Chapter 1

General introduction

Cover Photo: Drop by Drop ▶

500,000,000,000,000,000,000,000 NUCLEOTIDES AT RISK

DNA encodes the genetic instructions for life. It is for every organism of vital importance to safeguard its genetic information and transmit this faithfully to its progeny. Unfaithful replication of the genome and erroneous repair of damaged DNA lead to mutations in this genetic information. These mutations can be seen as a double-edged sword: on the one hand mutations are the driving force behind evolution and result in genetic diversity, which is beneficial for maintenance of the species, but on the other hand mutations can lead to reduced fitness of an individual organism, for instance due to mutation-induced uncontrolled growth of cells (cancer).

Keeping in mind that a haploid human genome consists of approximately $3.3x10^9$ base pairs (and thus $1.3x10^{10}$ nucleotides for a diploid genome) and that a human body is estimated to consist of ~$3.7x10^{13}$ cells (Bianconi et al., 2013), this means that an average person contains at least ~$5x10^{23}$ nucleotides providing inheritable genetic information. It is overwhelming to realize that all these nucleotides are copied from a single original template (the genome of a fertilized oocyte) with an extremely low mutation-rate. A recent study, in which the sequenced genomes of 78 Icelandic parent-offspring trios was used, estimated that the *de novo* mutation rate is $1.2x10^{-8}$ per nucleotide per generation (Kong et al., 2012). This low mutation rate is even more mind-boggling when realizing that replicative polymerases encounter many obstacles during replication and nucleotides are under a constant attack of endogenous and exogenous DNA damaging sources. For instance, it has been anticipated that human cells may experience up to 10^5 spontaneous DNA lesions per cell per day (Hoeijmakers, 2009; Lindahl, 1993). How all these nucleotides are protected against mutations has been under investigation for decades. Particularly the fact that every cancer is the consequence of one or more mutations in a genome has led to extensive research in the fields of genome stability and cancer genomics. Nevertheless, despite an officially declared war on cancer (National Cancer Act 1971), and an army of scientists, many questions in these fields of research remain to be answered, since cancer is still responsible for one in eight deaths worldwide to date (ACS, 2013).

In the first part of this introduction I will focus on aspects related to genome stability; the sources of genome instability and the currently known pathways that protect the genome against instability. Special emphasis will be on DNA polymerases and helicases, which will be of relevance in particular for chapters 4 and 5 of this thesis. The second part of this introduction will provide an up-to-date overview about the current knowledge of the two main themes in this thesis: microsatellites (short tandem repeats with units of 1-8 base pairs long) and G-quadruplexes (stable secondary structures with guanines as main building blocks). Finally, since this thesis describes various model systems to study genome instability, also a short overview will be provided discussing the main advantages and limitations of each system.

ENDOGENOUS AND EXOGENOUS HAZARDS TO DNA

Genomic integrity is threatened by various types of DNA damage and processes that can lead to unwanted changes in the DNA and loss of genetic information. Below, an overview is provided of the most prominent threats to DNA. A distinction is made between exogenous and endogenous sources that can ultimately lead to genome instability. A schematic overview is provided in Figure 1.

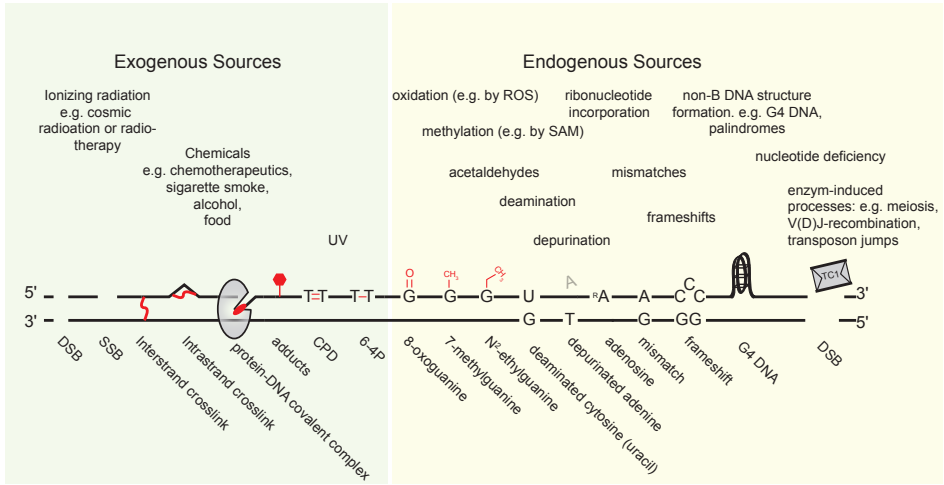

Figure 1 | Exogenous and endogenous sources that can lead to genomic instability. An overview of prominent exogenous and endogenous sources/processes that can lead to DNA modifications and ultimately to genomic instability. Examples of common types of DNA damages, replication errors and other replication blocking structures are shown that all need to be accurately resolved, repaired, or bypassed to prevent genomic instability. DSB, double-strand break; SSB, single-strand break; CPD, cyclobutane pyrimidine dimer; 6-4P, 6-4 photoproduct; ROS, reactive oxygen species; SAM, S-adenosylmethionine.

Exogenous Sources

Exogenous DNA damage is caused by physical and chemical sources from outside of the cell. Two well-known sources of physical DNA damage are ionizing radiation (IR) and ultraviolet (UV) light from sunlight. IR (for example from cosmic radiation or radiotherapy) can lead to oxidation of nucleotides and generate single-strand breaks (SSB) or, even more toxic, double-strand breaks (DSB) (Ciccia and Elledge, 2010). The most prominent lesions induced by UV light are pyrimidine dimers and 6-4 photoproducts. It has been estimated that cells exposed to the sun's UV light can suffer from up to 10^5 lesions per cell per day (Hoeijmakers, 2009).

An example of a chemical source that causes DNA damage is cigarette smoke. Estimates vary between 45 and 1,000 bulky aromatic DNA adducts per cell in tissues that are exposed to cigarette smoke (Ciccia and Elledge, 2010; Lindahl and Barnes, 2000). Also foods can contain DNA-damaging chemicals, such as aflatoxins in contaminated peanuts and heterocyclic amines in burnt meat (Wogan et al., 2004). Other chemical sources that can inflict severe DNA damage are chemical agents that are used in cancer chemotherapy. Commonly used therapies consist of toxic crosslinking agents such as mitomycin C (MMC) and cisplatin, which generate intrastrand and interstrand crosslinks (covalent links between nucleotides of the same or a different DNA strand, respectively). Other agents such as camptothecin (CPT) and etoposide trap topoisomerases in a covalent complex with DNA, which results in high numbers of SSBs and DSBs.

