Ribonucleotides in DNA

- application in genome-wide DNA polymerase tracking and physiological role in eukaryotes

Katrin Kreisel

Department of Medical Biochemistry and Cell Biology
Institute of Biomedicine
Sahlgrenska Academy, University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2021

Cover illustration: ‘Ribonucleotides in DNA’ created by Katrin Kreisel using BioRender.com.

Ribonucleotides in DNA - application in genome-wide DNA polymerase tracking and physiological role in eukaryotes

katrin.kreisel@gu.se

ISBN 978-91-8009-184-8 (PRINT)
ISBN 978-91-8009-185-5 (PDF)
http://hdl.handle.net/2077/68065

Printed in Borås, Sweden 2021
Printed by Stema Specialtryck AB

Mit dem Wissen wächst der Zweifel.

~ Johann Wolfgang von Goethe

Ribonucleotides in DNA

- application in genome-wide DNA polymerase tracking and physiological role in eukaryotes

Katrin Kreisel

Department of Medical Biochemistry and Cell Biology
Institute of Biomedicine
Sahlgrenska Academy, University of Gothenburg
Gothenburg, Sweden

ABSTRACT

The genetic code in the eukaryotic cell is stored in the form of DNA, which is more resistant to hydrolysis than RNA. Replication fidelity and DNA repair mechanisms are in place to ensure genomic integrity to preserve the information encoded. Despite DNA polymerases' discrimination against ribonucleotides, they are frequently incorporated into DNA and even in the presence of efficient ribonucleotide removal pathways, ribonucleotides may remain stably incorporated in the DNA.

Ribonucleotides can be used as a marker of DNA replication enzymology by using HydEn-seq, a next-generation sequencing technique for the genome-wide mapping of ribonucleotides. I aimed to elucidate the activities of the specialized translesion synthesis DNA polymerase η in yeast. By using a steric gate variant that incorporates more ribonucleotides and by tracking those ribonucleotides, I determined a lagging strand preference dependent on its C-terminus in Paper I. The findings suggest a possible extension of the 'division of labor' among replicative polymerases to the specialized polymerases.

Moreover, I was interested in the physiological role of incorporated ribonucleotides and used an extension of the HydEn-seq method outlined in Paper II, to map and quantitate ribonucleotides simultaneously. By investigating ribonucleotide incorporation into mouse mitochondrial DNA (mtDNA) in Paper III, we found that ribonucleotides are acquired mostly up until adulthood and are not connected to age-related mtDNA instability, suggesting relatively good tolerance of incorporated ribonucleotides in mtDNA.

To gain a more comprehensive view on incorporated ribonucleotides in the DNA of mammals, I mapped and quantitated incorporated ribonucleotides in nuclear DNA (nDNA) and mtDNA from murine blood, bone marrow, brain, heart, kidney, liver, lung, muscle and spleen in Paper IV. I found tissue-dependent variations in the number and the identity of incorporated ribonucleotides and marked differences between nDNA and mtDNA. The ribonucleotide distribution in both types of DNA was non-random and in nDNA affected by the proximity of genomic features, which in most cases increased the number of embedded ribonucleotides locally as compared to random positions in the nDNA.

The thesis extends the knowledge of DNA polymerase η's activity and the physiological role that incorporated ribonucleotides play in DNA. This more detailed characterization of the incorporated ribonucleotides genome-wide is a basic requirement for the understanding of diseases associated with genome instability, such as certain types of cancers or Aicardi-Goutières syndrome.

Keywords: Ribonucleotides, DNA instability, DNA polymerase eta, nuclear DNA, mitochondrial DNA

ISBN 978-91-8009-184-8 (PRINT)
ISBN 978-91-8009-185-5 (PDF)

http://hdl.handle.net/2077/68065

SAMMANFATTNING

Den genetiska koden i eukaryota celler lagras i form av DNA, vilket är stabilare än RNA och mindre känsligt för hydrolys. Replikationsnoggranhet och mekanismer för DNA reparation upprätthåller genomets integritet och säkerställer att replikeringen av DNA sker korrekt. Trots att DNA-polymeraser, vilka replikerar DNA:t, vanligtvis kan särskilja mellan deoxyribonukleotider (DNA:s byggstenar) och ribonukleotider (RNA:s byggstenar), inkorporeras ibland ribonukleotider i DNA-strängen som inte alltid tas bort av de processer som ska upptäcka och ta bort dessa. Dessa ribonukleotider inkorporeras då stabilt i DNA-strängen och blir kvar.

Ribonukleotider i DNA kan användas för att kartlägga DNA-polymerasernas enzymologi. Genom att använda en specialiserad sekvenseringsmetod (HydEn-seq), som kartlägger inkorporerade ribonukleotider i hela genomet, var mitt mål att fastställa aktiviteten hos DNA polymeras η (pol η), vilket är ett specialiserat translesionssyntes-polymeras i jäst. Genom att försämra pol η förmåga att välja bort ribonukleotider under DNA syntesen kunde jag fastställa att pol η är mest aktiv på DNA-strängen som byggs diskontinuerligt, den så kallade ”lagging strand”. Fyndet, vilket redovisas i delarbete I, implicerar att ”fördelningen av arbetskraft” man talar om mellan de replikerande polymeraserna i viss utsträckning kanske även gäller för de specialiserade polymeraserna.

I delarbete II använde jag en modifierad version av HydEn-seq som möjliggör både kartläggning och kvantifiering av ribonukleotider i genomet samtidigt, för att undersöka vilken fysiologisk roll de inkorporerade ribonukleotiderna har. När vi i delarbete III undersökte mitokondriellt DNA (mtDNA) från möss i varierade åldrar, kunde vi konstatera, att åldersrelaterad genominstabilitet inte orsakas av felaktigt inkorporerade ribonukleotider vilket tyder på att ribonukleotider i mtDNA är vältolererade.

För att ytterligare förstå vilken roll inkorporerade ribonukleotider spelar i däggdjurs DNA, kartlade och kvantifierade jag inkorporerade ribonukleotider i både nukleärt DNA (nDNA) och mtDNA från blod, benmärg, hjärna, hjärta, lever, lunga, mjälte, muskel och njure från mus i delarbete IV. Både antalet ribonukeotider och vilken basidentitet dessa hade varierade mellan olika vävnader och skiljde sig tydligt mellan mtDNA och nDNA i samma vävnad. Förekomsten av inkorporerade ribonukleotider var icke-slumpmässig, i nDNA ökade ofta förekomsten av

inkorporerade ribonukleotider runt områden med genomisk funktion, jämfört med slumpmässigt utvalda områden i det nukleära genomet.

Sammanfattningsvis bidrar avhandlingens resultat till utökad kunskap om DNA polymeras η aktiviteten och den fysiologiska roll inkorporerade ribonukleotider spelar för genomets integritet, vilket är grundläggande för att förstå sjukdomar associerade med genominstabilitet så som vissa typer av cancer och Aicardi-Goutières syndrom.

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. **Kreisel, K**, Engqvist, MKM, Kalm, J, Thompson, LJ, Boström, M, Navarrete, C, McDonald, JP, Larsson, E, Woodgate, R, Clausen, AR. DNA polymerase η contributes to genome-wide lagging strand synthesis. *Nucleic Acids Research*, 2019; 47(5): 2425-2435

II. **Kreisel, K**, Engqvist, MKM, Clausen, AR. Simultaneous mapping and quantitation of ribonucleotides in human mitochondrial DNA. *Journal of Visualized Experiments* 2017; 129: e56551

III. Wanrooij, PH, Tran, P, Thompson, LJ, Carvalho, G, Sharma, S, **Kreisel, K**, Navarrete, C, Feldberg, A, Watt, DL, Nilsson AK, Engqvist, MKM, Clausen, AR, Chabes, A. Elimination of rNMPs from mitochondrial DNA has no effect on its stability. *Proceedings of the National Academy of Sciences of the United States of America* 2020; 117(25): 14306-14313

IV. **Kreisel, K**, Kalm, J, Bandaru, S, Ala, C, Akyürek, L, Clausen, AR. Stably incorporated ribonucleotides in murine tissues: quantitation, base identity and distribution in nuclear and mitochondrial DNA. *(to be submitted)*

CONTENT

ABBREVIATIONS

A	Adenine
AGS	Aicardi-Goutières syndrome
AMP	Adenosine monophosphate
AP-site	Apurinic/apyrimidinic site
ATP	Adenosine triphosphate
BER	Base Excision Repair
C	Cytosine
CMG	Cdc45-MCM-GINS
CMP	Cytidine monophosphate
CPD	Cyclobutane pyrimidine dimer
D-loop	Displacement loop
dAMP	Deoxyadenosine monophosphate
dATP	Deoxyadenosine triphosphate
dCMP	Deoxycytidine monophosphate
dGMP	Deoxyguanosine monophosphate
DNA	Deoxyribonucleic acid
dNMP	Deoxyribonucleoside monophosphate
dNTP	Deoxyribonucleoside triphosphate
DSB	Double strand break
dTMP	Deoxythymidine monophosphate
Exo1	Exonuclease 1
FEN1	Flap Endonuclease 1
G	Guanine
G4	G-quadruplex
GMP	Guanosine monophosphate
HR	Homologous Recombination
ICL	Interstrand crosslink
MCM	Minichromosome Maintenance
MGME1	Mitochondrial Genome Maintenance Exonuclease 1
MMR	Mismatch Repair
MSH	MutS Homolog
mtDNA	Mitochondrial DNA
mtSSB	Mitochondrial single-stranded DNA-binding protein
nDNA	Nuclear DNA
NER	Nucleotide Excision Repair
NHEJ	Non-Homologous End-Joining
NTP	Nucleoside triphosphate
OriH	Origin of heavy strand synthesis

OriL	Origin of light strand synthesis
8-oxoG	7,8-dihydro-8-oxo-deoxyguanine
PARP	Poly(ADP-ribose) polymerase
Pol	Polymerase
6-4PP	Pyrimidine (6-4) pyrimidone photoproducts
R-loop	D-loop-like structure with an RNA transcript
RER	Ribonucleotide Excision Repair
RNA	Ribonucleic acid
RNS	Reactive nitrogen species
rNTP	Ribonucleoside triphosphate
ROS	Reactive oxygen species
RPA	Replication Protein A
rRNA	Ribosomal RNA
SSB	Single strand break
SSBR	Single Strand Break Repair
ssDNA	Single-stranded DNA
T	Thymine
TLS	Translesion synthesis
Top1	Topoisomerase 1
Top2	Topoisomerase 2
tRNA	Transfer RNA
TSS	Transcription start site
UMP	Uridine monophosphate
UV	Ultraviolet
XP	*Xeroderma pigmentosum*

1 INTRODUCTION

Despite having been discovered over 150 years ago[1], deoxyribonucleic acid (DNA), the central hereditary molecule of all known life forms[2] and connected molecular machineries, that replicate, repair and transcribe it, remain to be fully understood even today. As the body of knowledge grows, new mechanisms are discovered that either promote or impede genome stability[3]. In turn, based on genome instability or impairment of appropriate repair processes, mechanisms connected to aging and disease[4-6] are uncovered. In this thesis, aspects of genome replication and instability involving incorporated ribonucleotides in *Saccharomyces cerevisiae*, *Mus musculus* and *Homo sapiens* (henceforth called yeast, mouse and human, respectively) genomes were studied.

1.1 DNA

Deoxyribonucleic acid is the central hereditary molecule in all living cells[2]. With the exception of mature erythrocytes and cornified cells like hair and nails where the previously present DNA is degraded in a controlled manner[7,8], each living cell receives and maintains a copy of the full genetic code[2]. DNA was first isolated and documented in 1869 by Friedrich Miescher, who produced a first DNA precipitate while he isolated and described the proteins that constituted pus cells. Miescher already then speculated that the substance which he termed "nuclein" had a central role to play in the cell[1]. 75 years later, experiments by Avery *et al.* demonstrated that an attenuated avirulent strain of *Pneumococcus* could be transformed into a virulent strain by exposure to the DNA extracted from a virulent strain, implicating that DNA as opposed to proteins may function as the genetic material[9]. In 1953, the double-helical structure and canonical base-pairing were prominent discoveries by Franklin *et al.*[10] and Watson and Crick[11], followed by a surge of fundamental findings: among others the identification of a "DNA synthesizing enzyme", a DNA polymerase from *Escherichia coli*[12], the cracking of the genetic code of how DNA-encoded sequences of ribonucleic acid (RNA) base triplets called codons correspond to amino acids[13], the discovery of restriction enzymes that can cleave specific sites in the DNA[14,15], DNA sequencing methods[16,17], *in vitro* amplification of DNA by polymerase chain reaction (PCR)[18] and more, all of which enable modern research in genetics and related fields. DNA consists of the four deoxyribonucleoside monophosphates (dNMPs), deoxyadenosine monophosphate (dAMP), deoxythymidine monophosphate (dTMP), deoxyguanosine monophosphate (dGMP) and

deoxycytidine monophosphate (dCMP), linked together covalently to form long polynucleotide strands. DNA typically occurs as a double strand of two such chains that are oriented anti-parallelly[2]. The sequences of the dNMPs in each strand are complementary to each other, such that an adenine (A) pairs with a thymine (T) and a guanine (G) would pair with a cytosine (C) via hydrogen bonds, as proposed by Watson and Crick in 1953[19]. DNA may also assume noncanonical structures other than the B-form duplex and contain noncanonical base-pairing, both of which can affect genomic stability[20,21]. Noncanonical structures are for example cruciform DNA (Figure 1 A), A-DNA, Z-DNA (Figure 1 B), triplex (Figure 1 C), G-quadruplex (G4, Figure 1 D), i-motif, hairpin or slipped DNA (Figure 1 E), some of which are formed through noncanonical Hoogsteen hydrogen bonds[22].

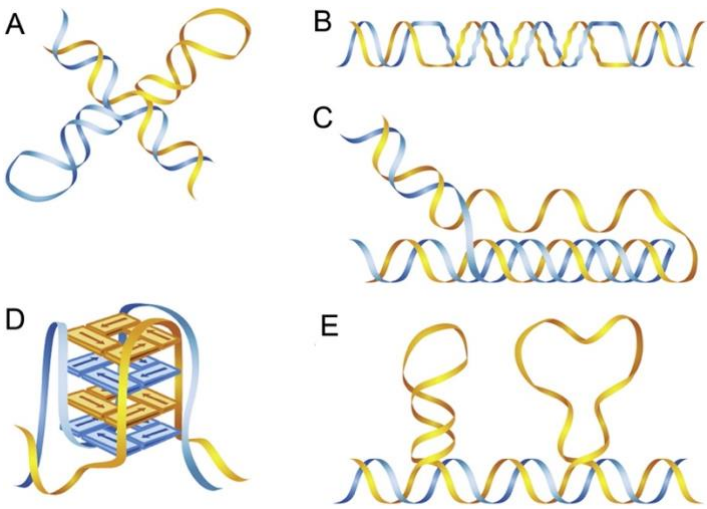


Figure 1: Examples of noncanonical DNA structures. *(A) Cruciform DNA. (B) Z-DNA. (C) Triplex DNA. (D) G-quadruplex. (E) Slipped DNA. (Figure from Zhao et al. (2010)[21] with permission.)*

1.1.1 NUCLEAR DNA

The eukaryotic nucleus contains most of the genetic material as nuclear DNA (nDNA), while a small number of genes is encoded by the mitochondrial DNA (mtDNA, see section 1.1.3). The eukaryotic nDNA is typically organized in several linear chromosomes and their number varies across species (Table 1)[23]. Somatic mammalian cells are diploid and carry two copies of each chromosome (autosome) and two sex chromosomes, while yeast cells can be haploid or diploid and can readily switch between mating types a and α[24].

Table 1: Comparison of genome sizes, chromosome numbers and genes between human, mouse and yeast. *Data for the reference genomes of Homo sapiens (GRCh38.p13), Mus musculus (GRCm39) and Saccharomyces cerevisiae (SacCer3) were retrieved from the RefSeq database[25] and the Saccharomyces genome database SGD (yeastgenome.org).*

	Human	**Mouse**	**Yeast**
Genome size [bp]	~ 3 Billion	~ 2.7 Billion	~ 12 Million
Total number of chromosomes	46 (22, X, Y) (diploid)	40 (19, X, Y) (diploid)	32/16 (diploid/haploid)
Genes	~ 38,000	~ 40,000	~ 6,600

The information contained in eukaryotic genomes is versatile. Unlike prokaryotic genomes where the vast majority of the DNA is protein-coding, only a small fraction of eukaryotic genomes contains protein-coding genes, which can be transcribed to mRNA and translated into proteins[26,27]. The ENCODE project showed that the protein-coding sequences cover only about 1.2% of the human genome, but interestingly those sequences are spread out and span about 40% of the genome from promoter to poly(A) tail[28]. The non-coding DNA was once termed "junk DNA"[29], but the understanding of eukaryotic genomes has since progressed to comprehend more of its complexity and uncover more of its functions. According to ENCODE, RNAs "cover" about 62% of the human genome; among them transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), short RNAs (microRNAs), small interfering RNAs, long non-coding RNAs and pseudogenes[28]. Other genomic features were identified as protein-binding sites, transcription start sites (TSS) or CpG dinucleotide methylation sites associated with epigenetic regulation and chromosome-interacting regions[30]. This illustrates that much of the eukaryotic genome's complexity lies within the non-coding regions.

1.1.2 NUCLEAR DNA REPLICATION

In order to equip eukaryotic daughter cells with a full set of chromosomes, the nDNA has to be replicated correctly in the DNA synthesis phase (S phase) of the cell cycle before cell division occurs during mitosis (M phase)[31]. The replication is initiated at specific positions in the genome called origins of replication (henceforth called origin). Origin firing is a temporally and spatially coordinated process, dividing the genome into replication domains[32]. In yeast, origin firing takes place at specific sequence motifs which are recognized by the Origin Recognition Complex[33,34]. In contrast, many possible positions of origins were identified in mammalian genomes, where some positions are more likely

to be initiated that contain regulatory elements such as TSS and enhancers or showed DNase I hypersensitivity[35]. Through the concerted action and involvement of a range of initiation factors a replisome is formed: two Minichromosome Maintenance (MCM) helicases are loaded in proximity of an Origin Recognition Complex onto the double strand forming a double hexamer[36]. Additional accessory factors including DNA polymerase ε (pol ε) join MCM to form a preinitiation complex[37]. The MCM helicases of the double hexamer are then separated and converted through melting of the double strand, conformational changes and recruitment of Cdc45 and GINS into their active form, encircling a single strand. MCM with Cdc45 and the multi-unit GINS complex on a single-stranded DNA (ssDNA) are called the Cdc45-MCM-GINS (CMG) complex which constitutes the basis for replisome assembly[37,38]. Divergent movement of the CMG helicases from the origin exposes ssDNA and forms the beginning of the replication bubble[36]. Coating of the ssDNA by Replication Protein A (RPA) provides the starting point for the DNA polymerase α (pol α)-primase complex to initiate DNA synthesis. The pol α-primase complex synthesizes an RNA-DNA primer. While the limitation to about 10 nucleotides of the RNA portion of the primer seems to be sterically regulated[39], the mechanism for limiting the DNA portion of the primer to about 20-30 nucleotides remains to be solved. Models of a possible mechanism propose either the removal of pol α by Replication Factor C[40] or by the conformational change as DNA synthesis progresses from A- to B-form DNA for which pol α has lower binding affinity[41]. The CMG complex is moving in 3´ to 5´ direction on the parental strand. It is associated with pol ε which is performing the leading strand DNA synthesis in a continuous manner (Figure 2)[42-46], though some evidence supports the idea that DNA polymerase δ (pol δ) is being used as the DNA polymerase for both strands[47,48]. Recent findings in yeast suggest a role of pol δ in leading strand DNA replication initiation but the bulk of the leading strand still being replicated by pol ε[49,50]. In contrast to the leading strand, lagging strand synthesis has to be performed discontinuously, since a portion of the stand has to be revealed first, before 5´ to 3´ synthesis may occur. DNA synthesis on the lagging strand is initiated by the pol α-primase complex, as well, which provides the RNA primer and limited elongation with DNA[39,51]. These primers are extended by pol δ whose nucleotide incorporation rate is accelerated by Proliferating Cell Nuclear Antigen (PCNA)[52]. Lagging strand synthesis is completed by primer removal during Okazaki fragment maturation discussed in more detail in section 1.5.3. In yeast, replication termination sites are usually found in the middle of two origins of replication where two replication forks meet. They are mostly determined by the timing of origin firing rather than specific termination sequences[53].

