. a. Kunkel, "Evidence that errors made by dna polymerase alpha are corrected by dna polymerase delta," *Current Biology*, vol. 16, no. 2, pp. 202–207, 2006. 14

[125] a. Morrison, J. B. Bell, T. a. Kunkel, and a. Sugino, "Eukaryotic dna polymerase amino acid sequence required for 3'—-5' exonuclease activity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 21, pp. 9473–9477, 1991. 17, 135

[126] A. Morrison, A. L. Johnson, L. H. Johnston, and A. Sugino, "Pathway correcting dna replication errors in saccharomyces cerevisiae.," *Embo Journal*, vol. 12, no. 4, pp. 1467–1473, 1993. 17

[127] S. A. Lujan, J. S. Williams, Z. F. Pursell, A. A. Abdulovic-Cui, A. B. Clark, S. A. Nick McElhinny, and T. A. Kunkel, "Mismatch repair balances leading and lagging strand dna replication fidelity," *PLoS Genetics*, vol. 8, no. 10, pp. e1003016–e1003016, 2012. 17, 30

[128] L. N. Williams, A. J. Herr, and B. D. Preston, "Emergence of dna polymerase epsilon antimutators that escape error-induced extinction in yeast," *Genetics*, vol. 193, no. 3, pp. 751–770, 2013. 17, 18, 31, 32, 77, 113, 116, 131, 187

[129] A. J. Herr, M. Ogawa, N. A. Lawrence, L. N. Williams, J. M. Eggington, M. Singh, R. A. Smith, and B. D. Preston, "Mutator suppression and escape from replication error-induced extinction in yeast," *Plos Genetics*, vol. 7, no. 10, p. e1002282, 2011. 17, 77, 131, 187

[130] J. M. Fortune, Y. I. Pavlov, C. M. Welch, E. Johansson, P. M. Burgers, and T. A. Kunkel, "Saccharomyces cerevisiae dna polymerase delta: high fidelity for base substitutions but lower fidelity for single- and multi-base deletions," *Journal of Biological Chemistry*, vol. 280, no. 33, pp. 29980–29987, 2005. 17

[131] M. W. Schmitt, R. N. Venkatesan, M. J. Pillaire, J. S. Hoffmann, J. M. Sidorova, and L. A. Loeb, "Active site mutations in mammalian dna polymerase delta alter accuracy and replication fork progression," *Journal of Biological Chemistry*, vol. 285, no. 42, pp. 32264–32272, 2010. 17

[132] P. V. Shcherbakova, Y. I. Pavlov, O. Chilkova, I. B. Rogozin, E. Johansson, and T. a. Kunkel, "Unique error signature of the four-subunit yeast dna polymerase epsilon," *Journal of Biological Chemistry*, vol. 278, no. 44, pp. 43770–43780, 2003. 17

[133] L. N. Williams, L. Marjavaara, G. M. Knowels, E. M. Schultz, E. J. Fox, A. Chabes, and A. J. Herr, "dntp pool levels modulate mutator phenotypes of error-prone dna polymerase epsilon variants," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 19, pp. E2457–E2466, 2015. 17, 132

[134] A. Datta, J. L. Schmeits, N. S. Amin, P. J. Lau, K. Myung, and R. D. Kolodner, "Checkpoint-dependent activation of mutagenic repair in saccharomyces cerevisiae pol3-01 mutants," *Molecular Cell*, vol. 6, no. 3, pp. 593–603, 2000. 17

[135] R. A. Ganai, X.-P. Zhang, W.-D. Heyer, and E. Johansson, "Strand displacement synthesis by yeast dna polymerase epsilon," *Nucleic Acids Research*, vol. 44, no. 17, pp. 8229–8240, 2016. 18

[136] K. Lee, E. Tosti, and W. Edelmann, "Mouse models of dna mismatch repair in cancer research," *DNA Repair*, vol. 38, pp. 140–146, 2016. 18

[137] R. E. Goldsby, N. a. Lawrence, L. E. Hays, E. a. Olmsted, X. Chen, M. Singh, and B. D. Preston, "Defective dna polymerase-delta proofreading causes cancer susceptibility in mice," *Nature Medicine*, vol. 7, no. 6, pp. 638–639, 2001. 18

[138] R. E. Goldsby, L. E. Hays, X. Chen, E. A. Olmsted, W. B. Slayton, G. J. Spangrude, and B. D. Preston, "High incidence of epithelial cancers in mice deficient for dna polymerase delta proofreading," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 24, pp. 15560–15565, 2002. 18

[139] T. M. Albertson, M. Ogawa, J. M. Bugni, L. E. Hays, Y. Chen, Y. Wang, P. M. Treuting, J. A. Heddle, R. E. Goldsby, and B. D. Preston, "Dna polymerase epsilon and delta proofreading suppress discrete mutator and cancer phenotypes in mice," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 40, pp. 17101–17104, 2009. 18, 19, 31, 77, 78, 79, 135, 136, 148, 163

[140] A. Uchimura, Y. Hidaka, T. Hirabayashi, M. Hirabayashi, and T. Yagi, "Dna polymerase delta is required for early mammalian embryogenesis," *PloS one*, vol. 4, no. 1, p. e4184, 2009. 18

[141] A. Uchimura, M. Higuchi, Y. Minakuchi, M. Ohno, A. Toyoda, A. Fujiyama, I. Miura, S. Wakana, J. Nishino, and T. Yagi, "Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice," *Genome Research*, vol. 25, no. 8, pp. 1125–1134, 2015. 18

[142] S. D. McCulloch, R. J. Kokoska, O. Chilkova, C. M. Welch, E. Johansson, P. M. Burgers, and T. A. Kunkel, "Enzymatic switching for efficient and accurate translesion dna replication," *Nucleic Acids Research*, vol. 32, no. 15, pp. 4665–4675, 2004. 19

[143] R. Yoshida, K. Miyashita, M. Inoue, A. Shimamoto, Z. Yan, A. Egashira, E. Oki, Y. Kakeji, S. Oda, and Y. Maehara, "Concurrent genetic alterations in dna polymerase proofreading and mismatch repair in human colorectal cancer," *European journal of human genetics : EJHG*, vol. 19, no. 3, pp. 320–325, 2011. 19, 22

[144] C. Palles, J.-B. Cazier, K. M. Howarth, E. Domingo, A. M. Jones, P. Broderick, Z. Kemp, S. L. Spain, E. Guarino, E. Guarino Almeida, I. Salguero, A. Sherborne, D. Chubb, L. G. Carvajal-Carmona, Y. Ma, K. Kaur, S. Dobbins, E. Barclay, M. Gorman, L. Martin, M. B. Kovac, S. Humphray, A. Lucassen, C. C. Holmes, D. Bentley, P. Donnelly, J. Taylor, C. Petridis, R. Roylance, E. J. Sawyer, D. J. Kerr, S. Clark, J. Grimes, S. E. Kearsey, H. J. W. Thomas, G. McVean, R. S. Houlston, and I. Tomlinson, "Germline mutations affecting the proofreading domains of pole and pold1

predispose to colorectal adenomas and carcinomas," *Nature Genetics*, vol. 45, no. 2, pp. 136–144, 2013. 19, 20, 21, 22, 27, 28, 29, 30, 31, 33, 112, 134, 136, 165, 182

[145] I. Spier, S. Holzapfel, J. Altmller, B. Zhao, S. Horpaopan, S. Vogt, S. Chen, M. Morak, S. Raeder, K. Kayser, D. Stienen, R. Adam, P. Nrnberg, G. Plotz, E. Holinski-Feder, R. P. Lifton, H. Thiele, P. Hoffmann, V. Steinke, and S. Aretz, "Frequency and phenotypic spectrum of germline mutations in pole and seven other polymerase genes in 266 patients with colorectal adenomas and carcinomas," *International Journal of Cancer*, vol. 137, no. 2, pp. 320–331, 2015. 20, 21

[146] D. Chubb, P. Broderick, M. Frampton, B. Kinnersley, A. Sherborne, S. Penegar, A. Lloyd, Y. P. Ma, S. E. Dobbins, and R. S. Houlston, "Genetic diagnosis of high-penetrance susceptibility for colorectal cancer (crc) is achievable for a high proportion of familial crc by exome sequencing," *J Clin Oncol*, vol. 33, no. 5, pp. 426–432, 2015. 20, 21

[147] L. Valle, E. Hernndez-Illn, F. Bellido, G. Aiza, A. Castillejo, M.-I. Castillejo, M. Navarro, N. Segu, G. Vargas, C. Guarinos, M. Juarez, X. Sanjun, S. Iglesias, C. Alenda, C. Egoavil, n. Segura, M.-J. Juan, M. Rodriguez-Soler, J. Brunet, S. Gonzlez, R. Jover, C. Lzaro, G. Capell, M. Pineda, J. L. Soto, and I. Blanco, "New insights into pole and pold1 germline mutations in familial colorectal cancer and polyposis," *Human Molecular Genetics*, vol. 23, no. 13, pp. 3506–3512, 2014. 20, 21

[148] F. a. Elsayed, C. M. Kets, D. Ruano, B. van den Akker, A. R. Mensenkamp, M. Schrumpf, M. Nielsen, J. T. Wijnen, C. M. Tops, M. J. Ligtenberg, H. F. Vasen, F. J. Hes, H. Morreau, and T. van Wezel, "Germline variants in pole are associated with early onset mismatch repair deficient colorectal cancer," *European journal of human genetics : EJHG*, vol. 23, no. 8, pp. 1080–1084, 2014. 20, 31, 77, 112

[149] L. G. Aoude, E. Heitzer, P. Johansson, M. Gartside, K. Wadt, A. L. Pritchard, J. M. Palmer, J. Symmons, A.-M. Gerdes, G. W. Montgomery, N. G. Martin, I. Tomlinson, S. Kearsey, and N. K. Hayward, "Pole mutations in families predisposed to cutaneous melanoma," *Familial Cancer*, vol. 14, no. 4, pp. 621–628, 2015. 20, 21, 23