Endogenous sources

Besides threats from outside the cell, also endogenous reactive molecules and processes that take place during cellular and DNA metabolism endanger the genetic code. Since DNA is a chemically reactive molecule, it is exposed to processes like hydrolysis, oxidation and nonenzymatic methylation(Lindahl, 1993). A major threat of DNA hydrolysis is the generation of predominantly apurinic sites (~2,000 - 10,000 lesions/cell/day (Lindahl and Nyberg, 1972)) and deaminated cytosines (~100 - 500 lesions/cell/day (Lindahl and Barnes, 2000)). Oxidation contributes to the generation of oxidized base lesions, such as 8-oxoguanines. During normal cellular metabolism many reactive oxygen species (ROS) are generated, which, in turn, lead to oxidation of DNA bases. Non-enzymatic methylation of bases is mainly caused by the molecule *S*-adenosylmethionine (SAM). SAM is used as cofactor in many cellular transmethylation reactions and SAM-induced methylation of bases is estimated to occur at rates of ~600 and ~4000 lesions/cell/day (for 3-methyladenine and 7-methylguanine, respectively (Lindahl and Barnes, 2000)). Other reactive molecules are aldehydes, which are common byproducts formed during cellular metabolism (e.g. lipid peroxidation (O'Brien et al., 2005)) and histone demethylation (Rosado et al., 2011).

Also during DNA replication there are several processes that can lead to unwanted changes in the genetic code. Formation of secondary structures (e.g. palindromes and G-quadruplexes), mismatches and frameshifts are all associated with replication and require repair to prevent loss of genetic information. Until recently, an underrated class of DNA damage is the incorporation of ribonucleotides during DNA replication. A recent study in *Saccharomyces cerevisiae* describes that the leading-strand polymerase ε (pol-ε) incorporates one ribonucleotide monophosphate (rNMP) per 1,250 deoxyribonucleotide monophosphates (dNMP), and the lagging-strand polymerase δ (pol-δ) one rNMP per 5,000dNMPs (McElhinny et al., 2010; Nick McElhinny et al., 2010). Although it is thought that "mis"-incorporation of these rNMPs may have a biological function (as will be discussed later), these rNMPs need to be removed to prevent base-substitutions and fork collapse.

Finally, another prominent class of endogenous DNA damage is the formation of DSBs during active processes such as V(D)J-recombination, meiosis, transposon

jumping and uncoiling of DNA by topoisomerase II. Altogether, it has been estimated that a dividing human cell suffers from ~10 DSBs per day (Lieber, 2010).

THE DNA-DAMAGE RESPONSE

To counteract the threats posed by DNA damage, organisms have evolved an elaborate genome maintenance apparatus to sense DNA lesions, signal their presence and promote their repair, collectively often termed the DNA-Damage Response (DDR) (Jackson and Bartek, 2009). The DDR is a signal transduction pathway that consists of a well-orchestrated interplay between a plethora of enzymes which determine the cell's fate: survival, replicative senescence or death (Hoeijmakers, 2009). Below, a concise overview will be provided of the main signaling routes of the DDR and the specific repair-pathways that are recruited to resolve the damage.

DNA-damage signaling

Two key players in the DDR are the protein kinases ATM (ataxia-telangiectasia mutated) and ATR (ATM and Rad3-related) (Cimprich and Cortez, 2008; Shiloh, 2003). An important function of ATM together with its regulator the MRN-complex (Mre11, Rad50 and NBS1) is sensing the presence of DSBs, while ATR together with ATRIP (ATR Interacting Protein) senses replication protein A (RPA)-coated single-strand DNA (ssDNA) generated by resected DBSs or stalled replication forks (Matsuoka et al., 2007). Both kinases then phosphorylate proteins to set a signaling cascade in motion that includes the checkpoint kinases Chk1 and Chk2, which in turn activate a second wave of phosphorylation events. This whole cascade of signaling events is believed to be important for at least two particular reasons: first, it results in reduced cyclin-dependent kinases (CDK) activity which slows down or halts cell-cycle progression and allows more time for repair before going into the next phase of the cell-cycle. Second, ATM/ATR signaling enhances repair by recruiting repair proteins to the damage and activating DNA repair proteins by post-translational modifications (Jackson and Bartek, 2009). Besides ATM and ATR also PARP1 and PARP2 are important players in the DDR. Both poly(ADP-ribose) polymerases are believed to be one of the earliest responders in the DDR: within seconds they catalyze the addition of poly (ADP-ribose) chains on proteins at SSBs and DSBs and thereby recruit DDR factors to the chromatin at breaks (reviewed in (Pines et al., 2013)).

When the damage is repaired effectively, the DDR is inactivated and the cell can progress with its normal function. However, if the damage cannot be repaired, chronic DDR signaling can lead to genomic instability, cellular senescence or apoptosis. An important player in this process is the tumor suppressor TP53. A recent study in which 3,281 tumors across 12 tumor types were genome-wide sequenced, illustrated the importance of this transcription factor once again: in more than 40% of the tumors, a mutation was found in TP53 (Kandoth et al., 2013).

Base excision repair

Base excision repair (BER) is an important DNA repair pathway that is responsible for the removal of non-helix-distorting base lesions, such as oxidized, alkylated and deaminated bases. BER can be subdivided in two pathways: short-patch BER and long-patch BER. During short-patch BER only a single nucleotide is repaired, while during long-patch BER a repair tract of approximately two to eleven nucleotides is produced (Pascucci et al., 1999). The basis of BER can be summarized in four basic steps: first, the recognition and removal of the damage by a DNA glycosylase. Second, the cleavage of the DNA backbone by a DNA AP endonuclease or AP lyase resulting in a single nucleotide gap in the DNA. Next, this gap is filled by a DNA polymerase, and finally the gap is sealed by a DNA ligase (Robertson et al., 2009). This pathway was discovered in *Escherichia coli* nearly 40 years ago (Lindahl, 1974), but it became quickly apparent that this pathway was conserved among other species. In human BER, several glycosylases are involved such as 8-Oxoguanine DNA glycosylase (OGG1) and Uracil-DNA glycosylase (UNG). APEX1, APEX2 and Flap structure-specific endonuclease 1 (FEN1) function as endonucleases, and predominantly DNA polymerase beta (POLβ) and Ligase 3 (LIG3) act as required polymerase and ligase, respectively. In addition, also PARP1 and PARP2 are involved in BER and act as sensors and signal transducers for lesions. For an elaborate review about BER see reference (Robertson et al., 2009).

Nucleotide excision repair (NER)

Whereas BER is active at small base adducts, nucleotide excision repair (NER) targets the more bulky lesions that distort the structure of the DNA helix (Cleaver et al., 2009). Lesions that are repaired by NER are for instance pyrimidine dimers and 6-4 photoproducts induced by UV-light. NER is often subclassified into two branches: transcription-coupled NER (TC-NER) and global-genome NER (GG-NER). The main difference between these two classes is the way the lesion is detected, while subsequent repair is executed via a similar mechanism. In TC-NER, transcription-blocking lesions result in the stalling of RNA polymerase II (RNAPII) and the subsequent recruitment of Cockayne Syndrome protein A and B (CSA and CSB, respectively). In GG-NER, helix-distorting lesions are recognized by protein complexes that are encoded by the genes from Xeroderma Pigmentosum complementation group C and E (genes XPC and XPE, respectively). Upon lesion detection, the DNA is opened via the multifunctional protein complex TFIIH (transcription factor II human). Next, re-annealing of the DNA is prevented by RPA and XPA, followed by incision of the DNA on both sides of the lesion by endonucleases ERCC1-XPF and XPG. This results in the excision of the damage as part of a 22-30 base pair (bp) oligo. The resulting gap is filled predominantly by polymerases δ and ε, and ligation is performed by ligase I and III.

