# Defects in Replicative DNA Polymerases Linked to Cancer Predisposition and Tumour Development




Sibyl Bertrand

Christ Church College

University of Oxford

A thesis submitted for the degree of

*Doctor of Philosophy*

Trinity 2020

To C.

Even in your all shattered genome, unfortunately no alteration was found to be targetable. I don't know why your case touched me so much. Maybe it is because even through a file, I got to know a bit of you and who you could have become. I hope you still enjoy dancing.

*Science can flourish only in an atmosphere of free speech.*

*Creativity is the residue of wasted time.*

*Everybody is a genius. But if you judge a fish by its ability to climb a tree, it will live its whole life believing that it is stupid.*

Albert Einstein (1879-1955)

## Acknowledgements

This manuscript is the result of a collaborative effort. I would not have been able to produce it without the help, support, and trust of all whom I met during my time as a PhD student. Even if it is impossible to cite all of you, I would like to mention the ones who dealt with me almost on a daily basis. Thanks to you, I stayed in the boat, afloat, and moved on, even in the stormiest waters.

First of all, comes my supervisor P$^{r}$ Stephen Kearsey. To start with, thank you for welcoming me into your group. I did not know what was ahead when I met you. But even now, knowing, I would have taken the exact same decision. You have been badly hit by life in the past few years. This is how unfair and revolting life can be. Nonetheless, you astonish me by your unshakeable support, presence and patience when it came to move on with things. You are a true example of resilience, humility, and strength, and I am grateful I got to see you as an example of these virtues. Then comes my second supervisor P$^{r}$ Catherine Green. The Tinbergen building shut down was a true disaster, but working with you offered me the opportunity to absorb your knowledge, energy and get inspiration from your incisive critical thinking. I am truly grateful for your support, at any time, in any situation, even you being so busy. I would also like to deeply thank my College advisor P$^{r}$ Stephanie Cragg who has been a strong and unwavering support, particularly during my first year. A special thank to P$^{r}$ Sue Cotterill and D$^{r}$ Gilles Crevel for the opportunity to discover the cell-free world.

This PhD was not simple for many reasons. Nonetheless, thanks to this rather eventful course, I got to work in most of the scientific buildings around the University. Most importantly, I also got to meet great people who were part of them. In the Green

Building, my first group with Xianning, Cecilia, Johanna, Sophie, Jutta and Gijs (and not Rice). You made Hell more bearable and I love the friendship which resulted from it. In the Tinbergen Building, my second group, Nacho and Nagore. You guys also had your share of difficult time. I am amazed by your resilience in the face of adversity. You inspire me and I wish I extracted more knowledge on *pombe* from you. In the Wellcome Trust, my third group, Cath and Ross' ones. I had the privilege to expend my knowledge on DNA damage and DNA repair by attending your engaging lab meetings. Thank you, Jordan, Cathy, Hind, Kirstin, Peter, Daniela, Li-Hao, Maciej, Maria, and Marie for sharing your knowledge and insight into my project, and your resilience when it came to listen to *pombe* stuff. In the New Biochemistry building, my fourth group, Jane Mellor's one: Jack, Ülkü, Meredith, Anna, Silvia, Phil, and Harry. Thank you for the tea breaks, the lunches, the escape rooms, and your patience with my endless questions about the building facilities and managment. And last but not least, my final group, the ZRAB people. Seriously guys, you made my time here really fun! Thank you, Joana, Nacho, Andres, Jess, my alter-ego Ryan, Gerardo the matchmaker, Jack, Sarah, Connor, Thomas, Victor and all the others. Thank you for the chats, the picnics, the ultimate, the punt, the drinks, the fun. I would not have arrived into Stephen's group without one of his previous PhD students and *pombe* knowledgeable researcher: Chen-Chun. I will remember for a while you coming to me and asking if I was still looking for a new supervisor: 'I know the best one for you Sibyl'. Thank you. I also met along the way two other amazingly supportive researchers: Marketa and Michael. Marketa, thank you so much for your time, insight and support into the project. One result chapter would not have been possible without you. Michael, thank you for challenging my thinking and the

way I saw this project and science in general. Finally, Emily. Thank you for our Pol ε chats, your insight, and making me discover Oxford through the Thames.

The advantage of Oxford is the College system through which I got to meet some of my now best friends. Alina, Megan. You are so important to me. You saw me at my worst and even then, you provided me with shelter to take refuge, a confident ear to talk to, and discussions to broaden my mind. I wish we were together to celebrate this big achievement and I cannot wait for us to be reunited again! Koon, Valentina and Asmita. I am so lucky I met you. Cooking, baking, breakfasts, walks, cakes, berry picking, long philosophical, existential or trivial discussions. You provided stability and anchor when everything spiralled around. Amazing people hid all around us, not only at work or in College. Beccy. I did not know a smiley would lead me to such a funny, energetic, funky and patient person as you are. Thank you for correcting my English, and for introducing me to British habits and G&Ts. To whom I had the chance to live with and became important friends: Daniela, Miren and Arianna. We will celebrate my friends. We. Will. Celebrate. Of course, then comes my family: père, mère, frère et soeur. Always there.

The last authors are usually the ones who conducted the study. So here comes mon Tom Jedusor. Contrairement au “vrai” Tom, tu n’as pas cherché à me nuire. En bon journal intime magique, tu m’as tout simplement lu, conseillé et porté, toujours sans jugement. Imperturbable. Finally, to you, whose name appears exactly 343 times in this manuscript. I can’t thank you enough for the unbreakable support you gave me during this crazy difficult time. Not only you proofread my writing and (tried to) taught me R, you also fed me, reassured me, and did all you could to put a smile on my face. That is something to also deeply acknowledge.

## Declaration

I declare that the work presented in this thesis is wholly my own, except where stated otherwise, and that no portion of this work has been presented previously for another qualification at this or any other University.

## Nomenclature

Human genes are denoted in upper case italic lettering (e.g. *POLE*).

Human proteins are denoted in upper case lettering (e.g POLE).

Fission and budding yeast genes are in lower case italic lettering (e.g. *cdc20*).

Fission and budding yeast proteins are denoted with the first letter in upper case and the remaining letters in lowercase (e.g. Cdc20).

Species are denoted in italic lettering (e.g. *Schizosaccharomyces pombe*).

# Contents

**Acknowledgements** ..... **III**

**Declaration** ..... **VI**

**Nomenclature** ..... **VI**

**Contents** ..... **VII**

**List of Figures** ..... **XII**

**List of Tables** ..... **XV**

**List of Abbreviations** ..... **XVI**

**Abstract** ..... **XXII**

**1 Introduction** ..... **1**

**1.1 Avoiding Cancer development** ..... **1**

1.1.1 Proto-oncogene and tumour suppressor genes ..... 2

1.1.2 Molecular alterations ..... 2

1.1.3 Cell division ..... 4

1.1.4 Overcoming DNA damage in S-phase ..... 14

**1.2 DNA Polymerase ε** ..... **23**

1.2.1 A brief history of polymerases ..... 23

1.2.2 Pol ε or Pol δ – who is who? ..... 23

1.2.3 Pol ε – a complex holoenzyme ..... 24

1.2.4 Crystallographic structure of Pol ε ..... 28

1.2.5 Pol ε and cell cycle progression ..... 35

**1.3 Pol ε and Cancer** ..... **36**

1.3.1 Exonuclease domain mutations in budding yeast ..... 37
1.3.2 Exonuclease domain mutations in mouse ..... 37
1.3.3 Human cancers and Pol ε exonuclease domain mutations ..... 38
**1.4 Mutational Signatures ..... 50**
1.4.1 Mutational signature and Pol ε ..... 55
1.4.2 Mutational signatures and MMR ..... 57
1.4.3 Intertwined mutational processes ..... 60
**1.5 Study aims ..... 63**
**2 Materials and Methods ..... 64**
**2.1 *Schizosaccharomyces pombe* methods ..... 64**
2.1.1 Growth and maintenance ..... 64
2.1.2 Setting up crosses ..... 69
2.1.3 Tetrad dissection ..... 69
2.1.4 Random spore analysis ..... 69
2.1.5 Growth rate analysis ..... 70
2.1.6 Spot assay ..... 70
2.1.7 Light microscopy and cell length ..... 70
2.1.8 Fission yeast transformation ..... 71
2.1.9 DNA isolation from fission yeast ..... 71
2.1.10 FACS analysis ..... 72
2.1.11 Fluctuation analysis ..... 73
2.1.12 Mutation Accumulation Experiment ..... 77
**2.2 Bacterial Methods ..... 77**
2.2.1 Plasmid preparation from E. coli ..... 77
2.2.2 *Escherichia coli* transformation ..... 77
**2.3 Molecular Genetics Methods ..... 78**

2.3.1 PCR product purification ..... 79
2.3.2 DNA purification through gel extraction ..... 79
2.3.3 Enzymatic digestion reaction ..... 80
2.3.4 DNA ligation ..... 80
2.3.5 DNA analysis by gel electrophoresis ..... 80
2.3.6 Colony PCR ..... 80
2.3.7 3HA Tagging Process ..... 81
2.3.8 Sanger Sequencing ..... 82
2.3.9 Whole Genome Sequencing ..... 82
**2.4 Biochemistry Methods ..... 83**
2.4.1 TriChloroAcetic protein extraction ..... 83
2.4.2 SDS PAGE ..... 84
2.4.3 Western Blot ..... 84
2.4.4 dNTPs extraction and measurement ..... 85
**2.5 *In vitro* methods ..... 86**
2.5.1 Extension assay ..... 86
2.5.2 Exonuclease assay ..... 86
2.5.3 DNA Electrophoresis in denaturation conditions ..... 87
**2.6 Whole Genome Sequencing Analysis ..... 87**
**3 Phenotypic Characterisation of Novel Cancer Related Germline *POLE*$^{EDM}$ variants in Fission Yeast ..... 89**
**3.1 Introduction ..... 89**
**3.2 Results ..... 91**
3.2.1 Construction of germline *cdc20*$^{EDM}$ variants in *S. pombe* ..... 91
3.2.2 Newly generated *cdc20*$^{EDM}$ variants do not promote a diploid status ..... 97
3.2.3 Germline *cdc20*$^{EDM}$ variants show protein level expression similar to wild-type .. 99

3.2.4 Some *cdc20*$^{EDM}$ variants exhibit an increase in mutation rate ..... 100

3.2.5 Some germline *cdc20*$^{EDM}$ variants exhibit abnormally shaped and longer cells similar to the somatic *cdc20*$^{P287R}$ ..... 102

3.2.6 Germline *cdc20*$^{N364K}$ and somatic *cdc20*$^{P287R}$ exhibit a growth defect rate ..... 105

**3.3 Discussion ..... 107**

**4 Mechanism of hypermutation driven by *cdc20*$^{EDM}$ in fission yeast ..... 112**

**4.1 Introduction ..... 112**

**4.2 Results ..... 114**

4.2.1 *In vitro* study of human *POLE*$^{EDM}$ variants ..... 114

4.2.2 Do *cdc20*$^{EDM}$ strains that show a cell cycle defect rely on Rad3 for survival? ..... 119

4.2.3 Response of *cdc20*$^{EDM}$ mutants to DNA-damaging agents ..... 126

4.2.4 dNTP pools depletion ..... 132

4.2.5 Is Pol κ responsible for the *cdc20*$^{EDM}$ hypermutated phenotype? ..... 139

**4.3 Discussion ..... 141**

**5 Deep-Sequencing Analysis of Mutation Spectrum Caused by Novel Cancer Related Germline *POLE*$^{EDM}$ variants in Fission Yeast ..... 147**

**5.1 Introduction ..... 147**

**5.2 Results ..... 150**

5.2.1 *cdc20*$^{EDMs}$ and MMR deficiency ..... 150

5.2.2 Mutation Accumulation experiment ..... 151

5.2.3 Assessment of sequencing quality ..... 153

5.2.4 Mutation Burden ..... 163

5.2.5 Change in mutation incidence with increasing number of generations ..... 166

5.2.6 Mutational pattern ..... 171

**5.3 Discussion ..... 177**

**6 Discussion ..... 182**

**7 References ..... 191**

**Appendices ..... 210**

**7.1 Appendix I: Media Preparation ..... 210**

7.1.1 Luria Bertani (LB) ..... 210

7.1.2 Yeast Extract plus Supplement (YES) ..... 210

7.1.3 Pombe Glutamate Medium (PMG) ..... 210

7.1.4 Edinburgh minimal medium (EMM) ..... 211

7.1.5 Edinburgh minimal medium -nitrogen (EMM-N) ..... 211

7.1.6 Salt stock solution (50X) ..... 212

7.1.7 Vitamins stock solution (1000X) ..... 212

7.1.8 Minerals stock solution (1000X) ..... 212

## List of Figures

Figure 1.1 Mismatch repair pathway ..... 18

Figure 1.2 Structure of Pol **ε** holoenzyme ..... 30

Figure 1.3 Structure of Pol ε catalytic subunit ..... 32

Figure 1.4 Structure of Pol ε exonuclease domain highlighting the studied variants ..... 43

Figure 1.5 Studied *POLE*$^{EDMs}$ are highly conserved across species ..... 44

Figure 1.6 Mutational signature extraction, or the 'Cocktail Party Problem' ..... 53

Figure 1.7 Mutational signature 10 – Before and after ..... 57

Figure 1.8 Mutational signatures related to MMR-deficiency – Before and after ..... 59

Figure 1.9 Newly characterised, associated, and intertwined mutational signatures ..... 62

Figure 3.1 PCR-based site-directed mutagenesis process - Mutation incorporation ..... 94

Figure 3.2 PCR-based site-directed mutagenesis process – Cloning and transformation ..... 95

Figure 3.3 Sub-cloning process ..... 97

Figure 3.4 Ploidy status of newly generated strains ..... 98

Figure 3.5 Expression level of Cdc20 in Cdc20-3HA variant strains ..... 100

Figure 3.6 Fluctuation analysis - Canavanine assay ..... 101

Figure 3.7 Cell length of Cdc20$^{EDM}$ variant strains ..... 103

Figure 3.8 c*dc20*$^{EDM}$ variants phenotype ..... 104

Figure 3.9 Growth rate analysis of *cdc20*$^{EDM}$ variant strains ..... 106

Figure 4.1 *In vitro* study of human *POLE*$^{EDM}$ variants – holoenzyme activity ..... 116

Figure 4.2 *In vitro* study of human *POLE*$^{EDM}$ variants – Exonuclease activity ..... 118

Figure 4.3 Analysis of *cdc20*$^{EDM}$ *rad3*△ double mutants ..... 121

Figure 4.4 Analysis of *cdc20*$^{EDM}$ *rad3ts* double mutants – Spot assay ..... 123

Figure 4.5 Analysis of *cdc20*$^{EDM}$ *rad3ts* double mutants – Growth curve ..... 125

Figure 4.6 Temperature sensitivity of *cdc20*$^{EDM}$ strains ..... 127

Figure 4.7 Response of *cdc20*$^{EDM}$ strains to DNA double-strand breaks ..... 128

Figure 4.8 Response of *cdc20*$^{EDM}$ strains to UV-C radiation ..... 129

Figure 4.9 Response of *cdc20*$^{EDM}$ strains to 4- NQO ..... 130

Figure 4.10 Response of *cdc20*$^{EDM}$ strains to camptothecin ..... 131

Figure 4.11 Response of *cdc20*$^{EDM}$ strains to MMS ..... 132

Figure 4.12 Response of *cdc20*$^{EDM}$ strains to HU ..... 133

Figure 4.13 Study of *cdc20*$^{EDM}$ variants dNTP levels ..... 135

Figure 4.14 Study of *cdc20*$^{EDM}$ variants dNTP levels – Single dNTPs ..... 136

Figure 4.15 *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ might be synthetic lethal with *cdc22*$^{D57N}$ ..... 138

Figure 4.16 Pol κ is partially responsible for the increase in mutation rate in *cdc20*$^{N364K}$ variant ..... 140

Figure 5.1 *cdc20*$^{EDM}$ combined with *msh2Δ* ..... 151

Figure 5.2 Mutation accumulation experiment ..... 152

Figure 5.3 Sequencing quality check – Per base sequence quality ..... 155

Figure 5.4 Sequencing quality check – Per sequence quality sequence ..... 156

Figure 5.5 Sequencing quality check – Per base sequence content ..... 158

Figure 5.6 Sequencing quality check – Per sequence GC content ..... 159

Figure 5.7 Sequencing quality check – Per base N content ..... 160

Figure 5.8 Sequence quality check – Sequence length distribution ..... 161

Figure 5.9 Sequencing quality check – Sequence duplication levels ..... 162

Figure 5.10 SNVs discrepancies between reference and parental strain ..... 164

Figure 5.11 Indels discrepancies between reference and parental strain.....165

Figure 5.12 SNVs arising along generations .....167

Figure 5.13 Indels arising along generations.....169

Figure 5.14 Mutational Spectrum.....172

Figure 5.15 Mutational patterns.....175

Figure 5.16 Similarities between mutational patterns and COSMIC signatures.....176

## List of Tables

Table 1.1 Nomenclature Pol ε genes.....25

Table 2.1 Strains.....65

Table 2.2 Primers.....78

Table 2.3 Plasmids.....79

Table 2.4 Antibodies.....83

Table 3.1 Cdc20$^{EDMs}$ cloning.....93

Table 4.1 *cdc20*$^{EDM}$ variants dNTP levels.....135

Table 5.1 Mutation rate and fold-change – SNV.....168

Table 5.2 Mutation rate and fold-change – indels.....170

## List of Abbreviations

| | | |
|---|---|---|
| 4-NQO | : | 4-nitroquinoline-1-oxide |
| Å | : | angstrom |
| A, Ala | : | alanine |
| aa | : | amino acid |
| APC | : | anaphase promoting complex |
| ATP | : | adenosine triphosphate |
| ATR | : | ataxia telangectasia and RAD3-related protein |
| BER | : | base excision repair |
| bMMRD | : | biallelic mismatch repair deficiency |
| bp | : | base pair |
| C-ter | : | carboxyl-terminal |
| CDC | : | cell division cycle protein, e.g. CDC45 |
| CDK | : | cyclin dependent kinase |
| Cdt1 | : | chromatin licensing and DNA replication factor 1 |
| Cdt2 | : | denticleless protein homolog |
| CHK | : | checkpoint kinase, e.g. CHK1 |
| CMG | : | Cdc45 Mcm2-7 GINS complex |
| CORGI | : | colorectal gene identification |
| CRC | : | colorectal cancer |
| CRL | : | cullin-RING ubiquitin ligases, e.g. CRL2 |
| CTD | : | C-terminal domain |
| Ctf4 | : | chromosome transmission fidelity 4 |

| | | |
|---|---|---|
| D, Asp | : | aspartic acid |
| dATP | : | deoxyadenosine triphosphate |
| dCTP | : | deoxycytidine triphosphate |
| DDK | : | dfb4-dependent kinase |
| dG | : | deoxyguanosine |
| dGTP | : | deoxyguanosine triphosphate |
| dinB | : | damage induced B |
| DMSO | : | dimethylsulfoxide |
| DNA | : | deoxyribonucleic acid |
| dNDP | : | deoxyribonucleotides diphosphate |
| dNMP | : | deoxyribonucleotide monophosphate |
| dNTP | : | deoxyribonucleotide triphosphate |
| Dpb11 | : | DNA polymerase B 11 |
| dsDNA | : | double-stranded DNA |
| dTTP | : | deoxythymidine triphosphate |
| E, Glu | : | glutamic acid |
| *E. coli* | : | *Escherichia coli* |
| EDM | : | exonuclease domain mutation |
| EDTA | : | ethylenediamine tetraacetic acid |
| EMM | : | Edinburgh minimal medium |
| F, Phe | : | phenylalanine |
| FACS | : | fluorescence-activated cell sorting |
| Fen1 | : | flap structure endonuclease 1 |
| g | : | gravity |

| | | |
|---|---|---|
| G418 | : | geneticin |
| gDNA | : | genomic DNA |
| GINS | : | go-ichi-ni-san (5, 1, 2, 3 in Japanese for Sld5, Psf1, Psf2, and Psf3) |
| HPLC | : | high performance liquid chromatography |
| HRP | : | horseradish peroxidase |
| HU | : | hydroxyurea |
| I, Ile | : | isoleucine |
| indels | : | insertion and deletion |
| iPOND | : | isolation of proteins on nascent DNA |
| K, Lys | : | lysine |
| kg | : | kilogram(s) |
| L, Leu | : | leucine |
| LB | : | Luria-Bertani |
| LiAc | : | lithium acetate |
| M | : | molar |
| M, Met | : | methionine |
| MA | : | mutation accumulation |
| Mbp | : | megabase pairs |
| MCM | : | mini chromosome maintenance, e.g. MCM2 |
| mg | : | milligram(s) |
| mL | : | millilitre(s) |
| MLH | : | mutL homolog, e.g. MLH1 |
| MMR | : | mismatch repair |
| MMS | : | methyl methane sulfonate |

| | | |
|---|---|---|
| Mrc1 | : | mediator of the replication checkpoint 1 |
| MSH | : | mutS homolog, e.g. MSH2 |
| MSI | : | microsatellite instability |
| MSS | : | microsatellite stable |
| N-ter | : | amino-terminal |
| N, Asn | : | asparagine |
| $N^2$ | : | dinitrogen |
| NDP | : | ribonucleotide diphosphate |
| NER | : | nucleotide excision repair |
| ng | : | nanogram(s) |
| NMF | : | nonnegative matrix factorisation |
| NTD | : | N-terminal domain |
| ORC | : | origin replication complex |
| P-domain | : | processivity domain |
| P, Pro | : | proline |
| PAGE | : | polyacrylamide gel electrophoresis |
| PCNA | : | proliferating cell nuclear antigen |
| PCR | : | polymerase chain reaction |
| PEG | : | polyethylene glycol |
| PIP-box | : | PCNA interacting protein box |
| PMG | : | pombe glutamate medium |
| Pms | : | postmeiotic segregation increased, e.g Pms1 |
| Pol $\alpha$ | : | polymerase alpha |
| Pol $\delta$ | : | polymerase delta |

| | | |
|---|---|---|
| Pol ε | : | polymerase epsilon |
| PPAP | : | polymerase proofreading-associated polyposis |
| Pre-IC | : | pre-initiation complex |
| Pre-LC | : | pre-loading complex |
| Pre-RC | : | pre-replication complex |
| Psf | : | partner of Sld five, e.g. Psf1 |
| PVDF | : | polyvinylidene difluoride |
| R, Arg | : | arginine |
| RAD | : | radiation resistant protein, e.g. RAD53 |
| RECQL4 | : | recQ like helicase 4 |
| RER | : | ribonucleotide excision repair |
| RFC | : | replication factor C |
| RNA | : | ribonucleic acid |
| rNMP | : | ribonucleotide monophosphate |
| RNR | : | ribonucleotide reductase |
| ROS | : | radical oxygen species |
| RPA | : | replication protein A |
| rpm | : | rotation per minute |
| S, Ser | : | serine |
| S-CDK | : | S phase-specific cyclin-dependent kinase |
| *S. cerevisiae* | : | *Saccharomyces cerevisiae* |
| *S. pombe* | : | *Schizosaccharomyces pombe* |
| SBS | : | single base substitution |
| SCF | : | SKP1-Cullin-F-Box |

| | | |
|---|---|---|
| SD | : | standard deviation |
| SDS | : | sodium dodecyl sulphate |
| Sld | : | synthetically lethal with dpb11-1, e.g. Sld5 |
| SNV | : | single nucleotide variant |
| ssDNA | : | single-stranded DNA |
| TBE | : | tris borate EDTA |
| TBST | : | tris buffer saline tween |
| TCA | : | trichloroacetic |
| TCGA | : | The Cancer Genome Atlas |
| TE | : | tris EDTA |
| TLS | : | translesion synthesis |
| TOPBP1 | : | DNA topoisomerase II binding protein 1 |
| TP53 | : | tumour protein 53 |
| UV | : | ultra-violet |
| V, Val | : | valine |
| WGS | : | whole genome sequencing |
| XRCC1 | : | X-ray repair cross complementing 1 |
| Y, Tyr | : | tyrosine |
| YES | : | yeast extract plus supplement |

## Abstract

Somatic and germline mutations in the exonuclease domain of the replicative DNA polymerase epsilon (*POLE*$^{EDM}$) can contribute to cancer development. The ultramutator phenotype observed in the *POLE*$^{EDM}$ tumours is thought to be due to alteration of the polymerase proofreading ability. Notably, while some *POLE*$^{EDM}$ variants show a mutation rate similar to a 3'-5' exonuclease-dead version, others have a mutation rate several orders of magnitude higher than exonuclease defective polymerases. Thus, it is likely that the mechanism of mutagenesis involves some processes other, or in addition to, loss of proofreading. As many *POLE*$^{EDM}$ variants are unclassified, and to better assess their pathogenicity, we constructed in fission yeast carrying germline *POLE*$^{EDM}$ variants identified in familial cases of colorectal cancer. Our combined *in vitro* and *in vivo* experimental approach confirmed that mutagenesis differs amongst *POLE*$^{EDM}$ variants, and that this capability is not due to a simple loss of exonuclease activity. Although previous studies identified an unbalanced dNTP levels as the origin of hypermutator phenotypes, our data demonstrate that this does not apply in our study. Our exploration via deep-sequencing performed on hypermutated strains confirmed that concurrent *POLE*$^{EDM}$ and mismatch repair (MMR) deficiency leads to a previously characterised specific mutational signature. Intriguingly, a second mutational signature related with coexisting Pol $\delta$ alteration and MMR deficiency was also identified. Finally, further investigation via genetic interactions lead us to suspect the translesion synthesis polymerase Pol $\kappa$ as a potential cause for the hypermutated phenotype and the mutational signatures observed. In summary, this work clarifies the tumour driver ability of germline *POLE*$^{EDMs}$, and has the potential to assist genetic counselling and *POLE*$^{EDM}$ targeted treatments.

# Chapter 1

## Introduction

### 1.1 Avoiding Cancer development

Cells are programmed to maintain an appropriate balance between quiescence, cell division and apoptosis. Disruption of the balance can lead to an uncontrolled cell proliferation, potentially giving rise to tumours. In case of malignancy, the tumour has the capacity to escape the primary neoplasm, invade other tissues, and metastasizing to other sites. Tumour development is a multistep process. After collation of decades of knowledge on neoplasia development, Douglas Hanahan and Robert Weinberg published a review of the biological capabilities acquired during cancer development – 'The Hallmarks of Cancer' (Hanahan *et al.*, 2000). These characteristics include evading apoptosis, self-sufficiency in growth signals, evading growth suppressors, enabling limitless replicative potential, sustained angiogenesis, and activating tissue invasion and metastasis. These functional capabilities are acquired via two enabling characteristics – tumour-promoting inflammation, and genome instability and mutations (Hanahan *et al.*, 2011). Possessing these competences would enable evolving populations of premalignant cells to reach the biological end-points necessary for cancer emergence.

### 1.1.1 Proto-oncogene and tumour suppressor genes

The hallmarks previously described are acquired after alteration at a molecular level of genes safeguarding the integrity of our genome. Two categories of genes play a central role in oncogenesis: proto-oncogenes and tumour suppressor genes (Pecorino, 2012). First, molecular alterations of a proto-oncogene lead to an oncogenic state, getting abnormally activated and acting in a dominant manner, as one allele altered is sufficient for oncogenic effect. Second, tumour suppressor genes can be divided into two categories: gatekeeper and caretaker (Kinzler *et al.*, 1997). Gatekeeper genes encode proteins involved in cell cycle progression, inhibiting growth or promoting apoptosis. Inactivation of gatekeeper genes enables progression through the cell cycle, even if a dysfunction is present. Caretaker genes are involved in maintaining the integrity of the genome, for instance via DNA repair, preventing accumulation of mutations. In contrast to proto-oncogenes, bi-allelic inactivation is required for a tumour suppressor gene to trigger a potential oncogenic effect.

### 1.1.2 Molecular alterations

Changes affecting the expression of proto-oncogenes and tumour suppressor genes can take different forms, starting with aneuploidy or chromosomal abnormality. Due to defective chromosome segregation or chromosome breakage, whole chromosomes or large segments of chromosomes can be lost, altering the copy number of tumour suppressor genes. Focal amplification of small chromosomic portions (< 10 Mbp) can lead to increased oncogene expression, while chromosome translocations can promote

the fusion of two distinct genes, resulting in a chimeric protein whose function is altered, potentially oncogenic.

Changes on proto-oncogenes or tumour suppressor genes at the nucleotide level can either activate their oncogenic potential or inactivate their functions, respectively. Single nucleotide changes include substitutions which can be either silent, missense or nonsense mutations. The degenerate genetic code enables in some cases a mutation to not impact the amino acid coded, thus being silent and having no impact on the coded protein. A nonsense mutation however results in a truncated and altered version of a protein. Finally, a missense mutation leads to a different amino acid being incorporated, potentially altering the function of the protein. This can lead to an oncogene activation, or suppression of the function of a tumour suppressor gene. The epigenetic state of a gene, determined by cytosine methylation and histone modification, can also impact specific genes without changing their DNA sequence, activating or silencing their expression.

Since the first division of a fertilised egg, the genome is constantly insulted, threatening its integrity. Exogenous sources such as exposure to carcinogens or UV irradiations can lead to damaged bases, crosslinks, or breaks, potentially stalling the replication fork during DNA synthesis. Endogenous processes can also alter the genome such as spontaneous deamination of bases, or damages induced by free-radicals (ROS) or alkylating agents. Another source of mutation is DNA replication errors. Even though, less than one mutation arises after each human cell division (Alberts *et al.*, 2015). This is thanks to a panoply of safeguarding pathways limiting replications errors, correcting errors if present, repairing if damages, and finally preventing cell division and promoting apoptosis if damage to the cell is beyond a recovery threshold.

### 1.1.3 Cell division

#### 1.1.3.1 Cell cycle

Cell division is a tightly regulated process divided into two main phases: interphase which comprises $G_1$, S- and $G_2$ phases, and M-phase for mitosis. The S-phase is dedicated to the accurate and faithful synthesis of the nuclear genome, while the gap phases – $G_1$ and $G_2$ – are transition periods when the internal and external environments are monitored to ensure that all needs are met prior to cell cycle progression. Each transition from one phase to the next is triggered by the cell cycle control system. Indeed, cell cycle progression is based on a connected series of irreversible on/off biochemical switches, triggering entrance into the next phase of the cycle, with no return. The control system blocks the transition from one phase to another if defects are detected, either enabling time for the cell to recover and fix the issue, or trigger apoptosis if impossibility to overcome the damage. The on/off switch relies on the dance between a family of protein kinases known as cyclin-dependent kinases – CDKs, and its activating dancing partners appearing and disappearing as the cell cycle moves on – the cyclins. While the CDKs concentration is constant throughout the cell cycle, cyclins concentration goes up and down along the cell cycle phases, allowing cyclical assembly and activation of cyclin-CDK complexes at specific stages. There are three main classes of cyclins: 1) $G_0/G_1$ help to trigger the entrance of the cell into the cell cycle by activating CDKs and exiting from $G_0$; 2) S-cyclins help to stimulate chromosome duplication and bind CDKs soon after progression through $G_1$; 3) M-cyclins activate CDKs that stimulate the $G_2$/M-phase transition. Another factor regulating the cell cycle is the Anaphase Promoting Complex, also called Cyclosome (APC/C), a multisubunit E3 ubiquitin ligase

(Sudakin *et al.*, 1995). Once ubiquitinated, APC/C substrates are directed towards the proteasome for degradation. Within the cell cycle, APC/C mediates S-phase and $G_2$/M-phase cyclin degradation (Siddiqui *et al.*, 2013). Interestingly, $G_1$-CDKs are not targeted by the APC/C, but are ubiquitylated by the SKP1-Cullin-F-Box (SCF) complex instead for degradation during S-phase (Ang *et al.*, 2005). However, CDKs and APC/C regulate each other. Indeed, $G_1$-CDK phosphorylates the APC/C adaptor Cdh1, hence prevents APC/C function and allows S-CDK (S phase-specific cyclin-dependent kinase) activity accumulation. Furthermore, M-CDKs-dependent phosphorylation of the APC/C interacting factor Cdc20 results in activation of the E3 ligase, ubiquitinating S- and $G_2$/M-CDKs, allowing activation of the APC/C (reviewed in (Peters, 1998).

#### 1.1.3.2 DNA replication

In 1953, J. Watson and F. Crick hypothesised that DNA possesses a double helix conformation, with a backbone made of deoxyribose and phosphates, with complementary bases A:T and G:C linked by two and three hydrogen bonds, respectively (Watson *et al.*, 1953). This hypothesis also suggested that duplication of the genetic material would take place in a semiconservative way, which was later confirmed by the classic Meselson-Stahl experiment (Meselson *et al.*, 1958). DNA replication is a complex process and can be divided into four steps where loading into chromatin of different transient factors gives rise to different complexes, tightly regulated by the cell cycle. It is initiated with 1) origin licencing in $G_1$ with the recruitment of inactive helicases – or called pre-Replication Complex (pre-RC) – at the origins. Then, during the transition from $G_1$ to S-phase, the pre-RC builds up through the recruitment of additional factors via the pre-Loading Complex (pre-LC) to become a pre-Initiation Complex (pre-IC), a helicase

ready to be activated. During S-phase, 2) origin firing occurs via helicases activation, the replisome moves on and 3) DNA synthesis occurs. Finally, 4) DNA replication terminates when the two replication forks converge.

***Origin licencing***

Origin licencing starts in late mitosis/early $G_1$ phase, when APC/C is highly active (Limas *et al.*, 2019), preventing activity from S- and $G_2$/M-phase cyclins. The Origin Replication Complex (ORC, comprising the six subunits ORC1-6) combined with Cdc6 (Duzdevich *et al.*, 2015) is loaded onto the origins. The Mini Chromosome Maintenance helicase complex (MCM2-7, containing the six subunits MCM2-7), in complex with Cdt1, is then recruited to ORC-Cdc6 (Sun *et al.*, 2013; Yuan *et al.*, 2017). Followed by Cdt1 (Ticau *et al.*, 2017), Cdc6 (Ticau *et al.*, 2015) is then released, subsequently followed by the loading of a second MCM2-7-Cdt1 complex via Cdc6 (Ticau *et al.*, 2015). Finally, the loading factors Cdc6, Cdt1, then ORC, are released from the origins (Ticau *et al.*, 2017), leaving a symmetrical head-to-head double hexamer surrounding double-stranded DNA (dsDNA) corresponding to an inactive helicase – or called pre-RC – at the end of $G_1$ phase (Deegan & Diffley, 2016).

In order to shift from an inactive to an active helicase, additional factors are required at the origins. These factors are recruited onto the pre-RC during S-phase, subsequently forming the pre-IC, which consists of the addition of Cdc45, GINS (protein complex composed of Sld5, Psf1, Psf2, and Psf3), Pol ε, Sld2, and Sld3 (Tanaka & Araki, 2013; L. Zou *et al.*, 1998). The recruitment of the pre-IC mainly relies on two cyclin kinases: DDK (Dbf4-Dependent Cdc7 Kinase) and S-CDK (Labib, 2010; Tanaka *et al.*, 2010). Phosphorylation of Mcm4 and Mcm6 by DDK (Randell *et al.*, 2010; Sheu *et al.*,

2010) promotes interaction and recruitment of Sld3 and Cdc45 at the helicase (Deegan, Yeeles*, et al.*, 2016). S-CDK activation leads to Sld2 phosphorylation (Tanaka *et al.*, 2007; Zegerman *et al.*, 2007). Phosphorylated Sld2 leads to the recruitment of Dpb11, in addition to GINS and Pol ε, altogether constituting the pre-Loading Complex (pre-LC) (Muramatsu *et al.*, 2010). While Dpb11 interacts with the GINS complex (Tanaka, Komeda*, et al.*, 2013) the carboxyl-terminal (C-ter) of the catalytic subunit of Pol ε (Pol2) interacts with Sld2 and Dpb11 (Muramatsu *et al.*, 2010). Sld3 (located on the inactive helicase) phosphorylation by S-CDK leads to the recruitment of the pre-LC (Sld2, Dpb11, GINS and Pol ε) via interaction with Dbp11 to the pre-RC (Tanaka *et al.*, 2007; Zegerman *et al.*, 2007) using Sld3 as an anchor point. Although Sld2, Sld3 and Dpb11 are essential for helicase activation, these factors are not components of the CMG complex. Still under investigation how the helicase is activated and how these factors are dissociated, although Mcm10 seems to play an important role (R. Georgescu *et al.*, 2017).

Pol ε seems to play a crucial role in replication initiation in human cells as it physically interacts with GINS, suggesting a direct link between Pol ε and the DNA helicase (Bermudez *et al.*, 2011). Furthermore, Pol ε is required in fission yeast for assembly and activation of the replicative helicase as mutations disrupting the polymerase disturb the full helicase progression away from origins (Handa *et al.*, 2012). Pol ε is also required for assembly and activation of the CMG helicase complex constituted of the Cdc45, Mcm2-7 and GINS subunits (Sengupta *et al.*, 2013).

***Origin firing***

Origin firing consists of the conversion of the two MCM2-7 hexamers surrounding dsDNA into two active helicases. The MCM2-7 double hexamer surrounding dsDNA

transitions to a single MCM2-7 hexamer encircling single-stranded DNA (ssDNA) on the leading strand of the two divergent forks moves away from the origins in a bidirectional manner. Even though the mechanism of origin firing is still under investigation, in budding yeast it seems to require stimulation via Mcm10 which leads to DNA unwinding (Looke *et al.*, 2017). Exposure of ssDNA allows priming of DNA replication, as well as the recruitment of additional components of the replication machinery – or replisome –, including polymerase alpha (Pol $\alpha$) which initiates DNA synthesis.

***DNA replication***

Helicase activation leads to DNA unwinding. In order to prevent DNA degradation and reannealing, ssDNA exposure triggers RPA coating. This provides sites for recruitment of Pol $\alpha$-primase, which initiates both leading- and lagging-strand synthesis. The Pol $\alpha$-RNA primase complex contains two catalytic activities: a primase and a 5'-3' polymerase activity. Pol $\alpha$ synthetises ~ 10 nucleotides via its primase activity, before switching to its polymerase activity for further 20-30 nucleotides synthetisation (Bhagavan *et al.*, 2015). Two additional DNA polymerases – Pol $\delta$ and Pol $\varepsilon$ – are then substituted to Pol $\alpha$-primase to perform the bulk of DNA replication. The main consensus for the division of labour between the two family-B polymerases is that Pol $\varepsilon$ replicates in a continuous fashion the main bulk of the leading strand, while Pol $\delta$ deals with the lagging strand, synthesised discontinuously. Supported by *in vivo* (Miyabe *et al.*, 2011; Nick McElhinny *et al.*, 2008; Pursell *et al.*, 2007), yet this model was challenged when it was suggested that Pol $\delta$ might synthesise a small fraction of the leading strand replication events, close to replication origins, in place of Pol $\varepsilon$ (Daigaku *et al.*, 2015).

Complementary approaches are used to decipher DNA replication. In recent years, elegant *in vitro* studies reconstituted step by step the eukaryotic replisome, defining the minimum set of proteins required to initiate origin-dependent DNA replication in budding yeast and confirming the role of Pol ε in replicating the leading strand (J. T. Yeeles *et al.*, 2015). In a subsequent publication, Pol δ and PCNA were added to the list of replication factors, which allowed to start to clarify the role of Pol δ in leading strand synthesis (J. T. P. Yeeles *et al.*, 2017). In line with previous *in vivo* findings showing that Pol δ might synthesise leading strand around replication origins (Daigaku *et al.*, 2015) it was confirmed that Pol ε cannot function on lagging strand on the contrary of Pol δ (R. E. Georgescu *et al.*, 2014; Nick McElhinny *et al.*, 2008; J. T. P. Yeeles *et al.*, 2017), in addition to Pol δ which contributes to the initiation of the leading strand synthesis after Pol α priming, before switching to Pol ε for more extensive synthesis of the leading strand (Aria *et al.*, 2018; J. T. Yeeles *et al.*, 2015; J. T. P. Yeeles *et al.*, 2017). *In vivo* studies later confirmed these findings (Garbacz *et al.*, 2018; Z. X. Zhou *et al.*, 2019).

That Pol δ could participate in the leading strand synthesis was previously illustrated by the fact that budding yeast lacking both Pol ε polymerase and exonuclease activity were relying on Pol δ, suggesting that Pol δ can synthesise leading strand, although poorly and leading to genome instability (Dua *et al.*, 1999; Garbacz *et al.*, 2018; Kesti *et al.*, 1999).

The replisome also contains a pentameric clamp loader – Replication Factor C (RFC) – which assembles the circular sliding clamp Proliferating Cell Nuclear Antigen – PCNA – on DNA. While a single PCNA is present for each leading strand, there is one PCNA for each Okazaki fragments on the lagging strand (Boehm *et al.*, 2016).

***Leading strand replication***

Initiated by Pol α, DNA synthesis occurs in a continuous manner in eukaryotes on the leading strand, as the active helicase exposes a 3'-5' template allowing continuous 5'-3' DNA synthesis by Pol ε. As previously mentioned, Pol ε is loaded onto the helicase prior to Pol α during the pre-IC recruitment, possibly via the C-ter domain of the second subunit of Pol ε, which in budding yeast interacts with Mcm5 (Langston *et al.*, 2014; Sun *et al.*, 2015). Further studies have shown that Pol ε interacts with replisome factors such as Psf1 and the amino-terminal (N-ter) of the second subunit of Pol ε in fission yeast (Sengupta *et al.*, 2013). Furthermore, evidences demonstrate that Pol ε moves together with the replication fork (Aparicio *et al.*, 1997), and its high replication rate (Ganai, Osterman*, et al.*, 2015) could be explained by its structure which encircles nascent dsDNA (Hogg *et al.*, 2014).

***Lagging strand replication***

The lagging strand is replicated in a semi-discontinuous way. Repeated primings occur as the 3'-5' ssDNA lagging strand template DNA is exposed after helicase activity. Following priming by Pol α, RFC is recruited to the 3'-OH end of the nascent primer, enabling subsequent recruitment of PCNA to displace Pol α. Pol δ is recruited, allowing further elongation of the lagging strand, the replicative polymerase activity enhanced by the presence of PCNA (Syvaoja, 1990) . Ultimately, Pol δ reaches the 5'-end of the preceding Okazaki fragment and strand displacement occurs. The rate of synthesis of the enzyme drastically decreases, allowing the RNA primers of the last Okazaki fragment synthesised to be displaced, generating a 5'-flap. This 5'-flap is then excised by Fen1, initiating a nick-translation process: the nascent flap is cut, and strand displacement

synthesis restarts, making this process iterative and releasing predominantly mononucleotides. Once complete degradation of the RNA primer, ligation of the nick can occur via DNA ligase 1.

***A brief note on PCNA, Pol δ and Pol ε***

The sliding clamp PCNA is a large trimeric ring-shaped molecule wrapped around dsDNA. Loaded onto DNA by the activity of the clamp loader RFC, PCNA has been described as associating with DNA polymerases and preventing dissociation of these from duplex DNA during DNA replication (Boehm *et al.*, 2016; Slade, 2018). PCNA modulates the activity of both Pol δ and Pol ε, although via distinct mechanisms. Indeed, while Pol δ activity is PCNA-dependent (Chui *et al.*, 1995), the presence of PCNA only stimulates the activity of Pol ε, already highly processive even in absence of PCNA, via facilitating the binding of the primer end to the polymerase, preventing the unproductive binding of ssDNA (Burgers, 1991). The catalytic subunit of Pol ε possesses a canonical PIP-box motif (Dua *et al.*, 2002) related to PCNA binding which could lead to think of possible direct interaction between Pol ε and PCNA. However, this sequence is located within the first half of the N-ter domain of the catalytic subunit of the protein, adjacent to the polymerase domain, hidden and difficult of access, suggesting that it does not directly interact with PCNA, later confirmed by surface plasmon resonance techniques showing no physical interaction between Pol ε and PCNA, on the contrary of Pol δ (Chilkova *et al.*, 2007). It is not until recently that a cryo-electron microscopy study of the budding yeast Pol ε holoenzyme contrasted this finding by showing a direct interaction between PCNA and the N-ter domain of the catalytic domain of Pol ε (Yuan *et al.*, 2020). Although alteration of Pol ε PIP box does not affect cell growth in budding yeast, it increases cell

sensitivity to replication stress (Dua *et al.*, 2002), thus highlighting the importance of PCNA and Pol ε interaction.

***Replication fork convergence – DNA replication termination***

Once replisomes encounter one another, the helicase faces the 5' end of the last initiated Okazaki fragment from the converging replication fork. Replication fork convergence is a field of study which requires still further investigation. Nevertheless, current models suggest that the CMG passes onto dsDNA, translocating along leading strand, not unwinding it (Dewar & Walter, 2017). Unloading of the CMG complex is one of the latest events in DNA replication, occurring only once encircling dsDNA. Even if the mechanism is still not fully understood, final dissociation of the CMG from the dsDNA seems to require Mcm7 ubiquitylation via E3 ubiquitin ligase $SCF^{Dia2}$ in budding yeast (Duxin *et al.*, 2014). In metazoans, such as worms or frog eggs, $CRL2^{LRR1}$ seems to be the ubiquitin ligase promoting MCM7 ubiquitylation and CMG unloading at the end of S-phase (Dewar, Low*, et al.*, 2017; Maric *et al.*, 2014). CMG unloading is followed by the ligation of the 3'-leading and 5'-lagging strand ends. The disassembly of the rest of the replisome factors is still an open question. Nevertheless, is it likely that these proteins are passively unloaded as an indirect consequence of the CMG dissociation. This is supported by the idea that Pol ε and the CMG dissociate with similar kinetics, and that blocking the CMG unloading leads to retention of most replisome components at the fork, including Pol ε (Dewar & Walter, 2017).

#### 1.1.3.3 dNTP pool regulation

DNA synthesis is semi-conservative, requiring therefore synthesis during the cell cycle of the four types of deoxyribonucleotide triphosphate (dNTPs) – dATP, dTTP, dGTP, dCTP. These are obtained via two distinct pathways: *de novo* synthesis and salvage. Taking place in the cytoplasm, *de novo* synthesis relies on ribonucleotide reductase (RNR) which converts ribonucleotide diphosphate (NDP) to deoxyribonucleotides diphosphate (dNDP), which subsequently are phosphorylated to dNTP. Budding yeast RNR is a holoenzyme composed of a large – Rnr1 – and two small subunits – Rnr2 and Rnr4. RNR also possesses a minor isoform where Rnr1 is replaced by Rnr3. RNR activity goes up and down following the cell cycle. On one hand, transcription of RNR factors increase at the $G_1$/S-phase transition and peak in the middle of S-phase for dNTPs supply (Mathews, 2015). On the other hand, possibly in order to limit unscheduled DNA replication, Rnr2 level decreases during $G_1$ and $G_2$ phases via ubiquitin-mediated degradation (A. L. Chabes *et al.*, 2003). Another regulation of dNTPs synthesis lies in the two allosteric sites of RNR which also regulate its activity and the relative amounts of dNTPs synthetised. The first allosteric site controls the overall activity of RNR. dATP binding to the enzyme leads to its inactivation, while ATP binding stimulates catalysis. Specific mutations in this allosteric site leads to elimination of dATP feedback inhibition and to an overall elevated dNTP pool, as it has been observed in *rnr1-D57N* budding yeast strain (A. Chabes *et al.*, 2003), *cdc22-D57N* fission yeast strain (Fleck *et al.*, 2013) and mouse (Caras *et al.*, 1988). The second allosteric site, or called the 'specific site', controls the selection of the NDP to catalyse, ensuring the dNTP pool balance (Guarino *et al.*, 2014).

The second pathway participating in the dNTP pools regulation is the salvage pathway which 'reutilises' the nucleosides and bases that arise through nucleic acid degradation. The nucleosides are processed by nucleoside kinases while the bases are salvaged by reaction with phosphoryltransferases and usually takes place in both mitochondria and cytoplasm (Mathews, 2015).

Careful regulation of dNTP levels is vital to first, maintain the right supply of dNTPs required during S-phase for full and accurate genome duplication, and second, to avoid genomic instability. Intracellular dNTP pool balance results from the coordination of synthesis and degradation of dNTPs, critical for DNA repair and DNA synthesis fidelity (Bester *et al.*, 2011). Indeed, imbalanced dNTP supplies is associated with DNA replication stress and altered replication fork velocity (Clifford *et al.*, 2014), and lead to an increased mutation rate, the mutation frequencies related to the type of dNTP ratios altered (Schmidt *et al.*, 2019). Furthermore, an overall excess of dNTPs was associated with hypermutator phenotype and cancer development (Kohnken *et al.*, 2015), which could be due to a less efficient proofreading (J. S. Williams *et al.*, 2016). Altogether, this suggests an intertwined relationship between replication fidelity and cell cycle metabolism.

### 1.1.4 Overcoming DNA damage in S-phase

#### 1.1.4.1 DNA Repair in general

Thanks to a fantastic evolutionary process, the cells have developed alternatives to cope with DNA damage – DNA repair. Facing the palette of different alterations on DNA, a wide range of DNA repair pathways are available to locate, signal, and repair the

anomaly. DNA repair can be divided into two main categories: direct and indirect. Direct reversal of the DNA damage corresponds for instance to $O_6$-methylguanine DNA methyltransferase which removes the methyl groups on $O_6$-methylguanine and restores non-altered guanine. Indirect DNA repair is a post-replicative process which takes place after synthesis of the daughter strand. Indirect DNA repair can be sub-categorised as follows: base-excision repair (BER), nucleotide excision repair (NER), ribonucleotide excision repair (RER), mismatch repair (MMR) and cross-link repair. BER is the main pathway involved in repairing base modifications such as 8-oxoguanine lesion resulting from guanine oxidation, sites of loss of bases and single stranded break. NER participate in the removal of bulky DNA lesions due to UV exposure forming thymidine dimers or 6,4-photoproducts for instance. Via the RER pathway, RNaseH2 excises ribonucleotides by cleaving portions of DNA-RNA hybrids. There are in addition two pathways for repair of double strand breaks, namely homologous recombination and non-homologous end joining.

#### 1.1.4.2 Mismatch Repair in particular

Mismatch Repair is mainly responsible for the correction of replication-error mismatches. For instance, microsatellites – repeated sequences of 1-6 nucleotides widely distributed along the genome – are thought to be generated from slippage of replicative DNA polymerases. Microsatellite Instability (MSI) – microsatellite mutations, or alteration in microsatellite lengths – is due to defect in MMR. The Lynch syndrome, an autosomal dominant hereditary disease in which patients are prone to early development of colon cancer, often carry genomic alterations impairing MMR genes, such as silencing of MLH1 by sporadic hyper-methylation (Boland *et al.*, 2008).

Thoroughly studied in *Escherichia coli* (*E. coli*), MMR factors are nonetheless highly conserved between prokaryotes and eukaryotes (Figure 1.1). MMR process can be divided into four distinct steps. First, the mismatch is recognised. Second, factors required for repair are recruited, in addition to a nick into duplex DNA, initiating the third step, which is excision of the incorrect sequence. Re-synthetisation, or gap-filling of the newly synthesised strand is the final step. Mismatches can be divided into two distinct classes: base-base mispairs, and unpaired insertion and deletion (indels). In prokaryotes, MMR is initiated with the recognition of the distorted DNA by MutS. The factor MutL is then recruited, and acts as a dynamic ATP-activated clamp with a central channel that can capture DNA (Hsieh *et al.*, 2008). MutL acts as a 'Janus' (Sameer *et al.*, 2014) of the MMR system: it mediates interactions between MutS and other protein complexes of the MMR, creating a link between the mismatch recognised by MutS and the protein responsible for the excision of the mismatch. Together, these proteins licence the excision repair pathway by stimulating the endonuclease activity of MutH. MutH targets the cleavage of the unmethylated strand, assuring the correction of the mismatch which has occurred on the newly (thus unmethylated) synthetised strand. Excision of the mispaired base occurs via nicks on the 3' or 5' side. The resulting DNA gap is then filled in and ligated, restoring the initial sequence prior to mismatch.

The MMR repair process in eukaryotes follows similar steps. The mismatch recognition involves MutS Homologues (MSH) complexes: MutS$\alpha$ and MutS$\beta$. MutS$\alpha$ corresponds to the Msh2-Msh6 duplex which recognises more readily single base-base and 1-2 base indels mismatches. MutS$\beta$ involves the protein Msh2 in complex with Msh3. Similarly to MutS$\alpha$, it recognises single base indels, but also larger ones (up to 15 bp). There are three distinct MutL homolog complexes in eukaryotes, each composed of

two proteins: MutLα (MLH1-PMS2 in human, Mlh1-Pms1 in budding yeast), MutLβ (Mlh1-Mlh2) and MutLγ (Mlh1-Mlh3). MutLα is the most active of the three complexes in human and plays the major role in the system (Kunkel *et al.*, 2005), while MutLβ and MutLγ mainly participate in the repair of some base-base mispairs and indels, probably redundant with the action of MutLα (Cannavo *et al.*, 2005). DNA cleavage happens via Pms2 or Mlh3, whose latent endonuclease activity is activated by PCNA (Kadyrov *et al.*, 2006). The excision step in eukaryotes is still unclear. Indeed, Exo1 is to date the only exonuclease identified to be involved in MMR. Physically interacting with Msh2 and Mlh1 (Tran *et al.*, 2001), its 5'-3' exonuclease activity can support 5'-3' MMR excision. However, for 3'-5' MMR excision, things are still unclear on that aspect. One alternative suggested is the strand displacement capacity of Pol δ from a nick (Hsieh *et al.*, 2017).

Once excision is performed, gap-filling with a high-fidelity polymerase takes place, followed by ligation with DNA ligase I. PCNA loaded by RFC at a 3' terminus, Pol δ or Pol ε can use the RPA coated gapped DNA to carry out synthesis. It is still unclear which polymerase is favoured and in which context. However, *in vitro* studies revealed that Pol δ was identified as necessary for MMR (Longley *et al.*, 1997) while later, Pol ε was also identified to robustly support gap filling reaction required for both 5' and 3' nick-directed MMR reaction reconstituted *in vitro* (Bowen *et al.*, 2017).

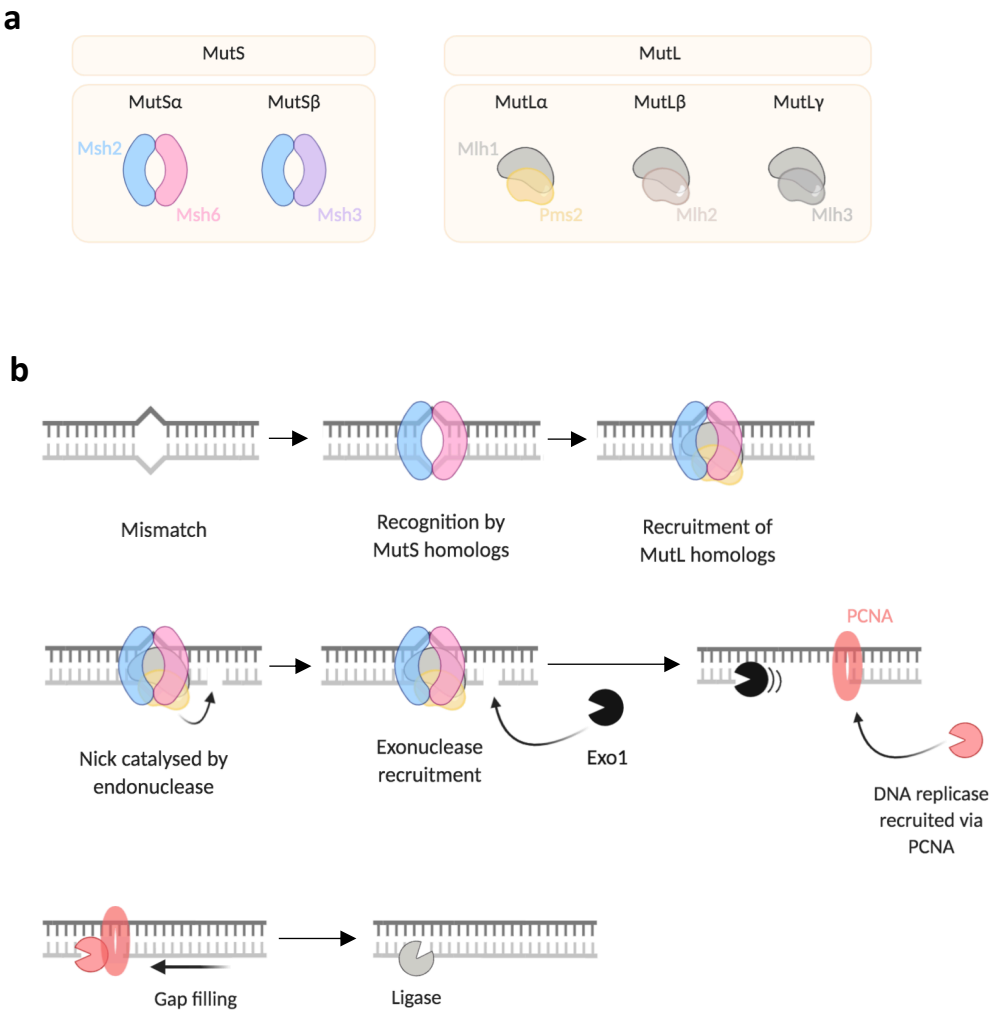



**Figure 1.1 Mismatch repair pathway**

(**a**) MMR factors. MutS homolog complexes are MutSα and MutSβ while MutL homolog complexes are MutLα, MutLβ and MutLγ. (**b**) Schematic representation of the MMR pathway when detection of a base-base mismatch. First the distortion of the DNA is recognised by MutS homologs enabling recruitment of the appropriate MutL homologs. Then, a DNA nick is catalysed by the endonuclease recruited via the MutL homologs. This allows recruitment of the exonuclease Exo1, excising DNA in a 5'-3' direction and removing the mismatch. Finally, a DNA replicase is recruited to fill in the gap created, followed by a final ligation.

Although in *E. coli* – a gram-negative bacterium – the newly synthesised strand is distinguished from the parental one thanks to transient unmethylated adenine (Lahue *et al.*, 1989), in gram-positive bacterium and eukaryotes the absence of this feature led to question how strand discrimination occurs in such organisms. One possibility, later confirmed by *in vitro* and *in vivo* assays, was that MMR could be directed to nascent DNA by strand discontinuity such as gap between Okazaki fragments (Holmes *et al.*, 1990; Nick McElhinny, Kissling*, et al.*, 2010; Thomas *et al.*, 1991), Pms2 associating with PCNA and introducing nicks which could be used as loading site for Exo1 (Pena-Diaz *et al.*, 2010; Pluciennik *et al.*, 2010). Yet easy to apply in the context of discontinuous lagging strand, nascent strand in a continuous leading strand where PCNA is not as present as for each of the Okazaki fragments, raised the question of strand discrimination in a leading strand context where strand discontinuity is readily available.

Pol ε and Pol δ incorporate not only nucleotides, but also ribonucleotides during the DNA replication process. While Pol δ incorporates 1 ribonucleotide every 5,000 bases synthetised, Pol ε incorporates about four times more ribonucleotides per base synthesised (Nick McElhinny, Kumar*, et al.*, 2010; Nick McElhinny, Watts*, et al.*, 2010; J. S. Williams *et al.*, 2016). Although Pol ε exonuclease activity is efficient on rNMP, its efficiency is much lower than for base-base mismatches (J. S. Williams *et al.*, 2012). Hence each rNMP left on the nascent strand leaving a 'mark' on the leading strand. Targeted by RER, RNAse-H2 generates strand breaks through ribonucleotides removal used by MMR system as an entry site for exonucleolytic degradation of the error-containing strand (Ghodgaonkar *et al.*, 2013; Lujan *et al.*, 2013).

#### 1.1.4.3 Translesion Synthesis

The road the replisome rides during DNA synthesis, or the pavement the replicative polymerases walks, is not always straight and well aligned. There are obstacles and unforeseen events along the way. Sometimes, it is better to keep going – even limping – than to fully stop. Similarly, replication forks can encounter numerous obstacles, such as thymidine dimers, bulky adducts, damaged bases, or secondary structures such as hairpin or G-quadruplexes elements. Replicative DNA polymerases might struggle to replicate these elements, possibly leading to replication fork stalling. If unresolved, accumulation of stalled replication forks can lead to even worst scenarios, including chromosomal rearrangements, cell cycle arrest, and ultimately, cell death. However, 'difficult-to-replicate' sequences can be traversed by specialised polymerases with the help of additional factors. This bypassing process is called Translesion Synthesis (TLS), a process which allows the cell to tolerate genomic lesions. The only negative aspect of the TLS is the inaccuracy of the replicative enzymes involved. The first reason explaining the error-prone aspect of these enzymes is that none of them possesses a proofreading domain which would enable the polymerase to excise misincorporated bases. The second reason resides in the tertiary structure of the TLS polymerases. Crystallographic studies revealed that TLS polymerases possess the classic right-hand shape (describe more thoroughly in section 1.2.4) observed for replicative polymerases, with fingers, palm and thumb domains (Ling *et al.*, 2001; Silvian *et al.*, 2001; B. L. Zhou *et al.*, 2001). Similar to other replicative polymerases, the catalytic residues for polymerisation are located in the palm domain; although contrastingly, the remaining domains are much smaller and stubby. This renders the active site more spacious to accommodate large

bulky DNA lesions that would be unable to fit into the tighter constraints of the polymerases with higher fidelity (Ling *et al.*, 2001).

There are seven mammalian enzymes with substantial TLS activity. Four belong to the Y-family of replicative polymerases: Pol ι, Pol η, Pol κ and Rev1; two belong to the A-family: Pol θ and Pol ν; and finally Pol ξ, part of the B-family (Lange *et al.*, 2011). Once the replisome faces a lesion which cannot be replicated, the appropriate TLS polymerase is chosen according to the lesion. Even if the mechanism of selection of a particular TLS polymerase is still under investigation, it seems that the recruitment takes place via interaction with the replicative sliding processivity clamp PCNA. Stalled DNA replication triggers monoubiquitylation of PCNA via $CRL4^{Cdt2}$ E3 ubiquitin ligase, creating a binding site for TLS polymerases from the Y-family which all possess a specific PIP-box allowing direct interaction with the clamp (Terai *et al.*, 2010). Once extension beyond the lesion executed, the TLS polymerase is replaced by a replicative polymerase to resume replication.

Even if regulated by key protein-protein interactions, TLS polymerases are still error-prone, even on small stretches of DNA not containing any lesions (Shcherbakova *et al.*, 2006). In the context of stalled replication fork, the recruitment of Pol ξ leads to a mutator phenotype (Northam *et al.*, 2006). It has been suggested that an increased mutation rate observed in cells exhibiting an increased dNTP pools could be due to activation of TLS (Aye *et al.*, 2015). Consistently, inactivation of TLS polymerases reduces the mutation rate associated with an increase in dNTP pools (Fleck *et al.*, 2013; Gon *et al.*, 2011). Thus, one would suspect that TLS polymerases have a unique mutagenic signature.