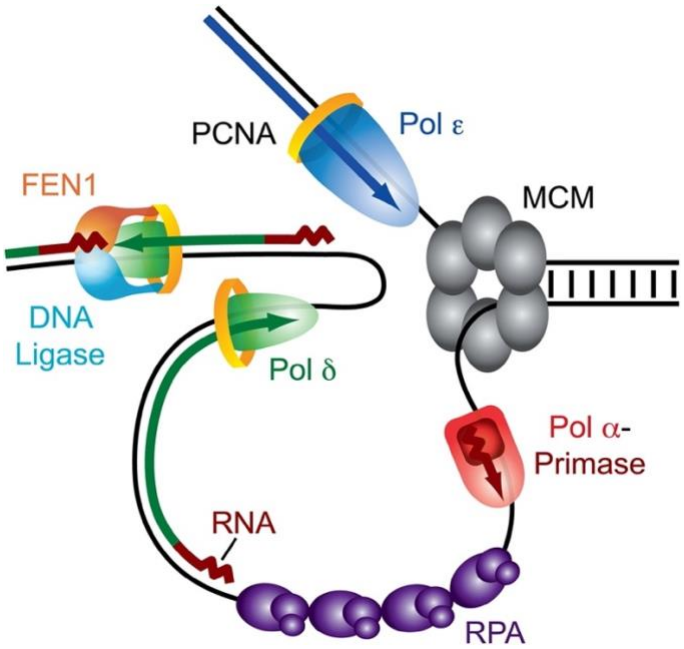


Figure 2: Eukaryotic replication fork. *Minichromosome Maintenance (MCM) helicase (grey) unwinds the parental double strand. The leading strand (blue arrow) is continuously synthesized by DNA polymerase ε (pol ε, blue), while lagging strand synthesis is produced discontinuously by repeated RNA primer (red lines) synthesis by the DNA polymerase α (pol α)-primase complex (red) and extension by DNA polymerase δ (pol δ, green). Okazaki fragment maturation is facilitated by Flap Endonuclease 1 (FEN1, orange) and DNA ligase (light blue). Proliferating Cell Nuclear Antigen (PCNA, yellow) functions as an accessory unit or processivity factor. The Replication Protein A (RPA, purple) coats the exposed parental strand. (Figure from Nick McElhinny et al. (2008)[54] with permission.)*

In humans, replication initiation and termination were found to be co-localized with transcription start and termination sites, respectively, to ensure coordination with the transcription machinery at highly transcribed regions[55]. Termination occurs when two forks converge and leading strands are ligated to the last Okazaki fragment of the respective opposite replication fork. Topoisomerase 2 (Top2) seems to be of importance in fork convergence, where Topoisomerase 1 (Top1) can no longer relieve positive supercoiling[56]. The replication fork probably rotates along the double strand instead, resulting in precatenanes behind it, which are likely resolved by Top2[56-58]. Recent experiments in *Xenopus laevis* egg extracts suggest that CMG complexes can go past each other before being unloaded and that the presence of DNA structure at the CMG complex suppresses its ubiquitination resulting in disassembly during

replication[59]. Disassembly is facilitated by poly-ubiquitination of the CMG complex and subsequent separation of the complex subunits[60,61].

1.1.3 MITOCHONDRIAL DNA

Mitochondria are cellular organelles believed to have originated from an endosymbiotic α-proteobacterium in an archaeal-derived host cell. While the proto-eukaryotic genome increased over time, the mitochondrial genome reduced, only a few genes remained (Table 2) and part of the mitochondrial proteins are encoded on the nDNA[62].

***Table 2: Comparison of eukaryotic mitochondrial DNAs**. Human[63] and mouse[64] mtDNAs are more similar in size and coding genes than the larger yeast[65] mtDNA.*

	Human	**Mouse**	**Yeast**
mtDNA size [kb]	16.5	16.3	85.8
Protein-coding genes	13	13	19
tRNA	22	22	24
rRNA	2	2	2

The cell organelle efficiently generates adenosine triphosphate (ATP) and contributes to the cell's metabolism[66]. Depending on species and tissue, hundreds or thousands of mitochondria are present in each cell. Each mitochondrion in turn contains about 1 to 10 and between 50 to 200 copies of mtDNA in animals and yeast, respectively[67]. As opposed to the biparental inheritance of chromosomal genes, mtDNA is inherited only maternally[68,69]. MtDNA in mammals is about 16.5 kb long but unlike nDNA has a circular form (Figure 3), reminiscent of its bacterial origin[62]. Due to the differences in base composition of each mtDNA strand, the strands could be separated on a density gradient. This gave rise to the designations of "heavy strand" and "light strand"[70]. Yeast has a considerably bigger mtDNA, which is mainly caused by introns and non-coding regions, since only six more protein-coding genes and two more tRNAs are encoded compared to the human and murine mtDNA (Table 2)[71]. It further differs from mammalian mtDNA in that it occurs predominantly in linear form, though it was long believed to be circular as well[72,73]. In contrast to nDNA, mtDNA in mammals (Figure 3) contains very few non-coding sequences, such as the origin of light strand synthesis (OriL) and the displacement loop (D-loop), which encompasses the origin of heavy strand synthesis (OriH) and accounts almost entirely for the difference in length between human and mouse mtDNA[74]. Why certain

genes stay encoded on the mtDNA, while others migrated to the nDNA, is a debated issue. Currently proposed hypotheses include that hydrophobic proteins are difficult to transport to the mitochondria[75,76], that certain gene products could be toxic in the cytosol[77], that mitochondrial genes use noncanonical codons[78] or that colocalization of gene and gene product is required for regulatory purposes[79].

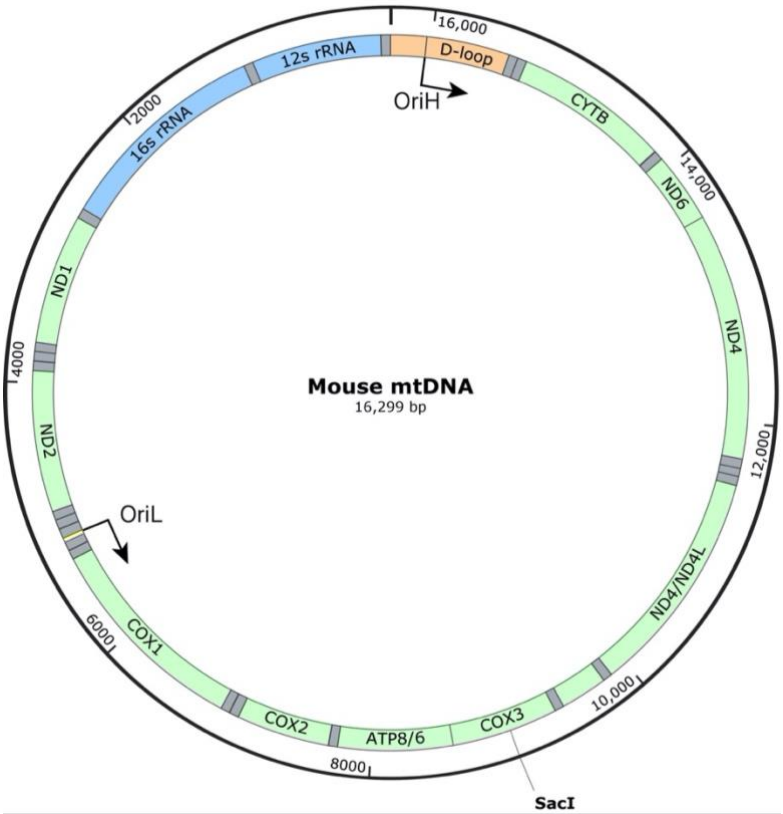


Figure 3: Map of the mouse mitochondrial DNA. *The mouse mtDNA encodes for 13 mRNAs (light green), 22 tRNAs (grey) and 2 rRNAs (blue). A prematurely terminated nascent heavy strand can form a triple-stranded structure, the displacement loop (D-loop, orange), which is also encompassing the heavy strand origin of replication (OriH). The light strand origin of replication (OriL, light yellow) is located at 5,160-5,191 in a short non-coding region. A single SacI cleavage site is situated at position 9,047. Figure based on RefSeq accession NC_005089.1 and created with SnapGene® software (from Insightful Science, snapgene.com).*

1.1.4 MITOCHONDRIAL DNA REPLICATION

The mtDNA replication is mechanistically distinct from nDNA replication and involves factors specific to the mitochondrion. Interestingly, part of the mtDNA replication machinery seems to have originated from bacteriophages[80]. The currently favored model for mtDNA replication in mammals is the strand displacement model (Figure 4)[81,82].

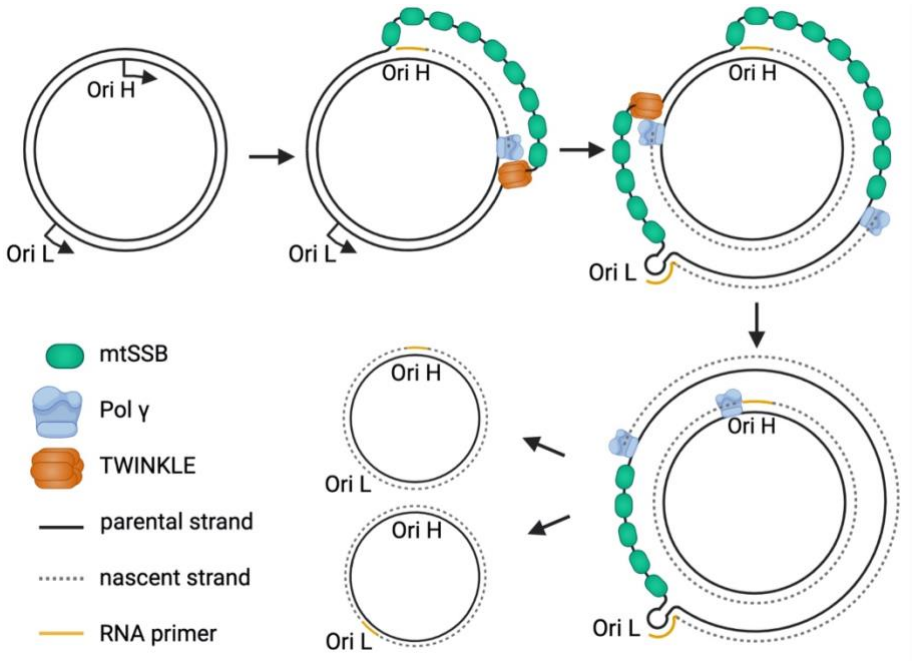


Figure 4: Strand displacement model of mtDNA replication. *Replication is initiated from an RNA primer (yellow line near OriH) stemming from a prematurely terminated transcript between the Light Strand Promoter and Conserved Sequence Box II near OriH. TWINKLE (orange) unwinds the double strand ahead of DNA polymerase γ (pol γ, blue), which extends the nascent heavy strand (dotted, grey line). Meanwhile, the exposed parental heavy strand is bound by the mitochondrial single-stranded DNA binding protein (mtSSB, green). Once OriL on the parental heavy strand is single-stranded, it can form a stem loop near OriL which mtSSB cannot bind. Mitochondrial RNA polymerase provides an RNA primer at the stem loop (yellow line near OriL) which is extended by pol γ (blue) to form the nascent light strand (dotted, grey line). (Figure based on Falkenberg (2018)[82].)*

According to this model, both strands are synthesized in a continuous manner, as revealed by 5´-end mapping of the nascent daughter strands[83]. The strands are replicated sequentially. First, replication is initiated at OriH, where prematurely ended transcripts at the Conserved Sequence Box II (CSBII) from the Light Strand Promoter form R-loops, D-loop-like

structures where the second strand is displaced by an RNA[84-87]. An interstrand G4 between the DNA strand and the transcript anchors the RNA, increases the R-loop's stability and likely contributes to the premature transcription termination at CSBII[88]. RNase H1 seems to be involved in processing of this RNA into a suitable RNA primer and allows the initiation of replication[89]. TWINKLE, the mtDNA helicase, unwinds the double strand[90] while the mitochondrial DNA polymerase γ (pol γ) synthesizes the new heavy strand[91,92]. During this DNA synthesis, the displaced parental heavy strand is bound by the mitochondrial single-stranded DNA-binding protein (mtSSB). The vast majority of replication attempts ends prematurely after about 605 nt. The 7S DNA can stay hybridized to the template, thereby forming a D-loop. The function of the D-loop is however unknown[63,93]. In about 5% of replication events, heavy strand synthesis continues and full replication is achieved[63]. Once the replication fork has passed OriL the parental heavy strand can form a stem loop which cannot be bound by mtSSB[94]. The mitochondrial RNA polymerase can then begin primer synthesis from a poly-T sequence, which is extended by pol γ after about 25 nt to begin light strand synthesis[95,96]. At this point, heavy and light strand synthesis proceed in parallel. Once both strands are completely synthesized, RNA primers need to be processed in order for DNA Ligase III to ligate the newly produced strands[97]. The primers at each origin are removed differently. Near OriH, 5´-ends of the nascent strand mapped to multiple positions suggest the removal of not only the RNA primer but also parts of the newly synthesized heavy strand[63]. RNase H1 and the Mitochondrial Genome Maintenance Exonuclease 1 (MGME1) seem to perform this process at OriH together: RNase H1 is thought to remove the RNA primer from Light Strand Promoter to CSBII whereas MGME1 removes the remaining primer and part of the nascent heavy strand[98,99]. MGME1 can only cleave ssDNA, hence the 5´-end of the nascent strand has to be displaced. One possibility is that the synthesis of the 7S RNA transcript, whose function is otherwise unknown, may be facilitating the required displacement[63,100,101]. Once the primer at the 5´-end is removed, the 5´- and 3'-end can be ligated. For ligation the 5´-end and the 3'-end need to be neighboring, which is achieved by the concerted actions of pol γ and MGME1: pol γ can extend or resect the 3´-end and the 5´-end (displaced by pol γ) can be cleaved by MGME1 until the appropriate substrate for DNA Ligase III is achieved[98,102]. At OriL, the RNA primer is almost entirely removed by RNase H1, only leaving 1-3 ribonucleotides[103]. A second nuclease is required to remove the remaining ribonucleotides before ligation; the responsible enzyme has however not yet been identified[104].

1.2 DNA POLYMERASES

The eukaryotic DNA polymerases can be classified in four DNA polymerase families: A, B, X and Y (Table 3).

Table 3: DNA polymerases from human, mouse and yeast. *Eukaryotic DNA polymerases are classified in 4 families: A, B, X and Y. DNA polymerases are involved in a variety of DNA transactions, including DNA replication, proof-reading, DNA repair mechanisms and translesion synthesis. Abbreviations: BIR: break-induced replication; mtDNA: mitochondrial DNA; nDNA: nuclear DNA; OFM: Okazaki fragment maturation; PrimPol: Primase and DNA-directed polymerase; TdT: terminal deoxynucleotidyl transferase; TLS: translesion synthesis; VDJ: lymphocyte receptor V, D and J gene segments. (Modified from McVey et al. (2016)[105].)*

DNA pol	Human	Mouse	Yeast	Functions
A family[106]				
Pol γ	✓	✓	✓	Replication, proof-reading & repair of mtDNA[91,92,107,108]
Pol θ	✓	✓		TLS, DNA repair[109-113]
Pol ν	✓	✓		TLS, DNA repair[114,115]
B family[116]				
Pol α	✓	✓	✓	Primer synthesis for nDNA replication, BIR[117,118]
Pol δ	✓	✓	✓	nDNA replication & proof-reading (lagging strand), OFM[118-121]
Pol ε	✓	✓	✓	nDNA replication & proof-reading (leading strand), DNA repair[117]
Pol ζ	✓	✓	✓	TLS, DNA repair[122-125]
X family[126]				
Pol λ	✓	✓	✓	TLS, DNA repair[127-130]
Pol β	✓	✓		TLS, DNA repair[112,131-133]
Pol μ	✓	✓		TLS, DNA repair[134-136]
TdT	✓	✓		VDJ recombination, immune adaption, DNA repair[136-138]
Y family[139]				
Pol η	✓	✓	✓	TLS, DNA repair, VDJ recombination[140-145]
Pol κ	✓	✓		TLS, DNA repair[146,147]
Pol ι	✓	✓		TLS, DNA repair[146,148,149]
Rev1	✓	✓	✓	coordinating TLS, dCMP transferase activity[122,150,151]
PrimPol	✓	✓		Lesion-skipping & repriming at stalled replication forks[152-155]

The A-family DNA polymerases mtDNA pol γ[156], DNA polymerases θ (pol θ)[157] and ν (pol ν)[158] are all related to Pol I in *Escherichia coli*. The three replicative polymerases in eukaryotes are pol α, pol δ and pol ε. The lagging strand is synthesized by pol α and δ in a discontinuous manner, while the leading strand is synthesized continuously by pol ε[43,44,54]. All three replicative polymerases belong to the B family of polymerases, which consist of a catalytic and a regulatory subunit, as well as accessory subunits[116] and can be found in yeast, mice and humans (Table 3). B-family DNA polymerases are considered the most common replicases and are found across all domains of life and even in some viruses[116,159]. The terminal deoxynucleotidyl transferase (TdT) and the DNA polymerases λ (pol λ), β (pol β) and μ (pol μ) make up the X family of DNA polymerases. While the primary sequence homology is a bit lower, overall structures of the X-family polymerases are similar[126]. Y-family DNA polymerases such as DNA polymerases η (pol η), κ (pol κ) and ι (pol ι) are all comprised of two subunits: one catalytic, one regulatory[139]. While B-family DNA polymerases almost exclusively serve as replicative polymerases, the DNA polymerases from the A, X and Y families are involved in a wide variety of cellular functions, such as repair and DNA damage tolerance pathways (see Table 3 and section 1.3.3).

1.2.1 DNA POLYMERASE η

Pol η is a specialized Y-family DNA polymerase, found in human, mouse and yeast cells. In analogy with other genes leading to sensitivity to UV radiation, deletion of the pol η gene in yeast was found to result in UV sensitivity, its transcription to be induced by exposure to UV radiation and was hence termed *rad30*[160]. It was later determined to facilitate error-free translesion synthesis (TLS) across UV-induced lesions, such as thymine-thymine *cis-syn* cyclobutane dimers, which would otherwise act as a replication barrier[161-163] but can be accommodated by the more spacious active site[164]. Pol η may however also facilitate error-prone TLS at other damaged bases, such as 7,8-dihydro-8-oxo-deoxyguanine (8-oxoG) which are usually repaired via more efficient mechanisms[165] and at pyrimidine (6-4) pyrimidone photoproducts (6-4PP)[166-168]. Interestingly, I found a distinct lagging strand bias for pol η activity in yeast and presented evidence for the lagging strand bias in humans as well[169].

Aside from the canonical function of TLS, various additional noncanonical cellular functions of pol η were discovered: pol η is involved in diversifying Ig genes by introducing A/T mutations in mice[170] and humans[171]. It is involved in maintaining chromosomal and common fragile site stability[172,173], which can otherwise cause double strand breaks (DSBs) that could promote cancer development[174]. A role of pol η in

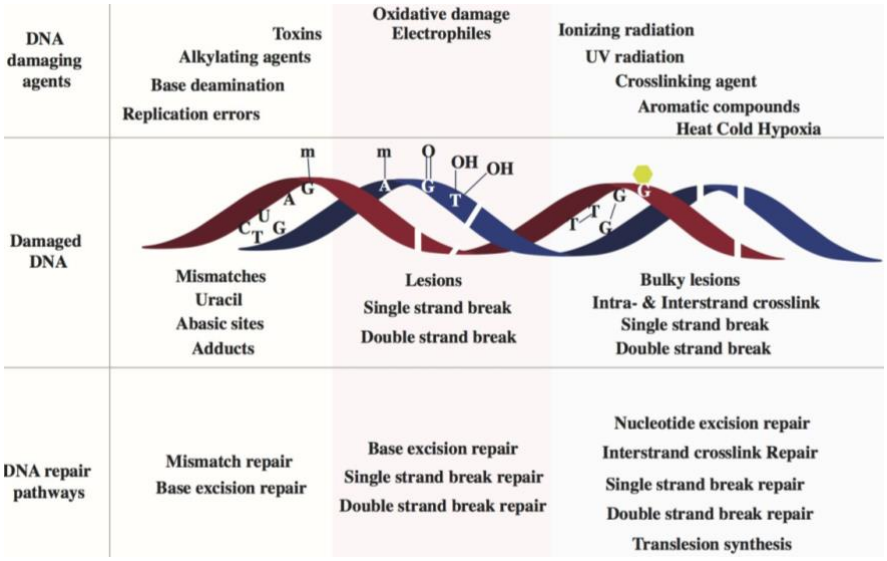
processing of oxidized ribonucleotides by NER was proposed based on deletions caused by 7,8-dihydro-8-oxo-riboguanosine in pol η deficient cells[175]. During homologous recombination (HR) a D-loop intermediate is formed by an invading DNA overhang which needs to be extended past the initial break based on the new homologous template[176]. In *in vitro* reconstitutions of HR, pol η was found to be able to perform extension of the invading strand similar to pol δ in those HR intermediates mediated by RAD51[177,178]. *In vivo* experiments in human cells confirmed an involvement of pol η and pol κ in HR[146], though pol η is probably not a strict requirement for HR[179,180]. In yeast, pol η seems to be the only TLS polymerase involved in the formation of damage-induced cohesion throughout the whole genome which is important for correct chromosome segregation and DSB repair[181] and it is independent of pol η's polymerase activity[182,183]. Moreover, pol η is implicated in regulating alternative lengthening of telomeres and facilitating telomer replication[184,185]. These canonical and noncanonical functions of pol η illustrate the wide variety of mechanisms it is involved in and suggest a fundamental role in maintaining a healthy level of genome stability.