Mismatch repair (MMR)

The mismatch repair machinery deals primarily with misincorporated nucleotides and insertion and deletions loops (IDLs) which are formed during DNA replication.

MMR improves the fidelity of DNA replication several orders of magnitude and its importance is illustrated by patients that suffer from defective MMR (known as the Lynch syndrome or hereditary nonpolyposis colon cancer (HNPCC)): patients develop colon cancer at an early age (with an average onset of 45 years of age), but also other tissues are predisposed to tumor formation (e.g. 40-60% of the female carriers will develop endometrial cancer) (Lynch et al., 2009).

MMR's main task is to remove sections of nascent strands containing mispaired nucleotides or IDLs. Similar to other repair pathways MMR can be divided in distinct stages: recognition of the lesion, removal of the lesion and finally filling and ligation of the gap. Key players in MMR for detecting mismatches and IDLs are the MutSα and the Mutsβ complexes. The MutSα-complex is a heterodimeric complex consisting of MSH2 and MSH6 that recognizes base-base mismatches and IDLs of 1-3 nucleotides. The MutSβ-complex comprises MSH2 and MSH3 and is involved in recognizing IDLs of approximately 1-12 nucleotides (Peña-Diaz and Jiricny, 2012). After recognition of the lesion both MutS-complexes undergo an ATP-dependent conformational switch, which converts them into a sliding clamp on the DNA. Subsequently and in cooperation with PCNA and RFC1, the MutLα heterodimer (consisting of the nucleases MLH1 and PMS2) is recruited and activated. Due to the endonuclease activity of PMS2 nicks are introduced around the lesion, which serve as entry site for the 5′ to 3′ exonuclease EXO1. EXO1 activity results in excision of the mismatched nucleotide and a single-stranded gap of approximately 150 bps. This gap is filled by high-fidelity polymerases δ and ε. The sequential steps of canonical MMR are illustrated in Figure 2.

Despite the discovery of the first mismatch repair gene already more than 50 years ago (Siegel and Bryson, 1963), ample questions remain in the field of mismatch repair. For example, how does the MMR machinery know which base in a DNA mispair is the incorrect one, in other words, how is the nascent strand recognized? It is known that in prokaryotes like *Escherichia coli* (*E. coli*) the nascent strand is recognized by the presence of unmethylated adenines, however strand recognition by methylation is not used by eukaryotes. Until recently, many researchers favored the hypothesis that the MMR in eukaryotes was directed to the nascent strand by the presence of strand discontinuities such as gaps between Okazaki fragments. However, recent studies (Ghodgaonkar et al., 2013; Lujan et al., 2013) provide strong evidence that the incorporation of ribonucleotides into the nascent strand during DNA synthesis guides the MMR machinery towards the nascent strand: removal of ribonucleotides by RNAse H2-dependent ribonucleotide excision repair (RER) creates an initiation site for MMR, and in this way the MMR machinery is directed to the error-containing nascent strand.

Another long-standing puzzle in the field of MMR is the identification of several colorectal and other cancers that are characterized by high microsatellite instability (MSI), a hallmark for defective MMR, but for which no genetic or epigenetic defect is detected in the known players of MMR. This raises the question whether there

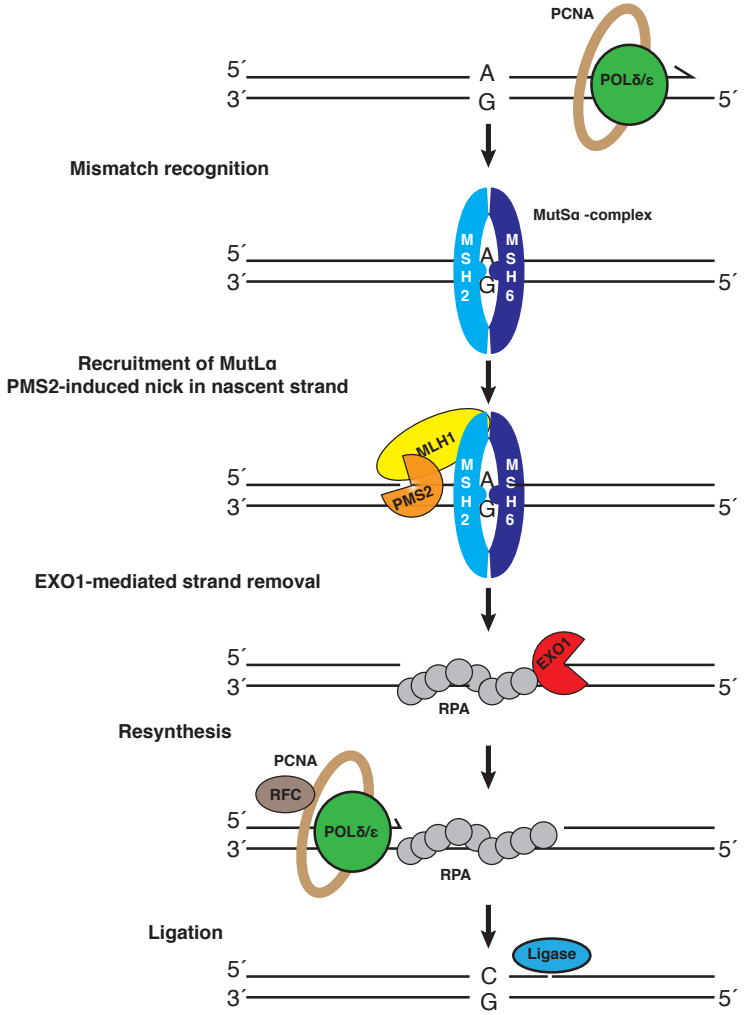

Figure 2 | Sequential steps of MMR in eukaryotes. See text for further details.

are more proteins involved in MMR than currently known. Indeed, a recent study by Li et al (Li et al., 2013) reports the involvement of such a novel player, namely the histone methyltransferase SETD2: cells lacking functional SETD2 display MSI and an elevated mutation frequency. The authors provide evidence that a SETD2-dependent epigenetic histone mark, H3K36me3, is required to recruit the Mutsα-complex to the chromatin.

These very recent discoveries showing important roles of chromatin organization and ribonucleotides incorporation/removal in MMR, underline that still a lot is incompletely understood in the field of MMR.