It has been suggested that TLS processes are responsible for many of the point mutations in the cells (Waters *et al.*, 2009) and perhaps contribute to the large increase in point mutations found in cancer genomes compared with normal tissues (Bielas *et al.*, 2006). For instance, Pol κ is known to be prone to making -1 frameshift mutations, although it can accurately and efficiently bypass a number of $N^2$-dG lesions (Sale, 2013). Initially identified in bacteria as Pol IV encoded by the gene *dinB* (Wagner *et al.*, 1999), Pol κ was identified as Pol IV human ortholog (Gerlach *et al.*, 1999; Ohashi *et al.*, 2000). Although it is widely conserved, no ortholog for Pol κ has been identified in budding yeast (Ohmori *et al.*, 2001). Even if not essential for viability, Pol $\kappa^{-/-}$ mice exhibit a shorter survival than its heterozygous counterparts, exhibiting spontaneous mutator phenotype in tissues including kidney, liver and lung (Stancel *et al.*, 2009). In cell lines, Pol κ depletion sensitises cells to apoptosis and chromosome aberrations after mitomycin C treatment, which can cross link $N^2$-dGs (Minko *et al.*, 2008). In addition, despite its low fidelity, Pol κ seems to protect against spontaneous mutagenesis (Stancel *et al.*, 2009), and is capable of bypassing thymine glycol residues, a major product of ROS-induced damage to DNA (Yoon *et al.*, 2010). Pol κ has also been identified as cooperating with Pol δ in NER for gap-filling step, being recruited by ubiquitinated PCNA and the scaffold protein XRCC1 (Ogi *et al.*, 2010). Also, it seems that Pol κ is more accurate than Pol δ in replicating microsatellites (Baptiste *et al.*, 2012). Taken together, it appears clear that there is a balance to maintain for the use of TLS polymerases, which excessive activity could be deleterious via an increased mutation rate.

## 1.2 DNA Polymerase ε

### 1.2.1 A brief history of polymerases

The term 'polymerase' was first used in 1957 in a scientific publication 'in order to facilitate reference in this report to the enzyme responsible for deoxyribonucleotide incorporation' (Kornberg, 1957). In this study, 1 kg of *E. coli* cell mass yielded 10 mg of the purified enzyme DNA polymerase I, identified as the actor catalysing the incorporation of dNTPs into DNA. Before 1975, Roman numerals were used to identify budding yeast DNA polymerases. Later on, a Greek alphabet nomenclature was used to categorise the different type of polymerases observed in higher eukaryotes (Weissbach *et al.*, 1975). Polymerases would be named accordingly to their order of discovery, following the alphabetic order. Due to the increase in newly identified polymerases and the similarities between budding yeast and human polymerases, this nomenclature was updated in 1989 during the Conference on Eukaryotic DNA Replication in Cold Spring Harbor, leading eventually to a revised nomenclature published in 1990 (Burgers *et al.*, 1990). Soon after, the number of polymerases identified in a broad range of cellular kingdoms expanded, and led to a new classification by family. Starting with three in 1991 (Ito *et al.*, 1991), there are today seven families of DNA polymerases (including reverse transcriptase such as telomerase) established, based on their amino acid sequence relationship (Filee *et al.*, 2002).

### 1.2.2 Pol ε or Pol δ – who is who?

For many years, Pol α was thought to be the only eukaryotic DNA polymerase (Yoneda *et al.*, 1965). It is only about 10 years later that came the isolation of the second of the

three main eukaryotic DNA replicative polymerases – Pol δ (Byrnes *et al.*, 1976). In addition, this enzyme was the first eukaryotic DNA polymerase identified with an additional activity other than 5'-3' polymerase – a 3'-5' exonuclease activity which reflects the capacity to proofread and excise mismatched nucleotides. Later, larger proteins with similar features were identified as isoforms of Pol δ, following the creation of two subclasses for Pol δ: Pol $\delta_1$ (M. Y. Lee *et al.*, 1980) and Pol $\delta_2$ (Crute *et al.*, 1986). The former one exhibited a low processivity with long templates and increased processivity in the presence of PCNA (Syvaoja, 1990); the subsequent one was highly processive even in the absence of PCNA and its activity was not enhanced by its presence (Focher *et al.*, 1988). It is only in 1990 that a final word was put to differentiate those two sub-classes of polymerases in mammalian cells – Pol $\delta_1$ was defined as Pol δ, and Pol $\delta_2$ as Pol ε (Burgers *et al.*, 1990; Syvaoja, 1990). Based on their amino acid sequences homologies, in mammalian cells the 15 DNA polymerases identified fall into five distinct families: A, B, C, X and Y (Lange *et al.*, 2011). The family B DNA polymerase is named after the product of the polB gene encoding *E. coli* DNA polymerase II. Pol ε, like Pol δ and Pol α, is a member of this family (Braithwaite *et al.*, 1993; Ito *et al.*, 1991).

### 1.2.3 Pol ε – a complex holoenzyme

Pol ε is a tetrameric holoenzyme with a stoichiometry of 1:1:1:1 (Chilkova *et al.*, 2003). The largest subunit contains both polymerisation and exonuclease catalytic activities, while the smaller subunits seem to play secondary roles, nonetheless important for the survival of the cell. Indeed, Pol ε appears to be involved in a number of cellular mechanisms other than DNA replication, such as cell cycle regulation, DNA damage

response, DNA repair, and maintenance of epigenetic states. *In vitro* and *in vitro* studies, helped to decipher the different aspects of this complex protein. In order to facilitate the reading flow and the differentiation between organisms, the subunits will be named using the human nomenclature. Please refer to Table 1.1 for the gene and protein names in different organisms.

| | Pol ε holoenzyme encoded gene names | | | |
|---|---|---|---|---|
| Human | POLE (2286 aa) | POLE2 (527 aa) | POLE3 (147 aa) | POLE4 (117 aa) |
| Mouse | Pole (2283 aa) | Pole2 (527 aa) | Pole3 (145 aa) | Pole4 (118 aa) |
| *S. cerevisiae* | pol2 (2222 aa) | dpb2 (689 aa) | dpb4 (196 aa) | dpb3 (201 aa) |
| *S. pombe* | cdc20 (2199 aa) | dpb2 (594 aa) | dpb4 (210 aa) | dpb3 (87 aa) |

**Table 1.1 Nomenclature Pol ε genes**

Pol ε subunits name in human, mouse, budding and fission yeast. Amino acid length indicated in brackets.

#### 1.2.3.1 POLE

Study of the amino acid sequence of the largest of the four subunits of Pol ε revealed a protein divided into two modules. While the N-ter carries both 5'-3' polymerase and 3'-5' exonuclease domain and is dispensable for viability (Dua *et al.*, 1999; Feng *et al.*, 2001; Kesti *et al.*, 1999), the C-ter does not harbour any known catalytic activity, although essential for cell survival (Dua *et al.*, 1999; Feng *et al.*, 2001). The C-ter of POLE however includes a cysteine rich sequence composed of two putative zinc fingers. Deletion of the region in between the two zinc-fingers regions is lethal in the budding yeast (Dua *et al.*,

1998), emphasising the importance of the C-ter domain in POLE for viability. Furthermore, the C-ter region is identified as an essential mediator for the interaction with the smaller subunits and required for the stability of the holoenzyme (Dua *et al.*, 1998, 1999) and is required for interaction with all other subunits in budding yeast (Morrison *et al.*, 1990).

POLE can be subject to post-translational modification as the N-ter is a potential target for SUMOylation (Cremona *et al.*, 2012). Two studies recently published back to back unravelled the fact that the lysine 571 located in the N-ter is a SUMO site, in parallel to the fact that C-ter domain contains a SUMO-interacting motif. SUMOylation of the subunit increases in response to methyl methane sulfonate and hydroxyurea treatment during S-phase in a RAD53-dependent manner (Meng *et al.*, 2019; Winczura *et al.*, 2019).

A close look at the amino acid sequence of POLE revealed that its C-ter module contains both exonuclease and polymerase domains present in the N-ter part of the protein, yet both inactive (Tahirov *et al.*, 2009). Nonetheless, a comparative analysis showed that both modules are only distantly related, ruling out a tandem duplication event. Phylogenetic and comparative analysis of eukaryotic and archaea B-family DNA polymerases indicated that the inactive domain could be the result of the insertion of a bacteria B-family polymerase between the active (N-ter) domain and the zinc fingers (present at the end of the C-ter), originated from archaeal B-family polymerase and archaeal D-family polymerase, respectively (Tahirov *et al.*, 2009).

#### 1.2.3.2 POLE2

The role of the second larger subunit in Pol ε is still unclear. Even if this subunit does not have any identified catalytic activity, its C-ter is essential for viability (Isoz *et al.*, 2012)

and impairment of the subunit can lead to an increase in the mutation rate in budding yeast (Jaszczur *et al.*, 2008; Jaszczur *et al.*, 2009; Kraszewska *et al.*, 2012). *In vitro* studies have tried to further dissect its role. In budding yeast, the C-ter of the second subunit interacts with the C-ter inter-zinc finger domain of the first subunit, stabilising the protein duplex (Dua *et al.*, 2000; Isoz *et al.*, 2012). However, the holoenzyme complex is stable when the second subunit is absent and does not impact replication efficiency *in vitro* (Isoz *et al.*, 2012). Furthermore, the second subunit is phosphorylated in early S-phase in a Cdc28-dependent manner, suggesting a potential regulatory role (Kesti *et al.*, 2004). In fission yeast it was identified as binding at replication origins in early S-phase, reinforcing the role of Pol ε in DNA replication regulation (Feng *et al.*, 2003).

#### 1.2.3.3 POLE3

This subunit has no identified catalytic activity. Non-essential in fission (He *et al.*, 2017), it however causes an increase in mutation rate when partially deleted in budding yeast, suggesting a role in maintaining chromosome replication fidelity (Araki, Hamatake, Morrison*, et al.*, 1991). Interacting with the fourth subunit of Pol ε holoenzyme, POLE3-POLE4 heteroduplex seems to play a role in epigenetic inheritance during DNA replication as the dimer associates with chromatin remodelers and histones, leading to loss of chromatin silencing after disruption of the dimer (He *et al.*, 2017). In mice, deletion of POLE3 has been associated with complete destabilisation of the Pol ε complex, leading to embryonic lethality in inbred strains and tumour predisposition in outbred strains (Bellelli *et al.*, 2018).

#### 1.2.3.4 POLE4

The smallest of the four subunits does not have any identified catalytic activity and is not essential in budding and fission yeasts (Ohya *et al.*, 2000; Spiga *et al.*, 2004). In budding yeast, *in vivo* depletion of this subunit slightly disturbs chromosomal DNA replication as it leads to S-phase delay (Ohya *et al.*, 2000), while depletion of the same subunit does not impair cell cycle progression in fission yeast (Spiga *et al.*, 2004). This subunit also contains a histone-fold motif, and interacts with POLE3 (Y. Li *et al.*, 2000). On a more structural aspect, absence of POLE4 in budding yeast leads to a less stable holoenzyme (Ohya *et al.*, 2000).

### 1.2.4 Crystallographic structure of Pol ε

The study of the structure of an enzyme can allow insight into its mechanism. The first DNA polymerase structure resolved by crystallography was the Klenow fragment of *E. coli* (Pol I) (Ollis *et al.*, 1985). This revealed a polymerase structure resembling the shape of a cupped right hand, consisting of a palm, thumb and fingers. While the palm contains the 5'-3' polymerase active site, the cleft created between the thumb and the fingers would hold dsDNA, with fingers and thumb moving in response to binding of DNA and nucleotides, switching from either an open or a close state. Notably, the 3'-5' exonuclease domain of Pol I is located 20 Å from the polymerase active site, suggesting a necessity for movement of the nascent DNA strand if proofreading is to occur. It has been suggested that a mismatch would result in the polymerase slowing down, followed by displacement of the primer template from the polymerase site to the exonuclease one, where the mis-incorporated nucleotide would be excised. The thumb, associating

with duplex DNA upstream of the polymerase active site, has been suggested to play a role in the balance between polymerase and exonuclease activity. Using the T4 DNA polymerase of bacteriophage, it was discovered that mutations in the thumb domain could be either mutator, if they cause the enzyme to spend more time in the polymerase mode (Stocki *et al.*, 1995), or antimutator, if they cause the enzyme to spend more time in the exonuclease/editing mode (Wu *et al.*, 1998).

Using macromolecular cryogenic electron macroscopy, the holoenzyme structure of *S. cerevisiae* Pol ε was set to be determined (Asturias *et al.*, 2006). The structure obtained to a 20-Å resolution showed two distinct modules joined by a flexible connexion. One globular domain would contain the catalytic subunit Pol2, the second domain having a shape of a tail constituted of Dpb2, Dpb3 and Dpb4, although in this study it was impossible to distinguish and identify each subunit and how they would interact with each other. Nonetheless, a 3.5-Å resolution structure of the same holoenzyme was very recently obtained, allowing a deeper insight into the enzyme and determine the juxtaposition of the subunits of Pol ε holoenzyme (Yuan *et al.*, 2020). The Similar to the previous structure, the main subunit Pol2 is divided into two main parts named NTD (N-terminal domain) and CTD (C-terminal domain). The NTD contains both catalytic functions of Pol ε and is organised into a toroid, a formation completely encircling DNA with a right-hand shape. On the second module sits Dpb2, barely interacting with NTD, Dpb3 or Dpb4. The two smallest subunits are intertwined and located in between the NTD and CTD. Wedged in the flexible joint, the Dpb3-Dpb4 subunits hence renders Pol ε in a rigid form on the contrary of a flexible form when absent, working as a bridge between the two Pol2 modules (Figure 1.2).

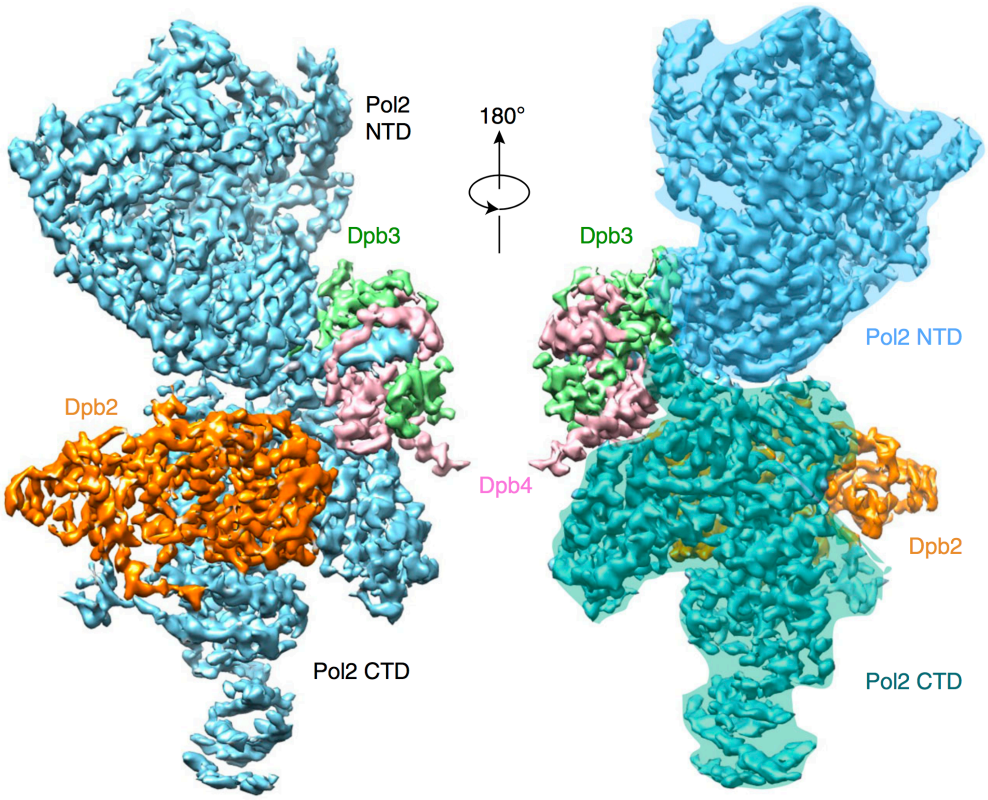



**Figure 1.2 Structure of Pol ε holoenzyme**

Front and back 3D map of *S. cerevisiae* Pol ε holoenzyme. Pol2 NTD and CTD domains are separated into two distinct lobes, Dpb3 and Dpb4 interacting and wedged in between, rendering the enzyme in a rigid form. Dpb2, the second subunit, sits on Pol2 CTD, the non-catalytic part of the main subunit of Pol ε. Adapted from and reprinted from Yuan et al., 2020 with creative commons attribution 4.0 international license (CC BY 4.0).

The crystallographic structure of the catalytic subunit of Pol ε remained unknown until 2014 when two groups solved the structure of its catalytic subunit in the budding yeast (Hogg *et al.*, 2014; Jain *et al.*, 2014). Both studies obtained the crystallographic structure of the N-ter region of the catalytic subunit (containing the exonuclease and polymerase domains) with an incoming dNTP. Jain and colleagues crystallised it with an incoming dCTP, while Hogg and colleagues used an incoming dATP in addition to the

3’OH missing from the primer terminus. In order to prevent degradation of the 3’ terminus of the DNA substrate during crystallisation, the residues D290 and E292 (budding yeast equivalent of D275 and E277 in human) from the exonuclease domain were mutated to alanine. Mutations of those residues were previously described as abrogating the exonuclease activity in budding yeast (*pol2-4* strain), as those two residues are required to coordinate the two divalent metal ions needed for catalysis of proofreading activity (Morrison *et al.*, 1991). The two resulting structures from both studies are very similar. Both confirmed that like Pol I, Pol ε possesses the characteristic right-hand shape shared with all DNA polymerases: a palm where the polymerase active site lies, fingers and thumb. In addition to those three domains, the exonuclease site, an NTD and a Processivity domain (P-domain) were also identified (Figure 1.3)

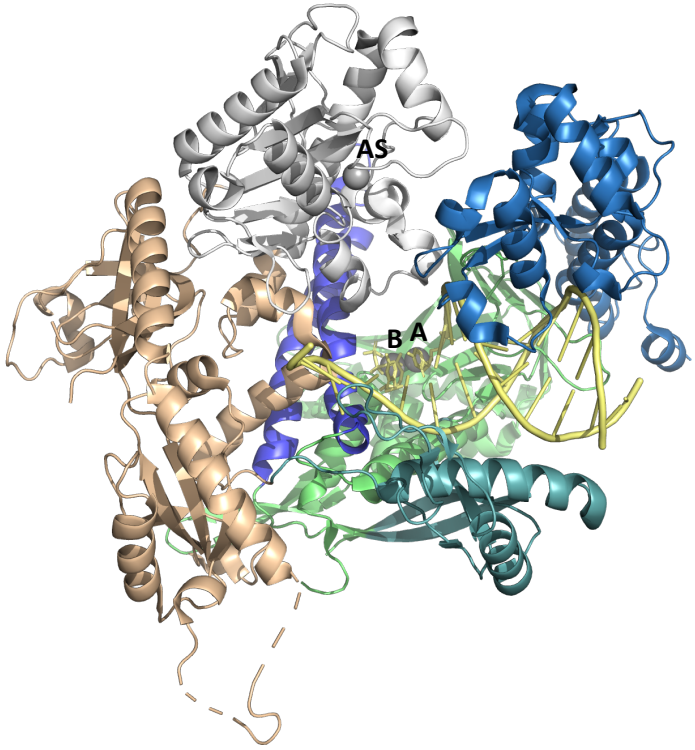



**Figure 1.3 Structure of Pol ε catalytic subunit**

*S. cerevisiae* Pol ε catalytic subunit crystallographic structure (Jain et al., 2014) (PDB ID 4PTF). The NTD is in light orange, exonuclease domain in grey with active site (AS), palm in green, thumb in blue, fingers in purple, and P-domain in cyan. The DNA is represented in yellow with the position for metal A and B necessary for the polymerisation reaction located in the palm domain.

#### 1.2.4.1 5'-3' polymerase activity of DNA polymerase ε

Pol ε possesses two main catalytic activities: 5'-3' DNA template-dependent DNA synthesis (polymerisation) and 3'-5' exonuclease activity (proofreading). Polymerisation by Pol ε happens in a processive way without dissociation of the polymerase from the DNA, and is initiated with the binding of a 3' primer template to the polymerase active site. The open position of the fingers allows binding of the correct complementary dNTP. A conformational change in both polymerase and DNA leads to a closed state of the polymerase, bringing the dNTP into position for catalysis, or more precisely, a

nucleophilic attack from the 3' hydroxyl primer to the $\alpha$-phosphate of the incoming dNTP. This phosphoryl reaction releases a pyrophosphate and is followed by the translocation of the polymerase to the newly incorporated nucleotide, serving as a new 3' primer terminus for the next round of catalysis. Two ions are involved in the catalysis reaction at position A and B (Figure 1.3). While metal ion A usually has a catalytic role in countering the growing charge due to the leaving pyrophosphate group, the metal ion at position B facilitates the dNTP binding and subsequent release of pyrophosphate group. Hogg *et al.* (2014) shows a single $Mg^{2+}$ at position B while Jain et al. show two $Ca^{2+}$ ions at positions A and B in the polymerase active site. The absence of metal A ion in Hogg *et al.* (2014) can however be explained by the lack of 3'OH ligand at the primer terminus used in this study.

#### 1.2.4.2 3'-5' exonuclease activity of DNA polymerase ε

A proofreading mechanism requires that the exonuclease satisfies the following criteria: i) that it acts unidirectionally at the 3'-end terminus and in a non-processive way on its DNA substrate, ii) ssDNA is the optimal substrate and catalysis releases dNMPs preferentially excised from mispaired primer-terminus, and iii) it functionally interacts with the polymerase to enhance fidelity of the enzyme (Kunkel, 1988). The second catalytic activity of Pol $\varepsilon$ – 3'-5' exonuclease – responsible for proofreading activity, aims for the excision of a mismatching nucleotide previously incorporated during its polymerisation process. While the polymerase activity lies within the palm, the exonuclease domain of Pol $\varepsilon$ is approximately 40 Å away from it (Hogg *et al.*, 2014; Jain *et al.*, 2014). However, even given the distance, the 3' mismatch terminus primer is

translocated from the polymerase domain to the exonuclease domain via an intramolecular mechanism (Ganai, Bylund*, et al.*, 2015). Once the mismatch removed, the polymerase switches back to polymerase mode and resumes DNA synthesis. Pol ε is a polymerase-associated 3'-5' exonuclease as its catalytic subunit possesses both activities (Kunkel, 1988). The exonuclease domain of Pol ε is located in the N-ter of its catalytic subunit. Its crystal structure reveals the exonuclease domain folded around a central β-sheet which contains the active site aspartic acid at position 290 and glutamic acid at position 292, together with a DNA-binding pocket where the last incorporated nucleotide could be excised. Excision will happen using a two-metal-ion-catalysed phosphoryl transfer, analogous to the one involved in polymerisation (Figure 1.3).

There are six exonuclease sequences called EXO motifs which are consensus sequences identified in enzymes from bacteriophage to human (Bernad *et al.*, 1989; Shevelev *et al.*, 2002). These motifs were confirmed to be important by site-directed mutagenesis (Derbyshire *et al.*, 1995). Pol ε contains five of these six EXO motifs (Shevelev *et al.*, 2002).

#### 1.2.4.3 Processivity domain of Pol ε

A new feature of Pol ε that the crystallographic structures unravelled is a domain that Hogg and colleagues named the P domain, after the suspicion that it might have an impact on processivity. Not previously reported in B-family polymerases, this domain forms an elongated 'arm' that extends outward from the palm toward the dsDNA, encircling the newly synthesised duplex DNA, suggesting that it could improve its

processivity (Figure 1.3). Confirming this hypothesis, partial deletion of the P-domain negatively impacts the polymerase activity (Hogg *et al.*, 2014).

#### 1.2.4.4 N-Terminal Domain of Pol ε

In addition to the palm, fingers and thumb characteristic of the B-family DNA polymerases, an NTD was also observed via crystallography (Figure 1.3). The function of NTD in Pol ε is yet to date not clear, although the NTD of Pol δ was suggested to bind both DNA and RNA (Swan *et al.*, 2009).

### 1.2.5 Pol ε and cell cycle progression

Pol ε being one of the major replicative DNA polymerases, it is legitimate to expect that it might play a role in cell cycle regulation. There are evidences that Pol ε participates in checkpoint activation, although its exact role remains unclear. First, a budding yeast mutation in the C-ter of Pol ε (glutamine at position 2195 converted to a stop codon), leads to a loss of S-phase checkpoint function (Navas *et al.*, 1995). Second, still in budding yeast, after UV irradiation Pol ε was identified as a transducer during S-phase, required to sense UV damage and replication blocks, as the phosphorylation of the effector Rad53 (human CHK2 ortholog) is Pol2-dependent (Navas *et al.*, 1996). Third, this time in fission yeast, Pol ε was identified as interacting with the transducer Mrc1, required for Cds1 (human CHK2 ortholog) phosphorylation, therefore S-phase checkpoint (Lou *et al.*, 2008). Indeed, under unperturbed conditions, Mrc1 was found to interact with both N-ter and C-ter of Pol2, while Mrc1 phosphorylation during S-phase would abolish binding to the N-ter of Pol2, but not the C-ter. Taken together, it was

suggested that Mrc1 might stabilise Pol ε at the replisome, enhancing interaction between Pol ε and the Mcm2-7 replication helicase. Furthermore, dephosphorylation of the second subunit of Pol ε is linked to proper S-phase progression and exit from mitosis (Voitenleitner *et al.*, 1997). MDM2 is a major repressor of p53 as it acts as an E3 ubiquitin ligase and targets p53 for degradation. It has been showed that the N-ter of MDM2 interacts with the C-ter of Pol ε (Vlatkovic *et al.*, 2000), and seems to stimulate Pol ε activity (Asahara *et al.*, 2003). Thus, it has been suggested that MDM2 and other factors could modulate Pol ε function in response to changing conditions, such as a reconfiguration of Pol ε after binding to MDM2, displacing Pol ε bound factors in response to stress, leading to a changing role of Pol ε for DNA repair and/or checkpoint control.

### 1.3 Pol ε and Cancer

As described earlier, DNA replication is a tightly regulated machinery. Not only DNA has to be duplicated once and only once, it also has to be accurately replicated. The replicative DNA Pol ε is an important part in this cellular orchestra. One of its main function is to accurately replicate the main bulk of our DNA. Pol ε is a remarkably accurate replicative polymerase thanks to its highly intrinsic base selectivity combined with its proofreading activity. As one could expect, inactivation of the exonuclease activity of human Pol ε increases the mutation rate for mispairs *in vitro* ($6.5 \times 10^{-4}$ for wild-type to $46 \times 10^{-4}$ for an exonuclease-dead version) (Korona *et al.*, 2011). Similar results were found for the equivalent mutation in budding yeast (from $2 \times 10^{-5}$ for wild-type to $24 \times 10^{-5}$ for an exonuclease-dead version) (Shcherbakova *et al.*, 2003). As

genomic instability and mutations acquisition are related to cancer development, it is legitimate to suspect Pol ε as a potential proto-oncogene or tumour suppressor gene.

### 1.3.1 Exonuclease domain mutations in budding yeast

Pol ε exonuclease activity was first inactivated in budding yeast (Morrison *et al.*, 1991). Comparing amino acid sequence of *E. coli* DNA Pol I and Pol2 in budding yeast, Morrison and colleagues identified a highly conserved region – FDIET –located in the EXO motif I of Pol ε. Related to DNA Pol I in *E. coli*, they decided to mutate the aspartic acid at position 290 and glutamic acid at position 292 to alanine in Pol2, as they are the corresponding active sites residues in Pol I. As expected, these two mutations abolished the exonuclease activity of Pol ε. The resulting strain (*pol2-4*) did not exhibit any growth defects or sensitivity to DNA-damaging agent, indicating that 3'-5' exonuclease activity is not essential in budding yeast. As observed in the *in vitro* assays, the strains exhibit an increase in spontaneous mutation rate varying from 5- to 43-fold. Similar observations were concluded from inactivation of the proofreading activity of Pol δ (Morrison *et al.*, 1993).

### 1.3.2 Exonuclease domain mutations in mouse

Eight more years had to pass before the establishment of the first link between lack of exonuclease activity in DNA polymerases and cancer. In 2001, engineered mice lacking the exonuclease activity of Pol δ exhibited an increase incidence in cancer in the homozygous $Pold1^{D400A/D400A}$ strains in parallel to an overall increase in spontaneous mutations (Goldsby *et al.*, 2002; Goldsby *et al.*, 2001). The first link between cancer

development and Pol ε appeared only 10 years ago. Similar to Goldsby and colleagues, mice were engineered in order to mutate the two previously identified exonuclease active sites of Pol ε (D272 and E274 in human and mouse, similar to *pol2-4* in budding yeast) (Albertson *et al.*, 2009). As observed for Pold1$^{D400A/D400A}$ strains, homozygous Pole$^{e/e}$ exhibit an increased incidence of cancer development, as well as an increase in the spontaneous mutation rate compared to the heterozygous version. Notably, inactivation of Pol ε or Pol δ exonuclease activity would lead to a unique spectrum of spontaneous cancers. Indeed, Pol δ mutant mice developed more readily thymic lymphomas and tail skin carcinomas compared to Pol ε mutant mice, whereas intestinal adenomas/adenocarcinomas and histiocytic sarcoma were more frequent in Pole$^{e/e}$ strains. This disparate phenotype suggests different function of Pol ε and Pol δ during chromosomal replication. In addition, embryonic lethality was observed for strains with both inactivation of proofreading activity in Pol ε and Mlh1 depletion inactivating the MMR repair pathway. This emphasises that mutations incorporated due to lack of proofreading are largely corrected by MMR.

### 1.3.3 Human cancers and Pol ε exonuclease domain mutations

#### 1.3.3.1 Somatic *POLE*$^{EDM}$

A link between Pol ε exonuclease capability and human cancer development was first established in 2011 via the targeted sequencing analysis of 76 colorectal carcinomas and six colorectal cell lines. In one of the tumours, a T>C transition in a TTT context, giving rise to a F397S amino acid change in the exonuclease domain of Pol ε was identified (Aller *et al.*, 2011). This amino acid is conserved in all B family DNA polymerases and

located in the EXO II motif, adjacent to the aspartate coordinating the $Mg^{2+}$ ion in the exonuclease site (Shamoo *et al.*, 1999). In addition, this tumour also carried a mutation in Mlh1, inactivating MMR, translated by MSI.

The Cancer Genome Atlas (TCGA) program was initiated in 2006 in a joint effort between the National Cancer Institute and the National Human Genome Research Institute with the common aim to obtain a comprehensive understanding of the genomic alterations that underlie all major cancers. Over the years, the TCGA has generated more than 2.5 petabytes of publicly accessible data. These data are a combination of whole genome and exome sequencing, transcriptome, epigenomic and proteomic data covering over 33 cancer types and matched normal and tumoral tissues. Thanks to this prodigious amount of data, the TCGA network set out to characterise alterations present in colorectal carcinomas (Cancer Genome Atlas, 2012). Combining whole genome and exome sequencing, the data showed that a subset of tumours (16 %) exhibits a mutation frequency of over 10 mutations per $10^6$ bases – called now hypermutated – 80 % of which were microsatellite unstable and epigenetically silenced Mlh1. Interestingly, 38 % of the hypermutated tumours (3 % of the full data set) harboured a mutation in the exonuclease domain of Pol ε. Even more striking, some tumours carrying mutations in the exonuclease domain were ultramutated, i.e. had a mutation frequency of over 100 per $10^6$ bases and also carried alterations in MMR genes. This led to the conclusion that MMR deficiency could be compatible with Pol ε proofreading deficiency (even if the functional alteration of those MMR genes in the TCGA study were not tested, not enabling us to confirm MMR impairment). A similar study at a lower scale with analysis of 72 colorectal cancer (CRC) exomes showed a subset of recurrent ultramutated tumours which were microsatellite stable (MSS) with

mutations located in the exonuclease domain of Pol ε (Seshagiri *et al.*, 2012). In 2013, the TCGA network published a similar analysis based on endometrial cancer (Cancer Genome Atlas Research *et al.*, 2013). Exome sequencing combined with DNA methylation analysis highlighted one subgroup representing 17 % of the studied endometrial tumours, with an ultramutated phenotype, and mutations located in the exonuclease domain of Pol ε. Taken together, identification of mutations in the proofreading domain of Pol ε associated with a spontaneous increase in mutation rate in human cancers suggests further investigation.

Numerous Pol ε exonuclease domain mutations (EDM) are reported in the literature. However, not all tumours carrying a mutation in the exonuclease domain of Pol ε exhibit an either hyper or ultramutated phenotype. Nonetheless, some *POLE*$^{EDMs}$ stand out as they are frequently observed and present features likely to indicate pathogenicity.

***Most common somatic POLE*$^{EDM}$ *– POLE*$^{P286R}$**

*POLE*$^{P286R}$ is one of the most recurrent somatic *POLE* variants observed in a wide range of cancers such as colorectal, endometrial, breast cancer and gliobastomas, and is associated with an ultramutated phenotype (Rayner *et al.*, 2016). This mutation is potentially genetically dominant at it is found most of the time (if not always) in a heterozygous status. This suggests that *POLE*$^{P286R}$ is probably a neomorphic mutation, not exhibiting either a gain or loss of function, but producing a new function not intrinsic to the wild-type allele. Located just at the edge of the EXO I motif, within the DNA binding pocket (Figure 1.4), this proline is highly conserved throughout a wide range of species (Figure 1.5) An analogous mutation in haploid budding yeast shows a strong

mutator phenotype exceeding that of proofreading-deficient mutants (Kane *et al.*, 2014), a phenotype also observed in the diploid state (Barbari *et al.*, 2018).

Conditional knock-in mice for *POLE*$^{P286R}$ showed that heterozygous carriers are more cancer-prone than the wild-type cohort, where the most common malignancies observed are lung adenocarcinomas and aggressive $CD3^{+}CD19^{-}$ thymus-based T cell lymphomas (H. D. Li *et al.*, 2018). The hemizygous *POLE*$^{P286R}$ state is predominantly lethal as very rare escaper animal survived (2 out of 164 live-born progenies were hemizygous) but rapidly succumbed to aggressive T cell lymphomas. Finally, an overall increase in the mutation rate was observed for heterozygous mouse embryonic fibroblasts, with an even higher rate for the hemizygous status.

To better characterise the pathogenic mechanism of this variant, cell-free experiments were performed in order to assess its exonuclease activity. The human catalytic subunit of Pol ε carrying the P286R mutation shows no exonuclease activity, similar to the exonuclease-dead version of POLE (Shinbrot *et al.*, 2014). A more recent study used the purified budding yeast holoenzyme Pol ε carrying the analogue of P286R and showed that actually, this variant contains residual exonuclease activity. *In vitro* assays unexpectedly show the Pol2-P301R (budding yeast ortholog for human POLE-P286R) holoenzyme replicates more accurately than the exonuclease dead version, although *in vivo* the mutation rate for the *pol2-P301R* strain is still higher than the exonuclease null version (Xing *et al.*, 2019). In parallel to this recent work, a companion study published the crystal structure of the budding yeast analogue *POLE*$^{P286R}$, shedding light on the mutagenic effect of this variant (Parkash *et al.*, 2019). It seems that the presence of the arginine acts like an arm impeding the space supposedly to be occupied by the 3'-terminal nucleotide of the primer when Pol ε is in editing mode. In addition,

this mutation affects the metal binding at the exonuclease. Taken together, this mutation is more likely to impair catalysis, but also to prevent proper positioning of the residue to excise within the exonuclease site; while looking at the exonuclease dead version, the mutation prevents hydrolysis but does not create steric hindrance for the movement of the 3'end to the exonuclease site.

Thus, the current hypothesis suggested to explain the mutagenic property of *POLE*$^{P286R}$ is that the variant is unable to accommodate ssDNA in the exonuclease site, therefore 'forced' to remain in polymerase mode. This results in a higher activity, faster DNA synthesis, increased extension from non-corrected mismatch or bypass of secondary structures such as hairpins, resulting in an overall increase in mutation rate. However, this does not link the *in vitro* and *in vivo* experiments as Pol2-P301R replicates more accurately than the exonuclease dead version *in vitro*, whereas it is more error-prone *in vivo*.

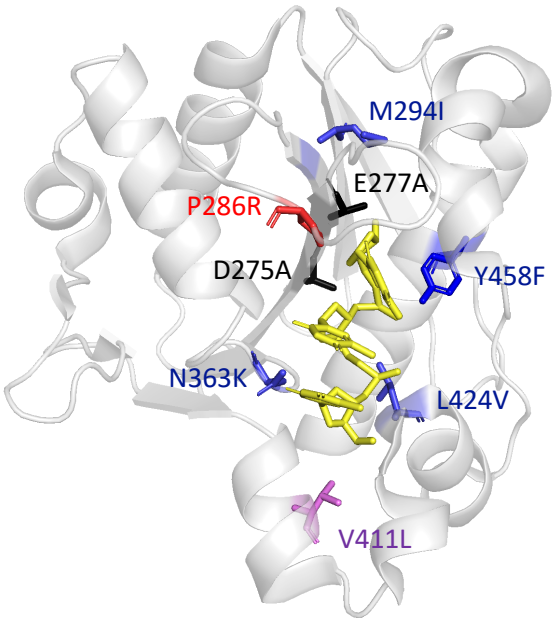



**Figure 1.4 Structure of Pol ε exonuclease domain highlighting the studied variants**

*S. cerevisiae* Pol ε exonuclease domain was isolated from PDB ID 4M8O (Hogg *et al.*, 2014) (grey) and combined with single-stranded DNA from the aligned bacteriophage T4 polymerase complex extracted from PDB ID 1NOY (Wang *et al.*, 1996) (yellow). To make the reading easier, the EDMs highlighted are the human equivalent. The position of the exonuclease active sites (D275 and E277) are indicated in black, while the germline and somatic EDM studied here are highlighted in blue and red, respectively. Note that V411L is highlighted in purple as described as both germline and somatic variant.

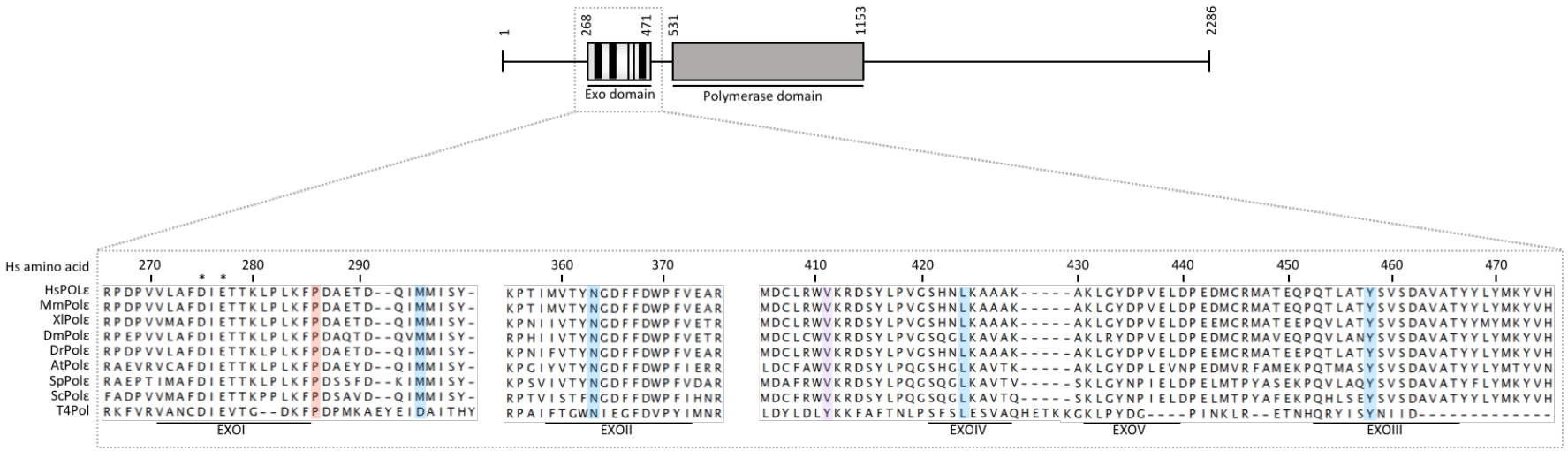



**Figure 1.5 Studied *POLE*$^{EDMs}$ are highly conserved across species**

Top panel: Schematic representation of the human catalytic subunit of Pol ε. The 2286 amino acid long subunit contains the two catalytic activities of Pol ε: 3'-5'exonuclease and 5'-3' polymerase activity represented by the EXO domain and Polymerase domain, respectively. Note that the exonuclease domain possesses five highly conserved domains called Exo motifs (I-V), highlighted in black. Bottom panel: Zoom in of the Pol ε exonuclease domain. Amino acid alignment of Pol ε using T-Coffee (Notredame *et al.*, 2000) as a Multiple Sequence Alignment software (default parameters; no matrix). Germline, somatic and germline/somatic EDM are highlighted in blue, red and violet, respectively, while the two catalytic sites are highlighted with a * and Exo motifs are underlined. Hs: human; Mm: mouse; Xl: *Xenopus laevis*; Dm: drosophilia; Dr: *Danio rerio*; At: *Arabidopsis thalinana*; Sp: *Schizosaccharomyces pombe*; Sc: *Saccharomyces cerevisiae*; T4Pol: T4 DNA polymerase.

### 1.3.3.2 Germline *POLE*$^{EDM}$

The studies previously mentioned are all related to somatic mutations identified in cancers. However, it was not established until recently if germline mutations in the exonuclease domain of Pol ε could be related to cancer predisposition. Linkage analysis in combination with whole genome sequencing of families with an apparent increased risk for CRC development – probably caused by heritable but yet unidentified genetic modifications – enabled a link between *POLE*$^{EDM}$ and CRC predisposition. Indeed, via the Oxford-Illumina WGS500 project, the *POLE*$^{L424V}$ variant was detected in a family with a high burden of CRC. (Palles *et al.*, 2013). Named Polymerase Proofreading-Associated Polyposis (PPAP), this syndrome is characterised by dominantly inherited predisposition to development of a variable number of colorectal adenomas and carcinomas, related to impairment of proofreading activity of Pol ε and Pol δ.

#### *POLE*$^{L424V}$

The variant *POLE*$^{L424V}$ has been described as both somatic and germline mutation: somatic mutation in endometrial (Cancer Genome Atlas Research *et al.*, 2013) and breast cancer (Cerami *et al.*, 2012), and germline as associated with CRC predisposition and glioblastoma (Palles *et al.*, 2013). The *POLE*$^{L424V}$ germline mutation associated with PPAP is heterozygous and highly penetrant. The EDM is located in the EXO IV motif, within the DNA binding pocket (Figure 1.4) and is highly conserved across species (Figure 1.5). Mutation of the equivalent residue in Pol δ leads to exonuclease activity depletion and an increase in spontaneous mutation rate in budding yeast, emphasizing the importance of the residue (Murphy *et al.*, 2006). *In vitro* assay using the human catalytic

subunit POLE-L424V showed serious exonuclease activity impairment (Shinbrot *et al.*, 2014). In budding yeast, the *pol2-L439V* strain (ortholog for human *POLE*$^{L424V}$) exhibits an increase in mutation rate compared to a wild-type version of Pol ε (Barbari *et al.*, 2018). Notably, the mutation rate is approximatively two-fold higher than that of the exonuclease-dead version of Pol ε, suggesting that loss of exonuclease activity only is not sufficient to explain the increase in spontaneous mutation rate for this variant.

***POLE*$^{M294I}$**

The *POLE*$^{M294I}$ variant does not appear in the literature. It was identified via the Oxford-Illumina WGS500 project as a heterozygous mutation – NM_006231:c.82G>A; pMet294Ile – in a 63-year old women who had CRC and liver metastases (Palles, unpublished data). Not located in any of the identified EXO motifs of Pol ε, the affected residue is however highly conserved across species (Figure 1.5), located in the DNA binding pocket of the exonuclease domain (Figure 1.4). It is also assessed as deleterious via Condel (Gonzalez-Perez *et al.*, 2011) and might destabilise the ion metal coordination and the protein structure (Hamzaoui *et al.*, 2020), suggesting further investigation in its role for cancer predisposition.

***POLE*$^{N363K}$**

In 2014, a new germline *POLE*$^{EDM}$ variant was identified in a Norwegian family with a high burden of CRC, in addition to a broad range of extra intestinal tumours such as ovarian, endometrial and brain tumours, and a case of pancreatic cancer (Rohlin *et al.*, 2014). The heterozygous mutation – NM_006231.2:c.1089C>A; pAsn363Lys – is located in the EXO II motif and is highly penetrant. The mutated residue is highly conserved

across species (Figure 1.5), and is predicted to have a profound effect on the substrate binding at the active site of the exonuclease domain as located in the DNA binding pocket (Hamzaoui *et al.*, 2020) (Figure 1.4). Indeed, the side chain $NH_2$-group from this asparagine can form a direct hydrogen bound with the 4' oxygen of the substrate backbone present in the DNA binding pocket of the exonuclease domain (Rohlin *et al.*, 2014). This interaction is important for exonuclease activity as it has been observed for Pol I in *E. coli*. Indeed, in DNA Pol I in *E. coli*, mutation of the residue to alanine decreases the rate of excision by 100-fold, as the alanine chain is rather small and does not impose steric hinderance (Beese *et al.*, 1991). However, it is not identified as playing a role in the protein structure nor ion metal coordination (Hamzaoui *et al.*, 2020). Finally, the tumours were MSS and no second hits were found.

A second family was reported with the same Pol ε alteration (Vande Perre *et al.*, 2018). It also exhibited a high burden in CRC, and the carriers presented a striking increased risk of brain cancer. Second hits were not looked for in this study, and no alterations in MMR associated genes were detected (MLH1, MSH2, MSH6 and PMS2), concordant with MSS observed.

Taken together, these two studies allowed to assess that the *POLE*$^{N363K}$ variant increases the incidence of glioblastoma to at least 17.4 % in the carrier population (Vande Perre *et al.*, 2018). For CRC incidence, combined family data increases the incidence of CRC to 69.5 % in the carrier population (Vande Perre *et al.*, 2018).

A recently published article reported the hypermutagenicity of this variant in budding yeast (Hamzaoui *et al.*, 2020). Unfortunately, up to now there is no *in vitro* analysis describing how this mutation could possibly alter proofreading capability.

***POLE*$^{V411L}$**

The *POLE*$^{V411L}$ variant was initially described as a somatic mutation observed in a broad range of cancers such as CRC, endometrial, glioblastoma, ovarian, and gastric cancer (Rayner *et al.*, 2016). However, it has recently been detected as a heterozygous germline mutation in the youngest reported PPAP patient with colorectal carcinoma and extensive polyposis (Wimmer *et al.*, 2017). This 14-year-old patient presented with symptoms indicating MMR deficiency. However, no germline mutations were detected in MUTYH or APC, nor in MLH1, MSH2 and PMS2. The patient also presented MSS which could be explained by the absence of mutations in the MMR genes tested. Nevertheless, sequencing of the exons of the exonuclease domain of Pol ε revealed the following variant: NM_006231.3:c.1231G>C; p.Val411Leu. Flanking the EXO IV motif of the exonuclease domain of Pol ε, this residue is highly conserved across species (Figure 1.5) and located in vicinity of the DNA binding pocket of the exonuclease domain (Figure 1.4) and is likely to have a functional impact on the protein. In addition, *in vitro* assays using the human catalytic subunit POLE-V411L shows substantially reduced exonuclease activity (Shinbrot *et al.*, 2014). Notably, even associated with spontaneous increase in mutation rate in tumour samples, in haploid budding yeast this variant does not exhibit an increased mutation rate compare to the wild-type version of Pol ε, which indicates that some functional aspects of Pol ε may not be fully conserved between mammalian cells and yeast (Barbari *et al.*, 2018).

***POLE*$^{Y458F}$**

For over two decades, a large family was followed at a Norwegian hospital as numerous of their members presented with a very high burden of cancer and adenomas (about 40

individuals over approximately 100 individuals in total) (Hansen *et al.*, 2015). The striking autosomal dominant inheritance pattern observed generation after generation in this family lead to the suspicion that a highly penetrant mutation was carried in this family. Initially thought to be affected with a polyposis syndrome, no germline mutations were to be found in either APC nor MUTYH, related to familial adenomatous polyposis and MUTYH-associated polyposis, respectively. No abnormalities were detected either in MMR genes at a germline level (MLH1, MSH2, MSH6 and PMS2). Exome sequencing eventually enabled the identification of the potential responsible mutation. Located in POLE, the germline NM_006231.2:c.1373A>T; p.Tyr458Phe variant was detected as a heterozygous mutation. This mutation seems to predispose to adenomas and carcinomas not only in the colon and rectum but also in pancreas, small intestine, stomach and ovaries. This missense mutation located in the EXO III motif of the exonuclease domain of Pol ε is suspected to have a functional impact on the protein. The mutated residue is highly conversed across species (Figure 1.5), and the exonuclease activity is impaired in equivalent mutations in other species, leading to decrease in polymerase fidelity and increase in mutation rate in polymerases from *E. coli*, T4 and ϕ29 bacteriophages (Abdus Sattar *et al.*, 1996; Derbyshire *et al.*, 1991; Soengas *et al.*, 1992). The exact role of this tyrosine is still unclear. However, structural data from *E. coli* DNA Pol I Klenow fragment (Freemont *et al.*, 1988) suggested that its function may be to orientate the attacking hydroxide ion necessary for the first step of excision during transition state from polymerase to exonuclease function. An additional study using *E. coli* DNA Pol I also described an increase binding of DNA substrate to the exonuclease site when the tyrosine residue is mutated into alanine (Lam *et al.*, 1998). Altogether, it seems that the tyrosine 458 of Pol ε is important for orientating the DNA substrate from

the binding conformation to the catalytically active conformation, making the DNA more accessible for hydrolysis (Figure 1.4). No second hits where investigated in this study, or microsatellite stability/instability status. Unfortunately, as yet there is no *in vitro* data describing how this mutation could alter proofreading capability.

## 1.4 Mutational Signatures

Cancer development can be related to Darwin's theory of evolution: first, heritable genetic variations can be acquired randomly and passed down from one generation to another, and second, natural selection promotes emergence of the fittest in a given environment. One asset for cancer development is genomic instability. Amongst genomic instability lies mutation acquisition, which happens all along the life span of an individual, as a result of exposure to endogenous or exogenous sources. One of the first exogenous sources identified as linked with cancer development was exposure to UV radiation and skin cancer (Baumann, 1939). Later was observed that UV light can lead to the formation of dimers of two adjacent pyrimidines, with preferences towards thymidine-thymidine dimers (Witkin, 1969) and that this DNA damage predominantly results in cytosine to thymidine (C>T), or cytosine-cytosine to thymidine-thymidine (CC>TT) mutations (Setlow *et al.*, 1966). However, it was still unclear whether the mutations caused by UV light were responsible for neoplastic expansion. The answer came from the combined analysis of skin cancer samples from patients with tumours originating in sun-exposed skin, and sequencing of the TP53 gene, already linked with neoplasia. Sanger sequencing revealed the characteristic C>T and CC>TT mutations patterns caused by UV irradiation, highlighting the fact that specific exogenous source of DNA damage can leave distinct scars or mutational imprints over the genome, these

potentially leading to cancer development when hitting a tumour suppressor gene, in this example (Brash *et al.*, 1991). There is a multitude of cancer types which, up to now, are still not associated with a causal factor. Furthermore, it is now clear that cancer does not arise due to a single molecular alteration, but to the cumulative effect over the lifespan of a cell. Knowing that our genome bears the scars of past carcinogenic contacts, examination of patterns of somatic mutations present in a cancer cell could allow identification of the aetiology of the operative mutational process, therefore its cause.

As previously mentioned, specific biological processes can generate specific catalogue of somatic mutations. This mutational process comprises both a DNA damage component, and a DNA repair one, ongoing or part of the past. Each mutational process leaves a characteristic imprint in the cancer genome – a mutational signature. At a given time, a cancer genome exhibits a mutational portrait which is the sum of all the mutational signatures resulting from the many mutational processes the cell was exposed to (Helleday *et al.*, 2014). In order to identify such signatures and unravel hidden mutational processes, it is necessary to considerably increase the scale of the catalogues of somatic mutations. Thanks to recent advance in sequencing technologies and collective efforts, researchers have increased the scale of mutation pattern studies. Instead of focusing on the sequencing of a unique reporter gene – like the TP53 –, whole genome, exome and transcriptome sequencing of numerous cancer types would allow identification of new driver mutations, as well as mutational processes hiding behind somatic mutations. In addition to collaborative efforts, algorithms have been developed to extract the mutational signatures from catalogues of somatic mutations. Indeed, mathematical modelling overcame the gap between mutational catalogues and mutational signatures (Alexandrov, Nik-Zainal, Wedge, Campbell*, et al.*, 2013). Explained

by a non-expert in mathematical modelling, extraction of a mutational signature could be related to the 'cocktail party problem'. Suggested by the British cognitive scientist Collin Cherry, this problem aimed to help to study the capacity of an individual to focus its attention on one conversation while many others are ongoing in a noisy room (Cherry, 1953). An individual is exposed to conversations, as a cell is exposed to mutational processes. Loudness of the conversations is the equivalent of exposure to a carcinogen. Recording of the conversations creates a catalogue of mutations, which are at that stage a jumbled composite of signatures. The aim of the mathematical approaches is to extract and identify unique conversations – the mutational signatures (Figure 1.6). One mathematical model which has been shown to extract such meaningful components is the nonnegative matrix factorisation (NMF) (D. D. Lee *et al.*, 1999). When applied to cancer mutation databases, this approach uses iterations to identify two matrices which, put together, can accurately reconstruct the distribution of the different mutations across the different mutation types. The first matrix is a group of specific mutation signatures that is defined by relative abundances of each mutation type, while the second matrix describes the strength at which the mutation signature occurs in each tumour in the analysed database. The final algorithm deciphers the minimal set of mutational signatures that optimally explains the proportion of each mutation type found in each cancer sample, and the method estimates the contribution of each signature to each cancer type.

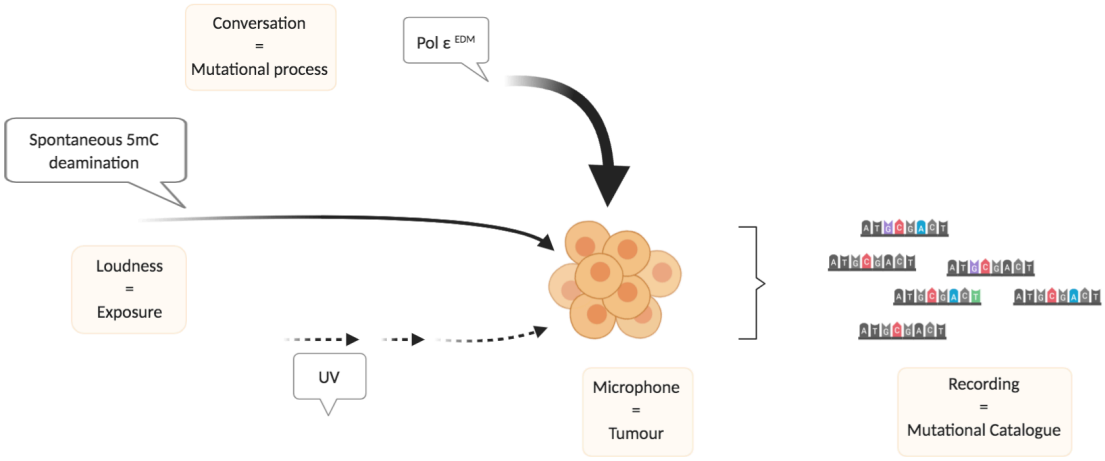



**Figure 1.6 Mutational signature extraction, or the 'Cocktail Party Problem'**

Schematic representation of acquisition of mutational catalogue prior to extraction of mutational signatures. Cells are exposed to mutational processes either all along their life (such as spontaneous 5mC deamination) or by intermittence (UV exposure). Some mutational processes lead to a strong burst in mutation (such as *POLE*$^{EDM}$). All together, these mutational processes can lead to tumour development, the tumour carrying the mutational catalogue.

This computational framework was first applied to a set of 21 breast cancer genomes (Nik-Zainal *et al.*, 2012). Mutations can be classified according to the following features: i) the mutated nucleotide, ii) the resulting one, iii) as well as the immediate sequence context at 3' and 5' end of the mutation point. These three classes can be incorporated into the mutational signature. There are by convention six classes of base substitutions referring to the pyrimidine of the mutated Watson-Crick base pair: C>A, C>G, C>T, T>A, T>C and T>G. Adding the sequence context gives an overall 96 channels of possible mutations. This is essential to help to distinguish mutational signatures that cause the same substitution in different contexts, such as for example substitutions due to UV exposure giving rise to C>T in TCC or CCC context, from C>T mutations due to spontaneous deamination of 5-mC predominantly taking place at CpG sites. After the catalogue of mutations from the set of 21 breast cancer genome was analysed using this

new computational approach, five independent single nucleotide substitution processes were identified. They seem to have been acting in combination, or simultaneously or during different phases of evolution of the cancer clone. Even if the advantage of this approach is that it is unbiased and hypothesis-free, the number of extracted signatures depends on the number of analysed samples, and the diversity and frequency of mutation signatures. Hence, low-frequency mutational signatures may be difficult to detect (such as a 'quiet' conversation). That is why an increase in scale for this type of analysis is essential.

In 2013, a study was published with the aim of applying this method to detect the repertoire of mutational signatures operating across spectrum of human neoplasia (Alexandrov, Nik-Zainal, Wedge, Aparicio*, et al.*, 2013). This consisted of the study of mutational processes across 30 different cancer type derived from 7042 primary cancers. In total, 4,938,362 somatic substitutions were fed into the algorithm. From this study, 22 validated distinct mutational signatures were extracted. It is important to mention that only 11 out of the 22 mutational signatures were associated with a mutational process. 11 remain unknown. Seven years later, improvement of high throughput sequencing techniques, and astonishing collaborative effort led to the re-analysis of mutational signatures, expanding the number of analysed tumour samples to 23,829, with an overall 79,793,266 somatic-single-base substitution (SBS) (Alexandrov *et al.*, 2020). Similar NMF methods were used, ending with the extraction of a total of 49 SBS signatures. Some of these mutational signatures were already extracted and known from previous analysis. However, the substantial increase in sample size certainly enabled the discovery of new mutational signatures, but also the separation of

overlapping ones, or decomposition of signatures related to specific DNA repair mechanisms, DNA damage, or related to replication processes.

### 1.4.1 Mutational signature and Pol ε

It is no surprise that, following the identification of *POLE*$^{EDMs}$ and its link with hyper- and ultramutated phenotype observed in tumour cells, a mutational signature was associated with this alteration (Alexandrov, Nik-Zainal, Wedge, Aparicio*, et al.*, 2013). Signature 10 is associated with impairment of Pol ε – more precisely with somatic *POLE*$^{P286R}$ and *POLE*$^{V411L}$ variants. It is predominantly operative in colorectal and uterine cancers. It is characterised by a high proportion of C>A transversions in a TCT context, C>T transitions in TCG context, as well as T>G transversions in a TTT context (Figure 1.7). In addition, it shows a transcriptional strand bias where the C>A and T>G substitutions are more readily taking place on the transcribed strand.

Very recent re-analysis of a larger set of catalogue of somatic mutations enabled the dissociation of the signature 10 into two distinct signatures named SBS10a and SBS10b (Alexandrov *et al.*, 2020) (Figure 1.7); these two signatures are usually found in samples exhibiting a hypermutated phenotype, also related to *POLE*$^{EDMs}$. SBS10a is characterised by the C>A transversion in TCT context, while SBS10b possesses the C>T transition in TCG context; both features observed in signature 10. Notably, they are usually associated with each other, in parallel with the SBS28 mutational signature, a signature of unknown aetiology mainly characterised by T>G transversions in a TTT context, the third feature characteristic of the signature 10. SBS28 contributes to a large number of somatic mutations when observed in samples with SBS10a/b signatures, whereas it contributes to only few mutations when SBS10a/b are absent. On a final note,

SBS10a exhibits a transcription strand bias for the C>A transversion in TCT context. These studies helped in determining the role of Pol ε in the mutagenesis process of neoplasms. Indeed, similarly to the signature 10, a strong strand bias is observed in tumour samples. In addition, a strand enrichment in mutations is observed in the vicinity of replication origins, strengthening the idea that Pol ε is in charge of replicating the leading strand.

To look deeper into the mechanism of mutagenesis of *POLE*$^{EDM}$ variants, combined mutational spectrum analysis of tumours carrying *POLE*$^{EDM}$ mutation and *in vitro* analysis revealed that mutational burden, even if strongly similar, can differ amongst variants (Haradhvala *et al.*, 2018; Shinbrot *et al.*, 2014). A subgroup of *POLE*$^{EDM}$ variants exhibits for most of them an ultramutated phenotype combined with an increased C>A transversions in a TCT context, as well as C>T transition in a TCG context, both features similarly observed in the signature 10. A second sub-group of *POLE*$^{EDM}$ did not exhibit an elevated context-dependent C>A and C>T mutation frequencies, suggesting a difference of behaviour related to the type of *POLE*$^{EDM}$ variant (Shinbrot *et al.*, 2014). This was confirmed for the -P286R and -V411L variants. Extraction of mutational signatures from endometrial samples carrying either POLE-P286R or -V411L variant alone led to 3 distinct signatures. Two mutational signatures strongly resembled signature 10 associated with *POLE*$^{EDM}$, whereas the third one was associated predominantly with POLE-V411L (Haradhvala *et al.*, 2018). Notably, it was shown that *POLE*$^{EDM}$ variants exhibits an enrichment in TP53 stop codon mutations (Hu Fang *et al.*, 2019; Shinbrot *et al.*, 2014).

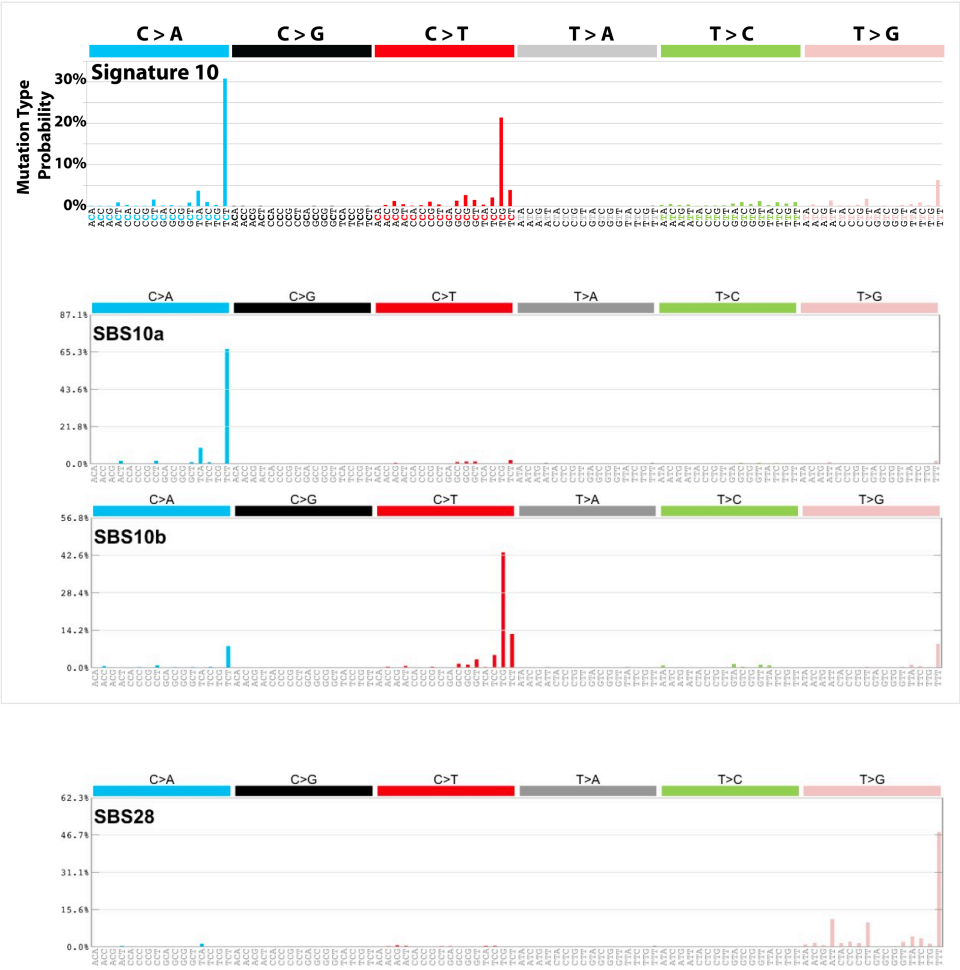



**Figure 1.7 Mutational signature 10 – Before and after**

Initial mutational signature associated with Pol ε mutation was unique signature 10 (Alexandrov, Nik-Zainal, Wedge, Aparicio*, et al.*, 2013) Re-analysis and increase in tumour samples enabled the extraction of 2 unique signatures, both associated with Pol ε mutation: SBS10a and SBS10b (Alexandrov et al., 2020). In addition, signature 28 is of unknown aetiology, but found in most samples with the signatures SBS10a and SBS10b.