[150] F. Bellido, M. Pineda, G. Aiza, R. Valds-Mas, M. Navarro, D. a. Puente, T. Pons, S. Gonzlez, S. Iglesias, E. Darder, V. Piol, J. L. Soto, A. Valencia, I. Blanco, M. Urioste, J. Brunet, C. Lzaro, G. Capell, X. S. Puente, and L. Valle, "Pole and pold1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance," *Genetics in Medicine*, vol. 18, no. 4, pp. 325–332, 2015. 20, 21, 23, 28

[151] R. Ferrer-Avargues, V. Diez-Obrero, E. Martin-Tomas, E. Hernandez-Illan, M. I. Castillejo, A. Codoner-Alejos, V. M. Barbera, A. B. Sanchez-Heras, A. Segura, M. J. Juan, I. Tena, A. Castillejo, and J. L. Soto, "Characterization of a novel pold1 missense founder mutation in a spanish population," *J Gene Med*, 2017. 20, 21, 23

[152] M. F. Hansen, J. Johansen, I. Bjrnevoll, A. E. Sylvander, K. S. Steinsbekk, P. Strom, A. K. Sandvik, F. Drabls, and W. Sjursen, "A novel pole mutation associated with cancers of colon, pancreas, ovaries and small intestine," *Familial Cancer*, vol. 14, no. 3, pp. 1–12, 2015. 20, 21

[153] C. Esteban-Jurado, D. Gimnez-Zaragoza, J. Muoz, S. Franch-Expsito, M. lvarez Barona, T. Ocaa, M. Cuatrecasas, S. Carballal, M. Lpez-Cern, M. Marti-Solano, M. Daz-Gay, T. van Wezel, A. Castells, L. Bujanda, J. Balmaa, V. Gonzalo, G. Llort, C. Ruiz-Ponte, J. Cubiella, F. Balaguer, R. Aligu, and S. Castellv-Bel, "Pole and pold1 screening in 155 patients with multiple polyps and early-onset colorectal cancer," *Oncotarget*, vol. 8, no. 16, pp. 26732–26743, 2017. 20, 21, 23

[154] A. Rohlin, T. Zagoras, S. Nilsson, U. Lundstam, J. Wahlstrm, L. Hultn, T. Martinsson, G. B. Karlsson, and M. Nordling, "A mutation in pole predisposing to a multi-tumour phenotype," *International Journal of Oncology*, vol. 45, no. 1, pp. 77–81, 2014. 20, 21

[155] K. Wimmer, A. Beilken, R. Nustede, T. Ripperger, B. Lamottke, B. Ure, D. Steinmann, T. Reineke-Plaass, U. Lehmann, J. Zschocke, L. Valle, C. Fauth, and C. P. Kratz, "A novel germline pole mutation causes an early onset cancer prone syndrome mimicking constitutional mismatch repair deficiency," *Familial Cancer*, vol. 16, no. 1, pp. 67–71, 2017. 20, 21, 28, 166

[156] S. Seshagiri, E. W. Stawiski, S. Durinck, Z. Modrusan, E. E. Storm, C. B. Conboy, S. Chaudhuri, Y. Guan, V. Janakiraman, B. S. Jaiswal, J. Guillory, C. Ha, G. J. P. Dijkgraaf, J. Stinson, F. Gnad, M. a. Huntley, J. D. Degenhardt, P. M. Haverty, R. Bourgon, W. Wang, H. Koeppen, R. Gentleman, T. K. Starr, Z. Zhang, D. a. Largaespada, T. D. Wu, and F. J. de Sauvage, "Recurrent r-spondin fusions in colon cancer," *Nature*, vol. 488, no. 7413, pp. 660–664, 2012. 20, 22, 23, 29, 30, 166

[157] T. C. G. A. Network, "Integrated genomic characterization of endometrial carcinoma," *Nature*, vol. 497, no. 7447, pp. 67–73, 2013. 20, 22, 23, 28, 31, 76, 167

[158] D. N. Church, S. E. W. Briggs, C. Palles, E. Domingo, S. J. Kearsey, J. M. Grimes, M. Gorman, L. Martin, K. M. Howarth, S. V. Hodgson, K. Kaur, J. Taylor, and I. P. M. Tomlinson, "Dna polymerase epsilon and delta exonuclease domain mutations in endometrial cancer," *Human Molecular genetics*, vol. 22, no. 14, pp. 2820–2828, 2013. 20, 22, 23, 29, 30, 31, 97, 112, 136, 166

[159] K. Murphy, H. Darmawan, A. Schultz, E. F. da Silva, and L. J. Reha-Krantz, "A method to select for mutator dna polymerase delta in saccharomyces cerevisiae," *Genome*, vol. 49, no. 4, pp. 403–410, 2006. 20, 21, 22, 33

[160] A. K. Abdus Sattar, T. C. Lin, C. Jones, and W. H. Konigsberg, "Functional consequences and exonuclease kinetic parameters of point mutations in bacteriophage t4 dna polymerase," *Biochemistry*, vol. 35, no. 51, pp. 16621–16629, 1996. 21, 22

[161] D. P. Kane and P. V. Shcherbakova, "A common cancer-associated dna polymerase epsilon mutation causes an exceptionally strong mutator phenotype, indicating fidelity defects distinct from loss of proofreading," *Cancer Research*, vol. 74, no. 7, pp. 1895–1901, 2014. 21, 22, 26, 27, 129, 130

[162] E. Shinbrot, E. E. Henninger, N. Weinhold, K. R. Covington, a. Y. Gksenin, N. Schultz, H. Chao, H. Doddapaneni, D. M. Muzny, R. a. Gibbs, C. Sander, Z. F.

Pursell, and D. a. Wheeler, "Exonuclease mutations in dna polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication," *Genome Research*, vol. 24, no. 11, pp. 1740–1750, 2014. 21, 22, 23, 25, 26, 27, 29, 30, 31, 32, 33, 76, 129, 130, 136, 166, 168, 179

[163] S. R. Barbari, D. P. Kane, E. A. Moore, and P. V. Shcherbakova, "Functional analysis of cancer-associated dna polymerase epsilon variants in saccharomyces cerevisiae," *G3 (Bethesda)*, vol. 8, no. 3, pp. 1019–1029., 2018. 21, 22, 26, 27, 30, 32, 76, 77, 130, 181

[164] A. Shlien, B. B. Campbell, R. D. Borja, L. B. Alexandrov, D. Merico, D. Wedge, P. V. Loo, P. S. Tarpey, P. Coupland, S. Behjati, A. Pollett, T. Lipman, A. Heidari, S. Deshmukh, N. Avitzur, B. Meier, M. Gerstung, Y. Hong, D. M. Merino, M. Ramakrishna, M. Remke, R. Arnold, G. B. Panigrahi, N. P. Thakkar, K. P. Hodel, E. E. Henninger, a. Y. Gksenin, D. Bakry, G. S. Charames, H. Druker, J. Lerner-ellis, M. Mistry, P. Dirks, A. Huang, S. W. Scherer, R. G. Grundy, C. Durno, M. Aronson, A. Gartner, M. S. Meyn, M. D. Taylor, Z. F. Pursell, C. E. Pearson, D. Malkin, P. A. Futreal, M. R. Stratton, and E. Bouffet, "Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers," *Nature Genetics*, vol. 47, no. 3, pp. 257–262, 2015. 22, 23, 25, 26, 28, 31, 131, 165, 166, 176, 186

[165] A. M. L. Jansen, T. van Wezel, B. E. W. M. van den Akker, M. Ventayol Garcia, D. Ruano, C. M. J. Tops, A. Wagner, T. G. W. Letteboer, E. B. Gomez-Garcia, P. Devilee, J. T. Wijnen, F. J. Hes, and H. Morreau, "Combined mismatch repair and pole/pold1 defects explain unresolved suspected lynch syndrome cancers," *Eur J Hum Genet*, vol. 24, no. 7, pp. 1089–1092, 2015. 22, 25

[166] E. Domingo, L. Freeman-Mills, E. Rayner, M. Glaire, S. Briggs, L. Vermeulen, E. Fessler, J. P. Medema, A. Boot, H. Morreau, T. van Wezel, G.-J. Liefers, R. A. Lothe, S. A. Danielsen, A. Sveen, A. Nesbakken, I. Zlobec, A. Lugli, V. H. Koelzer, M. D. Berger, S. Castellv-Bel, J. Muoz, M. de Bruyn, H. W. Nijman, M. Novelli, K. Lawson, D. Oukrif, E. Frangou, P. Dutton, S. Tejpar, M. Delorenzi, R. Kerr, D. Kerr, I. Tomlinson, and D. N. Church, "Somatic pole proofreading domain mutation, immune response, and prognosis in colorectal cancer: a retrospective, pooled biomarker study," *The Lancet Gastroenterology Hepatology*, vol. 1, no. 3, pp. 207–216, 2016. 22, 23, 28, 54, 166, 167, 168, 177, 179, 183, 184

[167] S.-M. Ahn, A. A. Ansari, J. Kim, D. Kim, S.-M. Chun, J. Kim, T. W. Kim, I. Park, C.-S. Yu, and S. J. Jang, "The somatic pole p286r mutation defines a unique subclass of colorectal cancer featuring hypermutation, representing a potential genomic biomarker for immunotherapy," *Oncotarget*, vol. 7, no. 42, pp. 68638–68649, 2016. 22, 28, 166