DNA damage tolerance pathway (DDT)

Despite the presence of numerous repair pathways, DNA lesions might escape repair. When the replication machinery encounters a blocking lesion, the replication fork is stalled which may ultimately result in replication fork collapse, genomic instability and toxicity. To avoid such a scenario, cells can trigger the DNA damage tolerance pathway leading to bypass of the lesion, thus tolerating the lesion to be unrepaired. In this way cells can proceed to complete DNA replication, which is of importance for cellular survival, while the lesion can be repaired at a later time point.

There are two major pathways implicated in DDT: translesion synthesis (TLS) and template switching (TS, also often termed damage avoidance (DA)). In TLS, specialized polymerases are recruited to the fork to replace the stalled replicative polymerase. In contrast to the normal replicative DNA polymerases δ and ε, TLS polymerases have a more open structure, and therefore they can directly bypass damaged bases or bulky adducts. This bypass, however, often comes with the cost of mutation induction, since TLS polymerases are notorious for operating in an error-prone fashion. Therefore TLS is considered an error-prone pathway. By contrast, template switching/damage avoidance is an error-free process: during template switching the undamaged sister chromatid is temporarily used.

The choice between using either TLS or template switching at a lesion is still not fully understood. It is clear though that the posttranslational modification of the homotrimer DNA sliding clamp PCNA plays an important role. It is generally believed that mono-ubiquitination of PCNA promotes TLS whereas polyubiquitination stimulates TS/DA. For more details and a comprehensive overview about DDT see (Ghosal and Chen, 2013).

Fanconi anemia (FA) pathway and the repair of DNA interstrand crosslinks

A form of damage that can never be bypassed is a DNA interstrand crosslink (ICL), a covalent chemical bond between two nucleotides of opposing DNA strands. ICL's are very toxic to cells since they prevent strand separation and therefore hamper essential biological processes such as DNA replication and transcription. The toxicity but also the therapeutically potential of ICL's became apparent when a ship, loaded with ± 60.000 tonnes of the ICL-agent nitrogen mustard, was bombed in the harbor of Bari during the Second World War (Deans and West, 2011). Many soldiers and civilians were exposed to the nitrogen mustard and autopsies on fatal casualties showed that the chemical specifically attacked the victims' white blood cells. This discovery led to the view that these kind of chemicals could be used as a potential treatment for patients that suffered from leukemia. More than 70 years later, ICL-agents such as cisplatin and mitomycin C (MMC) are still widely used in the clinic to treat leukemia and other types of cancer. Also in research laboratories these compounds are commonly used to investigate the molecular mechanisms that play a role in the repair of ICLs.

Many years of research has led to great insight into which genes and pathways are involved in ICL repair. Key in the dissection of the molecular mechanisms behind ICL repair was the identification of the underlying mutations in Fanconi anemia (FA) patients, as these patients showed severe sensitivity to ICL-agents. To date, 16 Fanconi anemia complementation groups (FANCA-FANCQ) have been connected to distinct genes and there are still patients in whom a mutation has yet to be identified (see for a complete list of identified genes and their functions references (Garaycoechea and Patel, 2014; Kottemann and Smogorzewska, 2013). Interestingly, several of these 16 genes were already known to play a role in other DNA repair pathways (e.g. FANCQ/XPF in NER, FANCO/RAD51C in homologous recombination). Studies in Xenopus extracts (Klein Douwel et al., 2014; Knipscheer et al., 2009) have lead to a model in which the interstrand crosslinks are removed by an elaborate interplay between FA proteins, excision repair, translesion synthesis and homologous recombination (HR) (see Figure 3) (Garaycoechea and Patel, 2014); i) Upon replication, two converging forks stall at the ICL. ii) Recruitment of the so-called "FA core complex" takes place and ensures iii) monoubiquitination of FANCD2 and FANCI. iv) Next, unhooking of the ICL is accomplished by the incision in one strand on both sides by the nucleases FANCQ/XPF-ERCC1 and FANCP/SLX4, which results in a broken chromatid (bottom strand in Figure 3c) and an intact chromatid containing the crosslink (top strand in Figure 3c). v) Translesion synthesis bypasses the lesion, and it is generally thought that NER removes the lesion in the top strand, and HR (involving FANCD1/BRCA2, FANCO/RAD51C, FANCN/PALB2 and FANCJ/BRIP1) is used to repair the broken chromatid, resulting in two complete repaired DNA duplexes.

Although it is nowadays clear why FA-defective patients are hypersensitive to exogenous administered ICL-agents, it is less well understood which endogenous source(s) are the cause of the clinical manifestations seen in FA-patients. FA patients are characterized by bone marrow failure, congenital abnormailities, infertility and a high risk to develop cancers. Remarkably, FA is phenotypically heterogeneous; some patients develop bone marrow failure at the onset of 3 years, whereas other patients with the exact same mutation may never suffer from bone marrow failure. Interestingly, recent studies by the group of Patel and Hira indicate that aldehydes may be an endogenous source that can result in genomic instability in the absence of a functional FA pathway ((Garaycoechea et al., 2012) and references therein). Aldehydes (like formaldehyde and acetaldehyde) can be formed during cellular metabolism (e.g. during DNA and histone methylation) and are able to form DNA adducts. Intriguingly, both mice and humans who are deficient for the FA-pathway and also in the breakdown of acetaldehyde due to a mutation in aldehyde dehydrogenase 2 (*Aldh2*), show dramatically increased manifestations of FA-related clinical features such as bone marrow failure (Hira et al., 2013; Langevin et al., 2011). This finding strongly suggest that adducts of aldehydes to the DNA are natural substrates for the FA-pathway. Whether adducts formed by these aldehydes are removed from the DNA in a similar way as ICLs is yet unknown. Furthermore, it will be interesting to find out

