Chapter 5

Discussion and future directions

5.1 Whole-genome sequencing as a flexible tool to address problems in cell biology

In this thesis, I have used whole-genome sequencing of model organisms to address questions in cell biology in DNA repair and replication. In the first part of this work, we have successfully used whole-genome sequencing to identify suppressor mutations in synthetic viability screens. In this type of experiment a selectable phenotype, usually due to a mutation, is alleviated by a second mutation. This allows inferences about a relationship between the two mutations and between the second mutation and the phenotype. These types of genetic interactions can be more informative than synthetic lethality, which sometimes arises due to the inactivation of two important, but unrelated pathways. While this type of screen has been utilised to uncover genetic interactions for decades, the identification of the secondary suppressor mutation is often labrious and time-consuming. The work in this thesis has shown that whole-genome sequencing can be utilised to correctly identify a suppressor mutation and that follow-up of these suppressors can yield relevant biological insight[801]. Currently, Dr. Fabio Puddu and I are validating suppressor mutations identified in a third *Saccharomyces cerevisiae* suppressor screen looking at proteins involved in replication stress response. Exploiting recent advances in culturing haploid mouse cells has also allowed us to extend this work to mammalian systems and demonstrate that the technique works to identify known suppressors to 6-thioguanine sensitivity as a proof-of-principle[1124]. Currently, we are extending this work to identify mutations alleviating the sensitivity to other chemicals of interest.

Suppressors identified in yeast usually arose without the need for mutagenesis, which com-

plicates the identification of suppressor mutations. However, it is known that the genetic background influences the pattern of spontaneously arising mutations, which may influence and limit the kind of suppressor that can be identified. Such differences in mutation patterns are the key interest in a nation-wide multi-institute project that the Jackson and Adams group are involved in (COMSIG). In an attempt to understand mutational processes in budding yeast, our group is identifying mutational patterns caused by deletion of any yeast gene.

As such we have propagated the *S. cerevisiae* gene deletion collection for a defined period of time and are sequencing strains to identify mutations acquired in that time frame. By adapting the analysis protocol I have developed for budding yeast genetic screens, I was able to identify acquired mutations in such mutation accumulation experiments. The acquired mutations will uncover mutational patterns and will generate a dataset from which mutational signatures can be extracted as it has been successfully demonstrated for human cancers [761–764]. It is expected that a catalogue correlating genetic defects and mutation patterns will in the future assist with elucidating the history and aetiology of cancer samples.

5.2 Polymerase mutations as drivers of mutagenesis

As part of the overarching effort to identify patterns of mutations associated with the loss of particular genes, I focused my attention on DNA polymerases. Tasked with duplicating the entire genome, they are prime candidates for sources of mutagenesis and recent work has identified mutations in DNA polymerase δ and ε as factors predisposing to familial colorectal cancer[768]. Further work identified more mutations in these DNA polymerases[768, 809, 810], but failed to identify which mutations had an impact on mutagenesis and which did not.

To examine the global effects of mutated DNA polymerases on genome stability, I used mutation accumulation experiments, propagating yeast strains carrying DNA polymerase mutations for a fixed amount of time. Whole-genome sequencing every sample at the beginning and end of the experiment allowed us to obtain lists of mutations accumulated by each sample. The design of the experiment was aided by similar experiments carried out in the group of Alain Nicolas, whose work with different budding yeast mutant provided information on the numbers of mutations expected in a wild-type strain[835]. Having used both mutation accumulation experiments and classic genetic reporter assays, I find that there is some agreement between these two methods, similar to what has been observed by Alain Nicolas[835]. However, whole-genome sequencing provides less variable data and more information, making it the better choice for this work.