In humans, pol η is of particular interest due to its role in the *Xeroderma pigmentosum* (XP) variant subgroup XP-V[186,187]. XP-V makes up about 23% of all XP cases which is characterized by sun sensitivity in 60% of patients and development of basal cell and squamous epithelial cell carcinoma and cutaneous melanoma at an average age of 8 years[188]. While other XP subgroups are caused by deficiencies in Nucleotide Excision Repair (NER), XP-V is based on the perturbation of pol η[189]. In both cases, the disruption of NER or pol η decreases the cells' ability to tolerate (ultraviolet (UV) light-induced) DNA damage by repairing or efficiently bypassing the damage, respectively[190]. Another aspect, relevant to human health is that pol η was found to facilitate resistance against anticancer therapeutics that induce interstrand crosslinks, probably by accommodating the lesions during TLS[191-193]. Therefore, pol η may be a valuable target for enhancing the treatments that are otherwise rendered ineffective[194].

1.3 GENOME INSTABILITY

Many factors play a role in maintaining genome stability and protecting from sources of genomic instability. Both endogenous and exogenous factors challenge the DNA integrity and numerous control and repair mechanisms have evolved to mitigate and tolerate those challenges (Table 4)[3,195], while genome instability is associated with aging and disease[5,196].

Table 4: Overview of DNA damaging agents, resulting DNA lesions and associated repair mechanisms. *Endo- and exogenous factors (upper row) cause a variety of different DNA damage types. For each category of DNA damage (middle row) a number of cellular mechanisms (lower row) have evolved to repair or tolerate the lesion. (Table from Chatterjee et al. (2017)[195] with permission.)*



1.3.1 EXOGENOUS SOURCES OF GENOME INSTABILITY

The most common sources for exogenous DNA damage are radiation or exposure to chemical agents. Sunlight, especially the contained UV light, can cause alterations in the DNA via direct absorption by the DNA or indirect mechanisms via non-DNA chromophores[197]. Directly absorbed UV light mainly causes cyclobutane pyrimidine dimers (CPDs) and to a lesser extend 6-4PPs (Figure 5)[197-201]. Single strand breaks (SSBs) and possibly DSBs, as well as DNA crosslinks can form via both direct or indirect pathways[197]. Oxidative damage of the DNA can stem from the interaction with reactive oxygen species (ROS) or reactive nitrogen species (RNS) which can be generated by photosensitized reactions[202] or the induction of cellular responses, the latter of which typically occur at a delay after exposure[197,203].

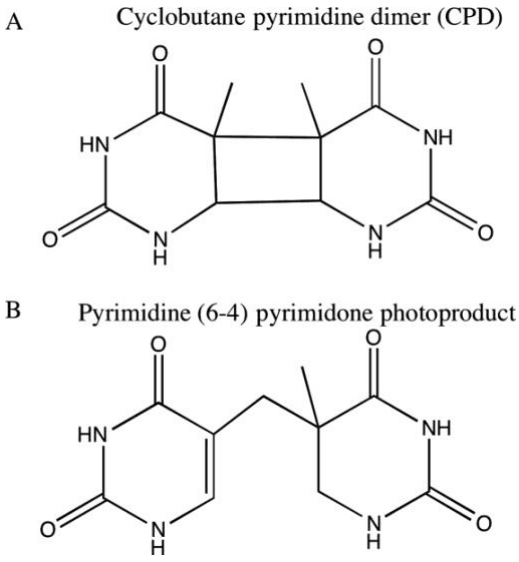


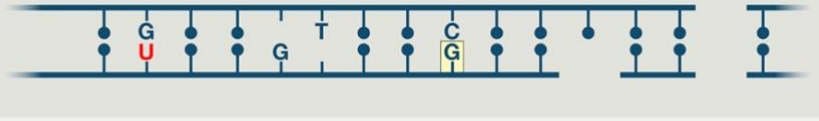
Figure 5: Representative structures of the main DNA lesions induced by UV radiation. *(A) Cyclobutane pyrimidine dimers (CPDs), here: cyclobutane thymine dimers. (B) Pyrimidine (6-4) pyrimidone photoproduct (6-4PP), here: thymine dimer linked via C4 and C6. (Figure from Chatterjee et al. (2017)[195] with permission.)*

Similarly, ionizing radiation such as used in radiotherapy or for diagnostics (X-rays, computer tomography scans, positron emission tomography scans) can directly introduce DSBs, while indirect damaging via ROS can result in SSBs, abasic sites, sugar modifications and base deamination[3,195,204]. Exogenous genotoxins include cigarette smoke, cancer therapeutics, environmental pollutants and contaminants, which can induce a variety of DNA damages[205]. Cigarette smoke causes oxidative damage because it contains free radicals and oxidants[206]. Chemicals commonly used in chemotherapy may alkylate DNA bases, crosslink DNA strands covalently or introduce SSBs and DSBs via the inhibition of topoisomerases[3].

1.3.2 ENDOGENOUS SOURCES OF GENOME INSTABILITY

Aside from exogenous challenges to the DNA, even more lesions are caused by a range of endogenous processes that are part of the normal cellular metabolism and affect the integrity of the DNA[207,208]. In human cells, approximately 70,000 lesions are caused by endogenous mechanisms per day, the majority of which are SSBs (Table 5)[209].

Table 5: Estimation of DNA lesions per cell and day. *Regular cellular functions and processes cause lesions in the DNA. The most common lesions of endogenous origin, their frequency and the predominant mutations they cause are listed. Abbreviations: 7,8-dihydro-8-oxo-deoxyguanine (8-oxoG), single strand break (SSB), double strand break (DSB). (Table from Tubbs et al. (2017)[209] with permission.)*



Damage	Cytosine deamination	Depurination/ depyrimidination	8-oxoG	SSB	DSB
Estimated frequency (per cell per day)	192	12,000/600	2,800	55,000	25
Predominant mutation	C>T	Substitutions	G>T	Substitutions, DBSs	Chromosome rearrangements

While some lesions may occur spontaneously, the underlying mechanisms for each lesion are often not as clear as for the lesions caused by the exogenous factors described above[210]. Spontaneous DNA damage may be mediated by the water present in the cell[211]. While the hydrolysis of the DNA backbone is slow, deamination of cytosine to uracil is estimated to occur around 100 to 500 times per cell per day[212]. Moreover, glycosidic bonds between the bases and the sugar-phosphate backbone are prone to hydrolysis, leading most often to depurination and less frequently to depyrimidination (Table 5)[213,214]. Regular cellular metabolism, such as oxidative phosphorylation for the energy production in mitochondria or the activity of NADPH oxidases and cytochrome P450 reductases are the sources for cellular ROS and RNS[215,216]. DNA damage caused by ROS or RNS is the most frequent type of damage and can lead to base oxidation, SSBs and DSBs[217]. 8-oxoG is the major oxidative base lesion observed, probably due to the low redox potential of guanosine[218,219].

Furthermore, a recent study by Xia *et al.* identified a number of proteins promoting spontaneous DNA damage in human cells[220]. The identified proteins, only 5.6% of which were known to be involved in DNA repair proteins, showed an overrepresentation among known cancer-driving genes and were associated with increased mutation rates when found in higher copy numbers. The authors suspect a role upstream of the known DNA repair pathways, through either promoting DNA damage and thereby overwhelming the available DNA repair capacity or by

downregulation or inhibition of DNA repair mechanisms. Moreover, they propose three possible mechanisms for endogenous DNA damage: 1) blockage of the replisome or fork reversal caused by transcription factor binding, 2) altered transmembrane transporter activities causing increased levels of ROS and 3) disruption of the replisome causing replication fork collapse[210].

The cell usually facilitates DNA methylation as a normal epigenetic mechanism which is often associated with the repression of transcription[221]. The cellular methyl group donor S-adenosylmethionine can however also spontaneously generate N7-methylguanine, N3-methyladenine and O^6-methylguanine residues. While N7-methylguanine is not considered a harmful lesion, N3-methyladenine can facilitate cytotoxicity through the inhibition of DNA synthesis. O^6-methylguanine produces G:C to A:T transitions and is therefore a highly mutagenic lesion[195].

In addition, DNA transactions, such as DNA replication, faulty chromosomal segregation and erroneous or impaired DNA repair need to be considered as potential endogenous sources of genome instability[222]. During DNA replication the fidelity of DNA polymerases when incorporating the nucleotides and the ability of some DNA polymerases to proof-read incorporated bases determines how often wrong bases with a potential for mutagenesis are introduced[223]. Replication fidelity is further increased by mismatch repair (MMR), which is discussed in more detail in section 1.3.3[224,225]. DNA polymerases' base selectivity can be affected by deoxyribonucleoside triphosphate (dNTP) pool imbalances, repetitive sequences that promote polymerase slippage, and other sequence effects including secondary structures[226,227]. Another challenge to fork progression is posed by the transcription machinery which can be met head-on or co-directionally. The activities of the replisome and the transcription machinery are usually well-regulated but some instances of their collision have been demonstrated and are associated with DNA damage or recombination[228]. Furthermore, R-loop formation involving the nascent RNA transcript can block the replication fork and is also associated with genomic instability[229]. For the mtDNA pol γ, the ROS-rich environment of the mitochondria also seems to affect fidelity by causing oxidative damage to its exonuclease domain which decreases its proof-reading ability[230]. Faulty replication fork progression may cause DSBs or ssDNA gaps, and chromosomal damages like elevated occurrence of sister chromatid exchange, hyper recombination, gross chromosomal rearrangements and even chromosome loss[222]. Topological stress, which

is normally resolved by the activity of suitable topoisomerases may also introduce DNA damage through the activity of cohesin that traps the topological stress near centromeres[231]. DNA repair mechanisms, while in place to repair or tolerate damage, can introduce errors for the sake of preventing greater damage to the DNA. For example, post-replicative repair through translesion DNA polymerases may allow replication past bulky lesions to maintain DNA integrity but comes at the cost of the lower DNA polymerase fidelity of TLS DNA polymerases[232]. DSBs are considered to be the most harmful DNA lesions, probably due to the mutagenic potential of the available repair mechanisms: Non-Homologous End-Joining (NHEJ) and HR (see section 1.3.3)[233]. Furthermore, base mismatches and their repair via MMR was found to be associated with repair-induced lesions in flanking regions of the original lesion[234].

Finally, the various ways of incorporation of ribonucleotides as described in section 1.4 and imperfect removal contribute to genome instability by forming SSBs through the hydrolysis via the 2´-hydroxyl group present in ribonucleotides. The majority of the estimated 70,000 lesions per cell and day are thought to be repaired efficiently and likely do not reflect permanent damage present in the DNA[209]. While I did not determine the daily frequency of incorporated ribonucleotides in DNA, I could estimate 5.2 million stably incorporated ribonucleotides in the murine nDNA in Paper IV, which makes this noncanonical nucleotide the most common lesion in DNA by at least two orders of magnitude and may in part explain the fact that SSBs are the most common other lesion (Table 5)[209].

1.3.3 MITIGATING MECHANISMS

As outlined above, the genome stability is challenged by a wide variety of stress factors that can cause a wide variety of DNA damages. Eukaryotic cells have therefore developed many mechanisms to repair or tolerate such damages to preserve the genomic integrity. DNA repair and tolerance mechanisms have to efficiently recognize the presence and type of lesion and select and facilitate appropriate repair[195,235], which will be described briefly in this section.

Direct reversal

Mammals contain enzymes that can directly reverse some of the DNA lesions arising from UV radiation or alkylation[195]. Direct reversal of O-alkylation of guanines and even interstrand crosslinks between guanines via alkyl-groups can be facilitated by O^6-alkylguanine-DNA alkyltransferase[236]. A family of O^6-alkylguanine-DNA alkyltransferase-homologous enzymes lacking the ability to reverse the damage, however,

sense and direct bulky alkylations to the NER pathway[209]. N-alkylation of bases can be directly reversed by AlkB-related α-ketoglutarate-dependent dioxygenases[237].

Base Excision Repair
Small lesions that are usually not causing significant structural distortions, including forms of oxidation (e.g. 8-oxoG), deamination, alkylation and apurinic/apyrimidinic sites (AP-sites) are recognized and repaired by Base Excision Repair (BER). A damaged base is removed by a DNA glycosylase or the process proceeds directly from an AP-site. Apurinic/apyrimidinic Endonuclease 1 makes an incision at the 5´-side of the AP-site's sugar moiety, freeing the remaining 5´-deoxyribose phosphate. This gap is then either filled by a single nucleotide during single-nucleotide BER or via strand-displacement DNA synthesis in long-patch BER. The resulting flap in long-patch BER can be removed by Flap Endonuclease 1 (FEN1) and the resulting nick from both BER pathways can be sealed by DNA Ligases I or III[238].

Nucleotide Excision Repair
Bulkier lesions such as CPDs and 6-4PPs or damage from genotoxic agents are typically repaired through NER, which is considered to be a very versatile repair pathway. The mechanisms of global genome NER and transcription-coupled NER are distinguished mechanistically in how lesions are recognized. Global genome NER is initiated via the recognition of genome-wide lesions, while transcription-coupled NER is triggered by a stalled RNA polymerase at a DNA lesion[239]. In brief, after pathway-specific recognition the pathways converge with the recruitment of transcription initiation complex TFIIH, which contains the helicases XPB and XPD that unwind about 30 nucleotides around the lesion[240]. A pre-incision complex is formed, which protects the free ssDNA on the intact strand. The ERCC1-XPF nuclease incises the strand with the lesion and DNA displacement synthesis by pol ε in replicating cells or pol δ and pol κ in non-replicating cells proceeds for a few nucleotides[241]. The resulting ssDNA flap is cleaved by the endonuclease XPG and DNA Ligase I or Ligase IIIα/XRCC1 seal the nick in replicating or quiescent cells, respectively[242].

Ribonucleotide Repair
The main repair pathway for incorporated ribonucleotides is Ribonucleotide Excision Repair (RER) and in its absence Top1-mediated ribonucleotide removal can serve as an alternative repair mechanism. Ribonucleotide repair is discussed in more detail in section 1.5.

Mismatch Repair

Contributing to the replication fidelity, as mentioned in the previous section, is the MMR pathway. MMR acts on base mismatches that were wrongfully produced during replication and not removed by proof-reading[225]. Mainly MutSα but also MutSβ, which are heterodimers of MutS Homolog (MSH) 2 and MSH6, and MSH2 and MSH3, respectively, recognize the mismatches or roadblocks while sliding along the DNA[243]. Upon recognition, MutLα, PCNA and Replication Factor C are recruited to the lesion. Moreover, the Exonuclease 1 (Exo1) is loaded onto the nascent strand for excision of the error[244]. Due to the 5´ to 3´ directionality of Exo1, an incision 5' of the mismatch is necessary for its activity and was found to be facilitated by MutLα[245]. Finally, DNA pol δ can synthesize DNA to fill the resulting gap and the remaining nicks are sealed by DNA Ligase I[246,247].

Interstrand Crosslink Repair

Covalent interstrand crosslinks (ICL) of bases in complementary strands may arise from a variety of endo- and exogenous agents including certain cancer therapeutics[195]. ICL repair follows varied pathways in quiescent cells, proliferating cells (replication-coupled) and in connection with transcription. Moreover, recent findings suggest even a lesion-specific variant of the ICL repair[248]. In brief, replication-coupled ICL repair is triggered at converging replisomes where the parental DNA strands are held together by the ICL. Separation of the parental strands or so-called "unhooking" can be facilitated through the Fanconi anaemia protein-dependent pathway, a NEIL3 DNA glycosilase-dependent pathway or in the case of acetaldehyde ICLs via direct reversal of the lesion on one of the strands. The resulting gap on the unhooked DNA strand is then filled either via HR and or TLS which are both discussed below[248,249]. In quiescent cells transcription-dependent and -independent ICL repair involves different pathways of ICL recognition, but the pathways converge during the first incisions 5' and 3' of the ICL. In this case, incisions are made by NER factors on one strand and the resulting gap is filled via TLS[250]. A second gap is produced in a similar fashion; now on the other strand with the attached ICL and the previously incised DNA stretch so that it is released. The resulting gap is filled by pol δ[251].

Single Strand Break Repair

SSB repair (SSBR) can be divided into long and short patch SSBR[195]. For long patch SSBR the SSBs are detected by Poly(ADP-ribose) polymerase (PARP) 1 and undergo subsequent end processing to remove any damages and generate 3´- and 5´-ends. A variety of enzymes, including polynucleotide kinase, apurinic/apyrimidinic endonuclease, pol β,

Tyrosyl DNA Phosphodiesterase 1, Aprataxin and FEN1, may facilitate this step to handle each possible terminal lesion appropriately. The resulting gap is filled by pol β in connection with pol δ and pol ε, and sealed by DNA Ligase I. During short patch SSBR the substrate is generated by the BER pathway, the gap is filled by pol β and DNA ligase III seals the remaining nick[195,252].

Double Strand Break Repair

The two major repair pathways for DSBs are NHEJ and HR. In mammalian cells, NHEJ can further be divided into canonical NHEJ and alternative end joining which serves as a backup pathway in the absence of NHEJ proteins[253]. During classic NHEJ, the Ku70-Ku80 heterodimer recognizes and binds the DSBs and protects the DSBs from degradation[254,255]. The Ku-heterodimer allows recruitment of components for the long-range synaptic complex that can turn into the short-range synaptic complex. The short-range synaptic complex ensures compatibility of the ends and allows ligation once any processing of incompatible terminal groups by the nuclease Artemis, terminal deoxynucleotidyl transferase, pol λ or pol μ has taken place[256]. Finally, strands are ligated by Ligase IV in complex with XRCC4 and XLF[257,258]. Unlike NHEJ, which can repair DSBs flawlessly but also has the potential to ligate mismatching DNA ends, HR uses strand invasion of the sister chromatid for the template-directed repair to facilitate DSB repair[195]. HR is initiated by the Mre11-Rad50-Nbs1 complex allowing the recruitment of the HR components. The DNA is initially resected to generate 3´-overhangs by the Mre11-Rad50-Nbs1 complex and RPA is loaded on the ssDNA overhang. Long range resection by Exo1 and the BLM helicase follows, creating a longer 3´-overhang that will invade the homologous DNA to form a D-loop upon sufficient base pairing[259]. The invading 3´-overhang can then be extended by pol δ through displacement DNA synthesis using the newly acquired template strand, though other TLS polymerases have also been suggested[105]. HR is mainly resolved via non-crossover synthesis-dependent strand annealing in somatic cells or double Holliday junction in dividing cells. Error-prone alternatives such as long-tract gene conversion and break-induced replication may occur when the other two pathways fail[253].

DNA Damage Tolerance

When a replicative polymerase is stalled at a bulky DNA lesion, the CMG helicase is typically not affected and can continue unwinding which leads to uncoupling of the helicase and the stalled polymerase, and produces long stretches of ssDNA[260]. Coating of the ssDNA with RPA triggers the ATR/Chk1 pathway which promotes cell cycle arrest, replication fork

stabilization and restarting of the DNA synthesis either by downstream repriming, lesion bypass by TLS polymerases, template switching or fork reversal and lesion repair[261]. TLS is thought to occur either on the fly at the replication fork or as post-replicative gap filling[232,262,263]. The bypass of DNA lesions by TLS polymerases typically has a lower fidelity than the replicative DNA polymerases, but can be accurate when synthesizing past certain lesions[264]. The TLS polymerases' lower fidelity stems from wider active sites which can encompass the DNA lesions on the parental DNA strand or allows them to fill ssDNA gaps[195,265]. This is a trade-off in preventing more catastrophic consequences of stalled forks, such as fork collapse which can result in a DSB[266]. The other major pathways of DNA damage tolerance are error-free: 1) the template switching mechanism which allows error-free synthesis across DNA lesions by utilizing the sister chromatid as the template reminiscent of the HR pathway[267] and 2) fork reversal which is also dependent on some HR factors[268].