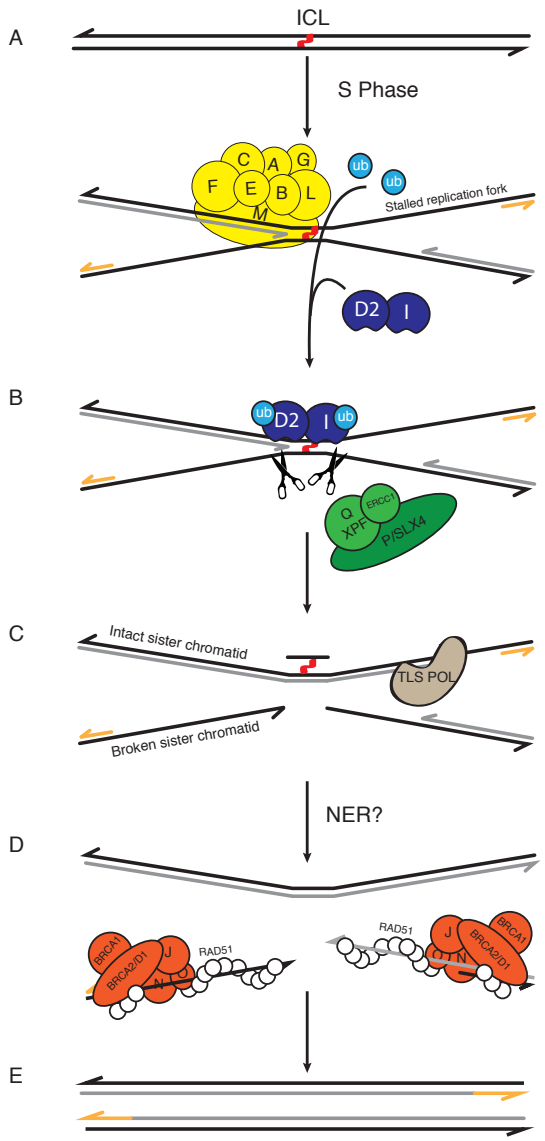

Figure 3 | Current view of Interstrand Crosslink Repair by the Fanconi Anemia Pathway. (A) Upon replication, the leading strands of two converging replication forks are blocked at an ICL (depicted in red). Recruitment of the FA core complex (indicated in yellow) takes place and ensures monoubiquitination of its substrates FANCD2 and FANCI. (B) Incision on both sides of the ICL is accomplished by the nucleases XPF (FANCQ), ERCC1 and SLX4 (FANCP), resulting in the uncoupling of the two sister chromatids. (C) Translesion synthesis ensures extension of the nascent strand beyond the ICL and it is thought that the crosslink is removed by NER. (D) HR is used to repair the broken chromatid. (E) Together, ICL-repair by the FA pathway results in two fully repaired DNA duplexes. Figure adapted from Garaycoechea & Patel, 2014.

whether exposure to different levels of these toxic metabolites could also explain the phenotypically heterogeneity found in FA-patients.

Double-strand break repair

DNA double-strand breaks (DSBs) are dangerous lesions for a cell. Inappropriate repair of DSBs can lead to loss of genomic information, inversions and, perhaps most hazardous of all, chromosome translocations, which can lead to the formation of oncogenic gene-fusions. Given that cells frequently endure DSBs and that DSBs are formed under various conditions (e.g. during replication, mitosis, meiosis), it is not surprisingly that cells have evolved several DSB repair pathways to preserve genomic integrity. Two prominent DSB repair pathways are homologous recombination (HR) and nonhomologous end-joining (NHEJ). HR is considered as an error-free process but requires the use of a homologous template. Therefore, HR can usually only take place for breaks that occur during or after DNA replication, when an identical sister chromatid is available as template (so during the S and G2 phase of the cell cycle). In contrast, the error-prone pathway NHEJ is able to ligate two broken ends of a chromosome without the need of a homologous template and can therefore also be active in cells in the G1-phase. A third pathway that is able to repair DSBs goes by a variety of names of which alternative NHEJ (alt-NHEJ) and microhomology-mediated end joining (MMEJ) are most commonly used. As the latter name implies, this pathway makes use of small pieces of homology to fuse two broken ends together. Below I will discuss these three pathways in more detail.

Homologous recombination (HR)

Crucial for homologous recombination is the generation of 3′ single-stranded DNA (ssDNA) tails that are required to find and invade a homologous template. Recent studies suggest a two-step model for the generation of these overhangs ((Symington, 2014) and references therein): first, the DSB is recognized, bound and processed by the MRE11-RAD50-NBS1 complex leaving a short 3′ overhang. In yeast, and likely in higher organisms, this 3′ short overhang is generated by endonuclease activity of Mre11 in cooperation with Sae2 (homolog of the human nuclease CtIP) 15-20nt downstream of the break, followed by 3′ to 5′ exonuclease activity of Mre11 towards the break (Cannavo and Cejka, 2014). The second step involves long-range resection by the nucleases EXO1 and DNA2 generating extensive tracts of ssDNA. To prevent the formation of secondary structures, the ssDNA becomes coated with the heterotrimeric complex RPA (RPA1, RPA2, RPA3).

After resection and coating of the ssDNA with RPA, loading of RAD51 takes place, aided by BRCA2, creating a nucleoprotein filament that is able to invade homologous duplex DNA (known as D-loop formation). After invasion, the strand can be extended by a DNA polymerase, dissociate (a process stimulated by helicase RTEL1) and re-anneal to the other end of the break (a process called synthesis dependent strand annealing, SDSA). Alternative to SDSA, double Holliday Junctions can be formed,

and endonucleases GEN1, MUS81/EME1, SLX1/SLX4 are required for resolving the intertwined DNA strands(Boulton et al., 2012; Ciccia and Elledge, 2010).

Unlike HR, which is error-free, an error-prone alternative when breaks are surrounded by repeat sequences is single strand annealing (SSA). Independent of RAD51, but catalyzed by RAD52, annealing of the resected strands can take place at the two repeats, followed by flap removal by XPF/ERCC1 and ligation (Ciccia and Elledge, 2010).

Classical Nonhomologous end-joining

In classical NHEJ (cNHEJ), the first step is the binding of the heterodimer Ku (Ku70 and Ku80). This step happens within seconds after DSB-formation, and binding of Ku prevents resection of the break ((Mahaney et al., 2009) and references therein). Next, Ku translocates inwards, allowing the recruitment of the protein kinase DNA-PKcs at the DNA termini, thereby assisting in tethering the broken ends together. To remove non-ligatable end groups or other lesions, processing may occur at the termini by among others the exonuclease ARTEMIS and polymerases λ and μ. Finally, the break is ligated by XRCC4 and Ligase IV (LIG4), with the help of XLF. Because frequently limited processing of the DNA ends takes places, NHEJ is characterized by small deletions or insertions and is therefore considered an error-prone repair pathway.

Alt-NHEJ/ MMEJ

Two main features characterize alt-NHEJ: first of all, it accomplishes the repair of a break without the requirement of the classical NHEJ factors such as Ku and Ligase IV. Second, the repair products are often characterized by excessive deletions, microhomology of 1-10 base pairs and templated insertions (Deriano and Roth, 2012). This pathway has very recently been subjected to increasing investigation and many questions remain to be addressed. For example, what determines whether a DSB is repaired via cNHEJ or Alt-NHEJ? One study suggests that perhaps PARP1 could play a role in this process by competing with Ku and thereby directing the repair of the break towards Alt-NHEJ (Wang et al., 2006). Furthermore it remains poorly understood which nucleases, polymerases and ligases are involved in processing and ligating the DNA termini.

Since many chromosomal translocations show characteristics of Alt-NHEJ (Decottignies, 2012), it is clear that better understanding of the process of Alt-NHEJ can be of great value.

DNA POLYMERASES AND HELICASES IN DNA REPAIR

DNA polymerases and helicases play an important role in DNA repair pathways. In chapter 4 and 5 of this thesis, I investigate the genetic consequences of G-quadruplex instability, and to which extent polymerases and helicases are involved in the prevention and the repair of G-quadruplex-induced DNA damage. In the following paragraphs, I provide a short introduction to the polymerases and helicases involved in DNA repair.