[168] M. Giannakis, X. J. Mu, S. A. Shukla, Z. R. Qian, O. Cohen, R. Nishihara, S. Bahl, Y. Cao, A. Amin-Mansour, M. Yamauchi, Y. Sukawa, C. Stewart, M. Rosenberg, K. Mima, K. Inamura, K. Nosho, J. A. Nowak, M. S. Lawrence, E. L. Giovannucci, A. T. Chan, K. Ng, J. A. Meyerhardt, E. M. Van Allen, G. Getz, S. B. Gabriel, E. S. Lander, C. J. Wu, C. S. Fuchs, S. Ogino, and L. A. Garraway, "Genomic correlates of immune-cell infiltrates in colorectal carcinoma," *Cell Reports*, vol. 15, no. 4, pp. 857–865, 2016. 22

[169] N. Kothari, J. K. Teer, A. M. Abbott, T. Srikumar, Y. Zhang, S. J. Yoder, A. S. Brohl, R. D. Kim, D. R. Reed, and D. Shibata, "Increased incidence of fbxw7 and pole proofreading domain mutations in young adult colorectal cancers," *Cancer*, vol. 122, no. 18, pp. 2828–2835, 2016. 22

[170] Z. K. Stadler, F. Battaglin, S. Middha, J. F. Hechtman, C. Tran, A. Cercek, R. Yaeger, N. H. Segal, A. M. Varghese, D. L. Reidy-Lagunes, N. E. Kemeny, E. E. Salo-Mullen, A. Ashraf, M. R. Weiser, J. Garcia-Aguilar, M. E. Robson, K. Offit, M. E. Arcila, M. F. Berger, J. Shia, D. B. Solit, and L. B. Saltz, "Reliable detection of mismatch repair deficiency in colorectal cancers using mutational load in next-generation sequencing panels," *Journal of Clinical Oncology*, vol. 34, no. 18, pp. 2141–2147, 2016. 22

[171] B. B. Campbell, N. Light, D. Fabrizio, M. Zatzman, F. Fuligni, R. de Borja, S. Davidson, M. Edwards, J. A. Elvin, K. P. Hodel, W. J. Zahurancik, Z. Suo, T. Lipman, K. Wimmer, C. P. Kratz, D. C. Bowers, T. W. Laetsch, G. P. Dunn, T. M. Johanns, M. R. Grimmer, I. V. Smirnov, V. Larouche, D. Samuel, A. Bronsema, M. Osborn, D. Stearns, P. Raman, K. A. Cole, P. B. Storm, M. Yalon, E. Opocher, G. Mason, G. A. Thomas, M. Sabel, B. George, D. S. Ziegler, S. Lindhorst, V. M. Issai, S. Constantini, H. Toledano, R. Elhasid, R. Farah, R. Dvir, P. Dirks, A. Huang, M. A. Galati, J. Chung, V. Ramaswamy, M. S. Irwin, M. Aronson, C. Durno, M. D. Taylor, G. Rechavi, J. M. Maris, E. Bouffet, C. Hawkins, J. F. Costello, M. S. Meyn, Z. F. Pursell, D. Malkin, U. Tabori, and A. Shlien, "Comprehensive analysis of hypermutation in human cancer," *Cell*, vol. 171, no. 5, pp. 1042–1056.e10, 2017. 22, 23, 26, 31, 32, 33, 131, 166, 176

[172] K. Hatakeyama, K. Ohshima, T. Nagashima, S. Ohnami, S. Ohnami, M. Serizawa, Y. Shimoda, K. Maruyama, Y. Akiyama, K. Urakami, M. Kusuhara, T. Mochizuki, and K. Yamaguchi, "Molecular profiling and sequential somatic mutation shift in hypermutator tumours harbouring pole mutations," *Scientific Reports*, vol. 8, no. 1, p. 8700, 2018. 22, 28

[173] R. Yaeger, W. K. Chatila, M. D. Lipsyc, J. F. Hechtman, A. Cercek, F. Sanchez-Vega, G. Jayakumaran, S. Middha, A. Zehir, M. T. A. Donoghue, D. You, A. Viale, N. Kemeny, N. H. Segal, Z. K. Stadler, A. M. Varghese, R. Kundra, J. Gao, A. Syed, D. M. Hyman, E. Vakiani, N. Rosen, B. S. Taylor, M. Ladanyi, M. F. Berger, D. B. Solit, J. Shia, L. Saltz, and N. Schultz, "Clinical sequencing defines the genomic landscape of metastatic colorectal cancer," *Cancer Cell*, vol. 33, no. 1, pp. 125–136.e3, 2018. 22, 23

[174] A. Zehir, R. Benayed, R. H. Shah, A. Syed, S. Middha, H. R. Kim, P. Srinivasan, J. Gao, D. Chakravarty, S. M. Devlin, M. D. Hellmann, D. A. Barron, A. M. Schram, M. Hameed, S. Dogan, D. S. Ross, J. F. Hechtman, D. F. DeLair, J. Yao, D. L. Mandelker, D. T. Cheng, R. Chandramohan, A. S. Mohanty, R. N. Ptashkin, G. Jayakumaran, M. Prasad, M. H. Syed, A. B. Rema, Z. Y. Liu, K. Nafa, L. Borsu, J. Sadowska, J. Casanova, R. Bacares, I. J. Kiecka, A. Razumova, J. B. Son, L. Stewart, T. Baldi, K. A. Mullaney, H. Al-Ahmadie, E. Vakiani, A. A. Abeshouse, A. V. Penson, P. Jonsson, N. Camacho, M. T. Chang, H. H. Won, B. E. Gross, R. Kundra, Z. J. Heins, H. W. Chen, S. Phillips, H. Zhang, J. Wang, A. Ochoa, J. Wills, M. Eubank, S. B.

Thomas, S. M. Gardos, D. N. Reales, J. Galle, R. Durany, R. Cambria, W. Abida, A. Cercek, D. R. Feldman, M. M. Gounder, A. A. Hakimi, J. J. Harding, G. Iyer, Y. Y. Janjigian, E. J. Jordan, C. M. Kelly, M. A. Lowery, L. G. T. Morris, A. M. Omuro, N. Raj, P. Razavi, A. N. Shoushtari, N. Shukla, T. E. Soumerai, A. M. Varghese, R. Yaeger, J. Coleman, B. Bochner, G. J. Riely, L. B. Saltz, H. I. Scher, P. J. Sabbatini, M. E. Robson, D. S. Klimstra, B. S. Taylor, J. Baselga, N. Schultz, D. M. Hyman, M. E. Arcila, D. B. Solit, M. Ladanyi, and M. F. Berger, "Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients," *Nature Medicine*, vol. 23, no. 6, pp. 703–713, 2017. 22, 23

[175] Y. R. Hussein, B. Weigelt, D. A. Levine, J. K. Schoolmeester, L. N. Dao, B. L. Balzer, G. Liles, B. Karlan, M. Kbel, C.-H. Lee, and R. A. Soslow, "Clinicopathological analysis of endometrial carcinomas harboring somatic pole exonuclease domain mutations," *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, vol. 28, no. 4, pp. 505–514, 2014. 22, 28, 167

[176] S. Jones, N. Stransky, C. L. McCord, E. Cerami, J. Lagowski, D. Kelly, S. V. Angiuoli, M. Sausen, L. Kann, M. Shukla, R. Makar, L. D. Wood, L. A. Diaz Jr, C. Lengauer, and V. E. Velculescu, "Genomic analyses of gynaecologic carcinosarcomas reveal frequent mutations in chromatin remodelling genes," *Nature Communications*, vol. 5, p. 5006, 2014. 22, 23, 169

[177] C. C. Billingsley, D. E. Cohn, D. G. Mutch, J. A. Stephens, A. A. Suarez, and P. J. Goodfellow, "Polymerase epsilon (pole) mutations in endometrial cancer: Clinical outcomes and implications for lynch syndrome testing," *Cancer*, vol. 121, no. 3, pp. 386–394, 2014. 22, 23, 28

[178] B. Meng, L. N. Hoang, J. B. McIntyre, M. A. Duggan, G. S. Nelson, C.-H. Lee, and M. Kbel, "Pole exonuclease domain mutation predicts long progression-free survival in grade 3 endometrioid carcinoma of the endometrium," *Gynecologic oncology*, vol. 134, no. 1, pp. 15–19, 2014. 22, 28, 167

[179] E. Stelloo, T. Bosse, R. A. Nout, H. J. MacKay, D. N. Church, H. W. Nijman, A. Leary, R. J. Edmondson, M. E. Powell, E. J. Crosbie, H. C. Kitchener, L. Mileshkin, P. M. Pollock, V. T. Smit, and C. L. Creutzberg, "Refining prognosis and identifying targetable pathways for high-risk endometrial cancer; a transportec initiative," *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, vol. 28, no. 6, pp. 836–844, 2015. 22, 23, 28, 167

[180] D. N. Church, E. Stelloo, R. a. Nout, N. Valtcheva, J. Depreeuw, N. Ter Haar, A. Noske, F. Amant, I. P. M. Tomlinson, P. J. Wild, D. Lambrechts, I. M. Jrgenliemk-Schulz, J. J. Jobsen, V. T. H. B. M. Smit, C. L. Creutzberg, and T. Bosse, "Prognostic significance of pole proofreading mutations in endometrial cancer," *Journal of the National Cancer Institute*, vol. 107, no. 1, pp. 1–8, 2015. 22, 28, 167, 183

[181] E. Z. Erson-Omay, a. O. Ca layan, N. Schultz, N. Weinhold, S. B. Omay, K. Ozduman, Y. Koksal, J. Li, A. Serin Harmanc, V. Clark, G. Carrion-Grant, J. Baranoski, C. Ca lar, T. Barak, S. Co kun, B. Baran, D. Kose, J. Sun, M. Bak rc o lu, J. Moliterno Gunel, M. N. Pamir, K. Mishra-Gorur, K. Bilguvar, K. Yasuno, A. Vortmeyer,

a. J. Huttner, C. Sander, and M. Gunel, "Somatic pole mutations cause an ultramutated giant cell high-grade glioma subtype with better prognosis," *Neuro-Oncology*, vol. 17, no. 10, pp. 1356–1364, 2015. 22, 23, 28, 165, 166, 167, 176, 179