Bioinformatic predictions and frequencies of mutations in the COSMIC dataset were used to prioritise three mutations in the human replicative polymerases, *POLE* S297F, *POLE* P286R and *POLE* V411L, and mutation accumulation experiments in *Saccharomyces cerevisiae* showed significant mutation increases for two of these mutations, *pol2-S312F* and *pol2*-P301R, which correspond to the human *POLE* S297F and *POLE* P286R, respectively. Other polymerase mutations conferring increases in mutation numbers are *pol2-L439V, pol2-M459K* and *pol3-S483N.*

Intriguingly, while the *POLE* V411L variant is the most commonly observed mutation, among the mutations studied in this thesis, in sequenced cancers, the budding yeast equivalent resulted in only a small, 1.4-fold increase over wild-type in diploid cells. As a comparison, the second most common mutation - *POLE* P286R (*pol2-P301R*) - resulted in a 27-fold increase over wild-type. This could either be due to a difference between the yeast and human version of *POLE* V411L or suggest that mutation frequency in cancers are not necessarily predictors of the severity of the resulting phenotype. It is also possible that the *POLE* V411L mutation promotes tumourigenesis by a manner other than mutation rate increases.

Exonuclease deficient strains, which are expected to show mutation rate increases, were included as a reference. The *pol3-01* and *pol2-4* alleles result in mutations of two acidic amino acids, involved in metal ion coordination, affecting proofreading, but not the polymerase activity of the encoded proteins. Considering that candidate polymerase mutations are located in the exonuclease domain and should affect the exonuclease activity, we expected mutation number increases to fall between wild-type and *pol3-01* or *pol2-4* strains. Surprisingly, the increases observed for *pol3-01* and *pol2-4* heterozygous diploid cells, were only 1.7- and 3.3-fold over wild type, respectively, meaning that every mutator strain identfied in this work showed mutation increases exceeding those observed in the corresponding exouclease deficient strain.

Recently, some of these findings have been validated by work from another group using classical reporter gene assays on strains carrying *pol2*-P301R or *pol2-4* mutations[888]. Here, mutation rate estimates for *pol2-P301R* also exceeded those from *pol2-4* strains.

How these polymerase mutations exert their mutagenic potential is a question that remains open. One possibility is that the *pol3-01* and *pol2-4* alleles may not be truly proofreading deficient. While this would explain how other mutations in the exonuclease domain could be more deleterious, it is unlikely since these alleles have been studied extensively *in vivo* and *in*

vitro [282, 303, 312, 338, 347, 807] .

Another possibility is that instead of a reduced exonuclease activity, these polymerase mutant strains actually have a hyperactive exonuclease, leading to removal of correctly paired nucleotides and idling.This could also explain why mutating the catalytic residues of the *POL2* exonuclease domain to alanine alleviated the mutator phenotype of strong mutators as for example *pol2-P301R*. However, it is possible that combining polymerase mutations with mutations in the exonuclease catalytic residues results in structural changes not present in the initial mutant protein. Thus, I have not excluded the possibility that the reduction in the mutator phenotype severity is due to the loss of exonuclease catalytic activity, specifically. Indeed, a polymerase that excises correctly paired nuclotides would lead to decreased processivity. In such a case mutagenesis could arise from a less accurate DNA polymerase having increased access to the replication fork to compensate for the less processive replicative polymerase.

As recent work has placed Polζ in the replisome[885, 886] and indicates it can take over for Polεin cases of destabilizing mutations in its subunits[883, 884], the catalytic subunit of Polζ, *REV3*, was a natural target for our work to identify the source of mutagenesis in polymerase mutant strains. Unlike the *dpb2* mutagenesis, which seems to be *REV3*-dependent, the *pol2-P301R*-dependent mutagenesis is potentiated in the absence of Rev3. Thus, it seems that, if anything, Rev3 is protective against *pol2-P301R*-dependent mutagenesis rather than introducing mutations. One hypothesis could be, that the missense mutation in the exonuclease domain does not just affect proofreading accuracy, but causes hyperactivity which leads to processivity decreases. A less processive, stalling polymerase could then be switched out more often in the replisome for Polζ, which as recent research indicates also plays a role in the replication of undamaged DNA, thus decreasing the access of the mutated polymerase epsilon to DNA during replication. If the mutagenesis is not entirely due to decreased proofreading - as indicated by mutating the exonuclease catalytic residues in mutated polymerase genes - and if it is not due to Polζ, the question remains which process introduces these mutations. It is possible that yet another polymerase is responsible, but it could also be that they are due to a loss of fidelity in the polymerase active site of Polε. The exonuclease and polymerase active site are on the same polypeptide and it is possible that a point mutation in one domain also affects the activity of the other. To test whether this occurs, I propose generating a mutated Polε, that carries the *P301R* mutation as well as mutations inactivating the catalytic activity of the polymerase domain. Since mutations in the catalytic residues of POL2 have been reported lethal[279], it is likley that this construct would also be lethal in haploid yeast cells, which is

why this work would be carried out in heterozygous diploids. If successful, these experiments could determine whether misincorporation, rather than deficient proofreading, by Polεcauses the increased mutagenesis in Polε mutants.