### 1.4.2 Mutational signatures and MMR

As mentioned earlier, MMR is the ultimate guardian which corrects replication errors. As one could expect, mutational signatures have been associated with MMR deficiency. Alexandrov, Nik-Zainal, Wedge, Aparicio*, et al.* (2013) identified four mutational signatures associated with defective MMR: signature 6, 15, 20 and 26 (Figure 1.8). The signatures 6, 15 and 20 exhibit mainly an overall increase in C>T transversions, however the trinucleotide context differs. Signature 6 is characterised by a C>T in NCG context,

which could be related to CpG where cytosines are more prone to methylation, leading to an increased prevalence of spontaneous 5mC deamination. This leads to formation of an uracil, and eventually to a thymidine after the next round of replication. Signature 15 is characterised by C>T transition in a GCN context, while signature 20 contributes to an increase in C>T transition primarily in a CCT context. Finally, signature 26 is characterised mainly by an overall increase T>C transitions. It seems also important to mention that the mutational signatures are present in different type of cancers. Indeed, while signature 6 is found in 17 cancer types, it is more common in uterine and colorectal cancers. Signatures 15, 20 and 26 are observed in stomach cancers, while signature 26 and 20 are often observed in breast cancers; finally, signature 26 is often observed in cervical cancer and uterine carcinoma.

Re-analysis and sampling increase enabled the identification of additional mutational signatures, as well as refinement of the already extracted ones (Alexandrov *et al.*, 2020). In total, seven SBS mutational signatures associated with MMR deficiency were extracted: SBS6, SBS14, SBS15, SBS20, SBS21, SBS26 and SBS44 (Figure 1.8) (Figure 1.9). While SBS44 was not previously characterised, the signatures SBS6, SBS15, SBS20 and SBS26 are similar to the previously extracted signatures 6, 15, 20, and 26, respectively. Overall, the SBS signatures show similar features compare to the initial ones; the notable difference is the decrease in contamination by other MMR deficiency signatures. Interestingly, signature 21 was initially of unknown origin, although identified as somehow similar to signature 26 and probably related to microsatellite unstable tumours. A note on SBS20. Initially labelled as associated with MMR deficiency only, increase in sampling enabled to label this mutational signature not only to defective MMR, but to concurrent Pol $\delta$ mutations and MMR defect.

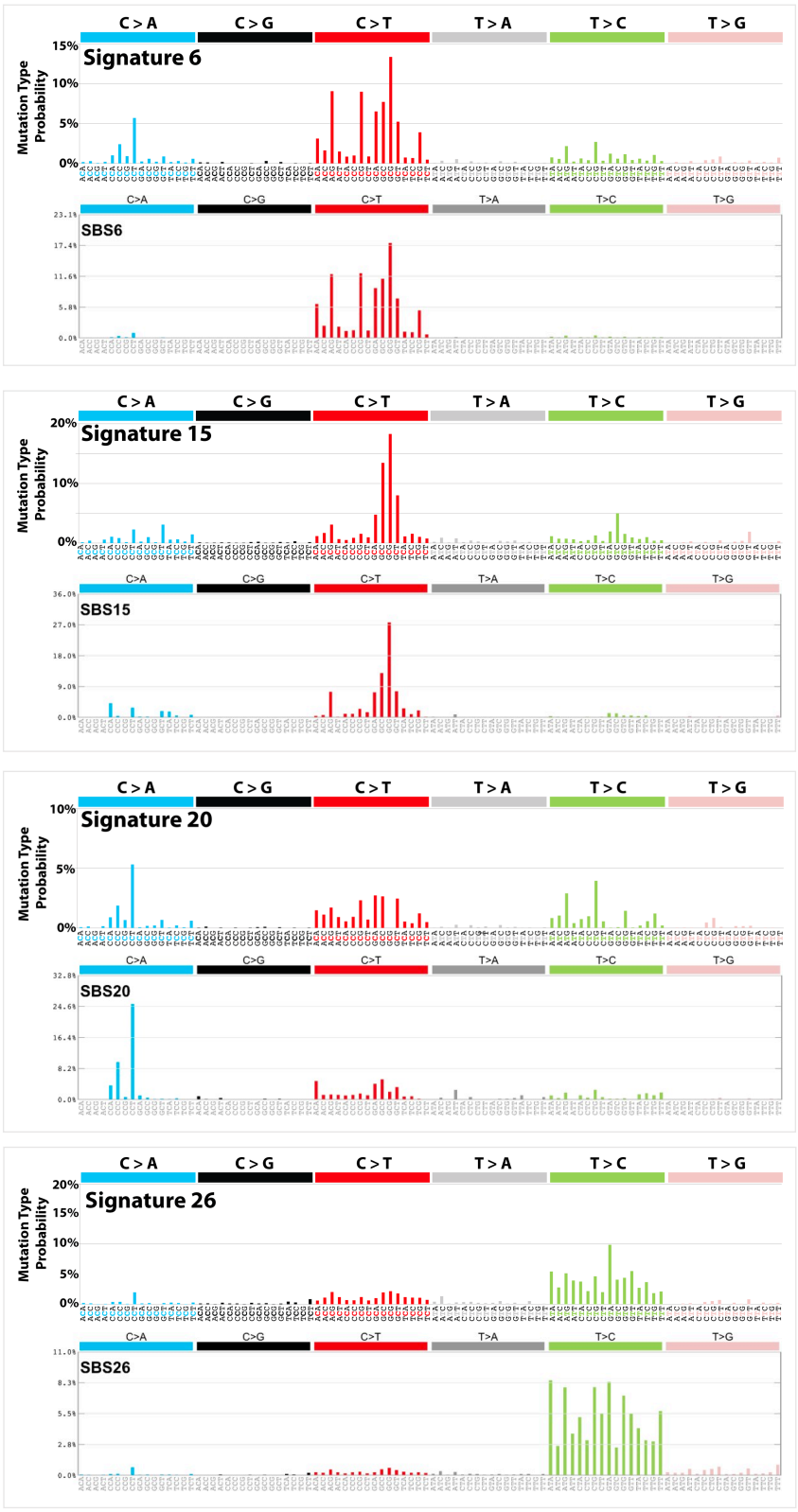



**Figure 1.8 Mutational signatures related to MMR-deficiency – Before and after**

Mutational signature 6, 15, 20 and 26 were the primary extracted mutational signatures associated with MMR-deficiency. Each grey frame contains the initial signature (top part; (Alexandrov, Nik-Zainal, Wedge, Aparicio*, et al.*, 2013)) and the newly extracted one (bottom part; (Alexandrov *et al.*, 2020)).

### 1.4.3 Intertwined mutational processes

Sampling analysis at a bigger scale led to the refinement of signature's aetiology (Alexandrov *et al.*, 2020). The aetiology of the signature 20 for instance, initially linked with MMR alteration, was refined and is now associated with concurrent Pol $\delta$ and MMR alteration. Another example is the signature 14 – initially of unknown aetiology – now associated with concurrent *POLE*$^{EDM}$ mutation and defective MMR pathway. SBS14 is characterised by C>A transversion in NCT context. C>T transitions initially observer in signature 14 were not reported in SBS14, most likely explained by a contamination by SBS5 in the initial analysis (Figure 1.9). Several tumours samples presenting evidences of MMR deficiency in combination with *POLE*$^{EDM}$ were previously reported in CRC and endometrial cancer (Cancer Genome Atlas, 2012; Cancer Genome Atlas Research *et al.*, 2013).

Looking deeper into those specific tumours, a study focused on inherited biallelic MMR deficiency (bMMRD) in children related to glioblastomas development and found tumours carrying both altered pathways (Shlien *et al.*, 2015). These brain tumours exhibited >200 mutations/Mb, and 7 out of the 10 tumours had a *POLE*$^{EDM}$ variant combined with MSS. Similar to the mutational signature 10 which was identified at that time, C>A in a TCT context, as well as T>G in a TTT context were characteristics of the mutation spectrum. Notably, this study suggested that the maximum mutation threshold a cell can bear is below 1-2 x $10^4$ mutations per exome. It appears clear that in the case of inherited bMMRD, acquisition of *POLE*$^{EDM}$ would lead to a rapid and explosive increase in the mutation burden. In a follow-up study including whole genome, exome and targeting sequencing, hypermutated phenotypes were looked at in childhood and adult cancers (Haradhvala *et al.*, 2018). Two classes of mutated

phenotypes were distinguished: hypermutated – exhibiting >10 mutations/Mb – and ultra-hypermutated – which exhibit >100 mutations/Mb. Notably, most of the ultra-hypermutated cancers carry both MMR deficiency and a *POLE*$^{EDM}$, while the hypermutated tumours would only carry either one of the altered pathways.

Study of the mutational signatures in these cancers enabled the following conclusions: some cancers had germline bMMRD, with later acquisition of *POLE*$^{EDM}$. Other cancers would conversely have a *POLE*$^{EDM}$ germline mutation, then followed by MMR deficiency. In the first case, the mutational signatures observed would mainly be related to MMR deficiency (signature 15), with a smaller portion related to *POLE*$^{EDM}$ signature 10. In the second case, the main mutational signature would be the signature 10, related to *POLE*$^{EDM}$, with a smaller proportion of the signature related to MMR (signature 15). Furthermore, not only these two signatures were observed: mutational signature 14 was also observed in the tumours carrying both alterations.

Finally, as mentioned earlier, some hypermutated tumours were identified in endometrial cancers (Cancer Genome Atlas Research *et al.*, 2013). Looking deeper into those specific tumours (Haradhvala *et al.*, 2018), *POLE*$^{EDMs}$ were observed in parallel to MLH1 silencing or truncated/frameshift mutations or deletion of either MLH1, MSH2, MSH6 or PSM2, part of the MMR pathway. Mutational signatures were extracted from samples carrying either only a *POLE*$^{EDM}$, MMR-deficiency, or both alterations. From the samples carrying *POLE*$^{EDMs}$ only, mutational signatures extracted resembled the signature 10. The mutational signatures related to MMR-deficient samples on the other hand resembled signatures 6 and 26. Interestingly, mutational signatures extracted from samples carrying both alterations exhibit minimal contributions from either the *POLE*$^{EDM}$ and MMR-deficiency signatures, but a distinct one. More surprisingly, this mutational

signature had strong similarities with the at-that-time unknown aetiology signature 14, later re-assessed as associated with concurrent *POLE*$^{EDM}$ and MMR-deficiency.

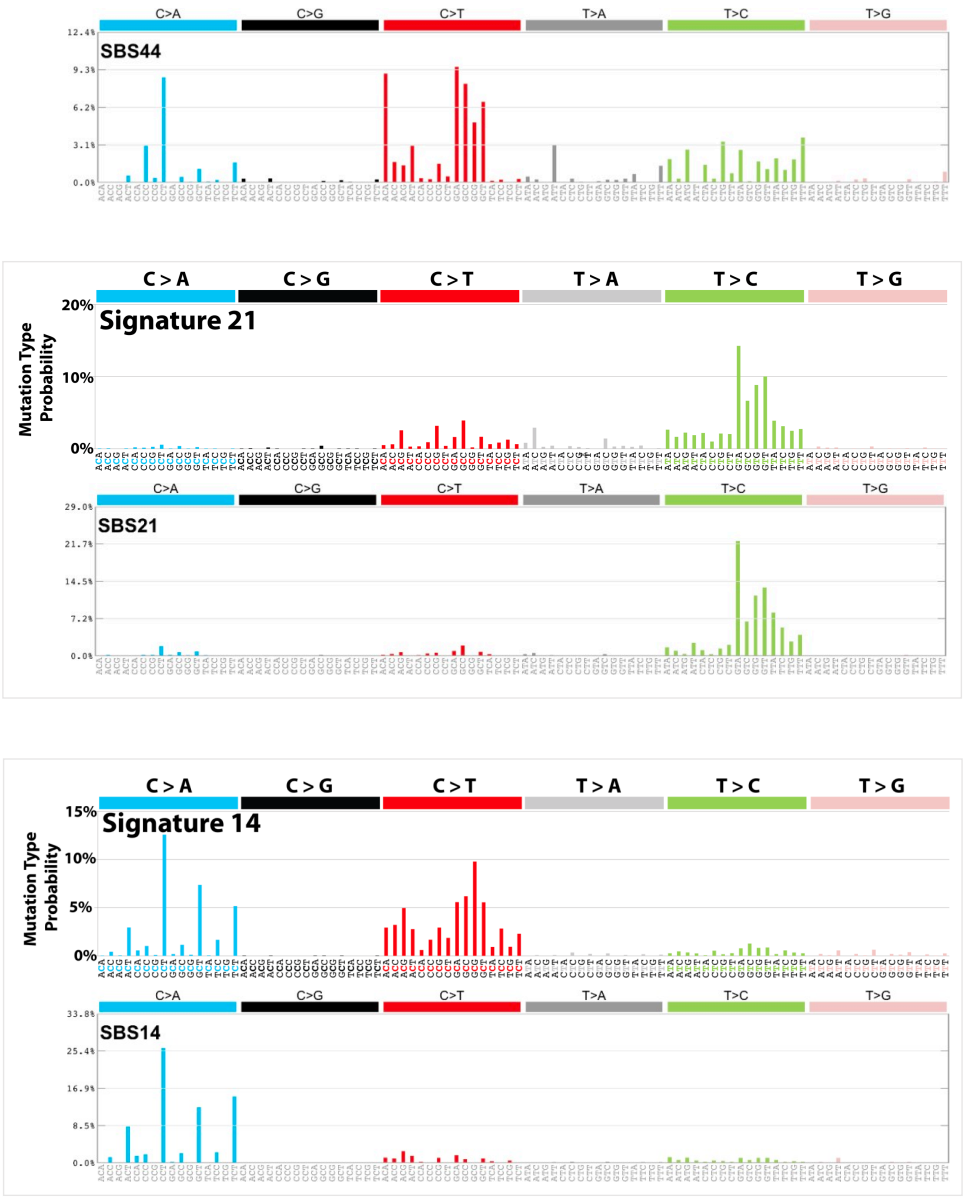



**Figure 1.9 Newly characterised, associated, and intertwined mutational signatures**

Each grey frame contains the initial signature (top part; (Alexandrov, Nik-Zainal, Wedge, Aparicio*, et al.*, 2013) and the newly extracted one (bottom part; (Alexandrov *et al.*, 2020). SBS44 as recently been newly associated with defective MMR pathway; not previously characterised. Signature 21 was initially of unknown aetiology, then established as related to MMR deficiency. Similarly to Signature 21, Signature 14 was initially of unknown aetiology although it is now associated with concurrent Pol ε mutation and defective MMR.

## 1.5 Study aims

The primary objective of this work is to study the pathogenic potential of germline *POLE*$^{EDMs}$. To do so, firstly, we constructed in fission yeast the relevant *POLE*$^{EDM}$ and aimed to evaluate their impact on mutation burden and cellular metabolism (addressed in chapter 3). Secondly, we sought to decipher the hypermutagenic mechanism of a subset of *POLE*$^{EDMs}$ whose spontaneous mutation rate appeared greater than those simply devoid of proofreading ability (addressed in chapter 4). Thirdly, we indented to study the mutational patterns of hypermutated strains combined with loss of MMR in order to corroborate clinical observations while investigating the *POLE*$^{EDM}$ mutational mechanism (addressed in chapter 5).

# Chapter 2

## Materials and Methods

### 2.1 *Schizosaccharomyces pombe* methods

#### 2.1.1 Growth and maintenance

Cells were grown on either restrictive minimal media or nutrient rich media (Moreno *et al.*, 1991; Petersen *et al.*, 2016) supplemented with the required amino acids, and at permissive temperature (26°C for temperature-sensitive strains, otherwise 30°C). All *in vivo* experiments were performed from strains freshly woken up as patches on an agar rich media plate. Cells were stored at 4°C for up to 7 days. Cell culture concentration was determined via Optical Density ($OD_{600}$) using an Eppendorf Biophotometer, where $OD_{600}$ = 1 corresponds to ca. $10^7$ cells/mL. If not otherwise stated, liquid cultures were initiated by inoculating 20 mL of media with a 1 µL loop of cells and incubated overnight in a water bath (Innova 3100 – New Brunswick Scientific) at permissive temperature with shaking at 150 rpm. If not otherwise stated, all *in vivo* experiments were performed on exponential growing culture. Overnight saturated cultures were refreshed at $OD_{600}$ = 0.3 and cells were allowed to cycle twice (4 hours) to get to exponential phase. Long term storage was achieved in glycerol at 25 % of a liquid growing culture or scratch of cell patches. Cells were frozen down and stored at -80°C in Nunc CryoTubes™.

## Table 2.1 Strains

| Number | Genotype | Source |
|---|---|---|
| 11 | ***diploid**; h-/h-* | Collection |
| 1123 | *h+ ura4-D18 ex TH 360* | Collection |
| 2758 | *h+ rad3△::clonat ade6-M210 leu1-32 ura4-D18* | Collection |
| 2840 | ***Parental strain/wild-type**; h- ade6-485* | Collection |
| 3061 | ***rad3ts**; h+ rad3ts ura4-D18 leu1-32 ade6 (M216/704)* | Collection |
| 3169 | ***msh2△**; h+ msh2::his3 ade6-424 his3-D1* | Collection |
| 3186 | ***cdc20$^{WT\text{-}construct}$**; h- pol2 (E) WT::kanMX6 ade6-485 ex 2840* | Collection |
| 3188 | ***cdc20$^{L425V}$-01**; h- pol2 (E) L425V::kanMX6 ade6-485 h- ex 2840 (human L424V)* | Collection |
| 3189 | ***cdc20$^{L425V}$-02**; h- pol2 (E) L425V::kanMX6 ade6-485 h- ex 2840 (human L424V)* | Collection |
| 3190 | ***cdc20$^{L425V}$-03**; h- pol2 (E) L425V::kanMX6 ade6-485 h- ex 2840 (human L424V)* | Collection |
| 3199 | ***cdc20$^{exonull}$**; h- pol2 (E) D276A/E278A::kanMX6 ade6-485* | Collection |
| 3201 | ***cdc20$^{V412L}$-01**; h- pol2 (E) V412L::kanMX6 ade6-485 h- ex 2840 (human V411L)* | Collection |
| 3202 | ***cdc20$^{V412L}$-02**; h- pol2 (E) V412L::kanMX6 ade6-485 h- ex 2840 (human V411L)* | Collection |
| 3203 | ***cdc20$^{V412L}$-03**; h- pol2 (E) V412L::kanMX6 ade6-485 h- ex 2840 (human V411L)* | Collection |
| 3218 | ***cdc20$^{P287R}$**; h- pol2 (E) P287R::kanMX6 ade6-485 h- ex 2840 (human P286R)* | Collection |
| 3252 | *top1::kanMX6 ura4 leu1 ade6* | Collection |
| 3285 | ***cdc22$^{D57N}$**; h+ cdc22-D57N* | Collection |
| 3643 | *h+ rad3△ leu1-32* | Collection |
| 3669 | ***cdc20$^{M295I}$-01**; h- pol2 (E) M295I::kanMX6 ade6-485 h- ex 2840 (human M294I)* | This study |
| 3670 | ***cdc20$^{M295I}$-02**; h- pol2 (E) M295I::kanMX6 ade6-485 h- ex 2840 (human M294I)* | This study |
| 3671 | ***cdc20$^{M295I}$-03**; h- pol2 (E) M295I::kanMX6 ade6-485 h- ex 2840 (human M294I)* | This study |
| 3672 | ***cdc20$^{N364K}$-01**; h- pol2 (E) N364K::kanMX6 ade6-485 h- ex 2840 (human N363K)* | This study |
| 3673 | ***cdc20$^{N364K}$-02**; h- pol2 (E) N364K::kanMX6 ade6-485 h- ex 2840 (human N363K)* | This study |
| 3674 | ***cdc20$^{N364K}$-03**; h- pol2 (E) N364K::kanMX6 ade6-485 h- ex 2840 (human N363K)* | This study |
| 3680 | ***cdc20$^{Y459F}$-01**; h- pol2 (E) Y459F::kanMX6 ade6-485 h- ex 2840 (human Y458F)* | This study |
| 3681 | ***cdc20$^{Y459F}$-02**; h- pol2 (E) Y459F::kanMX6 ade6-485 h- ex 2840 (human Y458F)* | This study |
| 3682 | ***cdc20$^{Y459F}$-03**; h- pol2 (E) Y459F::kanMX6 ade6-485 h- ex 2840 (human Y458F)* | This study |
| 3783 | ***kpa1△**; h+ kpa1::kanr ade6-M21 leu1-32 ura4-D18* | Collection (Chen *et al.*, 2014) |
| 3814 | ***Exo/rad3△-01**; pol2 (E) D276A/E278A::kanMX6 rad3△::clonat ade6? leu1-32? ura4-D18? clone 1* | This study |
| 3815 | ***Exo/rad3△-02**; pol2 (E) D276A/E278A::kanMX6 rad3△::clonat ade6? leu1-32? ura4-D18? clone 2* | This study |
| 3816 | ***Exo/rad3△-03**; pol2 (E) D276A/E278A::kanMX6 rad3△::clonat ade6? leu1-32? ura4-D18? clone 3* | This study |
| 3817 | ***P287R/rad3△-01**; pol2 (E) P287R::kanMX6 rad3△::clonat ade6? leu1-32? ura4-D18? clone 1* | This study |
| 3818 | ***P287R/rad3△-02**; pol2 (E) P287R::kanMX6 rad3△::clonat ade6? leu1-32? ura4-D18? clone 2* | This study |

| | | |
|---|---|---|
| 3819 | ***P287R/rad3⧋-03****; pol2 (E) P287R::kanMX6 rad3⧋::clonat ade6? leu1-32? ura4-D18? clone 3* | This study |
| 3820 | ***N364K/rad3⧋-01****; pol2 (E) N364K::kanMX6 rad3⧋::clonat ade6? leu1-32? ura4-D18? clone 1* | This study |
| 3821 | ***N364K/rad3⧋-02****; pol2 (E) N364K::kanMX6 rad3⧋::clonat ade6? leu1-32? ura4-D18? clone 2* | This study |
| 3822 | ***N364K/rad3⧋-03****; pol2 (E) N364K::kanMX6 rad3⧋::clonat ade6? leu1-32? ura4-D18? clone 3* | This study |
| 3823 | ***Exo/rad3ts-01****; pol2 (E) D276A/E278A::kanMX6 rad3ts ade6? leu1-32? ura4-D18? clone 1* | This study |
| 3824 | ***Exo/rad3ts-02****; pol2 (E) D276A/E278A::kanMX6 rad3ts ade6? leu1-32? ura4-D18? clone 2* | This study |
| 3825 | ***Exo/rad3ts-03****; pol2 (E) D276A/E278A::kanMX6 rad3ts ade6? leu1-32? ura4-D18? clone 3* | This study |
| 3826 | ***P287R/rad3ts-01****; pol2 (E) P287R::kanMX6 rad3ts ade6? leu1-32? ura4-D18? clone 1* | This study |
| 3827 | ***P287R/rad3ts-02****; pol2 (E) P287R::kanMX6 rad3ts ade6? leu1-32? ura4-D18? clone 2* | This study |
| 3828 | ***P287R/rad3ts-03****; pol2 (E) P287R::kanMX6 rad3ts ade6? leu1-32? ura4-D18? clone 3* | This study |
| 3829 | ***N364K/rad3ts-01****; pol2 (E) N364K::kanMX6 rad3ts ade6? leu1-32? ura4-D18? clone 1* | This study |
| 3830 | ***N364K/rad3ts-02****; pol2 (E) N364K::kanMX6 rad3ts ade6? leu1-32? ura4-D18? clone 2* | This study |
| 3831 | ***N364K/rad3ts-03****; pol2 (E) N364K::kanMX6 rad3ts ade6? leu1-32? ura4-D18? clone 3* | This study |
| 3832 | ***Exo/kpa1⧋-01****; pol2 (E) D276A/E278A::kanMX6 kpa1::kanr ade6? leu1-32? ura4-D18? clone 1* | This study |
| 3833 | ***Exo/kpa1⧋-02****; pol2 (E) D276A/E278A::kanMX6 kpa1::kanr ade6? leu1-32? ura4-D18? clone 2* | This study |
| 3834 | ***Exo/kpa1⧋-03****; pol2 (E) D276A/E278A::kanMX6 kpa1::kanr ade6? leu1-32? ura4-D18? clone 3* | This study |
| 3835 | ***N364K/kpa1⧋-01****; pol2 (E) N364K::kanMX6 kpa1::kanr ade6? leu1-32? ura4-D18? clone 1* | This study |
| 3836 | ***N364K/kpa1⧋-02****; pol2 (E) N364K::kanMX6 kpa1::kanr ade6? leu1-32? ura4-D18? clone 2* | This study |
| 3837 | ***N364K/kpa1⧋-03****; pol2 (E) N364K::kanMX6 kpa1::kanr ade6? leu1-32? ura4-D18? clone 3* | This study |
| 3838 | *h+ msh2::his3 ade6-424 his3-D1 - G0 - D1* | This study |
| 3839 | *h+ msh2::his3 ade6-424 his3-D1 - G100 - D1* | This study |
| 3840 | *h+ msh2::his3 ade6-424 his3-D1 - G150 - D1* | This study |
| 3841 | *h+ msh2::his3 ade6-424 his3-D1 - G200 - D1* | This study |
| 3842 | *h+ msh2::his3 ade6-424 his3-D1 - G0 - D2* | This study |
| 3843 | *h+ msh2::his3 ade6-424 his3-D1 - G100 - D2* | This study |
| 3844 | *h+ msh2::his3 ade6-424 his3-D1 - G150 - D2* | This study |
| 3845 | *h+ msh2::his3 ade6-424 his3-D1 - G200 - D2* | This study |
| 3846 | *h+ msh2::his3 ade6-424 his3-D1 - G0 - D3* | This study |
| 3847 | *h+ msh2::his3 ade6-424 his3-D1 - G100 - D3* | This study |
| 3848 | *h+ msh2::his3 ade6-424 his3-D1 - G150 - D3* | This study |
| 3849 | *h+ msh2::his3 ade6-424 his3-D1 - G200 - D3* | This study |
| 3850 | *h- pol2 (E) N364K::kanMX6 ade6-485 h- ex 2840 (human N363K) - G0 - D1* | This study |

| | | |
|---|---|---|
| 3851 | *h- pol2 (E) N364K::kanMX6 ade6-485 h- ex 2840 (human N363K) - G100 - D1* | This study |
| 3852 | *h- pol2 (E) N364K::kanMX6 ade6-485 h- ex 2840 (human N363K) - G150 - D1* | This study |
| 3853 | *h- pol2 (E) N364K::kanMX6 ade6-485 h- ex 2840 (human N363K) - G200 - D1* | This study |
| 3854 | *h- pol2 (E) N364K::kanMX6 ade6-485 h- ex 2840 (human N363K) - G0 - D2* | This study |
| 3855 | *h- pol2 (E) N364K::kanMX6 ade6-485 h- ex 2840 (human N363K) - G100 - D2* | This study |
| 3856 | *h- pol2 (E) N364K::kanMX6 ade6-485 h- ex 2840 (human N363K) - G150 - D2* | This study |
| 3857 | *h- pol2 (E) N364K::kanMX6 ade6-485 h- ex 2840 (human N363K) - G200 - D2* | This study |
| 3858 | *h- pol2 (E) N364K::kanMX6 ade6-485 h- ex 2840 (human N363K) - G0 - D3* | This study |
| 3859 | *h- pol2 (E) N364K::kanMX6 ade6-485 h- ex 2840 (human N363K) - G100 - D3* | This study |
| 3860 | *h- pol2 (E) N364K::kanMX6 ade6-485 h- ex 2840 (human N363K) - G150 - D3* | This study |
| 3861 | *h- pol2 (E) N364K::kanMX6 ade6-485 h- ex 2840 (human N363K) - G200 - D3* | This study |
| 3862 | *h- ade6-485 - G0 - D1* | This study |
| 3863 | *h- ade6-485 - G200 - D1* | This study |
| 3864 | *h- ade6-485 - G0 - D2* | This study |
| 3865 | *h- ade6-485 - G200 - D2* | This study |
| 3866 | *h- ade6-485 - G0 - D3* | This study |
| 3867 | *h- ade6-485 - G200 - D3* | This study |
| 3868 | *h- pol2 (E) D276A/E278A::kanMX6 ade6-485 - G0 - D1* | This study |
| 3869 | *h- pol2 (E) D276A/E278A::kanMX6 ade6-485 - G100 - D1* | This study |
| 3870 | *h- pol2 (E) D276A/E278A::kanMX6 ade6-485 - G150 - D1* | This study |
| 3871 | *h- pol2 (E) D276A/E278A::kanMX6 ade6-485 - G200 - D1* | This study |
| 3872 | *h- pol2 (E) D276A/E278A::kanMX6 ade6-485 - G0 - D2* | This study |
| 3873 | *h- pol2 (E) D276A/E278A::kanMX6 ade6-485 - G100 - D2* | This study |
| 3874 | *h- pol2 (E) D276A/E278A::kanMX6 ade6-485 - G150 - D2* | This study |
| 3875 | *h- pol2 (E) D276A/E278A::kanMX6 ade6-485 - G200 - D2* | This study |
| 3876 | *h- pol2 (E) D276A/E278A::kanMX6 ade6-485 - G0 - D3* | This study |
| 3877 | *h- pol2 (E) D276A/E278A::kanMX6 ade6-485 - G100 - D3* | This study |
| 3878 | *h- pol2 (E) D276A/E278A::kanMX6 ade6-485 - G150 - D3* | This study |
| 3879 | *h- pol2 (E) D276A/E278A::kanMX6 ade6-485 - G200 - D3* | This study |
| 3880 | *h- pol2 (E) Y459F::kanMX6 ade6-485 h- ex 2840 (human Y458F) - G0 - D1* | This study |
| 3881 | *h- pol2 (E) Y459F::kanMX6 ade6-485 h- ex 2840 (human Y458F) - G100 - D1* | This study |
| 3882 | *h- pol2 (E) Y459F::kanMX6 ade6-485 h- ex 2840 (human Y458F) - G150 - D1* | This study |
| 3883 | *h- pol2 (E) Y459F::kanMX6 ade6-485 h- ex 2840 (human Y458F) - G200 - D1* | This study |
| 3884 | *h- pol2 (E) Y459F::kanMX6 ade6-485 h- ex 2840 (human Y458F) - G0 - D2* | This study |
| 3885 | *h- pol2 (E) Y459F::kanMX6 ade6-485 h- ex 2840 (human Y458F) - G100 - D2* | This study |
| 3886 | *h- pol2 (E) Y459F::kanMX6 ade6-485 h- ex 2840 (human Y458F) - G150 - D2* | This study |
| 3887 | *h- pol2 (E) Y459F::kanMX6 ade6-485 h- ex 2840 (human Y458F) - G200 - D2* | This study |
| 3888 | *h- pol2 (E) Y459F::kanMX6 ade6-485 h- ex 2840 (human Y458F) - G0 - D3* | This study |
| 3889 | *h- pol2 (E) Y459F::kanMX6 ade6-485 h- ex 2840 (human Y458F) - G100 - D3* | This study |
| 3890 | *h- pol2 (E) Y459F::kanMX6 ade6-485 h- ex 2840 (human Y458F) - G150 - D3* | This study |
| 3891 | *h- pol2 (E) Y459F::kanMX6 ade6-485 h- ex 2840 (human Y458F) - G200 - D3* | This study |
| 3892 | *pol2 (E) D276A/E278A::kanMX6 msh2::his3 ade6? his3-D1? - G0 - D1* | This study |

| | | |
|---|---|---|
| 3893 | *pol2 (E) D276A/E278A::kanMX6 msh2::his3 ade6? his3-D1? - G50 - D1* | This study |
| 3894 | *pol2 (E) D276A/E278A::kanMX6 msh2::his3 ade6? his3-D1? - G100 - D1* | This study |
| 3895 | *pol2 (E) D276A/E278A::kanMX6 msh2::his3 ade6? his3-D1? - G150 - D1* | This study |
| 3896 | *pol2 (E) D276A/E278A::kanMX6 msh2::his3 ade6? his3-D1? - G200 - D1* | This study |
| 3897 | *pol2 (E) D276A/E278A::kanMX6 msh2::his3 ade6? his3-D1? - G0 - D2* | This study |
| 3898 | *pol2 (E) D276A/E278A::kanMX6 msh2::his3 ade6? his3-D1? - G50 - D2* | This study |
| 3899 | *pol2 (E) D276A/E278A::kanMX6 msh2::his3 ade6? his3-D1? - G100 - D2* | This study |
| 3900 | *pol2 (E) D276A/E278A::kanMX6 msh2::his3 ade6? his3-D1? - G150 - D2* | This study |
| 3901 | *pol2 (E) D276A/E278A::kanMX6 msh2::his3 ade6? his3-D1? - G200 - D2* | This study |
| 3902 | *pol2 (E) D276A/E278A::kanMX6 msh2::his3 ade6? his3-D1? - G0 - D3* | This study |
| 3903 | *pol2 (E) D276A/E278A::kanMX6 msh2::his3 ade6? his3-D1? - G50 - D3* | This study |
| 3904 | *pol2 (E) D276A/E278A::kanMX6 msh2::his3 ade6? his3-D1? - G100 - D3* | This study |
| 3905 | *pol2 (E) D276A/E278A::kanMX6 msh2::his3 ade6? his3-D1? - G150 - D3* | This study |
| 3906 | *pol2 (E) D276A/E278A::kanMX6 msh2::his3 ade6? his3-D1? - G200 - D3* | This study |
| 3907 | *pol2 (E) N364K::kanMX6 msh2::his3 ade6? his3-D1? - G0 - D1* | This study |
| 3908 | *pol2 (E) N364K::kanMX6 msh2::his3 ade6? his3-D1? - G50 - D1* | This study |
| 3909 | *pol2 (E) N364K::kanMX6 msh2::his3 ade6? his3-D1? - G0 - D2* | This study |
| 3910 | *pol2 (E) N364K::kanMX6 msh2::his3 ade6? his3-D1? - G50 - D2* | This study |
| 3911 | *pol2 (E) N364K::kanMX6 msh2::his3 ade6? his3-D1? - G0 - D3* | This study |
| 3912 | *pol2 (E) Y459F::kanMX6 msh2::his3 ade6? his3-D1? - G0 - D1* | This study |
| 3913 | *pol2 (E) Y459F::kanMX6 msh2::his3 ade6? his3-D1? - G50 - D1* | This study |
| 3914 | *pol2 (E) Y459F::kanMX6 msh2::his3 ade6? his3-D1? - G100 - D1* | This study |
| 3915 | *pol2 (E) Y459F::kanMX6 msh2::his3 ade6? his3-D1? - G150 - D1* | This study |
| 3916 | *pol2 (E) Y459F::kanMX6 msh2::his3 ade6? his3-D1? - G0 - D2* | This study |
| 3917 | *pol2 (E) Y459F::kanMX6 msh2::his3 ade6? his3-D1? - G50 - D2* | This study |
| 3918 | *pol2 (E) Y459F::kanMX6 msh2::his3 ade6? his3-D1? - G100 - D2* | This study |
| 3919 | *pol2 (E) Y459F::kanMX6 msh2::his3 ade6? his3-D1? - G0 - D3* | This study |
| 3920 | *pol2 (E) Y459F::kanMX6 msh2::his3 ade6? his3-D1? - G50 - D3* | This study |
| 3921 | *pol2 (E) Y459F::kanMX6 msh2::his3 ade6? his3-D1? - G100 - D3* | This study |
| 3922 | *pol2 (E) Y459F::kanMX6 msh2::his3 ade6? his3-D1? - G150 - D3* | This study |
| 3923 | ***Cdc20$^{+}$**; h- pol2 (E)-3HA ade6-485 ex 2840* | This study |
| 3924 | ***Cdc20$^{WT\text{-}construct}$**; h- pol2 (E) WT-3HA::kanMX6 ade6-485 ex 3186* | This study |
| 3925 | ***Cdc20$^{exonull}$**; h- pol2 (E) D276A/E278A-3HA::kanMX6 ade6-485 ex3199* | This study |
| 3926 | ***Cdc20$^{P287R}$**; h- pol2 (E) P287R-3HA::kanMX6 ade6-485 (human P286R) ex 3218* | This study |
| 3927 | ***Cdc20$^{M295I}$**; h- pol2 (E) M295I-3HA::kanMX6 ade6-485 (human M294I) ex 3669* | This study |
| 3928 | ***Cdc20$^{N364K}$**; h- pol2 (E) N364K-3HA::kanMX6 ade6-485 (human N363K) ex 3672* | This study |
| 3929 | ***Cdc20$^{V412L}$**; h- pol2 (E) V412L-3HA::kanMX6 ade6-485 (human V411L) ex 3201* | This study |
| 3930 | ***Cdc20$^{L425V}$**; h- pol2 (E) L425V-3HA::kanMX6 ade6-485 (human L424V) ex 3188* | This study |
| 3931 | ***Cdc20$^{Y459F}$**; h- pol2 (E) Y459F-3HA::kanMX6 ade6-485 (human Y458F) ex 3680* | This study |

### 2.1.2 Setting up crosses

The equivalent of a 1 µL loop of the $h^-$ strain cells were spread on a small patch on an EMM-N plate supplemented with the right amino acids. An equivalent amount of $h^+$ strain was mixed with the $h^-$ patch. The plate was incubated for 3-4 days at 26°C and monitored microscopically for asci formation.

### 2.1.3 Tetrad dissection

The equivalent of a 1 µL loop of mated cells were taken and re-suspended in 100 µL of rich media. 20 µL of the cells were spread on a side of a rich media plate prior to dissection using a Yeast Tetrad Dissection Workstation (Singer – Instrument MSM400). After dissection of the tetrads, the plates were incubated for 4 hours at 30°C to break down the cell wall of the asci and release the spores. After spore dissection, the plates were incubated at the relevant temperature until colonies appeared.

### 2.1.4 Random spore analysis

The equivalent of 1 µL loop of mated cells were re-suspended in 100 µL of 0.05 % of *Suc Helix pomatia* juice (SEPRACOR) and incubated overnight at 36°C to free the spores from the asci. Cells were then centrifuged 5 minutes at 1600 g and re-suspended in 100 µL of 30 % ethanol for 30 minutes to kill any remaining vegetative cells. Then, the cells were centrifuged 5 minutes at 1600 g before re-suspension in 200 µL of rich media. Spores were counted using a haemocytometer prior to plating on a rich media to obtain ca. 200 spores per plate, and incubated at permissive temperature until colonies appeared.

### 2.1.5 Growth rate analysis

Growth curves were obtained using the Bioscreen C MBR machine. Cultures in exponential phase were adjusted to 2 x $10^6$ cells/mL and 250 µL were dispensed in 4-8 replicate wells for each strain. The plate was then locked into the machine and incubated at 30°C with constant shaking at maximum amplitude. The $OD_{600}$ of each well was measured every 20 minutes for 22 hours. Duplication time was determined via analysis of the exponential phase of the plotted log2 graph time vs $OD_{600}$ using the Exponential Growth Equation available on Prism.

### 2.1.6 Spot assay

$10^7$/mL cells from an exponential growing culture were serial diluted 5 times at ¼. Cells were then spotted on plates using a replicator (6 columns x 8 rows) and incubated at the relevant temperature for 2, 3 or 5 days.

### 2.1.7 Light microscopy and cell length

8 x $10^6$ cells from an exponentially growing culture were pelleted and re-suspended in 50 µL of media before 10 µL being spotted on a slide (Menzel-Gläser Superfrost – ThermoScientific) and a coverslip was put on top of the suspension. Visualisation was done using an AxioPlan 2 microscope and pictures where taken using a Hamamatsu ORCA E camera system with Micro-Manager 1.3 software. Cell length was measured using ImageJ.

### 2.1.8 Fission yeast transformation

The Lithium Acetate method (Bahler *et al.*, 1998) was used for fission yeast transformation. Briefly, $10^7$ cells from an exponential growing culture were washed in water twice (centrifugation 5 minutes at 1600 g) before re-suspension in 1 mL of LiAc/TE (0.1 M LiAc, 10 mM Tris-HCl, 1 mM EDTA). Cells were then re-suspended at 2 x $10^9$ cells/mL in LiAc/TE before addition of 2 µL of sheared carrier DNA (Merk – D7656) at 10 mg/mL and 200 ng of transforming DNA. Cells were incubated 10 minutes at room temperature before adding 260 µL of 40 % PEG/LiAc/TE (40 % PEG4000, 0.1 M LiAc, 10 mM Tris-HCl, 1 mM EDTA) and incubated 1 hour at 30°C. Samples were then heat shocked at 42°C with 43 µL of DMSO for 5 minutes before plating on non-selective media. Following 24 hours incubation, the plates were replica-plated on selective media and incubated until colony formation (3-4 days).

### 2.1.9 DNA isolation from fission yeast

#### 2.1.9.1 QIAGEN® DNeasy Plant Mini Kit

20 mL of a saturated culture were pelleted (centrifugation 5 minutes at 1600 g) and washed in 1 mL of 10 mM of EDTA pH 8.9. 250 µL of EDTA/ß-mercapto-ethanol/zymolase (EDTA 10 mM pH 8.9, 1% ß-mercapto-ethanol, 2 mg/mL zymolase) were added to the cell pellet and incubated at 37°C for 1 hour with shaking 400 rpm (ThermomixerC – Eppendorf). Equal amount of glass beads acid-washed (Sigma G8772) and cell pellet were added prior to 4 cycles of 15 seconds at 4 m/s on the FastPrep-24 from mpbio™ to help to break cell wall. Following mechanistic cell break, a 10 minutes heat shock at 65°C

was performed. gDNA was then extracted and purified according to manufacturer guidelines for the DNeasy Plant Mini Kit from QIAGEN®.

#### 2.1.9.2 Isopropanol gDNA extraction

5 mL of a saturated culture were pelleted (centrifugation 5 minutes at 1600 g) before washing the cells in 1 mL of SCE solution (1 M sorbitol, 0.1 M $Na_3$Citrate pH 5.8, 0.01 M EDTA). The cells were then re-suspended in 150 µL of zymolase solution (SCE solution, 0.8 % ß-mercapto-ethanol, 0.6 mg/mL zymolase in 10 % glycerol and 0.1 M NaCl) and incubated for 1 hour at 37°C with shaking. 150 µL of lysis solution (10 % SDS, 0.1 M Tris pH 9.0, 0.01 M EDTA) were added prior to a 5 minutes heat shock at 65°C. SDS and proteins were precipitated after addition of 150 µL of 5 M KAc and incubation on ice for 1 hour. After a 10 minutes centrifugation at 16000 g at 4°C, 300 µL supernatant were taken and mixed with 200 µL of 5M $NH_4$Ac and 1 mL of isopropanol to allow DNA precipitation at -20°C for 10 minutes. DNA was then pelleted by 3 minutes centrifugation at 16000 g at 4°C. After the supernatant was poured off, DNA pellet was rehydrated in 90 µL of $H_20$ then precipitated a second time with 10 µL 5 M $NH_4$Ac and 200 µL of isopropanol. DNA was centrifuged for 2 minutes at 16000 g at 4°C and washed in 1 mL of 80 % ethanol. DNA pellet was allowed to dry before dissolving in $H_20$. A final centrifugation (1 minute at 16000 g) was used to clarify the DNA solution.

### 2.1.10 FACS analysis

$10^7$ cells from a growing exponential culture were re-suspended in 1 mL of ice-cold 70 % ethanol. For FACS processing, 2-3 x $10^6$ cells were washed in 4.5 mL of 10 mM EDTA pH

8.0 prior to incubation for 2 hours at 37°C with 0.5 mL EDTA 10 mM containing 0.1 mg/mL RNase A. 1 µM of Sytox Green were added to the samples for DNA labelling. To avoid doublets, cells were sonicated for 30 seconds using an Elmasonic S10.

### 2.1.11 Fluctuation analysis

Fluctuation analysis was first described by S. E. Luria and M. Delbrück in 1943 (Luria *et al.*, 1943). Their research focused on *E. coli B* resistance acquired when facing attack by bacteriophage $\alpha$. S.E. Luria and M. Delbrück noticed that if incubation was prolonged, liquid cultures inoculated with both bacteria and bacteriophage would first get clear due to viral bacterial lysis, but then turbidity would reappear, for the reason that resistance to bacteriophage would emerge. Two questions arose from this observation: 1) is the resistance acquired prior to viral contact or 2) does the viral contact trigger resistance? In other words, are the resistant mutations directed (happening under selection pressure, therefore after viral addition) or random (therefore before viral contact). To answer this question, the following process was suggested: cells were plated and grown under no stress conditions prior to be plated onto a selective plate containing the bacteriophage. After incubation, resistant colonies were counted. In the case of mutations acquired due to viral contact, the number of viable colonies would follow a Poisson distribution – with equal variance and mean. Alternatively, if the mutations were obtained before viral contact, the variance would be higher than the mean of the resistant colonies. This increase in variance would be due to the fact that early clones which acquired resistant mutations early enough would have sufficient time to multiply and pass down their selective advantage on to the next generation, while late mutations would confer resistance to only a small number of cells which would not have sufficient

time to multiply before plating. Therefore, the distribution would not be Poisson, but a so called Luria-Delbrück one (Foster, 2006). Observing a higher variance than the mean, therefore not a Poisson distribution, the conclusion drew from this study was that the bacterial mutations were happening randomly, before spotted on selective plates. In addition to this, mutation rates – the probability of a mutation occurring per cell division – were estimated. Mutations are assumed to arise in a stochastic manner. However, in the case of a fluctuation analysis, if a selective mutation appears early in a culture, the resistant progeny will expand and be translated as a high mutation rate. This phenomenon is known as 'jackpot' mutation/culture. The earlier a selective mutation arises, the more resistant mutants will be observed.

#### 2.1.11.1 Fluctuation analysis – Mutation rate calculation

Algorithms used to estimate mutation rate have been improved and updated since the original Luria-Delbrück paper. Initially based on the Luria-Delbrück model, then extended by Lea and Coulson (Lea *et al.*, 1949), the most widely used method in the literature (Foster, 2006; Rosche *et al.*, 2000; Zheng, 2002) seems to be the Ma-Sandri-Sarkar Maximum Likelihood method (Sarkar *et al.*, 1992). This recursive algorithm computes the Luria-Delbrück distribution with the Lea-Coulson generated function. Explained briefly, the algorithm uses the estimated m (number of mutations per culture) and generates a provisional distribution, then compares it to the experimental values using the likelihood function, and finally repeats this step until the best m is found (Rosche *et al.*, 2000). The mutation rate is then m divided by the average total number of cells plated (N).

### 2.1.11.2 Canavanine resistance

There are several methods available to assess the mutation rate by fluctuation analysis in *S. pombe*. The canavanine assay is one of them. Initiated by the *S. cerevisiae* community, the fluctuation assay uses canavanine resistance as a selective process. In the budding yeast, the *CAN1* gene codes for a permease (Ahmad *et al.*, 1986; Grenson *et al.*, 1966). When the permease is functional, it allows arginine uptake from the environment. As canavanine is an analogue of arginine, the permease would also allow canavanine uptake if present in the environment. Incorporation of canavanine in place of arginine during protein synthesis might produce structurally aberrant and malfunctioning proteins, leading eventually to cell death.

The mechanism of canavanine uptake in fission yeast is poorly understood. One publication from Fantes *et al.* (1984) investigates the mechanism of canavanine resistance. Summarised briefly, this study analysed two strains (*can1.1* and *can2.1*) resistant to canavanine contact. After backcross with a wild type strain, it appeared that more than one gene is involved in the canavanine resistance mechanism. Both strains seemed to carry a main canavanine resistance mutation possibly located in the permease *can1*$^{+}$ gene, the *CAN1* gene equivalent in the budding yeast; an additional enhancer mutation was also identified.

Study of the *can1.1* strain revealed that mutation in the *canx* gene plays a role in canavanine resistance in a nitrogen source-dependent manner. Indeed, the level of resistance was increased with glutamate as nitrogen source instead of ammonium. The nature of the canavanine resistance for the *can2.1* strain is more uncertain. The strain has a growth defect compared to *can1.1* and a wild type strain as the colonies were smaller under normal growing conditions and in contact with canavanine. In addition,

even if similar to the *can1.1* strain, its arginine intake is lower, it is irrespective of the nitrogen source. From this work, two arginine uptake systems emerged. The first relies on a single major gene (*can1*, similar to *CAN1*) which is specific for arginine and the main canavanine entrance route. The second system also allows arginine/canavanine uptake, but is repressed or inhibited by the presence of ammonium.

#### 2.1.11.3 Canavanine assay

Cells from an exponentially growing culture were set up at an initial concentration of $10^4$ cells/mL and 1 mL were dispensed in a 24 wells plate. In order to allow statistical analysis, for each strain cultures were replicated in 12 wells. Each culture grew up to saturation for 2 days. 150 µL of each culture were spotted on a PMG plate containing 80 µg/mL of canavanine and incubated for 12 days at 30°C to allow resistant colonies to emerge. In parallel to spotting on a canavanine plate, cell density was assessed by plating 50 µL of a $10^4$ dilution of two random wells/strains. Numbers of resistant colonies on selective plates were counted by naked-eye, with no 'cut-off' colony size. In order to avoid any misleading counting, plates were randomised and all colony sizes were counted. Number of mutations per cultures were calculated computing the Maximum Likelihood Estimator of mutations under the Luria-Delbruck Model using the rSalavador package (Zheng, 2017) (Newton.LD function). The mutation rates were calculated by dividing the number of mutations by the average number of viable cells plated/spot.

### 2.1.12 Mutation Accumulation Experiment

Studied strains were woken up as usual on a rich media plate and incubated for 2 days at 30°C. For each streaked ancestor, 3 random single colonies were picked up and used to start lineages. Each line was passed through a randomly selected isolated colony, streaked on a new rich medium plate and allow to grow for 3 days at 30°C. In total, each lineage went through up to 11 passages (ca. 218 generations). Doubling time used to estimate the number of generations was based on the Farlow *et al.* (2015) publication estimated at 6.6 generations/days. For each passage, part of the colony was used to continue the lineage while the rest was inoculated in 20 mL of rich liquid media and incubated overnight at 30°C with shaking. The following day, the cells were frozen down for storage at -80°C.

## 2.2 Bacterial Methods

### 2.2.1 Plasmid preparation from E. coli

Plasmids were extracted from *E. coli* according to the manufacturer guidelines using QiaPrep Spin MiniPrep Kit from QIAGEN®.

### 2.2.2 *Escherichia coli* transformation

New England Biolabs® 5-alpha competent *E. coli* cells (#C2988) were used for plasmid transformation with the Calcium Chloride method (J. Sambrook, 1989). Briefly, 50 µL of competent cells were thawed on ice before adding 1-5 µL of 100 pg - 1 µg of plasmid. The mixture was placed on ice for 30 minutes before a heat shock at 42°C for 30 seconds. Cells were placed back on ice before adding 950 µL of LB, then were incubated for 1 hour

at 37°C with vigorous shaking at 500 rpm (ThermomixerC). Finally, the cells were spread on agar LB plate with required selection and incubated at 37°C overnight. For ampicillin selection, final concentration was 100 µg/mL while kanamycin was 50 µg/mL. For long term storage, cells were stored at -80°C in 25% glycerol final concentration.

## 2.3 Molecular Genetics Methods

**Table 2.2 Primers**

| Number | Primer | Sequence (5'-3') | Source |
|---|---|---|---|
| 989 | cdc22-D57N-check_F | TTTGGATCCATGTTTGTATACAAAAGAGGTATGAAACAGG | Collection |
| 990 | cdc22-D57N-check_R | GGACGAGGAGTACCGGCATTG | Collection |
| 1083 | PN_F | tttagatctCTGATTCGTTTAAGCTTTTTCAGTTAATGGT | Collection |
| 1084 | PC_R | TggcgcgccGGTACATTGAGGCGACATCAAGATGG | Collection |
| 1088 | pFA6a_R | gGTGACCCGGCGGGGAC | Collection |
| 1089 | pFA6a_F | AAGCTTCGTACGCTGCAGGTCG | Collection |
| 1102 | cdc20-seq1_F | ACTGGCCATTTGTAGACGCACGTG | Collection |
| 1117 | cdc20-seq2_R | GACTTATATTCATCCTCTGCGTCCC | Collection |
| 1118 | cdc20-seq3_F | GCCATTGAAGTTTCCTGATAGCTCGTTCG | Collection |
| 1119 | cdc20-seq4_R | CTTGTTGGGAAGAATTAGCTTTCACAGC | Collection |
| 1120 | cdc20-seq5_R | GGGTAGAATTGCCTTTGGTGCTTTC | Collection |
| 1173 | msh2△_F | GAGAAGACGCTTCGTTTGTTGC | Collection |
| 1212 | PC-cdc20-M295I_F | AAATAATcATGATATCTTATATGATTGATGGCC | This study |
| 1213 | PN-cdc20-M295I_R | TCATgATTATTTTATCGAACGAGCTATCAGG | This study |
| 1243 | cdc20-screen2 | GGGCACATTAAGATCGCATGTTGCACAAAAC | This study |
| 1244 | pFA6a_F | GAAGCTTCGTACGCTGCAGGTCG | This study |
| 1321 | PC-cdc20-N364K_F | gTACAAgGGTGACTTCTTTGACTGGCCATTTG | This study |
| 1322 | PN-cdc20-N364K_R | CACCcTTGTAcGTAACTATTACTGAAGGCTTT | This study |
| 1369 | msh2△_F | GAAGTGCTATTGCTACCAGTTGC | This study |
| 1374 | cdc20-3HA_F | AACGTTTATCAATCTGTTGCTGATTTTTACGAGTTTTCTATATTGCAAAATTCTGTTCAGTCCATACTTTCTGTGCTGAACCGGATCCCCGGGTTAATTAA | This study |

| | | | |
|---|---|---|---|
| 1375 | cdc20-3HA_R | AGCAGATAAAATCAAAATAACTTATAAAATACATATAAAGACAAAAAAGCCAAACTCATGCTCATGGTCAGGCGAGAAATGGAATTCGAGCTCGTTTAAAC | This study |
| 1376 | cdc20-3HA-check_R | CGCATAGTCAGGAACATCGTATGGGTAAAAGATG | This study |
| 1377 | cdc20-3HA-check_F | ACAAATTCAAGAGTTTTCCACTCAGGCGG | This study |
| 1380 | kpa1△-check_F | GGCAGAGAAGCAGAGGACAC | This study (Chen *et al.*, 2014) |
| 1381 | kpa1△-check_R | GCTAGAACGTTTGGCTTTGC | This study (Chen *et al.*, 2014) |

**Table 2.3 Plasmids**

| Number | Plasmid | Source |
|---|---|---|
| 950 | pFA6a-KanMX6-cdc20-WT | Collection |
| 958 | pFA6a-KanMX6 | Collection (Wach *et al.*, 1997) |
| 990 | pFA6a-KanMX6-cdc20-N364K from #958 | This study |
| 991 | pFA6a-KanMX6-cdc20-M295I from #958 | This study |
| 992 | pFA6a-KanMX6-cdc20-Y459F from #950 | This study |
| 993 | pMK-RQ-Kana-cdc20-Y459F | This study |
| 998 | pFA6a-natMX6-3HA | Collection (Hentges *et al.*, 2005) |

### 2.3.1 PCR product purification

PCR products were purified according to manufacturer guidelines using QIAquick PCR Purification Kit from QIAGEN®.

### 2.3.2 DNA purification through gel extraction

DNA product purifications were executed according to manufacturer guidelines using QIAquick Gel Extraction Kit from QIAGEN®.

### 2.3.3 Enzymatic digestion reaction

Restriction enzymes from New England Biolabs® were incubated at 37°C for at least 2 hours with DNA and the appropriate restriction enzyme buffer at the concentration suggested by the manufacturer. Inactivation at 75°C for 15 minutes was performed after plasmid linearization prior to yeast transformation.

### 2.3.4 DNA ligation

Ligation reactions were performed according to the manufacturer guidelines using T4 DNA ligase from New England Biolabs®. Briefly, 20 µL of reaction containing 1:3 ratio of insert:plasmid, respectively, 1X T4 DNA ligase buffer and 1 µL of T4 DNA ligase, were incubated at room temperature for 1 hour prior to 10 minutes enzymatic inactivation at 65°C.

### 2.3.5 DNA analysis by gel electrophoresis

DNA products were mixed with 6X loading dye (10 mM Tris-HCl pH 7.6, 0.03 % bromophenol blue, 0.03 % xylene cyanol FF, 60 % glycerol, 60 mM EDTA) and migrated on an agarose gel made with 1X TAE (0.04 M Tris-Acetate and 0.01 M EDTA pH 8.0) using the Biorad MiniSub Cell GT system. MIDORI$^{Green}$ Advance was used up to 1:45000 dilution in gel to detect DNA under UV light (302 nm).

### 2.3.6 Colony PCR

Taq DNA Polymerase 1.1X Master Mix 1.5 mM $MgCl_2$ from VWR® (Tris-HCl pH 8.5, 2.2 mM $MgCl_2$, 0.11 % Tween 20, 0.22 mM of each dNTP, 0.11 units/µL Taq DNA polymerase

and stabilizer) was used for colony PCR in a 20 µL final volume combined with 0.5 µM final concentration of each primer. A small amount of colony picked with a tooth stick was used as a PCR template. The PCR programme used is as follow: denaturation at 94°C for 5 minutes followed by 30 cycles of 1) denaturation at 94°C for 30 seconds, 2) primers annealing at the required temperatures (usually between 55 and 65°C) for 1 minutes and 3) elongation at 68°C for 6 minutes. The program ends by a final elongation step at 68°C for 7 minutes and holding at 10°C.

### 2.3.7 3HA Tagging Process

The 3HA tag was amplified from the plasmid #998 with the primers #1374 and #1375, both primers containing homology sequence with C-ter of *cdc20* necessary for genomic integration. About 1 µg of plasmid #998 combined with 0.5 µM final concentration of each primers were included into a PCR mix containing 0.2 mM of each dNTP (BIO-39028 from meridianBIOSCIENCEC®), 1X Phusion™ HF buffer (MAN0012393 from Phusion™ High-Fidelity DNA Polymerase), 0.5 units of Phusion™ DNA Polymerase (MAN0012393 from Phusion™ High-Fidelity DNA Polymerase), in a final volume of 50 µL. The PCR programme used is as follow: denaturation at 98°C for 30 seconds followed by 35 cycles of 1) denaturation at 98°C for 10 seconds, 2) primers annealing at 50° for 30 seconds C) for 1 minutes and 3) elongation at 72°C for 2 minutes. The program ends by a final elongation step at 72°C for 7 minutes and holding at 4°C. The PCR product was then transformed into yeast as previously described (Materials and Methods 2.1.8), yet using 3 µg of transforming DNA instead of 200 ng. Following the 24 hours incubation on non-selective media, the plates were replica-plated on rich media containing clonat at 100 µg/mL and incubated until colony formation (3-4 days). Growing colonies were then

screened for tag insertion via colony PCR (Materials and Methods 2.3.6) with the primers #1376 and #1377 and annealing temperature at 55°C.

### 2.3.8 Sanger Sequencing

Sanger sequencing services from GATC Eurofins Genomics were used, according to manufacturer guidelines. Briefly, LIGHTRUN were performed on purified plasmid or PCR fragments concentrated at either 80-100 ng/µL or 20-80 ng/µL, respectively, and primer to initiate sequencing reaction added at 5 µmol/L.

### 2.3.9 Whole Genome Sequencing

Whole Genome Sequencing was performed by the Oxford Genomic Centre platform. Between 0.7 and 2.5 µg of gDNA were provided to the platform (0.7 - 2.5 ng/µL in 100 µL final volume). The platform processed the samples from quality assessment to sequencing. This process involves the following: gDNA was quantified using Qubit (Invitrogen) or PicoGreen (Thermo Fisher) on the FLUOstar OPTIMA plate reader (BMG Labtech) and the size profile analysed using eGel (Thermo Fisher, 1 % EX Agoarose). Input material was normalised to 300 ng prior to fragmentation and library preparation. Fragmentation was performed by mechanical shearing to an average size of 300 bp (Covaris; duty Cycle – 10 %, intensity – 5.00, cycles/Bursts – 200, time – 60 seconds). Libraries were prepared using the PrepX DNA library kit (#640102) from Takara according to manufacturer's instructions. Individual libraries were normalised using Qubit, and the size profile was analysed on the 2200 or 4200 TapeStation. Individual libraries were normalised and pooled together accordingly. The pooled library was

diluted to ~10 nM for storage. The 10 nM library was denatured and further diluted prior to loading on the sequencer. Paired end sequencing was performed using a Novaseq6000 150bp platform (Illumina, NovaSeq 6000 S2 Reagent Kit (300 cycles) (#20012860) Kit and 300 cycle SBS Kit).

## 2.4 Biochemistry Methods

**Table 2.4 Antibodies**

| Name | Biological source | Supplier | Reference | Used dilution | Comments |
|---|---|---|---|---|---|
| Anti-HA | Mouse | Biolegend | MMS-101R | 1/2000 | Clone 16B12 |
| Anti- tubulin $\alpha$ | Mouse | Sigma | T5168 | 1/2000 | Clone B-5-1-2 |
| Anti-mouse (HRP) | Horse | Vector Laboratory | Pi-2000 | 1/2000 | |
| Anti-FLAG M2 | Mouse | Sigma | F3165-2MG | 1/1000 | Used for in vitro assays |
| Anti-mouse (HRP) | Goat | AbCam | Ab97040 | 1/10000 | Used for *in vitro* assays |

### 2.4.1 TriChloroAcetic protein extraction

20 mL of exponentially growing cultures at $OD_{595}$ = 0.8-1 were centrifuged for 5 minutes at 1600 g at room temperature. Cells were washed in 1 mL of $H_2O$ prior to be re-suspended in 100 µL of 20 % TCA. Glass beads acid-washed (Sigma G8772) were added at equal cell pellet volume before the cells were sheared for 5 minutes at 4°C using a vortex Genie2. Samples were left on ice for 1 minute prior to a second shearing. Then, 900 µL of 5 % TCA were added before transfer of 800 µL of the sample into a new tube, and centrifuged for 10 minutes at 16000 g. The supernatant was discarded and the cell pellet was re-suspended in 100 µL of Laemmli 2X (120 mM Tris-HCl pH 6.8, 20 % glycerol, 4 % SDS, 10 % ß-mercapto-ethanol, 0.02 % bromo-phenol-blue) before boiling for 3

minutes at 90°C. After centrifugation for 2 minutes at 16249 g, the supernatant was transferred into a new tube and stored at -20°C.