1.4 RIBONUCLEOTIDE INCORPORATION

The two main pathways by which ribonucleotides can be incorporated into DNA are through synthesis of RNA primers that are needed for the initiation of DNA replication or through misincorporation by the replicative polymerases. In the nucleus, the pol α-primase complex is responsible for synthesizing a primer consisting of 7 to 12 ribonucleotides synthesized by the primase and is extended by pol α[39,51].

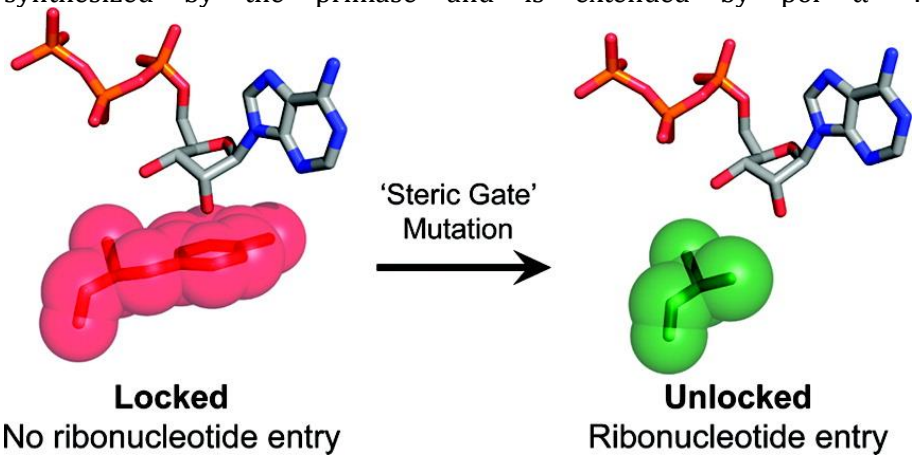


Figure 6: Steric ribonucleotide discrimination by a 'steric gate' residue. *DNA polymerases often, though not exclusively, exclude incoming ribonucleotides through a steric clash between a bulky "steric gate" residue and the 2´-hydroxyl group of the incoming ribonucleotide (left). Changing this residue to a smaller one unlocks the steric gate and allows*

for ribonucleotides to be incorporated during DNA synthesis more frequently. (Figure from Brown et al. (2011)[269] with permission.)

In the mitochondria the mitochondrial RNA polymerase provides a 25 to 27 nt RNA primer[270]. The bulk of DNA synthesis is then performed by extension of those primers by pol δ or pol ε in the nucleus[271], or pol γ in mitochondria[82]. The replicative DNA pols α, δ and ε are able to distinguish between incoming dNTPs and ribonucleoside triphosphates (rNTPs) by a so-called "steric gate" (Figure 6), but the discrimination against ribonucleotides is imperfect[269,272]. Hence, ribonucleotides are occasionally incorporated by the replicative DNA polymerases. Similarly, the mitochondrial DNA pol γ can incorporate ribonucleotides into mtDNA as well, even though it possesses a higher selectivity than the replicative polymerases[273]. In addition, other mechanisms may introduce ribonucleotides into the DNA. Specialized DNA polymerases involved in TLS or DNA repair pathways are also able to misincorporate ribonucleotides and it has been demonstrated for yeast pol ζ[274], murine TdT[275] and human pol β[276], pol η[277], pol λ[278], pol μ[279], pol θ[280], pol ι[281] and may be implicated for pol κ[282] and Rev1[283] as well. While incorporated ribonucleotides are typically seen as a threat to genome stability, ribonucleotides were found to be crucial in DSB repair via NHEJ[284,285] or HR[286]. Ribonucleotide incorporation may also temporarily serve to tolerate dNTP shortage[287] and was shown to be a discrimination mechanism for the nascent strand during MMR[288,289]. Moreover, incorporated ribonucleotides function as imprints to facilitate mating type switch in *Schizosaccharomyces pombe*[290,291].

1.5 RIBONUCLEOTIDE REPAIR

1.5.1 RIBONUCLEOTIDE EXCISION REPAIR

RER (Figure 7) is considered the main pathway by which ribonucleotides can be removed from nDNA. The endonuclease RNase H2 is active on single incorporated ribonucleotide[292,293] and can make an incision 5´ of it[294,295]. DNA synthesis by pol δ or pol ε then displaces the strand with the incorporated ribonucleotide and the resulting flap is excised by FEN1 or Exo1[294,295]. The resulting nick is subsequently sealed by DNA Ligase I[294]. A redundancy of the participating enzymes and the possible involvement of other components in RER was recently proposed based on *in vitro* experiments that suggested a role of the DDX3X protein, which showed a RNase H2-like activity, and of pol β and pol λ as alternatives to pol δ[296]. RER plays an important role in MMR because the transiently occurring

nick can act as a strand discrimination signal for a nearby mismatch[288,289]. While central to RER, RNase H2 has additional functionality in processing longer stretches of incorporated ribonucleotides or R-loops. This activity overlaps with RNase H1, which can process stretches of ribonucleotides or R-loops, but lacks the ability to incise single ribonucleotides[297].

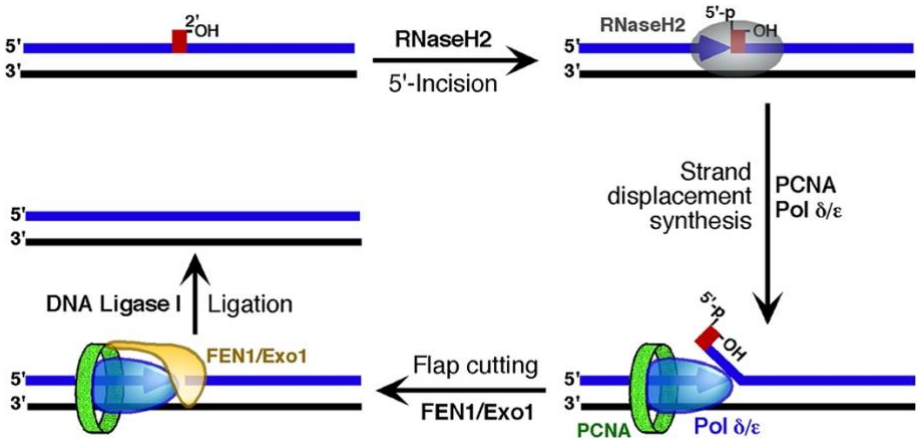


Figure 7: Schematic of Ribonucleotide Excision Repair. *RNase H2 (grey) performs an incision at the 5´-side of the incorporated ribonucleotide (red). Pol δ or pol ε (blue) can extend the resulting 3´-end during displacement DNA synthesis. The resulting flap can be cleaved by the Flap Endonuclease 1 (FEN1, yellow) or Exonuclease 1 (Exo1). The remaining nick is sealed by DNA Ligase I. (Figure from Sparks et al. (2012)[294] with permission.)*

1.5.2 TOP1-MEDIATED RIBONUCLEOTIDE REPAIR

An alternative pathway may process incorporated ribonucleotides in the absence of RNase H2. Top 1 relieves tension from transcriptional supercoiling by introducing a nick in double strand DNA, but was also shown to have an endoribonuclease activity, thereby facilitating a potentially mutagenic removal of incorporated ribonucleotides[298-302]. When Top1 incises DNA at a ribonucleotide, Top1 is bound covalently to the 3´-phosphate, leaving a 5´-hydroxyl group. The 2´-hydroxyl group of the incorporated ribonucleotide has then the possibility for a nucleophilic attack at the phosphate, forming a 2´-3´cyclic phosphate[303] and releasing Top1. Subsequent processing including a second incision by Top1 can be error-free or result in dinucleotide deletions at repetitive sequences[299,304]. If the second incision is however performed on the strand opposite of the incorporated ribonucleotide, a DSB is caused[298]. Both DSBs and deletions are obvious threats to genome stability and it remains unclear if and how a DNA nick with a 2´-3´-cyclic phosphate may be resolved.

1.5.3 PRIMER REMOVAL

Aside from the removal of incorporated single ribonucleotides, short stretches of ribonucleotides are frequently introduced in the form of primers for replication. Especially lagging strand replication, which requires the repeated initiation of DNA synthesis due to the discontinuous mode of DNA synthesis, involves the numerous, transient incorporations of RNA primers produced by the pol α/primase complex[37]. On the lagging strand, DNA synthesis by pol δ will eventually near the 5´-end of the previous Okazaki fragment. By continuing the DNA synthesis, the primer of the previous Okazaki fragment is displaced and a DNA flap created which is cleaved by FEN1[305]. Displacement of a longer flap is thought to be rare. Instead only few ribonucleotides from the primer are displaced and cleaved in multiple successive steps in a process called nick translation[37,306]. The process is completed by DNA ligase sealing the nick[307]. Alternatively, primers or R-loops may be recognized and removed by RNase H1 or RNase H2. RNA:DNA hybrids are largely removed by RNase H2, which seems to have unrestricted access to genome. RNase H1 can remove such hybrids as well, but the enzyme is spatially restricted[297,308].

1.5.4 RIBONUCLEOTIDE REPAIR IN MITOCHONDRIA

As described above, a number of redundant ribonucleotide removal pathways are available in the nucleus. In contrast, ribonucleotide removal in mtDNA seems to be limited to the removal of replication primers generated by the mitochondrial RNA polymerase[309,310]. Similar to the Okazaki fragment maturation, the removal involves pol γ which, like pol δ, possesses a 3´-5´exonucleolytic function balanced with its limited ability for strand displacement, contributing to the removal of the RNA primer[311,312]. As mentioned in section 1.1.4, RNaseH1 has been implicated in facilitating primer removal from mtDNA and seems to be part of a mechanism involving two nucleases[104]. It is localized to the mitochondria and its loss is detrimental to mtDNA genome replication, leading to embryonic lethality in mice[313,314]. The severity of RNase H1 deficiency may also be explained by the recent finding of its involvement in the initiation of mtDNA replication[89]. RER mediated by RNase H2 was eliminated as a possible removal mechanism in yeast mitochondria, where RNase H2 deficiency did not increase the number of incorporated ribonucleotides in mtDNA[309]. The impact of nucleoside triphosphate (NTP) pool imbalances on the incorporated ribonucleotides in mtDNA but not in nDNA unless RER was abolished, is further evidence suggesting the absence of RER in mitochondria[315].

1.6 RIBONUCLEOTIDES AND DISEASE

The importance of efficient ribonucleotide removal is illustrated by the fact that more than half of all Aicardi-Goutières syndrome (AGS) cases are characterized by mutations in each of the three subunits of RNase H2, the central enzyme of RER[316-318]. AGS is a neurological disorder that is usually passed on through autosomal recessive inheritance. The autoimmune disease systemic lupus erythematosus shares some similarities with AGS connected to impairment in RNase H2[319]. RNase H2-mediated RER is usually error-free, the backup pathway via Top1, however, is error-prone, suggesting a contribution to genome instability in absence of efficient RER[320]. In addition, stably incorporated ribonucleotides were found to increase UV-induced CPDs which pose a threat to genome stability as well[319]. Interestingly, some AGS patients with mutations in RNASEH2B had a later onset and milder neurological phenotypes as well as reduced mortality[321]. One proposed possible explanation might be that a certain threshold for the number of tolerable stably incorporated ribonucleotides exists[320]. A study in mice showed that an RNase H2A subunit variant deficient for ribonucleotide excision but proficient in the processing of RNA:DNA hybrids leads to embryonic lethality, like loss of the entire RNase H2B or RNase H2C subunits. This indicates that the defective removal of ribonucleotides and not the processing of R-loops is causative of the embryonic lethality. Furthermore, the RNase H2A G37S variant with reduced ability of ribonucleotide removal and RNA:DNA hybrid processing was viable, which suggests that there is a level of incorporated ribonucleotides that can be tolerated[322]. Loss of RNASEH2B was found to occur in chronic lymphocytic leukemia and other haematopoietic malignancies as a result of RNASEH2B's genomic position between two tumor suppressor loci that are frequently deleted[323]. Similarly, RNASEH2B loss was a side effect of deletions found in prostate cancer[324]. In these cases, disruption of RER seems to be not the disease cause but a consequence. The accumulation of ribonucleotides, however, is a feature that can potentially be exploited in therapies of these cancers, because they confer sensitivity to PARP inhibitors[325]. A contributing role of stably incorporated ribonucleotides was also suggested in the neurological phenotype of ataxia with oculomotor apraxia 1 where RNA:DNA hybrids accumulate[326]. In mitochondria, the absence of RNase H1 impairs the ability of RNA primer removal which is associated with adult-onset mitochondrial encephalomyopathy[313,327]. Stably incorporated ribonucleotides seem however to be relatively well-tolerated in mtDNA where no ribonucleotide removal pathway comparable to RER in the nucleus is present[309,315].

2 AIMS

The broad aims in the thesis were two-fold: The first aim was to investigate the application of ribonucleotide incorporation as means to genome-wide track specialized DNA polymerases that are not involved in the bulk of DNA synthesis during replication. Second, the aim was to gain a better understanding of the role stably incorporated ribonucleotides have in mammalian genomes *in vivo*. The first aim was addressed by Paper I, while different aspects of the second aim were addressed in Papers II, III and IV.

The publications had the following specific aims:

Paper I: To determine and characterize yeast DNA pol η's contribution in DNA synthesis genome-wide.

Paper II: To describe a modification to the HydEn-seq method which allows for simultaneous mapping and quantitation of incorporated ribonucleotides in DNA.

Paper III: To investigate the effects on mouse mtDNA stability of stably incorporated ribonucleotides *in vivo*.

Paper IV: To study stably incorporated ribonucleotides in nDNA and mtDNA from different murine tissues and the effects of genomic features on the distribution of those incorporated ribonucleotides.

3 RESULTS

3.1 PAPER I

DNA polymerase η contributes to genome-wide lagging strand synthesis

Eukaryotic DNA replication is characterized by the division of labor between the three replicative polymerases, pol α, pol δ and pol ε, that perform the majority of DNA synthesis. Pol ε continuously synthesizes the leading strand and pol α and pol δ facilitate lagging strand DNA synthesis[54]. Eukaryotes contain, however, a number of other DNA polymerases including pol η. Their roles are more specialized and related to DNA repair mechanisms or the bypass of lesions in the template DNA[105]. Where in the genome and how much those specialized polymerases contribute to DNA synthesis has not been studied in detail. The HydEn-seq method was previously used to precisely determine the activities of the yeast replicative DNA polymerases genome-wide. This approach requires the modification of the steric gate residue near the polymerases' active site, or a residue close to the steric gate, that decreases the polymerases' ability to discriminate against rNTPs and the absence of efficient RER. This leads to a much more frequent misincorporation of ribonucleotides during replicative DNA synthesis which are not readily repaired due to the RER deficiency. The HydEn-seq method can then be applied to map those incorporated ribonucleotides genome-wide, thereby tracking where the DNA polymerase in question was active[43].

To genome-wide map the activity of pol η, I used yeast strains with a steric gate variant of pol η in an RER and NER deficient background. Additionally, pol ζ was initially removed due to its overlapping functionality[328], though NER- and pol ζ-proficient strains yielded similar results. HydEn-seq libraries prepared from these strains were analyzed bioinformatically and compared to steric gate variants of the replicative polymerases. It was possible to detect incorporated ribonucleotides by pol η, despite the assumption that pol η, as a specialized TLS polymerase, is only involved in DNA synthesis to a limited extend. By comparing the mapped ribonucleotides from pol η to the patterns generated by steric gate variants of the replicative pol α, pol δ and pol ε, I could conclude lagging strand-specific activity for pol η. This discovery held also true near replication origins, where I found the replicative polymerases to transition between strands. Pol η activity clearly followed the same

transitions as pol α and pol δ and the opposite directionality was observed for pol ε. I hypothesized that pol η's C-terminus is involved in its recruitment to the lagging strand, based on the presence of PCNA-Interacting Protein or Rev1-Interacting Region motifs therin[329,330]. To determine whether pol η's lagging strand specificity is facilitated via the C-terminus, I investigated a truncated pol η steric gate variant where the last 10 C-terminal residues were deleted. This deletion led to the of loss of lagging strand-specific pol η activity, indicating that the C-terminus is indeed the interface needed for pol η's lagging strand recruitment. I further characterized the pol η steric gate variant by providing a modified RNase H2 which retains the ability to cleave at multiple consecutive ribonucleotides but is ineffective when encountering single incorporated ribonucleotides[331]. I observed a loss of lagging strand specificity in the presence of this RNase H2 variant, indicating that the steric gate variant of pol η predominantly incorporates short stretches of consecutive ribonucleotides. Moreover, I analyzed the sequence context of the detected ribonucleotides and found that a T-T dimer template was the most common position for ribonucleotides incorporated by the pol η steric gate variant. Pol η's preference for T on the template strand was further supported by our analysis of single base pair deletions in the *CAN1* gene. This preference is most likely explained by pol η's function in efficiently bypassing *cis-syn* T-T dimers[161]. Finally, we found also evidence that human pol η exhibits a lagging strand preference as well, by analyzing TA to TG mutations attributed to pol η[332] in whole genome sequencing data from melanoma patients.

3.2 PAPER II

Simultaneous mapping and quantitation of ribonucleotides in human mitochondrial DNA

There are several methods for the identification or quantitation of incorporated ribonucleotides in DNA. Ribonucleotide incorporation rates of DNA polymerases can be estimated by *in vitro* primer extension experiments[269,272], but do not reflect stably incorporated ribonucleotides in the genome due to ribonucleotide removal pathways. Approaches to quantitate stably incorporated ribonucleotides involves alkaline hydrolysis and visualizing resulting DNA fragments, allowing estimates of the number of stably incorporated ribonucleotides but not conclusions about the base identity of incorporated ribonucleotides[333]. Sequencing-based methods[334-336] including HydEn-seq[43] were able to identify incorporated ribonucleotides genome-wide but lacked the ability to

determine the absolute number of stably embedded ribonucleotides. To determine both the number and the base identity of incorporated ribonucleotides the HydEn-seq method was extended to include cleavage by a sequence specific endonuclease. In the HydEn-seq libraries generated to produce the data presented in this paper[315], HincII was utilized, but even other restriction enzymes may be used, depending on the target organism and whether mtDNA, nDNA or both will be the subject of the analysis.

This publication illustrates important aspects of the modified HydEn-seq method. I presented representative electropherograms of acceptable KOH- and KCl-treated libraries depicting clear differences in resulting fragment sizes. KCl-treated libraries should contain highly intact DNA consistent with free 5´-ends, showing very little to no DNA fragment below 1500 bp, whereas the KOH-treated libraries will typically result in a broad spectrum of DNA fragments ranging from around 300 bp to 2000 bp consistent with an increase in 5´-ends from hydrolysis at incorporated ribonucleotides. Furthermore, I provided validation for the HincII cleavage, where I detected 70% of all 5´-ends in KCl-treated libraries at the HincII cleavage site and this number is decreased to about 40% of all ends mapping to the HincII cleavage site in KOH-treated libraries, due to the alkaline hydrolysis at ribonucleotides as well. Lastly, I illustrated the ability to calculate absolute number of incorporated ribonucleotides by normalization to the reads at the HincII cleavage site and the base identity of them.

3.3 PAPER III

Elimination of rNMPs from mitochondrial DNA has no effect on its stability

The mitochondrial replicative pol γ is capable of stringent discrimination against rNTPs[273,337]. Still, stably incorporated ribonucleotides are found in mtDNA due to the lack of efficient removal mechanisms[309,315] and incorporation by pol γ is thought to be the main source for these ribonucleotides[309]. Though stably incorporated ribonucleotides in mtDNA seem well-tolerated, it is poorly understood how they may affect mtDNA stability. A role of ribonucleotides was implicated in previous studies of mtDNA depletion syndrome patient-derived cell lines[315] and in a mouse model for mtDNA disorders caused by the loss of function in the MPV17 protein[338].