[182] L. N. Hoang, M. K. McConechy, M. Kbel, M. Anglesio, J. Senz, M. Maassen, S. Kommoss, B. Meng, L. Postovit, L. E. Kelemen, A. Staebler, S. Brucker, B. Krmer, J. N. McAlpine, C. B. Gilks, D. G. Huntsman, and C.-H. Lee, "Polymerase epsilon exonuclease domain mutations in ovarian endometrioid carcinoma," *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society*, vol. 25, no. 7, pp. 1187–1193, 2015. 22, 28

[183] A. G. Robertson, J. Kim, H. Al-Ahmadie, J. Bellmunt, G. Guo, A. D. Cherniack, T. Hinoue, P. W. Laird, K. A. Hoadley, R. Akbani, M. A. A. Castro, E. A. Gibb, R. S. Kanchi, D. A. Gordenin, S. A. Shukla, F. Sanchez-Vega, D. E. Hansel, B. A. Czerniak, V. E. Reuter, X. Su, B. de Sa Carvalho, V. S. Chagas, K. L. Mungall, S. Sadeghi, C. S. Pedamallu, Y. Lu, L. J. Klimczak, J. Zhang, C. Choo, A. I. Ojesina, S. Bullman, K. M. Leraas, T. M. Lichtenberg, C. J. Wu, N. Schultz, G. Getz, M. Meyerson, G. B. Mills, D. J. McConkey, J. N. Weinstein, D. J. Kwiatkowski, and S. P. Lerner, "Comprehensive molecular characterization of muscle-invasive bladder cancer," *Cell*, vol. 171, no. 3, pp. 540–556.e25, 2017. 22

[184] M. van de Wetering, H. E. Francies, J. M. Francis, G. Bounova, F. Iorio, A. Pronk, W. van Houdt, J. van Gorp, A. Taylor-Weiner, L. Kester, A. McLaren-Douglas, J. Blokker, S. Jaksani, S. Bartfeld, R. Volckman, P. van Sluis, V. S. W. Li, S. Seepo, C. Sekhar Pedamallu, K. Cibulskis, S. L. Carter, A. McKenna, M. S. Lawrence, L. Lichtenstein, C. Stewart, J. Koster, R. Versteeg, A. van Oudenaarden, J. Saez-Rodriguez, R. G. J. Vries, G. Getz, L. Wessels, M. R. Stratton, U. McDermott, M. Meyerson, M. J. Garnett, and H. Clevers, "Prospective derivation of a living organoid biobank of colorectal cancer patients," *Cell*, vol. 161, no. 4, pp. 933–945, 2015. 22

[185] Y. Zou, F.-Y. Liu, H. Liu, F. Wang, W. Li, M.-Z. Huang, Y. Huang, X.-Q. Yuan, X.-Y. Xu, O.-P. Huang, and M. He, "Frequent pole1 p.s297f mutation in chinese patients with ovarian endometrioid carcinoma," *Mutation research. Fundamental and molecular mechanisms of mutagenesis*, vol. 761, pp. 49–52, 2014. 22, 23

[186] D. Temko, I. C. Van Gool, E. Rayner, M. Glaire, S. Makino, M. Brown, L. Chegwidden, C. Palles, J. Depreeuw, A. Beggs, C. Stathopoulou, J. Mason, A. M. Baker, M. Williams, V. Cerundolo, M. Rei, J. C. Taylor, A. Schuh, A. Ahmed, F. Amant, D. Lambrechts, V. T. Smit, T. Bosse, T. A. Graham, D. N. Church, and I. Tomlinson, "Somatic pole exonuclease domain mutations are early events in sporadic endometrial and colorectal carcinogenesis, determining driver mutational landscape, clonal neoantigen burden and immune response," *Journal of Pathology*, vol. 245, no. 3, pp. 283–296, 2018. 22, 166, 167, 168, 171, 176

[187] A. Wong, C. H. Kuick, W. L. Wong, J. M. Tham, S. Mansor, E. Loh, S. Jain, N. N. Vikas, S. H. Tan, S. H. Chan, S. T. Li, S. H. Chew, W. Hong, and J. Ngeow, "Mutation spectrum of pole and pold1 mutations in south east asian women presenting with grade 3 endometrioid endometrial carcinomas," *Gynecologic Oncology*, vol. 141, no. 1, pp. 113–120, 2016. 22, 23, 167

[188] A. Talhouk, M. K. McConechy, S. Leung, H. H. Li-Chang, J. S. Kwon, N. Melnyk, W. Yang, J. Senz, N. Boyd, A. N. Karnezis, D. G. Huntsman, C. B. Gilks, and J. N. McAlpine, "A clinically applicable molecular-based classification for endometrial cancers," *Br J Cancer*, vol. 113, no. 2, pp. 299–310, 2015. 22, 28, 167

[189] S. Bellone, F. Centritto, J. Black, C. Schwab, D. English, E. Cocco, S. Lopez, E. Bonazzoli, F. Predolini, F. Ferrari, D.-A. Silasi, E. Ratner, M. Azodi, P. E. Schwartz, and A. D. Santin, "Polymerase epsilon (pole) ultra-mutated tumors induce robust tumor-specific cd4+ t cell responses in endometrial cancer patients," *Gynecologic Oncology*, vol. 138, no. 1, pp. 11–17, 2015. 22

[190] J. M. Mehnert, A. Panda, H. Zhong, K. Hirshfield, S. Damare, K. Lane, L. Sokol, M. N. Stein, L. Rodriguez-Rodriquez, H. L. Kaufman, S. Ali, J. S. Ross, D. C. Pavlick, G. Bhanot, E. P. White, R. S. DiPaola, A. Lovell, J. Cheng, and S. Ganesan, "Immune activation and response to pembrolizumab in pole-mutant endometrial cancer," *The Journal of Clinical Investigation*, vol. 126, no. 6, pp. 2334–2340, 2016. 22, 167

[191] T. C. G. A. Network, "Comprehensive molecular characterization of gastric adenocarcinoma," *Nature*, vol. 513, no. 7517, pp. 202–209, 2014. 22

[192] N. The Cancer Genome Atlas Research, J. Kim, R. Bowlby, A. J. Mungall, A. G. Robertson, R. D. Odze, A. D. Cherniack, J. Shih, C. S. Pedamallu, C. Cibulskis, A. Dunford, S. R. Meier, J. Kim, B. J. Raphael, H.-T. Wu, A. M. Wong, J. E. Willis, A. J. Bass, S. Derks, K. Garman, S. J. McCall, M. Wiznerowicz, A. Pantazi, M. Parfenov, V. Thorsson, I. Shmulevich, V. Dhankani, M. Miller, R. Sakai, K. Wang, N. Schultz, R. Shen, A. Arora, N. Weinhold, F. Snchez-Vega, D. P. Kelsen, J. Zhang, I. Felau, J. Demchok, C. S. Rabkin, M. C. Camargo, J. C. Zenklusen, J. Bowen, K. Leraas, T. M. Lichtenberg, C. Curtis, J. A. Seoane, A. I. Ojesina, D. G. Beer, M. L. Gulley, A. Pennathur, J. D. Luketich, Z. Zhou, D. J. Weisenberger, R. Akbani, J.-S. Lee, W. Liu, G. B. Mills, W. Zhang, B. J. Reid, T. Hinoue, P. W. Laird, H. Shen, M. B. Piazuelo, B. G. Schneider, M. McLellan, A. Taylor-Weiner, M. Lawrence, K. Cibulskis, C. Stewart, G. Getz, E. Lander, S. B. Gabriel, L. Ding, M. D. McLellan, C. A. Miller, E. L. Appelbaum, M. G. Cordes, C. C. Fronick, L. A. Fulton, E. R. Mardis, R. K. Wilson, H. K. Schmidt, R. S. Fulton, A. Ally, M. Balasundaram, R. Carlsen, E. Chuah, N. Dhalla, R. A. Holt, S. J. M. Jones, K. Kasaian, D. Brooks, H. I. Li, Y. Ma, M. A. Marra, M. Mayo, R. A. Moore, K. L. Mungall, J. E. Schein, *et al.*, "Integrated genomic characterization of oesophageal carcinoma," *Nature*, vol. 541, p. 169, 2017. 22, 25

[193] A. S. Morrissy, L. Garzia, D. J. H. Shih, S. Zuyderduyn, X. Huang, P. Skowron, M. Remke, F. M. G. Cavalli, V. Ramaswamy, P. E. Lindsay, S. Jelveh, L. K. Donovan, X. Wang, B. Luu, K. Zayne, Y. Li, C. Mayoh, N. Thiessen, E. Mercier, K. L. Mungall, Y. Ma, K. Tse, T. Zeng, K. Shumansky, A. J. L. Roth, S. Shah, H. Farooq, N. Kijima, B. L. Holgado, J. J. Y. Lee, S. Matan-Lithwick, J. Liu, S. C. Mack, A. Manno, K. A. Michealraj, C. Nor, J. Peacock, L. Qin, J. Reimand, A. Rolider, Y. Y. Thompson, X. Wu, T. Pugh, A. Ally, M. Bilenky, Y. S. N. Butterfield, R. Carlsen, Y. Cheng, E. Chuah, R. D. Corbett, N. Dhalla, A. He, D. Lee, H. I. Li, W. Long, M. Mayo, P. Plettner, J. Q. Qian, J. E. Schein, A. Tam, T. Wong, I. Birol, Y. Zhao, C. C. Faria, J. Pimentel, S. Nunes, T. Shalaby, M. Grotzer, I. F. Pollack, R. L. Hamilton, X.-N. Li, A. E. Bendel, D. W. Fults, A. W. Walter, T. Kumabe, T. Tominaga, V. P.