### 2.4.2 SDS PAGE

Prior to loading, protein samples were boiled for 3 minutes at 90°C and put on ice for 2 minutes. Equal volumes of protein extracts from TCA protein extraction protocol were resolved on 3-8% Tris Acetate NuPAGE (ThermoScientifc) in reducing conditions (NuPAGE Tris Acetate SDS running buffer – ThermoScientific)

### 2.4.3 Western Blot

For Western Blot analysis, the proteins were transferred onto a PVDF membrane using the iBlot2 system and the large protein transfer program (20 V for 1 minute 30 seconds, 23 V for 5 minutes and 30 seconds, 25 V for 3 minutes). Following the transfer, the membranes were blocked with 5 % BSA in 1X TBST (20 mM Tris pH 7.5, 150 mM NaCl, 0.1 % Tween) for 1 hour at room temperature on a roller mixer. Probing with relevant antibody was performed in 2.5 % BSA 1X TBST according to Table 2.4 for either 1 hour at room temperature or overnight at 4°C on a roller mixer. Incubation with respective HRP-conjugated antibody was followed by several washes with 1X TBST. Immunodetection was performed by chemoluminescence HRP-substrate (ECL Prime Western Blot detection reagent – GE Healthcare) and imaged using the iBright CL1000 from Invitrogen.

### 2.4.4 dNTPs extraction and measurement

Samples were extracted for the measurement of dNTPs from exponentially growing cells. $10^9$ cells were washed with 10 mL of 3 % glucose with gentle inversion. Working on ice, 50 µL of ice-cold 10% TCA was added to the cells while vortexing. The solution was then left on ice for 5 minutes and vortexed for 1 minutes before being frozen at -80°C.

The hydrolysed samples were shipped on dry ice to the Bioanalysis Core, Department of Oncology for HPLC analysis where D$^r$ Lisa Folkes processed the samples. This process includes the following steps: samples were defrosted and spun to remove any precipitate. The acid extracts were diluted 5-fold in water and the acid removed by vortex mixing for 15 seconds with an equal volume of a 4:1 mixture of 1,1,2-trichlorotrifluoroethane:trioctylamine. The samples were spun in a microfuge to separate the layers and the top aqueous layer removed into an HPLC vial for analysis. HPLC used a Waters (Watford UK) separation module (e2695) with a waters photodiode array detector (2996) equipped with an ACE column (100 x 3.2 mm, 3 µm) at 35°C. Mobile phase consisted of A: 10 mM potassium dihydrogen orthophosphate ($KH_2PO_4$) , 10 mM tetrabutylammonium hydroxide (TBA), 10 % methanol, pH 6.9 and B: 30 mM $KH_2PO_4$, 6 mM TBA, 30 % methanol, pH 7, with a gradient of 25-80 % B in 20 minutes and a 25 minutes run time (similar to that described in (Decosterd *et al.*, 1999).

Area under peak were assessed for each dNTP. A calibration curve was performed in parallel for each dNTP, allowing gradient (slope) calculation for each of those. In order to get the final dNTP concentration for each sample, the area under peak was divided by the gradient of the corresponding dNTP, then reported to total number of cells/sample.

## 2.5 *In vitro* methods

### 2.5.1 Extension assay

Substrate for DNA processivity assay was obtained after annealing Cy5-labeled oligonucleotide Cy5 - 5'- TTTTCCCAGTCACGACGTTGG -3' to 42-mer template 5' – CGACAGCGACCGGCAACAGAAGGTTGCAGCACTGACCCTTTT -3'. The annealing was performed by incubating the labelled primer and template at a 1:1 ratio at 95°C for 20 minutes in the presence of 20 mM Tris pH 8, 10mM KCl and 0.01mM EDTA. The sample was then slowly cooled down until 40°C were reached (ca. 2 hours). The annealed sample was then stored at 4°C.

For extension assay, in a 20 µL-reaction, 12 fmol of substrate were added to 1X buffer (20 mM Tris pH 7.5, 150 µg/mL BSA, 1 mM DTT, 10 mM MgAc), dNTP at 0.125 mM each (dATP, dTTP, dCTP and dGTP) and Pol ε at the indicated concentration. The samples were incubated at 37°C for indicated times. Enzymatic reactions were quenched by adding EDTA at 0.1 M final concentration.

### 2.5.2 Exonuclease assay

Substrate for DNA exonuclease assay was obtained after annealing Cy5-labeled oligonucleotide Cy5 - 5'- TTTTCCCAGTCACGACGTTGG -3' to 34-mer template 5' - ACCGGCAGCAAAATGTTGCAGCACTGACCCTTTT -3', mismatching on the 3' - dGTP of the Cy5-labeled oligonucleotide. The annealing was performed by incubating the labelled primer and template at a 1:1 ratio at 95°C for 20 minutes in the presence of 20 mM Tris pH 8, 10mM KCl and 0.01mM EDTA and then slowly cooled down until 40°C were reached (ca. 2 hours).

For exonuclease assay, in a 20 µL-reaction, 12 fmol of substrate were added to 1X buffer (20 mM Tris pH 7.5, 150 µg/mL BSA, 1 mM DTT, 10 mM MgAc) and Pol ε at the indicated concentration. The samples were incubated at 37°C for indicated times. Enzymatic reactions were quenched by adding EDTA at 0.1 M final concentration. The exonuclease activity was also performed using the single-stranded Cy5-labeled oligonucleotide.

### 2.5.3 DNA Electrophoresis in denaturation conditions

Prior to electrophoresis, the DNA samples were mixed with equal volume of Ambion Gel Loading Buffer II (95 % Formamide, 18 mM EDTA, 0.025 % each of SDS, Xylene Cyanol, and Bromophenol Blue) then boiled for 15 minutes at 95°C and loaded directly onto gel. For the extension assay, the samples were loaded on a pre-run (15 minutes at 300 V) 12.5 % denaturing polyacrylamide gel containing 8 M urea in 1X TBE buffer. The samples were run for 25 minutes at 300 V. For exonuclease assay, the samples were loaded on pre-run (30 minutes at 300 V) 20 % denaturing polyacrylamide gel containing 8 M urea in 1X TBE buffer. The samples were run for 2 hours a 200 V. Detection was performed using the Odyssey CLx Imaging System from LI-COR.

## 2.6 Whole Genome Sequencing Analysis

Sequencing adaptors were trimmed from the raw sequencing files prior to alignment against the *S. pombe* genome ASM294v2.30 (Wood *et al.*, 2002) from Pombase (Lock *et al.*, 2019) using BWA (H. Li *et al.*, 2010). These steps were executed by the Bioinformatics Core of the Wellcome Trust Centre of Human Genomics.

The aligned reads were then processed using GATK (McKenna *et al.*, 2010) for duplicate removal (gatk MarkDuplicates --INPUT inputfile.bam --OUTPUT inputfile_MarkDup.bam --METRICS_FILE filemetrics.txt --TAGGING_POLICY All) prior to variant calling using HaplotypeCaller (Ryan Poplin, 2017) (gatk HaplotypeCaller --input inputfile_MarkDup.bam --output inputfile.vcf --reference pombe.ASM294v2.30.dna.genome.fa --sample-ploidy 1). Variants were then filtered using hard filtering parameters suggested by GATK Best Practices recommendations (gatk VariantFiltration --reference pombe.ASM294v2.30.dna.genome.fa --variant inputfile.vcf --output filtered.vcf --filter "MQ < 40.0" --filter-name "MQ40" --filter "QUAL < 30.0" --filter-name "QUAL30 " --filter "DP < 20.0" --filter-name "DP20").

# Chapter 3

## Phenotypic Characterisation of Novel Cancer Related Germline *POLE*$^{EDM}$ variants in Fission Yeast

### 3.1 Introduction

Identification of driver mutations is one of the many challenges cancer research faces. Advanced sequencing techniques combined with collaborative efforts in bioinformatics analysis enabled such identifications in the exonuclease domain of the replicative DNA polymerase epsilon (Cancer Genome Atlas, 2012; Cancer Genome Atlas Research *et al.*, 2013). The high mutation burden observed in the *POLE*$^{EDM}$ tumours is suspected to be due to altered proofreading ability of Pol ε (Rayner *et al.*, 2016), although there are distinct mutation rates (Barbari *et al.*, 2018; Hamzaoui *et al.*, 2020) and exonuclease activity amongst *POLE*$^{EDMs}$ (Barbari *et al.*, 2018; Shinbrot *et al.*, 2014; Xing *et al.*, 2019), which cannot be explained by simple lack of exonuclease activity. Notably, somatic *POLE*$^{EDMs}$ are associated with a favourable prognosis in glioblastoma (Erson-Omay *et al.*, 2015), endometrial (Church *et al.*, 2015) and colorectal cancer (Domingo *et al.*, 2016), while specific germline *POLE*$^{EDMs}$ predispose to CRC (Palles *et al.*, 2013).

Numerous germline *POLE*$^{EDMs}$ were recently identified in patient cohorts from Colorectal Gene Identification (CORGI) study (Wang *et al.*, 2013)and highly suspected to predispose carriers to cancer (Hansen *et al.*, 2015; Palles *et al.*, 2013; Rohlin *et al.*, 2014;

Vande Perre *et al.*, 2018; Wimmer *et al.*, 2017). Although, the diversity of mutations extracted, added to population polymorphism renders the distinction of causal from passenger mutations more difficult. As early detection of driver mutations may help to manage the risk of cancer development and to prevent severe forms of disease, it is vital to confirm the pathogenic effects of germline *POLE*$^{EDM}$ variants. Thus, it is essential to use *in vivo* models to first, confirm their pathogenic effect and second, allow further investigation into their pathogenic mechanisms to ultimately, explore therapeutic approaches.

*Schizosaccharomyces pombe* — or commonly named fission yeast — is an excellent model for studying DNA replication and damage. In contrast to *Saccharomyces cerevisiae*, since divergence of the two species from a common ancestor, *S. pombe* seems to have retained common features with metazoan cells. Fission yeast and mammalian cells possess a large number of orthologous genes, with some proteins actually more similar to their mammalian homologs than that of budding yeast (Sipiczki, 2000). Easy to genetically manipulate and with similar features to mammalian cells including the including the catalytic subunit of Pol ε – named Cdc20/POLE – we sought to construct human germline *POLE*$^{EDM}$ variants in fission yeast and evaluate the cellular response to altered Cdc20.

## 3.2 Results

### 3.2.1 Construction of germline *cdc20*$^{EDM}$ variants in *S. pombe*

Two approaches were used to construct in fission yeast the selected germline *POLE*$^{EDM}$ variants to further study their potential pathogenicity. The first approach used was PCR-based site-directed mutagenesis. In the second approach, the synthesis of the relevant part of *cdc20* was ordered from Invitrogen (GeneArt), then sub-cloned into an appropriate vector prior to transformation into fission yeast. The germline *POLE*$^{EDM}$ variants selected for further characterisation are *POLE*$^{M294I}$, *POLE*$^{N363K}$, *POLE*$^{V411L}$, *POLE*$^{L424V}$ and *POLE*$^{Y458F}$. The *POLE*$^{V411L}$ (*cdc20*$^{V412L}$ equivalent in *S. pombe*) and *POLE*$^{L424V}$ (*cdc20*$^{L425V}$ equivalent in *S. pombe*) strains were already available in the laboratory collection. Therefore, the following variants were left to construct in fission yeast: *POLE*$^{M294I}$ (*cdc20*$^{M295I}$ equivalent in *S. pombe*), *POLE*$^{N363K}$ (*cdc20*$^{N364K}$ equivalent in *S. pombe*) and *POLE*$^{Y458F}$ (*cdc20*$^{Y459F}$ equivalent in *S. pombe*). The somatic *POLE*$^{P286R}$ (*cdc20*$^{P287R}$ equivalent in *S. pombe*) variant was included in the analysis to allow comparison with a hypermutating variant, as well as a Cdc20 exonuclease-dead version, with D276A and E278A mutations (D290 and E292 human homolog).

#### 3.2.1.1 PCR-based site-directed mutagenesis

The PCR-based site-directed mutagenesis process starts with the incorporation of the *cdc20*$^{EDM}$ mutation into part of *cdc20*, using two mutated overlapping primers orientated forward and reverse, paired with two flanking oligos. In addition to the EDM, additional mutations were included into the mutagenic primers in order to incorporate a new restriction enzyme site (silent mutation), added to aid screening for incorporation

of the mutation (Table 3.1). This first round of PCR was performed on fission yeast wild-type gDNA, the mutagenic primers paired with either cdc20-N forward (#1083) or cdc20-C reverse (#1084). After purification via gel electrophoresis, the two PCR fragments were annealed prior to the second round of PCR using the primers cdc20-N forward (#1083) and cdc20-C reverse (#1084) to amplify the full-length fragment of *cdc20* which will be used for cloning (Figure 3.1 – b). The mutation incorporation was checked by digestion of the purified full-length fragment. The PCR products, in parallel with the plasmid pFA6a-KanMX6 (#958), were digested by the restriction enzymes BglII and AscI, since the primers cdc20-N forward (#1083) and cdc20-C reverse (#1084) contain the restriction sites BglII and AscI respectively. Those two restrictions sites allow incorporation of the mutated *cdc20* fragment into the pFA6a-KanMX6 (#958) (Figure 3.1 – c). The digested PCR products were cleaned up, and the plasmid backbone was purified by gel electrophoresis. Ligation was then performed, followed by *E. coli* transformation. After plasmid amplification and purification, the presence of the mutation was checked by enzymatic restriction. Positive plasmids were then sequenced further in order to confirm accuracy of the full cloned sequence (oligo used for Sanger sequencing #1102, #1117, #1118, #1119, #1120) prior to BamHI linearization and yeast transformation (Figure 3.2 – a). Positive clones were selected for geneticin (G418) resistance conferred by pFA6a-KanMX6 transformed plasmid. Colony PCR was performed to check whether the plasmid had integrated at the *cdc20* locus; the forward primer pFA6a (#1244) anneals against the plasmid sequence (upstream of *cdc20*), whereas the reverse primer cdc20_screen2 anneals to part of genomic *cdc20* downstream of the region in the transforming plasmid (#1243) (Figure 3.2 – b). For transformants that generated a positive PCR reaction, the fragments were digested with the relevant restriction enzyme

to check that the mutation had been retained (Table 3.1). As homologous recombination occurs during the transformation process, plasmid incorporation could lead to integration of the *cdc20*$^{EDM}$ either on the full-length genomic *cdc20* or the partial genomic *cdc20* as a result of the resolution of the Holliday junction (Figure 3.2). The positive clones were sequenced to confirm the presence of the EDM, and verify the absence of additional mutations.

**Table 3.1 Cdc20$^{EDMs}$ cloning**

| **Human POLE variant** | ***S. pombe* Cdc20 variant** | ***cdc20* mutation** | ***cdc20* codon variation** | **Mutagenic primers** | **Diagnostic restriction site mutation** | **Diagnostic restriction site** |
|---|---|---|---|---|---|---|
| M294I | M295I | c.885G>C | ATG>ATc | #1212/#1213 | c.885G>C | BspHI |
| N363K | N364K | c.1092T>G | AAT>AAg | #1321/#1322 | c.1086A>G | SnaBI |
| Y458F | Y459F | c.1376A>T | TAT>TtT | #993 (custom plasmid) | c.1371T>C; c.1374A>G | BmRI |

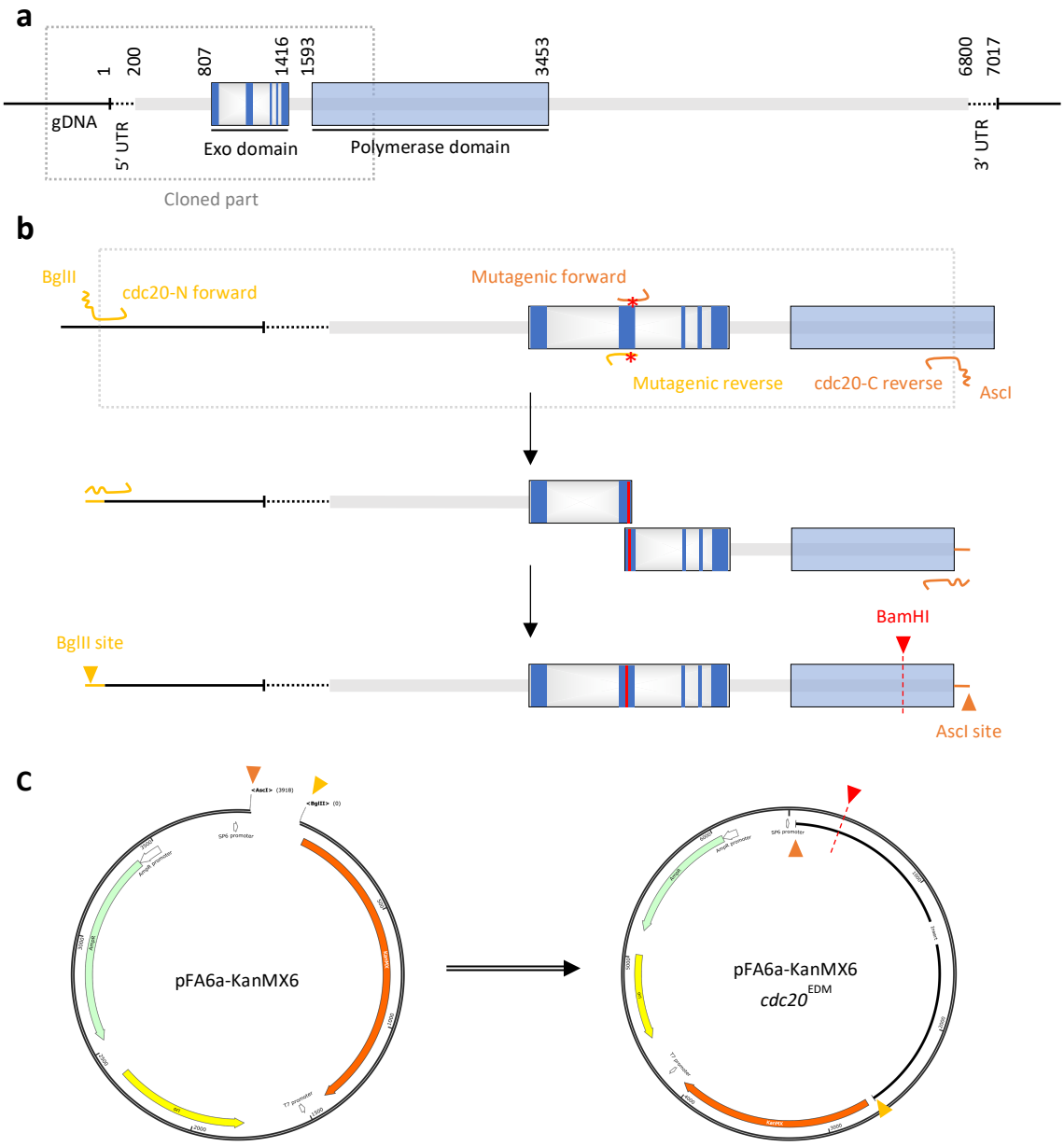



**Figure 3.1 PCR-based site-directed mutagenesis process - Mutation incorporation**

(**a**) Schematic representation of *cdc20*. The untranslated regions (5' and 3' UTR) of *cdc20* are represented by a black dotted line while non-coding sequences surrounding the gene are in plain black line. The Exo motifs are in dark blue within the Exo domain. The region cloned is designated by a grey box. (**b**) The *cdc20*$^{EDM}$ (represented by a red *, then a red |) is incorporated via two pairs of primers separately amplifying the N-ter (primer pair in light yellow; cdc20-N forward/Mutagenic reverse) and C-ter (primer pair in dark yellow; Mutagenic forward/cdc20-C reverse) of the fragment to clone. 10 bp homology between the two generated fragments allows a final amplification of the full fragment to clone (primer pair cdc20-N forward and cdc20-C reverse) (**c**) The restriction enzyme sites BglII and AscI were added via the primers cdc20-N forward and cdc20-C-reverse at the N-ter and C-ter of the cloned fragment, respectively. This allowed insertion of the cloned fragment into pFA6a-KanMX6 plasmid via restriction digestion. BamHI restriction site in indicated in red.

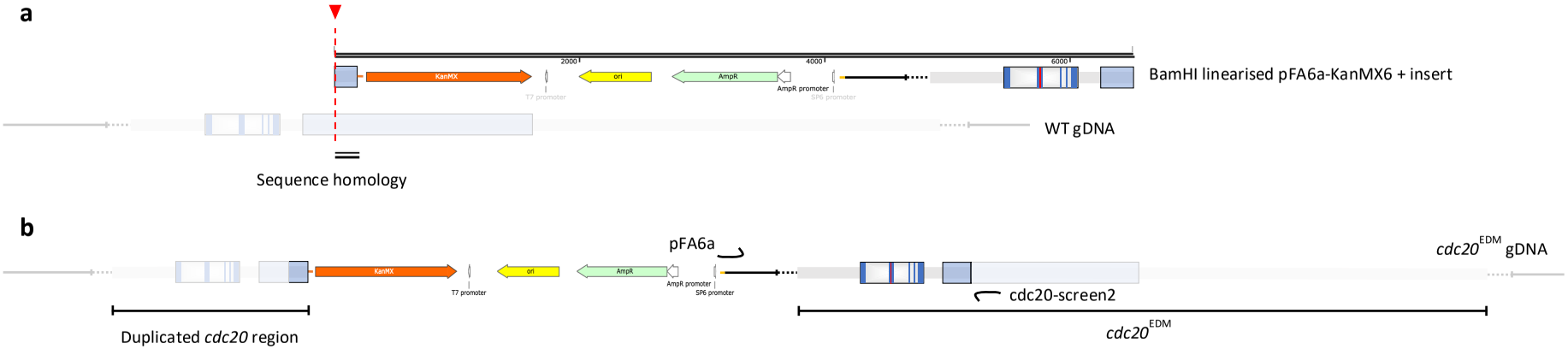



**Figure 3.2 PCR-based site-directed mutagenesis process – Cloning and transformation**

(**a**) The newly generated plasmid containing *cdc20*$^{EDM}$ is linearised via BamHI enzymatic digestion (red dotted line). During the transformation process, the linearised plasmid introduced into the cell is treated as a damaged DNA, inducing homologous recombination between the gDNA and the DNA fragment, leading to the insertion of the former one. Here is represented the resolution of the Holliday junction which results in the incorporation of the EDM in the full length genomic *cdc20*. (**b**) Final genomic construct of the *cdc20* variants. Note the duplicated region of *cdc20* upstream of the actual mutated *cdc20*. Colony PCR for positive transformant screening were performed with primers pFA6a_F (#1244) and cdc20-screen2 (#1243).

### 3.2.1.2 Sub-cloning

Sub-cloning approach using synthetic DNA was used to accelerate the cloning process of the germline *cdc20*$^{EDM}$ variants (Figure 3.3). To do so, a custom plasmid containing 1249 bp of the *cdc20* sequence including the mutated site was ordered from Invitrogen (GeneArt). The fragment to sub-clone was extracted via digestion with DraIII and SnaBI, while in parallel the pFA6a-KanMX6-cdc20-WT plasmid (#950) was digested with the same enzymes. Both the custom fragment and the backbone were purified by agarose gel migration prior to ligation via T4 ligase at a 3:1 ratio, respectively. Following transformation into *E. coli*, the plasmid was amplified, extracted and its structure checked by restriction enzyme digestion. Presence of the mutation was verified by enzymatic digestion (Table 3.1). As previously described, the plasmid was linearised with BamHI and transformed into the parental strain. G418 resistant transformants were screened for incorporation into the *cdc20* locus via colony PCR prior to sequencing to confirm the presence of the EDM, and confirm the absence of additional mutations.

It is important to note that both strain construction processes lead to a duplication of 2385 bp of the N-ter part of *cdc20* (Figure 3.2– b). This truncated portion of gene is however expected to be inactive as it is incomplete. In order to assess the impact of the strain construction process and the presence of the additional truncated version of *cdc20*, a 'WT-construct' (*cdc20*$^{WT\text{-}construct}$) strain was obtained in the same way the variants were, except that the cloned fragment is wild type.

The *cdc20*$^{M295I}$ and *cdc20*$^{N364K}$ strains were successfully obtained via PCR-based site-directed mutagenesis while the *cdc20*$^{Y459F}$ strains were obtained from synthetic DNA. No additional mutations were observed after sequencing of the cloned part of *cdc20* incorporated into the genomic version of *cdc20*.

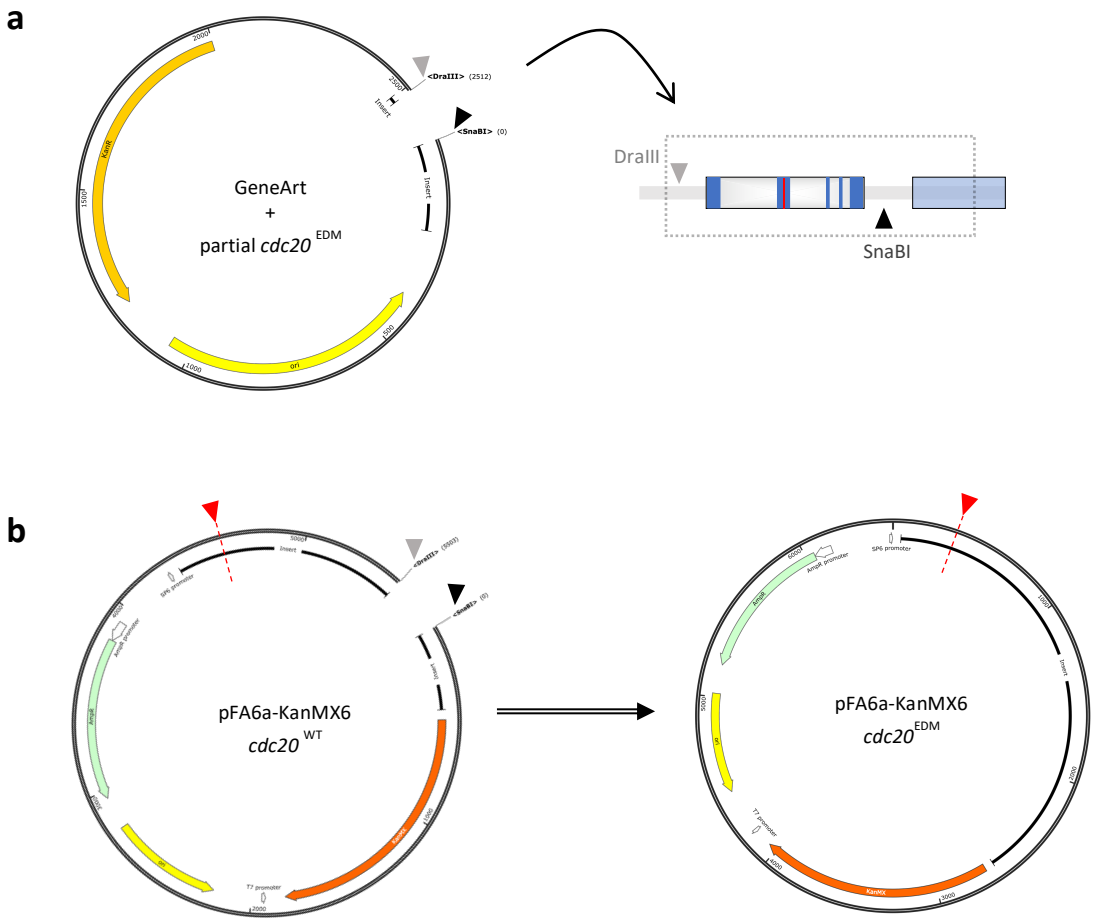



**Figure 3.3 Sub-cloning process**

(**a**) GeneArt plasmid is digested with DraIII and SnaBI restriction enzymes in order to extract the *cdc20*$^{EDM}$ fragment to sub-clone into pFA6a-KanMX6-*cdc20*$^{WT}$. (**b**) Once the *cdc20*$^{EDM}$ digested fragment purified, pFA6a-KanMX6-*cdc20*$^{WT}$ plasmid is digested with DraIII and SnaBI prior to ligation.

### 3.2.2 Newly generated *cdc20*$^{EDM}$ variants do not promote a diploid status

Previous work from our group showed that some mutator *S. pombe* strains — such as *cdc20*$^{P287R}$ — can diploidise, possibly as the mutator phenotype of this strain is more tolerated in a diploid state. In order to assess the ploidy of the newly generated variants, FACS were performed on log phase cultures, assessing the cell's DNA content. In addition to the newly generated variants, a diploid strain was added to the FACS analysis

to facilitate identification of the ploidy status. As observed in Figure 3.4, the diploid strain exhibits a first small, and a second, large peak, which correspond to 2C and 4C DNA content, respectively. The presence of a main 4C peak confirms the diploid status of the strain, although the presence of a – yet smaller – 2C peak is unexpected. A 2C peak in a log phase culture is characteristic of a haploid strain as observed for the wild-type in Figure 3.4. The presence of 2C DNA content in a diploid strain could be due to contamination of the diploid strain by a haploid one, or could be related to the difficulty to maintain *S. pombe* in a diploid state (Forsburg *et al.*, 2006). Nevertheless, the *cdc20*$^{EDM}$ variants constructed display a haploid status, indicative of a non-lethal impact of the variants on the strains.

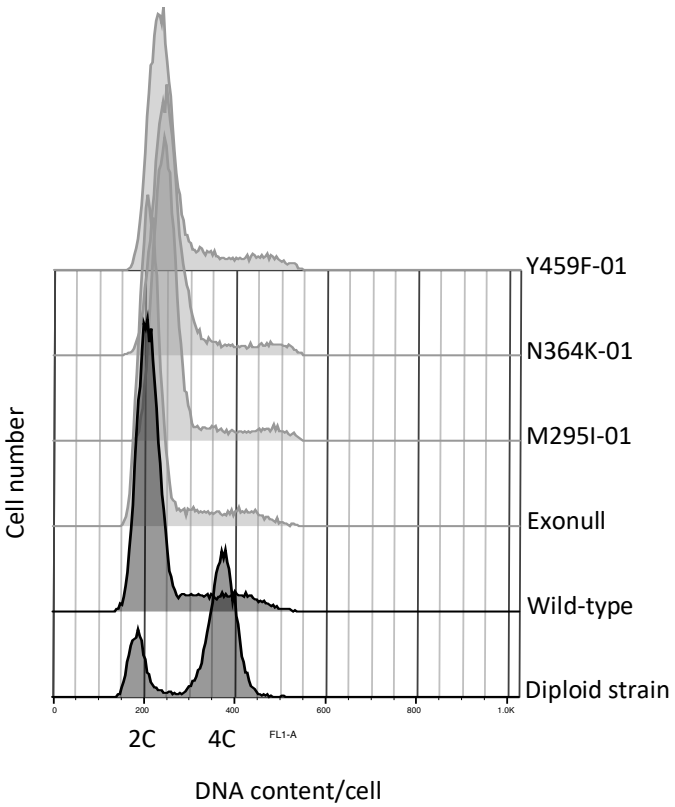



**Figure 3.4 Ploidy status of newly generated strains**

Exponentially growing cultures were treated with Sytox Green for DNA labelling prior to FACS analysis.

### 3.2.3 Germline *cdc20*$^{EDM}$ variants show protein level expression similar to wild-type

It is possible that there could be altered expression levels of Cdc20 in the *cdc20* variant strains, for instance if the mutation causes destabilisation of the protein. To examine this, western blotting was performed to see if steady state levels of the variant proteins were similar to wild-type. Since no antibody against Cdc20 was available, a 3HA tag was added to the C-ter of Cdc20. The 3HA tag was first amplified from the plasmid pFA6a-natMX6-3HA (#998) with a set of primers containing a 81 bp sequence homology to the C-ter of *cdc20* (#1374/#1375). Transformation of the purified PCR fragment into the *cdc20*$^{EDM}$ variant strains followed. Nourseothricin-resistant transformants were checked for correct insertion at the *cdc20* locus via colony PCR using the pair of primers cdc20-3HA-check (#1376/#1377). Protein extractions from exponential phase culture were performed on the newly generated strains before SDS-PAGE and western blotting analysis. The western blot presented in Figure 3.5 informs on a qualitative aspect, not a quantitative one, even if the experimental process was performed three time delivering the same outcome. Hence, analysis of the Figure 3.5 suggests that the protein levels of Cdc20 in all the *cdc20*$^{EDMs}$ strains are similar to that in a fully the wild type strain, as well as the *cdc20*$^{WT\text{-}construct}$ strain.

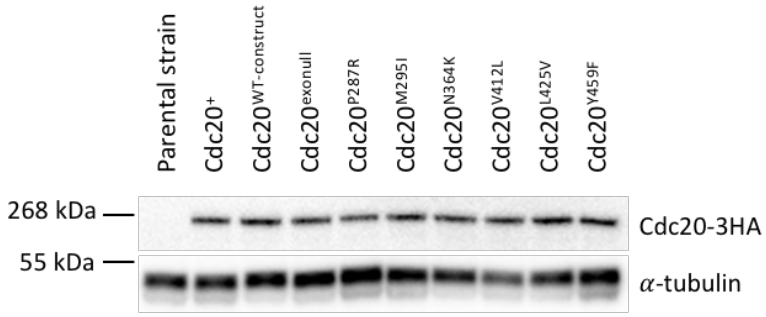



**Figure 3.5 Expression level of Cdc20 in Cdc20-3HA variant strains**

Addition of a 3HA tag at the C-ter of Cdc20 allowed study of the stability of the Cdc20$^{EDM}$ variant and their expression levels after cell lysates were probed by western blot using anti-HA antibody. $\alpha$-tubulin was used as a loading control. There are no notable differences in Cdc20-3HA expression level between the parental strain and variants, WT-construct included. One representative experiment out of three.

### 3.2.4 Some *cdc20*$^{EDM}$ variants exhibit an increase in mutation rate

The predominant question about the *cdc20*$^{EDM}$ variants used in this work is whether they show an increased mutation rate, and if so, similar to previously confirmed pathogenic mutations observed in human tumours. Several techniques are available to estimate the mutation rates. The two main ones are fluctuation analysis and mutation accumulation experiment. Fluctuation analysis is used in this chapter, while the mutation accumulation will be described in chapter 5. The mutation rates relative to the wild-type strain obtained for the different *cdc20*$^{EDM}$ variants are plotted in Figure 3.6. There is no impact on mutation rate for the *cdc20*$^{WT\text{-}construct}$ strain. However, there is an increase for the *cdc20*$^{exonull}$ strain compare to the wild-type, suggesting that the exonuclease activity of Cdc20 is important for faithful DNA synthesis. The *cdc20*$^{M295I}$, *cdc20*$^{V412L}$ and *cdc20*$^{L245V}$ strains exhibit a mutation rate similar to the strain devoid of exonuclease activity. Notably, the mutation rate for *cdc20*$^{P287R}$ has a significant increase compare to wild-type. This increase is even significantly greater than the 3'-5' exonuclease-dead version of *cdc20*, indicating that the lack of proofreading activity is not a sufficient

explanation for this phenotype. The same phenotype – however milder – is also observed for the *cdc20*$^{N364K}$ variants. In addition, one clone out of two for *cdc20*$^{Y459F}$ exhibits a significant increase in mutation rate compared to the wild-type strain, but not significant compare to *cdc20*$^{exonull}$.

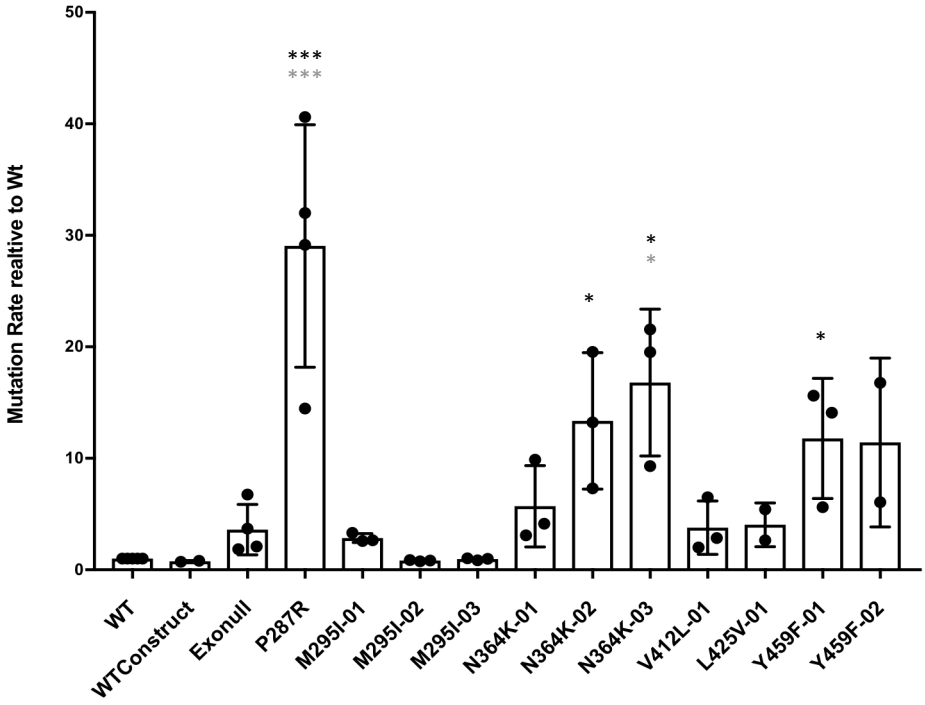



**Figure 3.6 Fluctuation analysis - Canavanine assay**

Mutation rates were determined by fluctuation analysis, using canavanine resistance. Values represent the mutation rate per $10^{-7}$ mutation/generation determined by the Ma-Sandri-Sarkar Maximum Likelihood Estimator Method using rSalvador (Zheng, 2017). When possible, three biological replicates were used. The experiments were performed at least twice for each sample. Analysis of values distribution using the Shapiro-Wilk normality test shows a normal distribution for all samples (except the construct strain, *cdc20*$^{L425V}$-01 and *cdc20*$^{Y459F}$-02 as there were only two values). Following this, a multiple comparison against the WT strain using the ordinary one-way ANOVA revealed that *cdc20*$^{P287R}$ has a significantly greater mutation rate compare to WT. This is also observed for two clones out of three of *cdc20*$^{N364K}$ and one clone out of two for *cdc20*$^{Y459F}$ (black asterisks). An additional multiple comparison using an ordinary one-way ANOVA was performed against the *cdc20*$^{exonull}$ strain. Only *cdc20*$^{P287R}$ and one clone out of three for *cdc20*$^{N364K}$ were exhibiting significantly increased mutation rate compared to *cdc20*$^{exonull}$ (grey asterisks). The data for the V412L-01 and L425V-01 samples were obtained by Ellen Heitzer. * P<0.05; ** P<0.01; **** P<0.0001

### 3.2.5 Some germline *cdc20*$^{EDM}$ variants exhibit abnormally shaped and longer cells similar to the somatic *cdc20*$^{P287R}$

Assessment of cell length of *S. pombe* can indicate a cell cycle defect. Indeed, elongated cells can arise if there is activation of the DNA damage checkpoint, as the cells will continue to grow but are delayed in dividing (Hayles *et al.*, 2013). As observed in Figure 3.7 and Figure 3.8, the wild-type strain exhibits a normal rod-shape in addition to a cell length expected between 7 and 14 µm (Zhang *et al.*, 2014). Similar features are observed for the *cdc20*$^{WT\text{-}construct}$ strain, *cdc20*$^{exonull}$, *cdc20*$^{M295I}$, *cdc20*$^{V412L}$ and two out of three clones for *cdc20*$^{L425V}$. Notably, the somatic variant *cdc20*$^{P287R}$ exhibits significantly elongated cells compare to the parental strain. Furthermore, some cells exhibit a branched shape which could be due to a cytokinesis defect. Furthermore, *cdc20*$^{N364K}$ exhibits a similar phenotype for two out of three clones with clear outliers. *cdc20*$^{Y459F}$ also shows a low frequency of elongated cells but the mean cell size is not significantly different to the wild-type one. Overall, these results suggest that the DNA damage/replication checkpoint is not activated in the *cdc20*$^{EDM}$ variant strains, with the exception of *cdc20*$^{P287R}$.

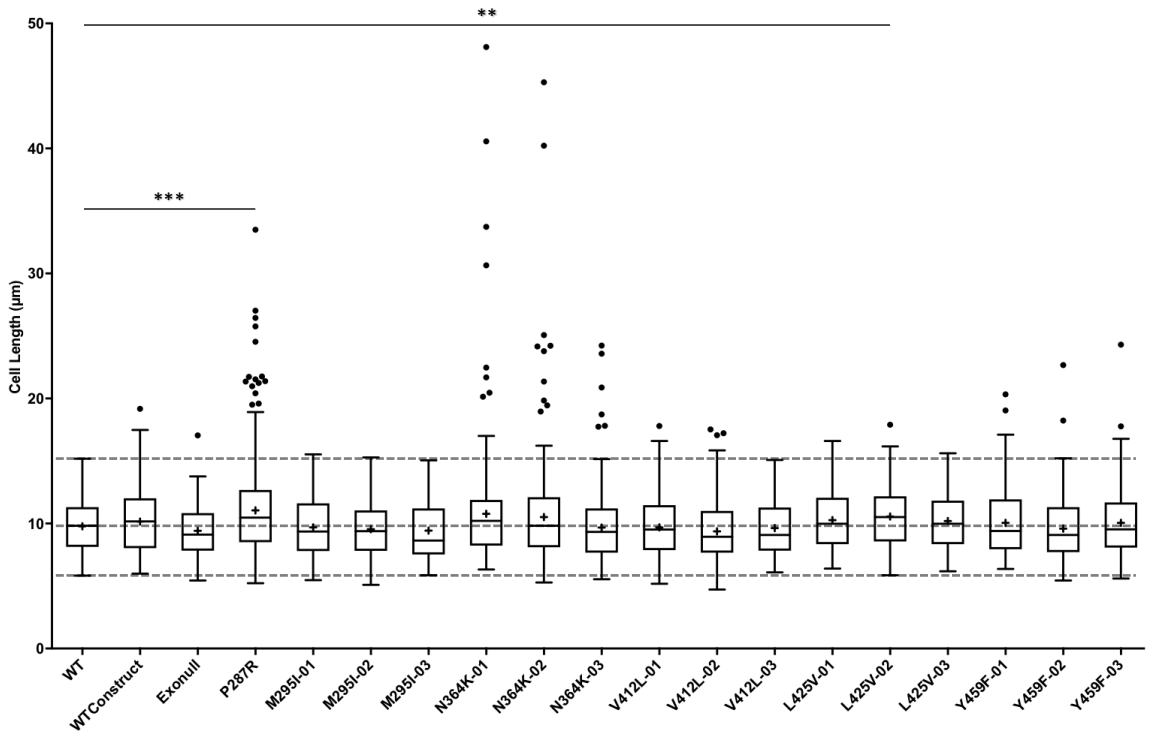



**Figure 3.7 Cell length of Cdc20$^{EDM}$ variant strains**

Cell lengths were obtained from exponentially growing cultures. Between 200 and 400 cells per sample were measured. All samples follow a non-normal distribution (Shapiro-Wilk test); therefore, a multiple comparison against the wild-type strain using the Krustal-Wallis test was performed (unpaired and non-parametric samples comparing more than 2 groups) * P<0.05; ** P<0.01; *** P<0.001. (Tukey graphical representation). This test reveals that *cdc20*$^{P287R}$ shows a significantly longer cell length, as well as one clone for *cdc20*$^{L425V}$. Interestingly, all three clones for *cdc20*$^{N364K}$ show some clear outliers with longer cell length, similar to the ones observed for *cdc20*$^{P287R}$.

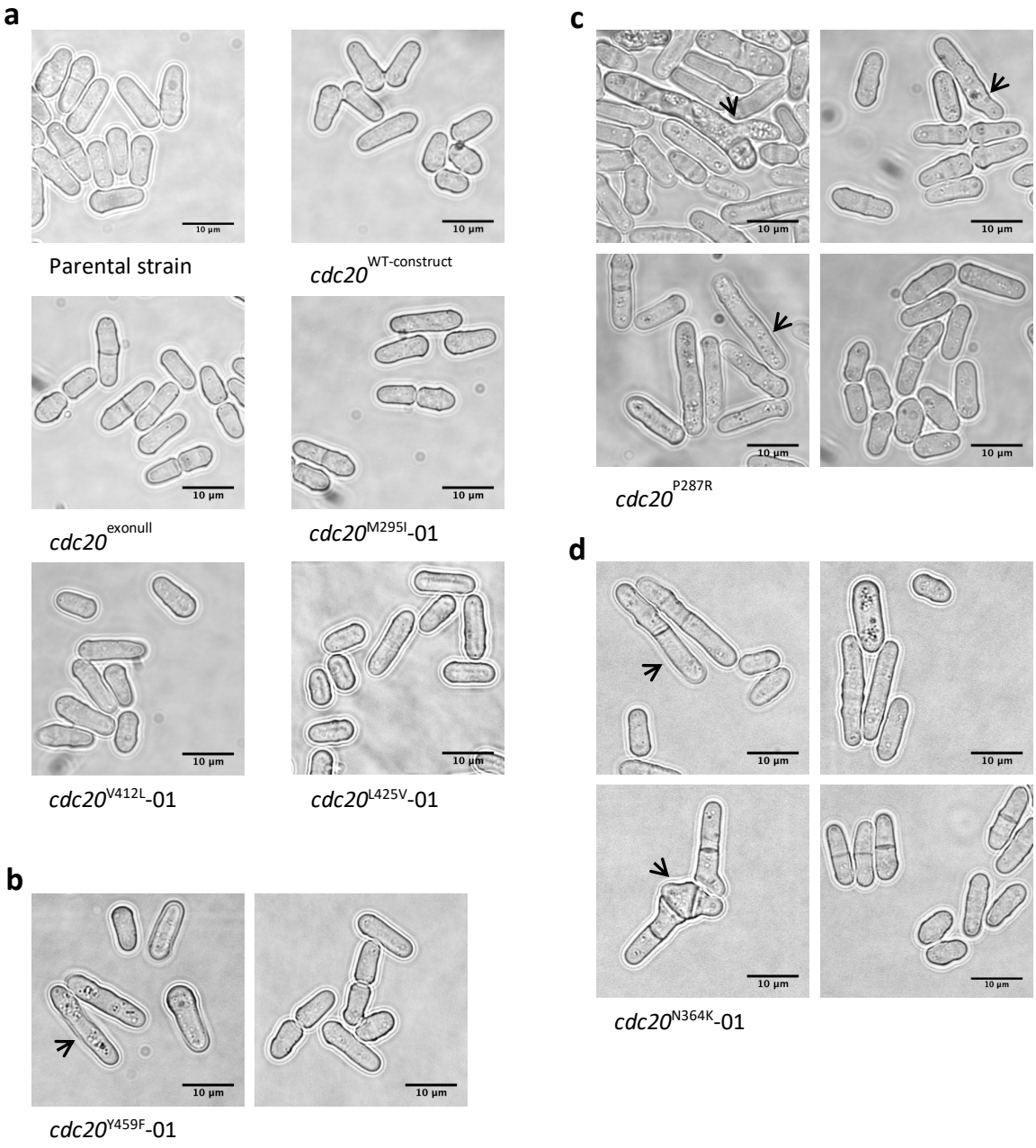



**Figure 3.8 c*dc20*$^{EDM}$ variants phenotype**

Cells exponentially growing were spotted on a slide prior to cell length measurement. Between 200 and 400 cells were measured per samples. (**a**) Parental strain, *cdc20*$^{WT\text{-}construct}$, *cdc20*$^{exonull}$, as well as *cdc20*$^{M295I}$, *cdc20*$^{V412L}$ and *cdc20*$^{L425V}$ exhibit a normal phenotype. (**b**) *cdc20*$^{Y459F}$ exhibit overall a normal phenotype except for some very few elongated cells. (**c**) The somatic mutation *cdc20*$^{P287R}$ exhibits normal cells but also numerous elongated cells as well as some other cells with abnormal features. (**d**) *cdc20*$^{N364K}$ exhibits a similar phenotype compare to *cdc20*$^{P287R}$ but milder.

### 3.2.6 Germline *cdc20*$^{N364K}$ and somatic *cdc20*$^{P287R}$ exhibit a growth defect rate

In order to study the impact of the *cdc20*$^{EDMs}$ on cell growth, growth curves were obtained for each of the three clones for each of the studied variants and doubling times were calculated. As shown in Figure 3.9, the wild-type strain exhibits a doubling time of 3 hours, similar to previously reported ones (Petersen *et al.*, 2016). The same values are observed for the *cdc20*$^{WT\text{-}construct}$, as well as for the *cdc20*$^{exonull}$, *cdc20*$^{M295I}$, *cdc20*$^{V412L}$ and *cdc20*$^{L425V}$. Notably, the somatic variant *cdc20*$^{P287R}$ exhibits a growth defect reflected by a significantly increase in the doubling time compare to wild-type. Similarly, *cdc20*$^{N364K}$ exhibits a growth defect and a significant increase in doubling time compare to wild-type, even if this difference is milder compare to *cdc20*$^{P287R}$. Further to note, the *cdc20*$^{Y459F}$ variant exhibits only one clone with significantly increased doubling time compare to wild-type. Taken together, these results show that loss of Cdc20 exonuclease activity does not impair growth, although some specific *cdc20*$^{EDMs}$ variants do to a certain degree.

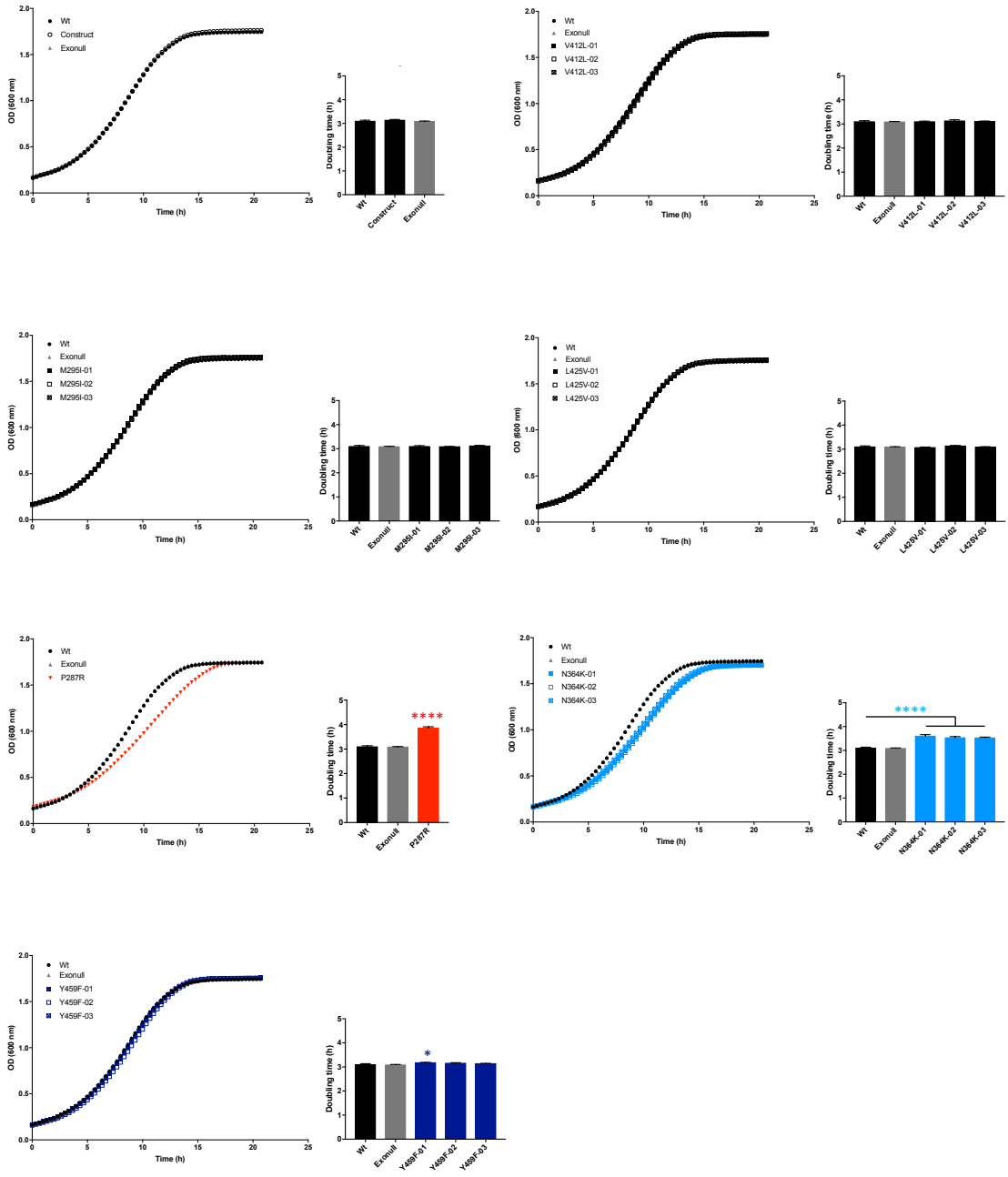



**Figure 3.9 Growth rate analysis of *cdc20*$^{EDM}$ variant strains**

Growth curves for the studied variants were obtained by absorbance recording every 20 minutes for 22h. The growth curves are representative figures from 3 independent experiments. Doubling times were calculated via Prism equation for exponential growth after a non-linear regression analysis: $(\ln(2)) / K$ (K = constant rate in exponential phase); n = 3; Analysis of values distribution using the Shapiro-Wilk normality test shows a normal distribution for all samples. Following this, a multiple comparison against the wild-type using the ordinary one-way ANOVA was performed; * P<0.05; **** P<0.0001.

## 3.3 Discussion

Numerous germline *POLE*$^{EDMs}$ were identified as potentially associated with cancer development predisposition (PPAP; see section 1.3). To further characterise their pathogenicity, the germline *POLE*$^{EDM}$ variants were constructed in fission yeast (see section 3.2.1). No homozygous *POLE*$^{EDMs}$ are yet to be identified in human cancer (Rayner *et al.*, 2016) and murine models with *POLE*$^{P286R}$ hemizygous status are, in most instances, embryonically lethal (H. D. Li *et al.*, 2018). This suggests that homozygous *POLE*$^{EDMs}$ might be lethal for the cell, although all studied variants are viable in a haploid status, as similarly reported in budding yeast (Barbari *et al.*, 2018). The difference of viability between mammalian and yeast cells could be explained by a difference in mutagenesis of Pol ε in human and yeasts.

The main oncogenic mechanism of *POLE*$^{EDMs}$ is thought to be an increase in mutation burden (see section 1.3). Alteration of one of the subunits of Pol ε can lead to an unstable holoenzyme, decreasing the fidelity of the enzyme, ultimately leading to an increase in mutation rate (Bellelli *et al.*, 2018; Isoz *et al.*, 2012; Ohya *et al.*, 2000). Our data suggest that this does not apply on the studied variants as no change of Cdc20 expression occurs in the constructed strains (Figure 3.5). A previously reported *in vitro* purification assay, where stability of the Pol2-P301R holoenzyme variant (Polε-P286R human ortholog) was not impaired, confirms our finding (Xing *et al.*, 2019).

Mutation in the exonuclease domain of Pol ε can impede the proofreading ability of the replicase, increasing the mutation burden after each round of cell division (see section 1.3). Notably, mutation rate varies depending to the *POLE*$^{EDM}$ variants (Barbari *et al.*, 2018). Our data exploring the mutation rate of the germline *POLE*$^{EDM}$ variants

emphasize this notion (Figure 3.6). There is no published study which looked at the spontaneous mutation rate of Pol ε-M294I. However, Pol2-M309R variant (Pol ε-M394R human ortholog) is predicted to destabilise the protein structure and the ion metal coordination necessary for the exonuclease activity (Hamzaoui *et al.*, 2020). Fluctuation analysis showed an increase in mutation rate similar to that of a wild-type strain, slightly below the proofreading deficient strain. Methionine and isoleucine are both amino acids with hydrophobic side chain while arginine is electrically charge. Thus, a change from methionine to arginine (M294R) is predicted to have a stronger impact on the protein function than a methionine to isoleucine change (M294I). Even if highly conserved through species (Figure 1.5) and identified as deleterious via Condel, our data suggest that it is unlikely to be pathogenic.

The mutation rate for the *cdc20*$^{V412L}$ and *cdc20*$^{L425V}$ variants shows similar results to *S. cerevisiae* constructs (Barbari *et al.*, 2018; Hamzaoui *et al.*, 2020). Equivalent to an exonuclease-dead Pol ε version, this reinforces the hypothesis that either additional alterations are necessary in human tumour to trigger oncogenesis, or that Pol ε possesses a distinct function in yeasts. The *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ strains were found to have a mutation rate greater than that of a strain devoid of exonuclease activity, as previously reported in budding yeast (Barbari *et al.*, 2018; Hamzaoui *et al.*, 2020; Xing *et al.*, 2019). For the first time, the variant Pol ε-Y458F, which in our study exhibits a mutation rate higher than that of Pol ε exonuclease-deficient strain, was studied outside of a clinical context. A mutation rate greater than that of exonuclease-dead Pol ε version suggests that, for some variants, simple loss of Pol ε exonuclease activity is not sufficient to explain the dramatic increase in mutation rate. The mutagenic ability of Pol ε-P286R was collaboratively studied in budding yeast (Parkash *et al.*, 2019; Xing *et al.*, 2019). The

crystal structure of Pol2$^{P301R}$ (Parkash *et al.*, 2019) revealed that the mutation impairs exonuclease activity via the creation of a steric hindrance for the movement of the 3'end primer terminus to the exonuclease site, forcing the enzyme to stay in polymerase mode, and extend from mismatches. Thus, the spontaneous mutation rate observed for the *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ variants support the fact that the mechanism of mutagenicity is not only related to a lack of proofreading activity but could also be due to a structural impairment of the catalytic activity of the enzyme. Taken together, it seems that the *POLE*$^{V411L}$ and *POLE*$^{L424V}$ variants might be impaired similarly as a proofreading deficient polymerase, while *POLE*$^{M294I}$ is affected to an even lesser extent. The variant *POLE*$^{N363K}$ and *POLE*$^{Y458F}$ are however suspected to have a similar impairment to *POLE*$^{P286R}$.

A note on the discrepancy of results observed within clones for the same *cdc20*$^{EDMs}$ variant in the fluctuation analysis. The discrepancies found could be explained by a number of factors, which first could be related to the experimental aspect of the fluctuation analysis. First is the sensitivity of the canavanine assay. Due to unforeseen events, the fluctuation analysis were performed in different laboratory environments and were spread through time. While in some environment the wild-type did not exhibit any colonies, in others some would start to appear. Second, colony counting was performed naked-eye, which could lead to explain high standard deviation as all colony size (even very small ones) were counted. Not related to the fluctuation analysis experiment, another aspect to consider is a possible chromosomal loop-out of the plasmid which carries the *cdc20*$^{EDM}$ studied, as contact with G418 for selection pressure to maintain the plasmid was performed only from transformation to storage of the strain. Cultures prior to fluctuation analysis did not include G418 selection, therefore

could lead to a heterogeneous population of cells carrying or not *cdc20*$^{EDM}$. It is also important to bear in mind that, in the case of a selection pressure maintained, the presence of a truncated wild-type version of the exonuclease domain could be used as a repair template, resulting in a truncated and non-functional, and full wild-type version of *cdc20*. Similarly, maintenance of the selection pressure could not prevent transient loop-out and reintegration of the plasmid at a different locus, possibly restoring *cdc20* wild-type and a truncated *cdc20*$^{EDM}$. Altogether, these genetic alterations could explain the clone-to-clone variations observed in the fluctuation analysis. Finally, the mechanism of canavanine resistance in *S. pombe* is uncertain. It is necessary to bear in mind that, given the signature of *POLE*$^{EDMs}$ observed in human cancer (Alexandrov *et al.*, 2020; Alexandrov, Nik-Zainal, Wedge, Aparicio*, et al.*, 2013), if several genes are related to canavanine resistance in our assays, some genes might be more likely to be mutated than others. It would be therefore necessary to perform additional fluctuation analysis, e.g. histidine reversion or 5-fluororototic acid resistance, to confirm the mutagenic phenotype observed in this study.

Increase in mutation rate could negatively impact the cellular metabolism and lead to checkpoint activation (Mertz *et al.*, 2015; L. N. Williams *et al.*, 2015). Our data support this hypothesis for the strains exhibiting an increase in spontaneous mutation rate greater than that of simple lack of proofreading activity, although there are disparities within this group. In order to refine our data, it might be necessary to first, investigate via cell cycle progression analysis (FACS) if a longer S-phase supports the observed growth defect, and second, use genetic interactions with DNA damage checkpoints, to determine the activated pathway used for the cell cycle regulation.

In summary, this work is a first exploration of the pathogenic potential of germline *POLE*$^{EDMs}$. Further characterisation needs to be performed to evaluate their pathogenicity.

# Chapter 4

## Mechanism of hypermutation driven by *cdc20*$^{EDM}$ in fission yeast

### 4.1 Introduction

Some of the *POLE*$^{EDMs}$ are associated with a hypermutated phenotype suspected to be related to oncogenic development (Cancer Genome Atlas, 2012; Cancer Genome Atlas Research *et al.*, 2013). In our study, some *cdc20*$^{EDM}$ variants constructed in fission yeast exhibit a phenotype similar to their human ortholog (see chapter 3). EDMs might impair proofreading ability (see section 1.3), although the catalytic activity differs amongst the *POLE*$^{EDM}$ tested (Shinbrot *et al.*, 2014; Xing *et al.*, 2019). Therefore, investigation of exonuclease activity of studied *POLE*$^{EDMs}$ is necessary to clarify their role in mutagenic ability.

Pol $\delta$ exonuclease-deficient budding yeast strains exhibit checkpoint activation (Datta *et al.*, 2000), similarly to *cdc20*$^{\Delta N\text{-ter}}$ which requires the DNA checkpoint control Rad3 for viability (Feng *et al.*, 2001). In budding yeast, the mutator phenotype observed in Pol $\varepsilon$ and Pol $\delta$ with altered base-selectivity is further amplified when dNTP levels are modulated (Mertz *et al.*, 2015; L. N. Williams *et al.*, 2015). Even if *pol2-P301R* strain does not exhibit altered dNTP pools (Xing *et al.*, 2019), the high mutation burden (Figure 3.6) and cell growth defect (Figure 3.7) observed in some of the hypermutated *cdc20*$^{EDMs}$ variants suggest further investigation on cell cycle defect and dNTP levels to explain their phenotype, as dNTP levels are tightly regulated by cell cycle (Mathews, 2015).