Beyond this, it is known that the composition and concentration of the dNTP pool is correlated with mutation rates[889]. While difficult to explain it is conceivable that an altered polymerase can lead to imbalances in nucleotide pools, which are known to affect the accuracy of DNA replication. In fact, recent work has shown that mutagenesis due to DNA polymerase mutations in the polymerase domain of the protein depends on *DUN1*, which is known to stimulate ribonucleotide reductase (RNR) activity, which in turn is responsible for precise regulation of dNTP pools[890–892]. In this model, defective polymerases lead to an accumulation of incomplete replication intermediates, which in turn leads to checkpoint activation[892]. Checkpoint activation increases dNTP levels via an activation of Dun1. At these increased dNTP levels, a mutated DNA polymerase will more readily extend the incomplete termini and likely make more misinsertions. While a lot of this work has focused on mutations in the polymerase domain of DNA polymerases, it will be interesting to explore if a similar mechanism opperates in our exoncuelase domain mutated strains. Considering that dNTP pool levels seem to correlate with polymerase mutator severity, targeting dNTP pools could be a target for therapy of polymerase-mutated cancers[891].

To test the interactions between the mutator phenotypes caused by mismatch repair (MMR) deficiency and that caused by polymerase mutations, we tried to obtain double mutants to examine how mutation numbers and two distinctive mutational patterns interact. By doing so, we found that *pol2-P301R, pol2-S312F* and *pol2-M459K* were lethal in combination with a deletion in *MSH2*, a key mismatch repair player. Similar results have been obtained for simultaneous loss of mismatch repair and exonuclease activity by others in haploid yeast and mice[338, 769, 893, 894]. In yeast, recent work has pointed to a threshold of mutation rates that are acceptable: any higher mutation rate results in replication error-induced extinction (EEX)[895, 896].

Interestingly, it has been reported that the phenoype of exonuclease domain mutations found in cancer depends MMR deficiency[897] and that mutation of *MSH2* and *MSH6* mutations in addition to exonuclease polymerase mutations is a common event[806]. In fact, there are known cases of children with inherited biallelic mismatch repair deficiency that acquired early somatic driver mutations in DNA polymerase ε or δ[898]. Of those polymerase mutations identified in the childrens' brain tumors one, *POLE S297F*, is included in this work as

well. Its yeast equivalent, *pol2-S312F*, is however lethal with *msh2*Δ in haploid cells. This discrepancy could either be due to the MMR status of the cancer cell, or to the fact that the yeast cells were haploid and did not have a wild-type DNA polymerase. Equally, it is possible that the cancer cells have acquired suppressor mutations that allow for this otherwise lethal combination.

The acquired mutations in polymerase mutant strains were used to visualise trinucleotide mutational patterns in the strongest mutators, *pol3-S483N* and *pol2-P301R,* and comapre them to the wild type. While the former shows a pattern fairly similar to the wild-type, the latter shows three distinct peaks among other more subtle differences to the wild-type. Mutational signature extraction predicts three signatures in the data and signatures obtained by different alogrithms produce similar, but subtly different results. While similarities between the human Signature 10 and the signature extracted from yeast polymerase ε mutants can be seen, there are striking differences between the two as well. That being said, COSMIC signatures are extracted from an amalgamation of patterns in vastly more mutational data. While the 8815 mutations used for signature extraction in this work are sufficient to extract signatures, the human cancers COSMIC signatures are derived from are based on hundreds of thousands of mutations. That and inherent difference between human and yeast genomes and mutational processes can account for these differences. Similar differences can be seen between mutational patterns in the yeast strains and the human cancer samples reported by Shibrot et al. [899]: while the *POLE-Pro286Arg* mutated samples do show similarly low relative levels of C:G>G:C mutations and high relative levels of C:G>A:T mutations, the human samples show very low contributions of T:A>A:T mutations, while in the yeast strains they contribute approximately a quarter of all mutations to the overall mutational pattern. They report that TCG→TTG and TCT→TAT mutations account for >50% of the mutations found. However, in the yeast samples this is not the case: the TCT→TAT is one of the most common changes, but it accounts for less than 10% of all changes, while the TCG→TTG mutation is not common in the mutated yeast cells. Again, this could be due to differences in observed mutation numbers, inherent biological differences between human and yeast or the fact that the human cancers with 100 mutations per Mb will have acquired more mutations which can further contribute to the mutation pattern. Further work will show whether these differences hold true when mutation accumulation experimenst are performed with mammalian cell lines.