Collins, Y.-J. Cho, C. Hoffman, D. Lyden, J. H. Wisoff, J. H. Garvin, D. S. Stearns, L. Massimi, U. Schller, J. Sterba, K. Zitterbart, S. Puget, O. Ayrault, S. E. Dunn, D. P. C. Tirapelli, C. G. Carlotti, H. Wheeler, A. R. Hallahan, W. Ingram, T. J. MacDonald, J. J. Olson, E. G. Van Meir, J.-Y. Lee, K.-C. Wang, *et al.*, "Divergent clonal selection dominates medulloblastoma at recurrence," *Nature*, vol. 529, p. 351, 2016. 22

[194] A. Stenzinger, N. Pfarr, V. Endris, R. Penzel, L. Jansen, T. Wolf, E. Herpel, A. Warth, F. Klauschen, M. Kloor, W. Roth, H. Blker, J. Chang-Claude, H. Brenner, M. Hoffmeister, and W. Weichert, "Mutations in pole and survival of colorectal cancer patients - link to disease stage and treatment," *Cancer Medicine*, vol. 3, no. 6, pp. 1527–1538, 2014. 22, 28, 169

[195] M. Kbel, B. Meng, L. N. Hoang, N. Almadani, X. Li, R. A. Soslow, C. B. Gilks, and C.-H. Lee, "Molecular analysis of mixed endometrial carcinomas shows clonality in most cases," *The American journal of surgical pathology*, vol. 40, no. 2, pp. 166–180, 2016. 22

[196] A. Maria, G. K. Chetan, M. K. Sibin, T. McKee, D. Merkler, R. K. V. L. Narasinga, P. Ribaux, J.-L. Blouin, P. Makrythanasis, V. B. Seplyarskiy, S. E. Antonarakis, and S. I. Nikolaev, "Germline pms2 and somatic pole exo mutations cause hypermutability of the leading dna strand in biallelic mismatch repair deficiency syndrome brain tumors," *The Journal of Pathology*, vol. 243, no. 3, pp. 331–341, 2017. 22, 166

[197] S. R. Barbari and P. V. Shcherbakova, "Replicative dna polymerase defects in human cancers: Consequences, mechanisms, and implications for therapy," *DNA Repair*, vol. 56, pp. 16–25, 2017. 20, 163

[198] S. Zhao, M. Choi, J. D. Overton, S. Bellone, D. M. Roque, E. Cocco, F. Guzzo, D. P. English, J. Varughese, S. Gasparrini, I. Bortolomai, N. Buza, P. Hui, M. Abu-Khalaf, A. Ravaggi, E. Bignotti, E. Bandiera, C. Romani, P. Todeschini, R. Tassi, L. Zanotti, L. Carrara, S. Pecorelli, D.-A. Silasi, E. Ratner, M. Azodi, P. E. Schwartz, T. J. Rutherford, A. L. Stiegler, S. Mane, T. J. Boggon, J. Schlessinger, R. P. Lifton, and A. D. Santin, "Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma," *Proceedings of the National Academy of Sciences*, vol. 110, no. 8, pp. 2916–2921, 2013. 23

[199] T. C. G. A. Network, "Widespread genetic heterogeneity in multiple myeloma: Implications for targeted therapy," *Cancer Cell*, vol. 25, no. 1, pp. 91–101, 2014. 25

[200] D. a. Korona, K. G. Lecompte, and Z. F. Pursell, "The high fidelity and unique error signature of human dna polymerase epsilon," *Nucleic Acids Research*, vol. 39, no. 5, pp. 1763–1773, 2011. 25

[201] S. Bakhsh, M. Kinloch, L. N. Hoang, R. Soslow, M. Kobel, C. H. Lee, J. N. McAlpine, M. K. McConechy, and C. B. Gilks, "Histopathological features of endometrial carcinomas associated with pole mutations: implications for decisions about adjuvant therapy," *Histopathology*, vol. 68, no. 6, pp. 916–924, 2015. 28

[202] I. C. Van Gool, J. E. H. Ubachs, E. Stelloo, C. D. de Kroon, J. J. Goeman, V. Smit, C. L. Creutzberg, and T. Bosse, "Blinded histopathological characterisation of pole exonuclease domain-mutant endometrial cancers: sheep in wolf's clothing," *Histopathology*, vol. 72, no. 2, pp. 248–258, 2018. 28

[203] P. Vande Perre, A. Siegfried, C. Corsini, D. Bonnet, C. Toulas, N. Hamzaoui, J. Selves, E. Chipoulet, J. S. Hoffmann, E. Uro-Coste, and R. Guimbaud, "Germline mutation p.n363k in pole is associated with an increased risk of colorectal cancer and giant cell glioblastoma," *Familial Cancer*, vol. 18, no. 2, pp. 173–178, 2018. 28

[204] S. Zhang, I. Mercado-Uribe, Z. Xing, B. Sun, J. Kuang, and J. Liu, "Generation of cancer stem-like cells through the formation of polyploid giant cancer cells," *Oncogene*, vol. 33, no. 1, pp. 116–128, 2014. 28

[205] I. C. Van Gool, F. a. Eggink, L. Freeman-Mills, E. Stelloo, E. Marchi, M. de Bruyn, C. Palles, R. a. Nout, C. D. de Kroon, E. M. Osse, P. Klenerman, C. L. Creutzberg, I. P. M. Tomlinson, V. T. H. B. M. Smit, H. W. Nijman, T. Bosse, and D. N. Church, "Pole proofreading mutations elicit an anti-tumor immune response in endometrial cancer," *Clinical Cancer Research*, vol. 21, no. 14, pp. 3347–3355, 2015. 28, 31, 112, 167, 177

[206] A. D. Santin, S. Bellone, F. Centritto, J. Schlessinger, and R. Lifton, "Improved survival of patients with hypermutation in uterine serous carcinoma," *Gynecologic Oncology Reports*, vol. 12, pp. 3–4, 2015. 28

[207] I. C. Van Gool, T. Bosse, and D. N. Church, "Pole proofreading mutation, immune response and prognosis in endometrial cancer," *Oncoimmunology*, vol. 5, no. 3, p. e1072675, 2016. 28, 167

[208] Z. Song, G. Cheng, C. Xu, W. Wang, Y. Shao, and Y. Zhang, "Clinicopathological characteristics of pole mutation in patients with non-small-cell lung cancer," *Lung Cancer*, vol. 118, pp. 57–61, 2018. 28

[209] N. A. Rizvi, M. D. Hellmann, A. Snyder, P. Kvistborg, V. Makarov, J. J. Havel, W. Lee, J. Yuan, P. Wong, T. S. Ho, M. L. Miller, N. Rekhtman, A. L. Moreira, F. Ibrahim, C. Bruggeman, B. Gasmi, R. Zappasodi, Y. Maeda, C. Sander, E. B. Garon, T. Merghoub, J. D. Wolchok, T. N. Schumacher, and T. A. Chan, "Mutational landscape determines sensitivity to pd-1 blockade in nonsmall cell lung cancer," *Science (New York, N.Y.)*, vol. 348, no. 6230, pp. 124–128, 2015. 28

[210] N. J. Haradhvala, J. Kim, Y. E. Maruvka, P. Polak, D. Rosebrock, D. Livitz, J. M. Hess, I. Leshchiner, A. Kamburov, K. W. Mouw, M. S. Lawrence, and G. Getz, "Distinct mutational signatures characterize concurrent loss of polymerase proofreading and mismatch repair," *Nature Communications*, vol. 9, no. 1, p. 1746, 2018. 30, 31, 32, 131, 166, 176

[211] Z. Wang, X. Wu, and E. C. Friedberg, "Dna repair synthesis during base excision repair in vitro is catalyzed by dna polymerase epsilon and is influenced by dna polymerases alpha and delta in saccharomyces cerevisiae," *Molecular Cell Biology*, vol. 13, no. 2, pp. 1051–1058, 1993. 30

[212] E. Parlanti, P. Fortini, P. Macpherson, J. Laval, and E. Dogliotti, "Base excision repair of adenine/8-oxoguanine mispairs by an aphidicolin-sensitive dna polymerase in human cell extracts," *Oncogene*, vol. 21, no. 34, pp. 5204–5212, 2002. 30

[213] Y. Matsumoto, K. Kim, and D. F. Bogenhagen, "Proliferating cell nuclear antigen-dependent abasic site repair in xenopus laevis oocytes: an alternative pathway of base excision dna repair," *Molecular Cell Biology*, vol. 14, no. 9, pp. 6187–6197, 1994. 30

[214] P. Fortini, E. Parlanti, O. M. Sidorkina, J. Laval, and E. Dogliotti, "The type of dna glycosylase determines the base excision repair pathway in mammalian cells," *Journal of Biological Chemistry*, vol. 274, no. 21, pp. 15230–15236, 1999. 30

[215] N. Sabouri and E. Johansson, "Translesion synthesis of abasic sites by yeast dna polymerase epsilon," *Journal of Biological Chemistry*, vol. 284, no. 46, pp. 31555–31563, 2009. 30

[216] R. C. Poulos, J. Olivier, and J. W. H. Wong, "The interaction between cytosine methylation and processes of dna replication and repair shape the mutational landscape of cancer genomes," *Nucleic Acids Research*, vol. 45, no. 13, pp. 7786–7795, 2017. 31

[217] M. Tomkova, M. McClellan, S. Kriaucionis, and B. Schuster-Bckler, "Dna replication and associated repair pathways are involved in the mutagenesis of methylated cytosine," *DNA Repair*, vol. 62, pp. 1–7, 2018. 31