Translesion synthesis polymerases have been identified as possible factors partly responsible for mutated phenotype observed in replicative polymerases (Northam *et al.*, 2006). Although this does not apply for Pol2-P301R in budding yeast (Xing *et al.*, 2019), recent work from our group on the same Pol ε-EDM variant suggests that the response might be different in fission yeast. Not all TLS polymerases are conserved between budding and fission yeast. One of the TLS polymerases, absent in budding yeast, does in fission yeast impact the mutation rate of *cdc20*$^{P287R}$ (I. Soriano, unpublished observations), prompting use to investigate the role of such polymerase in our hypermutated germline variants.

Altogether, this chapter disentangled the mechanism of hypermutation observed in *POLE*$^{EDM}$ germline variants.

## 4.2 Results

### 4.2.1 *In vitro* study of human *POLE*$^{EDM}$ variants

As observed in the previous results chapter, some *POLE*$^{EDM}$ variants exhibit an increase in mutation rate, for some greater than that which results from a simple loss of exonuclease activity. This leads to question whether the studied EDMs are causing loss of exonuclease activity, but also possibly other of the polymerase functions impairment. In order to address this question, exonuclease activity was assessed *in vitro* thanks to a collaborative work with the P$^{r}$ Sue Cotterill. D$^{r}$ Gilles Crevel, post-doctoral researcher in P$^{r}$ Cotterill's laboratory, graciously generated the human holoenzyme variants. To do so, Sf9 cells were co-infected with 4 baculoviruses, each expressing one of the human Pol ε holoenzyme subunits (POLE, POL2, POLE3 and POLE4). The second subunit, POLE2, was tagged at its C-ter with a FLAG peptide, allowing purification via affinity chromatography. The Pol ε exonuclease-dead version (subsequently referred as Pol ε-exonull) was obtained after mutation of the D275 and E277 residues to alanine, residues critical for exonuclease activity (see section 1.2.4.2).

#### 4.2.1.1 Activity assay

Prior to exonuclease activity assessment, the activity of the holoenzyme batches was evaluated. To do so, a 21-mer oligo labelled with Cy5 at its 5' end was annealed with a 42-mer oligo. Incubated at 37°C, time points were taken prior to electrophoresis of the samples on acrylamide gels in denaturing conditions. Activity was evaluated in a time course assessing the rate of appearance of a full length 42 bp product. As observed in Figure 4.1, for Pol ε wild-type version an increase in the intensity of the 42-mer band is

visible along the timeline of time-points. A similar pattern was observed for all the studied variants. However, the intensity of the final time-point differs amongst some of the studied variants. There is no notable difference between the wild-type and the exonuclease-dead version. Pol ε-M294I and -V411L show a similar pattern even if the enzymatic concentration is almost five times lower than the wild-type protein. Pol ε-P286R and -N363K variants show a lower intensity of synthesised oligo which could be translated by a polymerase activity impairment. However, it is important to mention that different batches of the same enzyme may have different activities simply reflecting the variability in preserving enzyme activity during the purification process. Considering that this qualitative assay aims to evaluate the activity of the purified holoenzymes, quantities were adjusted prior to *in vitro* exonuclease assays. Thus, for Pol ε wild-type, -exonull, -M294I, and -L424V, the same concentrations were used as in this assay, while Pol ε-P286R, -N363K and -V411L concentrations were doubled to obtain comparable activities.

```
5′ Cy5 - TTTTCCCAGTCACGACGTTGG
    3′ - AAAAGGGTCAGTGCTGCAACCTTCTGTTGCCGGTCGCTGTCG
```

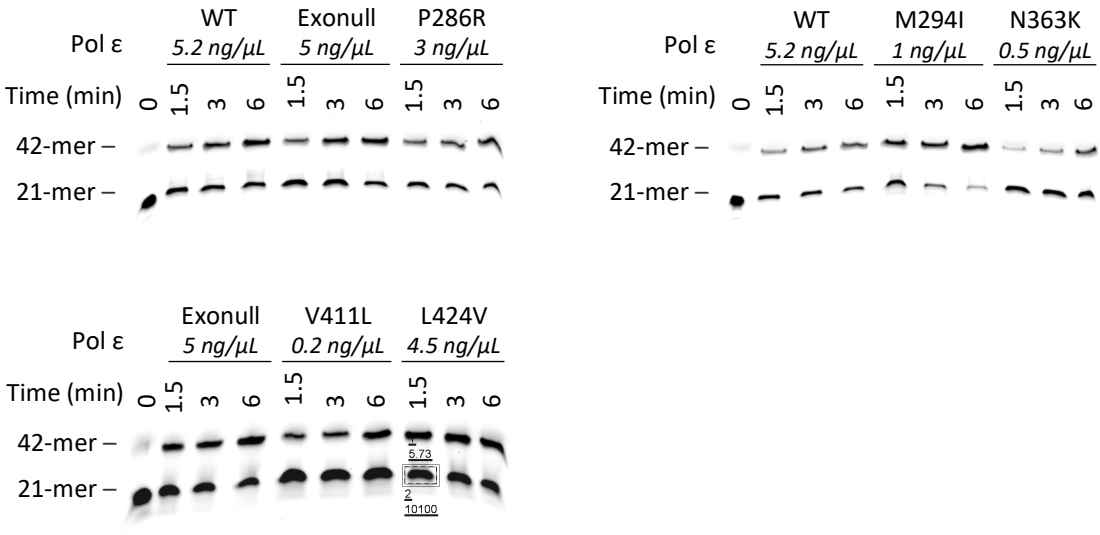



**Figure 4.1 *In vitro* study of human *POLE*$^{EDM}$ variants – holoenzyme activity**

DNA polymerase activity was assayed through the extension of the indicated substrate. Fully extended substrate would be of 42 bp, against 21 bp for un-extended. Incubating at 37°C for the indicated time points, enzymatic reactions were stopped by addition of EDTA at final concentration of 0.1 M. Samples were then run on acrylamide gel under denaturing conditions. Representative of two independent experiments.

#### 4.2.1.2 Exonuclease activity

After adjustment of the concentrations of the Pol ε variants, assessment of their exonuclease activity was performed using two different substrates. The first substrate (Figure 4.2 – a) was obtained after annealing of a 22-mer labelled with Cy5 at its 5'end, with a 34-mer which contains a A:G mismatch at its 3'-end which should trigger exonuclease activity. The assay was carried out in the absence of dNTPs to prevent extension of the primer. After incubation, the substrate was run on an acrylamide gel in denaturing conditions.

The wild-type version of Pol ε already exhibits oligo degradation after 1.5 minutes of incubation at 37°C. Nearly no full-length oligo was left visible after 8 minutes of incubation. Contrastingly, Pol ε-exonull version does not exhibit any exonuclease activity at all, similarly to Pol ε-P286R, even after 8 minutes of incubation (Figure 4.2 – a).

The second substrate used was an un-annealed version of the 5'-Cy5 substrate (Figure 4.2 – b), as exonuclease activity can be triggered either by dsDNA exposing a mismatch or ssDNA alone. Exonuclease activity was assayed as previously described, although with longer incubation time to assess potential residual activity. As observed for the annealed oligos, exonuclease activity is already visible after 1.5 minutes incubation for the wild-type protein, and barely no full-length oligo detected after 15 minutes incubation (Figure 4.2 – b). No exonuclease activity is visible for Pol ε-exonull, even after extended incubation. A smear is however detected, possibly related to enzymatic contamination of the sample (also, this assay was performed twice and no exonuclease activity was observed on the second assay, even after 2 hours of incubation). The variants Pol ε-M294I exhibits an exonuclease activity similar to the wild-type version, while Pol ε-L424V exhibits an impaired one. As observed in Figure 4.2 – a, Pol ε-P286R does not exhibit exonuclease activity even at the 2 hours timepoint. The same phenotype is observed for the variants Pol ε-N363K and -V411L. Overall, these data confirm D275 and E277 as essential for the exonuclease activity of Pol ε. Furthermore, Pol ε-P286R is confirmed as being devoid of exonuclease activity, similarly to the exonuclease-dead version of the protein. In addition, Pol ε-N363K and -V411L are the only germline variants exhibiting a similar loss of exonuclease activity.

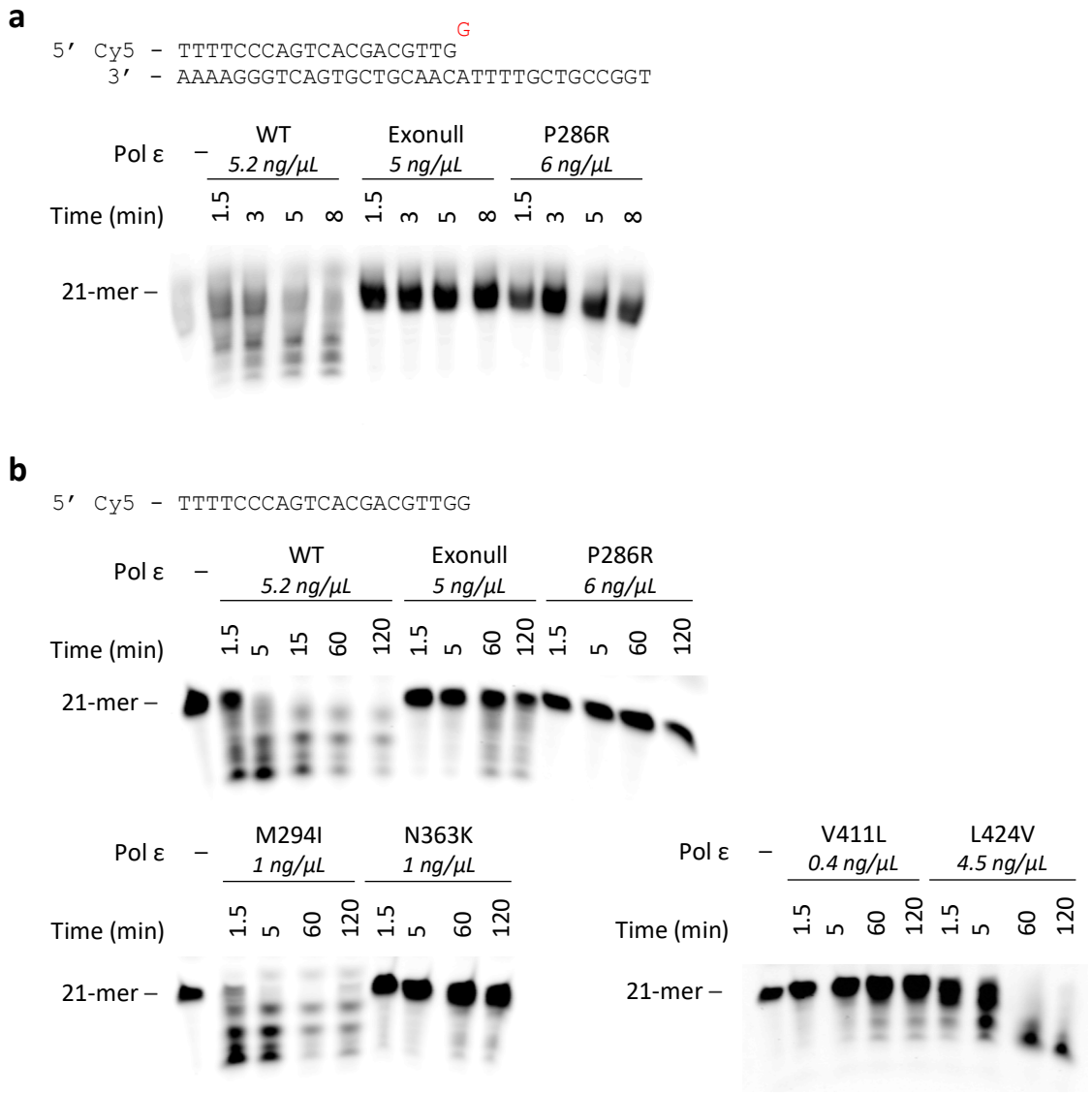



**Figure 4.2 *In vitro* study of human *POLE*$^{EDM}$ variants – Exonuclease activity**

(**a**) Exonuclease activity was assayed on 12 fmol of substrate containing a terminal A:G mismatch. Incubating at 37°C for the indicated time points, enzymatic reactions were stopped by addition of EDTA at final concentration of 0.1 M. Samples were then run on acrylamide gel under denaturing condition. Representative of two independent experiments. (**b**) Exonuclease activity was assayed with 12 fmol of single-stranded oligonucleotide substrate. Incubating at 37°C for the indicated time points, enzymatic reactions were stopped by addition of EDTA at final concentration of 0.1 M. Samples were then run on acrylamide gel under denaturing condition. Representative of two independent experiments.

#### 4.2.1.3 *In vitro* assays limitations

It is important to mention several aspects of the *in vitro* assays which might renders the result interpretation sensitive. First, the variability of enzymatic concentration amongst Pol ε-EDMs used in the assays, in addition to the lack of quantification, for both activity and exonuclease assay, render the result interpretation difficult. Second, in light of the recent Xing *et al.* (2019) which showed that Pol2-P301R variant has a higher polymerase activity compared to a wild-type version, questions whether normalising the enzyme quantity based on the polymerase activity is the right way of doing.

### 4.2.2 Do *cdc20*$^{EDM}$ strains that show a cell cycle defect rely on Rad3 for survival?

As previously noticed, some *POLE*$^{EDM}$ variants exhibit a longer cell length phenotype which could be explained by a delayed cell cycle (see section 3.2.5). Rad3 is involved in recognition of DNA damages and activation of checkpoint in fission yeast (Humphrey, 2000). In conjunction with Rad9, Rad1 and Hus1, this ATR-homologue associates with damaged DNA, leading to a cascade of phosphorylation events and consequences which include delay to mitotic entry. In order to evaluate if *cdc20*$^{EDM}$ variants with cell cycle defect rely on Rad3 for survival, *cdc20*$^{exonull}$, *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ strains were crossed with a *rad3Δ* or *rad3ts* strains and double mutants were derived by tetrad dissection.

#### 4.2.2.1 *cdc20*$^{EDM}$ variants with cell cycle defect are synthetically sick with *rad3Δ*

As shown in Figure 4.3 – a, viable *rad3Δ* double mutants for all three studied *POLE*$^{EDMs}$ were obtained. Notably for *cdc20*$^{P287R}$, double mutants show smaller colony sizes.

Growth rates were measured on three double mutants in parallel to the single mutants. As observed in Figure 4.3 – b/c, the *rad3Δ* strain exhibits an increase in doubling time, similar to the ones observed for *cdc20*$^{exonull}$ double mutants. For the *cdc20*$^{P287R}$ double mutants, disparity is observed amongst the three clones. One strain shows no sign of growth rate defect compare to the single *cdc20*$^{P287R}$ strain, whereas the two others show clear sickness. A similar pattern is observed for *cdc20*$^{N364K}$, although the effect on growth rate is milder compared to *cdc20*$^{P287R}$.

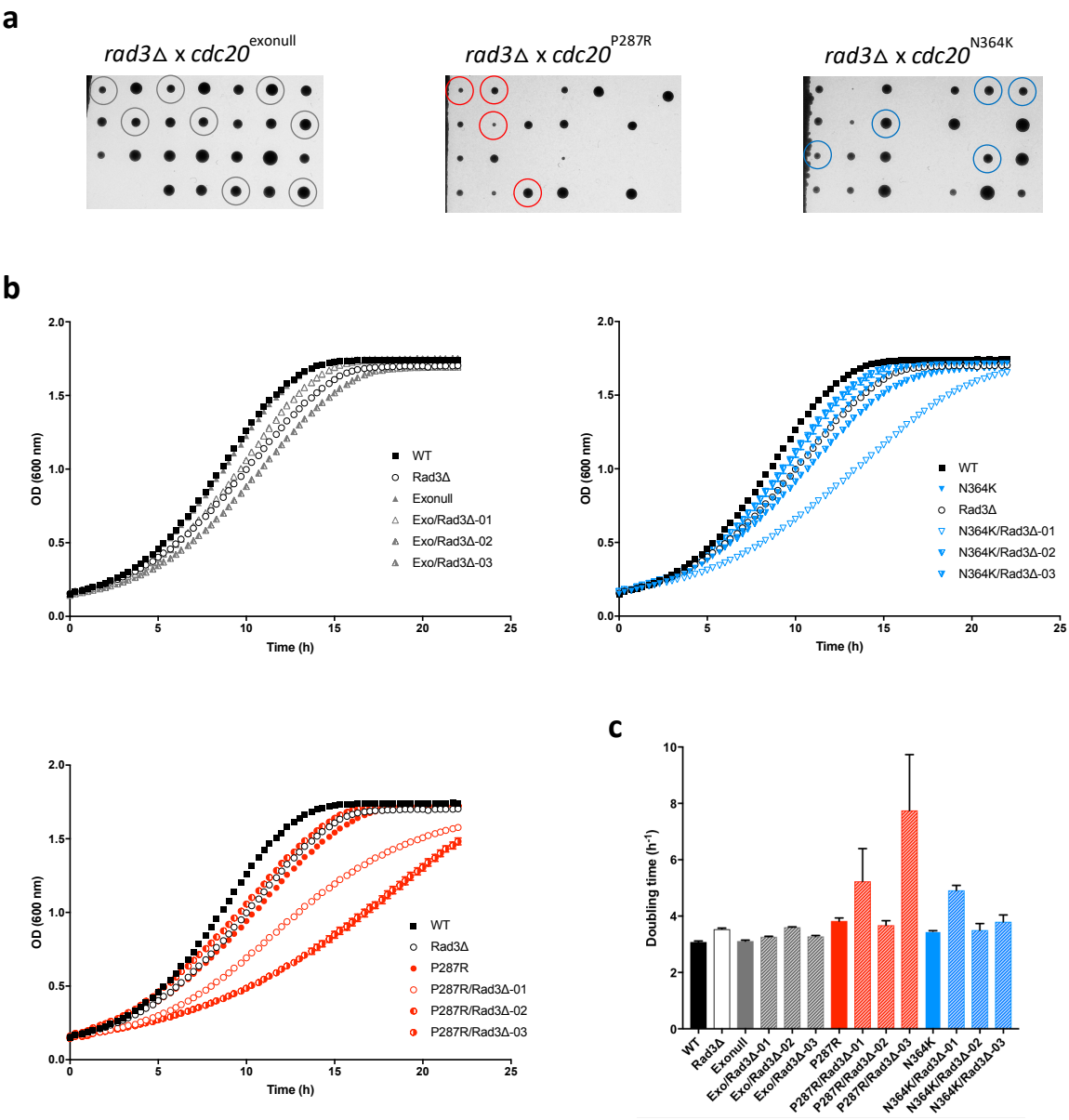



**Figure 4.3 Analysis of *cdc20*$^{EDM}$ *rad3*Δ double mutants**

(**a**) Tetrad analysis of *rad3*Δ and either *cdc20*$^{exonull}$, *cdc20*$^{P287R}$, or *cdc20*$^{N364K}$. Double mutants are circled in grey, red and blue, respectively. (**b**) Growth curves for the studied strains were obtained by absorbance recoding every 20 minutes for 22h. Representative figures from 3 independent experiments. (**c**) Doubling times were calculated via Prism equation for exponential growth after a non-linear regression analysis: $(\ln(2)) \,/\, K$ (K = constant rate in exponential phase); n = 3.

#### 4.2.2.2 In contrast to *cdc20*$^{N364K}$, *cdc20*$^{P287R}$ is synthetically lethal with *rad3ts*

One reason for the variability on phenotype of the *rad3Δ cdc20*$^{P287R}$ and *rad3Δ cdc20*$^{N364K}$ could be the emergence of suppressor mutations that alleviate the phenotype of the double mutants. Hence, a second strategy was used to study the requirement of Rad3 in a *cdc20*$^{EDM}$ background, namely to use a *rad3ts* allele which would allow the double mutant phenotype to be studied immediately after Rad3 inactivation. *cdc20*$^{exonull}$, *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ were crossed with a *rad3ts* strain and tetrad dissected. *cdc20*$^{EDM}$ variants were identified by G418$^{R}$, while *rad3ts* ones showed sensitivity to HU when incubated at 36°C. Spot assays and growth rates assessment for the double mutants at restrictive temperatures were analysed. In contrast to *cdc20*$^{N364K}$, the *cdc20*$^{P287R}$ mutation is synthetically lethal with *rad3ts* (Figure 4.4). This analysis also confirms that *cdc20*$^{exonull}$ does not require Rad3 for survival. On the other hand, *cdc20*$^{N364K}$ does not show synthetic lethality when combined with *rad3ts*.

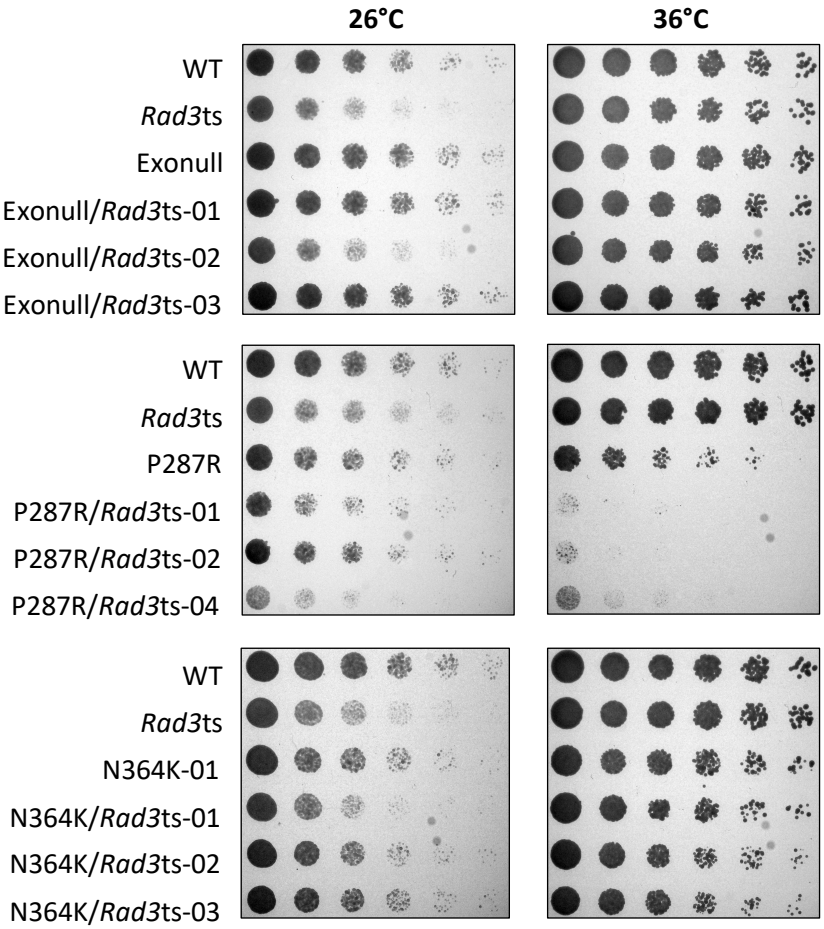



**Figure 4.4 Analysis of *cdc20*$^{EDM}$ *rad3ts* double mutants – Spot assay**

Indicated strains were spotted onto rich YE3S media agar plates and incubated for two days at indicated temperature. One representative experiment out of two.

A growth rate experiment was also performed on the *rad3ts* double mutants. Ahead of the 22 hours of growth rate recording, the cell cultures were refreshed for 3 hours at 26°C prior to 2 hours incubation at 36°C. This process aimed to inactivate Rad3 prior to the start of growth recording. As observed in Figure 4.5, the *cdc20*$^{P287R}$ *rad3ts* strains exhibit a growth rate defect in contrast to the *cdc20*$^{exonull}$ and *cdc20*$^{N364K}$ double mutants. However, *cdc20*$^{P287R}$ double mutants exhibit a striking synthetic sickness for two of the double mutant clones, whereas one clone has a shorter doubling time than *cdc20*$^{P287R}$ single mutant.

All together, these data suggest that the response of *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ to Rad3 inhibition varies depending on the type of Rad3 inactivation, or the time allocated to the strains to potentially develop suppressor mutations that would allow them to grow. Indeed, the *cdc20*$^{P287R}$ mutation shows a dependency on Rad3 for survival in a *rad3ts* context, possibly as this context gives little time to the double mutant strains to acquire suppressor mutations. *cdc20*$^{N364K}$ seems to not rely on Rad3 for survival, only being sick with Rad3Δ, probably as its effect on Pol ε function is less severe than *cdc20*$^{P287R}$.

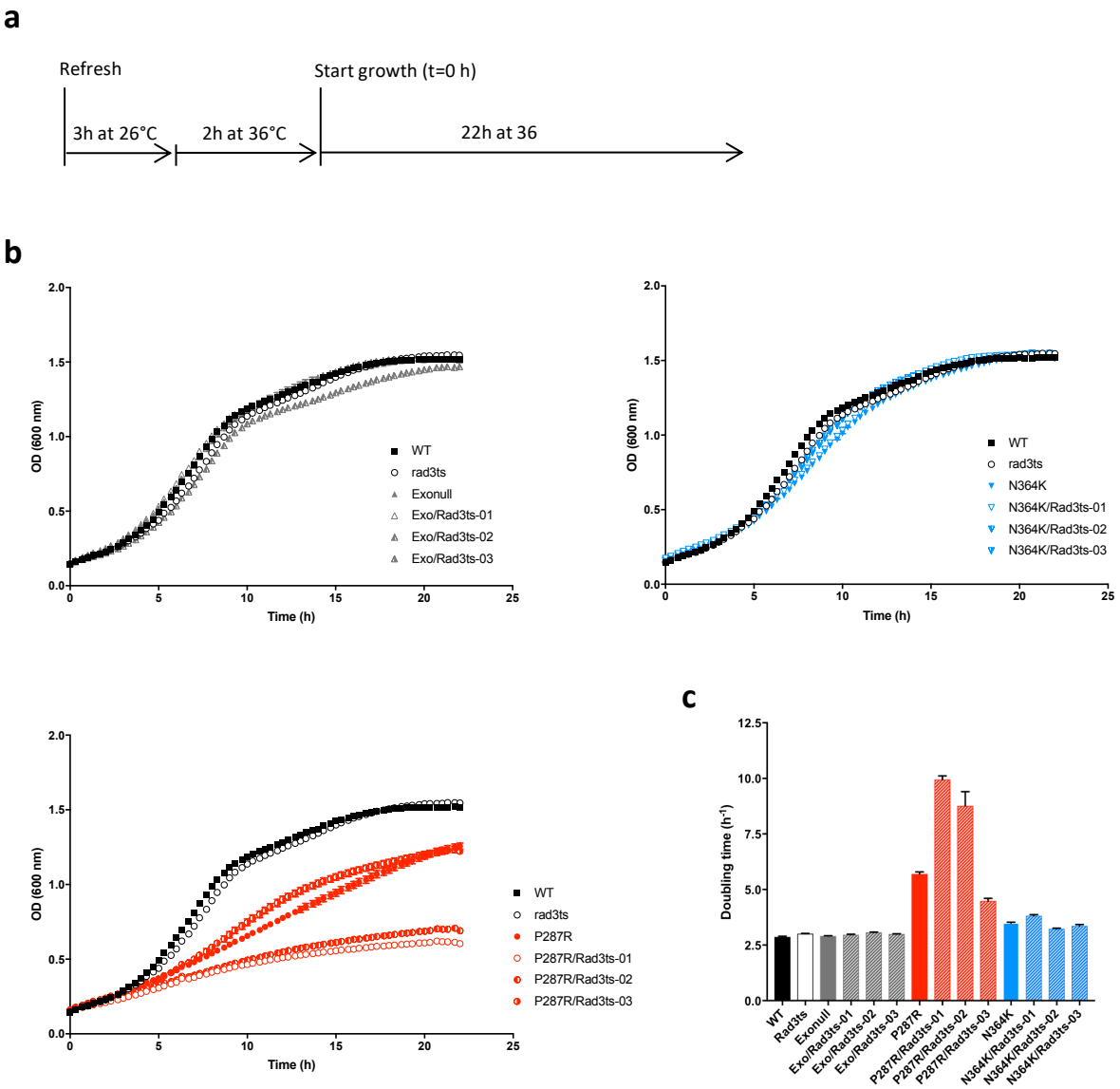



**Figure 4.5 Analysis of *cdc20*$^{EDM}$ *rad3ts* double mutants – Growth curve**

(**a**) Experimental timeline. After being refreshed, the cells were incubated first at permissive temperature for 3h, then at 36°C for 2h in order to initiate Rad3 destabilisation. Growth was then recorded at 36°C for 22h. (**b**) Growth curves for the studied strains were obtained by absorbance recoding every 20 minutes for 22h. Representative figures from 3 independent experiments. (**c**) Doubling times were calculated via Prism equation for exponential growth after a non-linear regression analysis: $(\ln(2)) \,/\, K$ (K = constant rate in exponential phase); n = 3.

### 4.2.3 Response of *cdc20*$^{EDM}$ mutants to DNA-damaging agents

The replication fork encounters numerous obstacles during DNA synthesis. As Pol ε is part of the replisome, a mutation in one of its catalytic domains could possibly alter replication (see section 1.1.3.2). The *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ strains have a longer doubling time, although this is not simply due to loss of exonuclease activity, since a similar phenotype is absent in the strains expressing the exonull version of *cdc20* (Figure 3.9). In addition to effect on mutation rates (Figure 3.6), *cdc20*$^{EDM}$ variants may be sensitive to DNA-damaging agents. Fission yeast strain lacking the N-ter of Cdc20 are sensitive to UV-C and MMS, suggesting that the strains are defective in DNA repair (Feng *et al.*, 2001). This could reflect the involvement of Pol ε in DNA repair *per se*, or that Pol ε variants could be more sensitive to template damaged during S-phase. To test whether the strains are sensitive either to replication blockage or DNA-damaging agents, spot assays were performed on rich media combined with a broad range of compounds of known functional consequence. Temperature sensitivity of the strains was also examined in case the EDM mutations rendered the polymerase less stable. A note to the reader: in order to facilitate the reading, the same figure for the YE3S control plate incubated 2 days at 30°C was used for each of the conditions tested.

#### 4.2.3.1 *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ exhibit a milder temperature sensitive phenotype

The optimum growth temperature for *S. pombe* lies between 30°C and 32°C. Incubation at 36°C however allows identification of temperature-sensitive mutants, usually exhibiting a growth defect due to mutations affecting the correct folding of the encoded protein. Growth at 36°C compared to 30°C reveals a clear growth defect for *cdc20*$^{P287R}$

compared to the parental strain (Figure 4.6). None of the other strains showed a clear difference between the two temperature, except maybe slightly for *cdc20*$^{N364K}$.

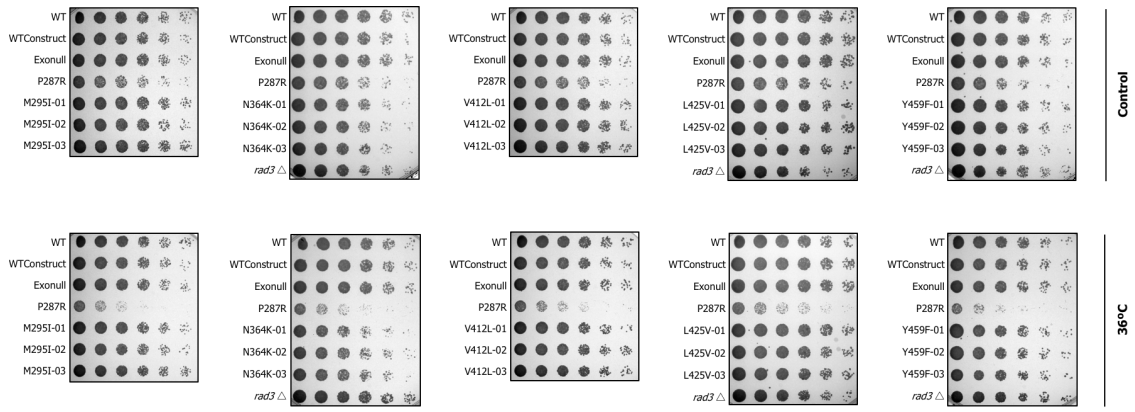



**Figure 4.6 Temperature sensitivity of *cdc20*$^{EDM}$ strains**

Indicated strains were spotted onto rich YE3S media agar plates and incubated for two days. The control plate was incubated at 30°C. One representative experiment out of 3.

### 4.2.3.2 *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ exhibit sensitivity to bleocin

Isolated from *Streptomyces verticillus*, the antibiotic bleocin produces DNA double-strand breaks. When spot assays were performed to assess the sensitivity/resistance of the studied strains, growth was performed at 26°C instead of 30°C as bleocin gets inactivated at higher temperature. When in contact with the drug, the parental strain as well as *cdc20*$^{WT\text{-}construct}$ and *cdc20*$^{exonull}$ do not show any sensitivity. However, *cdc20*$^{P287R}$ exhibits a clear sensitivity to the drug compared to the parental strain. In addition, *cdc20*$^{N364K}$ shows a similar phenotype, although milder (Figure 4.7).

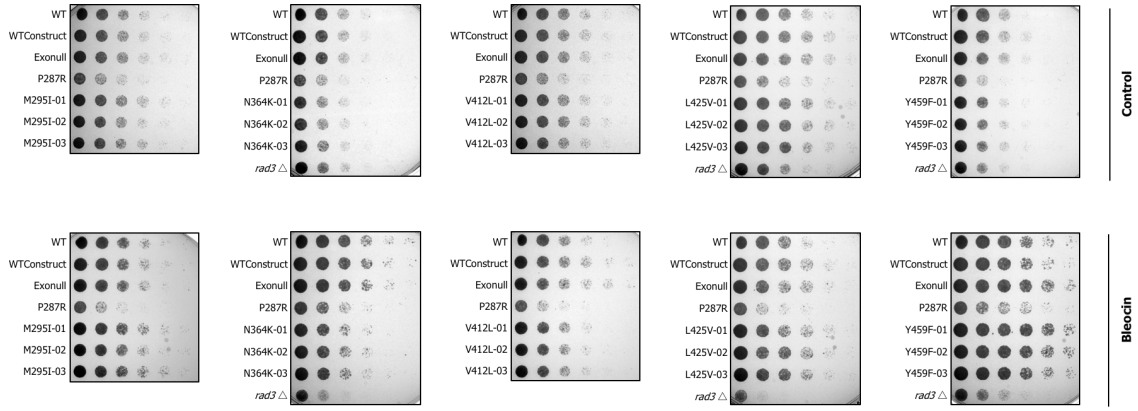



**Figure 4.7 Response of *cdc20*$^{EDM}$ strains to DNA double-strand breaks**

Indicated strains were spotted onto rich YE3S media agar plates with or without bleocin at 0.5 mg/mL, and incubated for two days at 26°C. One representative experiment out of 3.

### 4.2.3.3 *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ exhibit sensitivity to ultraviolet (UV-C) radiation

UV irradiation induces formation of a number of lesions in DNA, including cyclobutane-pyrimidine dimers formed between two adjacent pyrimidines, generating toxic and mutagenic DNA lesions. Replacement of the damaged DNA with new and undamaged nucleotides is usually done by NER, although *S. pombe* can also rely on UV Damage Endonuclease pathway, an alternative excision repair pathway of UV-damaged DNA (Yonemasu *et al.*, 1997). In order to evaluate how *cdc20*$^{EDM}$ strains cope with this type of DNA damage, spotted cells were irradiated with 200 J/m$^2$ UV-C and allowed to grow for two days (Figure 4.8). Compared to the control plates at 30°C, neither the parental strain, nor *cdc20*$^{WT\text{-}construct}$, or *cdc20*$^{exonull}$ show sensitivity to UV-C radiation. A similar phenotype is observed for the *cdc20*$^{M295I}$, *cdc20*$^{V412L}$, *cdc20*$^{L425V}$ and *cdc20*$^{Y459F}$ strains. However, slight sensitivity is observed for the *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ strains.

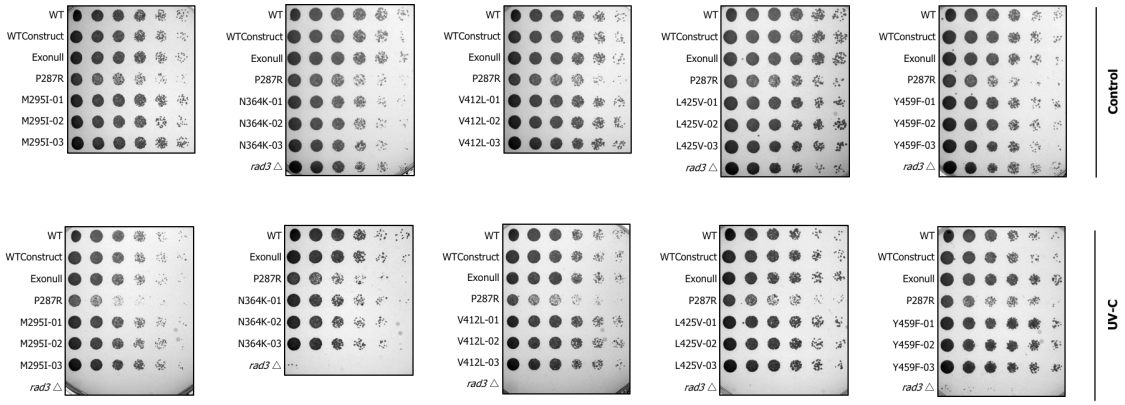



**Figure 4.8 Response of *cdc20*$^{EDM}$ strains to UV-C radiation**

Indicated strains were spotted onto rich YE3S media agar plates and either irradiated with UV at 200J/m$^2$ or not, and were incubated for two days at 30°C. One representative experiment out of 3.

### 4.2.3.4 *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ exhibit sensitivity to 4-NitroQuinoline-1-Oxide

4-NitroQuinoline-1-Oxide (4-NQO) mimics biological effects of UV light on DNA. After metabolic conversion to 4-hydroxyamoniquinolone 1 oxide, stable bulky adducts are formed on purines (Miller, 1970). Additional DNA damage may be generated through the production of reactive oxygen species. Compared to the drug-free plate, a clear sensitivity is present for the parental strain. A similar phenotype is observed for the *cdc20*$^{WT\text{-}construct}$, *cdc20*$^{exonull}$, *cdc20*$^{M295I}$, *cdc20*$^{V412L}$, *cdc20*$^{L425V}$ and *cdc20*$^{Y459F}$strains. However, *cdc20*$^{P287R}$ shows a clear sensitivity to 4-NQO compared to the wild-type strain. Finally, *cdc20*$^{N364K}$ also exhibits sensitivity to the drug compared to the parental strain, but this is milder compared to *cdc20*$^{P287R}$ (Figure 4.9).

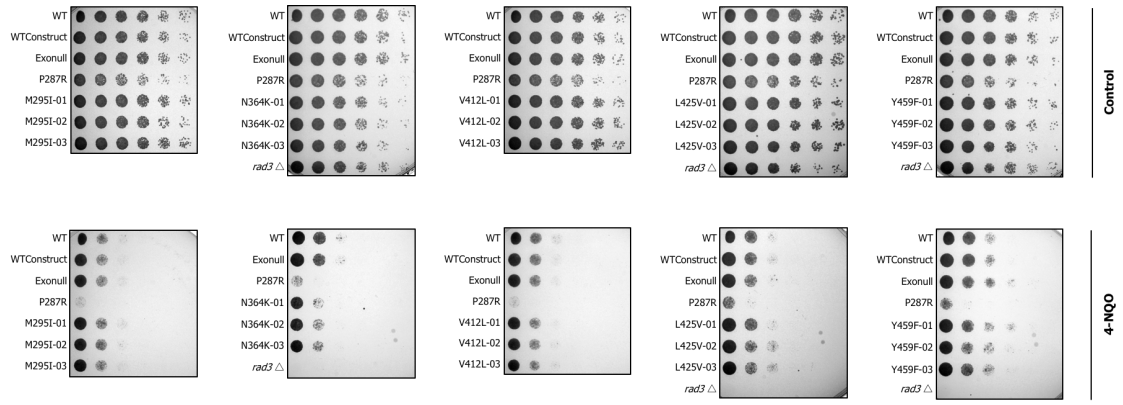



**Figure 4.9 Response of *cdc20*$^{EDM}$ strains to 4- NQO**

Indicated strains were spotted onto rich YE3S media agar plates with or without 0.5 µmol/L of 4-NQO, and were incubated for two days at 30°C. One representative experiment out of 3.

### 4.2.3.5 *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ exhibit sensitivity to Topoisomerase I blockage

Originally isolated from the bark and stem of the *Camptotheca acuminata* tree, camptothecin derivatives are now used in chemotherapies, e.g. irinotecan and topotecan. Camptothecin binds irreversibly to the complex formed between the topoisomerase I and DNA trapping the enzyme in a covalent linkage with the DNA. This complex will therefore form obstacles to the replication fork. Furthermore, given that DNA topoisomerase I affects the DNA topologic state during transcription and replication, inhibition of the enzyme might therefore impact transcription and replication, although deletion of the topoisomerase I gene is not lethal in fission yeast. No camptothecin sensitivity was shown for the *cdc20*$^{WT\text{-}construct}$, *cdc20*$^{exonull,}$ *cdc20*$^{M294I}$, *cdc20*$^{V412L}$, *cdc20*$^{L425V}$ or *cdc20*$^{Y459F}$ strains. However, clear sensitivity is displayed for the *cdc20*$^{N364K}$ strain, even if milder compared to *cdc20*$^{P287R}$ (Figure 4.10).

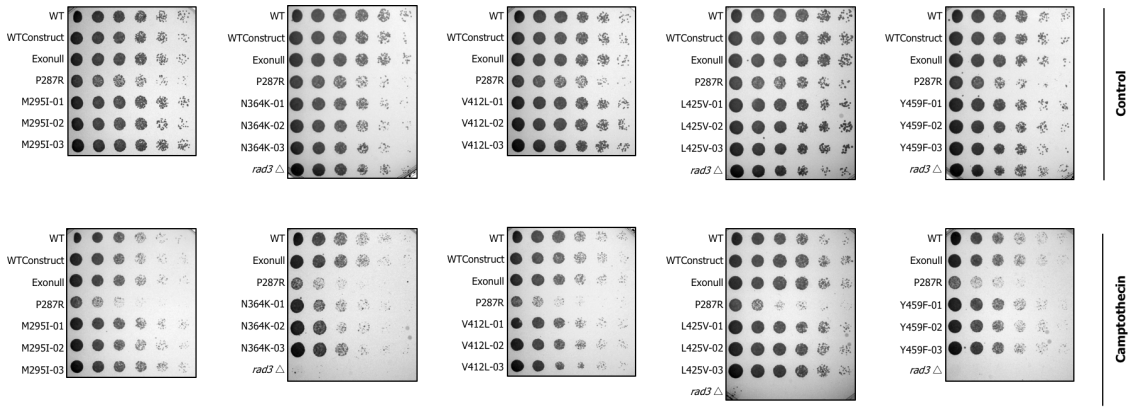



**Figure 4.10 Response of *cdc20*$^{EDM}$ strains to camptothecin**

Indicated strains were spotted onto rich YE3S media agar plates with or without 15 µml/L of camptothecin, and were incubated for two days at 30°C. One representative experiment out of 3.

### 4.2.3.6 *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ exhibit sensitivity to DNA methylation

Methyl Methane Sulfonate (MMS) is an alkylating agent, adding methyl groups to guanine, but also to adenine. Compared with the drug-free plate, all the strains grow less well in presence of MMS (Figure 4.11). However, *cdc20*$^{P287R}$ shows a striking sensitivity to DNA methylation compared to the parental strain. *cdc20*$^{N364K}$ strain maybe exhibit a very mild sensitivity.

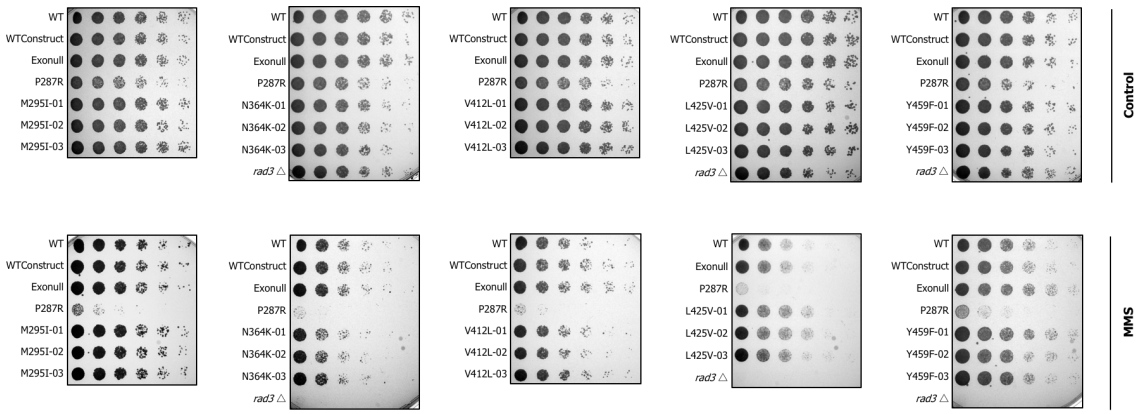



**Figure 4.11 Response of *cdc20*$^{EDM}$ strains to MMS**

Indicated strains were spotted onto rich YE3S media agar plates with or without 0.01 % of MMS, and were incubated for two days at 30°C. One representative experiment out of 3.

To summarise, the response to DNA-damaging agent differs amongst *cdc20*$^{EDM}$ variants. Compared to a wild-type Pol ε, lack of Pol ε exonuclease activity does not impair differently the response to DNA damage. Neither does the *cdc20*$^{M295I}$, *cdc20*$^{V412L}$, and *cdc20*$^{L425V}$. Nonetheless, the somatic variant *cdc20*$^{P287R}$ exhibits some striking impairment in DNA-damage response, like *cdc20*$^{N364K}$, although to a milder extend.

### 4.2.4 dNTP pools depletion

#### 4.2.4.1 *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ exhibit resistance to dNTP pools depletion

dNTP pools depletion leads to replication fork stalling (Pai *et al.*, 2017). In order to study the behaviour of *cdc20*$^{EDM}$ variants in this condition, spot assays were performed on rich media containing hydroxyurea (HU), which inhibits ribonucleoside diphosphate reductase, necessary for dNTP synthesis. Compared to growth on rich media alone, the

parental strain exhibits a clear growth defect when in contact with HU (Figure 4.12). Compared to the parental strain, the *cdc20*$^{\text{WT-construct}}$ and *cdc20*$^{\text{exonull}}$ strains do not exhibit an additional sensitivity to HU, similarly to *cdc20*$^{\text{M294I}}$, *cdc20*$^{\text{V412L}}$, *cdc20*$^{\text{L425V}}$ and *cdc20*$^{\text{Y459F}}$. However, *cdc20*$^{\text{P287R}}$ seems to show some HU resistance. Curiously, the variant *cdc20*$^{\text{N364K}}$ shows microcolonies on the spot patches, possibly reflecting suppressor mutations that facilitate growth on the HU plates.

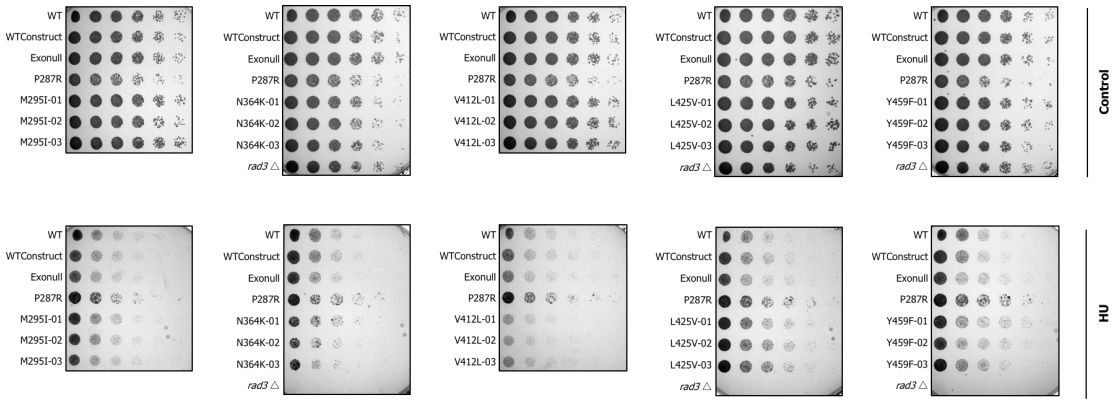



**Figure 4.12 Response of *cdc20*$^{\text{EDM}}$ strains to HU**

Indicated strains were spotted onto rich YE3S media agar plates with or without 7.5 mmol/L of HU and were incubated for two days at 30°C. One representative experiment out of 3.

### 4.2.4.2 Does dNTP pool alteration explain hypermutated phenotype and HU resistance of some *cdc20*$^{\text{EDM}}$ strains?

It has previously been reported that alteration in the polymerase domain of Pol ε lead to an overall increase in cellular dNTP pool in budding yeast (L. N. Williams *et al.*, 2015). A similar observation was made for Pol δ (Mertz *et al.*, 2015). These studies suggested that checkpoint-dependent extension of dNTP pools contributed to the increase in the

mutation rate, leading to synthetic lethality when dNTPs synthesis is not down-regulated (L. N. Williams *et al.*, 2015). Furthermore, dNTPs imbalance can also lead to an increase in mutation rate and can be synthetically enhanced with polymerase proofreading defects (Schmidt *et al.*, 2019). Hence, the hypothesis emerging from these studies is whether errors made by defective replicative polymerases could induce checkpoint-dependent expansion of dNTP pool, further increasing mutation rate. As observed in Figure 4.12, *cdc20*$^{P287R}$ resists to dNTP depletion. One hypothesis to explain this phenotype would be that the strains have an already increased dNTP pool level, possibly triggered by DNA damage checkpoint activation. Therefore, dNTP pool levels for the studied variant were determined by HPLC (Figure 4.13) (Figure 4.14) (Table 4.1). An RNR mutant (*cdc22*$^{D57N}$) was used as a positive control, as this strain was described with an alleviated allosteric feedback inhibition, leading to an overall increase in dNTP levels in fission (Fleck *et al.*, 2013) and budding yeast (A. Chabes *et al.*, 2003). As expected, the *cdc22*$^{D57N}$ strain exhibits an overall increase in dNTP levels as previously described (Fleck *et al.*, 2013). However, compared to the wild-type strain, none of the studied Pol ε variants exhibit an elevated dNTP pool similar to *cdc22*$^{D57N}$. Thus, resistance to HU observed in the *cdc20*$^{P287R}$ and *cdc20*$^{N64K}$ strains is not explained by an already elevated dNTP pools.

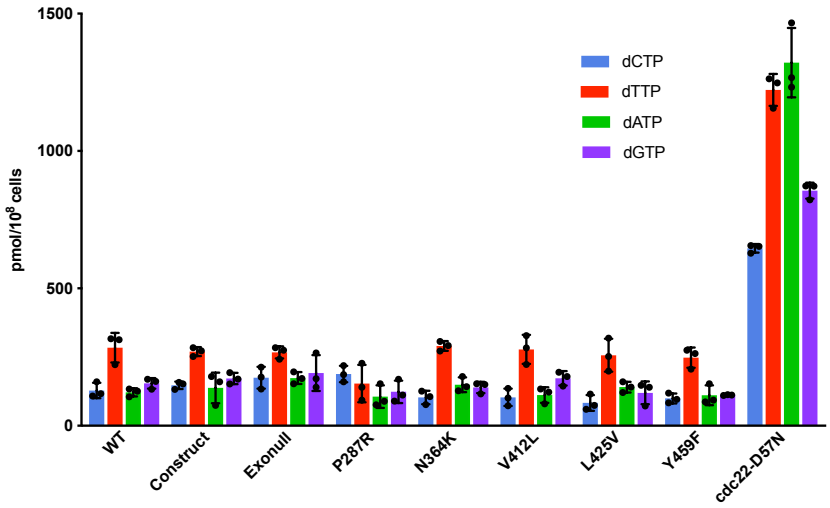



**Figure 4.13 Study of *cdc20*$^{EDM}$ variants dNTP levels**

dNTP concentration levels in *cdc20*$^{EDM}$ variants from exponential growth. Data show each value from 3 independent experiments and SD.

**Table 4.1 *cdc20*$^{EDM}$ variants dNTP levels**

| Strain | dCTP | dTTP | dATP | dGTP | Reference |
|---|---|---|---|---|---|
| WT | 1 | 1 | 1 | 1 | This study |
| WTConstruct | 1.2 | 1 | 1.2 | 1.1 | This study |
| Exonull | 1.4 | 1 | 1.4 | 1.3 | This study |
| P287R | 1.5 | 0.6 | 0.9 | 0.8 | This study |
| N364K | 0.8 | 1.1 | 1.2 | 0.9 | This study |
| V412L | 0.8 | 1 | 1 | 1.1 | This study |
| L425V | 0.7 | 0.9 | 1.2 | 0.8 | This study |
| Y459F | 0.8 | 0.9 | 1 | 0.7 | This study |
| *cdc22*$^{D57N}$ | 5.2 | 4.4 | 10.9 | 5.6 | This study |
| *cdc22*$^{D57N}$ | 5.8 | 10.3 | 12.1 | 6.2 | Fleck et. al. 2013 |
| *RNR1*$^{D57N}$ | 1.6 | 1.8 | 1.6 | 2 | Chabes et. al. 2003 |

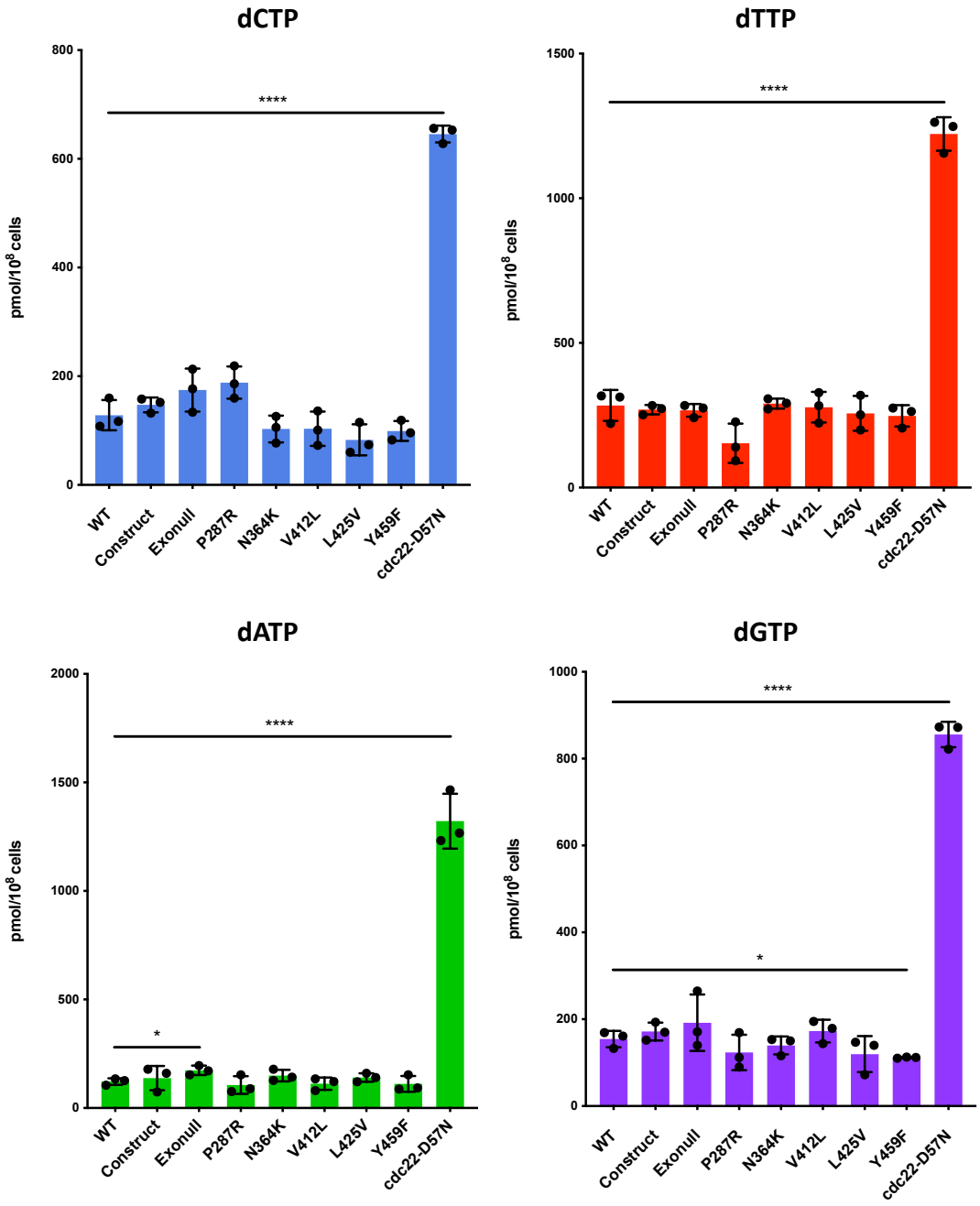



**Figure 4.14 Study of *cdc20*$^{EDM}$ variants dNTP levels – Single dNTPs**

dNTP concentration levels in *cdc20*$^{EDM}$ variants from exponential growth. Data show each value from 3 independent experiments and SD for each of the dNTPs. Analysis of values distribution using the Shapiro-Wilk normality test shows a normal distribution for all samples. Following this, a t-test was performed between wild-type and variant strains. * P<0.05; **** P<0.0001.

#### 4.2.4.3 Synthetic lethality between increased dNTP pools and hypermutated *cdc20*$^{EDM}$ variants

dNTP pool alteration can lead to an increase in mutation rate (Schmidt *et al.*, 2019). As observed in Figure 3.6, some *cdc20*$^{EDM}$ strains possess a hypermutated phenotype. However, every cell has a mutational threshold beyond which they become inviable. Thus, altered/increased dNTP levels combined with *cdc20*$^{EDM}$ hypermutated background could hypothetically lead to synthetic lethality, reaching the mutational threshold. To test this hypothesis, tetrad dissection of crosses between the *cdc20*$^{P287R}$ or *cdc20*$^{N364K}$ strains, with a *cdc22*$^{D57N}$ strain, were carried out, including *cdc20*$^{exonull}$ as a control (Figure 4.15). After growth of the spores for about 5 days, colonies were genotyped as followed: first, replica plating was performed on rich media containing G418 to identify *cdc20*$^{EDM}$ strains; second, identification of *cdc22*$^{D57N}$ variants was performed, only on G418$^{R}$ colonies. Tracking of the *cdc22*$^{D57N}$ variant started with colony PCR with the primers #989/#990 framing the *cdc22*$^{D57N}$ mutation, followed by digestion of the PCR product by DdeI restriction enzyme, as a silent mutation was incorporated alongside the *cdc22*$^{D57N}$ mutation (Fleck *et al.*, 2013). For *cdc20*$^{exonull}$, double mutants were clearly identified and the colony sizes suggest that they do not show any growth defect. However, for *cdc20*$^{P287R}$, no double mutants were detected through genotyping. Nevertheless, looking closely at colonies under the microscope, microcolonies were observed for possible double mutants. The same conclusion was drawn for *cdc20*$^{N364K}$.

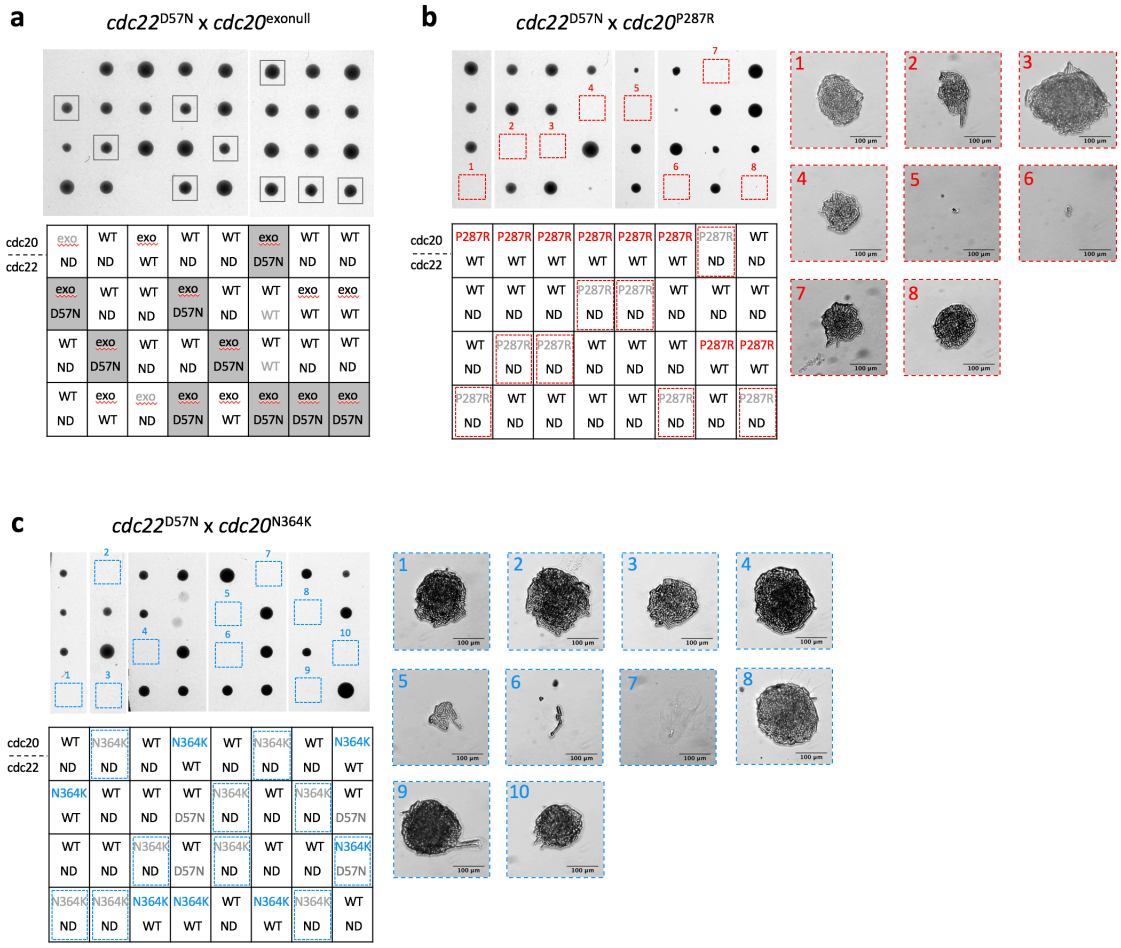



**Figure 4.15 *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ might be synthetic lethal with *cdc22*$^{D57N}$**

(**a**) Tetrad dissection of cross between *cdc22*$^{D57N}$ and *cdc20*$^{exonull}$. Double mutants are framed in grey. Genotypes indicated in grey are inferred while plain are confirmed. *cdc20* genotyping was performed via G418$^{R}$ analysis while *cdc22* genotyping was performed only on G418$^{R}$ colonies. (**b**) Tetrad dissection of cross between *cdc22*$^{D57N}$ and *cdc20*$^{P287R}$. Suspected double mutants are framed in red and colony pictures were taken. Genotypes indicated in grey are inferred while plain are confirmed. *cdc20* genotyping was performed via G418$^{R}$ analysis while *cdc22* genotyping was performed only on G418$^{R}$ colonies. (**c**) Tetrad dissection of cross between *cdc22*$^{D57N}$ and *cdc20*$^{N364K}$. Suspected double mutants are framed in blue and colony pictures were taken. Genotypes indicated in grey are inferred while plain are confirmed. *cdc20* genotyping was performed via G418$^{R}$ analysis while *cdc22* genotyping was performed only on G418$^{R}$ colonies.