To investigate the effects of stably incorporated ribonucleotides in mtDNA, the experimental approach to utilize alkaline hydrolysis through KOH-treatment of DNA to approximate incorporated ribonucleotides was validated. When comparing the effects of RNase H2- and KOH-treatment on DNA fragment length, both showed an almost identical decrease in fragment size, indicative of ribonucleotides being the main cause of alkali-sensitive sites in mtDNA. Comparing mtDNA from mouse embryos to mtDNA from spleen, heart, skeletal muscle, liver and brain of adult mice, we found that mtDNA from embryos contained the lowest number of incorporated ribonucleotides while those in mtDNA from tissues in adult mice were higher and varied between tissues. By measuring rNTPs and dNTPs in mouse embryos and spleen and skeletal muscle from adult mice via high-performance liquid chromatography, rNTP/dNTP ratios were determined to investigate whether NTP pools determine which ribonucleotides are incorporated in mtDNA. In accordance with previous findings[339], the ATP/deoxyadenosine triphosphate (dATP) ratio was by far the highest and was also found to vary the most between tissues. With respect to the nucleotide-specific discrimination ability of pol γ[273] this suggests that the ATP/dATP pools affect the ribonucleotide content in mtDNA to the greatest extent and confirms that the incorporated ribonucleotides in mtDNA reflect the present NTP pools[315]. Based on the observation that ribonucleotides were least common in mtDNA from embryos in comparison to mtDNA from adult tissues, an age-related increase of ribonucleotides in mtDNA was hypothesized. The hypothesis was tested by comparing the ribonucleotide content in mtDNA from embryos, pups, adult and aged mice. Interestingly, pups showed an intermediate level between the number of incorporated ribonucleotides determined in mtDNA from embryos and skeletal muscle in adult and aged mice, while there was no significant difference observed between the amounts in adult and aged mice. In contrast, the number of incorporated ribonucleotides in mtDNA from the hearts of pups was identical to the amount in embryos, but was increased significantly in mtDNA from hearts of adult mice. Both results show, however, that ribonucleotides are introduced up to adulthood, but do not increase in mtDNA of mice while aging. To further investigate how the ribonucleotide content in mtDNA reflects the present NTP pools, the ribonucleotide content of SAMHD1-deficient mice was investigated. SAMHD1 has the ability to hydrolyze all four dNTPs[340], its absence thereby increases dNTP pools. While the ribonucleotide content in nDNA was not affected by the deletion of SAMHD1, the increase in dNTPs was directly reflected in a decrease of incorporated ribonucleotides in mtDNA. Using HydEn-seq, I determined the base identity of incorporated ribonucleotides of mtDNA and nDNA of

SAMHD1-proficient and -deficient mice: Since SAMHD1 deficiency led to very few incorporated ribonucleotides in mtDNA, the detected distribution of base identity changed dramatically to reflect that of random nicks, by markedly decreasing incorporated rAMP, while loss of SAMHD1 had no effect on the ribonucleotide distribution in nDNA. Lastly, whether incorporated ribonucleotides contributed to mtDNA instability in aging was determined: We found that the mtDNA copy number was slightly higher in adult mice with SAMHD1 deficiency but not in old adults or aged mice. Furthermore, no differences in the number of deletions could be detected between aged SAMHD1-proficient and -deficient mice and no difference in life span of these animals was observed, indicating that incorporated ribonucleotides in mtDNA are not responsible for age-related mtDNA instability.

3.4 PAPER IV

Stably incorporated ribonucleotides in murine tissues: quantitation, base identity and distribution in nuclear and mitochondrial DNA

Stably incorporated ribonucleotides in DNA pose a possible threat to genome integrity due to their inherent predisposition to hydrolysis via the 2´-hydroxyl group[341]. While some beneficial roles of ribonucleotides have been found, as described earlier in section 1.4, their physiological relevance remains to be fully explained. Ribonucleotides are frequently misincorporated into nDNA and mtDNA[272,309]. In case of nDNA, efficient removal by RER was assumed[342], while ribonucleotides seem to be more tolerated in mtDNA where efficient removal pathways are lacking[309,315].

To gain more comprehensive knowledge of the significance of incorporated ribonucleotides in the DNA of healthy wild-type mammals, I used the approach outlined in Paper II to map and quantitate incorporated ribonucleotides in nDNA and mtDNA from nine different mouse tissues: blood, bone marrow, brain, heart, kidney, liver, lung, muscle and spleen. DNA samples were cleaved with SacI before preparing HydEn-seq libraries to allow for ribonucleotide quantitation. I determined the absolute numbers of incorporated ribonucleotides in nDNA and mtDNA and found statistically significant differences in the number of incorporated ribonucleotides among some of the investigated tissues. I found approximately 1 ribonucleotide per kb embedded in nDNA, while 1 to 8 ribonucleotides per kb were present in mtDNA, depending on the tissue. Analysing the base identity of incorporated ribonucleotides, I

found distinct differences between the ribonucleotides found in nDNA compared to mtDNA. Incorporated ribonucleotides in nDNA were found to be somewhat closer to the base composition of nDNA, with guanosine monophosphate (GMP) and adenosine monophosphate (AMP) incorporated proportionally to their occurrence in nDNA, but an about 2-fold overrepresentation of cytidine monophosphate (CMP) and an underrepresentation of uridine monophosphate (UMP). In contrast to the ribonucleotide content in nDNA, ribonucleotides in mtDNA were most frequently AMP, consistent with earlier findings and the high ATP/dATP ratio[343], while GMP was only overrepresented in mtDNA from kidney and lung and otherwise proportional to the mtDNA base composition, CMP was underrepresented and UMP was almost not detected at all. To study whether certain genomic features promote or prevent stably incorporated ribonucleotides in nDNA, I performed analyses of ribonucleotides near known or predicted positions of enhancers, tRNAs, promoters, protein binding sites, segmental duplications, CpG islands, microsatellites, TSS and intrastrand G4s. To compare the incorporated ribonucleotides near these features to the ribonucleotide distribution in the genome, I also generated a data set consisting of 21,000 positions by picking 1,000 random positions from each chromosome and subjected it to the same normalisation steps and binning. For genomic features present only on one of the strands, the analysis was split to reflect ribonucleotides on the same strand as the genomic feature or the opposite strand than the feature. Interestingly, I found that enhancers, promoters, TSS and protein binding sites have increased numbers of ribonucleotides in their proximity on the same and the opposite strand. Furthermore, tRNAs showed a marked increase on the same, but not the opposite strand. G4s showed decreased ribonucleotides just before the G4 on the same strand and a sharp increase in incorporated ribonucleotides on the opposite strand. For the analysis of segmental duplications, CpG islands and microsatellites, ribonucleotides on both strands were analysed. While CpG islands did not seem to change the number of incorporated ribonucleotides, segmental duplications showed a decrease in them. The analysis of microsatellites yielded an interesting pattern, where ribonucleotides were underrepresented at the very start position of a microsatellite but markedly increased above the background by about 2-fold at approximately 250 bp up- and downstream of this position and declining to background levels as represented by the random data set towards the periphery at about 500 bp up- and downstream of the feature. To investigate the ribonucleotide distribution in mtDNA, I calculated the mean number of incorporated ribonucleotides at each mtDNA position for the heavy and light strand. I found that ribonucleotides showed a non-

random distribution and were present at distinct hotspots. Moreover, I determined distinct peaks of ribonucleotides near OriH and OriL which may be the remnants of incomplete primer removal. These studies provide evidence that ribonucleotides are not only transiently inserted but also stay stably incorporated in the nuclear genome where RER pathways are present. Moreover, this may imply that ribonucleotides have a functional role in certain parts of the eukaryotic genome.

4 CONCLUDING REMARKS

The importance of balance between genome stability and instability for the evolutionary adaption or diversification of species and the flourishment of organisms is indisputable. A degree of variation in the genome allows for development of beneficial traits and adaption[344]. Imperfections in genome maintenance and repair mechanisms are contributors to this variation. An excess of changes or alterations, especially when affecting critical cellular functions, are however detrimental and associated with aging[5,6] and disease[4,345,346].

The body of knowledge on mechanisms of how DNA damage may come about and can be repaired by cellular repair mechanisms is continuously growing, as many questions surrounding them remain to be explored. This thesis ties together aspects surrounding genome maintenance and stability. In Paper I, I expanded the knowledge on pol η activity in the yeast genome and demonstrated that the division of labor among replicative DNA polymerases might extend to the specialized polymerases. The HydEn-seq method proved useful for tracking the activity of a specialized DNA polymerase and the method can be utilized to determine the genome-wide activities of the other known DNA polymerases. Moreover, a similar approach may be useful in determining DNA polymerase activities in humans, where additional specialized DNA polymerases are present. My findings also showed that pol η's lagging strand activity was abolished upon the loss of its C-terminus. Since the PCNA-Interacting Protein or Rev1-Interacting Region motif located in the C-terminus is the interface for interaction with PCNA and Rev1, my findings implicate them as immediate candidates to study the underlying mechanisms of pol η recruitment to the lagging strand.

I described an important advance in the HydEn-seq method in Paper II, allowing for simultaneous mapping and quantitation of incorporated ribonucleotides. Contrary to estimates based on DNA fragment sizes resulting from alkaline hydrolysis or RNase H2-treatment, this method provides the possibility for direct and genome-wide measurement while also obtaining base identity information. The possible applications are broad, as the HydEn-seq method and its expansion with restriction enzyme cleavage, can be adjusted to DNA from any given organism, requiring only a reference genome in the analysis of the acquired sequencing data. Using this method to produce atlases of the DNA polymerases' activities could paint a more comprehensive picture of the replication process, help define the roles of the specialized polymerases

and could potentially uncover overlapping or exclusive functionalities. The landscapes of incorporated ribonucleotides in genomes, their number and identity may also provide species-specific insight as to what levels of incorporated ribonucleotide are common and well-tolerated and which are detrimental.

MtDNA replication and maintenance are distinct from the processes replicating and maintaining nDNA[347]. Ribonucleotides seem to be more readily tolerated in mtDNA where effective mechanisms for their removal are lacking[309,315]. Our findings in Paper III suggest that mtDNA instability related to aging is not caused by incorporated ribonucleotides which accumulated in the mtDNA until adulthood but remained largely unchanged in older and aged mice. This finding would point at the idea that incorporated ribonucleotides in mtDNA are by enlarge a byproduct of replication by pol γ rather than actively fulfilling a specific function. How these incorporated ribonucleotides are tolerated in the mtDNA could be the subject of further inquiries. It is for example conceivable that mtDNA-interacting proteins (e.g. nucleoid-associated proteins)[348] stabilize ribonucleotides in the mtDNA by sterically hindering self-hydrolysis via the 2´-hydroxyl group.

I determined the number, identity and distribution of ribonucleotides in nDNA and mtDNA of nine mouse tissues in Paper IV and observed tissue-specific variations, in agreement with earlier findings of tissue-specific NTP pools. In extension of my experiments, it would be of interest to determine the precise NTP pools from the same biological materials. Moreover, it would be interesting to explore whether ribonucleotides contribute to tissue-specific genome instability based on the local NTP pools and repair conditions[349]. Upon analysis of the ribonucleotide distribution near genomic features, I determined distinct overrepresentation of ribonucleotides compared to random genomic positions near most of the studied genomic features. The generated data allows for further cross-referencing with other genomic features or regions of interest and poses a starting point for further experiments that explain my findings. Since I found an increase in ribonucleotides near protein binding sites, it would be interesting to test the hypothesis that ribonucleotides can be masked from RER recognition e.g. by reconstituting RER *in vitro* and determining the ribonucleotide removal efficiency in the presence and absence of DNA binding proteins. My findings near G-quadruplex DNA may in part be explained by my observation that CMP is 2- to 3-fold more frequently present in proportion to the base composition of nDNA than GMP. I see, however, about a 10-

fold increase of ribonucleotides opposite of intrastrand G-quadruplex DNA, which is probably not explained by the C-rich sequence alone. This marked overrepresentation of ribonucleotides on the opposite strand may point at substrate recognition being the main determining factor for efficient ribonucleotide removal (or lack thereof), as this strand likely remains single-stranded, may be bound by single strand binding proteins or may be able to form secondary structures itself[350,351] which could prevent recognition by RER.

Taken together, the work presented in this thesis has provided insights into the specialized pol η and the ribonucleotide landscape in nDNA and mtDNA of mammals. It may serve as the starting point for further lines of investigation in the pursuit of understanding the processes involved in maintaining or threatening genome stability.

5 ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to **Anders R. Clausen**, my supervisor. Thank you, for enabling me to learn and grow while performing exciting research! Throughout my years in the Clausen Lab, I've always felt valued not only as a researcher, but as a person. You've been a fantastic mentor and team leader! During tough times, you've never failed to understand, help and support me - a trait I've always been thankful for. It's been truly a pleasure to be part of your research group and to work with you. This is a chapter of my life I will always remember fondly. Thank you so much! I would also say you've succeeded in supervising your very first PhD student! ☺

I would also like to thank my co-supervisor **Claes Gustafsson** and all our collaborators, especially in **Levent Akyürek's Lab** who made my project possible!

Next, I must thank you, **Kerryn**. You've been one of the few 'constants' in an ever-changing research environment. You are an endless source of happiness and positive energy and I could always count on you as my favorite lab neighbor and dear friend. Thank you for all the chats, game nights, apple crumbles, amazing cakes and your help and support through difficult times. I really appreciate you!

Josephine, thank you for our wonderful time together. It's been lovely to start my journey in the Clausen Lab at the same time as you and I really missed you in the second half of my PhD. You've been such a dear friend, office mate and colleague! Thanks for all the talks, the lunches, the laughter and the Swedish support! Vielen Dank für diese unvergessliche Zeit!

I would also like to thank all the **Clausen Lab members**, present and past. **Alireza**, it's been such a pleasure to have you as a colleague and friend. Thanks for all our talks and your support! You've always been so kind and friendly and made the lab so welcoming! Thank you, **Delaney** for our time together during your project. You've been such a lovely office friend and colleague! **Liam**, thanks for easing my way into the world of bioinformatics! Thanks for the good conversations and time in the office! Thank you, **Mahmood,** for your positive energy, making the lab go around and keeping our fika traditions alive! **Martin**, we didn't spend too much time together, but you were probably right, there was a bioinformatician

in me, I just didn't know it yet! **Emma**, **Andranik**, **Yuika**, thank you all for the good times we've spend in the lab!

My heartfelt thanks goes to the **Larsson Lab members** past and present, pre-pandemic a.k.a. **"Lunch Bunch"**! I've always felt connected to you all, since our group is not so big! My thanks especially to **Ari**, **Markus**, **Martin**, **Susanne**, **Swaraj** and **Arman** for the endless lunches and fun game nights together! You've made this such a special experience and I've really missed you all while working from home!

Koviljka, though you are no longer here to see this, thank you for all our Swedish conversations. You've always been so kind!

I would also like to thank the **Falkenberg Lab** and the **Gustafsson Lab members**! Thanks for always helping out when our lab was in need of something, teaching me in the radio lab and the fun times spend together at the DNA transaction conferences!

*

I would like to thank all my **friends** from near and far who have shared this journey with me!

Cristiana, thank you so much for your friendship throughout this path! You've always been there with an open ear and supported me through all hardships! I have always loved our regular lunches, "PhD meetings" and felt strengthened and hopeful afterwards! I couldn't have done it without you!

Thanks also to everyone in the **"Lundberg lab crew"** (even though we all ended up somewhere else). It's been such a great pleasure to make you as my friends here in Sweden! Liebe "**Würzburger"**, **Laura**, **Öz**, **Isa**, **Anka**, **Fabi**, **Helen**, **Nicole**, es war eine lange Reise und ich bin dankbar, dass ihr alle ein Teil davon wart! Tack till alla mina **vänner** som har varit med en del av resan hit! **Kevin**, **Johannes**, **Fredrik**, **Marcus**, tack för alla roliga tider vi har haft - vi behöver definitivt spela mer DnD!

Liebe **Mama**, liber **Papa**, ohne eure Liebe, Hilfe und Unterstützung wäre ich niemals so weit gekommen. Ihr wart immer meine größten Fans, egal wohin mich die Wissenschaft geführt hat! Ihr habt immer an mich geglaubt und mich angefeuert, egal wie schwer ich es auch manchmal hatte. Ich danke euch von ganzem Herzen!

Oma, **Mamadote**, **Dote** und **Alexander**, vielen Dank für eure Unterstützung durch all meine Studien, die mich nach Schweden geführt haben! Es macht mich so stolz, dass ich diesen Erfolg mit euch teilen darf!

Lepski Opa, **Kreisel Oma** und **Opa**, ich wünschte ihr wärt hier und könnten mich jetzt sehen! Ihr fehlt mir und es wäre so schön gewesen, wenn ihr all das hier miterlebt hättet!

Ett hjärtligt tack till min "**svenska familj**"! **Carina** och **Lennart**, **Niklas** och **Patricia**, **Thomas**, **Micaela** och **Liam**, **Leif** och **Edith**, tack för att ni alltid har fått mig att känna mig som hemma här i Sverige! Ert stöd har verkligen gjort jättestor skillnad!

*

Last but not least, I would like to thank my beloved partner **Joakim**. I don't even know where to begin, but I know for certain that I couldn't have done this without you. You've been by my side from the very beginning of this journey and have supported me tirelessly throughout all the challenges and difficulties I've faced! I couldn't have wished for a better person by my side and I'm truly thankful for all you have done for me. I also cannot imagine how dehydrated and poorly fed I would have been! ☺
You've been an endless source of love, joy and support! You always know how to lighten my day and make me smile. Thank you so very much!

6 REFERENCES

1. Dahm R. Friedrich Miescher and the discovery of DNA. *Dev Biol.* 2005;278(2):274-288.
2. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. *Molecular Biology of the Cell.* 4th ed. New York: Garland Science; 2002.
3. Ciccia A, Elledge SJ. The DNA damage response: making it safe to play with knives. *Mol Cell.* 2010;40(2):179-204.
4. Aguilera A, Gomez-Gonzalez B. Genome instability: a mechanistic view of its causes and consequences. *Nat Rev Genet.* 2008;9(3):204-217.
5. Niedernhofer LJ, Gurkar AU, Wang YS, Vijg J, Hoeijmakers JHJ, Robbins PD. Nuclear genomic instability and aging. *Annu Rev Biochem.* 2018;87:295-322.
6. Vijg J, Dong X, Milholland B, Zhang L. Genome instability: a conserved mechanism of ageing? In: Kirkwood TBL, Korolchuk VI, eds. *Biological Mechanisms of Ageing.* Vol 61. London: Portland Press Ltd; 2017:305-315.
7. Felka T, Lemke J, Lemke C, Michel S, Liehr T, Claussen U. DNA degradation during maturation of erythrocytes - molecular cytogenetic characterization of Howell-Jolly bodies. *Cytogenet Genome Res.* 2007;119(1-2):2-8.
8. Fischer H, Szabo S, Scherz J, et al. Essential role of the keratinocyte-specific endonuclease DNase1L2 in the removal of nuclear DNA from hair and nails. *J Invest Dermatol.* 2011;131(6):1208-1215.
9. Avery OT, Macleod CM, McCarty M. Studies on the chemical nature of the substance inducting transformation of pneumococcal types - induction of transformation by a deoxyribonucleic acid fraction isolated from *Pneumococcus* type III. *J Exp Med.* 1944;79(2):137-158.
10. Franklin RE, Gosling RG. The structure of sodium thymonucleate fibres. I. The influence of water content. *Acta Crystallogr.* 1953;6(8-9):673-677.
11. Watson JD, Crick FHC. Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid. *Nature.* 1953;171(4356):737-738.
12. Kornberg A. Biologic synthesis of deoxyribonucleic acid. *Science.* 1960;131(3412):1503-1508.
13. Nirenberg MW, Leder P. RNA codewords and protein synthesis - the effect of trinucleotides upon the binding of sRNA to ribosomes. *Science.* 1964;145(363):1399-1407.

14. Smith HO, Wilcox KW. A restriction enzyme from *Hemophilus influenzae*. I. Purification and general properties. *J Mol Biol.* 1970;51(2):379-391.
15. Meselson M, Yuan R. DNA restriction enzyme from *E. coli*. *Nature.* 1968;217(5134):1110-1114.
16. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA.* 1977;74(12):5463-5467.
17. Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci USA.* 1977;74(2):560-564.
18. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harb Sym.* 1986;51:263-273.
19. Watson JD, Crick FHC. Genetical implications of the structure of deoxyribonucleic acid. *Nature.* 1953;171(4361):964-967.
20. Wang GL, Vasquez KM. Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. *DNA Repair.* 2014;19:143-151.
21. Zhao JH, Bacolla A, Wang GL, Vasquez KM. Non-B DNA structure-induced genetic instability and evolution. *Cell Mol Life Sci.* 2010;67(1):43-62.
22. Kaushik M, Kaushik S, Roy K, et al. A bouquet of DNA structures: emerging diversity. *Biochem Biophys Rep.* 2016;5:388-395.
23. Ishikawa F, Naito T. Why do we have linear chromosomes? A matter of Adam and Eve. *Mutation Research-DNA Repair.* 1999;434(2):99-107.
24. Duina AA, Miller ME, Keeney JB. Budding yeast for budding geneticists: a primer on the *Saccharomyces cerevisiae* model system. *Genetics.* 2014;197(1):33-48.
25. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 2016;44(D1):D733-D745.
26. Brown TA. *Genomes 2nd edition.* United Kingdom: BIOS Scientific Publishers Ltd; 2002.
27. Lee H, Zhang ZL, Krause HM. Long noncoding RNAs and repetitive elements: junk or intimate evolutionary partners? *Trends in Genetics.* 2019;35(12):892-902.
28. Dunham I, Kundaje A, Aldred SF, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414):57-74.
29. Kuska B. Should scientists scrap the notion of junk DNA? *J Natl Cancer I.* 1998;90(14):1032-1033.