[218] R. Vaser, S. Adusumalli, S. N. Leng, M. Sikic, and P. C. Ng, "Sift missense predictions for genomes," *Nature Protocols*, vol. 11, no. 1, pp. 1–9, 2016. 33, 97

[219] I. A. Adzhubei, S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov, and S. R. Sunyaev, "A method and server for predicting damaging missense mutations," *Nature Methods*, vol. 7, no. 4, pp. 248–249, 2010. 33, 97

[220] J. M. Schwarz, D. N. Cooper, M. Schuelke, and D. Seelow, "Mutationtaster2: mutation prediction for the deep-sequencing age," *Nature Methods*, vol. 11, no. 4, pp. 361–362, 2014. 33

[221] W. E. Glaab and K. R. Tindall, "Mutation rate at the hprt locus in human cancer cell lines with specific mismatch repair-gene defects," *Carcinogenesis*, vol. 18, no. 1, pp. 1–8, 1997. 41, 93, 94

[222] J. Nieminuszczy, R. A. Schwab, and W. Niedzwiedz, "The dna fibre technique tracking helicases at work," *Methods*, vol. 108, pp. 92–98, 2016. 43, 107

[223] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "Nih image to imagej: 25 years of image analysis," *Nature Methods*, vol. 9, no. 7, pp. 671–675, 2012. 44, 108

[224] S. J. Knight, C. M. Lese, K. S. Precht, J. Kuc, Y. Ning, S. Lucas, R. Regan, M. Brenan, A. Nicod, N. M. Lawrie, D. L. Cardy, H. Nguyen, T. J. Hudson, H. C. Riethman, D. H. Ledbetter, and J. Flint, "An optimized set of human telomere clones for studying telomere integrity and architecture," *American journal of human genetics*, vol. 67, no. 2, pp. 320–332, 2000. 44

[225] G. E. Truett, P. Heeger, R. L. Mynatt, A. A. Truett, J. A. Walker, and M. L. Warman, "Preparation of pcr-quality mouse genomic dna with hot sodium hydroxide and tris (hotshot)," *Biotechniques*, vol. 29, no. 1, pp. 52, 54, 2000. 48

[226] A. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, and S. G. Rozen, "Primer3new capabilities and interfaces," *Nucleic Acids Research*, vol. 40, no. 15, pp. e115–e115, 2012. 50, 99

[227] L. Zhang, "Immunohistochemistry versus microsatellite instability testing for screening colorectal cancer patients at risk for hereditary nonpolyposis colorectal cancer syndrome. part ii. the utility of microsatellite instability testing," *J Mol Diagn*, vol. 10, no. 4, pp. 301–307, 2008. 54

[228] K. A. Gould, C. Luongo, A. R. Moser, M. K. McNeley, N. Borenstein, A. Shedlovsky, W. F. Dove, K. Hong, W. F. Dietrich, and E. S. Lander, "Genetic evaluation of candidate genes for the mom1 modifier of intestinal neoplasia in mice," *Genetics*, vol. 144, no. 4, pp. 1777–1785, 1996. 54, 160

[229] J. A. Minton, S. E. Flanagan, and S. Ellard, "Mutation surveyor: software for dna sequence analysis," *Methods Mol Biol*, vol. 688, pp. 143–153, 2011. 57

[230] K. J. Livak and T. D. Schmittgen, "Analysis of relative gene expression data using real-time quantitative pcr and the 2deltadeltact method," *Methods*, vol. 25, no. 4, pp. 402–408, 2001. 64

[231] Y. Lallemand, V. Luria, R. Haffner-Krausz, and P. Lonai, "Maternally expressed pgk-cre transgene as a tool for early and uniform activation of the cre site-specific recombinase," *Transgenic Research*, vol. 7, no. 2, pp. 105–112, 1998. 70, 144, 204

[232] N. Barker, J. H. van Es, J. Kuipers, P. Kujala, M. van den Born, M. Cozijnsen, A. Haegebarth, J. Korving, H. Begthel, P. J. Peters, and H. Clevers, "Identification of stem cells in small intestine and colon by marker gene lgr5," *Nature*, vol. 449, no. 7165, pp. 1003–1007, 2007. 70, 151, 152, 204

[233] A. R. Moser, H. C. Pitot, and W. F. Dove, "A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse," *Science*, vol. 247, no. 4940, pp. 322–324, 1990. 70, 137, 157, 163, 204

[234] R. Rudling, A. B. Hassan, J. Kitau, N. Mandir, and R. A. Goodlad, "A simple device to rapidly prepare whole mounts of murine intestine," *Cell Proliferation*, vol. 39, no. 5, pp. 415–420, 2006. 71

[235] M. Martin, "Cutadapt removes adapter sequences from high-throughput sequencing reads," *Genome Research*, vol. 17, no. 1, p. 3, 2011. 73, 74, 101

[236] A. Wilm, P. P. Aw, D. Bertrand, G. H. Yeo, S. H. Ong, C. H. Wong, C. C. Khor, R. Petric, M. L. Hibberd, and N. Nagarajan, "Lofreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets," *Nucleic Acids Research*, vol. 40, no. 22, pp. 11189–11201, 2012. 73, 74, 98, 101

[237] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin, "The sequence alignment/map format and samtools," *Bioinformatics*, vol. 25, no. 16, pp. 2078–2079, 2009. 74, 101

[238] A. Rimmer, H. Phan, I. Mathieson, Z. Iqbal, S. R. F. Twigg, W. G. S. Consortium, A. O. M. Wilkie, G. McVean, and G. Lunter, "Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications," *Nature Genetics*, vol. 46, no. 8, pp. 912–918, 2014. 74

[239] W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek, and F. Cunningham, "The ensembl variant effect predictor," *Genome Biology*, vol. 17, no. 1, p. 122, 2016. 74, 101, 173

[240] H.-D. Li, I. Cuevas, M. Zhang, C. Lu, M. M. Alam, Y.-X. Fu, M. J. You, E. A. Akbay, H. Zhang, and D. H. Castrillon, "Polymerase-mediated ultramutagenesis in mice produces diverse cancers with high mutational load," *The Journal of Clinical Investigation*, vol. 128, no. 9, pp. 4179–4191, 2018. 76, 77, 78, 79, 135, 136, 148, 179

[241] K. P. Hodel, R. de Borja, E. E. Henninger, B. B. Campbell, N. Ungerleider, N. Light, T. Wu, K. G. LeCompte, A. Y. Goksenin, B. A. Bunnell, U. Tabori, A. Shlien, and Z. F. Pursell, "Explosive mutation accumulation triggered by heterozygous human pol epsilon proofreading-deficiency is driven by suppression of mismatch repair," *Elife*, vol. 7, p. e32692, 2018. 77, 79, 129, 131, 185, 186

[242] M. Tsuda, K. Terada, M. Ooka, K. Kobayashi, H. Sasanuma, R. Fujisawa, T. Tsurimoto, J. Yamamoto, S. Iwai, K. Kadoda, R. Akagawa, S. N. Huang, Y. Pommier, J. E. Sale, S. Takeda, and K. Hirota, "The dominant role of proofreading exonuclease activity of replicative polymerase epsilon in cellular tolerance to cytarabine (ara-c)," *Oncotarget*, vol. 8, no. 20, pp. 33457–33474, 2017. 77, 79, 80, 117, 132

[243] I. C. Van Gool, E. Rayner, E. M. Osse, R. A. Nout, C. L. Creutzberg, I. Tomlinson, D. N. Church, V. T. Smit, N. de Wind, T. Bosse, and M. Drost, "Adjuvant treatment for pole proofreading domain-mutant cancers: sensitivity to radiotherapy, chemotherapy, and nucleoside analogs," *Clinical Cancer Research*, vol. 24, no. 13, pp. 3197–3203, 2018. 77, 78, 79, 80, 117, 132

[244] C. M. Galmarini, J. R. Mackey, and C. Dumontet, "Nucleoside analogues and nucleobases in cancer treatment," *The Lancet Oncology*, vol. 3, no. 7, pp. 415–424, 2002. 80

[245] E. J. Fox and L. A. Loeb, "Lethal mutagenesis: Targeting the mutator phenotype in cancer," *Seminars in Cancer Biology*, vol. 20, no. 5, pp. 353–359, 2010. 80

[246] H. Deveau, J. E. Garneau, and S. Moineau, "Crispr/cas system and its role in phage-bacteria interactions," *Annu Rev Microbiol*, vol. 64, pp. 475–493, 2010. 80

[247] P. D. Hsu, D. A. Scott, J. A. Weinstein, F. A. Ran, S. Konermann, V. Agarwala, Y. Q. Li, E. J. Fine, X. B. Wu, O. Shalem, T. J. Cradick, L. A. Marraffini, G. Bao, and F. Zhang, "Dna targeting specificity of rna-guided cas9 nucleases," *Nature Biotechnology*, vol. 31, no. 9, pp. 827–832, 2013. 80, 81, 82

[248] P. D. Hsu, E. S. Lander, and F. Zhang, "Development and applications of crispr-cas9 for genome engineering," *Cell*, vol. 157, no. 6, pp. 1262–1278, 2014. 80, 86

[249] Z. Mao, M. Bozzella, A. Seluanov, and V. Gorbunova, "Comparison of nonhomologous end joining and homologous recombination in human cells," *DNA Repair*, vol. 7, no. 10, pp. 1765–1771, 2008. 80

[250] L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, and F. Zhang, "Multiplex genome engineering using crispr/cas systems," *Science*, vol. 339, no. 6121, pp. 819–823, 2013. 81, 82, 84, 128

[251] F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, and F. Zhang, "Genome engineering using the crispr-cas9 system," *Nature Protocols*, vol. 8, no. 11, pp. 2281–2308, 2013. 81, 129