Polymerases in DNA repair

Both repair and bypass of damaged DNA often require DNA polymerase activity. The mammalian genome encodes at least 16 DNA polymerases, which can be subdivided in four main families (A, B, X and Y, see also table 1). Replication of undamaged DNA is performed by polymerases from the B-family, including Pol α, Pol δ and Pol ε. Initiation of DNA synthesis depends on the Pol α –primase complex. The primase synthesizes an oligo of 7-12 ribonucleotides, which is then elongated by pol α with ±20-30 deoxyribonucleotides (Muzi-Falconi et al., 2003). Next, the Pol α –primase complex is substituted by either Pol δ or Pol ε, which will in their turn elongate the synthesized RNA-DNA hybrid in the lagging or leading strand, respectively. For a long time it was thought that in human cells the Pol α –primase complex was the sole complex that could initiate *de novo* synthesis. However, a recent study discovered a second primase, named PRIMPOL, which is furthermore a TLS polymerase and is involved in the repriming of DNA synthesis at stalled replication forks (García-Gómez et al., 2013; Mourón et al., 2013).

Polymerases from the Y-family (Pol η, Pol κ, Pol ι, and REV1) are mainly involved in translesion synthesis. These Y-family polymerases have a more open catalytic active site (compared to Pol δ and Pol ε), which allows them to synthesize DNA past damaged nucleotides. A major function of Pol η is to bypass UV-induced CPDs and defective Pol η will lead to the cancer predisposition disease Xeroderma Pigmentosum. Pol κ and Pol ι are mainly involved in the bypass of N^2-dG adducted sites and dA templates, respectively (Sale et al., 2012). REV1 can only incorporate dC residues opposite abasic cites and dG. In addition, REV1 plays an important role in the bypass of G-quadruplex structures in chicken cells (Sarkies et al., 2010).

In higher eukaryotes, the X-family of polymerases consists of four members, namely Pol β, Pol λ, Pol μ and TdT. Remarkably, some organisms, such as *C. elegans* and *D. melanogaster*, appear to be devoid of any X-family polymerase (Uchiyama et al., 2009). Pol β plays an important role in BER, whereas the other three polymerase are implicated in NHEJ (Yamtich and Sweasy, 2010). Notably, Pol μ and TdT are mainly expressed in lymphoid tissues and are thought to play an important role in V(D)J-recombination.

Three polymerases belong to the group of A-family polymerases. One member, Pol γ, is dedicated for the replication of mitochondrial DNA. Pol ν is considered to be a proficient TLS polymerase for the accurate bypass of thymine glycols (Takata et al., 2006). The last member of the A-family polymerases is named Pol θ, also known as POLQ. Pol θ is a large 290kDA protein and is characterized by an N-terminal ATPase-helicase like domain and a C-terminal polymerase domain, flanking a large central

Table 1 | An overview of mammalian and *C. elegans* genes involved in DNA repair pathways. Also an overview is provided of all DNA polymerases and a selection of prominent helicases and their polarity.

Mammalian Homolog	(putative) *C. elegans* homolog		
HR		**ICL /Fanconi Anemia pathway**	
MRE11	*mre-11*	*FANCA*	-
RAD50	*rad-50*	*FANCB*	-
NBS1	-	*FANCC*	-
CtIP	*com-1*	*FANCD1/BRCA2*	*brc-2*
EXO1	*exo-1*	*FANCD2*	*fcd-2*
DNA2	*dna-2*	*FANCE*	-
RPA1-3	*rpa-1 - 3*	*FANCF*	-
RAD51	*rad-51*	*FANCG/XRCC9*	-
RAD52	-	*FANCI*	*fnci-1*
BRCA1	*brc-1*	*FANCJ/BRIP1/BACH1*	*dog-1*
BRCA2	*brc-2*	*FANCL/POG*	-
GEN1	*gen-1*	*FANCM*	*fncm-1*
SLX1	*slx-1*	*FANCN/PALB2*	-
SLX4/FANCP	*slx-4*	*FANCO/RAD51C*	-
MUS81	*mus-81*	*FANCP/SLX4*	*slx-4/him-18*
EME1	*eme-1*	*FANCQ/XPF*	*xpf-1*
		FAN1	*fan-1*
NHEJ			
KU70	*cku-70*	**NER**	
KU80	*cku-80*	*DDB1*	*ddb-1*
LIG4	*lig-4*	*DDB2*	-
DNAPK	-	*ERCC1*	*ercc-1*
XRCC4	-	*ERCC4/XPF/FANCQ*	*xpf-1*
Artemis	-	*ERCC5/XPG*	*xpg-1*
XLF	-	*CSA*	-
		CSB	*csb-1*
MMR		*LIG1*	*lig-1*
MSH2	*msh-2*	*LIG3*	K07C5.3
MSH3	-	*PCNA*	*pcn-1*
MSH6	*msh-6*	*RFC1-5*	*rfc-1 – 4*, F44B9.8
MLH1	*mlh-1*	*RPA1-3*	*rpa-1 -3*
PMS2	*pms-2*	*XPA*	*xpa-1*
EXO1	*exo-1*	*XPB/ERCC3*	Y66D12A.15
		XPC	*xpc-1*
		XPD/ERCC2	Y50D7A.2

Table 1 | An overview of mammalian and *C. elegans* genes involved in DNA repair pathways. Also an overview is provided of all DNA polymerases and a selection of prominent helicases and their polarity. (*Continued*)

Mammalian Homolog	(putative) *C. elegans* homolog
BER	
-	*apn-1*
APE1	*exo-3*
NTHL1	*nth-1*
UDG/UNG	*ung-1*
Polymerase β	-
LIG3	K07C5.3
FEN1	*crn-1*
PARP1	*pme-1*
PARP2	*pme-2*

Polymerases		**Helicases**	
Family A		RecQ-family	
Pol γ	*polg-1*	*RECQ1* (3′-5′)	K02F3.12
Pol ν	-	*BLM* (3′-5′)	*him-6*
Pol θ /POLQ	*polq-1*	*WRN* (3′-5′)	*wrn-1*
		RECQ4 (3′-5′)	-
Family B		*RECQ5* (3′-5′)	*rcq-5*
Pol α	*div-1*		
Pol δ	F10C2.4	Fe-S cluster	
Pol ε	F33H2.5	*FANCJ* (5′-3′)	*dog-1*
Pol ζ / REV3	Y37B11A.2	*XPD* (5′-3′)	Y50D7A.2
		RTEL1 (5′-3′)	*rtel-1*
Family X		*DDX11/CHL1* (5′-3′)	*chl-1*
Pol β	-		
Pol λ	-	others	
Pol μ	-	*PIF1* (5′-3′)	*pif-1*
TdT	-	*XPB* (3′-5′)	Y66D12A.15
		DNA2 (3′-5′)	*dna-2*
Family Y		*ATRX*	*xnp-1*
Pol η	*polh-1*	*HELQ* (3′-5′)	*helq-1*
Pol κ	*polk-1*		
Pol ι	-		
REV1	*rev-1*		

domain. Pol θ is highly expressed in the testis, placental tissue and hematopoietic cells (Seki et al., 2003; Shima et al., 2004) (Kawamura et al., 2004). Overexpression of Pol θ has been observed in several cancers and correlates with a lower patient survival rate(Higgins et al., 2010; Lemée et al., 2010). Pol θ is a low fidelity polymerase but has the capability to extend DNA from minimally paired primers(Arana et al., 2008; Seki et al., 2004; Yousefzadeh et al., 2014). As research progresses, more and more functions are ascribed to Pol θ. Pol θ is implicated in the bypass of abasic sites and thymic glycols (Seki et al., 2004; Yoon et al., 2014), functioning as a backup polymerase in BER (Asagoshi et al., 2012; Prasad et al., 2009; Yoshimura et al., 2006), linked to Alt-NHEJ and ICL repair in *D. Melanogaster* (Chan et al., 2010; Harris et al., 1996) and found to be involved in the timing of firing of origins of replication (Fernandez-Vidal et al., 2014).