### 4.2.5 Is Pol κ responsible for the *cdc20*$^{EDM}$ hypermutated phenotype?

TLS polymerases have a more 'open' polymerase active site and lack proofreading activity (see section 1.1.4.3), making them suited to bypass damaged templates. However, dysregulation of TLS in human cells leads to increased mutation rates that may contribute to tumourigenesis (Hedglin *et al.*, 2017). Pol κ is the most accurate of the Y-family polymerases on undamaged DNA (Johnson *et al.*, 2000) and can bypass many lesions, albeit with relatively low efficiency. It is not conserved in *S. cerevisiae*, although it is in *S. pombe*.

Previous work from our group discovered that deletion of *kpa1* (encoding Pol κ in fission yeast) in a *cdc20*$^{P287R}$ background leads to a drastic decrease in the mutation rate, which is reduced to near *cdc20*$^{exonull}$ levels (I. Soriano, unpublished results). In order to test whether the same phenotype is observed for the hypermutated germline variant *cdc20*$^{N364K}$, double mutant strains were obtained via cross and tetrad dissection with a *kpa1Δ* strain (Chen *et al.*, 2014). Tracking of *cdc20*$^{EDM}$ genotype was done via colony PCR (primers #1243 and #1244) followed by enzymatic digestion (Table 3.1), while tracking of *kpa1Δ* genotype was done via colony PCR only (primers #1380 and #1381). Canavanine fluctuation assay was then performed on at least three of the double mutants. The *cdc20*$^{exonull}$ double mutant was use as a control. As observed on Figure 4.16, the mutation rate reported relative to the parental strain shows an even lower mutation rate for the *kpa1Δ* strain than for the wild-type strain. As previously observed (Figure 3.6), lack of Cdc20 exonuclease activity leads to an increase in mutation rate, even greater for the *cdc20*$^{N364K}$ strain. Deletion of *kpa1* in a *cdc20*$^{exonull}$ background leads to a decrease in mutation rate to the level of the *kpa1Δ* single mutant. In a *cdc20*$^{N364K}$

background, *kpa1* depletion leads to a decrease in mutation rate similar to the one observed for *cdc20*$^{exonull}$ for at least two out of three clones. This analysis shows that deletion of *kpa1* does reduce the mutation rate compared to the single *cdc20*$^{N364K}$ mutant, further implying that lack of proofreading is not the only factor responsible for the hypermutating phenotype.

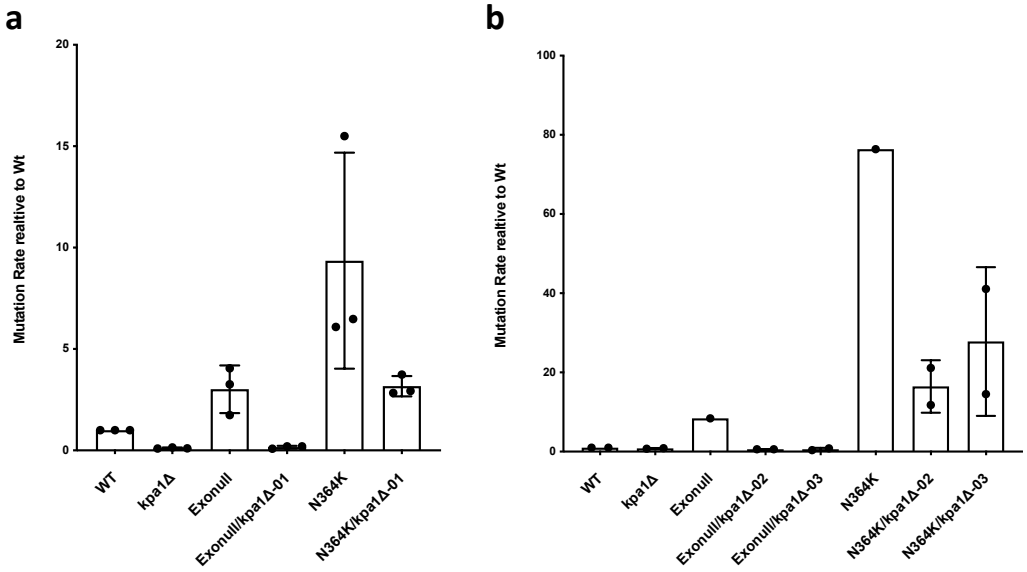



**Figure 4.16 Pol κ is partially responsible for the increase in mutation rate in *cdc20*$^{N364K}$ variant**

Mutation rates were determined by fluctuation analysis, using canavanine resistance. Values represent the mutation rate per $10^7$ cells determined by the Ma-Sandri-Sarkar Maximum Likelihood Estimator using rSalvador (Zheng, 2017) with SD. The figures (**a**) and (**b**) are distinct in order to facilitate the interpretation due to the difference of scale. (**a**) Experiment was performed three times for each sample. Analysis of values distribution using the Shapiro-Wilk normality test shows a normal distribution for all samples, except for *cdc20*$^{exonull}$ *kpa1Δ-01* strain. Following this, a t-test was performed between *cdc20*$^{N364K}$ and *cdc20*$^{N364K}$ *kpa1Δ-01*. No significantly difference was observed. (**b**) Experiment was performed two times for each sample, except *cdc20*$^{exonull}$ and *cdc20*$^{N364K}$. No statistical tests were performed on this sample as the sampling is too small.

## 4.3 Discussion

*POLE*$^{EDMs}$ can impede 3'-5' exonuclease activity (see section 1.3). Notably, the exonuclease activity is modulated depending on the amino acid mutated (Shinbrot *et al.*, 2014), but also varies within organisms of origin, and whether this activity was assessed from the catalytic subunit alone or from the full holoenzyme of Pol ε (Shinbrot *et al.*, 2014; Xing *et al.*, 2019). In Xing *et al.* (2019), Pol2-P301R exhibits residual exonuclease activity while in Shinbrot *et al.* (2014), POLE-P286R does not. POLE-V411L and POLE-L424V displayed reduced exonuclease activity compared to wild-type, illustrating the discrepancy of exonuclease activity amongst Pol ε-EDMs (Shinbrot *et al.*, 2014). Using Pol ε human holoenzyme, Pol ε-M294I shows exonuclease activity levels similar to a wild-type protein, while Pol ε-L424V exhibits a reduced, but does not eliminated exonuclease activity (Figure 4.2). Pol ε-N363K and Pol ε-V411L variants exhibit a complete lack of exonuclease activity, similar to the exonuclease-dead version (Figure 4.2). Contrary to Xing *et al.* (2019), residual exonuclease activity for Pol ε-P286R was not detected in our assay (Figure 4.2), even with extended incubation. The discrepancy between results could be explained by several factors. First, previously reported assays used the catalytic subunit of either budding yeast or human Pol ε, rather than the holoenzyme as used in our study. The presence of other Pol ε subunits could potentially modulate the 3'-5' exonuclease activity in the variants as depletion of either of the secondary subunits leads to a less stable holoenzyme (Bellelli *et al.*, 2018; Dua *et al.*, 2000; Ohya *et al.*, 2000). Second, even if the studied EDMs are highly conserved between human and yeast, the overall homology between the human catalytic subunit of Pol ε and *S. cerevisiae* and *S. pombe* is of 43 % and 47 %, respectively, i.e. yeasts may

not be a perfect model for analysis the human variants. Therefore, Pol ε might have diverged from common ancestor between mammal cells and yeasts, resulting in different function and role due to possible difference in structure. Finally, the *in vitro* assays presented in 4.2.1 retain some limitations such as the normalisation of the enzyme quantity based on the polymerase activity, as the Pol2-P301R variant exhibits an increase in polymerase activity (Xing *et al.*, 2019). Furthermore, the absence of quantification and the variability of enzymatic concentrations used render the interpretation difficult. Altogether, Hence, it might be worth to repeat the experiments with normalised enzymatic concentration amongst Pol ε-EDMs.

Some of the hypermutated *cdc20*$^{EDM}$ variants show sign of growth defect and possible checkpoint activation, possibly due to their spontaneous increased mutation rate (Figure 3.7 and Figure 3.9). *S. pombe* relies on Rad3 for DNA damage recognition and checkpoint activation (Humphrey, 2000). As the *cdc20*$^{\Delta N\text{-ter}}$ strain exhibits a delayed cell cycle and required the presence DNA checkpoint control Rad3 for viability (Feng *et al.*, 2001), we tested whether the hypermutated *cdc20*$^{EDM}$ strains relied on Rad3 for viability. *rad3Δ* led to synthetic sickness for *cdc20*$^{P287R}$ and only had a subtle impact on *cdc20*$^{N364K}$ double mutants. However, *rad3ts* double mutant showed a clear synthetic lethality for *cdc20*$^{P287R}$, while *cdc20*$^{N364K}$ exhibited no synthetic lethality, nor sickness. The difference observed between the *rad3Δ* and *rad3ts* contexts could be explained by the mutagenic ability of *cdc20*$^{P287R}$ and the experimental time frame. Indeed, the construction of *cdc20*$^{EDM}$ *rad3Δ* double mutant allowed sufficient time for acquisition of suppressor mutations compared to the *rad3ts* double mutant where the experimental time frame was more restrained. These results are consistent with the elongated cell length (Figure 3.7) as well as reduced growth rate previously observed for these two

mutagenic strains (Figure 3.9). As Rad3 presence is required, it is presumably activated. Looking at the activation of Rad3 downstream effectors either via assessment of their phosphorylation status or genetic interaction would point towards either defects during S-phase (Cds1) or mitotic defects (Chk1) (Humphrey, 2000). DNA replication checkpoint would be activated following DNA replication errors due to malfunctioning Pol ε and the increase in mismatches or stalled replication forks. Examine whether ssDNA presence would increase due to stalled fork via RPA foci tracking would enable us to test this hypothesis (Y. Zou *et al.*, 2006).

In budding yeast, one of the downstream effectors of DNA damage or replication blockage is the Dun1 effector kinase (Z. Zhou *et al.*, 1993). One of the functions of Dun1 is the modulation of dNTP levels via negative regulation of RNR (Northam *et al.*, 2006). Mutator phenotype observed for proofreading-deficient Pol δ was partially Dun1-dependent which activation triggered an increased level of dNTPs (Mertz *et al.*, 2015). Similar conclusion was observed for *pol2-4* mutator cells, for which *dun1Δ* restores the mutation rate to normal levels (L. N. Williams *et al.*, 2015). The hypermutated strains *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ are resistant (at different levels) to hydroxyurea (Figure 4.12), possibly reflecting the emergence of suppressor mutations via mutagenic phenotype. Higher dNTP levels would explain the resistance phenotype and correlates with *pol2-M644G* strain which exhibits an increased level of dNTPs (L. N. Williams *et al.*, 2015). However, none of the studied strains show any dramatic modulation on dNTP levels (Figure 4.13), which confirms previous observations for *pol2-4* (L. N. Williams *et al.*, 2015) and *pol2-P301R* (Xing *et al.*, 2019) strains. The discrepancy of results could be related to the type of Pol ε alteration as either they affect the exonuclease (*pol2-4* and *cdc20*$^{EDMs}$ strains) or the polymerase base selectivity function (*pol2-M644G* strain).

The *pol2-M644G* strain is synthetically lethal with an increase in dNTPs, probably due to a cumulative effect of mutagenicity which leads to a lethal threshold (Kohnken *et al.*, 2015; L. N. Williams *et al.*, 2015). Even if none of the *cdc20*$^{EDM}$ strains exhibit an increased dNTP pool compared to wild-type, we investigated whether a combination of an increase in dNTP pool would be deleterious when combined with a *cdc20*$^{EDM}$ hypermutated phenotype. No viable double mutant for *cdc20*$^{P287R}$ nor *cdc20*$^{N364K}$ combined with *cdc22*$^{D57N}$ were obtained from tetrad dissection, presumably as the double mutant has a mutation rate which leads to error-induced extinction. These findings suggest a potential line of treatment which could specifically target ultramutated tumours carrying *POLE*$^{EDM}$ by increasing dNTP levels.

Similarly to *cdc20*$^{exonull}$ (Figure 4.12), cells lacking the N-ter of Cdc20 do not show sensitivity nor resistance to hydroxyurea (Feng *et al.*, 2001). Nevertheless, *cdc20*$^{\Delta N\text{-ter}}$ are sensitive to MMS and UV-C suggesting deficiency in DNA repair (Feng *et al.*, 2001). Similarly, *cdc20*$^{EDM}$ strains were exposed to different DNA-damaging agents to address whether EDMs affect DNA replication or DNA repair. The behavioural response differs amongst *cdc20*$^{EDM}$ variant (see section 4.2.3). On one hand, *cdc20*$^{exonull}$, as well as *cdc20*$^{M295I}$, *cdc20*$^{V412L}$, *cdc20*$^{L425V}$ and *cdc20*$^{Y459F}$ are not sensitive to DNA-damaging agents, including UV irradiation, MMS, camptothecin, bleocin, and 4-NQO, and respond similarly to a wild-type strain. On the other hand, *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ (to a milder extend) show sensitivity to the aforementioned components. These results suggest a DNA repair deficiency for some particular *cdc20*$^{EDMs}$ and emphasises the potential role of Pol ε in DNA repair (Pursell *et al.*, 2008). Furthermore, the observed sensitivity could be explained by the Pol ε-EDM difficulties in by-passing difficult to replicate stretch of

DNA. Although, this does not correlate with the fact that Pol2-P301R can readily by-pass hairpin structure and extend from mismatches in a cell-free context (Xing *et al.*, 2019).

Translesion synthesis participate in damage tolerance and enable the replisome to cope with difficult-to-replicate templates (Sale, 2013). For example, altered Pol $\alpha$, Pol $\delta$ and Pol $\varepsilon$ see their spontaneous mutation rate decrease when the TLS polymerase Pol $\xi$ is depleted (Northam *et al.*, 2006). Although TLS polymerases were not accounted responsible for the ultramutated phenotype observed in Pol2-P301R (Xing *et al.*, 2019), previous work from our group noticed that depletion of the TLS polymerase Pol $\kappa$ – not conserved in budding yeast – drastically lower *cdc20*$^{P287R}$ mutation rate back to a *cdc20*$^{exonull}$ strain level (I. Soriano, unpublished results). The same principle was applied to the *cdc20*$^{exonull}$ and *cdc20*$^{N364K}$ strain, showing similar results (Figure 4.16). Notably, the *cdc20*$^{exonull}$ *kpa1Δ* strain sees its mutation rate lowering back to a wild-type, or even below wild-type level. This suggests that Pol $\kappa$ might be the polymerase called for rescue when Pol $\varepsilon$ is malfunctioning, and the responsible for the increased mutation rate observed in a *cdc20*$^{EDM}$ context. Similar observations were published about altered Pol $\alpha$ in fission yeast: the increased mutation rate triggered checkpoint activation, further increasing the expression of Pol $\kappa$ which later was identified as the responsible for the elevated mutation burden (Kai *et al.*, 2003). Thus, it would be interesting to look at the expression of Pol $\kappa$ in a *cdc20*$^{EDM}$ context and see if there is any induced expression. Furthermore, recruitment of the Pol $\kappa$ could be looked at via assessment of PCNA ubiquitylation status or the contribution of Pol $\kappa$ noncanonical PIP-box motif (Ogi *et al.*, 2010). Direct recruitment at the replication fork could also be looked at via immunofluorescent, pull-down, or iPOND (isolation of proteins on nascent DNA).

Altogether, this chapter gives insight into the mechanism of hypermutation observed for some *POLE*$^{EDM}$ variants, and potential lines of exploration for *POLE*$^{EDM}$ targeted treatment.

# Chapter 5

## Deep-Sequencing Analysis of Mutation Spectrum Caused by Novel Cancer Related Germline *POLE*$^{EDM}$ variants in Fission Yeast

### 5.1 Introduction

Constantly subjected to damage, the genome bears the scars of mutagenic mechanisms. Recent advances in sequencing technologies combined with formidable collective efforts enabled the extraction of mutational signatures from catalogues of mutations obtained from multiple cancer types (Alexandrov *et al.*, 2020; Alexandrov, Nik-Zainal, Wedge, Aparicio*, et al.*, 2013). Some of these mutational signatures are associated with one or several mutagenic pathways. For instance, the signature 10 was initially associated with *POLE*$^{P286R}$ and *POLE*$^{V411L}$ alterations, while the signatures 6, 15, 20 and 26 where related to MMR-deficiency (Alexandrov, Nik-Zainal, Wedge, Aparicio*, et al.*, 2013). The increase in data scale allowed the refinement of some of the mutational signatures (Alexandrov *et al.*, 2020). For instance, the signature 10 was subcategorised into two signatures – SBS10a and SBS10b. In addition, signatures associated with MMR-deficiency were refined, while signature 21 of previously unknown aetiology was newly associated with MMR deficiency. Notably, signature 14 which was of unknown aetiology (Alexandrov, Nik-Zainal, Wedge, Aparicio*, et al.*, 2013), was refined and is now associated with concurrent *POLE*$^{EDM}$ and MMR deficiency (Alexandrov *et al.*, 2020; Haradhvala *et al.*, 2018).

Similar studies focused on the extraction of mutational signatures present in tumours carrying either *POLE*$^{EDM}$, MMR deficiency, or both alterations (Haradhvala *et al.*, 2018; Hodel *et al.*, 2020; Shlien *et al.*, 2015). Prior to the publication of Alexandrov *et al.* (2020), these studies identified a signature similar to the signature 14, in samples carrying both *POLE*$^{EDM}$ and MMR deficiency, which does not represent a simple combination of the Pol ε and MMR signatures. In addition, looking deeper into the mutational signatures of specific *POLE*$^{EDMs}$, distinct signatures (although quite similar) within *POLE*$^{EDMs}$ were identified, possibly indicating a distinct mutational mechanism for each *POLE*$^{EDM}$ (Haradhvala *et al.*, 2018; Hodel *et al.*, 2020).

Construction of germline *POLE*$^{EDMs}$ in fission yeast showed that, similar to a certain extent to the *cdc20*$^{P287R}$ variant, the *cdc20*$^{N364K}$ and *cdc20*$^{Y459F}$ strains exhibit a mutation rate greater than that of *cdc20*$^{exonull}$ (Figure 3.6). The signature 10 being associated with *POLE*$^{EDMs}$ variants (Alexandrov *et al.*, 2020; Alexandrov, Nik-Zainal, Wedge, Aparicio*, et al.*, 2013) and observed for the *POLE*$^{P286R}$ and *POLE*$^{V411L}$ in particular (Alexandrov, Nik-Zainal, Wedge, Aparicio*, et al.*, 2013; Haradhvala *et al.*, 2018; Hodel *et al.*, 2020), the question whether a similar mutational signature would be observed for the germline *POLE*$^{N363K}$ and *POLE*$^{Y458F}$ has to be assessed to examine the mutagenesis process of these variants.

To explore this question, a mutation accumulation (MA) experiment was set up, passaging cells in a bottleneck fashion, up to approximately 200 generations to generate a catalogue of mutations prior to mutational pattern analysis. A MMR deficient strain (*msh2Δ*) was also included in order to study the impact of MMR impairment in a *cdc20*$^{EDM}$ context, and determine if the signature 14 is detected in the double mutants. In addition to the initial starting point (generation 0) and the final one (generation 200),

intra-generational time points were also processed for whole genome sequencing (WGS). The aim was to assess the pace of mutation appearance amongst *cdc20*$^{EDMs}$, bearing in mind the fact that depletion of MMR might severely increase the mutation rate up to a lethal threshold.

Altogether, this chapter aims to explore the mutagenesis of *POLE*$^{EDMs}$ variants in *S. pombe*.

## 5.2 Results

### 5.2.1 *cdc20*$^{EDMs}$ and MMR deficiency

Inactivation of Pol ε exonuclease activity combined with MMR was identified as embryonically lethal in mice (Albertson *et al.*, 2009), although human tumours were identified with both alterations (Haradhvala *et al.*, 2018; Shlien *et al.*, 2015). Notably, *pol2-4* strain (*cdc20*$^{exonull}$ equivalent) did not exhibit any synthetic lethality with *msh6Δ* while *pol2-P301R* (*cdc20*$^{P287R}$ equivalent) did (Xing *et al.*, 2019). In MMR pathway, Msh2 is one of the most upstream factors as being the common factor of MutS, signalling mismatches. Therefore, its deletion would abrogate the MMR, not correcting anymore the replications errors inserted by Pol ε. The strains *cdc20*$^{exonull}$, *cdc20*$^{N364K}$ and *cdc20*$^{Y459F}$ were crossed with an *msh2Δ* strain prior to tetrad dissection. Genotyping for *cdc20*$^{EDM}$ was deduced from G418 resistance, while *msh2Δ* genotyping was performed via colony PCR using the primers #1369 and #1173. As observed in Figure 5.1, *cdc20*$^{exonull}$ *msh2Δ* double mutants are viable, similarly to *cdc20*$^{Y459F}$. Notably, *cdc20*$^{N364K}$ *msh2Δ* double mutants exhibit smaller colony sizes, possibly reflecting synthetic sickness. Taken together, this suggests that MMR is dispensable, at least on a short-term basis, in a *cdc20*$^{EDM}$ context, although potential sickness is noticed for the variants with the highest mutated phenotype.

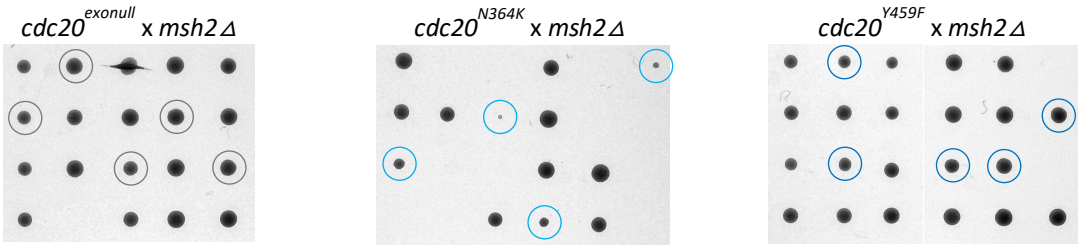


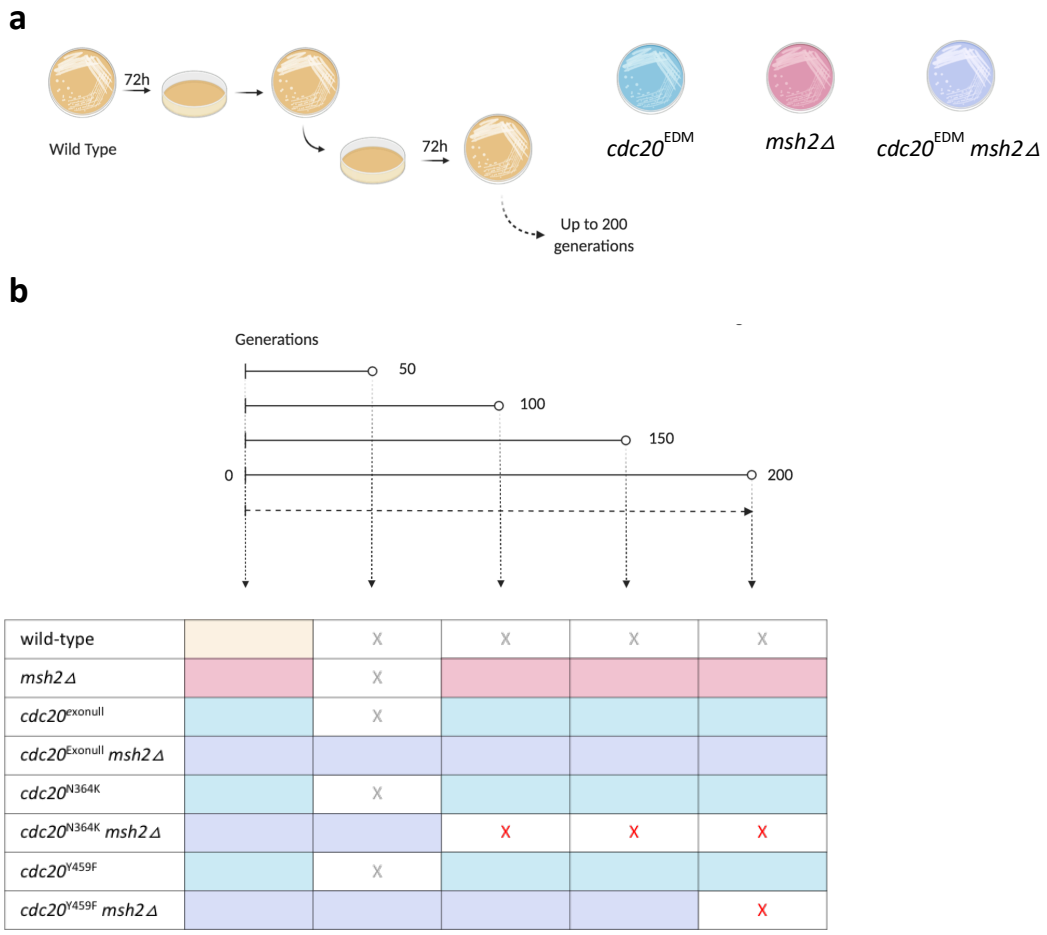

**Figure 5.1 *cdc20*$^{EDM}$ combined with *msh2Δ***

Tetrad analysis of *msh2Δ* and either *cdc20*$^{exonull}$, *cdc20*$^{N364K}$, or *cdc20*$^{Y459F}$. Double mutants are circled in grey, light blue and dark blue, respectively.

### 5.2.2 Mutation Accumulation experiment

Mutation accumulation experiments were carried out in triplicate as described in section 2.1.12. Briefly, for the parental strain, as well as for each of the single mutant *cdc20*$^{exonull}$, *cdc20*$^{N364K}$, *cdc20*$^{Y459F}$, and *msh2Δ* strains, single colonies were re-streaked every 3 days up to 200 generations. For the generations 0, 100, 150 and 200, half of the colony selected was used to carry on the MA, while the other half was stored prior to genomic DNA extraction for WGS. The same process was performed for the double mutants, with an additional time point at generation 50. The *cdc20*$^{exonull}$ *msh2Δ* strain could be passaged up to 200 generations, while only 2 out of 3 of the replicates for the *cdc20*$^{N364K}$ *msh2Δ* strains reached 50 generations. No viable cells grew past that point. Similarly but less dramatically, 2 out of 3 *cdc20*$^{Y459F}$ *msh2Δ* strains went up to 150 generations. No viable cells were observed past this point (Figure 5.2). In total, 85 strains were generated, including: 5 single mutants (wild-type, *msh2Δ*, *cdc20*$^{exonull}$, *cdc20*$^{N364K}$, *cdc20*$^{Y459F}$), 3 double mutants (*cdc20*$^{exonull}$ *msh2Δ*, *cdc20*$^{N364K}$ *msh2Δ*, *cdc20*$^{Y459F}$ *msh2Δ*), the different generational time points, and finally 3 lineages (replicates) for each genotype. Genomic DNAs were extracted prior to sending to the Oxford Genomic Centre (Wellcome Centre for Human Genetics) for library generation and sequencing on a

NovaSeq6000 (Illumina) in a 150-paired-end fashion. The Bioinformatics core performed the alignment of the reads on the fission yeast genome (ASM294v2) (see section 2.6) and provided the FASTQ and BAM files for further analysis.




| | | | | | |
|---|---|---|---|---|---|
| wild-type | | X | X | X | X |
| *msh2Δ* | | X | | | |
| *cdc20*$^{exonull}$ | | X | | | |
| *cdc20*$^{Exonull}$ *msh2Δ* | | | | | |
| *cdc20*$^{N364K}$ | | X | | | |
| *cdc20*$^{N364K}$ *msh2Δ* | | | X | X | X |
| *cdc20*$^{Y459F}$ | | X | | | |
| *cdc20*$^{Y459F}$ *msh2Δ* | | | | | X |

**Figure 5.2 Mutation accumulation experiment**

(**a**) Strains were woken up, streaked on rich media plate and allowed growth for 48 hours. A single colony was then picked and streaked on a new rich media plate and allowed growth for 72 hours. This process was repeated until approximately 200 generations were reached. For each strain, when starting the MA experiment, 3 lineages were created from 3 single colonies from a single plate carrying cells woken-up form storage at -80°C. (**b**) Generational time points were stored for further characterisation. Wild-type strain time points were for generation 0 and 200. Single mutant time points were for generation 0, 100, 150 and 200. Double mutant time points were for generation 0, 50, 100, 150 and 200. However, only *cdc20*$^{exonull}$ *msh2Δ* reached that point. *cdc20*$^{N364K}$ *msh2Δ* could not grow beyond 50 generations (only 2 out of the 3 lineages could grow up to 50 generations) and *cdc20*$^{Y459F}$ *msh2Δ* could not reach 200 generations (only 2 out of 3 lineages reached 150 generations).

### 5.2.3 Assessment of sequencing quality

Sequencing analysis was initiated by looking at the quality of the sequencing data. Indeed, bias could emerge from library preparation or sequencing issues. Quality of the sequencing can be evaluated via the FASTQ files, where the FASTQ format is a text-based format storing both sequences and the corresponding quality score. Quality control can be done using the FastQC application which has been developed by the Babraham Institute (Andrews, 2010). Briefly, the software takes the FASTQ files and runs a series of tests on them, and generates a comprehensive Quality Control (QC) report including several modules allowing investigation on different sequence quality aspects. The Quality Control modules are produced on a per-sample basis. Working with a rather big sample batch, instead of going through each of the samples and checking each of the modules (which would be time consuming and repetitive, increasing the risks of missing a warning) an alternative option used here is MultiQC (Ewels *et al.*, 2016). This tool was designed to parse and compile summary statistics and log files generated by other bioinformatics tools – such as FastQC – creating a concise visualisation, enabling a quality check in one glance.

#### 5.2.3.1 Per Base Sequence Quality

Derived from Sanger sequencing, the Phred score (or Q-score) indicates the probability that a given base is called incorrectly by the sequencer (Richterich, 1998). The Q-score is related to the following formulae: $Q = -10 \log^{10} P$, which can be translated as follows: for example, if the Phred is assigned a Q-score of 30 (or Q30) to a base, the probability P to call the wrong one is of 1 in 1000, equivalent of a 99.9 % of base call accuracy. Each

base sequenced within a read is allocated a quality value (Q-score). The Per Base Sequence Quality graph indicates the quality of the raw reads for each position within reads. It is usual to see a decrease in the sequencing quality of towards the 3' end of the reads. Indeed, in Illumina technology, sequencing takes place in cycles where each of the four nucleotides is coupled to a fluorescent signal. The fluorescence is detected at each cycle, intensified by the thousands of amplified templates within a cluster. In addition, each nucleotide is also coupled with a blocker, therefore enabling the synthesis of only one nucleotide at a time, thus supposedly the detection of a unique fluorescence signal at every cycle. After fluorescence detection, the blocker is eliminated, allowing further DNA synthesis. However, phasing – or the inefficient blocker removal – amplifies over time. This leads to the desynchronisation of the fluorescent signal within a cluster, degrading the signal detected by the camera, ultimately translated by a decrease in the read quality. Usually the quality score is divided into 3 categories: very good (>28), good (between 20 and 28) and calls of poor quality (<20). In this set of data, the Per Base Sequence Quality shows no flaw: it stays above 28 along the 151 length long sequences, for all samples (Figure 5.3).

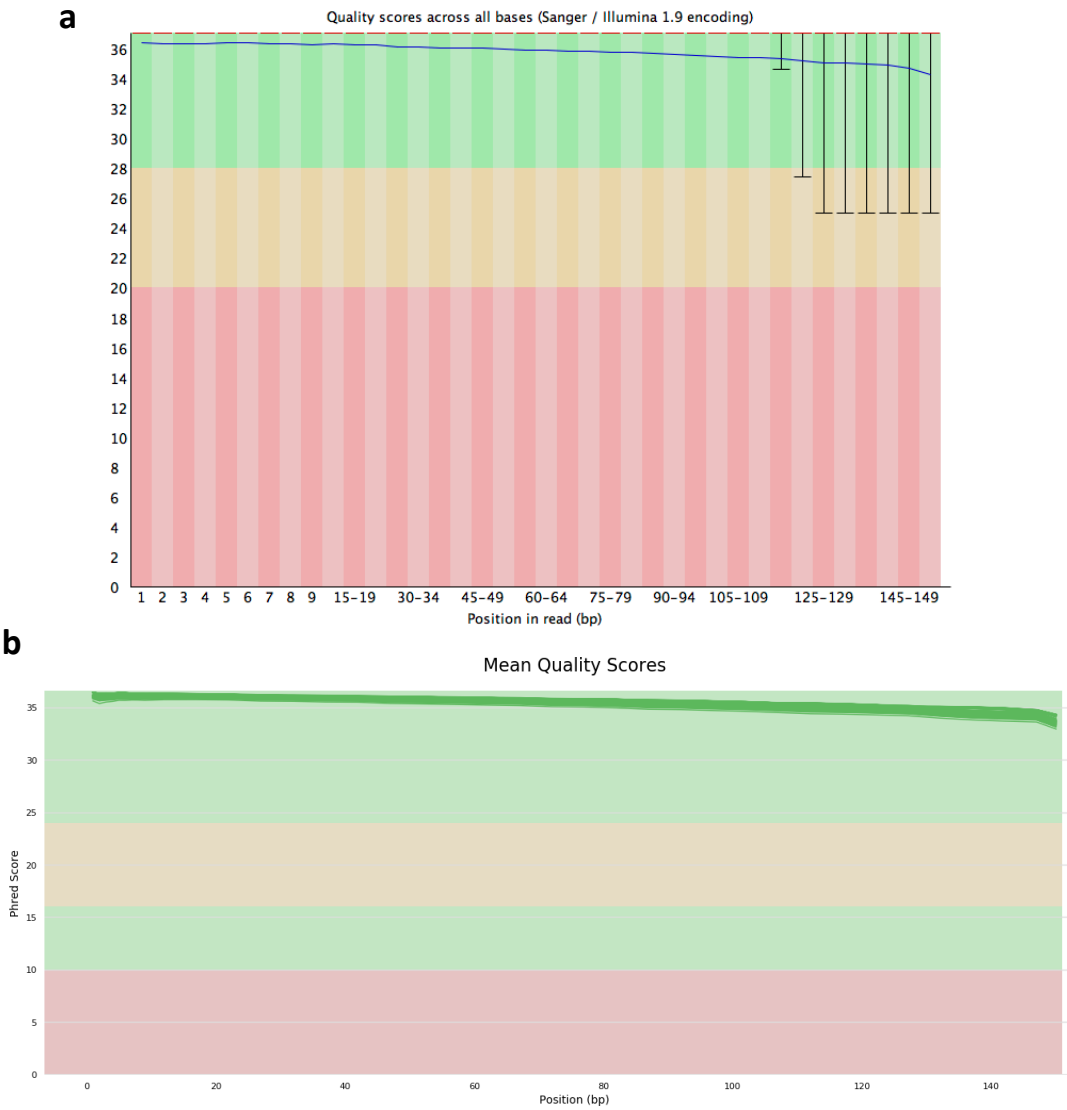



**Figure 5.3 Sequencing quality check – Per base sequence quality**

Modules used for sequencing quality check were generated using FastQC. Once generated for every samples, the modules were compiled using MultiQC. For clarity, each module is represented for a single sample (here wild-type generation 0 duplicate 1) (**a**) and the MultiQC output for the full set of samples (**b**).

### 5.2.3.2 Per Sequence Quality Score

A low Q-score (below Q20) can lead to an increase in false-positive variant calls, whereas when the Q-score reaches Q30, virtually all of the reads will be error-free. In order to see if a subset of the reads has low quality values, the Per Sequence Quality Score graph

represents the quality score distribution over all sequences reported. As observed in Figure 5.4, the average quality per read peaks at 36, translated an overall very high base call accuracy.

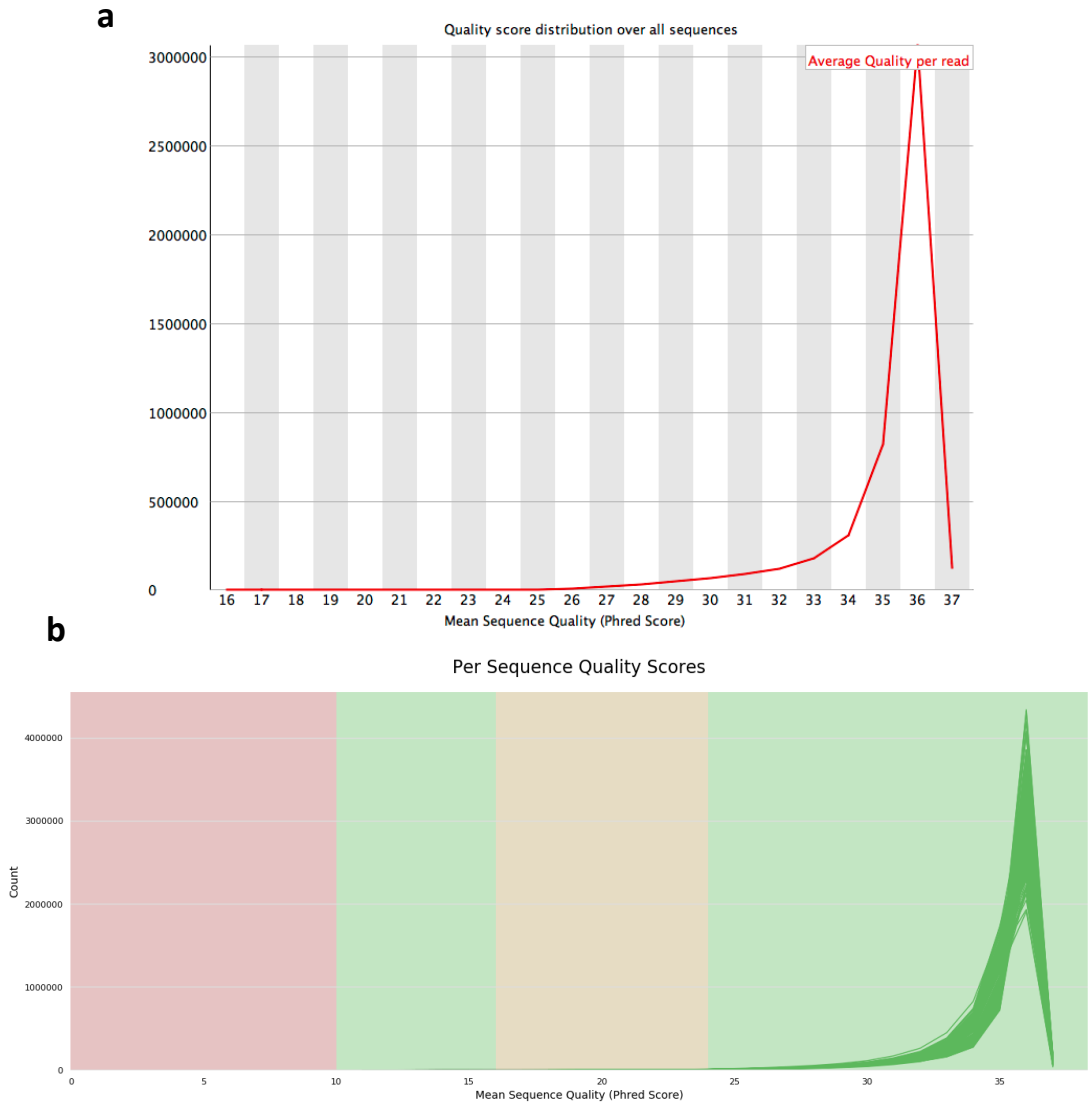



**Figure 5.4 Sequencing quality check – Per sequence quality sequence**

Modules used for sequencing quality check were generated using FastQC. Once generated for every samples, the modules were compiled using MultiQC. For clarity, each module is represented for a single sample (here wild-type generation 0 duplicate 1) (**a**) and the MultiQC output for the full set of samples (**b**).

#### 5.2.3.3 Per Base Sequence Content

This plot reports the proportion of each of the four bases at a given position within all reads of a FASTQ file. One should expect a parallel distribution of all four bases across the reads as a biased distribution could be related to a biased sequence distribution such as adapters not trimmed out, or in RNA sequencing data. The FastQC graph shows an overall straight line for each of the four bases, translating no bias in the sequence distribution. The plot provided by MultiQC is a stack of each of the 85 samples, one might find difficult to interpret. Looked closely, strain by strain, no discrepancy was observed (Figure 5.5).

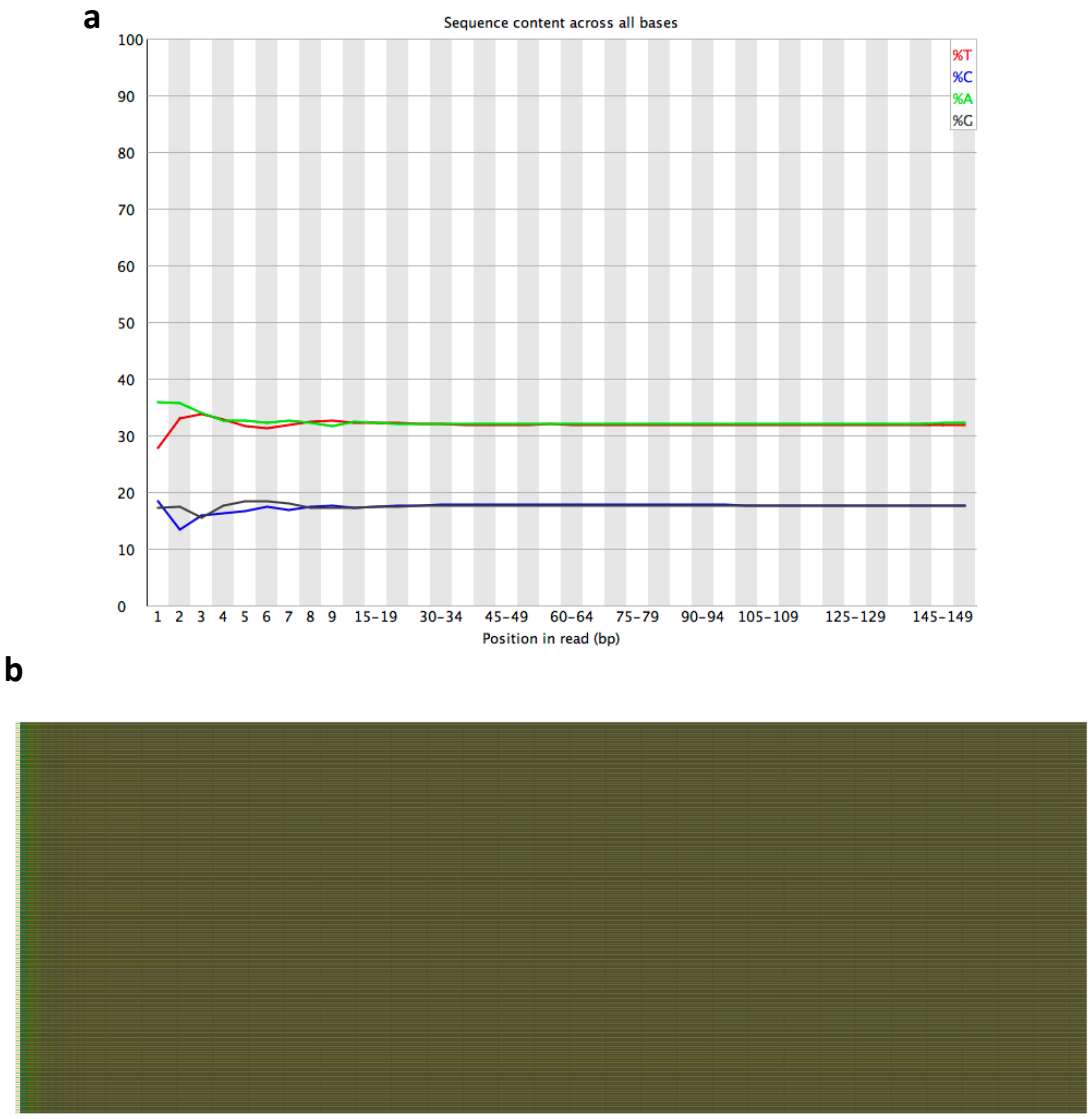



**Figure 5.5 Sequencing quality check – Per base sequence content**

Modules used for sequencing quality check were generated using FastQC. Once generated for every samples, the modules were compiled using MultiQC. For clarity, each module is represented for a single sample (here wild-type generation 0 duplicate 1) (**a**) and the MultiQC output for the full set of samples (**b**).

#### 5.2.3.4 Per Sequence GC Content

This graph plots the GC content across the whole length of each sequence present in a FASTQ file. The distribution is expected to follow a normal distribution, with the central peak at the overall GC content of the underlying genome in the case of a WGS. The GC

content in fission yeast is expected to be around 36 % as described in Behringer *et al.* (2015). The MultiQC analysis confirm this percentage as observed in Figure 5.6.

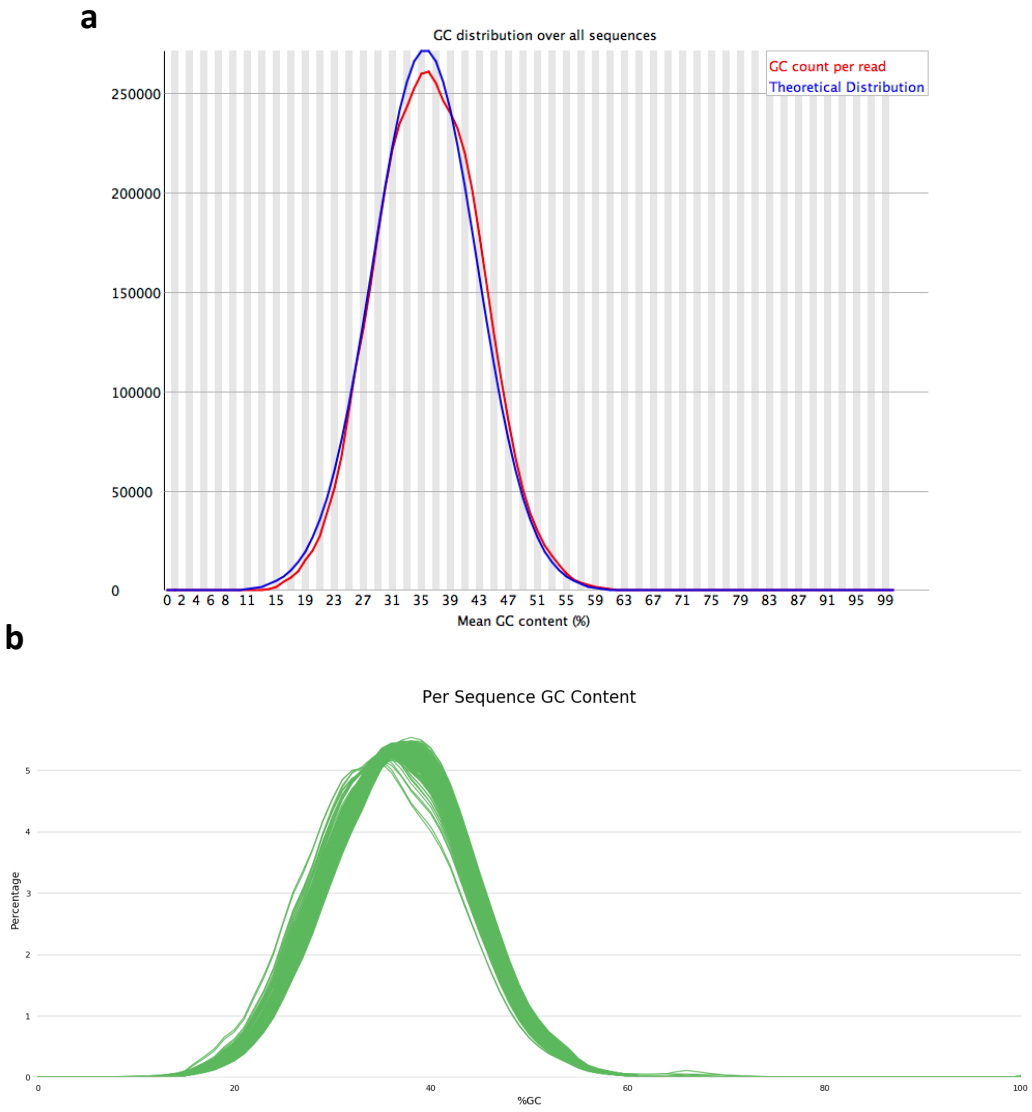



**Figure 5.6 Sequencing quality check – Per sequence GC content**

Modules used for sequencing quality check were generated using FastQC. Once generated for every samples, the modules were compiled using MultiQC. For clarity, each module is represented for a single sample (here wild-type generation 0 duplicate 1) (**a**) and the MultiQC output for the full set of samples (**b**).

### 5.2.3.5 Per Base N Content

N bases are bases not recognised by the sequencer, or base unable to be called with sufficient confidence. It is not unusual to see an increase in N base appearing towards the end of the reads, related to the phasing effect. As shown in Figure 5.7, barely no N bases were identified in the sequencing data.

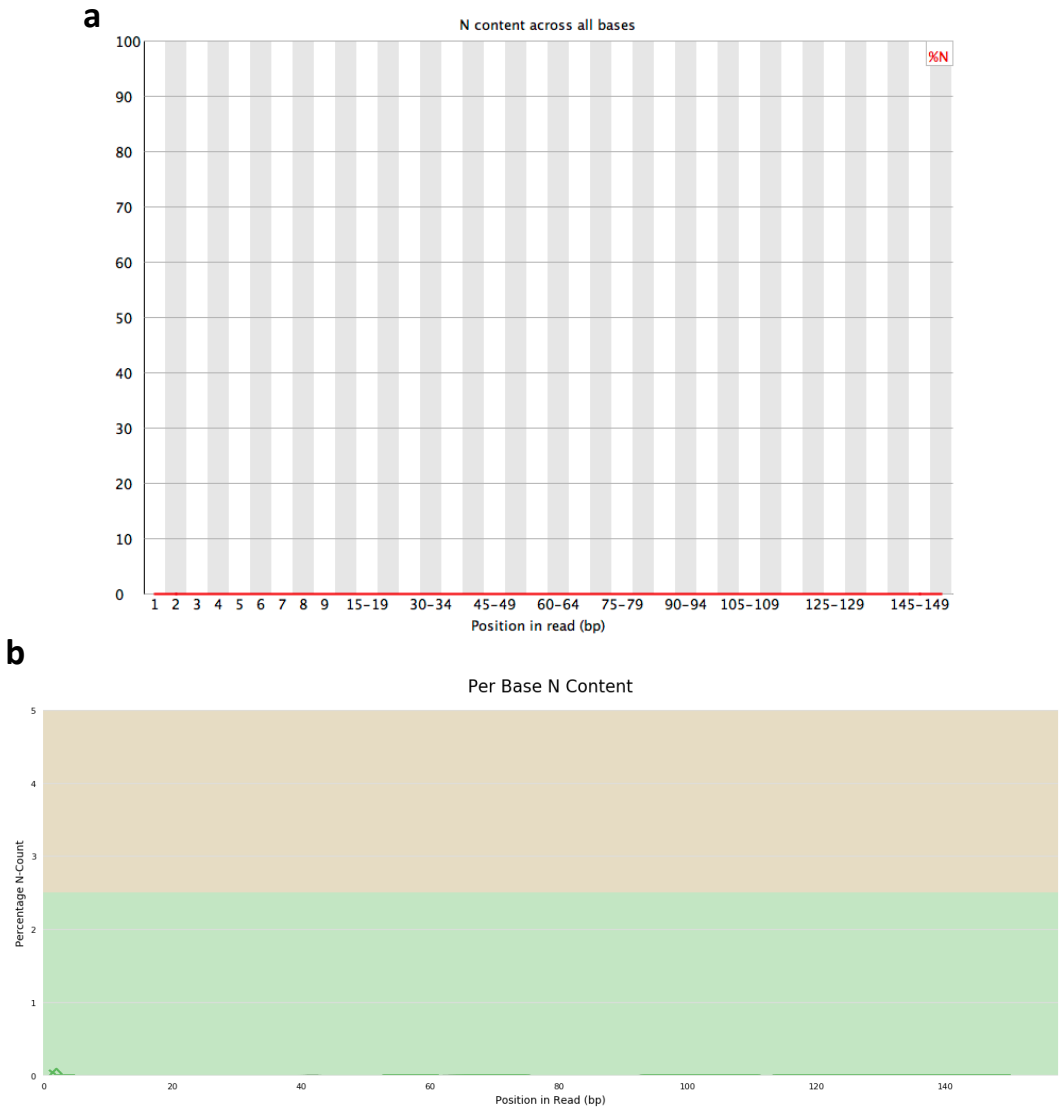



**Figure 5.7 Sequencing quality check – Per base N content**

Modules used for sequencing quality check were generated using FastQC. Once generated for every samples, the modules were compiled using MultiQC. For clarity, each module is represented for a single sample (here wild-type generation 0 duplicate 1) (**a**) and the MultiQC output for the full set of samples (**b**).

### 5.2.3.6 Sequence Length Distribution

Figure 5.8 aims to shows the distribution of the read length in the analysed FASTQ file. As the sequencing performed was 151 cycles of sequencing, it is expected to find a 151 bp length for all samples analysed.

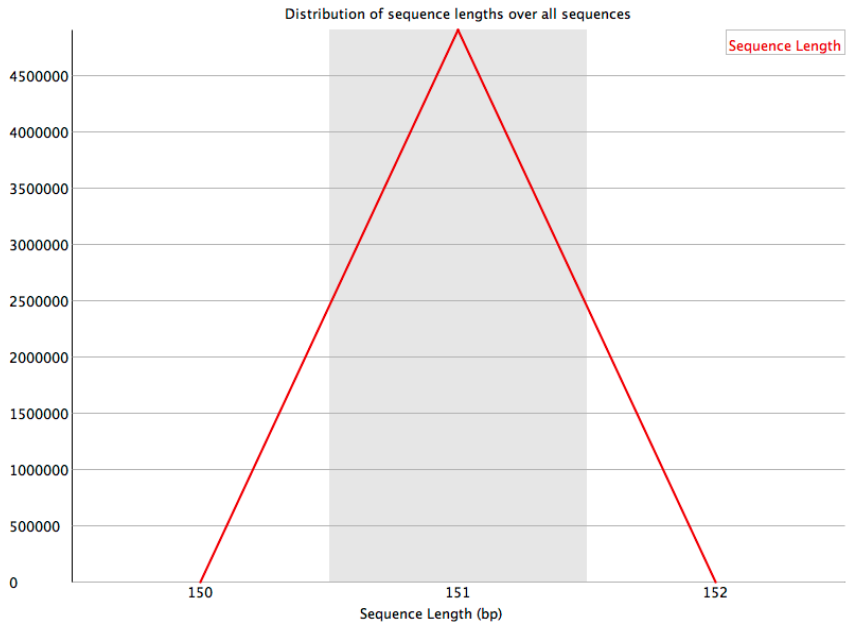



**Figure 5.8 Sequence quality check – Sequence length distribution**

Modules used for sequencing quality check were generated using FastQC. Once generated for every samples, the modules were compiled using MultiQC. Module is represented for a single sample (here wild-type generation 0 duplicate 1) and the MultiQC output simply indicated that all samples have sequences of a single length of 151 bp.

### 5.2.3.7 Sequence Duplication Levels

Random fragmentation of the genome prior to library preparation is expected to lead to unique sets of reads. However, read duplication bias can be observed as a result of PCR amplification during the library preparation process. The Sequence Duplication Levels modules counts the degree of reads duplicated within a library. Conventionally, a

warning flag is raised when non-unique sequences make up more than 20% of the total sequences. As observed in Figure 5.9, the overall duplicated sequences stay below 20 % of the total sequences, indicating the overall uniqueness of each reads sequenced.

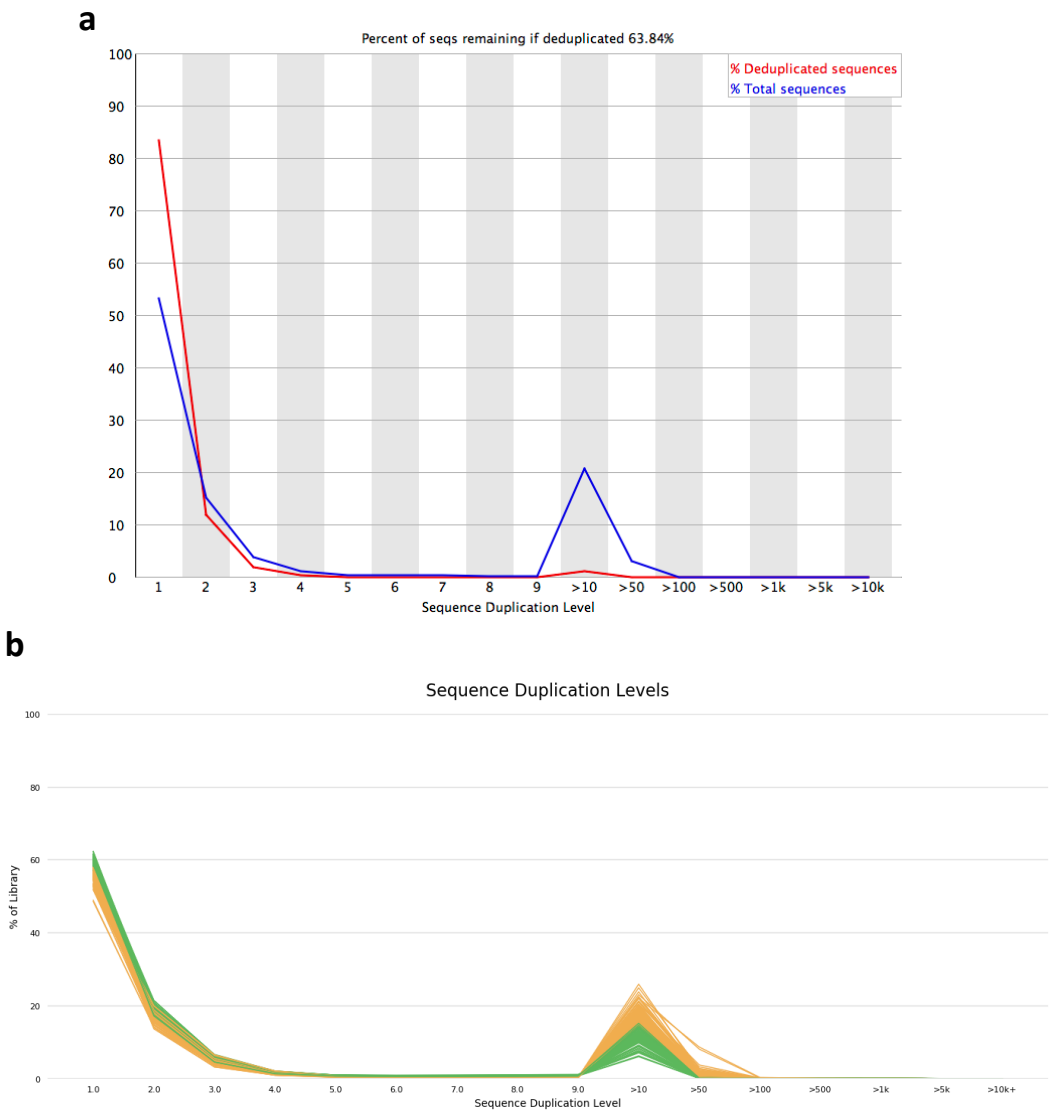



**Figure 5.9 Sequencing quality check – Sequence duplication levels**

Modules used for sequencing quality check were generated using FastQC. Once generated for every samples, the modules were compiled using MultiQC. For clarity, each module is represented for a single sample (here wild-type generation 0 duplicate 1) (**a**) and the MultiQC output for the full set of samples (**b**).

#### 5.2.3.8 Overrepresented Sequences

A sequence is considered overrepresented if it accounts for more than 0.1 % of the total reads reported. The overrepresented sequences module lists these potential reads, and tries to identify the origin of the highly represented sequences. No overexpressed sequence found in our set of data. Overall, the qualitative analysis of the sequencing data shows a very good quality, allowing us to move forward with confidence for variant calling.

### 5.2.4 Mutation Burden

#### 5.2.4.1 Differences between the reference genome and the parental strain

Each of the 85 samples was processed in the same way from alignment to variant filtering (see Materials and Methods 2.6). The first enquiry was to quantify the number of single nucleotide variant (SNV) present in the wild-type parental strain called by HaplotypeCaller (Ryan Poplin, 2017) against the reference genome used for alignment (ASM294v2 from Pombase) (Lock *et al.*, 2019; Wood *et al.*, 2002). Approximately 100 SNVs were extracted from the wild-type parental strain at generation 0 (Figure 5.10 – a) which indicates differences between the reference genome and the parental strain used in this study. Notably, the *cdc20*$^{N364K}$ strain shows a higher number of SNVs compared to the parental strain even at generation 0. This is also the case for the double mutant strain lacking MMR. This is indicative of additional mutations not related to the strain background, nor to the genome of reference used for the variant calling. To determine whether the SNVs called in the parental strain are the same ones called in the constructed strains, the wild-type SNVs at generation 0 (position and type) were

subtracted from each of the non-parental strains, according to their replicate group. A replicate group is constituted of one replicate of each strain, i.e. wild-type parental, *cdc20*$^{exonull}$, *cdc20*$^{N364K}$, *cdc20*$^{Y459F}$, *msh2Δ*, *cdc20*$^{exonull}$ *msh2Δ*, *cdc20*$^{N364K}$ *msh2Δ* and *cdc20*$^{Y459F}$ *msh2Δ*. There is in total 3 groups of replicates: 1, 2 and 3. After removal of the 'parental SNVs', there is a clear decrease in the number of SNVs for all strains (Figure 5.10 – b). The *cdc20*$^{exonull}$ SNV level goes down to almost basal level while the levels for *cdc20*$^{N364K}$ and *cdc20*$^{Y459F}$ indicates additional mutations which cannot be explained by the parental strain background. This could however be explained by the mutagenic power previously observed in these strains (Figure 3.6). Notably, the *msh2Δ* strains also exhibit a number of SNV compare to the wild-type ones, even after removal of the 'parental SNVs' (Figure 5.10 – b), which could reflect the different strain background, as *msh2Δ* was not derived from the wild-type strain, as the *cdc20*$^{EDM}$ strains are.