Due to the limitations of variant calling algorithms when it comes to analyse repetitive sequences, acquired mutations were only identifed from non-repetitive sequences. However,

repetitive regions can provide valuable information such as copy number estimates for rDNA repeats.

In this thesis, I detailed our approach to estimate rDNA repeat number from whole-genome sequencing data. Validation using strains of known rDNA copy number shows that our approach estimates copy number as accurately if not better than Southern blotting. While we could not identify striking deviations from copy numbers in wild-type strains, this technique could be used to identify as of yet unknown regulators of rDNA copy number in budding yeast.

5.3 Future directions

While this work provides initial insights into the effects of a collection of DNA polymerase mutants, further work will address how exactly these mutations cause mutation rate increases. For example it will be interesting to explore the difference between haploid, heterozygous and homozygous diploid mutants. Furthermore, it will be important to measure protein levels in heterozygous diploid cells to determine whether different mutation rates result from different ratios of wild-type versus mutated proteins. Additionally, the production of recombinant polymerases willallow us to determine whether these have defect in polymerisation, exonuclease activity or processivity *in vitro*.

To further validate the synthetic lethality between polymerase mutants and mismatch repair deficiency, I am planning to confirm it using plasmid eviction of a *MSH2*-carrying plasmid from a strain carrying both a polymerase mutation and a genomic deletion of *MSH2*. To define a possible mutation load threshold, combinations of other polymerase mutations, such as the *pol2-L439V*, and mismatch repair deficiency will be attempted in haploid cells. Considering their severe mutator phenotypes in single mutant cells, haploid double mutants of *pol2-P301R* and *pol3-S483N* may be inviable.

Indeed, if lethality arises from the increased mutation rate leading to an increased chance of mutating essential genes, mutation combinations that are lethal in a haploid background, could be viable in a diploid background. For instance, homozygous *msh2*Δ combined with heterozygous *pol2-S312F* could be viable, which would be in agreement with the existence of cancer cells with this genotype. Thus, using plasmid eviction I will attempt to construct diploid cells that are mismatch repair deficient and carry one of the polymerase mutations.

Considering our success with suppressor screens, we are also considering a screen to identify mutations that suppress or enhance the strong mutation rates of some polymerase mutations. The former could identfiy the manner in which these mutations operate, the latter could

identify targets for a synthetic lethality approach to treatment of affected patients. Suppressors for the synthetic lethality of mismatch repair deficiency and some polymerase mutations could also be aimed for.

While much of my work on mutation rates and signature has focused on polymerase mutants, it is entirely possible to extend this work to virtually any budding yeast mutant. In fact, as part of the COMSIG consortium, we will screen the entire yeast gene deletion collection for genes likely to regulate rDNA copy number maintenance. Beyond that we are on track to identify mutator phenotypes and mutational signatures for a wide array of nuclear gene deletions.

In summary, there are many questions we are looking forward to answer relating to the mutagenic potential of mutated DNA polymerases. We expect to uncover more answers by exploring genetic interactions with other mutations and examining key DNA polymerase mutations in mammalian systems. Additionally, our work will investigate the mutagenic potential of hundreds of genes and their associated mutational signatures. Hopefully, as we accumulate more information about how mutated proteins or their absence shapes a cell's genome we will learn more about fundamental biological processes and the contributions such altered proteins can make to a cancer genome.