30. Djebali S, Davis CA, Merkel A, et al. Landscape of transcription in human cells. *Nature.* 2012;489(7414):101-108.
31. Harashima H, Dissmeyer N, Schnittger A. Cell cycle control across the eukaryotic kingdom. *Trends in Cell Biol.* 2013;23(7):345-356.
32. Rivera-Mulia JC, Gilbert DM. Replicating large genomes: divide and conquer. *Mol Cell.* 2016;62(5):756-765.
33. Marahrens Y, Stillman B. A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science.* 1992;255(5046):817-823.
34. Bell SP, Kobayashi R, Stillman B. Yeast origin recognition complex functions in transcription and silencing and DNA replication. *Science.* 1993;262(5141):1844-1849.
35. Prioleau M-N, MacAlpine DM. DNA replication origins - where do we begin? *Genes Dev.* 2016;30:1683-1697.
36. Deegan TD, Diffley JFX. MCM: one ring to rule them all. *Curr Opin Struc Biol.* 2016;37:145-151.
37. Burgers PMJ, Kunkel TA. Eukaryotic DNA replication fork. *Annu Rev Biochem.* 2017;86:417-438.
38. Boos D, Frigola J, Diffley JFX. Activation of the replicative DNA helicase: breaking up is hard to do. *Curr Opin Cell Biol.* 2012;24(3):423-430.
39. Baranovskiy AG, Babayeva ND, Zhang YB, et al. Mechanism of concerted RNA-DNA primer synthesis by the human primosome. *J Biol Chem.* 2016;291(19):10006-10020.
40. Yuzhakov A, Kelman Z, Hurwitz J, O'Donnell M. Multiple competition reactions for RPA order the assembly of the DNA polymerase δ holoenzyme. *Embo J.* 1999;18(21):6189-6199.
41. Perera RL, Torella R, Klinge S, Kilkenny ML, Maman JD, Pellegrini L. Mechanism for priming DNA synthesis by yeast DNA polymerase α. *eLife.* 2013;2.
42. Shinbrot E, Henninger EE, Weinhold N, et al. Exonuclease mutations in DNA polymerase ε reveal replication strand specific mutation patterns and human origins of replication. *Genome Res.* 2014;24(11):1740-1750.
43. Clausen AR, Lujan SA, Burkholder AB, et al. Tracking replication enzymology *in vivo* by genome-wide mapping of ribonucleotide incorporation. *Nat Struct Mol Biol.* 2015;22(3):185-191.
44. Pursell ZF, Isoz I, Lundstrom EB, Johansson E, Kunkel TA. Yeast DNA polymerase ε participates in leading-strand DNA replication. *Science.* 2007;317(5834):127-130.
45. Georgescu RE, Langston L, Yao NY, et al. Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nat Struct Mol Biol.* 2014;21(8):664-670.

46. Yeeles JTP, Janska A, Early A, Diffley JFX. How the eukaryotic replisome achieves rapid and efficient DNA replication. *Mol Cell.* 2017;65(1):105-116.
47. Johnson RE, Klassen R, Prakash L, Prakash S. A major role of DNA polymerase δ in replication of both the leading and lagging DNA strands. *Mol Cell.* 2015;59(2):163-175.
48. Stillman B. Reconsidering DNA polymerases at the replication fork in eukaryotes. *Mol Cell.* 2015;59(2):139-141.
49. Garbacz MA, Lujan SA, Burkholder AB, et al. Evidence that DNA polymerase δ contributes to initiating leading strand DNA replication in *Saccharomyces cerevisiae*. *Nat Commun.* 2018;9:1-11.
50. Daigaku Y, Keszthelyi A, Muller CA, et al. A global profile of replicative polymerase usage. *Nat Struct Mol Biol.* 2015;22(3):192-198.
51. Pellegrini L. The Pol α-Primase Complex. In: MacNeill S, ed. *The Eukaryotic Replisome: a Guide to Protein Structure and Function.* Dordrecht: Springer; 2012.
52. Mondol T, Stodola JL, Galletto R, Burgers PM. PCNA accelerates the nucleotide incorporation rate by DNA polymerase δ. *Nucleic Acids Res.* 2019;47(4):1977-1986.
53. McGuffee SR, Smith DJ, Whitehouse I. Quantitative, genome-wide analysis of eukaryotic replication initiation and termination. *Mol Cell.* 2013;50(1):123-135.
54. Nick McElhinny SA, Gordenin DA, Stith CM, Burgers PMJ, Kunkel TA. Division of labor at the eukaryotic replication fork. *Mol Cell.* 2008;30(2):137-144.
55. Chen YH, Keegan S, Kahli M, et al. Transcription shapes DNA replication initiation and termination in human cells. *Nat Struct Mol Biol.* 2019;26(1):67-77.
56. Keszthelyi A, Minchell NE, Baxter J. The causes and consequences of topological stress during DNA replication. *Genes.* 2016;7(12):1-13.
57. Heintzman DR, Campos LV, Byl JAW, Osheroff N, Dewar JM. Topoisomerase II is crucial for fork convergence during vertebrate replication termination. *Cell Rep.* 2019;29(2):422-436.
58. Fachinetti D, Bermejo R, Cocito A, et al. Replication termination at eukaryotic chromosomes is mediated by Top2 and occurs at genomic loci containing pausing elements. *Mol Cell.* 2010;39(4):595-605.
59. Low E, Chistol G, Zaher MS, Kochenova OV, Walter JC. The DNA replication fork suppresses CMG unloading from chromatin before termination. *Genes Dev.* 2020;34(21-22):1534-1545.

60. Moreno SP, Bailey R, Campion N, Herron S, Gambus A. Polyubiquitylation drives replisome disassembly at the termination of DNA replication. *Science.* 2014;346(6208):477-481.
61. Maric M, Maculins T, De Piccoli G, Labib K. Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication. *Science.* 2014;346(6208):440-452.
62. Roger AJ, Munoz-Gomez SA, Kamikawa R. The origin and diversification of mitochondria. *Curr Biol.* 2017;27(21):R1177-R1192.
63. Falkenberg M, Gustafsson CM. Mammalian mitochondrial DNA replication and mechanisms of deletion formation. *Crit Rev Biochem Mol Biol.* 2020;55(6):509-524.
64. Kazachkova N, Ramos A, Santos C, Lima M. Mitochondrial DNA damage patterns and aging: revising the evidences for humans and mice. *Aging Dis.* 2013;4(6):337-350.
65. Foury F, Roganti T, Lecrenier N, Purnelle B. The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *Febs Lett.* 1998;440(3):325-331.
66. Youle RJ. Mitochondria - striking a balance between host and endosymbiont. *Science.* 2019;365(6454):655-664.
67. Cole LW. The evolution of per-cell organelle number. *Front Cell Dev Biol.* 2016;4:1-7.
68. Giles RE, Blanc H, Cann HM, Wallace DC. Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci USA.* 1980;77(11):6715-6719.
69. Hutchison CA, Newbold JE, Potter SS, Edgell MH. Maternal inheritance of mammalian mitochondrial DNA. *Nature.* 1974;251(5475):536-538.
70. Berk AJ, Clayton DA. Mechanism of mitochondrial DNA replication in mouse L-cells - asynchronous replication of strands, segregation of circular daughter molecules, aspects of topology and turnover of an initiation sequence. *J Mol Biol.* 1974;86(4):801-824.
71. Freel KC, Friedrich A, Schacherer J. Mitochondrial genome evolution in yeasts: an all-encompassing view. *Fems Yeast Res.* 2015;15(4):1-9.
72. Williamson D. Timeline - the curious history of yeast mitochondrial DNA. *Nat Rev Genet.* 2002;3(6):475-481.
73. Solieri L. Mitochondrial inheritance in budding yeasts: towards an integrated understanding. *Trends Microbiol.* 2010;18(11):521-530.
74. Bibb MJ, Vanetten RA, Wright CT, Walberg MW, Clayton DA. Sequence and gene organization of mouse mitochondrial DNA. *Cell.* 1981;26(2):167-180.

75. de Grey A. Forces maintaining organellar genomes: is any as strong as genetic code disparity or hydrophobicity? *Bioessays.* 2005;27(4):436-446.
76. Bjorkholm P, Harish A, Hagstrom E, Ernst AM, Andersson SGE. Mitochondrial genomes are retained by selective constraints on protein targeting. *Proc Natl Acad Sci USA.* 2015;112(33):10154-10161.
77. Martin W, Schnarrenberger C. The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. *Curr Genet.* 1997;32(1):1-18.
78. Wolstenholme DR. Genetic novelties in mitochondrial genomes of multicellular animals. *Curr Opin Genet Dev.* 1992;2(6):918-925.
79. Allen JF. Why chloroplasts and mitochondria retain their own genomes and genetic systems: colocation for redox regulation of gene expression. *Proc Natl Acad Sci USA.* 2015;112(33):10231-10238.
80. Shutt TE, Gray MW. Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends Genet.* 2006;22(2):90-95.
81. Robberson DL, Vinograd J, Kasamatsu H. Replication of mitochondrial DNA - circular replicative intermediates in mouse L-cells. *Proc Natl Acad Sci USA.* 1972;69(3):737-741.
82. Falkenberg M. Mitochondrial DNA replication in mammalian cells: overview of the pathway. *Essays Biochem.* 2018;62(3):287-296.
83. Tapper DP, Clayton DA. Mechanism of replication of human mitochondrial DNA - localization of the 5´-ends of nascent daughter strands. *J Biol Chem.* 1981;256(10):5109-5115.
84. Chang DD, Clayton DA. Priming of human mitochondrial DNA replication occurs at the light strand promoter. *Proc Natl Acad Sci USA.* 1985;82(2):351-355.
85. Chang DD, Hauswirth WW, Clayton DA. Replication priming and transcription initiate from precisely the same site in mouse mitochondrial DNA. *Embo J.* 1985;4(6):1559-1567.
86. Gillum AM, Clayton DA. Mechanism of mitochondrial DNA replication in mouse L-cells - RNA priming during the initiation of heavy-strand synthesis. *J Mol Biol.* 1979;135(2):353-368.
87. Pham XH, Farge G, Shi YH, Gaspari M, Gustafsson CM, Falkenberg M. Conserved sequence box II directs transcription termination and primer formation in mitochondria. *J Biol Chem.* 2006;281(34):24647-24652.
88. Wanrooij PH, Uhler JP, Shi YH, Westerlund F, Falkenberg M, Gustafsson CM. A hybrid G-quadruplex structure formed

between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop. *Nucleic Acids Res.* 2012;40(20):10334-10344.
89. Posse V, Al-Behadili A, Uhler JP, et al. RNase H1 directs origin-specific initiation of DNA replication in human mitochondria. *Plos Genet.* 2019;15(1):1-20.
90. Korhonen JA, Gaspari M, Falkenberg M. TWINKLE has 5′ → 3′ DNA helicase activity and is specifically stimulated by mitochondrial single-stranded DNA-binding protein. *J Biol Chem.* 2003;278(49):48627-48632.
91. Bolden A, Noy GP, Weissbach A. DNA polymerase of mitochondria is a γ polymerase. *J Biol Chem.* 1977;252(10):3351-3356.
92. Kaguni LS. DNA polymerase γ, the mitochondrial replicase. *Annu Rev Biochem.* 2004;73:293-320.
93. Nicholls TJ, Minczuk M. In D-loop: 40 years of mitochondrial 7S DNA. *Exp Gerontol.* 2014;56:175-181.
94. Fuste JM, Shi YH, Wanrooij S, et al. *In vivo* occupancy of mitochondrial single-stranded DNA binding protein supports the strand displacement mode of DNA replication. *Plos Genet.* 2014;10(12):1-11.
95. Fuste JM, Wanrooij S, Jemt E, et al. Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication. *Mol Cell.* 2010;37(1):67-78.
96. Wong TW, Clayton DA. *In vitro* replication of human mitochondrial DNA: accurate initiation at the origin of light-strand synthesis. *Cell.* 1985;42(3):951-958.
97. Lakshmipathy U, Campbell C. The human DNA ligase III gene encodes nuclear and mitochondrial proteins. *Mol Cell Biol.* 1999;19(5):3869-3876.
98. Uhler JP, Thorn C, Nicholls TJ, et al. MGME1 processes flaps into ligatable nicks in concert with DNA polymerase gamma during mtDNA replication. *Nucleic Acids Res.* 2016;44(12):5861-5871.
99. Kornblum C, Nicholls TJ, Haack TB, et al. Loss-of-function mutations in MGME1 impair mtDNA replication and cause multisystemic mitochondrial disease. *Nat Genet.* 2013;45(2):214-219.
100. Amalric F, Merkel C, Gelfand R, Attardi G. Fractionation of mitochondrial RNA from HeLa cells by high-resolution electrophoresis under strongly denaturing conditions. *J Mol Biol.* 1978;118(1):1-25.
101. Jemt E, Persson O, Shi YH, et al. Regulation of DNA replication at the end of the mitochondrial D-loop involves the helicase

TWINKLE and a conserved sequence element. *Nucleic Acids Res.* 2015;43(19):9262-9275.
102. Ruhanen H, Ushakov K, Yasukawa T. Involvement of DNA ligase III and ribonuclease H1 in mitochondrial DNA replication in cultured human cells. *B B A - Mol Cell Res.* 2011;1813(12):2000-2007.
103. Lima WF, Rose JB, Nichols JG, et al. Human RNase H1 discriminates between subtle variations in the structure of the heteroduplex substrate. *Mol Pharmacol.* 2007;71(1):83-91.
104. Al-Behadili A, Uhler JP, Berglund AK, et al. A two-nuclease pathway involving RNase H1 is required for primer removal at human mitochondrial OriL. *Nucleic Acids Res.* 2018;46(18):9471-9483.
105. McVey M, Khodaverdian VY, Meyer D, Cerqueira PG, Heyer WD. Eukaryotic DNA polymerases in homologous recombination. *Ann Rev Genet.* 2016;50:393-421.
106. Shanbhag V, Sachdev S, Flores JA, Modak MJ, Singh K. Family A and B DNA polymerases in cancer: opportunities for therapeutic interventions. *Biology.* 2018;7(1):1-16.
107. Chen XJ. Mechanism of homologous recombination and implications for aging-related deletions in mitochondrial DNA. *Microbiol Mol Biol Rev.* 2013;77(3):476-496.
108. Krasich R, Copeland WC. DNA polymerases in the mitochondria: a critical review of the evidence. *Front Biosci-Landmark.* 2017;22:692-709.
109. Wang Z, Song YD, Li SB, et al. DNA polymerase θ (POLQ) is important for repair of DNA double-strand breaks caused by fork collapse. *J Biol Chem.* 2019;294(11):3909-3919.
110. Ceccaldi R, Liu JC, Amunugama R, et al. Homologous recombination-deficient tumours are dependent on pol θ-mediated repair. *Nature.* 2015;518(7538):258-U306.
111. Mateos-Gomez PA, Gong FD, Nair N, Miller KM, Lazzerini-Denchi E, Sfeir A. Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature.* 2015;518(7538):254-U285.
112. Yoshimura M, Kohzaki M, Nakamura J, et al. Vertebrate pol θ and pol β cooperate in base excision repair of oxidative DNA damage. *Mol Cell.* 2006;24(1):115-125.
113. Kent T, Chandramouly G, McDevitt SM, Ozdemir AY, Pomerantz RT. Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase theta. *Nat Struct Mol Biol.* 2015;22(3):230-237.
114. Moldovan GL, Madhavan MV, Mirchandani KD, McCaffrey RM, Vinciguerra P, D'Andrea AD. DNA polymerase POLN

participates in cross-link repair and homologous recombination. *Mol Cell Biol.* 2010;30(4):1088-1096.
115. Takata K, Tomida J, Reh S, et al. Conserved overlapping gene arrangement, restricted expression, and biochemical activities of DNA polymerase ν (POLN). *J Biol Chem.* 2015;290(40):24278-24293.
116. Kazlauskas D, Krupovic M, Guglielmini J, Forterre P, Venclovas C. Diversity and evolution of B-family DNA polymerases. *Nucleic Acids Res.* 2020;48(18):10142-10156.
117. Wang X, Ira G, Tercero JA, Holmes AM, Diffley JFX, Haber JE. Role of DNA replication proteins in double-strand break-induced recombination in *Saccharomyces cerevisiae*. *Mol Cell Biol.* 2004;24(16):6891-6899.
118. Holmes AM, Haber JE. Double-strand break repair in yeast requires both leading and lagging strand DNA polymerases. *Cell.* 1999;96(3):415-424.
119. Li X, Stith CM, Burgers PM, Heyer WD. PCNA is required for initiation of recombination-associated DNA synthesis by DNA polymerase δ. *Mol Cell.* 2009;36(4):704-713.
120. Maloisel L, Bhargava J, Roeder GS. A role for DNA polymerase δ in gene conversion and crossing over during meiosis in *Saccharmyces cerevisiae*. *Genetics.* 2004;167:1133-1142.
121. Byrnes JJ, Downey KM, Black VL, So AG. A new mammalian DNA polymerase with 3′ to 5′ exonuclease activity: DNA polymerase δ. *Biochemistry* 1976;15:2817-2823.
122. Sharma S, Hicks JK, Chute CL, et al. REV1 and polymerase ζ facilitate homologous recombination repair. *Nucleic Acids Res.* 2012;40(2):682-691.
123. Li J, Holzschu DL, Sugiyama T. PCNA is efficiently loaded on the DNA recombination intermediate to modulate polymerase δ, η, and ζ activities. *Proc Natl Acad Sci USA.* 2013;110(19):7672-7677.
124. Gan GN, Wittschieben JP, Wittschieben BO, Wood RD. DNA polymerase zeta (pol ζ) in higher eukaryotes. *Cell Res.* 2008;18(1):174-183.
125. Roberts SA, Sterling J, Thompson C, et al. Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol Cell.* 2012;46(4):424-435.
126. Yamtich J, Sweasy JB. DNA polymerase family X: function, structure, and cellular roles. *B B A - Proteins Proteom.* 2010;1804(5):1136-1150.
127. Maga G, Villani G, Ramadan K, et al. Human DNA polymerase λ functionally and physically interacts with proliferating cell nuclear antigen in normal and translesion DNA synthesis. *J Biol Chem.* 2002;277(50):48434-48440.