[252] F. A. Ran, P. D. Hsu, C.-Y. Lin, J. S. Gootenberg, S. Konermann, A. E. Trevino, D. A. Scott, A. Inoue, S. Matoba, Y. Zhang, and F. Zhang, "Double nicking by rna-guided crispr cas9 for enhanced genome editing specificity," *Cell*, vol. 154, no. 6, pp. 1380–1389, 2013. 81

[253] J.-H. Oh, J.-L. Ku, K.-A. Yoon, H.-J. Kwon, W.-H. Kim, H.-S. Park, K.-S. Yeo, S.-Y. Song, J.-K. Chung, and J.-G. Park, "Establishment and characterization of 12 human colorectal-carcinoma cell lines," *International Journal of Cancer*, vol. 81, no. 6, pp. 902–910, 1999. 83

[254] J. G. Tate, S. Bamford, H. C. Jubb, Z. Sondka, D. M. Beare, N. Bindal, H. Boutselakis, C. G. Cole, C. Creatore, E. Dawson, P. Fish, B. Harsha, C. Hathaway, S. C. Jupe, C. Y. Kok, K. Noble, L. Ponting, C. C. Ramshaw, C. E. Rye, H. E. Speedy, R. Stefancsik, S. L. Thompson, S. Wang, S. Ward, P. J. Campbell, and S. A. Forbes, "Cosmic: the catalogue of somatic mutations in cancer," *Nucleic Acids Research*, vol. 47, no. D1, pp. D941–D947, 2018. 83

[255] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock, "Gene ontology: tool for the unification of biology. the gene ontology consortium," *Nature Genetics*, vol. 25, no. 1, pp. 25–29, 2000. 88

[256] M. Uhln, L. Fagerberg, B. M. Hallstrm, C. Lindskog, P. Oksvold, A. Mardinoglu, s. Sivertsson, C. Kampf, E. Sjstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, and F. Pontn, "Tissue-based map of the human proteome," *Science*, vol. 347, no. 6220, p. 1260419, 2015. 88

[257] M. Toyota, Y. Sasaki, A. Satoh, K. Ogi, T. Kikuchi, H. Suzuki, H. Mita, N. Tanaka, F. Itoh, J.-P. J. Issa, K.-W. Jair, K. E. Schuebel, K. Imai, and T. Tokino, "Epigenetic inactivation of chfr in human tumors," *Proceedings of the National Academy of Sciences*, vol. 100, no. 13, pp. 7818–7823, 2003. 92

[258] X. Yu, K. Minter-Dykhouse, L. Malureanu, W.-M. Zhao, D. Zhang, C. J. Merkle, I. M. Ward, H. Saya, G. Fang, J. van Deursen, and J. Chen, "Chfr is required for tumor suppression and aurora a regulation," *Nature Genetics*, vol. 37, no. 4, pp. 401–406, 2005. 92

[259] T. Hinoi, A. Akyol, B. K. Theisen, D. O. Ferguson, J. K. Greenson, B. O. Williams, K. R. Cho, and E. R. Fearon, "Mouse model of colonic adenoma-carcinoma progression based on somatic apc inactivation," *Cancer Research*, vol. 67, no. 20, pp. 9721–9730, 2007. 99

[260] J. T. Robinson, H. Thorvaldsdttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov, "Integrative genomics viewer," *Nature Biotechnology*, vol. 29, no. 1, pp. 24–26, 2011. 101

[261] E. Afgan, D. Baker, B. Batut, M. van den Beek, D. Bouvier, M. ech, J. Chilton, D. Clements, N. Coraor, B. A. Grning, A. Guerler, J. Hillman-Jackson, S. Hiltemann, V. Jalili, H. Rasche, N. Soranzo, J. Goecks, J. Taylor, A. Nekrutenko, and D. Blankenberg, "The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update," *Nucleic Acids Research*, vol. 46, no. W1, pp. W537–W544, 2018. 101, 102

[262] W. Nyhan, "Leschnyhan syndrome," *eLS*, 2012. 101

[263] J.-B. Renaud, C. Boix, M. Charpentier, A. De Cian, J. Cochennec, E. Duvernois-Berthet, L. Perrouault, L. Tesson, J. Edouard, R. Thinard, Y. Cherifi, S. Menoret, S. Fontanire, N. de Croz, A. Fraichard, F. Sohm, I. Anegon, J.-P. Concordet, and C. Giovannangeli, "Improved genome editing efficiency and flexibility using modified oligonucleotides with talen and crispr-cas9 nucleases," *Cell Reports*, vol. 14, no. 9, pp. 2263–2272, 2016. 129

[264] S. Lin, B. T. Staahl, R. K. Alla, and J. A. Doudna, "Enhanced homology-directed human genome engineering by controlled timing of crispr/cas9 delivery," *eLife*, vol. 3, p. e04766, 2014. 129

[265] T. Maruyama, S. K. Dougan, M. C. Truttmann, A. M. Bilate, J. R. Ingram, and H. L. Ploegh, "Increasing the efficiency of precise genome editing with crispr-cas9 by inhibition of nonhomologous end joining," *Nature Biotechnology*, vol. 33, no. 5, pp. 538–542, 2015. 129

[266] G. Li, X. Zhang, C. Zhong, J. Mo, R. Quan, J. Yang, D. Liu, Z. Li, H. Yang, and Z. Wu, "Small molecules enhance crispr/cas9-mediated homology-directed genome editing in primary cells," *Scientific Reports*, vol. 7, no. 1, p. 8943, 2017. 129

[267] K. Jayathilaka, S. D. Sheridan, T. D. Bold, K. Bochenska, H. L. Logan, R. R. Weichselbaum, D. K. Bishop, and P. P. Connell, "A chemical compound that stimulates the human homologous recombination protein rad51," *Proceedings of the National Academy of Sciences*, vol. 105, no. 41, pp. 15848–15853, 2008. 129

[268] T. M. Mertz, S. Sharma, A. Chabes, and P. V. Shcherbakova, "Colon cancer-associated mutator dna polymerase delta variant causes expansion of dntp pools increasing its own infidelity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 19, pp. E2467–E2476, 2015. 132

[269] T. Gallardo, L. Shirley, G. B. John, and D. H. Castrillon, "Generation of a germ cell-specific mouse transgenic cre line, vasa-cre," *Genesis*, vol. 45, no. 6, pp. 413–417, 2007. 135

[270] R. Fodde, R. Smits, and H. Clevers, "Apc, signal transduction and genetic instability in colorectal cancer," *Nature Reviews Cancer*, vol. 1, no. 1, pp. 55–67, 2001. 136, 137

[271] H. Clevers, K. M. Loh, and R. Nusse, "Stem cell signaling. an integral program for tissue renewal and regeneration: Wnt signaling and stem cell control," *Science*, vol. 346, no. 6205, p. 1248012, 2014. 136

[272] J.-H. Kim, S.-Y. Park, Y. Jun, J.-Y. Kim, and J.-S. Nam, "Roles of wnt target genes in the journey of cancer stem cells," *International journal of molecular sciences*, vol. 18, no. 8, p. 1604, 2017. 136

[273] O. M. Sieber, I. P. Tomlinson, and H. Lamlum, "The adenomatous polyposis coli (apc) tumour suppressor–genetics, function and disease," *Mol Med Today*, vol. 6, no. 12, pp. 462–469, 2000. 136, 138

[274] I. S. Nathke, "The adenomatous polyposis coli protein: the achilles heel of the gut epithelium," *Annu Rev Cell Dev Biol*, vol. 20, pp. 337–366, 2004. 136

[275] N. S. Fearnhead, M. P. Britton, and W. F. Bodmer, "The abc of apc," *Human Molecular Genetics*, vol. 10, no. 7, pp. 721–733, 2001. 136

[276] A. R. Moser, E. M. Mattes, W. F. Dove, M. J. Lindstrom, J. D. Haag, and M. N. Gould, "Apcmin, a mutation in the murine apc gene, predisposes to mammary carcinomas and focal alveolar hyperplasias," *Proceedings of the National Academy of Sciences*, vol. 90, no. 19, pp. 8977–8981, 1993. 138

[277] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins, "Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega," *Mol Syst Biol*, vol. 7, p. 539, 2011. 140

[278] J. Hellemans, G. Mortier, A. De Paepe, F. Speleman, and J. Vandesompele, "qbase relative quantification framework and software for management and automated analysis of real-time quantitative pcr data," *Genome Biology*, vol. 8, no. 2, p. R19, 2007. 143

[279] J. Schuijers, L. G. van der Flier, J. van Es, and H. Clevers, "Robust cre-mediated recombination in small intestinal stem cells utilizing the olfm4 locus," *Stem cell reports*, vol. 3, no. 2, pp. 234–241, 2014. 151

[280] S. M. Soyal, A. Mukherjee, K. Y. S. Lee, J. Li, H. Li, F. J. DeMayo, and J. P. Lydon, "Cre-mediated recombination in cell lineages that express the progesterone receptor," *Genesis*, vol. 41, no. 2, pp. 58–66, 2005. 153, 204

[281] C. Luongo, A. R. Moser, S. Gledhill, and W. F. Dove, "Loss of apc in intestinal adenomas from min mice," *Cancer Research*, vol. 54, no. 22, pp. 5947–5952, 1994. 156

[282] W. F. Dietrich, E. S. Lander, J. S. Smith, A. R. Moser, K. A. Gould, C. Luongo, N. Borenstein, and W. Dove, "Genetic identification of mom-1, a major modifier locus affecting min-induced intestinal neoplasia in the mouse," *Cell*, vol. 75, no. 4, pp. 631–639, 1993. 160

[283] M. MacPhee, K. P. Chepenik, R. A. Liddell, K. K. Nelson, L. D. Siracusa, and A. M. Buchberg, "The secretory phospholipase a2 gene is a candidate for the mom1 locus, a major modifier of apcmin-induced intestinal neoplasia," *Cell*, vol. 81, no. 6, pp. 957–966, 1995. 160