Helicases in DNA repair

Helicases are ATP-dependent motor proteins that are able to unwind duplex nucleic acids. Various cancers and genetic disorders are linked to helicase defects, which illustrates their importance. Their prominence is furthermore marked by the great number of helicases found; a recent computational study reported 95 human genes encoding for helicases, of which 64 and 31 are thought to be RNA and DNA helicases, respectively (Umate et al., 2011). Based on motifs and consensus sequences, helicases have been classified in two larger superfamilies (SF1 and SF2) and four smaller superfamilies (SF3-6) (Singleton et al., 2007). Other classifications are based on whether the helicase acts on single or double-strand DNA (indicated by α and β, respectively) and by their polarity; type A helicases translocate in a 3′ to 5′ direction and type B helicases from 5′ to 3′. Most helicases discussed in this thesis belong to the SF2 family. Two prominent subclasses of the SF2 family are the RecQ family and Fe-S family. The RecQ family consists of five 3′ to 5′ helicases (RECQL1, BLM, WRN, RECQL4 and RECQL5). They are highly conserved and required for genome stability (Chu and Hickson, 2009). Defects in BLM, WRN and RECQL4 are linked to syndromes that predispose to cancer (Bloom syndrome, Werner syndrome and Rothmund-Thomson syndrome, respectively). RecQ helicases are primarily related to the repair of DSBs, fork regression and Holliday junction branch migration (Brosh, 2013). Furthermore there is biochemical and *in vivo* data that BLM and WRN are involved in the unwinding and bypass of G4 DNA (Fry and Loeb, 1999; Sarkies et al., 2012; Sun et al., 1998).

In contrast to the RecQ family, helicases belonging to the Fe-S family have a 5′ to 3′ polarity and are characterized by a conserved iron-sulfur (Fe-S) cluster. Although the exact role of the Fe-S cluster remains to be elucidated, it is thought that its redox properties are used to scan the genome for DNA damage (Wu and Brosh, 2012). Four DNA helicases belong to the Fe-S family (XPD, FANCJ/BRIP1/BACH1, RTEL1 and DDX11/CHL1/ChiR1) and are all implicated in autosomal recessive genetic disorders. XPD plays an important role in NER and is linked to Xeroderma Pigmentosum.

DDX11 is connected with the Warsaw Breakage Syndrome and is important for sister chromatid cohesion during DNA repair (van der Lelij et al., 2010). Regulator of telomere elongation helicase 1 (RTEL1) is indispensable for the maintenance of telomeres and for the dismantling of D-loop recombination intermediates (Barber et al., 2008; Vannier et al., 2013; 2012). Defective RTEL1 leads to dyskeratosis congenital (Ballew et al., 2013). Homozygous mutations in FANCJ lead to Fanconi Anemia (Levitus et al., 2005), whereas female carriers of monoallelic mutations in FANCJ have an elevated risk to develop breast cancer (Hiom, 2009). Several interacting proteins have been described for FANCJ of which the most prominent are BRCA1, MLH1, MRE11, RPA and FANCD2 (Cantor et al., 2001; Chen et al., 2014; Guillemette et al., 2014; Sommers et al., 2014; Suhasini et al., 2013). As research progresses, functions are described for FANCJ in ICL (Levitus et al., 2005), MMR, NER (Guillemette et al., 2014), displacement of DNA-protein blocks (Sommers et al., 2014) and unwinding of G-quadruplexes (Bosch et al., 2014; Cheung et al., 2002; Kruisselbrink et al., 2008; Sarkies et al., 2012; Schwab et al., 2013) and thereby maintaining genomic and epigenetic stability.

Other helicases that are considered as important genome caretakers are PIF1, ATRX, DNA2, and XPB. Although they function in different pathways such as Break-Induced Replication (PIF1) (Wilson et al., 2013) and NER (XPB), they share the property to unwind G4 DNA (Gray et al., 2014; Law et al., 2010; Lin et al., 2013; Ribeyre et al., 2009).

For a more detailed overview about DNA helicases and their role in DNA repair and cancer see reference (Brosh, 2013).

MICROSATELLITES AND G-QUADRUPLEX STRUCTURES

Apart from DNA damage, genomic integrity is endangered by DNA sequences that are difficult to replicate. In this thesis, I focus on two of such sequences: microsatellites and G-quadruplex sequences. In the next section, I will describe these mutagenic sequences in more detail.

Microsatellites

More than 40% of the human genome consists of repeats (Gemayel et al., 2010). Although historically these repeats were seen as “junk DNA” and therefore ignored, we now start to realize that their importance has been severely misjudged; changes in repeat-length can lead to phenotypic changes and many diseases, particularly neurodegenerative diseases.

Repeats can be categorized in two main groups: interspersed and tandem repeats (TR). Interspersed repeats are remnants of transposons and, as the name suggests, are interspersed throughout the genome. Tandem repeats consist of a short DNA sequence, named a “unit”, that is repeated several times right next to the other. Tandem repeats can

be subdivided in microsatellites, minisatelites and megasatellites. This classification is based on the length of the unit. Although definitions vary, microsatellites (also known as short tandem repeats) are repeats with units of 1-8 nucleotides, minisatellites are 9-135 nucleotides, megasatellites consist of untis greater than 135 nucleotides (Gemayel et al., 2010). With the completion of sequencing and assembly of the human genome, it appeared that approximately 17% of all genes contain a TR in their open reading frame (ORF). Of these TRs, microsatellites have been under great investigation, since they are thought to influence processes such as gene expression, chromatin organization and recombination at hotspots (Li et al., 2002).

Approximately 3% of the genome consists of microsatellites and the distribution of microsatellites appears to be nonrandom (Katti et al., 2001; Lander et al., 2001). The majority of microsatellites consist of mono-, di-, tri- and tetranucleotide repeats, of which the dinucleotide repeats are the most prominent. Microsatellites can be extremely unstable; mutation rates vary between 10^{-3} and 10^{-7} per cell division but rates above 10^{-2} have been described (Gemayel et al., 2010). Since their discovery in the early 1980s, it quickly became apparent that their instability and subsequent polymorphisms make microsatellites excellent markers for genome mapping and population genetics.