128. Bebenek K, Pedersen LC, Kunkel TA. Structure-function studies of DNA polymerase λ. *Biochemistry.* 2014;53(17):2781-2792.
129. Pardo B, Ma E, Marcand S. Mismatch tolerance by DNA polymerase Pol4 in the course of nonhomologous end joining in *Saccharomyces cerevisiae*. *Genetics.* 2006;172(4):2689-2694.
130. Lee K, Lee SE. *Saccharomyces cerevisiae* Sae2- and tell-dependent single-strand DNA formation at DNA break promotes microhomology-mediated end joining. *Genetics.* 2007;176(4):2003-2014.
131. Plug AW, Clairmont CA, Sapi E, Ashley T, Sweasy JB. Evidence for a role for DNA polymerase β in mammalian meiosis. *Proc Natl Acad Sci USA.* 1997;94(4):1327-1331.
132. Puigvert JC, Sanjiv K, Helleday T. Targeting DNA repair, DNA metabolism and replication stress as anti-cancer strategies. *Febs J.* 2016;283(2):232-245.
133. Vaisman A, Chaney SG. The efficiency and fidelity of translesion synthesis past cisplatin and oxaliplatin GpG adducts by human DNA polymerase β. *J Biol Chem.* 2000;275(17):13017-13025.
134. Havener JM, McElhinny SAN, Bassett E, Gauger M, Ramsden DA, Chaney SG. Translesion synthesis past platinum DNA adducts by human DNA polymerase μ. *Biochemistry.* 2003;42(6):1777-1788.
135. Covo S, Blanco L, Livneh Z. Lesion bypass by human DNA polymerase μ reveals a template-dependent, sequence-independent nucleotidyl transferase activity. *J Biol Chem.* 2004;279(2):859-865.
136. Lieber MR. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem.* 2010;79:181-211.
137. Gouge J, Rosario S, Romain F, Poitevin F, Beguin P, Delarue M. Structural basis for a novel mechanism of DNA bridging and alignment in eukaryotic DSB DNA repair. *Embo J.* 2015;34(8):1126-1142.
138. Motea EA, Berdis AJ. Terminal deoxynucleotidyl transferase: the story of a misguided DNA polymerase. *B B A - Proteins Proteom.* 2010;1804(5):1151-1166.
139. Yang W. An overview of Y-family DNA polymerases and a case study of human DNA polymerase η. *Biochemistry.* 2014;53(17):2793-2803.
140. Kawamoto T, Araki K, Sonoda E, et al. Dual roles for DNA polymerase η in homologous DNA recombination and translesion DNA synthesis. *Mol Cell.* 2005;20(5):793-799.
141. Gali VK, Balint E, Serbyn N, Frittmann O, Stutz F, Unk I. Translesion synthesis DNA polymerase η exhibits a specific

RNA extension activity and a transcription-associated function. *Sci Rep.* 2017;7:1-17.
142. Boldinova EO, Ignatov A, Kulbachinskiy A, Makarova AV. The active site residues Gln55 and Arg73 play a key role in DNA damage bypass by *S. cerevisiae* Pol η. *Sci Rep.* 2018;8:1-11.
143. Bassett E, King NM, Bryant MF, et al. The role of DNA polymerase η in translesion synthesis past platinum-DNA adducts in human fibroblasts. *Cancer Res.* 2004;64(18):6469-6475.
144. Ling H, Boudsocq F, Plosky BS, Woodgate R, Yang W. Replication of a cis-syn thymine dimer at atomic resolution. *Nature.* 2003;424(6952):1083-1087.
145. Di Nola JM, Neuberger MS. Molecular mechanisms of antibody somatic hypermutation. In: *Ann. Rev. Biochem.* Vol 76. Palo Alto: Annual Reviews; 2007:1-22.
146. Sebesta M, Burkovics P, Juhasz S, et al. Role of PCNA and TLS polymerases in D-loop extension during homologous recombination in humans. *DNA Repair.* 2013;12(9):691-698.
147. Tonzi P, Yin YD, Wei C, Lee T, Rothenberg E, Huang TT. Translesion polymerase κ-dependent DNA synthesis underlies replication fork recovery. *eLife.* 2018;7:1-28.
148. Vaisman A, Woodgate R. Mysterious and fascinating: DNA polymerase ι remains enigmatic 20 years after its discovery. *DNA Repair.* 2020;93.
149. Tissier A, McDonald JP, Frank EG, Woodgate R. Pol ι, a remarkably error-prone human DNA polymerase. *Genes Dev.* 2000;14(13):1642-1650.
150. Mito E, Mokhnatkin JV, Steele MC, et al. Mutagenic and recombinagenic responses to defective DNA polymerase δ are facilitated by the Rev1 protein in pol3-t mutants of *Saccharomyces cerevisiae*. *Genetics.* 2008;179(4):1795-1806.
151. Nayak S, Calvo JA, Cong K, et al. Inhibition of the translesion synthesis polymerase REV1 exploits replication gaps as a cancer vulnerability. *Sci Adv.* 2020;6(24).
152. Bailey LJ, Bianchi J, Doherty AJ. PrimPol is required for the maintenance of efficient nuclear and mitochondrial DNA replication in human cells. *Nucleic Acids Research.* 2019;47(8):4026-4038.
153. Boldinova EO, Wanrooij PH, Shilkin ES, Wanrooij S, Makarova AV. DNA Damage Tolerance by Eukaryotic DNA Polymerase and Primase PrimPol. *International Journal of Molecular Sciences.* 2017;18(7).
154. Mouron S, Rodriguez-Acebes S, Martinez-Jimenez MI, et al. Repriming of DNA synthesis at stalled replication forks by human PrimPol. *Nat Struct Mol Biol.* 2013;20(12):1383-1389.

155. Wan L, Lou JM, Xia YS, et al. hPrimpol1/CCDC111 is a human DNA primase-polymerase required for the maintenance of genome integrity. *Embo Rep.* 2013;14(12):1104-1112.
156. Lecrenier N, VanDerBruggen P, Foury F. Mitochondrial DNA polymerases from yeast to man: a new family of polymerases. *Gene.* 1997;185(1):147-152.
157. Seki M, Marini F, Wood RD. POLQ (Pol θ), a DNA polymerase and DNA-dependent ATPase in human cells. *Nucleic Acids Res.* 2003;31(21):6117-6126.
158. Marini F, Kim N, Schuffert A, Wood RD. POLN, a nuclear PolA family DNA polymerase homologous to the DNA cross-link sensitivity protein Mus308. *J Biol Chem.* 2003;278(34):32014-32019.
159. Kazlauskas D, Krupovic M, Venclovas C. The logic of DNA replication in double-stranded DNA viruses: insights from global analysis of viral genomes. *Nucleic Acids Res.* 2016;44(10):4551-4564.
160. Roush AA, Suarez M, Friedberg EC, Radman M, Siede W. Deletion of the *Saccharomyces cerevisiae* gene RAD30 encoding an *Escherichia coli* DinB homolog confers UV radiation sensitivity and altered mutability. *Mol Gen Genet.* 1998;257(6):686-692.
161. Johnson RE, Prakash S, Prakash L. Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, pol η. *Science.* 1999;283(5404):1001-1004.
162. McCulloch SD, Kokoska RJ, Masutani C, Iwai S, Hanaoka F, Kunkel TA. Preferential *cis-syn* thymine dimer bypass by DNA polymerase η occurs with biased fidelity. *Nature.* 2004;428(6978):97-100.
163. Yoon JH, Prakash L, Prakash S. Highly error-free role of DNA polymerase η in the replicative bypass of UV-induced pyrimidine dimers in mouse and human cells. *Proc Natl Acad Sci USA.* 2009;106(43):18219-18224.
164. Biertumpfel C, Zhao Y, Kondo Y, et al. Structure and mechanism of human DNA polymerase η. *Nature.* 2010;476(7360):360-360.
165. Haracska L, Yu SL, Johnson RE, Prakash L, Prakash S. Efficient and accurate replication in the presence of 7,8-dihydro-8-oxoguanine by DNA polymerase η. *Nat Genet.* 2000;25(4):458-461.
166. Bresson A, Fuchs RPP. Lesion bypass in yeast cells: pol η participates in a multi-DNA polymerase process. *Embo J.* 2002;21(14):3881-3887.

167. Johnson RE, Haracska L, Prakash S, Prakash L. Role of DNA polymerase η in the bypass of a (6-4) TT photoproduct. *Mol Cell Biol.* 2001;21(10):3558-3563.
168. Quinet A, Martins DJ, Vessoni AT, et al. Translesion synthesis mechanisms depend on the nature of DNA damage in UV-irradiated human cells. *Nucleic Acids Res.* 2016;44(12):5717-5731.
169. Kreisel K, Engqvist MKM, Kalm J, et al. DNA polymerase η contributes to genome-wide lagging strand synthesis. *Nucleic Acids Res.* 2019;47(5):2425-2435.
170. Delbos F, Aoufouchi S, Faili A, Weill JC, Reynaud CA. DNA polymerase η is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. *J Exp Med.* 2007;204(1):17-23.
171. Zeng XM, Winter DB, Kasmer C, Kraemer KH, Lehmann AR, Gearhart PJ. DNA polymerase η is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. *Nat Immunol.* 2001;2(6):537-541.
172. Bergoglio V, Boyer AS, Walsh E, et al. DNA synthesis by pol η promotes fragile site stability by preventing under-replicated DNA in mitosis. *J Cell Biol.* 2013;201(3):395-408.
173. Rey L, Sidorova JM, Puget N, et al. Human DNA polymerase η is required for common fragile site stability during unperturbed DNA replication. *Mol Cell Biol.* 2009;29(12):3344-3354.
174. Bignell GR, Greenman CD, Davies H, et al. Signatures of mutation and selection in the cancer genome. *Nature.* 2010;463(7283):893-U861.
175. Sassa A, Tada H, Takeishi A, et al. Processing of a single ribonucleotide embedded into DNA by human nucleotide excision repair and DNA polymerase η. *Sci Rep.* 2019;9:1-10.
176. Roy U, Greene EC. Demystifying the D-loop during DNA recombination. *Nature.* 2020;586(7831):677-678.
177. McIlwraith MJ, Vaisman A, Liu YL, Fanning E, Woodgate R, West SC. Human DNA polymerase η promotes DNA synthesis from strand invasion intermediates of homologous recombination. *Mol Cell.* 2005;20(5):783-792.
178. Sneeden JL, Grossi SM, Tappin I, Hurwitz J, Heyer WD. Reconstitution of recombination-associated DNA synthesis with human proteins. *Nucleic Acids Res.* 2013;41(9):4913-4925.
179. Limoli CL, Giedzinski E, Cleaver JE. Alternative recombination pathways in UV-irradiated XP variant cells. *Oncogene.* 2005;24(23):3708-3714.
180. Nicolay NH, Carter R, Hatch SB, et al. Homologous recombination mediates S-phase-dependent radioresistance in

cells deficient in DNA polymerase η. *Carcinogenesis.* 2012;33(11):2026-2034.
181. Strom L, Sjogren C. DNA damage-induced cohesion. *Cell Cycle.* 2005;4(4):536-539.
182. Enervald E, Lindgren E, Katou Y, Shirahige K, Strom L. Importance of pol η for damage-induced cohesion reveals differential regulation of cohesion establishment at the break site and genome-wide. *Plos Genet.* 2013;9(1):1-17.
183. Wu PS, Enervald E, Joelsson A, et al. Post-translational regulation of DNA polymerase η, a connection to damage-induced cohesion in *Saccharomyces cerevisiae*. *Genetics.* 2020;216(4):1009-1022.
184. Pope-Varsalona H, Liu FJ, Guzik L, Opresko PL. Polymerase η suppresses telomere defects induced by DNA damaging agents. *Nucleic Acids Res.* 2014;42(21):13096-13109.
185. Garcia-Exposito L, Bournique E, Bergoglio V, et al. Proteomic profiling reveals a specific role for translesion DNA polymerase η in the alternative lengthening of telomeres. *Cell Rep.* 2016;17(7):1858-1871.
186. Johnson RE, Kondratick CM, Prakash S, Prakash L. hRAD30 mutations in the variant form of *Xeroderma pigmentosum*. *Science.* 1999;285(5425):263-265.
187. Masutani C, Kusumoto R, Yamada A, et al. The XPV (*Xeroderma pigmentosum* variant) gene encodes human DNA polymerase η. *Nature.* 1999;399(6737):700-704.
188. Lehmann J, Seebode C, Martens MC, Emmert S. *Xeroderma pigmentosum* - facts and perspectives. *Aktuelle Derm.* 2018;44(5):232-236.
189. Lehmann AR, Kirkbell S, Arlett CF, et al. *Xeroderma pigmentosum* cells with normal levels of excision repair have a defect in DNA synthesis after UV-irradiation. *Proc Natl Acad Sci USA.* 1975;72(1):219-223.
190. Lerner LK, Francisco G, Soltys DT, et al. Predominant role of DNA polymerase η and p53-dependent translesion synthesis in the survival of ultraviolet-irradiated human cells. *Nucleic Acids Res.* 2017;45(3):1270-1280.
191. Albertella MR, Green CM, Lehmann AR, O'Connor MJ. A role for polymerase η in the cellular tolerance to cisplatin-induced damage. *Cancer Res.* 2005;65(21):9799-9806.
192. Srivastava AK, Han CH, Zhao R, et al. Enhanced expression of DNA polymerase η contributes to cisplatin resistance of ovarian cancer stem cells. *Proc Natl Acad Sci USA.* 2015;112(14):4411-4416.

193. Masutani C, Kusumoto R, Iwai S, Hanaoka F. Mechanisms of accurate translesion synthesis by human DNA polymerase η. *Embo J.* 2000;19(12):3100-3109.
194. Saha P, Mandal T, Talukdar AD, et al. DNA polymerase η: a potential pharmacological target for cancer therapy. *J Cell Physiol.*236:4106-4120.
195. Chatterjee N, Walker GC. Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mole Mutagen.* 2017;58(5):235-263.
196. Hasty P, Campisi J, Hoeijmakers J, van Steeg H, Vijg J. Aging and genome maintenance: lessons from the mouse? *Science.* 2003;299(5611):1355-1359.
197. Schuch AP, Moreno NC, Schuch NJ, Menck CFM, Garcia CCM. Sunlight damage to cellular DNA: focus on oxidatively generated lesions. *Free Radical Biol Med.* 2017;107:110-124.
198. Douki T. Pyrimidine (6-4) pyrimidone photoproducts in UVA-irradiated DNA: photosensitization or photoisomerization? *Chemphotochem.* 2020;4(4):294-299.
199. Jiang Y, Rabbi M, Kim M, et al. UVA generates pyrimidine dimers in DNA directly. *Biophys J.* 2009;96(3):1151-1158.
200. Mouret S, Philippe C, Gracia-Chantegrel J, et al. UVA-induced cyclobutane pyrimidine dimers in DNA: a direct photochemical mechanism? *Org Biomol Chem.* 2010;8(7):1706-1711.
201. Besaratinia A, Yoon JI, Schroeder C, Bradforth SE, Cockburn M, Pfeifer GP. Wavelength dependence of ultraviolet radiation-induced DNA damage as determined by laser irradiation suggests that cyclobutane pyrimidine dimers are the principal DNA lesions produced by terrestrial sunlight. *Faseb J.* 2011;25(9):3079-3091.
202. Wondrak GT, Jacobson MK, Jacobson EL. Endogenous UVA-photosensitizers: mediators of skin photodamage and novel targets for skin photoprotection. *Photochem Photobiol Sci.* 2006;5(2):215-237.
203. Cadet J, Douki T, Ravanat JL. Oxidatively generated damage to cellular DNA by UVB and UVA radiation. *Photochem Photobiol.* 2015;91(1):140-155.
204. Borrego-Soto G, Ortiz-Lopez R, Rojas-Martinez A. Ionizing radiation-induced DNA injury and damage detection in patients with breast cancer. *Genet Mol Biol.* 2015;38(4):420-432.
205. Vromman V, Maghuin-Rogister G, Vleminckx C, Saegerman C, Pussemier L, Huyghebaert A. Risk ranking priority of carcinogenic and/or genotoxic environmental contaminants in food in Belgium. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess.* 2014;31(5):872-888.

206. Chen ZH, Wang DP, Liu X, et al. Oxidative DNA damage is involved in cigarette smoke-induced lung injury in rats. *Environ Health Prev Med.* 2015;20(5):318-324.
207. Lindahl T, Barnes DE. Repair of endogenous DNA damage. *Cold Spring Harb Sym.* 2000;65:127-133.
208. Jackson AL, Loeb LA. The contribution of endogenous sources of DNA damage to the multiple mutations in cancer. *Mutat Res-Fund Mol M.* 2001;477(1-2):7-21.
209. Tubbs A, Nussenzweig A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell.* 2017;168(4).
210. Xia J, Chiu LY, Nehring RB, et al. Bacteria-to-human protein networks reveal origins of endogenous DNA damage. *Cell.* 2019;176(1-2):127-143.
211. Gates KS. An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals. *Chem Res Toxicol.* 2009;22(11):1747-1760.
212. Lindahl T. Instability and decay of the primary structure of DNA. *Nature.* 1993;362(6422):709-715.
213. Lindahl T, Nyberg B. Rate of depurination of native deoxyribonucleic acid. *Biochemistry.* 1972;11(19):3610–3618.
214. Lindahl T, Karlstrom O. Heat-induced depyrimidination of deoxyribonucleic acid in neutral solution. *Biochemistry.* 1973;12(25):5151-5154.
215. Kowaltowski AJ, de Souza-Pinto NC, Castilho RF, Vercesi AE. Mitochondria and reactive oxygen species. *Free Rad Biol Med.* 2009;47(4):333-343.
216. Nolfi-Donegan D, Braganza A, Shiva S. Mitochondrial electron transport chain: oxidative phosphorylation, oxidant production, and methods of measurement. *Redox Biol.* 2020;37.
217. De Bont R, van Larebeke N. Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis.* 2004;19(3):169-185.
218. Sassa A, Odagiri M. Understanding the sequence and structural context effects in oxidative DNA damage repair. *DNA Repair.* 2020;93.
219. Jovanovic SV, Simic MG. One-electron redox potentials of purines and pyrimidines. *J Phys Chem.* 1986;90(5):974-978.
220. Santivasi WL, Xia F. Ionizing Radiation-Induced DNA Damage, Response, and Repair. *Antioxidants & Redox Signaling.* 2014;21(2):251-259.
221. Bogdanovic O, Lister R. DNA methylation and the preservation of cell identity. *Curr Opin Genet Dev.* 2017;46:9-14.
222. Aguilera A, Garcia-Muse T. Causes of genome instability. *Annu Rev Genet.* 2013;47:1-32.

223. Bebenek A, Ziuzia-Graczyk I. Fidelity of DNA replication - a matter of proofreading. *Curr Genet.* 2018;64(5):985-996.
224. Kunkel TA, Cold Spring Harbor Lab P. Evolving views of DNA replication (in)fidelity. *Cold Spring Harb Quant Biol.* 2009;74:91-101.
225. Kunkel TA, Erie DA. Eukaryotic mismatch repair in relation to DNA replication. *Annu Rev Genet.* 2015;49:291-313.
226. Murat P, Guilbaud G, Sale JE. DNA polymerase stalling at structured DNA constrains the expansion of short tandem repeats. *Genome Biol.* 2020;21(1):1-26.
227. de Paz AM, Cybulski TR, Marblestone AH, et al. High-resolution mapping of DNA polymerase fidelity using nucleotide imbalances and next-generation sequencing. *Nucleic Acids Res.* 2018;46(13).
228. Helmrich A, Ballarino M, Nudler E, Tora L. Transcription-replication encounters, consequences and genomic instability. *Nat Struct Mol Biol.* 2013;20(4):412-418.
229. Hamperl S, Cimprich KA. Conflict resolution in the genome: how transcription and replication make it work. *Cell.* 2016;167(6):1455-1467.
230. Anderson AP, Luo XM, Russell W, Yin YW. Oxidative damage diminishes mitochondrial DNA polymerase replication fidelity. *Nucleic Acids Res.* 2020;48(2):817-829.
231. Minchell NE, Keszthelyi A, Baxter J. Cohesin causes replicative DNA damage by trapping DNA topological stress. *Mol Cell.* 2020;78(4):739-751.
232. Powers KT, Washington MT. Eukaryotic translesion synthesis: choosing the right tool for the job. *DNA Repair.* 2018;71:127-134.
233. Rodgers K, McVey M. Error-prone repair of DNA double-strand breaks. *J Cell Physiol.* 2015;231(1):15-24.
234. Chen J, Furano AV. Breaking bad: The mutagenic effect of DNA repair. *DNA Repair.* 2015;32:43-51.
235. Harper JW, Elledge SJ. The DNA damage response: ten years after. *Mol Cell.* 2007;28(5):739-745.
236. Pegg AE. Multifaceted roles of alkyltransferase and related proteins in DNA repair, DNA damage, resistance to chemotherapy, and research Tools. *Chem Res Toxicol.* 2011;24(5):618-639.
237. Yi CQ, He C. DNA repair by reversal of DNA damage. *Cold Spring Harb Perspect Biol.* 2013;5(1).
238. Beard WA, Horton JK, Prasad R, Wilson SH. Eukaryotic base excision repair: new approaches shine light on mechanism. *Annu Rev Biochem.* 2019;88:137-162.

239. Spivak G. Nucleotide excision repair in humans. *DNA Repair.* 2015;36:13-18.
240. Fuss JO, Tainer JA. XPB and XPD helicases in TFIIH orchestrate DNA duplex opening and damage verification to coordinate repair with transcription and cell cycle via CAK kinase. *DNA Repair.* 2011;10(7):697-713.
241. Ogi T, Limsirichaikul S, Overmeer RM, et al. Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. *Mol Cell.* 2010;37(5):714-727.
242. Scharer OD. Nucleotide excision repair in eukaryotes. *Cold Spring Harb Perspect Biol.* 2013;5(10):1-20.
243. Aden K, Bartsch K, Dahl J, et al. Epithelial RNase H2 maintains genome integrity and prevents intestinal tumorigenesis in mice. *Gastroenterology.* 2019;156(1):145–159.
244. Genschel J, Modrich P. Mechanism of 5′-directed excision in human mismatch repair. *Mol Cell.* 2003;12(5):1077-1086.
245. Kadyrov FA, Dzantiev L, Constantin N, Modrich P. Endonucleolytic function of MutLα in human mismatch repair. *Cell.* 2006;126(2):297-308.
246. Longley MJ, Pierce AJ, Modrich P. DNA polymerase δ is required for human mismatch repair *in vitro*. *J Biol Chem.* 1997;272(16):10917-10921.
247. Zhang YB, Yuan FH, Presnell SR, et al. Reconstitution of 5′-directed human mismatch repair in a purified system. *Cell.* 2005;122(5):693-705.
248. Amunugama R, Walter JC. A new varietal of DNA interstrand crosslink repair. *Cell Res.* 2020;30(6):459-460.
249. Wu RA, Semlow DR, Kamimae-Lanning AN, et al. TRAIP is a master regulator of DNA interstrand crosslink repair. *Nature.* 2019;567(7747):267-+.
250. Wood RD. Mammalian nucleotide excision repair proteins and interstrand crosslink repair. *Environ Mol Mutagen.* 2010;51(6):520-526.
251. Hashimoto S, Anai H, Hanada K. Mechanisms of interstrand DNA crosslink repair and human disorders. *Genes Environ.* 2016;38:1-8.
252. McKinnon PJ, Caldecott KW. DNA strand break repair and human genetic disease. *Annu Rev Genom Hum Genet.* 2007;8:37-55.
253. Scully R, Panday A, Elango R, Willis NA. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat Rev Mol Cell Biol.* 2019;20(11):698-714.
254. Mimitou EP, Symington LS. Ku prevents Exo1 and Sgs1-dependent resection of DNA ends in the absence of a functional MRX complex or Sae2. *Embo J.* 2010;29(19):3358-3369.