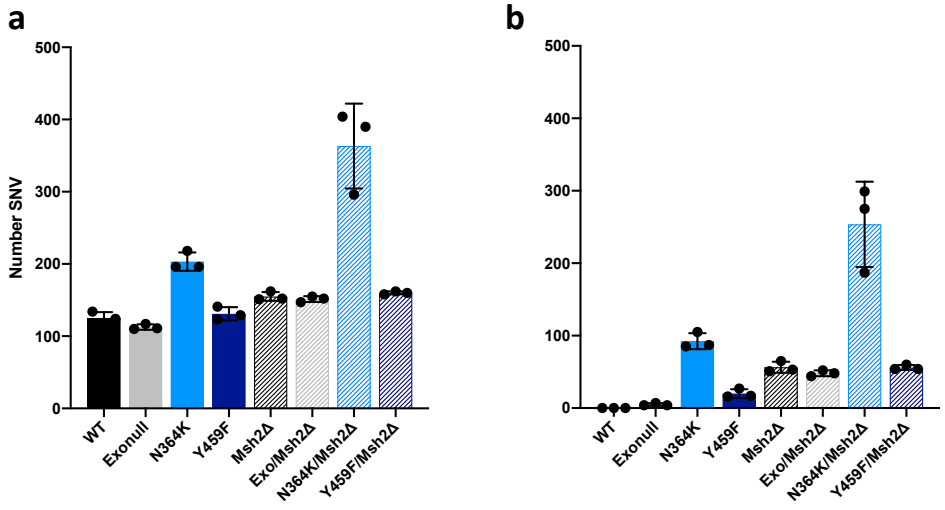



**Figure 5.10 SNVs discrepancies between reference and parental strain**

(**a**) Number of SNVs present at generation 0 in all strains; (**b**) SNVs left after removal of the parental SNVs; n=3

The same analysis was performed for indels. About 200 indels were detected in the parental strain compared to the reference genome and a similar number was also found in the *cdc20*$^{EDM}$ strains (Figure 5.11– a). This indicates that, similarly to the SNVs, there are differences between the parental strain used in our study and the reference genome. Notably, the *msh2Δ* single mutant strain clearly shows an increase in the number of indels compared to the parental strain, presumably due to MMR deficiency. This is also observed for the *msh2Δ* double mutant strains. Working in the same logic as for the SNVs, the indels present in the parental strain were also found in the subsequently constructed strains (Figure 5.11– b), implying that the same background is found between parental strains and subsequent constructs. Nevertheless, as previously observed in (Figure 5.10 – b), the number of indels is also higher in the *msh2Δ* strain, even after removal of the 'parental indels', illustrating the difference of strain origins between the *msh2Δ* and the wild-type strain.

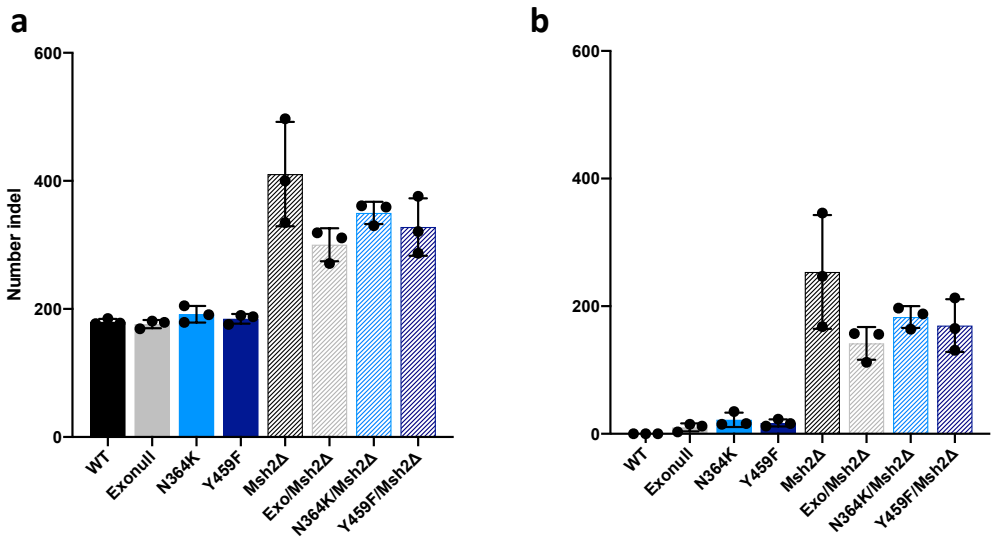



**Figure 5.11 Indels discrepancies between reference and parental strain**

(**a**) Number of indels present at generation 0 in all strain; (**b**) indels left after removal of the ones detected in parental strain. n=3

### 5.2.5 Change in mutation incidence with increasing number of generations

Even at the start of the MA experiment, the *cdc20*$^{EDM}$ strains exhibit an already higher number of SNVs and indels compared to the parental strain, this phenotype was even more drastic when combined with MMR-deficiency (Figure 5.10). To examine the pace of appearance of these alterations, the mean number of SNVs and indels were reported for each studied strain for each generation available. Very few SNVs were acquired for the wild-type parental strain, as well as for the MMR-deficient and the *cdc20*$^{exonull}$ ones, their mutation rates being quite in the same range (Figure 5.12 and Table 5.1). However, *cdc20*$^{Y459F}$ exhibits a steady increase in SNVs appearance, even higher for *cdc20*$^{N364K}$. The double mutants clearly show a striking increase of SNVs, *cdc20*$^{N364K}$ being the one with the greater slope, indicating a stronger mutation rate per generation compared to *cdc20*$^{Y459F}$ and *cdc20*$^{exonull}$. Notably, it seems that the gradients level off for the *cdc20*$^{exonull}$ *msh2Δ*, *cdc20*$^{Y459F}$ *msh2Δ*, and *cdc20*$^{N364K}$ strains which might be translated by a threshold for viable number of SNVs. Furthermore, although with difficulties as only two out of three replicates survive after 50 generations, *cdc20*$^{N364K}$ msh2*Δ* double mutant strains manage to go slightly above this threshold observed for the single mutant strains.

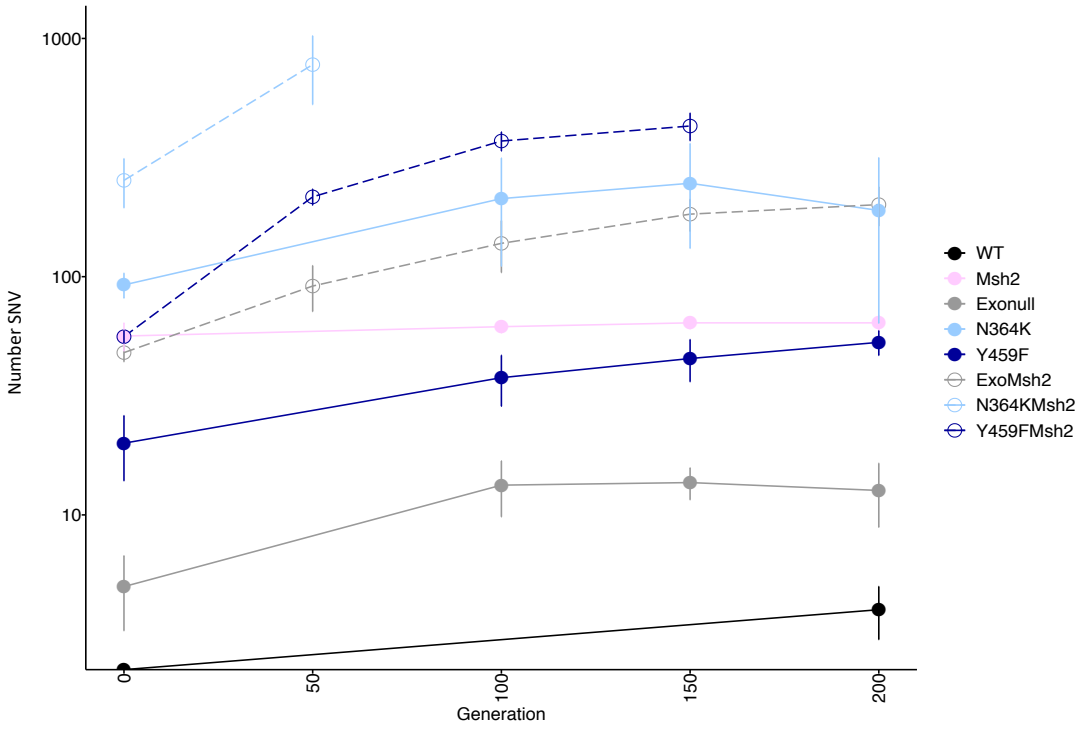



**Figure 5.12 SNVs arising along generations**

The numbers of SNVs were defined for each strain for each generation, each group of replicates processed separately. Then the SNVs present in the parental strain for generation 0 were subtracted from all samples within one group of replicate. The same process was applied on the two additional replicates. The plot represents on a log scale the resulted mean number of SNVs, with the SD from three replicates at the exception of *cdc20*$^{N364K}$ *msh2Δ* at generation 50 and *cdc20*$^{Y459F}$ *msh2Δ* at generation 150 for which only two values are available.

| Strain | Average mutation rate (n=3) | Fold-change (relative to wild-type) |
|---|---|---|
| wild-type | 1.4 x $10^{-09}$ ±0.4 | 1 |
| *msh2⊿* | 4.2 x $10^{-09}$ ±2.24 | 3.1 |
| *cdc20*$^{exonull}$ | 2.7 x $10^{-09}$ ±1.8 | 2.2 |
| *cdc20*$^{N364K}$ | 3.5 x $10^{-08}$ ±4.1 | 24.7 |
| *cdc20*$^{Y459F}$ | 1.2 x $10^{-08}$ ±0.7 | 8.6 |
| *cdc20*$^{exonull}$ *msh2⊿* | 5.5 x $10^{-08}$ ±14 | 39.0 |
| *cdc20*$^{N364K}$*msh2⊿* | 7.0 x $10^{-07}$ ±380 | 539.8 |
| *cdc20*$^{Y459F}$ *msh2⊿* | 2.0 x $10^{-07}$ ±49 | 145 |

**Table 5.1 Mutation rate and fold-change – SNV**

Mutation rate and fold change calculated for the highest generation of each strains, i.e. 200 generations, except for *cdc20*$^{N364K}$ *msh2⊿* (two replicates at generation 50 and one replicate at generation 0) and for *cdc20*$^{Y459F}$ *msh2⊿* (two replicates at generation 150 and one at generation 100). The mutation rate was obtained by the division of the number of SNVs for the highest generation by the number of generations, further divided by the fission yeast genome size, giving the number of SNV per generation per bp synthetised (e.g. (number of indels at generation 200/200)/14 Mb).

The same analysis was performed for indels accumulated along generations (Figure 5.13). Contrary to the parental strain and the single mutant for c*dc20*$^{EDMs}$, the *msh2⊿* strain shows a steady increase of indels along generations with a mutation rate of 5.6 x $10^{-08}$ ±28 indels/base/generation (Table 5.2). When compared to a *cdc20*$^{EDMs}$, the rate of indels appearing is in the same range for all double mutants, and higher than the single msh2⊿ strain. It appears very clear as well here that the curves level off at the same level, probably translated a threshold of viability.

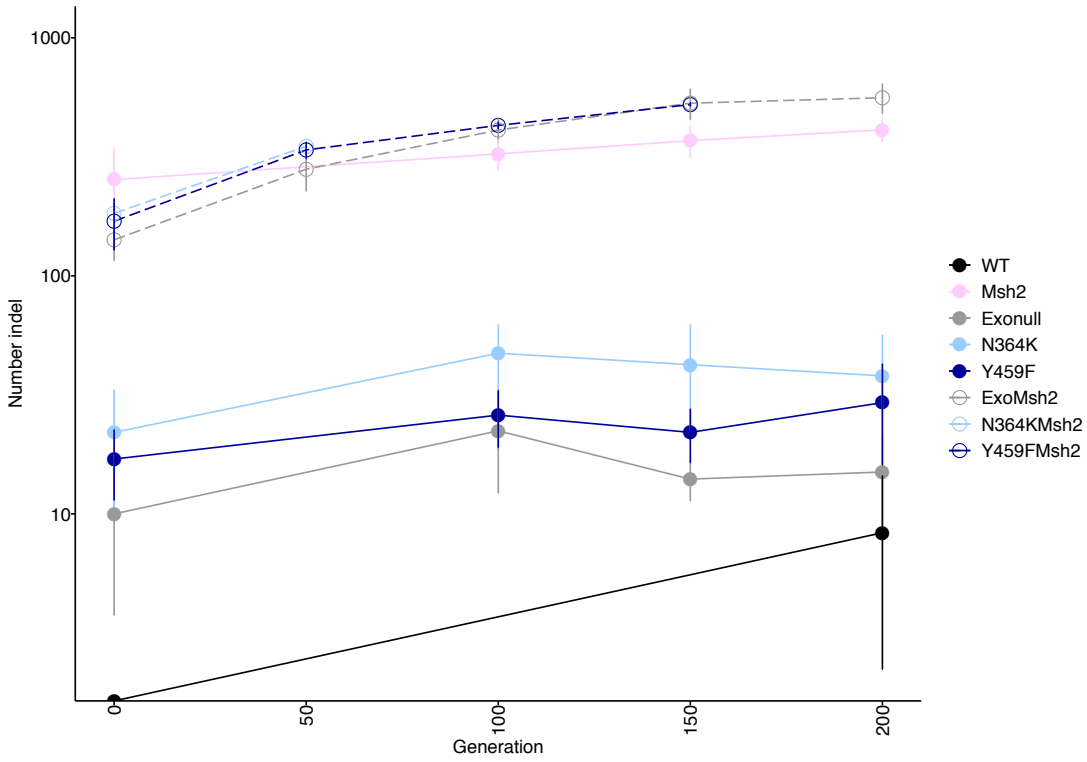



**Figure 5.13 Indels arising along generations**

The numbers of indels were defined for each strain for each generation, each group of replicates processed separately. Then the indels present in the parental strain for generation 0 were subtracted from all samples within one group of replicate. The same process was applied on the two additional replicates. The plot represents on a log scale the resulted mean number of indels, with the SD from three replicates at the exception of *cdc20*$^{N364K}$ *msh2Δ* at generation 50 and *cdc20*$^{Y459F}$ *msh2Δ* at generation 150 for which only two values are available.

| Strain | Average mutation rate (n=3) | Fold-change (relative to wild-type) |
|---|---|---|
| wild-type | 3.0 x $10^{-09}$ ±2.24 | 1 |
| *msh2⊿* | 5.6 x $10^{-08}$ ±28 | 44 |
| *cdc20*$^{exonull}$ | 3.5 x $10^{-09}$ ±1.4 | 2.5 |
| *cdc20*$^{N364K}$ | 6.9 x $10^{-09}$ ±7.4 | 4.8 |
| *cdc20*$^{Y459F}$ | 4.4 x $10^{-09}$ ±3.4 | 3.1 |
| *cdc20*$^{exonull}$ *msh2⊿* | 1.5 x $10^{-07}$ ±20 | 108 |
| *cdc20*$^{N364K}$*msh2⊿* | 2.3 x $10^{-07}$ ±9.1 | 183 |
| *cdc20*$^{Y459F}$ *msh2⊿* | 1.7 x $10^{-07}$ ±15 | 125 |

**Table 5.2 Mutation rate and fold-change – indels**

Mutation rate and fold change calculated for the highest generation of each strains, i.e. 200 generations, except for *cdc20*$^{N364K}$ *msh2⊿* (two replicates at generation 50 and one replicate at generation 0) and for *cdc20*$^{Y459F}$ *msh2⊿* (two replicates at generation 150 and one at generation 100). The mutation rate was obtained by the division of the number of indels for the highest generation by the number of generations, further divided by the fission yeast genome size, giving the number of SNV per generation per bp synthetized (e.g. (number of indels at generation 200/200)/14 Mb).

Taken together, it is interesting to see that *msh2⊿* strain does not exhibit an increase in SNVs acquisition, although it shows a steady increase in indels, related to the impairment of MMR. This suggests that in fission yeast, MMR does not fully rely on Msh2 for corrections of mismatches but might have an alternative pathway to maintain the mutation rate at a level similar to a MMR-proficient strain. Nevertheless, combination of MMR-deficiency and cdc20EDM indicates that replication-errors are corrected by MMR.

### 5.2.6 Mutational pattern

Specific mutational signatures were previously associated with *POLE*$^{EDM}$ variants, as well as MMR deficiency (Alexandrov *et al.*, 2020; Alexandrov, Nik-Zainal, Wedge, Aparicio*, et al.*, 2013). Also, mutation signatures were identified as specifically associated with a given *POLE*$^{EDM}$ (Haradhvala *et al.*, 2018; Shlien *et al.*, 2015). In addition, a distinct mutational signature was associated with concurrent *POLE*$^{EDM}$ and MMR alteration (Alexandrov *et al.*, 2020; Haradhvala *et al.*, 2018; Shlien *et al.*, 2015). With the purpose of establishing whether these apply also to the studied *cdc20*$^{EDMs}$, the type of mutations present at the oldest generational point were analysed in two complementary ways. The first aimed to examine the type of SNVs alone (mutational spectrum), while the second includes the 5' and 3' immediate context surrounding the alteration (mutational pattern). The SNVs were extracted for the highest generations, followed by subtraction of the SNVs present in the parental strain at generation 0 for each group of replicates.

The mutational spectrum (Figure 5.14) for parental wild-type strain is difficult to interpret as the number of SNVs present is very small and not clear and strong pattern emerges from it. The *msh2Δ* strain has no clear pattern either, although the ratio of C>G and T>G transversions are the smallest ones compared to the other type of mutations. Lack of MMR sees the proportion of T>C transition increase for the *cdc20*$^{exonull}$ and *cdc20*$^{Y459F}$ variants. The *cdc20*$^{N364K}$ strain however sees its number of alterations increase, but there is no change in the ratio of mutation type.

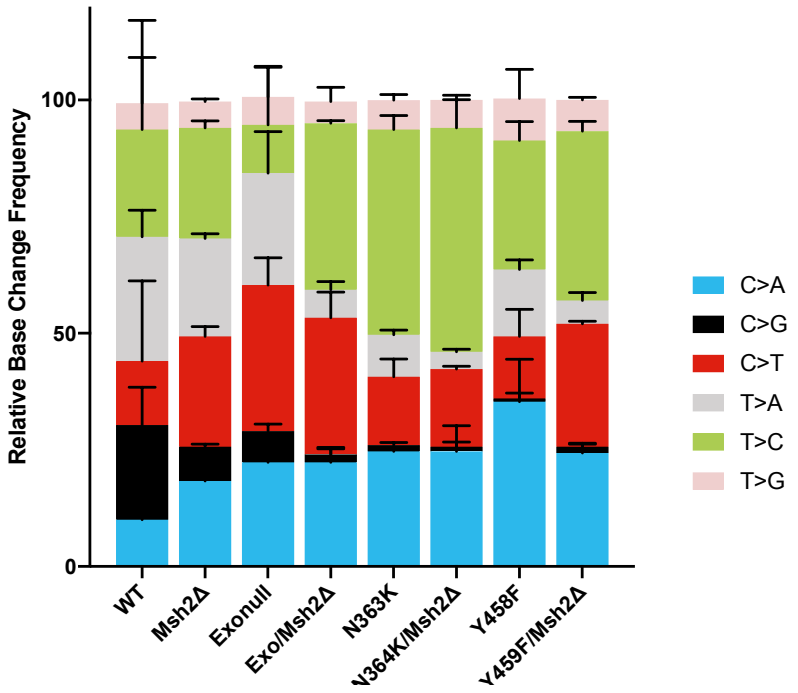



**Figure 5.14 Mutational Spectrum**

SNVs present at the latest generations for each of the sequenced strains were extracted. Then the SNVs present at the generation 0 in the wild-type parental strain were subtracted within each group of replicates. The average relative base change frequency was then plotted according to the type of mutations (C>A, C>G, C>T, T>A, T>C and T>G) with the SD from the three groups of replicates. Average number of alterations per strain: wild-type = 8; *msh2Δ* = 162; *cdc20*$^{exonull}$ = 16; *cdc20*$^{exonull}$ *msh2Δ* = 196; *cdc20*$^{N364K}$ = 198; *cdc20*$^{N364K}$ *msh2Δ* = 583; *cdc20*$^{Y459F}$ = 54; and *cdc20*$^{Y459F}$ *msh2Δ* = 419

To distinguish mutational signatures that might cause the same type of mutations, the mutation context is considered. To increase the analysis power, the SNVs were pooled for each strain, constituting a catalogue of variants per strain. Still, the number of alterations was too low to extract a mutational signature. So instead, mutational patterns were extracted from the catalogues of variants using the R package MutationalPatterns (Blokzijl *et al.*, 2018) (Figure 5.15). Analysis of the mutational profiles indicates a rather flat one for the MMR-deficient strain, while some distinct peaks emerge for the *cdc20*$^{EDM}$ strains, peaks evolving when MMR is depleted in these

strains. After the publication from Alexandrov, Nik-Zainal, Wedge, Aparicio*, et al.* (2013), 22 validated mutational signatures were identified and rendered available on COSMIC.

In order to assess if any of the mutational patterns obtained in this present study could be related to one of the COSMIC mutational signatures, a cosine heatmap was produced using MutationalPatterns package (Blokzijl *et al.*, 2018). The cosine coefficient aims to evaluate similarities – the higher the cosine coefficient, the more similar the compared entities are. As observed in Figure 5.16, the *cdc20*$^{EDMs}$ cluster together while the *msh2Δ* strain and double mutants cluster in a second group. As previously mentioned, it is important to emphasise the quite small number of alterations present in the *msh2Δ* and *cdc20*$^{EDMs}$ single mutant strains, therefore to be aware that the interpretation of the cosine heatmap might be slightly delicate. Nevertheless, the number of mutations being much higher in the *cdc20*$^{EDMs}$ *msh2Δ* double mutant strains allows more confidence in the data interpretation. Nonetheless, the cosine coefficients selected as relevant would be the ones higher than 0.75. Looking at the *msh2Δ* strain, the highest cosine score is related to the signature 5 and 8, signatures of currently unknown aetiology. None of the *cdc20*$^{EDMs}$ strains exhibit any cosine coefficient higher than 0.75. However, for the *cdc20*$^{EDMs}$ *msh2Δ* double mutant strains, *cdc20*$^{exonull}$ *msh2Δ* and *cdc20*$^{Y459F}$ *msh2Δ* show a clear high cosine coefficient for the signatures 14 and 20. Signature 14 was previously established as linked with concurrent *POLE*$^{EDM}$ and defective MMR, while signature 20 is associated with concurrent Pol $\delta$ mutations and defective MMR. *cdc20*$^{N364K}$ *msh2Δ* strain exhibits a high cosine factor for those same signatures, however not as high. Moreover, it exhibits a clear high cosine coefficient for signature 12 (which is currently of unknown aetiology) and signature 26 which is associated with defective MMR.

The signature 14 is characterized by increase in C>A transversion in NCT context. Going back to the mutational patterns after identification of a high cosine coefficient for the *cdc20*$^{EDMS}$ *msh2Δ* double mutant strains, we can actually notice the related peaks in the corresponding strains. As similar trend is observed in all three replicates. The characteristic alterations observed in signature 10 related to *POLE*$^{EDM}$ (Alexandrov, Nik-Zainal, Wedge, Aparicio*, et al.*, 2013) are C>A transversion in TCT context, C>T transition in TCG context, and T>G transversion in TTT context. Alexandrov *et al.* (2020) subcategorized this signature into 2, with SBS10a representing the C>A transversion in TCT context and SBS10b representing the C>T transition in TCG context. The T>G transversion in TTT context is since then related to signature SBS28, often observed with the signatures SBS10a/b.

Taken together, these data indicate that the signature 10 associated with *POLE*$^{EDMs}$ in human cancers is not the most prevalent mutational pattern detected in the *cdc20*$^{EDMs}$ of our study. Nonetheless, the signature 14, previously associated with *POLE*$^{EDMs}$ and MMR deficiency is clearly present in all *cdc20*$^{EDM}$ strains, and amplified in a MMR deficient context. Surprisingly, the signature 20, associated with *POLD1*$^{EDMs}$ and MMR deficiency is as strongly present as the signature 14 in all *cdc20*$^{EDMs}$, and amplified in a MMR deficient background.

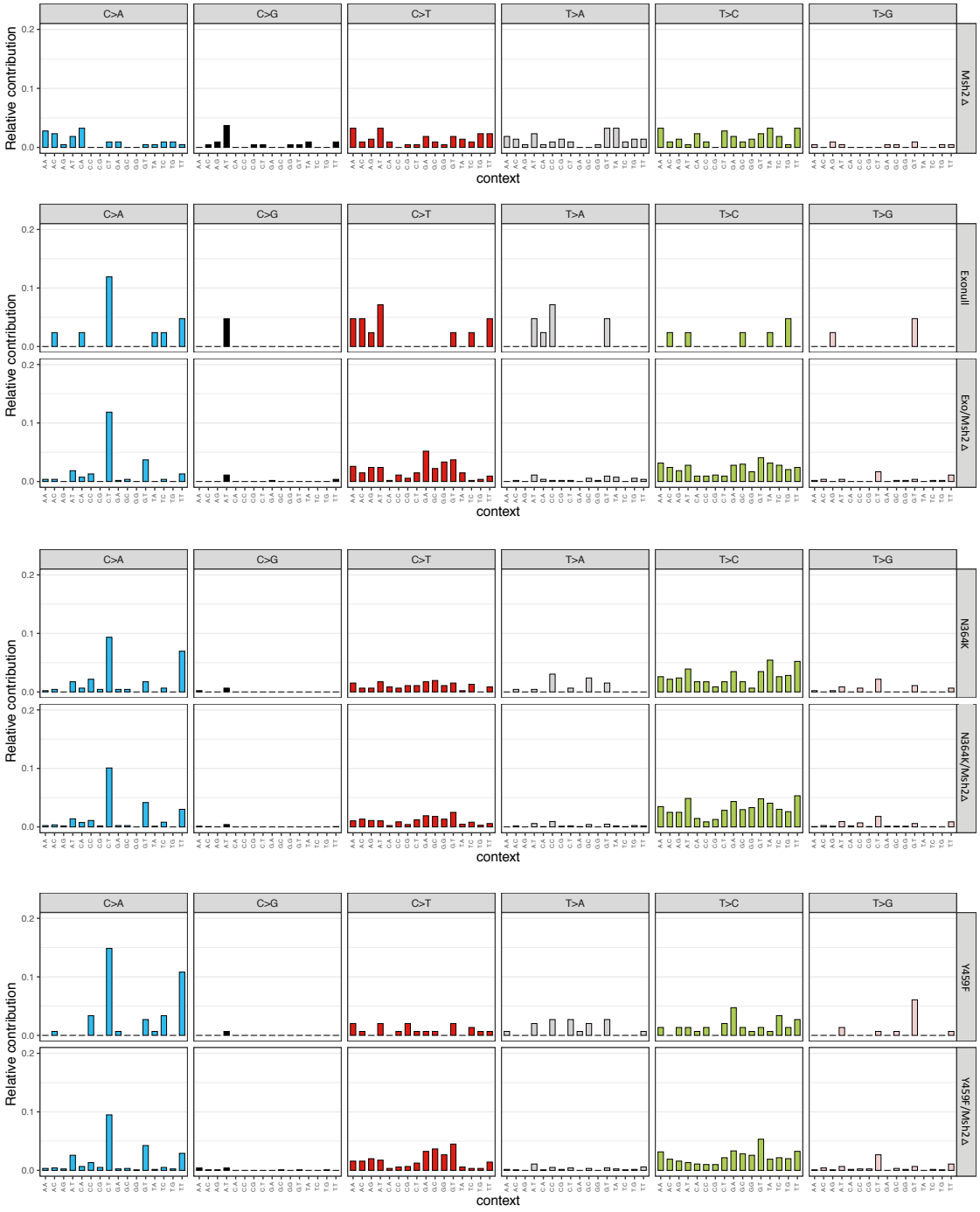



**Figure 5.15 Mutational patterns**

The SNVs extracted from the highest generations as described in 5.2.4 were pooled together for each strain. The mutational patterns were extracted from these catalogues of SNVs and represented with the relative proportion of 96 trinucleotides contexts for C>A, C>G, C>T, T>A, T>C and T>G. The total number of mutations analysed are as followed: *msh2Δ* = 215; *cdc20*$^{exonull}$ = 42; *cdc20*$^{N364K}$ = 460; *cdc20*$^{Y459F}$ = 148; *cdc20*$^{exonull}$ *msh2Δ* = 540; *cdc20*$^{N364K}$ *msh2Δ* = 1727 and *cdc20*$^{Y459F}$ *msh2Δ* = 1202

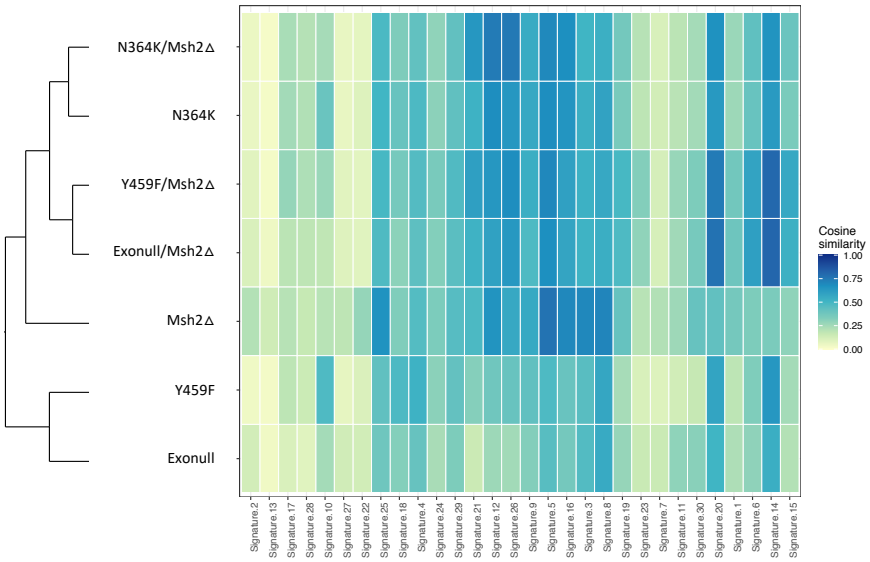



**Figure 5.16 Similarities between mutational patterns and COSMIC signatures**

Mutational patterns established for previously were compared to the COSMIC signatures using MutationalPatterns. The samples are clustered according to their similar mutational patterns. The number of variants used to obtain the heatmap are the same as stated in Figure 5.15.

## 5.3 Discussion

*POLE*$^{EDM}$ alterations are associated with a specific mutational signature (Alexandrov *et al.*, 2020; Alexandrov, Nik-Zainal, Wedge, Aparicio*, et al.*, 2013). Notably, *POLE*$^{EDM}$ variants display distinct mutational signatures resembling the signature 10 (Haradhvala *et al.*, 2018; Hodel *et al.*, 2020). Combined with MMR-deficiency, *POLE*$^{EDM}$ tumours exhibit a distinct mutational signature which is different from a combined MMR-deficiency and *POLE*$^{EDMs}$ signatures (Alexandrov *et al.*, 2020; Haradhvala *et al.*, 2018). Budding yeast strains lacking Pol ε exonuclease activity are viable in a MMR-depleted background (Xing *et al.*, 2019). Nonetheless, the *pol2-P301R* variant is synthetically lethal with *msh6Δ*, emphasising that, contrary to simple lack of proofreading ability, the hypermutated phenotype of *pol2*$^{P301R}$ reaches a viability threshold when replication-error correction is missing (Xing *et al.*, 2019). Similarly to Xing *et al.* (2019), the *cdc20*$^{exonull}$ strain is viable in a MMR-deficient background (Figure 5.1). Even if it exhibits a hypermutated phenotype, although not to the level of *cdc20*$^{P287R}$ (Figure 3.6), the *cdc20*$^{Y459F}$ variant shows no synthetic lethality with complete MMR abrogation (Figure 5.1). Notably, *cdc20*$^{N364K}$ variant is not synthetically lethal but exhibits sickness when combined with *msh2Δ* (Figure 5.1). Altogether, these data emphasise the discrepancy of mutagenesis ability amongst *cdc20*$^{EDM}$ variants, as the *cdc20*$^{Y459F}$ and *cdc20*$^{N364K}$ are probably less mutagenic than *cdc20*$^{P287R}$, as the germline variants can support the absence of MMR.

Even though *cdc20*$^{N364K}$ and *cdc20*$^{Y459F}$ msh2*Δ* double mutants are viable, only the *cdc20*$^{exonull}$ *msh2Δ* double mutants could grow up to 200 generations (Figure 5.2). Indeed, the *cdc20*$^{Y459F}$ *msh2Δ* strains could not grow beyond 150 generations. An even

more drastic phenotype was observed for the *cdc20*$^{N364K}$ *msh2Δ* strains, which could not go beyond 50 generations. This could also be explained by the mutagenic phenotype already observed for the *cdc20*$^{Y459F}$ and *cdc20*$^{N364K}$ strains. Already mutagenic, the *cdc20*$^{Y459F}$ strains would accumulate mutations along generations, most likely rectified by the MMR pathway. This would apply for *cdc20*$^{N364K}$, also at a more striking pace, as it was observed to be more mutagenic than *cdc20*$^{Y459F}$ (Figure 3.6). This result reinforces that fact that *cdc20*$^{N364K}$ is more mutagenic than *cdc20*$^{Y459F}$, and that simple loss of exonuclease activity is not sufficient for the mutagenic phenotype observed. This also strongly suggests that errors induced by *Cdc20* are corrected by the MMR pathway.

Mutation accumulation experiments enabled to assess the spontaneous mutation rate of *S. pombe* at approximately 2.0 x $10^{-10}$ mutations/base/generation (Behringer *et al.*, 2015; Farlow *et al.*, 2015). Through these studies was also highlighted the fact that strains diverged from the one which sequence is publicly available (Wood *et al.*, 2002).. Our data indicates a higher mutation rate for the parental strain used in our study (Table 5.1) than the one previously reported. Nevertheless, similarly to Behringer *et al.* (2015) our data show that the parental strain used in our study exhibits variations when compared to the reference genome used for mapping (Figure 5.10 and Figure 5.11). This could be explained by the natural divergence of the parental strain from the one used to obtain the genomic sequence of fission yeast in 2002, and the natural emergence of mutations even in quiescence (Gangloff *et al.*, 2017). Besides alterations between the reference genome and the parental strain, additional SNVs and indels were already present at the start of the MA experiment in the *cdc20*$^{EDM}$ strains, reinforcing the mutagenic phenotype of the *cdc20*$^{EDM}$ strains (Figure 5.10). Notably, the strains lacking Msh2 exhibit a clear increase in indels, presumably due to MMR deficiency (Figure 5.11).

Analysis of the mutagenic capability of the *cdc20*$^{EDM}$ variants over time further confirmed this notion of mutagenic property in unchallenged condition (see section 5.2.5). *cdc20*$^{N364K}$ is the more mutagenic variant compared to *cdc20*$^{exonull}$ and *cdc20*$^{Y459F}$ as this is translated by an overall higher mutation rate at the highest generation obtained, as well as a clear lethality after 50 generations when combined with MMR deficiency. It seems that the *cdc20*$^{N364K}$ *msh2Δ* strains reach more quickly a mutagenic threshold, explaining why the double mutants could not grow beyond 50 generations, while *cdc20*$^{Y459F}$ *msh2Δ* strains could go up to 150 generations (Figure 5.12). Notably, the level off of the indels acquired over generation might indicate the threshold number of indels a cell can cope with before error-induced extinction.

MMR-deficiency in human cancer is associated with characteristic mutational signatures (see section 1.4.2). Unexpectedly, none of these signatures were to be observed in fission yeast *msh2Δ* strains (Figure 5.15 and Figure 5.16). Instead, the mutational pattern shows a link with the signatures 5, 25 and 16 predominantly, all currently of unknown aetiology (Alexandrov *et al.*, 2020). The fact that the mutational patterns observed for the *msh2Δ* strain does not correspond to any previously characterised MMR deficiency mutational signatures could be explained by first the difference of organism in which these patterns were studied. Indeed, the COSMIC signatures are extracted from human tumours, while here the fission yeast is used as a model organism. The second explanation could be the quite low number of SNVs used to obtain the mutational patterns, leading to a potentially inaccurate result. Farlow *et al.* (2015) and Behringer *et al.* (2015) passaged the cells for 1716 and 1952 generations, respectively, whereas our study passaged up to 200 generations. Prolongating the MA

experiment would provide a more extending catalogue of mutations, thus ultimately a more accurate mutational pattern.

The signature SB10a and SBS10b, characterised by C>A transversion in a TCT context and C>T in a TCG context, respectively, are associated with *POLE*$^{EDM}$ alterations in human cancer (Alexandrov *et al.*, 2020). In our study, the *cdc20*$^{EDM}$ single mutants do not exhibit a strong similarity with the signature 10 (Figure 5.16), although some weak similarity could be observed such as the presence of C>A transversions in a TCT context for *cdc20*$^{N364K}$ and *cdc20*$^{Y459F}$ (Figure 5.15). The fact that the signature 10 is not the prominent one observed in the *cdc20*$^{EDM}$ single mutants is unexpected. Similarly to the *msh2Δ* strain, this could be explained by the difference in model organism used in this study, or low number of SNVs used to generate the mutational patterns. Although, it would be interested to assess the mutational signature for *cdc20*$^{P287R}$ and/or *cdc20*$^{V412L}$ present in fission yeast to clarify this issue (Haradhvala *et al.*, 2018). Furthermore, mutational signatures extracted from human tumours are *POLE*$^{EDMs}$-dependent (Haradhvala *et al.*, 2018; Hodel *et al.*, 2020; Shlien *et al.*, 2015) prompting to investigate in the future whether *POLE*$^{N363K}$ have also a distinct mutational signature in human tumours. Nonetheless, the mutational patterns of the *cdc20*$^{EDM}$ variants show an association with the signature 14 and 20 (Figure 5.16), associated with concurrent *POLE*$^{EDM}$ and MMR-deficiency, and concurrent *POLD1*$^{EDM}$ and MMR-deficiency, respectively (Alexandrov *et al.*, 2020). The loss of Msh2 seems to reinforce and enhance these patterns (Figure 5.16). The fact that the signature 14 is already present in the *cdc20*$^{EDM}$ single mutants, the signal amplified in a MMR-deficient background, suggests that the *cdc20*$^{EDMs}$ possess an intrinsic mutational signature, amplified when these replication-errors are left uncorrected in absence of MMR. Saturation of the MMR in

hypermutated strains could explain why some of these *cdc20*$^{EDM}$ intrinsic mutations are left uncorrected in the single mutants. Contrary to our result, the signature 14 was characterised predominantly in tumours carrying both *POLE*$^{EDM}$ and MMR-deficiency (Haradhvala *et al.*, 2018; Hodel *et al.*, 2020), suggesting a difference of mutational process between mammal cells and fission yeast in a Pol ε-EDM context. The detection of the signature 20 in our study is unexpected as it is associated with the lagging strand replicative DNA polymerase (Alexandrov *et al.*, 2020). Notably, the signature 20 is predominantly characterised by a C>A transversion in a CCT context (Figure 1.8), similarly to one of the main peaks characteristic of the signature 14 (Figure 1.9). The similarity could possibly explain the strong association of the signature 20 to the *cdc20*$^{EDM}$ variants, as well as a possible common mechanism of mutagenicity between *POLE*$^{EDMs}$ and *POLD1*$^{EDMs}$.

In summary, these results confirm that the *cdc20*$^{EDM}$ mutagenicity is shaped by the nature of EDM, and relies on MMR. Noteworthy, our data reveal a possible common mechanism of pathogenicity between *POLE*$^{EDMs}$ and *POLD1*$^{EDMs}$, opening potential exploration for *POLD1*$^{EDMs}$ related cancers.

# Chapter 6

## Discussion

Recent findings have highlighted the involvement of mutations in the exonuclease domain (EDM) of the replicative DNA Pol ε in contributing to the development of a broad range of cancers (reviewed in Rayner *et al.* (2016). Large scale sequencing studies led to the identification of several *POLE*$^{EDMs}$ (Cancer Genome Atlas, 2012; Cancer Genome Atlas Research *et al.*, 2013). Study of these variants suggest that the pathogenic mechanism is related to loss of proofreading capability, promoting genome instability via an increased mutation rate (reviewed in Barbari *et al.* (2017). Some somatic variants are already validated as pathogenic (reviewed in Rayner *et al.* (2016), while germline ones are associated with predisposition to cancer development (Palles *et al.*, 2013). Nonetheless, validation of some germline *POLE*$^{EDMs}$ is still missing, and characterisation of the potential pathogenic mechanism of these variants is essential as this would enable a more informed genetic counselling, prediction to treatment response and exploration of new *POLE*$^{EDMs}$-targeted treatments (Church *et al.*, 2015; Domingo *et al.*, 2016).

To investigate the pathogenicity of human *POLE*$^{EDMs}$, *S. pombe* strains carrying human germline equivalent of human *POLE*$^{EDMs}$ were constructed (see section 3.2.1). As previously reported in budding yeast (Barbari *et al.*, 2018; Shcherbakova *et al.*, 2003; Xing *et al.*, 2019), *cdc20*$^{P287R}$ exhibits an increase in mutation rate greater than the

equivalent *cdc20* strain only devoid of exonuclease activity (Figure 3.6). Similar results were found for *cdc20*$^{N364K}$, as previously observed in budding yeast (Hamzaoui *et al.*, 2020), and the *cdc20*$^{Y459F}$ variant. The newly characterised *cdc20*$^{M295I}$ has a mutation rate lying between a wild-type and an exonuclease-dead version of Cdc20 (Figure 3.6), although this might be explained by the similar nature between methionine and isoleucine, thus less likely to have a structural impact on Cdc20. The *cdc20*$^{V412L}$ and *cdc20*$^{L425V}$ strains in our study exhibit a mutation rate similar to the *cdc20*$^{exonull}$ strain (Figure 3.6), similar to equivalent constructs in budding yeast (Barbari *et al.*, 2018; Hamzaoui *et al.*, 2020). Altogether, our data confirm previously reported observations that first, POLE$^{EDMs}$ have a distinct mutagenicity related to the type of alteration and second, simple loss of exonuclease activity is not sufficient to explain the drastic increase in mutation rate observed for some *POLE*$^{EDM}$ variants.

Although loss of exonuclease activity does not seem to be the sole responsible of the hypermutated phenotype observed for some POLE$^{EDMs}$ (Shinbrot *et al.*, 2014) (see section 3.2.4), a recent publication reported residual activity for the budding yeast holoenzyme Pol ε-P286R (Xing *et al.*, 2019), encouraging further investigation of the exonuclease activity of the Pol ε-EDMs. Using *in vitro* assays, the holoenzymes Pol ε-P286R, -N363K and -V411L exhibit a complete loss of exonuclease activity, while Pol ε-L424V is impaired and Pol ε-M294I is similar to a wild-type protein (see section 4.2.1). Although similar to some extent with previously published reports (Shinbrot *et al.*, 2014), these results contradict recently published data from similar experiments using the holoenzyme from budding yeast where Pol ε-P301R retains some exonuclease activity (Xing *et al.*, 2019). This discrepancy of results could first be due to the different model organisms used. Indeed, diverging almost 1 billion years ago from a common

ancestor, Pol ε in *S. cerevisiae* and human cells retain common molecular characteristics, although these are not entirely conserved (see section 4.3). Second, the non-catalytic subunits of Pol ε seems to play an important role in stabilising the holoenzyme (Araki, Hamatake, Johnston*, et al.*, 1991; Dua *et al.*, 2000; Isoz *et al.*, 2012; Y. Li *et al.*, 2000; Ohya *et al.*, 2000), also involved in the fidelity of the enzyme (Jaszczur *et al.*, 2008; Jaszczur *et al.*, 2009; Kraszewska *et al.*, 2012). Taken together, these results confirm that the hypermutagenic *POLE*$^{EDMS}$ phenotype is not only due to lack of proofreading ability. Furthermore, the discrepancy observed between our study and already reported ones prompts for exploration of the role of Pol ε subunits in different model organisms.

To investigate the pathogenic mechanism of the hypermutated germline *cdc20*$^{EDM}$ variants, phenotypic characterisation of the constructed strains revealed a growth defect for the *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ strains (see sections 3.2.5 and 3.2.6). This phenotype was a first hint toward possible checkpoint activation and mitotic delay present in these strains. Indeed, alteration of the DNA damage sensor Rad3 revealed that *cdc20*$^{P287R}$, but not *cdc20*$^{N364K}$, relies on Rad3 presence for survival (see section 4.2.2), possibly as mutation accumulation would trigger checkpoint activation. Previously published data obtained from budding yeast showed that checkpoint activation due to Pol ε lack of proofreading could contribute to a mutator phenotype (L. N. Williams *et al.*, 2015). As our data clearly demonstrate that this does not apply for *cdc20*$^{P287R}$ as it is synthetically sick with *rad3Δ* (Figure 4.3) and synthetically lethal in a *rad3ts* background (Figure 4.4), it would be interested to test whether the *cdc20*$^{exonull}$ and *cdc20*$^{N364K}$ strains decrease their mutation rate in a *rad3Δ* background. Although this seems unlikely as growth defect for *cdc20*$^{N364K}$ is not restored after *rad3Δ*. The

*cdc20*$^{P287R}$ Rad3-dependent phenotype needs to be investigated. Assessment of the cell cycle progression for *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ would enable to determine when delay happens, and which downstream effector of Rad3 is phosphorylated. The prediction would be a longer S-phase and Chk1 phosphorylation, as the malfunctioning polymerase would might cause replication stress and activating a $G_2$/M phase delayed entry. The synthetic lethality observed between the *cdc20*$^{P287R}$ and *rad3Δ* opens the door to targeted treatment, as silencing of *POLE* and *POLE2* sensitised CRC cell lines to Chk1 inhibition (Rogers *et al.*, 2020). It is important to further investigate the difference of behaviour amongst *cdc20*$^{EDMs}$ as their response to treatment might differ.

Cell cycle defects and checkpoint activation were detected in Pol δ and Pol ε base-selectivity deficient strains (Mertz *et al.*, 2015; L. N. Williams *et al.*, 2015). Investigation of the origin and impact of such phenotype led to the identification of a positive feedback loop related to a mutator phenotype. Pol ε base-selectivity deficient enzyme is more likely to incorporate errors, triggering cell cycle delay and checkpoint activation (L. N. Williams *et al.*, 2015). This in turns leads to an increase in dNTP levels, which itself would further amplify the mutation rate of the strain (A. Chabes *et al.*, 2003). The fact that *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ strains show an increased mutation rate (Figure 3.6), possible checkpoint activation (see section 4.2.2) and resistance to hydroxyurea (Figure 4.12) prompted for investigation of the dNTP levels. However, no drastic difference was observed in dNTP pools amongst *cdc20*$^{EDM}$ variants (see section 4.2.4.2). This is consistent with previously published results for the *S. cerevisiae pol2-4* strain, devoid of exonuclease activity for Pol ε (L. N. Williams *et al.*, 2015). Therefore, the mutagenic mechanism of impaired Pol ε proofreading activity does not seem to be linked to an increase in dNTP levels, at least in *S. pombe*. It would be interesting however to test

whether the hydroxyurea-resistant strains are more prone to an increase in dNTP pools under hydroxyurea treatment.

One of the main challenges in cancer treatment is to specifically target cancer cells. A possible strategy investigated here is to use the mutagenic capacity of the c*dc20*$^{EDM}$ variants as an Achilles' heel and to force the mutation rate up to a lethal threshold by increasing the dNTP levels. Minor dNTP pools elevation dramatically decreases the replication fidelity of exonuclease-dead Pol ε (L. N. Williams *et al.*, 2015), while dNTP pools modulation influence the mutation rate (Schmidt *et al.*, 2019). In our study, only microcolonies were observed when an increase in dNTP pools was affected using the *cdc22*$^{D57N}$ mutation in a *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ background. Division stopped after a couple of divisions only, probably indicative of a replication error catastrophe (Morrison *et al.*, 1993). A fluctuation analysis would be interesting to perform on the *cdc20*$^{exonull}$ *cdc22*$^{D57N}$ double mutant strain to examine whether the mutation rate is elevated compared to the single *cdc20*$^{exonull}$ and *cdc22*$^{D57N}$ mutants. This opens the possibility of testing the use of exogenous deoxyribonucleosides in the Pol ε-P286R mouse model and to see if this is effective in specifically targeting highly mutated tumours.

The *POLE*$^{EDMs}$ could impair not only the exonuclease activity of the holoenzyme, but also its stability or capacity to deal with difficult-to-replicate sequences (Xing *et al.*, 2019). In our study, *cdc20*$^{P287R}$ and *cdc20*$^{N364K}$ (to a milder extent) are temperature sensitive, indicative of an unstable protein (Figure 4.6). Furthermore, sensitivity to DNA-damaging agents such as camptothecin (Figure 4.10), bleocin (Figure 4.7), UV (Figure 4.8 and Figure 4.9) and MMS (Figure 4.11), was observed for these same variants. These

data go against a cell-free experiment where Pol2-P301R is able to by-pass secondary structure such as hairpin and to extend from a mismatch while the exonuclease dead version stays stalled at the replication fork (Xing *et al.*, 2019). Altogether, the divergence of results for *in vitro* and *in vivo* data suggests that there might be an additional cellular factor that might impede the proper function of Pol ε during replication *in vivo*.

Difficult-to-replicate sequences are obstacles the replisome is facing during DNA synthesis and TLS polymerases are suited to deal with these obstacles (Sale, 2013). The mutation rates of strains expressing altered Pol ε show a reduction in their spontaneous mutation rates when Pol ξ is depleted (Northam *et al.*, 2006). Although TLS polymerases were not demonstrated to be responsible for the hypermutated phenotype observed in Pol2-P301R (Xing *et al.*, 2019) one TLS polymerase – Pol κ – is not conserved in budding yeast, although it is present in mammalian cells and fission yeast. Depletion of Pol κ in a *cdc20*$^{exonull}$ strain restores its mutation rate to a wild-type level and lowers the mutation rate of *cdc20*$^{N364K}$ to a level of a *cdc20*$^{exonull}$ strain (Figure 4.16). A similar result was previously observed for the *cdc20*$^{P287R}$ strain (I. Soriano, unpublished results), suggesting that Pol κ might be partially responsible for the hypermutated phenotype observed in some *POLE*$^{EDMs}$. Careful regulation of polymerases switching during DNA replication is important as TLS polymerases are error-prone (Hedglin *et al.*, 2017; Waters *et al.*, 2009). This regulation is possible via PCNA ubiquitination and sumoylation which can be triggered by replication defects (Ogi *et al.*, 2010). In the case of *cdc20*$^{EDMs}$, it would be interesting to investigate whether PCNA plays a role in the mutagenic activity of Pol κ via a Pol κ noncanonical PIP-box and/or ubiquitination of the replicative sliding clamp.

The MMR pathway is the final guardian of the genome when replication errors occur. Previously published work showed synthetic lethality between *pol2-P301R* and *msh6Δ*, however not with *pol2-4* (L. N. Williams *et al.*, 2015; Xing *et al.*, 2019), illustrating the stronger mutagenic capacity of *POLE*$^{P286R}$ compared to simple lack of exonuclease activity. Our data confirm the fact that hypermutated strains rely on MMR for replication-errors correction. Indeed, on a short-term basis (Figure 5.1) neither the *cdc20*$^{N364K}$, nor the *cdc20*$^{Y459F}$ variants showed lethality with *msh2Δ*. However, accumulation of generations led to a lethal threshold for both *cdc20*$^{N364K}$ and *cdc20*$^{Y459F}$, as the first could not grow beyond 50 generations, and the second beyond 150 generations. Taken together, this further indicate that replication-errors are corrected by MMR and that *cdc20*$^{EDMs}$ have a distinct mutational power.

Specific mutational signatures were previously associated with *POLE*$^{EDMs}$ (Alexandrov *et al.*, 2020; Alexandrov, Nik-Zainal, Wedge, Aparicio*, et al.*, 2013; Haradhvala *et al.*, 2018; Hodel *et al.*, 2020). Against expectations, the signature 10 was not the prevalent signature observed the *cdc20*$^{EDM}$ variants (see section 5.2.6). However, the signatures 14 was detected in both *cdc20*$^{N364K}$ and *cdc20*$^{Y459F}$, signal amplified in absence of MMR (see section 5.2.6), confirming the association of the signature 14 to concurrent *POLE*$^{EDMs}$ and MMR-deficiency (Alexandrov *et al.*, 2020; Haradhvala *et al.*, 2018; Hodel *et al.*, 2020), but suggesting an intrinsic mutational signature for *cdc20*$^{EDMs}$. The signature 20 – associated with *POLD1*$^{EDMs}$ and MMR-deficiency (Alexandrov *et al.*, 2020) – detected in our study at a similar level as the signature 14 suggests a similar mutational process between exonuclease domain altered Pol ε and Pol δ. However, the common C>A transversion in a CCT context (Figure 1.8)

observed in both signature 14 and 20 indicate that the intrinsic mutational signature might be mainly characterised by this alteration.

Altogether, our data – which first aim was to evaluate the pathogenicity of germline *POLE*$^{EDM}$ variants using fission yeast – shows that *POLE*$^{M294I}$ is probably non-pathogenic as it exhibits a phenotype similar to a wild-type strain. The *POLE*$^{V411L}$ and *POLE*$^{L424V}$ seem to have a similar impairment as an exonuclease-dead POLE version, although the fact that *POLE*$^{V411L}$ is associated with the signature 10 and *POLE*$^{L424V}$ is associated with PPAP syndrome leads to question whether a second hit such as MMR-deficiency is required to trigger pathogenicity. Lastly, the *POLE*$^{N363K}$ and *POLE*$^{Y458F}$ clearly demonstrate pathogenic capacities, although different from each other, and to a different degree compared to the somatic *POLE*$^{P286R}$ variant.

Investigation of the pathogenic mechanism of the hypermutated germline variant *cdc20*$^{EDM}$ in fission yeast led to the following hypothesis: specific Pol ε or Pol δ exonuclease domain alterations lead to incorporation of replication-errors which would trigger cell cycle delay and checkpoint activation. Encountering difficult-to-replicate sequences and less prone to bypass these elements, the TLS polymerase Pol κ is recruited and being error-prone, incorporate an even greater number of mutations. This would result in the mutational signature 14 or 20, and maybe more precisely in the increase in C>A transversion in a CCT context. To test this hypothesis, mutation accumulation experiment in a *cdc20*$^{EDMs}$ *msh2Δ kpa1Δ* triple mutant and identification of its mutational patterns would clarify as whether Pol κ is responsible for the signature 14 and 20. In addition, DNA template synthesis using *POLE*$^{EDMs}$ variants in *in vitro* assays

would help to determine the 'true' *POLE*$^{EDM}$ signature. The same could be performed for Pol κ.

In conclusion, this study provides insight into the pathogenic mechanism of germline *POLE*$^{EDMs}$ variants and initiates potential lines of future investigations for *POLE*$^{EDM}$ targeted treatments.

## References


Abdus Sattar, A. K., Lin, T. C., Jones, C., & Konigsberg, W. H. (1996). Functional consequences and exonuclease kinetic parameters of point mutations in bacteriophage T4 DNA polymerase. *Biochemistry, 35*(51), 16621-16629. doi:10.1021/bi961552q

Ahmad, M., & Bussey, H. (1986). Yeast arginine permease: nucleotide sequence of the *CAN1* gene. *Curr Genet, 10*(8), 587-592. doi:10.1007/bf00418125

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2015). *Molecular Biology of the Cell* (6th ed.). United States of America: Garland Science Taylor & Francis Group.

Albertson, T. M., Ogawa, M., Bugni, J. M., Hays, L. E., Chen, Y., Wang, Y., . . . Preston, B. D. (2009). DNA polymerase epsilon and delta proofreading suppress discrete mutator and cancer phenotypes in mice. *Proc Natl Acad Sci U S A, 106*(40), 17101-17104. doi:10.1073/pnas.0907147106

Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Tian Ng, A. W., Wu, Y., . . . Consortium, P. (2020). The repertoire of mutational signatures in human cancer. *Nature, 578*(7793), 94-101. doi:10.1038/s41586-020-1943-3

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., . . . Stratton, M. R. (2013). Signatures of mutational processes in human cancer. *Nature, 500*(7463), 415-421. doi:10.1038/nature12477

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J., & Stratton, M. R. (2013). Deciphering signatures of mutational processes operative in human cancer. *Cell Rep, 3*(1), 246-259. doi:10.1016/j.celrep.2012.12.008

Aller, P., Duclos, S., Wallace, S. S., & Doublie, S. (2011). A crystallographic study of the role of sequence context in thymine glycol bypass by a replicative DNA polymerase serendipitously sheds light on the exonuclease complex. *J Mol Biol, 412*(1), 22-34. doi:10.1016/j.jmb.2011.07.007

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online].

Ang, X. L., & Wade Harper, J. (2005). SCF-mediated protein degradation and cell cycle control. *Oncogene, 24*(17), 2860-2870. doi:10.1038/sj.onc.1208614

Aparicio, O. M., Weinstein, D. M., & Bell, S. P. (1997). Components and dynamics of DNA replication complexes in S. cerevisiae: redistribution of MCM proteins and Cdc45p during S phase. *Cell, 91*(1), 59-69. doi:10.1016/s0092-8674(01)80009-x

Araki, H., Hamatake, R. K., Johnston, L. H., & Sugino, A. (1991). *DPB2*, the gene encoding DNA polymerase II subunit B, is required for chromosome replication in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A, 88*(11), 4601-4605. doi:10.1073/pnas.88.11.4601

Araki, H., Hamatake, R. K., Morrison, A., Johnson, A. L., Johnston, L. H., & Sugino, A. (1991). Cloning *DPB3*, the gene encoding the third subunit of DNA polymerase II of *Saccharomyces cerevisiae*. *Nucleic Acids Res, 19*(18), 4867-4872. doi:10.1093/nar/19.18.4867

Aria, V., & Yeeles, J. T. P. (2018). Mechanism of Bidirectional Leading-Strand Synthesis Establishment at Eukaryotic DNA Replication Origins. *Mol Cell*. doi:10.1016/j.molcel.2018.10.019

Asahara, H., Li, Y., Fuss, J., Haines, D. S., Vlatkovic, N., Boyd, M. T., & Linn, S. (2003). Stimulation of human DNA polymerase epsilon by MDM2. *Nucleic Acids Res, 31*(9), 2451-2459. doi:10.1093/nar/gkg342

Asturias, F. J., Cheung, I. K., Sabouri, N., Chilkova, O., Wepplo, D., & Johansson, E. (2006). Structure of *Saccharomyces cerevisiae* DNA polymerase epsilon by cryo-electron microscopy. *Nat Struct Mol Biol, 13*(1), 35-43. doi:10.1038/nsmb1040

Aye, Y., Li, M., Long, M. J., & Weiss, R. S. (2015). Ribonucleotide reductase and cancer: biological mechanisms and targeted therapies. *Oncogene, 34*(16), 2011-2021. doi:10.1038/onc.2014.155

Bahler, J., Wu, J. Q., Longtine, M. S., Shah, N. G., McKenzie, A., 3rd, Steever, A. B., . . . Pringle, J. R. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast, 14*(10), 943-951. doi:10.1002/(SICI)1097-0061(199807)14:10<943::AID-YEA292>3.0.CO;2-Y

Baptiste, B. A., & Eckert, K. A. (2012). DNA polymerase kappa microsatellite synthesis: two distinct mechanisms of slippage-mediated errors. *Environ Mol Mutagen, 53*(9), 787-796. doi:10.1002/em.21721

Barbari, S. R., Kane, D. P., Moore, E. A., & Shcherbakova, P. V. (2018). Functional Analysis of Cancer-Associated DNA Polymerase epsilon Variants in *Saccharomyces cerevisiae*. *G3 (Bethesda)*. doi:10.1534/g3.118.200042

Barbari, S. R., & Shcherbakova, P. V. (2017). Replicative DNA polymerase defects in human cancers: Consequences, mechanisms, and implications for therapy. *DNA Repair (Amst), 56*, 16-25. doi:10.1016/j.dnarep.2017.06.003

Baumann, H. P. R. a. C. A. (1939). Tumor Production in Mice with Ultraviolet Irradiation. *Cancer Research, 35*(1). doi:10.1158/ajc.1939.55

Beese, L. S., & Steitz, T. A. (1991). Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J, 10*(1), 25-33.

Behringer, M. G., & Hall, D. W. (2015). Genome-Wide Estimates of Mutation Rates and Spectrum in *Schizosaccharomyces pombe* Indicate CpG Sites are Highly Mutagenic Despite the Absence of DNA Methylation. *G3 (Bethesda), 6*(1), 149-160. doi:10.1534/g3.115.022129

Bellelli, R., Borel, V., Logan, C., Svendsen, J., Cox, D. E., Nye, E., . . . Boulton, S. J. (2018). Pol ε Instability Drives Replication Stress, Abnormal Development, and Tumorigenesis. *Mol Cell, 70*(4), 707-721 e707. doi:10.1016/j.molcel.2018.04.008

Bermudez, V. P., Farina, A., Raghavan, V., Tappin, I., & Hurwitz, J. (2011). Studies on human DNA polymerase epsilon and GINS complex and their role in DNA replication. *J Biol Chem, 286*(33), 28963-28977. doi:10.1074/jbc.M111.256289

Bernad, A., Blanco, L., Lazaro, J. M., Martin, G., & Salas, M. (1989). A conserved 3'----5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases. *Cell, 59*(1), 219-228.

Bester, A. C., Roniger, M., Oren, Y. S., Im, M. M., Sarni, D., Chaoat, M., . . . Kerem, B. (2011). Nucleotide deficiency promotes genomic instability in early stages of cancer development. *Cell, 145*(3), 435-446. doi:10.1016/j.cell.2011.03.044

Bhagavan, N. V., & Ha, C. (2015). *DNA Replication, Repair and Mutagenesis* (2nd ed.): Academic Press.

Bielas, J. H., Loeb, K. R., Rubin, B. P., True, L. D., & Loeb, L. A. (2006). Human cancers express a mutator phenotype. *Proc Natl Acad Sci U S A, 103*(48), 18238-18242. doi:10.1073/pnas.0607057103

Blokzijl, F., Janssen, R., van Boxtel, R., & Cuppen, E. (2018). MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med, 10*(1), 33. doi:10.1186/s13073-018-0539-0

Boehm, E. M., Gildenberg, M. S., & Washington, M. T. (2016). The Many Roles of PCNA in Eukaryotic DNA Replication. *Enzymes, 39*, 231-254. doi:10.1016/bs.enz.2016.03.003

Boland, C. R., Koi, M., Chang, D. K., & Carethers, J. M. (2008). The biochemical basis of microsatellite instability and abnormal immunohistochemistry and clinical behavior in Lynch syndrome: from bench to bedside. *Fam Cancer, 7*(1), 41-52. doi:10.1007/s10689-007-9145-9

Bowen, N., & Kolodner, R. D. (2017). Reconstitution of *Saccharomyces cerevisiae* DNA polymerase epsilon-dependent mismatch repair with purified proteins. *Proc Natl Acad Sci U S A, 114*(14), 3607-3612. doi:10.1073/pnas.1701753114

Braithwaite, D. K., & Ito, J. (1993). Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res, 21*(4), 787-802. doi:10.1093/nar/21.4.787

Brash, D. E., Rudolph, J. A., Simon, J. A., Lin, A., McKenna, G. J., Baden, H. P., . . . Ponten, J. (1991). A role for sunlight in skin cancer: UV-induced p53 mutations in squamous cell carcinoma. *Proc Natl Acad Sci U S A, 88*(22), 10124-10128. doi:10.1073/pnas.88.22.10124

Burgers, P. M. (1991). *Saccharomyces cerevisiae* replication factor C. II. Formation and activity of complexes with the proliferating cell nuclear antigen and with DNA polymerases delta and epsilon. *J Biol Chem, 266*(33), 22698-22706.