[284] J. G. Fox, *The mouse in biomedical research*, vol. 3 of *American College of Laboratory Animal Medicine series*. Burlington, Ma. London: Academic Press, 2nd ed., 2007. 162

[285] J. M. Ward, "Lymphomas and leukemias in mice," *Experimental and Toxicologic Pathology*, vol. 57, no. 5, pp. 377–381, 2006. 162

[286] Y. Xie, H. Yang, C. Cunanan, K. Okamoto, D. Shibata, J. Pan, D. E. Barnes, T. Lindahl, M. McIlhatton, R. Fishel, and J. H. Miller, "Deficiencies in mouse myh and ogg1 result in tumor predisposition and g to t mutations in codon 12 of the k-ras oncogene in lung tumors," *Cancer Research*, vol. 64, no. 9, pp. 3096–3102, 2004. 163, 182

[287] S. Hirano, Y. Tominaga, A. Ichinoe, Y. Ushijima, D. Tsuchimoto, Y. Honda-Ohnishi, T. Ohtsubo, K. Sakumi, and Y. Nakabeppu, "Mutator phenotype of mutyh-null mouse embryonic stem cells," *Journal of Biological Chemistry*, vol. 278, no. 40, pp. 38121–38124, 2003. 163

[288] M. T. A. Ocampo, W. Chaung, D. R. Marenstein, M. K. Chan, A. Altamirano, A. K. Basu, R. J. Boorstein, R. P. Cunningham, and G. W. Teebor, "Targeted deletion of mnth1 reveals a novel dna repair enzyme activity," *Molecular and Cellular Biology*, vol. 22, no. 17, pp. 6111–6121, 2002. 163, 182

[289] Y. Yamada, K. Hata, Y. Hirose, A. Hara, S. Sugie, T. Kuno, N. Yoshimi, T. Tanaka, and H. Mori, "Microadenomatous lesions involving loss of apc heterozygosity in the colon of adult apc(min/+) mice," *Cancer Research*, vol. 62, no. 22, pp. 6367–6370, 2002. 164, 187, 188

[290] Y. Yamada and H. Mori, "Multistep carcinogenesis of the colon in apcmin/+ mouse," *Cancer Science*, vol. 98, no. 1, pp. 6–10, 2007. 164

[291] C. V. Rao, Y. M. Yang, M. V. Swamy, T. Liu, Y. Fang, R. Mahmood, M. Jhanwar-Uniyal, and W. Dai, "Colonic tumorigenesis in bubr1+/-apcmin/+ compound mutant mice is linked to premature separation of sister chromatids and enhanced genomic instability," *Proc Natl Acad Sci U S A*, vol. 102, no. 12, pp. 4365–4370, 2005. 164

[292] K. Aoki, Y. Tamai, S. Horiike, M. Oshima, and M. M. Taketo, "Colonic polyposis caused by mtor-mediated chromosomal instability in apc+/delta716 cdx2+/ compound mutant mice," *Nature Genetics*, vol. 35, no. 4, pp. 323–330, 2003. 164

[293] J. E. Dancey, P. L. Bedard, N. Onetto, and T. J. Hudson, "The genetic basis for cancer treatment decisions," *Cell*, vol. 148, no. 3, pp. 409–420, 2012. 165

[294] M. R. Stratton, P. J. Campbell, and P. A. Futreal, "The cancer genome," *Nature*, vol. 458, no. 7239, pp. 719–724, 2009. 165

[295] L. A. Loeb, "Mutator phenotype may be required for multistage carcinogenesis," *Cancer Research*, vol. 51, no. 12, pp. 3075–3079, 1991. 166

[296] L. A. Loeb, "Human cancers express a mutator phenotype: Hypothesis, origin, and consequences," *Cancer Research*, vol. 76, no. 8, pp. 2057–2059, 2016. 166

[297] N. J. Llosa, M. Cruise, A. Tam, E. C. Wicks, E. M. Hechenbleikner, J. M. Taube, R. L. Blosser, H. Fan, H. Wang, B. S. Luber, M. Zhang, N. Papadopoulos, K. W. Kinzler, B. Vogelstein, C. L. Sears, R. A. Anders, D. M. Pardoll, and F. Housseau, "The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints," *Cancer Discov*, vol. 5, no. 1, pp. 43–51, 2015. 172

[298] L. J. Reha-Krantz, "Dna polymerase proofreading: Multiple roles maintain genome stability," *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1804, no. 5, pp. 1049–1063, 2010. 181

[299] V. Parkash, Y. Kulkarni, J. ter Beek, P. V. Shcherbakova, S. C. L. Kamerlin, and E. Johansson, "Structural consequence of the most frequently recurring cancer-associated substitution in dna polymerase epsilon," *Nature Communications*, vol. 10, no. 1, p. 373, 2019. 181

[300] X. Xing, D. P. Kane, C. R. Bulock, E. A. Moore, S. Sharma, A. Chabes, and P. V. Shcherbakova, "A recurrent cancer-associated substitution in dna polymerase epsilon produces a hyperactive enzyme," *Nature Communications*, vol. 10, no. 1, p. 374, 2019. 181

[301] L. Lipton, S. E. Halford, V. Johnson, M. R. Novelli, A. Jones, C. Cummings, E. Barclay, O. Sieber, A. Sadat, M. L. Bisgaard, S. V. Hodgson, L. A. Aaltonen, H. J. Thomas, and I. P. Tomlinson, "Carcinogenesis in myh-associated polyposis follows a distinct genetic pathway," *Cancer Research*, vol. 63, no. 22, pp. 7595–7599, 2003. 182

[302] M. A. Glaire, M. Brown, D. N. Church, and I. Tomlinson, "Cancer predisposition syndromes: lessons for truly precision medicine," *The Journal of Pathology*, vol. 241, no. 2, pp. 226–235, 2017. 183

[303] F. Sanchez-Vega, M. Mina, J. Armenia, W. K. Chatila, A. Luna, K. C. La, S. Dimitriadoy, D. L. Liu, H. S. Kantheti, S. Saghafinia, D. Chakravarty, F. Daian, Q. Gao, M. H. Bailey, W. W. Liang, S. M. Foltz, I. Shmulevich, L. Ding, Z. Heins, A. Ochoa, B. Gross, J. Gao, H. Zhang, R. Kundra, C. Kandoth, I. Bahceci, L. Dervishi, U. Dogrusoz, W. Zhou, H. Shen, P. W. Laird, G. P. Way, C. S. Greene, H. Liang, Y. Xiao, C. Wang, A. Iavarone, A. H. Berger, T. G. Bivona, A. J. Lazar, G. D. Hammer, T. Giordano, L. N. Kwong, G. McArthur, C. Huang, A. D. Tward, M. J. Frederick, F. McCormick, M. Meyerson, E. M. Van Allen, A. D. Cherniack, G. Ciriello, C. Sander, and N. Schultz, "Oncogenic signaling pathways in the cancer genome atlas," *Cell*, vol. 173, no. 2, pp. 321–337.e10, 2018. 183

[304] S. Spranger, R. Bao, and T. F. Gajewski, "Melanoma-intrinsic beta-catenin signalling prevents anti-tumour immunity," *Nature*, vol. 523, no. 7559, pp. 231–235, 2015. 183

[305] T. Zhan, N. Rindtorff, and M. Boutros, "Wnt signaling in cancer," *Oncogene*, vol. 36, no. 11, pp. 1461–1473, 2017. 183

[306] A. E. McCart, N. K. Vickaryous, and A. Silver, "Apc mice: Models, modifiers and mutants," *Pathology - Research and Practice*, vol. 204, no. 7, pp. 479–490, 2008. 188

[307] T. Sato, D. E. Stange, M. Ferrante, R. G. J. Vries, J. H. van Es, S. van den Brink, W. J. van Houdt, A. Pronk, J. van Gorp, P. D. Siersema, and H. Clevers, "Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and barrett's epithelium," *Gastroenterology*, vol. 141, no. 5, pp. 1762–1772, 2011. 188

[308] M. Boretto, B. Cox, M. Noben, N. Hendriks, A. Fassbender, H. Roose, F. Amant, D. Timmerman, C. Tomassetti, A. Vanhie, C. Meuleman, M. Ferrante, and H. Vankelecom, "Development of organoids from mouse and human endometrium showing endometrial epithelium physiology and long-term expandability," *Development*, vol. 144, no. 10, pp. 1775–1786, 2017. 188

[309] L. Broutier, A. Andersson-Rolf, C. J. Hindley, S. F. Boj, H. Clevers, B.-K. Koo, and M. Huch, "Culture and establishment of self-renewing human and mouse adult liver and pancreas 3d organoids and their genetic manipulation," *Nature Protocols*, vol. 11, p. 1724, 2016. 188

[310] A. D. DeWard, J. Cramer, and E. Lagasse, "Cellular heterogeneity in the mouse esophagus implicates the presence of a nonquiescent epithelial stem cell population," *Cell Reports*, vol. 9, no. 2, pp. 701–711, 2014. 188

[311] J. Lee, R. Bscke, P. C. Tang, B. H. Hartman, S. Heller, and K. R. Koehler, "Hair follicle development in mouse pluripotent stem cell-derived skin organoids," *Cell Reports*, vol. 22, no. 1, pp. 242–254, 2018. 188

[312] M. Jager, F. Blokzijl, V. Sasselli, S. Boymans, R. Janssen, N. Besselink, H. Clevers, R. van Boxtel, and E. Cuppen, "Measuring mutation accumulation in single human adult stem cells by whole-genome sequencing of organoid cultures," *Nature Protocols*, vol. 13, no. 1, pp. 59–78, 2018. 188, 189, 190