255. Mimori T, Hardin JA. Mechanism of interaction between Ku protein and DNA. *J Biol Chem.* 1986;261(22):375-379.
256. Mladenov E, Magin S, Soni A, Iliakis G. DNA double-strand-break repair in higher eukaryotes and its role in genomic instability and cancer: Cell cycle and proliferation-dependent regulation. *Semin Cancer Biol.* 2016;37-38:51-64.
257. Nick McElhinny SA, Snowden CM, McCarville J, Ramsden DA. Ku recruits the XRCC4-ligase IV complex to DNA ends. *Mol Cell Biol.* 2000;20(9):2996-3003.
258. Andres SN, Vergnes A, Ristic D, Wyman C, Modesti M, Junop M. A human XRCC4-XLF complex bridges DNA. *Nucleic Acids Res.* 2012;40(4):1868-1878.
259. Holloman WK. Unraveling the mechanism of BRCA2 in homologous recombination. *Nat Struct Mol Biol.* 2011;18(7):748-754.
260. Byun TS, Pacek M, Yee MC, Walter JC, Cimprich KA. Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. *Genes Dev.* 2005;19(9):1040-1052.
261. Saldivar JC, Cortez D, Cimprich KA. The essential kinase ATR: ensuring faithful duplication of a challenging genome. *Nat Rev Mol Cell Biol.* 2017;18(12):783-783.
262. Sale JE, Lehmann AR, Woodgate R. Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat Rev Mol Cell Biol.* 2012;13(3):141-152.
263. Hedglin M, Benkovic SJ. Eukaryotic translesion DNA synthesis on the leading and lagging strands: unique detours around the same obstacle. *Chem Rev.* 2017.
264. Sale JE. Translesion DNA synthesis and mutagenesis in eukaryotes. *Cold Spring Harb Perspect Biol.* 2013;5(3).
265. Quinet A, Lerner LK, Martins DJ, Menck CFM. Filling gaps in translesion DNA synthesis in human cells. *Mutat Res - Gen Tox En.* 2018;836:127-142.
266. Zeman MK, Cimprich KA. Causes and consequences of replication stress. *Nat Cell Biol.* 2014;16(1):2-9.
267. Branzei D, Szakal B. DNA damage tolerance by recombination: molecular pathways and DNA structures. *DNA Repair.* 2016;44:68-75.
268. Neelsen KJ, Lopes M. Replication fork reversal in eukaryotes: from dead end to dynamic response. *Nat Rev Mol Cell Biol.* 2015;16(4):207-220.
269. Brown JA, Suo ZC. Unlocking the sugar "steric gate" of DNA polymerases. *Biochemistry.* 2011;50(7):1135-1142.
270. Wanrooij S, Fuste JM, Farge G, Shi YH, Gustafsson CM, Falkenberg M. Human mitochondrial RNA polymerase primes

lagging-strand DNA synthesis in vitro. *Proc Natl Acad Sci USA.* 2008;105(32):11122-11127.
271. Burgers PMJ. Polymerase dynamics at the eukaryotic DNA replication fork. *J Biol Chem.* 2009;284(7):4041-4045.
272. Nick McElhinny SA, Watts BE, Kumar D, et al. Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc Natl Acad Sci USA.* 2010;107(11):4949-4954.
273. Kasiviswanathan R, Copeland WC. Ribonucleotide discrimination and reverse transcription by the human mitochondrial DNA polymerase. *J Biol Chem.* 2011;286(36):31490-31500.
274. Makarova AV, McElhinny SAN, Watts BE, Kunkel TA, Burgers PM. Ribonucleotide incorporation by yeast DNA polymerase ζ. *DNA Repair.* 2014;18:63-67.
275. Boule JB, Rougeon F, Papanicolaou C. Terminal deoxynucleotidyl transferase indiscriminately incorporates ribonucleotides and deoxyribonucleotides. *J Biol Chem.* 2001;276(33):31388-31393.
276. Cavanaugh NA, Beard WA, Batra VK, Perera L, Pedersen LG, Wilson SH. Molecular insights into DNA polymerase deterrents for ribonucleotide insertion. *J Biol Chem.* 2011;286(36):31650-31660.
277. Mentegari E, Crespan E, Bavagnoli L, et al. Ribonucleotide incorporation by human DNA polymerase η impacts translesion synthesis and RNase H2 activity. *Nucleic Acids Res.* 2017;45(5):2600-2614.
278. Brown JA, Fiala KA, Fowler JD, et al. A novel mechanism of sugar selection utilized by a human X-family DNA polymerase. *J Mol Biol.* 2010;395(2):282-290.
279. Nick McElhinny SA, Ramsden DA. Polymerase μ is a DNA-directed DNA/RNA polymerase. *Mol Cell Biol.* 2003;23(7):2309-2315.
280. Hogg M, Sauer-Eriksson AE, Johansson E. Promiscuous DNA synthesis by human DNA polymerase θ. *Nucleic Acids Res.* 2012;40(6):2611-2622.
281. Donigan KA, McLenigan MP, Yang W, Goodman MF, Woodgate R. The steric gate of DNA polymerase ι regulates ribonucleotide incorporation and deoxyribonucleotide fidelity. *J Biol Chem.* 2014;289(13):9136-9145.
282. Nevin P, Engen JR, Beuning PJ. Steric gate residues of Y-family DNA polymerases DinB and pol κ are crucial for dNTP-induced conformational change. *DNA Repair.* 2015;29:65-73.

283. Brown JA, Fowler JD, Suo ZC. Kinetic basis of nucleotide selection employed by a protein template-dependent DNA polymerase. *Biochemistry.* 2010;49(26):5504-5510.
284. Pryor JM, Conlin MP, Carvajal-Garcia J, et al. Ribonucleotide incorporation enables repair of chromosome breaks by nonhomologous end joining. *Science.* 2018;361(6407):1126-1129.
285. Martin MJ, Garcia-Ortiz MV, Esteban V, Blanco L. Ribonucleotides and manganese ions improve non-homologous end joining by human Polμ. *Nucleic Acids Res.* 2013;41(4):2428-2436.
286. Liu S, Hua Y, Wang J, et al. RNA polymerase III is required for the repair of DNA double-strand breaks by homologous recombination. 2021;184:1-16.
287. Pai CC, Kearsey SE. A critical balance: dNTPs and the maintenance of genome stability. *Genes.* 2017;8(2):1-14.
288. Ghodgaonkar MM, Lazzaro F, Olivera-Pimentel M, et al. Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair. *Mol Cell.* 2013;50(3):323-332.
289. Lujan SA, Williams JS, Clausen AR, Clark AB, Kunkel TA. Ribonucleotides are signals for mismatch repair of leading-strand replication errors. *Mol Cell.* 2013;50(3):437-443.
290. Vengrova S, Dalgaard JZ. RNase-sensitive DNA modification(s) initiates *S. pombe* mating-type switching. *Gene Dev.* 2004;18(7):794-804.
291. Yamada-Inagawa T, Klar AJS, Dalgaard JZ. *Schizosaccharomyces pombe* switches mating type by the synthesis-dependent strand-annealing mechanism. *Genetics.* 2007;177(1):255-265.
292. Eder PS, Walder JA. Ribonuclease H from K563 human erythroleukamia cells - purification, characterization, and substrate specificity. *J Biol Chem.* 1991;266(10):6472-6479.
293. Eder PS, Walder RY, Walder JA. Substrate specificity of human RNase H1 and its role in excision repair of ribose residues misincorporated in DNA. *Biochimie.* 1993;75(1-2):123-126.
294. Sparks JL, Chon H, Cerritelli SM, et al. RNase H2-Initiated Ribonucleotide Excision Repair. *Mol Cell.* 2012;47(6):980-986.
295. Rydberg B, Game J. Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts. *Proc Natl Acad Sci USA.* 2002;99(26):16654-16659.
296. Riva V, Garbelli A, Casiraghi F, et al. Novel alternative ribonucleotide excision repair pathways in human cells by

DDX3X and specialized DNA polymerases. *Nucleic Acids Res.* 2020;48(20):11551-11565.
297. Cerritelli SM, Crouch RJ. Ribonuclease H: the enzymes in eukaryotes. *Febs J.* 2009;276(6):1494-1505.
298. Huang S-yN, Williams JS, Arana ME, Kunkel TA, Pommier Y. Topoisomerase I-mediated cleavage at unrepaired ribonucleotides generates DNA double-strand breaks. *Embo J.* 2017;36:361-373.
299. Kim N, Huang SYN, Williams JS, et al. Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. *Science.* 2011;332(6037):1561-1564.
300. Sekiguchi J, Shuman S. Site-specific ribonuclease activity of eukaryotic DNA topoisomerase I. *Mol Cell.* 1997;1(1):89-97.
301. Williams JS, Smith DJ, Marjavaara L, Lujan SA, Chabes A, Kunkel TA. Topoisomerase 1-mediated removal of ribonucleotides from nascent leading-strand DNA. *Mol Cell.* 2013;49(5):1010-1015.
302. Williams JS, Clausen AR, Lujan SA, et al. Evidence that processing of ribonucleotides in DNA by topoisomerase 1 is leading-strand specific. *Nat Struct Mol Biol.* 2015;22(4):291-U235.
303. Cedergren R, Lang BF, Gravel D. A mechanism for the RNA-catalyzed formation of 5´-phosphates - the origin of nucleases. *Febs Lett.* 1987;226(1):63-66.
304. Sparks JL, Burgers PM. Error-free and mutagenic processing of topoisomerase 1-provoked damage at genomic ribonucleotides. *Embo J.* 2015;34(9):1259-1269.
305. Rossi ML, Bambara RA. Reconstituted Okazaki fragment processing indicates two pathways of primer removal. *J Biol Chem.* 2006;281(36):26051-26061.
306. Stith CM, Sterling J, Resnick MA, Gordenin DA, Burgers PM. Flexibility of eukaryotic Okazaki Fragment maturation through regulated strand displacement synthesis. *J Biol Chem.* 2008;283(49):34129-34140.
307. Williams JS, Tumbale PP, Arana ME, Rana JA, Williams RS, Kunkel TA. High-fidelity DNA ligation enforces accurate Okazaki fragment maturation during DNA replication. *Nat Commun.* 2021;12(1):1-12.
308. Zimmer AD, Koshland D. Differential roles of the RNases H in preventing chromosome instability. *Proc Natl Acad Sci USA.* 2016;113(43):12220-12225.
309. Wanrooij PH, Engqvist MKM, Forslund JME, et al. Ribonucleotides incorporated by the yeast mitochondrial DNA polymerase are not repaired. *P Natl Acad Sci USA.* 2017;114(47):12466-12471.

310. Wanrooij PH, Chabes A. Ribonucleotides in mitochondrial DNA. *Febs Lett.* 2019;593(13):1554-1565.
311. He Q, Shumate CK, White MA, Molineux IJ, Yin YW. Exonuclease of human DNA polymerase γ disengages its strand displacement function. *Mitochondrion.* 2013;13(6):592-601.
312. Macao B, Uhler JP, Siibak T, et al. The exonuclease activity of DNA polymerase γ is required for ligation during mitochondrial DNA replication. *Nat Commun.* 2015;6:1-10.
313. Holmes JB, Akman G, Wood SR, et al. Primer retention owing to the absence of RNase H1 is catastrophic for mitochondrial DNA replication. *P Natl Acad Sci USA.* 2015;112(30):9334-9339.
314. Cerritelli SM, Frolova EG, Feng CG, Grinberg A, Love PE, Crouch RJ. Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. *Mol Cell.* 2003;11(3):807-815.
315. Berglund AK, Navarrete C, Engqvist MKM, et al. Nucleotide pools dictate the identity and frequency of ribonucleotide incorporation in mitochondrial DNA. *Plos Genet.* 2017;13(2).
316. Rice G, Patrick T, Parmar R, et al. Clinical and molecular phenotype of Aicardi-Goutières syndrome. *Am J Hum Genet.* 2007;81(4):713-725.
317. Rice GI, Reijns MAM, Coffin SR, et al. Synonymous mutations in RNASEH2A create cryptic splice sites impairing RNase H2 enzyme function in Aicardi-Goutières syndrome. *Hum Mut.* 2013;34(8):1066-1070.
318. Crow YJ, Leitch A, Hayward BE, et al. Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutières syndrome and mimic congenital viral brain infection. *Nat Genet.* 2006;38(8):910-916.
319. Gunther C, Kind B, Reijns MAM, et al. Defective removal of ribonucleotides from DNA promotes systemic autoimmunity. *J Clin Invest.* 2015;125(1):413-424.
320. Kellner V, Luke B. Molecular and physiological consequences of faulty eukaryotic ribonucleotide excision repair. *Embo J.* 2020;39(3).
321. Crow YJ, Chase DS, Schmidt JL, et al. Characterization of human disease phenotypes associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1. *Am J Med Genet A.* 2015;167(2):296-312.
322. Uehara R, Cerritelli SM, Hasin N, et al. Two RNase H2 mutants with differential rNMP processing activity reveal a threshold of ribonucleotide tolerance for embryonic development. *Cell Rep.* 2018;25(5):1135-1145.

323. Klein U, Lia M, Crespo M, et al. The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. *Cancer Cell.* 2010;17(1):28-40.
324. Abeshouse A, Ahn J, Akbani R, et al. The molecular taxonomy of primary prostate cancer. *Cell.* 2015;163(4):1011-1025.
325. Zimmermann M, Murina O, Reijns MAM, et al. CRISPR screens identify genomic ribonucleotides as a source of PARP-trapping lesions. *Nature.* 2018;559(7713):285–289.
326. Tumbale P, Williams JS, Schellenberg MJ, Kunkel TA, Williams RS. Aprataxin resolves adenylated RNA-DNA junctions to maintain genome integrity. *Nature.* 2014;506(7486):111–115.
327. Reyes A, Melchionda L, Nasca A, et al. RNASEH1 mutations impair mtDNA replication and cause adult-onset mitochondrial encephalomyopathy. *Am J Hum Genet.* 2015;97(1):186-193.
328. Livneh Z, Ziv O, Shachar S. Multiple two-polymerase mechanisms in mammalian translesion DNA synthesis. *Cell Cycle.* 2010;9(4):729-735.
329. Boehm EM, Powers KT, Kondratick CM, Spies M, Houtman JCD, Washington MT. The proliferating cell nuclear antigen (PCNA)-interacting protein (PIP) motif of DNA polymerase η mediates its interaction with the C-terminal domain of Rev1. *J Biol Chem.* 2016;291(16):8735-8744.
330. Haracska L, Kondratick CM, Unk I, Prakash S, Prakash L. Interaction with PCNA is essential for yeast DNA polymerase η function. *Mol Cell.* 2001;8(2):407-415.
331. Donigan KA, Cerritelli SM, McDonald JP, Vaisman A, Crouch RJ, Woodgate R. Unlocking the steric gate of DNA polymerase η leads to increased genomic instability in *Saccharomyces cerevisiae*. *DNA Repair.* 2015;35:1-12.
332. Supek F, Lehner B. Clustered mutation signatures reveal that error-prone DNA repair targets mutations to active genes. *Cell.* 2017;170(3):534-547.
333. Clausen AR, Williams JS, Kunkel TA. Measuring ribonucleotide incorporation into DNA i*n vitro* and i*n vivo*. In: Vengrova S, Dalgaard J, eds. *DNA Replication. Methods in Molecular Biology.* Vol 1300. New York, NY: Humana Press; 2013:123-139.
334. Koh KD, Balachander S, Hesselberth JR, Storici F. Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA. *Nat Methods.* 2015;12(3):251.
335. Keszthelyi A, Daigaku Y, Ptasinska K, Miyabe I, Carr AM. Mapping ribonucleotides in genomic DNA and exploring replication dynamics by polymerase usage sequencing (Pu-seq). *Nat Protoc.* 2015;10(11):1786-1801.

336. Ding J, Taylor MS, Jackson AP, Reijns MAM. Genome-wide mapping of embedded ribonucleotides and other noncanonical nucleotides using emRiboSeq and EndoSeq. *Nat Protoc.* 2015;10(9):1433-1444.
337. Forslund JME, Pfeiffer A, Stojkovic G, Wanrooij PH, Wanrooij S. The presence of rNTPs decreases the speed of mitochondrial DNA replication. *Plos Genet.* 2018;14(3).
338. Moss CF, Dalla Rosa I, Hunt LE, et al. Aberrant ribonucleotide incorporation and multiple deletions in mitochondrial DNA of the murine MPV17 disease model. *Nucleic Acids Res.* 2017;45(22):12808–12815.
339. Kong ZQ, Jia SD, Chabes AL, et al. Simultaneous determination of ribonucleoside and deoxyribonucleoside triphosphates in biological samples by hydrophilic interaction liquid chromatography coupled with tandem mass spectrometry. *Nucleic Acids Res.* 2018;46(11).
340. Hansen EC, Seamon KJ, Cravens SL, Stivers JT. GTP activator and dNTP substrates of HIV-1 restriction factor SAMHD1 generate a long-lived activated state. *Proc Natl Acad Sci USA.* 2014;111(18):E1843-E1851.
341. Li YF, Breaker RR. Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2 '-hydroxyl group. *J Am Chem Soc.* 1999;121(23):5364-5372.
342. Reijns MAM, Rabe B, Rigby RE, et al. Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell.* 2012;149(5):1008-1022.
343. Wanrooij PH, Tran P, Thompson LJ, et al. Elimination of rNMPs from mitochondrial DNA has no effect on its stability. *P Natl Acad Sci USA.* 2020;117(25).
344. Fumagalli M, Sironi M. Human genome variability, natural selection and infectious diseases. *Curr Opin Immunol.* 2014;30:9-16.
345. Sharma P, Sampath H. Mitochondrial DNA integrity: role in health and disease. *Cells.* 2019;8(2):21.
346. Pikor L, Thu K, Vucic E, Lam W. The detection and implication of genome instability in cancer. *Cancer Metas Rev.* 2013;32(3-4):341-352.
347. Yasukawa T, Kang D. An overview of mammalian mitochondrial DNA replication mechanisms. *J Biochem.* 2018;164(3):183-193.
348. Lee SR, Han J. Mitochondrial nucleoid: shield and switch of the mitochondrial genome. *Oxid Med Cell Longev.* 2017;2017.
349. Sun S, Osterman MD, Li M. Tissue specificity of DNA damage response and tumorigenesis. *Cancer Biol Med.* 2019;16(3):396-414.

350. Lacroix L, Lienard H, Labourier E, et al. Identification of two human nuclear proteins that recognise the cytosine-rich strand of human telomeres *in vitro*. *Nucleic Acids Res.* 2000;28(7):1564-1575.
351. Abou Assi H, Garavis M, Gonzalez C, Damha MJ. i-motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.* 2018;46(16):8038-8056.

7 APPENDIX

Paper I

Kreisel, K, Engqvist, MKM, Kalm, J, Thompson, LJ, Boström, M, Navarrete, C, McDonald, JP, Larsson, E, Woodgate, R, Clausen, AR. DNA polymerase η contributes to genome-wide lagging strand synthesis. *Nucleic Acids Research*, 2019; 47(5): 2425-2435

and supplemental information.

Paper II

Kreisel, K, Engqvist, MKM, Clausen, AR. Simultaneous mapping and quantitation of ribonucleotides in human mitochondrial DNA. *Journal of Visualized Experiments* 2017; 129: e56551

and supplemental information.

Paper III

Wanrooij, PH, Tran, P, Thompson, LJ, Carvalho, G, Sharma, S, **Kreisel, K**, Navarrete, C, Feldberg, A, Watt, DL, Nilsson AK, Engqvist, MKM, Clausen, AR, Chabes, A. Elimination of rNMPs from mitochondrial DNA has no effect on its stability. *Proceedings of the National Academy of Sciences of the United States of America* 2020; 117(25): 14306-14313

and supplemental information.

Paper IV

Kreisel, K, Kalm, J, Bandaru, S, Ala, C, Akyürek, L, Clausen, AR. Stably incorporated ribonucleotides in murine tissues: quantitation, base identity and distribution in nuclear and mitochondrial DNA. *(to be submitted)*

and supplemental information.