Burgers, P. M., Bambara, R. A., Campbell, J. L., Chang, L. M., Downey, K. M., Hubscher, U., . . . Spadari, S. (1990). Revised nomenclature for eukaryotic DNA polymerases. *Eur J Biochem, 191*(3), 617-618. doi:10.1111/j.1432-1033.1990.tb19165.x

Byrnes, J. J., Downey, K. M., Black, V. L., & So, A. G. (1976). A new mammalian DNA polymerase with 3' to 5' exonuclease activity: DNA polymerase delta. *Biochemistry, 15*(13), 2817-2823. doi:10.1021/bi00658a018

Cancer Genome Atlas, N. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature, 487*(7407), 330-337. doi:10.1038/nature11252

Cancer Genome Atlas Research, N., Kandoth, C., Schultz, N., Cherniack, A. D., Akbani, R., Liu, Y., . . . Levine, D. A. (2013). Integrated genomic characterization of endometrial carcinoma. *Nature, 497*(7447), 67-73. doi:10.1038/nature12113

Cannavo, E., Marra, G., Sabates-Bellver, J., Menigatti, M., Lipkin, S. M., Fischer, F., . . . Jiricny, J. (2005). Expression of the MutL homologue hMLH3 in human cells and its role in DNA mismatch repair. *Cancer Res, 65*(23), 10759-10766. doi:10.1158/0008-5472.CAN-05-2528

Caras, I. W., & Martin, D. W., Jr. (1988). Molecular cloning of the cDNA for a mutant mouse ribonucleotide reductase M1 that produces a dominant mutator phenotype in mammalian cells. *Mol Cell Biol, 8*(7), 2698-2704. doi:10.1128/mcb.8.7.2698

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., . . . Schultz, N. (2012). The cBio cancer genomics portal: an open platform for exploring

multidimensional cancer genomics data. *Cancer Discov, 2*(5), 401-404. doi:10.1158/2159-8290.CD-12-0095

Chabes, A., Georgieva, B., Domkin, V., Zhao, X., Rothstein, R., & Thelander, L. (2003). Survival of DNA damage in yeast directly depends on increased dNTP levels allowed by relaxed feedback inhibition of ribonucleotide reductase. *Cell, 112*(3), 391-401. doi:10.1016/s0092-8674(03)00075-8

Chabes, A. L., Pfleger, C. M., Kirschner, M. W., & Thelander, L. (2003). Mouse ribonucleotide reductase R2 protein: a new target for anaphase-promoting complex-Cdh1-mediated proteolysis. *Proc Natl Acad Sci U S A, 100*(7), 3925-3929. doi:10.1073/pnas.0330774100

Chen, J. S., Beckley, J. R., McDonald, N. A., Ren, L., Mangione, M., Jang, S. J., . . . Gould, K. L. (2014). Identification of new players in cell division, DNA damage response, and morphogenesis through construction of *Schizosaccharomyces pombe* deletion strains. *G3 (Bethesda), 5*(3), 361-370. doi:10.1534/g3.114.015701

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America, 12*(5), 5.

Chilkova, O., Jonsson, B. H., & Johansson, E. (2003). The quaternary structure of DNA polymerase epsilon from *Saccharomyces cerevisiae*. *J Biol Chem, 278*(16), 14082-14086. doi:10.1074/jbc.M211818200

Chilkova, O., Stenlund, P., Isoz, I., Stith, C. M., Grabowski, P., Lundstrom, E. B., . . . Johansson, E. (2007). 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 Res, 35*(19), 6588-6597. doi:10.1093/nar/gkm741

Chui, G., & Linn, S. (1995). Further characterization of HeLa DNA polymerase epsilon. *J Biol Chem, 270*(14), 7799-7808. doi:10.1074/jbc.270.14.7799

Church, D. N., Stelloo, E., Nout, R. A., Valtcheva, N., Depreeuw, J., ter Haar, N., . . . Bosse, T. (2015). Prognostic significance of *POLE* proofreading mutations in endometrial cancer. *J Natl Cancer Inst, 107*(1), 402. doi:10.1093/jnci/dju402

Clifford, R., Louis, T., Robbe, P., Ackroyd, S., Burns, A., Timbs, A. T., . . . Schuh, A. (2014). *SAMHD1* is mutated recurrently in chronic lymphocytic leukemia and is involved in response to DNA damage. *Blood, 123*(7), 1021-1031. doi:10.1182/blood-2013-04-490847

Cremona, C. A., Sarangi, P., Yang, Y., Hang, L. E., Rahman, S., & Zhao, X. (2012). Extensive DNA damage-induced sumoylation contributes to replication and repair and acts in addition to the mec1 checkpoint. *Mol Cell, 45*(3), 422-432. doi:10.1016/j.molcel.2011.11.028

Crute, J. J., Wahl, A. F., & Bambara, R. A. (1986). Purification and characterization of two new high molecular weight forms of DNA polymerase delta. *Biochemistry, 25*(1), 26-36. doi:10.1021/bi00349a005

Daigaku, Y., Keszthelyi, A., Muller, C. A., Miyabe, I., Brooks, T., Retkute, R., . . . Carr, A. M. (2015). A global profile of replicative polymerase usage. *Nat Struct Mol Biol, 22*(3), 192-198. doi:10.1038/nsmb.2962

Datta, A., Schmeits, J. L., Amin, N. S., Lau, P. J., Myung, K., & Kolodner, R. D. (2000). Checkpoint-dependent activation of mutagenic repair in *Saccharomyces cerevisiae pol3-01* mutants. *Mol Cell, 6*(3), 593-603. doi:10.1016/s1097-2765(00)00058-7

Decosterd, L. A., Cottin, E., Chen, X., Lejeune, F., Mirimanoff, R. O., Biollaz, J., & Coucke, P. A. (1999). Simultaneous determination of deoxyribonucleoside in the presence of ribonucleoside triphosphates in human carcinoma cells by high-performance liquid chromatography. *Anal Biochem, 270*(1), 59-68. doi:10.1006/abio.1999.4066

Deegan, T. D., & Diffley, J. F. (2016). MCM: one ring to rule them all. *Curr Opin Struct Biol, 37*, 145-151. doi:10.1016/j.sbi.2016.01.014

Deegan, T. D., Yeeles, J. T., & Diffley, J. F. (2016). Phosphopeptide binding by Sld3 links Dbf4-dependent kinase to MCM replicative helicase activation. *EMBO J, 35*(9), 961-973. doi:10.15252/embj.201593552

Derbyshire, V., Grindley, N. D., & Joyce, C. M. (1991). The 3'-5' exonuclease of DNA polymerase I of *Escherichia coli*: contribution of each amino acid at the active site to the reaction. *EMBO J, 10*(1), 17-24.

Derbyshire, V., Pinsonneault, J. K., & Joyce, C. M. (1995). Structure-function analysis of 3'-->5'-exonuclease of DNA polymerases. *Methods Enzymol, 262*, 363-385. doi:10.1016/0076-6879(95)62030-3

Dewar, J. M., Low, E., Mann, M., Raschle, M., & Walter, J. C. (2017). CRL2(Lrr1) promotes unloading of the vertebrate replisome from chromatin during replication termination. *Genes Dev, 31*(3), 275-290. doi:10.1101/gad.291799.116

Dewar, J. M., & Walter, J. C. (2017). Mechanisms of DNA replication termination. *Nat Rev Mol Cell Biol, 18*(8), 507-516. doi:10.1038/nrm.2017.42

Domingo, E., Freeman-Mills, L., Rayner, E., Glaire, M., Briggs, S., Vermeulen, L., . . . Church, D. N. (2016). Somatic *POLE* proofreading domain mutation, immune response, and prognosis in colorectal cancer: a retrospective, pooled biomarker study. *Lancet Gastroenterol Hepatol, 1*(3), 207-216. doi:10.1016/S2468-1253(16)30014-0

Dua, R., Edwards, S., Levy, D. L., & Campbell, J. L. (2000). Subunit interactions within the *Saccharomyces cerevisiae* DNA polymerase epsilon (pol epsilon ) complex. Demonstration of a dimeric pol epsilon. *J Biol Chem, 275*(37), 28816-28825. doi:10.1074/jbc.M002376200

Dua, R., Levy, D. L., & Campbell, J. L. (1998). Role of the putative zinc finger domain of *Saccharomyces cerevisiae* DNA polymerase epsilon in DNA replication and the S/M checkpoint pathway. *J Biol Chem, 273*(45), 30046-30055. doi:10.1074/jbc.273.45.30046

Dua, R., Levy, D. L., & Campbell, J. L. (1999). 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. *J Biol Chem, 274*(32), 22283-22288. doi:10.1074/jbc.274.32.22283

Dua, R., Levy, D. L., Li, C. M., Snow, P. M., & Campbell, J. L. (2002). *In vivo* reconstitution of *Saccharomyces cerevisiae* DNA polymerase epsilon in insect cells. Purification and characterization. *J Biol Chem, 277*(10), 7889-7896. doi:10.1074/jbc.M108546200

Duxin, J. P., Dewar, J. M., Yardimci, H., & Walter, J. C. (2014). Repair of a DNA-protein crosslink by replication-coupled proteolysis. *Cell, 159*(2), 346-357. doi:10.1016/j.cell.2014.09.024

Duzdevich, D., Warner, M. D., Ticau, S., Ivica, N. A., Bell, S. P., & Greene, E. C. (2015). The dynamics of eukaryotic replication initiation: origin specificity, licensing, and firing at the single-molecule level. *Mol Cell, 58*(3), 483-494. doi:10.1016/j.molcel.2015.03.017

Erson-Omay, E. Z., Caglayan, A. O., Schultz, N., Weinhold, N., Omay, S. B., Ozduman, K., . . . Gunel, M. (2015). Somatic *POLE* mutations cause an ultramutated giant cell high-grade glioma subtype with better prognosis. *Neuro Oncol, 17*(10), 1356-1364. doi:10.1093/neuonc/nov027

Ewels, P., Magnusson, M., Lundin, S., & Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics, 32*(19), 3047-3048. doi:10.1093/bioinformatics/btw354

Fantes, P. A., & Creanor, J. (1984). Canavanine resistance and the mechanism of arginine uptake in the fission yeast *Schizosaccharomyces pombe*. *J Gen Microbiol, 130*(12), 3265-3273. doi:10.1099/00221287-130-12-3265

Farlow, A., Long, H., Arnoux, S., Sung, W., Doak, T. G., Nordborg, M., & Lynch, M. (2015). The Spontaneous Mutation Rate in the Fission Yeast *Schizosaccharomyces pombe*. *Genetics, 201*(2), 737-744. doi:10.1534/genetics.115.177329

Feng, W., & D'Urso, G. (2001). *Schizosaccharomyces pombe* cells lacking the amino-terminal catalytic domains of DNA polymerase epsilon are viable but require the DNA damage checkpoint control. *Mol Cell Biol, 21*(14), 4495-4504. doi:10.1128/MCB.21.14.4495-4504.2001

Feng, W., Rodriguez-Menocal, L., Tolun, G., & D'Urso, G. (2003). *Schizosacchromyces pombe* Dpb2 binds to origin DNA early in S phase and is required for chromosomal DNA replication. *Mol Biol Cell, 14*(8), 3427-3436. doi:10.1091/mbc.e03-02-0088

Filee, J., Forterre, P., Sen-Lin, T., & Laurent, J. (2002). Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins. *J Mol Evol, 54*(6), 763-773. doi:10.1007/s00239-001-0078-x

Fleck, O., Vejrup-Hansen, R., Watson, A., Carr, A. M., Nielsen, O., & Holmberg, C. (2013). Spd1 accumulation causes genome instability independently of ribonucleotide reductase activity but functions to protect the genome when deoxynucleotide pools are elevated. *J Cell Sci, 126*(Pt 21), 4985-4994. doi:10.1242/jcs.132837

Focher, F., Spadari, S., Ginelli, B., Hottiger, M., Gassmann, M., & Hubscher, U. (1988). Calf thymus DNA polymerase delta: purification, biochemical and functional properties of the enzyme after its separation from DNA polymerase alpha, a DNA dependent ATPase and proliferating cell nuclear antigen. *Nucleic Acids Res, 16*(14A), 6279-6295. doi:10.1093/nar/16.14.6279

Forsburg, S. L., & Rhind, N. (2006). Basic methods for fission yeast. *Yeast, 23*(3), 173-183. doi:10.1002/yea.1347

Foster, P. L. (2006). Methods for determining spontaneous mutation rates. *Methods Enzymol, 409*, 195-213. doi:10.1016/S0076-6879(05)09012-9

Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R., & Steitz, T. A. (1988). Cocrystal structure of an editing complex of Klenow fragment with DNA. *Proc Natl Acad Sci U S A, 85*(23), 8924-8928. doi:10.1073/pnas.85.23.8924

Ganai, R. A., Bylund, G. O., & Johansson, E. (2015). Switching between polymerase and exonuclease sites in DNA polymerase epsilon. *Nucleic Acids Res, 43*(2), 932-942. doi:10.1093/nar/gku1353

Ganai, R. A., Osterman, P., & Johansson, E. (2015). Yeast DNA polymerase catalytic core and holoenzyme have comparable catalytic rates. *J Biol Chem, 290*(6), 3825-3835. doi:10.1074/jbc.M114.615278

Gangloff, S., Achaz, G., Francesconi, S., Villain, A., Miled, S., Denis, C., & Arcangioli, B. (2017). Quiescence unveils a novel mutational force in fission yeast. *Elife, 6*. doi:10.7554/eLife.27469

Garbacz, M. A., Lujan, S. A., Burkholder, A. B., Cox, P. B., Wu, Q., Zhou, Z. X., . . . Kunkel, T. A. (2018). Evidence that DNA polymerase delta contributes to initiating leading strand DNA replication in *Saccharomyces cerevisiae*. *Nat Commun, 9*(1), 858. doi:10.1038/s41467-018-03270-4

Georgescu, R., Yuan, Z., Bai, L., de Luna Almeida Santos, R., Sun, J., Zhang, D., . . . O'Donnell, M. E. (2017). Structure of eukaryotic CMG helicase at a replication fork and implications to replisome architecture and origin initiation. *Proc Natl Acad Sci U S A, 114*(5), E697-E706. doi:10.1073/pnas.1620500114

Georgescu, R. E., Langston, L., Yao, N. Y., Yurieva, O., Zhang, D., Finkelstein, J., . . . O'Donnell, M. E. (2014). Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nat Struct Mol Biol, 21*(8), 664-670. doi:10.1038/nsmb.2851

Gerlach, V. L., Aravind, L., Gotway, G., Schultz, R. A., Koonin, E. V., & Friedberg, E. C. (1999). Human and mouse homologs of *Escherichia coli* DinB (DNA polymerase IV), members of the UmuC/DinB superfamily. *Proc Natl Acad Sci U S A, 96*(21), 11922-11927. doi:10.1073/pnas.96.21.11922

Ghodgaonkar, M. M., Lazzaro, F., Olivera-Pimentel, M., Artola-Boran, M., Cejka, P., Reijns, M. A., . . . Jiricny, J. (2013). Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair. *Mol Cell, 50*(3), 323-332. doi:10.1016/j.molcel.2013.03.019

Goldsby, R. E., Hays, L. E., Chen, X., Olmsted, E. A., Slayton, W. B., Spangrude, G. J., & Preston, B. D. (2002). High incidence of epithelial cancers in mice deficient for DNA polymerase delta proofreading. *Proc Natl Acad Sci U S A, 99*(24), 15560-15565. doi:10.1073/pnas.232340999

Goldsby, R. E., Lawrence, N. A., Hays, L. E., Olmsted, E. A., Chen, X., Singh, M., & Preston, B. D. (2001). Defective DNA polymerase-delta proofreading causes cancer susceptibility in mice. *Nat Med, 7*(6), 638-639. doi:10.1038/88963

Gon, S., Napolitano, R., Rocha, W., Coulon, S., & Fuchs, R. P. (2011). Increase in dNTP pool size during the DNA damage response plays a key role in spontaneous and induced-mutagenesis in *Escherichia coli*. *Proc Natl Acad Sci U S A, 108*(48), 19311-19316. doi:10.1073/pnas.1113664108

Gonzalez-Perez, A., & Lopez-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet, 88*(4), 440-449. doi:10.1016/j.ajhg.2011.03.004

Grenson, M., Mousset, M., Wiame, J. M., & Bechet, J. (1966). Multiplicity of the amino acid permeases in *Saccharomyces cerevisiae*. I. Evidence for a specific arginine-transporting system. *Biochim Biophys Acta, 127*(2), 325-338. doi:10.1016/0304-4165(66)90387-4

Guarino, E., Salguero, I., & Kearsey, S. E. (2014). Cellular regulation of ribonucleotide reductase in eukaryotes. *Semin Cell Dev Biol, 30*, 97-103. doi:10.1016/j.semcdb.2014.03.030

Hamzaoui, N., Alarcon, F., Leulliot, N., Guimbaud, R., Buecher, B., Colas, C., . . . Clauser, E. (2020). Genetic, structural, and functional characterization of *POLE* polymerase proofreading variants allows cancer risk prediction. *Genet Med*. doi:10.1038/s41436-020-0828-z

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell, 100*(1), 57-70.

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell, 144*(5), 646-674. doi:10.1016/j.cell.2011.02.013

Handa, T., Kanke, M., Takahashi, T. S., Nakagawa, T., & Masukata, H. (2012). DNA polymerization-independent functions of DNA polymerase epsilon in assembly and progression of the replisome in fission yeast. *Mol Biol Cell, 23*(16), 3240-3253. doi:10.1091/mbc.E12-05-0339

Hansen, M. F., Johansen, J., Bjornevoll, I., Sylvander, A. E., Steinsbekk, K. S., Saetrom, P., . . . Sjursen, W. (2015). A novel *POLE* mutation associated with cancers of colon, pancreas, ovaries and small intestine. *Fam Cancer, 14*(3), 437-448. doi:10.1007/s10689-015-9803-2

Haradhvala, N. J., Kim, J., Maruvka, Y. E., Polak, P., Rosebrock, D., Livitz, D., . . . Getz, G. (2018). Distinct mutational signatures characterize concurrent loss of polymerase proofreading and mismatch repair. *Nat Commun, 9*(1), 1746. doi:10.1038/s41467-018-04002-4

Hayles, J., Wood, V., Jeffery, L., Hoe, K. L., Kim, D. U., Park, H. O., . . . Nurse, P. (2013). A genome-wide resource of cell cycle and cell shape genes of fission yeast. *Open Biol, 3*(5), 130053. doi:10.1098/rsob.130053

He, H., Li, Y., Dong, Q., Chang, A. Y., Gao, F., Chi, Z., . . . Li, F. (2017). Coordinated regulation of heterochromatin inheritance by Dpb3-Dpb4 complex. *Proc Natl Acad Sci U S A, 114*(47), 12524-12529. doi:10.1073/pnas.1712961114

Hedglin, M., & Benkovic, S. J. (2017). Eukaryotic Translesion DNA Synthesis on the Leading and Lagging Strands: Unique Detours around the Same Obstacle. *Chem Rev, 117*(12), 7857-7877. doi:10.1021/acs.chemrev.7b00046

Helleday, T., Eshtad, S., & Nik-Zainal, S. (2014). Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet, 15*(9), 585-598. doi:10.1038/nrg3729

Hentges, P., Van Driessche, B., Tafforeau, L., Vandenhaute, J., & Carr, A. M. (2005). Three novel antibiotic marker cassettes for gene disruption and marker switching in *Schizosaccharomyces pombe*. *Yeast, 22*(13), 1013-1019. doi:10.1002/yea.1291

Hodel, K. P., Sun, M. J. S., Ungerleider, N., Park, V. S., Williams, L. G., Bauer, D. L., . . . Pursell, Z. F. (2020). *POLE* Mutation Spectra Are Shaped by the Mutant Allele Identity, Its Abundance, and Mismatch Repair Status. *Mol Cell*. doi:10.1016/j.molcel.2020.05.012

Hogg, M., Osterman, P., Bylund, G. O., Ganai, R. A., Lundstrom, E. B., Sauer-Eriksson, A. E., & Johansson, E. (2014). Structural basis for processive DNA synthesis by yeast DNA polymerase varepsilon. *Nat Struct Mol Biol, 21*(1), 49-55. doi:10.1038/nsmb.2712

Holmes, J., Jr., Clark, S., & Modrich, P. (1990). Strand-specific mismatch correction in nuclear extracts of human and *Drosophila melanogaster* cell lines. *Proc Natl Acad Sci U S A, 87*(15), 5837-5841. doi:10.1073/pnas.87.15.5837

Hsieh, P., & Yamane, K. (2008). DNA mismatch repair: molecular mechanism, cancer, and ageing. *Mech Ageing Dev, 129*(7-8), 391-407. doi:10.1016/j.mad.2008.02.012

Hsieh, P., & Zhang, Y. (2017). The Devil is in the details for DNA mismatch repair. *Proc Natl Acad Sci U S A, 114*(14), 3552-3554. doi:10.1073/pnas.1702747114

Hu Fang, J. A. B., Rebecca C. Poulos, Riku Katainen, Lauri A., & Aaltonen, J. W. H. W., M. (2019). Mutational processes of distinct *POLE* exonuclease domain mutants drive an enrichment of a specific *TP53* mutation in colorectal cancer. *bioRxiv*.

Humphrey, T. (2000). DNA damage and cell cycle control in *Schizosaccharomyces pombe*. *Mutat Res, 451*(1-2), 211-226. doi:10.1016/s0027-5107(00)00051-8

Isoz, I., Persson, U., Volkov, K., & Johansson, E. (2012). The C-terminus of Dpb2 is required for interaction with Pol2 and for cell viability. *Nucleic Acids Res, 40*(22), 11545-11553. doi:10.1093/nar/gks880

Ito, J., & Braithwaite, D. K. (1991). Compilation and alignment of DNA polymerase sequences. *Nucleic Acids Res, 19*(15), 4045-4057. doi:10.1093/nar/19.15.4045

J. Sambrook, E. F. F., T. Maniatis. (1989). *Molecular cloning: a laboratory manual* (Vol. 2nd edition).

Jain, R., Rajashankar, K. R., Buku, A., Johnson, R. E., Prakash, L., Prakash, S., & Aggarwal, A. K. (2014). Crystal structure of yeast DNA polymerase epsilon catalytic domain. *PLoS One, 9*(4), e94835. doi:10.1371/journal.pone.0094835

Jaszczur, M., Flis, K., Rudzka, J., Kraszewska, J., Budd, M. E., Polaczek, P., . . . Fijalkowska, I. J. (2008). Dpb2p, a noncatalytic subunit of DNA polymerase epsilon, contributes to the fidelity of DNA replication in *Saccharomyces cerevisiae*. *Genetics, 178*(2), 633-647. doi:10.1534/genetics.107.082818

Jaszczur, M., Rudzka, J., Kraszewska, J., Flis, K., Polaczek, P., Campbell, J. L., . . . Jonczyk, P. (2009). Defective interaction between Pol2p and Dpb2p, subunits of DNA polymerase epsilon, contributes to a mutator phenotype in *Saccharomyces cerevisiae*. *Mutat Res, 669*(1-2), 27-35. doi:10.1016/j.mrfmmm.2009.04.012

Johnson, R. E., Washington, M. T., Haracska, L., Prakash, S., & Prakash, L. (2000). Eukaryotic polymerases iota and zeta act sequentially to bypass DNA lesions. *Nature, 406*(6799), 1015-1019. doi:10.1038/35023030

Kadyrov, F. A., Dzantiev, L., Constantin, N., & Modrich, P. (2006). Endonucleolytic function of MutLalpha in human mismatch repair. *Cell, 126*(2), 297-308. doi:10.1016/j.cell.2006.05.039

Kai, M., & Wang, T. S. (2003). Checkpoint activation regulates mutagenic translesion synthesis. *Genes Dev, 17*(1), 64-76. doi:10.1101/gad.1043203

Kane, D. P., & Shcherbakova, P. V. (2014). A common cancer-associated DNA polymerase epsilon mutation causes an exceptionally strong mutator phenotype, indicating fidelity defects distinct from loss of proofreading. *Cancer Res, 74*(7), 1895-1901. doi:10.1158/0008-5472.CAN-13-2892

Kesti, T., Flick, K., Keranen, S., Syvaoja, J. E., & Wittenberg, C. (1999). DNA polymerase epsilon catalytic domains are dispensable for DNA replication, DNA repair, and cell viability. *Mol Cell, 3*(5), 679-685. doi:10.1016/s1097-2765(00)80361-5

Kesti, T., McDonald, W. H., Yates, J. R., 3rd, & Wittenberg, C. (2004). Cell cycle-dependent phosphorylation of the DNA polymerase epsilon subunit, Dpb2, by the Cdc28 cyclin-dependent protein kinase. *J Biol Chem, 279*(14), 14245-14255. doi:10.1074/jbc.M313289200

Kinzler, K. W., & Vogelstein, B. (1997). Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature, 386*(6627), 761, 763. doi:10.1038/386761a0

Kohnken, R., Kodigepalli, K. M., & Wu, L. (2015). Regulation of deoxynucleotide metabolism in cancer: novel mechanisms and therapeutic implications. *Mol Cancer, 14*, 176. doi:10.1186/s12943-015-0446-6

Kornberg, A. (1957). Enzymatic synthesis of deoxyribonucleic acid. *Harvey Lect, 53*, 83-112.

Korona, D. A., Lecompte, K. G., & Pursell, Z. F. (2011). The high fidelity and unique error signature of human DNA polymerase epsilon. *Nucleic Acids Res, 39*(5), 1763-1773. doi:10.1093/nar/gkq1034

Kraszewska, J., Garbacz, M., Jonczyk, P., Fijalkowska, I. J., & Jaszczur, M. (2012). Defect of Dpb2p, a noncatalytic subunit of DNA polymerase varepsilon, promotes error prone replication of undamaged chromosomal DNA in *Saccharomyces cerevisiae*. *Mutat Res, 737*(1-2), 34-42. doi:10.1016/j.mrfmmm.2012.06.002

Kunkel, T. A. (1988). Exonucleolytic proofreading. *Cell, 53*(6), 837-840. doi:10.1016/s0092-8674(88)90189-4

Kunkel, T. A., & Erie, D. A. (2005). DNA mismatch repair. *Annu Rev Biochem, 74*, 681-710. doi:10.1146/annurev.biochem.74.082803.133243

Labib, K. (2010). How do Cdc7 and cyclin-dependent kinases trigger the initiation of chromosome replication in eukaryotic cells? *Genes Dev, 24*(12), 1208-1219. doi:10.1101/gad.1933010

Lahue, R. S., Au, K. G., & Modrich, P. (1989). DNA mismatch correction in a defined system. *Science, 245*(4914), 160-164. doi:10.1126/science.2665076

Lam, W. C., Van der Schans, E. J., Joyce, C. M., & Millar, D. P. (1998). Effects of mutations on the partitioning of DNA substrates between the polymerase and 3'-5' exonuclease sites of DNA polymerase I (Klenow fragment). *Biochemistry, 37*(6), 1513-1522. doi:10.1021/bi9720181

Lange, S. S., Takata, K., & Wood, R. D. (2011). DNA polymerases and cancer. *Nat Rev Cancer, 11*(2), 96-110. doi:10.1038/nrc2998

Langston, L. D., Zhang, D., Yurieva, O., Georgescu, R. E., Finkelstein, J., Yao, N. Y., . . . O'Donnell, M. E. (2014). CMG helicase and DNA polymerase epsilon form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication. *Proc Natl Acad Sci U S A, 111*(43), 15390-15395. doi:10.1073/pnas.1418334111

Lea, D. E., & Coulson, C. A. (1949). The distribution of the numbers of mutants in bacterial populations. *J Genet, 49*(3), 264-285. doi:10.1007/bf02986080

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature, 401*(6755), 788-791. doi:10.1038/44565

Lee, M. Y., Tan, C. K., So, A. G., & Downey, K. M. (1980). Purification of deoxyribonucleic acid polymerase delta from calf thymus: partial characterization of physical properties. *Biochemistry, 19*(10), 2096-2101. doi:10.1021/bi00551a015

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics, 26*(5), 589-595. doi:10.1093/bioinformatics/btp698

Li, H. D., Cuevas, I., Zhang, M., Lu, C., Alam, M. M., Fu, Y. X., . . . Castrillon, D. H. (2018). Polymerase-mediated ultramutagenesis in mice produces diverse cancers with high mutational load. *J Clin Invest*. doi:10.1172/JCI122095

Li, Y., Pursell, Z. F., & Linn, S. (2000). Identification and cloning of two histone fold motif-containing subunits of HeLa DNA polymerase epsilon. *J Biol Chem, 275*(40), 31554.

Limas, J. C., & Cook, J. G. (2019). Preparation for DNA replication: the key to a successful S phase. *FEBS Lett, 593*(20), 2853-2867. doi:10.1002/1873-3468.13619

Ling, H., Boudsocq, F., Woodgate, R., & Yang, W. (2001). Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell, 107*(1), 91-102. doi:10.1016/s0092-8674(01)00515-3

Lock, A., Rutherford, K., Harris, M. A., Hayles, J., Oliver, S. G., Bahler, J., & Wood, V. (2019). PomBase 2018: user-driven reimplementation of the fission yeast database provides rapid and intuitive access to diverse, interconnected information. *Nucleic Acids Res, 47*(D1), D821-D827. doi:10.1093/nar/gky961

Longley, M. J., Pierce, A. J., & Modrich, P. (1997). DNA polymerase delta is required for human mismatch repair in vitro. *J Biol Chem, 272*(16), 10917-10921. doi:10.1074/jbc.272.16.10917

Looke, M., Maloney, M. F., & Bell, S. P. (2017). Mcm10 regulates DNA replication elongation by stimulating the CMG replicative helicase. *Genes Dev, 31*(3), 291-305. doi:10.1101/gad.291336.116

Lou, H., Komata, M., Katou, Y., Guan, Z., Reis, C. C., Budd, M., . . . Campbell, J. L. (2008). Mrc1 and DNA polymerase epsilon function together in linking DNA replication and the S phase checkpoint. *Mol Cell, 32*(1), 106-117. doi:10.1016/j.molcel.2008.08.020

Lujan, S. A., Williams, J. S., Clausen, A. R., Clark, A. B., & Kunkel, T. A. (2013). Ribonucleotides are signals for mismatch repair of leading-strand replication errors. *Mol Cell, 50*(3), 437-443. doi:10.1016/j.molcel.2013.03.017

Luria, S. E., & Delbruck, M. (1943). Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics, 28*(6), 491-511.

Maric, M., Maculins, T., De Piccoli, G., & Labib, K. (2014). Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication. *Science, 346*(6208), 1253596. doi:10.1126/science.1253596

Mathews, C. K. (2015). Deoxyribonucleotide metabolism, mutagenesis and cancer. *Nat Rev Cancer, 15*(9), 528-539. doi:10.1038/nrc3981

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., . . . DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res, 20*(9), 1297-1303. doi:10.1101/gr.107524.110

Meng, X., Wei, L., Peng, X. P., & Zhao, X. (2019). Sumoylation of the DNA polymerase epsilon by the Smc5/6 complex contributes to DNA replication. *PLoS Genet, 15*(11), e1008426. doi:10.1371/journal.pgen.1008426

Mertz, T. M., Sharma, S., Chabes, A., & Shcherbakova, P. V. (2015). Colon cancer-associated mutator DNA polymerase delta variant causes expansion of dNTP pools increasing its own infidelity. *Proc Natl Acad Sci U S A, 112*(19), E2467-2476. doi:10.1073/pnas.1422934112

Meselson, M., & Stahl, F. W. (1958). The Replication of DNA in *Escherichia Coli*. *Proc Natl Acad Sci U S A, 44*(7), 671-682. doi:10.1073/pnas.44.7.671

Miller, J. A. (1970). Carcinogenesis by chemicals: an overview--G. H. A. Clowes memorial lecture. *Cancer Res, 30*(3), 559-576.

Minko, I. G., Harbut, M. B., Kozekov, I. D., Kozekova, A., Jakobs, P. M., Olson, S. B., . . . Lloyd, R. S. (2008). Role for DNA polymerase kappa in the processing of N2-N2-guanine interstrand cross-links. *J Biol Chem, 283*(25), 17075-17082. doi:10.1074/jbc.M801238200

Miyabe, I., Kunkel, T. A., & Carr, A. M. (2011). The major roles of DNA polymerases epsilon and delta at the eukaryotic replication fork are evolutionarily conserved. *PLoS Genet, 7*(12), e1002407. doi:10.1371/journal.pgen.1002407

Moreno, S., Klar, A., & Nurse, P. (1991). Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol, 194*, 795-823. doi:10.1016/0076-6879(91)94059-l

Morrison, A., Araki, H., Clark, A. B., Hamatake, R. K., & Sugino, A. (1990). A third essential DNA polymerase in *S. cerevisiae*. *Cell, 62*(6), 1143-1151.

Morrison, A., Bell, J. B., Kunkel, T. A., & Sugino, A. (1991). Eukaryotic DNA polymerase amino acid sequence required for 3'----5' exonuclease activity. *Proc Natl Acad Sci U S A, 88*(21), 9473-9477.

Morrison, A., Johnson, A. L., Johnston, L. H., & Sugino, A. (1993). Pathway correcting DNA replication errors in *Saccharomyces cerevisiae*. *EMBO J, 12*(4), 1467-1473.

Muramatsu, S., Hirai, K., Tak, Y. S., Kamimura, Y., & Araki, H. (2010). CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol $\varepsilon$, and GINS in budding yeast. *Genes Dev, 24*(6), 602-612. doi:10.1101/gad.1883410

Murphy, K., Darmawan, H., Schultz, A., Fidalgo da Silva, E., & Reha-Krantz, L. J. (2006). A method to select for mutator DNA polymerase deltas in *Saccharomyces cerevisiae*. *Genome, 49*(4), 403-410. doi:10.1139/g05-106

Navas, T. A., Sanchez, Y., & Elledge, S. J. (1996). *RAD9* and DNA polymerase epsilon form parallel sensory branches for transducing the DNA damage checkpoint signal in *Saccharomyces cerevisiae*. *Genes Dev, 10*(20), 2632-2643. doi:10.1101/gad.10.20.2632

Navas, T. A., Zhou, Z., & Elledge, S. J. (1995). DNA polymerase epsilon links the DNA replication machinery to the S phase checkpoint. *Cell, 80*(1), 29-39.

Nick McElhinny, S. A., Gordenin, D. A., Stith, C. M., Burgers, P. M., & Kunkel, T. A. (2008). Division of labor at the eukaryotic replication fork. *Mol Cell, 30*(2), 137-144. doi:10.1016/j.molcel.2008.02.022

Nick McElhinny, S. A., Kissling, G. E., & Kunkel, T. A. (2010). Differential correction of lagging-strand replication errors made by DNA polymerases {alpha} and {delta}. *Proc Natl Acad Sci U S A, 107*(49), 21070-21075. doi:10.1073/pnas.1013048107

Nick McElhinny, S. A., Kumar, D., Clark, A. B., Watt, D. L., Watts, B. E., Lundstrom, E. B., . . . Kunkel, T. A. (2010). Genome instability due to ribonucleotide incorporation into DNA. *Nat Chem Biol, 6*(10), 774-781. doi:10.1038/nchembio.424

Nick McElhinny, S. A., Watts, B. E., Kumar, D., Watt, D. L., Lundstrom, E. B., Burgers, P. M., . . . Kunkel, T. A. (2010). Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc Natl Acad Sci U S A, 107*(11), 4949-4954. doi:10.1073/pnas.0914857107

Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., . . . Breast Cancer Working Group of the International Cancer Genome, C. (2012). Mutational processes molding the genomes of 21 breast cancers. *Cell, 149*(5), 979-993. doi:10.1016/j.cell.2012.04.024

Northam, M. R., Garg, P., Baitin, D. M., Burgers, P. M., & Shcherbakova, P. V. (2006). A novel function of DNA polymerase zeta regulated by PCNA. *EMBO J, 25*(18), 4316-4325. doi:10.1038/sj.emboj.7601320

Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol, 302*(1), 205-217. doi:10.1006/jmbi.2000.4042

Ogi, T., Limsirichaikul, S., Overmeer, R. M., Volker, M., Takenaka, K., Cloney, R., . . . Lehmann, A. R. (2010). Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. *Mol Cell, 37*(5), 714-727. doi:10.1016/j.molcel.2010.02.009

Ohashi, E., Ogi, T., Kusumoto, R., Iwai, S., Masutani, C., Hanaoka, F., & Ohmori, H. (2000). Error-prone bypass of certain DNA lesions by the human DNA polymerase kappa. *Genes Dev, 14*(13), 1589-1594.

Ohmori, H., Friedberg, E. C., Fuchs, R. P., Goodman, M. F., Hanaoka, F., Hinkle, D., . . . Woodgate, R. (2001). The Y-family of DNA polymerases. *Mol Cell, 8*(1), 7-8. doi:10.1016/s1097-2765(01)00278-7

Ohya, T., Maki, S., Kawasaki, Y., & Sugino, A. (2000). Structure and function of the fourth subunit (Dpb4p) of DNA polymerase epsilon in *Saccharomyces cerevisiae*. *Nucleic Acids Res, 28*(20), 3846-3852.

Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., & Steitz, T. A. (1985). Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature, 313*(6005), 762-766. doi:10.1038/313762a0

Pai, C. C., & Kearsey, S. E. (2017). A Critical Balance: dNTPs and the Maintenance of Genome Stability. *Genes (Basel), 8*(2). doi:10.3390/genes8020057

Palles, C., Cazier, J. B., Howarth, K. M., Domingo, E., Jones, A. M., Broderick, P., . . . Tomlinson, I. (2013). Germline mutations affecting the proofreading domains of *POLE* and *POLD1* predispose to colorectal adenomas and carcinomas. *Nat Genet, 45*(2), 136-144. doi:10.1038/ng.2503

Parkash, V., Kulkarni, Y., Ter Beek, J., Shcherbakova, P. V., Kamerlin, S. C. L., & Johansson, E. (2019). Structural consequence of the most frequently recurring cancer-associated substitution in DNA polymerase epsilon. *Nat Commun, 10*(1), 373. doi:10.1038/s41467-018-08114-9

Pecorino, L. (2012). *Molecular Biology of Cancer: Machnisms, Targets and Therapeutics*: OUP Oxford.

Pena-Diaz, J., & Jiricny, J. (2010). PCNA and MutLalpha: partners in crime in triplet repeat expansion? *Proc Natl Acad Sci U S A, 107*(38), 16409-16410. doi:10.1073/pnas.1011692107

Peters, J. M. (1998). SCF and APC: the Yin and Yang of cell cycle regulated proteolysis. *Curr Opin Cell Biol, 10*(6), 759-768. doi:10.1016/s0955-0674(98)80119-1

Petersen, J., & Russell, P. (2016). Growth and the Environment of *Schizosaccharomyces pombe*. *Cold Spring Harb Protoc, 2016*(3), pdb top079764. doi:10.1101/pdb.top079764

Pluciennik, A., Dzantiev, L., Iyer, R. R., Constantin, N., Kadyrov, F. A., & Modrich, P. (2010). PCNA function in the activation and strand direction of MutLalpha endonuclease in mismatch repair. *Proc Natl Acad Sci U S A, 107*(37), 16066-16071. doi:10.1073/pnas.1010662107

Pursell, Z. F., Isoz, I., Lundstrom, E. B., Johansson, E., & Kunkel, T. A. (2007). Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science, 317*(5834), 127-130. doi:10.1126/science.1144067

Pursell, Z. F., & Kunkel, T. A. (2008). DNA polymerase epsilon: a polymerase of unusual size (and complexity). *Prog Nucleic Acid Res Mol Biol, 82*, 101-145. doi:10.1016/S0079-6603(08)00004-4

Randell, J. C., Fan, A., Chan, C., Francis, L. I., Heller, R. C., Galani, K., & Bell, S. P. (2010). Mec1 is one of multiple kinases that prime the Mcm2-7 helicase for phosphorylation by Cdc7. *Mol Cell, 40*(3), 353-363. doi:10.1016/j.molcel.2010.10.017

Rayner, E., van Gool, I. C., Palles, C., Kearsey, S. E., Bosse, T., Tomlinson, I., & Church, D. N. (2016). A panoply of errors: polymerase proofreading domain mutations in cancer. *Nat Rev Cancer, 16*(2), 71-81. doi:10.1038/nrc.2015.12

Richterich, P. (1998). Estimation of errors in "raw" DNA sequences: a validation study. *Genome Res, 8*(3), 251-259. doi:10.1101/gr.8.3.251

Rogers, R. F., Walton, M. I., Cherry, D. L., Collins, I., Clarke, P. A., Garrett, M. D., & Workman, P. (2020). CHK1 inhibition is synthetically lethal with loss of B-family DNA polymerase function in human lung and colorectal cancer cells. *Cancer Res*. doi:10.1158/0008-5472.CAN-19-1372

Rohlin, A., Zagoras, T., Nilsson, S., Lundstam, U., Wahlstrom, J., Hulten, L., . . . Nordling, M. (2014). A mutation in *POLE* predisposing to a multi-tumour phenotype. *Int J Oncol, 45*(1), 77-81. doi:10.3892/ijo.2014.2410

Rosche, W. A., & Foster, P. L. (2000). Determining mutation rates in bacterial populations. *Methods, 20*(1), 4-17. doi:10.1006/meth.1999.0901

Ryan Poplin, V. R.-R., Mark A. DePristo, Tim J. Fennell, Mauricio O. Carneiro, Geraldine A. Van der Auwera, David E. Kling, Laura D. Gauthier, Ami Levy-Moonshine, David Roazen, Khalid Shakir, Joel Thibault, Sheila Chandran, Chris Whelan, Monkol Lek, Stacey Gabriel, Mark J. Daly, Benjamin Neale, Daniel G. MacArthur, Eric Banks. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*. doi:https://doi.org/10.1101/201178

Sale, J. E. (2013). Translesion DNA synthesis and mutagenesis in eukaryotes. *Cold Spring Harb Perspect Biol, 5*(3), a012708. doi:10.1101/cshperspect.a012708

Sameer, A. S., Nissar, S., & Fatima, K. (2014). Mismatch repair pathway: molecules, functions, and role in colorectal carcinogenesis. *Eur J Cancer Prev, 23*(4), 246-257. doi:10.1097/CEJ.0000000000000019

Sarkar, S., Ma, W. T., & Sandri, G. H. (1992). On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica, 85*(2), 173-179.

Schmidt, T. T., Sharma, S., Reyes, G. X., Gries, K., Gross, M., Zhao, B., . . . Hombauer, H. (2019). A genetic screen pinpoints ribonucleotide reductase residues that sustain dNTP homeostasis and specifies a highly mutagenic type of dNTP imbalance. *Nucleic Acids Res, 47*(1), 237-252. doi:10.1093/nar/gky1154

Sengupta, S., van Deursen, F., de Piccoli, G., & Labib, K. (2013). Dpb2 integrates the leading-strand DNA polymerase into the eukaryotic replisome. *Curr Biol, 23*(7), 543-552. doi:10.1016/j.cub.2013.02.011

Seshagiri, S., Stawiski, E. W., Durinck, S., Modrusan, Z., Storm, E. E., Conboy, C. B., . . . de Sauvage, F. J. (2012). Recurrent R-spondin fusions in colon cancer. *Nature, 488*(7413), 660-664. doi:10.1038/nature11282

Setlow, R. B., & Carrier, W. L. (1966). Pyrimidine dimers in ultraviolet-irradiated DNA's. *J Mol Biol, 17*(1), 237-254. doi:10.1016/s0022-2836(66)80105-5

Shamoo, Y., & Steitz, T. A. (1999). Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell, 99*(2), 155-166. doi:10.1016/s0092-8674(00)81647-5

Shcherbakova, P. V., & Fijalkowska, I. J. (2006). Translesion synthesis DNA polymerases and control of genome stability. *Front Biosci, 11*, 2496-2517. doi:10.2741/1985

Shcherbakova, P. V., Pavlov, Y. I., Chilkova, O., Rogozin, I. B., Johansson, E., & Kunkel, T. A. (2003). Unique error signature of the four-subunit yeast DNA polymerase epsilon. *J Biol Chem, 278*(44), 43770-43780. doi:10.1074/jbc.M306893200

Sheu, Y. J., & Stillman, B. (2010). The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. *Nature, 463*(7277), 113-117. doi:10.1038/nature08647

Shevelev, I. V., & Hubscher, U. (2002). The 3' 5' exonucleases. *Nat Rev Mol Cell Biol, 3*(5), 364-376. doi:10.1038/nrm804

Shinbrot, E., Henninger, E. E., Weinhold, N., Covington, K. R., Goksenin, A. Y., Schultz, N., . . . Wheeler, D. A. (2014). Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. *Genome Res, 24*(11), 1740-1750. doi:10.1101/gr.174789.114

Shlien, A., Campbell, B. B., de Borja, R., Alexandrov, L. B., Merico, D., Wedge, D., . . . Biallelic Mismatch Repair Deficiency, C. (2015). Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. *Nat Genet, 47*(3), 257-262. doi:10.1038/ng.3202

Siddiqui, K., On, K. F., & Diffley, J. F. (2013). Regulating DNA replication in eukarya. *Cold Spring Harb Perspect Biol, 5*(9). doi:10.1101/cshperspect.a012930

Silvian, L. F., Toth, E. A., Pham, P., Goodman, M. F., & Ellenberger, T. (2001). Crystal structure of a DinB family error-prone DNA polymerase from *Sulfolobus solfataricus*. *Nat Struct Biol, 8*(11), 984-989. doi:10.1038/nsb1101-984

Sipiczki, M. (2000). Where does fission yeast sit on the tree of life? *Genome Biol, 1*(2), REVIEWS1011. doi:10.1186/gb-2000-1-2-reviews1011

Slade, D. (2018). Maneuvers on PCNA Rings during DNA Replication and Repair. *Genes (Basel), 9*(8). doi:10.3390/genes9080416

Soengas, M. S., Esteban, J. A., Lazaro, J. M., Bernad, A., Blasco, M. A., Salas, M., & Blanco, L. (1992). Site-directed mutagenesis at the Exo III motif of phi 29 DNA polymerase; overlapping structural domains for the 3'-5' exonuclease and strand-displacement activities. *EMBO J, 11*(11), 4227-4237.

Spiga, M. G., & D'Urso, G. (2004). Identification and cloning of two putative subunits of DNA polymerase epsilon in fission yeast. *Nucleic Acids Res, 32*(16), 4945-4953. doi:10.1093/nar/gkh811

Stancel, J. N., McDaniel, L. D., Velasco, S., Richardson, J., Guo, C., & Friedberg, E. C. (2009). Polk mutant mice have a spontaneous mutator phenotype. *DNA Repair (Amst), 8*(12), 1355-1362. doi:10.1016/j.dnarep.2009.09.003

Stocki, S. A., Nonay, R. L., & Reha-Krantz, L. J. (1995). Dynamics of bacteriophage T4 DNA polymerase function: identification of amino acid residues that affect switching

between polymerase and 3' --> 5' exonuclease activities. *J Mol Biol, 254*(1), 15-28. doi:10.1006/jmbi.1995.0595

Sudakin, V., Ganoth, D., Dahan, A., Heller, H., Hershko, J., Luca, F. C., . . . Hershko, A. (1995). The cyclosome, a large complex containing cyclin-selective ubiquitin ligase activity, targets cyclins for destruction at the end of mitosis. *Mol Biol Cell, 6*(2), 185-197. doi:10.1091/mbc.6.2.185

Sun, J., Evrin, C., Samel, S. A., Fernandez-Cid, A., Riera, A., Kawakami, H., . . . Li, H. (2013). Cryo-EM structure of a helicase loading intermediate containing ORC-Cdc6-Cdt1-MCM2-7 bound to DNA. *Nat Struct Mol Biol, 20*(8), 944-951. doi:10.1038/nsmb.2629

Sun, J., Shi, Y., Georgescu, R. E., Yuan, Z., Chait, B. T., Li, H., & O'Donnell, M. E. (2015). The architecture of a eukaryotic replisome. *Nat Struct Mol Biol, 22*(12), 976-982. doi:10.1038/nsmb.3113

Swan, M. K., Johnson, R. E., Prakash, L., Prakash, S., & Aggarwal, A. K. (2009). Structural basis of high-fidelity DNA synthesis by yeast DNA polymerase delta. *Nat Struct Mol Biol, 16*(9), 979-986. doi:10.1038/nsmb.1663

Syvaoja, J. E. (1990). DNA polymerase epsilon: the latest member in the family of mammalian DNA polymerases. *Bioessays, 12*(11), 533-536. doi:10.1002/bies.950121106

Tahirov, T. H., Makarova, K. S., Rogozin, I. B., Pavlov, Y. I., & Koonin, E. V. (2009). Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol epsilon and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors. *Biol Direct, 4*, 11. doi:10.1186/1745-6150-4-11

Tanaka, S., & Araki, H. (2010). Regulation of the initiation step of DNA replication by cyclin-dependent kinases. *Chromosoma, 119*(6), 565-574. doi:10.1007/s00412-010-0291-8

Tanaka, S., & Araki, H. (2013). Helicase activation and establishment of replication forks at chromosomal origins of replication. *Cold Spring Harb Perspect Biol, 5*(12), a010371. doi:10.1101/cshperspect.a010371

Tanaka, S., Komeda, Y., Umemori, T., Kubota, Y., Takisawa, H., & Araki, H. (2013). Efficient initiation of DNA replication in eukaryotes requires Dpb11/TopBP1-GINS interaction. *Mol Cell Biol, 33*(13), 2614-2622. doi:10.1128/MCB.00431-13

Tanaka, S., Umemori, T., Hirai, K., Muramatsu, S., Kamimura, Y., & Araki, H. (2007). CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature, 445*(7125), 328-332. doi:10.1038/nature05465

Terai, K., Abbas, T., Jazaeri, A. A., & Dutta, A. (2010). CRL4(Cdt2) E3 ubiquitin ligase monoubiquitinates PCNA to promote translesion DNA synthesis. *Mol Cell, 37*(1), 143-149. doi:10.1016/j.molcel.2009.12.018

Thomas, D. C., Roberts, J. D., & Kunkel, T. A. (1991). Heteroduplex repair in extracts of human HeLa cells. *J Biol Chem, 266*(6), 3744-3751.

Ticau, S., Friedman, L. J., Champasa, K., Correa, I. R., Jr., Gelles, J., & Bell, S. P. (2017). Mechanism and timing of Mcm2-7 ring closure during DNA replication origin licensing. *Nat Struct Mol Biol, 24*(3), 309-315. doi:10.1038/nsmb.3375

Ticau, S., Friedman, L. J., Ivica, N. A., Gelles, J., & Bell, S. P. (2015). Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. *Cell, 161*(3), 513-525. doi:10.1016/j.cell.2015.03.012

Tran, P. T., Simon, J. A., & Liskay, R. M. (2001). Interactions of Exo1p with components of MutLalpha in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A, 98*(17), 9760-9765. doi:10.1073/pnas.161175998

Vande Perre, P., Siegfried, A., Corsini, C., Bonnet, D., Toulas, C., Hamzaoui, N., . . . Guimbaud, R. (2018). Germline mutation p.N363K in POLE is associated with an increased risk of colorectal cancer and giant cell glioblastoma. *Fam Cancer*. doi:10.1007/s10689-018-0102-6

Vlatkovic, N., Guerrera, S., Li, Y., Linn, S., Haines, D. S., & Boyd, M. T. (2000). MDM2 interacts with the C-terminus of the catalytic subunit of DNA polymerase epsilon. *Nucleic Acids Res, 28*(18), 3581-3586. doi:10.1093/nar/28.18.3581

Voitenleitner, C., Fanning, E., & Nasheuer, H. P. (1997). Phosphorylation of DNA polymerase alpha-primase by cyclin A-dependent kinases regulates initiation of DNA replication in vitro. *Oncogene, 14*(13), 1611-1615. doi:10.1038/sj.onc.1200975

Wach, A., Brachat, A., Alberti-Segui, C., Rebischung, C., & Philippsen, P. (1997). Heterologous HIS3 marker and GFP reporter modules for PCR-targeting in *Saccharomyces cerevisiae*. *Yeast, 13*(11), 1065-1075. doi:10.1002/(SICI)1097-0061(19970915)13:11<1065::AID-YEA159>3.0.CO;2-K

Wagner, J., Gruz, P., Kim, S. R., Yamada, M., Matsui, K., Fuchs, R. P., & Nohmi, T. (1999). The *dinB* gene encodes a novel *E. coli* DNA polymerase, DNA pol IV, involved in mutagenesis. *Mol Cell, 4*(2), 281-286. doi:10.1016/s1097-2765(00)80376-7

Wang, J., Carvajal-Carmona, L. G., Chu, J. H., Zauber, A. G., Collaborators, A. P. C. T., Kubo, M., . . . Bertagnolli, M. M. (2013). Germline variants and advanced colorectal adenomas: adenoma prevention with celecoxib trial genome-wide association study. *Clin Cancer Res, 19*(23), 6430-6437. doi:10.1158/1078-0432.CCR-13-0550

Wang, J., Yu, P., Lin, T. C., Konigsberg, W. H., & Steitz, T. A. (1996). Crystal structures of an NH2-terminal fragment of T4 DNA polymerase and its complexes with single-stranded DNA and with divalent metal ions. *Biochemistry, 35*(25), 8110-8119. doi:10.1021/bi960178r

Waters, L. S., Minesinger, B. K., Wiltrout, M. E., D'Souza, S., Woodruff, R. V., & Walker, G. C. (2009). Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol Mol Biol Rev, 73*(1), 134-154. doi:10.1128/MMBR.00034-08

Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature, 171*(4356), 737-738. doi:10.1038/171737a0

Weissbach, A., Baltimore, D., Bollum, F., Gallo, R., & Korn, D. (1975). Nomenclature of eukaryotic DNA polymerases. *Science, 190*(4212), 401-402. doi:10.1126/science.1179222

Williams, J. S., Clausen, A. R., Nick McElhinny, S. A., Watts, B. E., Johansson, E., & Kunkel, T. A. (2012). Proofreading of ribonucleotides inserted into DNA by yeast DNA polymerase varepsilon. *DNA Repair (Amst), 11*(8), 649-656. doi:10.1016/j.dnarep.2012.05.004

Williams, J. S., Lujan, S. A., & Kunkel, T. A. (2016). Processing ribonucleotides incorporated during eukaryotic DNA replication. *Nat Rev Mol Cell Biol, 17*(6), 350-363. doi:10.1038/nrm.2016.37

Williams, L. N., Marjavaara, L., Knowels, G. M., Schultz, E. M., Fox, E. J., Chabes, A., & Herr, A. J. (2015). dNTP pool levels modulate mutator phenotypes of error-prone DNA polymerase epsilon variants. *Proc Natl Acad Sci U S A, 112*(19), E2457-2466. doi:10.1073/pnas.1422948112

Wimmer, K., Beilken, A., Nustede, R., Ripperger, T., Lamottke, B., Ure, B., . . . Kratz, C. P. (2017). A novel germline *POLE* mutation causes an early onset cancer prone syndrome mimicking constitutional mismatch repair deficiency. *Fam Cancer, 16*(1), 67-71. doi:10.1007/s10689-016-9925-1

Winczura, A., Appanah, R., Tatham, M. H., Hay, R. T., & De Piccoli, G. (2019). The S phase checkpoint promotes the Smc5/6 complex dependent SUMOylation of Pol2, the catalytic subunit of DNA polymerase epsilon. *PLoS Genet, 15*(11), e1008427. doi:10.1371/journal.pgen.1008427

Witkin, E. M. (1969). Ultraviolet-induced mutation and DNA repair. *Annu Rev Microbiol, 23*, 487-514. doi:10.1146/annurev.mi.23.100169.002415

Wood, V., Gwilliam, R., Rajandream, M. A., Lyne, M., Lyne, R., Stewart, A., . . . Nurse, P. (2002). The genome sequence of *Schizosaccharomyces pombe*. *Nature, 415*(6874), 871-880. doi:10.1038/nature724

Wu, Z., Nagano, I., & Takahashi, Y. (1998). The detection of Trichinella with polymerase chain reaction (PCR) primers constructed using sequences of random amplified polymorphic DNA (RAPD) or sequences of complementary DNA encoding excretory-secretory (E-S) glycoproteins. *Parasitology, 117 ( Pt 2)*, 173-183. doi:10.1017/s0031182098002881

Xing, X., Kane, D. P., Bulock, C. R., Moore, E. A., Sharma, S., Chabes, A., & Shcherbakova, P. V. (2019). A recurrent cancer-associated substitution in DNA polymerase epsilon produces a hyperactive enzyme. *Nat Commun, 10*(1), 374. doi:10.1038/s41467-018-08145-2

Yeeles, J. T., Deegan, T. D., Janska, A., Early, A., & Diffley, J. F. (2015). Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature, 519*(7544), 431-435. doi:10.1038/nature14285

Yeeles, J. T. P., Janska, A., Early, A., & Diffley, J. F. X. (2017). How the Eukaryotic Replisome Achieves Rapid and Efficient DNA Replication. *Mol Cell, 65*(1), 105-116. doi:10.1016/j.molcel.2016.11.017

Yoneda, M., & Bollum, F. J. (1965). Deoxynucleotide-Polymerizing Enzymes of Calf Thymus Gland. I. Large Scale Purification of Terminal and Replicative Deoxynucleotidyl Transferases. *J Biol Chem, 240*, 3385-3391.

Yonemasu, R., McCready, S. J., Murray, J. M., Osman, F., Takao, M., Yamamoto, K., . . . Yasui, A. (1997). Characterization of the alternative excision repair pathway of UV-damaged DNA in *Schizosaccharomyces pombe*. *Nucleic Acids Res, 25*(8), 1553-1558. doi:10.1093/nar/25.8.1553

Yoon, J. H., Bhatia, G., Prakash, S., & Prakash, L. (2010). Error-free replicative bypass of thymine glycol by the combined action of DNA polymerases kappa and zeta in human cells. *Proc Natl Acad Sci U S A, 107*(32), 14116-14121. doi:10.1073/pnas.1007795107

Yuan, Z., Georgescu, R., Schauer, G. D., O'Donnell, M. E., & Li, H. (2020). Structure of the polymerase epsilon holoenzyme and atomic model of the leading strand replisome. *Nat Commun, 11*(1), 3156. doi:10.1038/s41467-020-16910-5

Yuan, Z., Riera, A., Bai, L., Sun, J., Nandi, S., Spanos, C., . . . Li, H. (2017). Structural basis of Mcm2-7 replicative helicase loading by ORC-Cdc6 and Cdt1. *Nat Struct Mol Biol, 24*(3), 316-324. doi:10.1038/nsmb.3372

Zegerman, P., & Diffley, J. F. (2007). Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature, 445*(7125), 281-285. doi:10.1038/nature05432

Zhang, J., Hulme, L., & Liu, J. L. (2014). Asymmetric inheritance of cytoophidia in *Schizosaccharomyces pombe*. *Biol Open, 3*(11), 1092-1097. doi:10.1242/bio.20149613

Zheng, Q. (2002). Statistical and algorithmic methods for fluctuation analysis with SALVADOR as an implementation. *Math Biosci, 176*(2), 237-252. doi:10.1016/s0025-5564(02)00087-1

Zheng, Q. (2017). rSalvador: An R Package for the Fluctuation Experiment. *G3 (Bethesda), 7*(12), 3849-3856. doi:10.1534/g3.117.300120

Zhou, B. L., Pata, J. D., & Steitz, T. A. (2001). Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain. *Mol Cell, 8*(2), 427-437. doi:10.1016/s1097-2765(01)00310-0

Zhou, Z., & Elledge, S. J. (1993). *DUN1* encodes a protein kinase that controls the DNA damage response in yeast. *Cell, 75*(6), 1119-1127. doi:10.1016/0092-8674(93)90321-g

Zhou, Z. X., Lujan, S. A., Burkholder, A. B., Garbacz, M. A., & Kunkel, T. A. (2019). Roles for DNA polymerase delta in initiating and terminating leading strand DNA replication. *Nat Commun, 10*(1), 3992. doi:10.1038/s41467-019-11995-z

Zou, L., & Stillman, B. (1998). Formation of a preinitiation complex by S-phase cyclin CDK-dependent loading of Cdc45p onto chromatin. *Science, 280*(5363), 593-596. doi:10.1126/science.280.5363.593

Zou, Y., Liu, Y., Wu, X., & Shell, S. M. (2006). Functions of human replication protein A (RPA): from DNA replication to DNA damage and stress responses. *J Cell Physiol, 208*(2), 267-273. doi:10.1002/jcp.20622

# Appendices

## 7.1 Appendix I: Media Preparation

Media were autoclaved using the Classic Prestige Medical autoclave.

For agar media, 22.5 g/L of agar were added to liquid media.

### 7.1.1 Luria Bertani (LB)

10 g/L bacto-tryptone

5 g/L g yeast extract

10 g/L g NaCl

### 7.1.2 Yeast Extract plus 3 Supplement (YE3S)

5 g/L g of Yeast Extract

30 g/L g of glucose

0.2 g/L of amino acid (supplements leucine, adenine and uracil)

### 7.1.3 Pombe Glutamate Medium (PMG)

3 g/L potassium hydrogen phthallate

2.2 g/L $Na_2HPO4$

3.75 g/L L-glutamic acid monosodium salt

20 g/L glucose

0.2 g/L of amino acid (supplement)

1X salt stock

1X vitamins

1X minerals

1 mM NaCl

### 7.1.4 Edinburgh minimal medium (EMM)

3 g/L potassium hydrogen phtallate ($C_8H_5KO_4$)

2.2 g/L disodium phosphate ($Na_2HPO4$)

5 g/L ammonium chloride ($NH_4Cl$)

20 g/L glucose

0.2 g/L of amino acid (supplement)

1X salt stock

1X vitamins

1X minerals

1 mM NaCl

### 7.1.5 Edinburgh minimal medium -nitrogen (EMM-N)

3 g/L potassium hydrogen phtallate ($C_8H_5KO_4$)

2.2 g/L disodium phosphate ($Na_2HPO4$)

20 g/L glucose

1X salt stock

1X vitamins

1X minerals

1 mM NaCl

### 7.1.6 Salt stock solution (50X)

52.5 g/L magnesium chloride hexahydrate ($MgCl_2$)

0.74 g/L calcium chloride dihydrate ($CaCl_2$)

50 g/L potassium chloride (KCl)

2 g/L sodium sulfate ($Na_2SO_4$)

Autoclave and store at room temperature.

### 7.1.7 Vitamins stock solution (1000X)

1 g/L sodium pantothenate ($C_9H_{16}NNaO_5$)

10 g/L nicotinic acid ($C_6H_5NO_2$)

10 g/L myo-inositol ($C_6H_{12}O_6$)

10 mg/L biotin

Autoclave and store at 4°C.

### 7.1.8 Minerals stock solution (1000X)

5 g/L boric acid ($H_3BO_3$)

4 g/L manganese sulfate monohydrate ($MnSO_4$)

4 g/L zinc sulfate heptahydrate ($ZnSO_4$)

2 g/L iron chloride hexahydrate ($FeCl_2$)

0.4 g/L molybdic acid ($H_2MBO_4$)

1 g/L potassium iodide (KI)

0.4 g/L copper sulfate ($CuSO_4$)

10 g/L citric acid

Autoclave and store at 4°C.