Two major models are proposed for microsatellite expansion and contraction: recombination and strand-slippage during replication (Gemayel et al., 2010). In the latter, the nascent strand denatures from the template strand during synthesis of the microsatellite and then pairs with another part of the repeat sequence. This can lead to a loop either in the nascent strand or in the template strand, resulting in expansion or contraction of the microsatellite, respectively. Notably, in most cases, these loops are recognized by the mismatch repair machinery, thereby preventing microsatellite instability. During a recombination event, unequal crossover or gene conversion can lead to expansion/contraction of a microsatellite.

Microsatellites are abundant in the genome, are very instable, linked to important biological functions and connected to disease. These aspects stress the need to elucidate the factors and mechanisms involved in microsatellite instability.

G-quadruplex structures

In 1910, the German chemist Ivar Bang made the observation that guanylic acid formed gels when kept at high concentrations (Bang, 1910). More than fifty years thereafter Gellert and colleagues (Gellert et al., 1962) found an explanation for this unusual physical property; using X-ray diffraction techniques they showed that guanylic acids can assemble into tetrameric structures, also named G-quartets (see Figure 4a+b). In this configuration each of the four guanine molecules is a donor and acceptor of two hydrogen bonds and in the central cavity a metal cation plays an important role in stabilizing the G-quartet. Stacking of multiple G-quartets forms a stable G4 structure (hereafter also named G-quadruplex and G4 DNA).

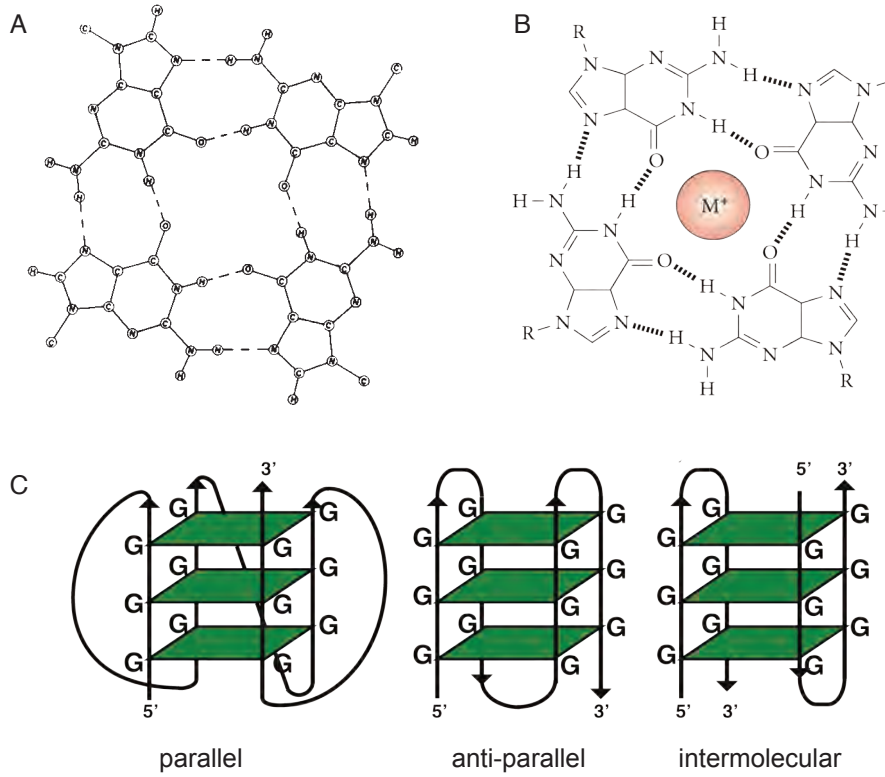

Figure 4 | G-quadruplex DNA. (A) Proposed arrangement of interacting guanines in a G-quartet by Gellert et al in 1962. (B) Current view of interactions in a G-quartet. M+ denotes a monovalent cation. Illustration adapted from Bochman et al, 2012. (C) Schematic representation of intramolecular G-quadruplex structures in a parallel and anti-parallel conformation (left and middle panels). The right panel illustrates the topology of an intermolecular G-qaudruplex structure formed by dimerization of two strands. Illustration adapted from Tarsounas & Tijsterman, 2013.

G-quadruplexes come in many flavours: they can form within one strand (intramolecular) or from two or more strands (intermolecular), strands can run in a parallel or antiparallel orientation (Figure 4c), and various loop structures can form by the nucleotide linkers between the stacks. Additionally, G-quadruplexes can consist of DNA or RNA molecules or a combination of both. The stability of a G-quadruplex depends on multiple factors: the number of G-quartets formed, the size of the loops and the nature of the stabilizing cation.

Although the formation of G-quadruplexes was shown *in vitro*, many researchers remained skeptic about their presence *in vivo* and about a potential biological function. However, this skepticism is likely greatly reduced with the publication of some seminal publications in the last couple of years. Below, I will briefly introduce some of these groundbreaking publications that demonstrate that G-quadruplexes are

present *in vivo*, cause genomic and epigenetic instability, have a biological function and are linked to several diseases.

G-quadruplexes and genomic instability

In 2002 the lab of Lansdorp identified a helicase-defective worm that triggered deletions upstream guanine-rich DNA (Cheung et al., 2002). They named the gene encoding the helicase *dog-1* (for deletion of guanine-rich DNA), which later turned out to be the homolog of human FANCJ (Youds et al., 2008). Six years later our lab demonstrated that solely guanine-rich sequences that match the G-quadruplex consensus motif ($G_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}$) lead to the induction of deletions (Kruisselbrink et al., 2008). Besides in worms, G-quadruplex sequences appeared to cause genomic instability in yeast (Piazza et al., 2012; Ribeyre et al., 2009) and in human cells (Rodriguez et al., 2012). In addition, it has been shown that G-quadruplex sequences are enriched in breakpoints in cancer genomes (De and Michor, 2011). Studies in chicken cells furthermore show that G-quadruplexes can lead to epigenetic instability (Sarkies et al., 2010; 2012; Schwab et al., 2013). Finally, a study in mice shows that G-quadruplex formation endangers telomere integrity (Vannier et al., 2012).

Evidence of G-quadruplex formation *in vivo*

Besides the previously described publications that (indirectly) imply that G-quadruplexes must form *in vivo*, several labs have attempted to visualize G-quadruplexes *in vivo* with help of antibodies. In 2001 a study provided evidence that telomeres of the single-celled eukaryote *Stylonychia lemnae* formed G-quadruplex structures *in vivo*. Recent studies with newly developed antibodies showed the presence of G-quadruplexes in mammalian cells *in vivo* (Biffi et al., 2013; Henderson et al., 2013). G-quadruplexes appear to be enriched in replicating cells and cancerous tissue (Biffi et al., 2013; 2014).

G-quadruplexes and their function

One of the first publications describing an *in vivo* function for G-quadruplexes was a study performed in the bacteria *Neisseria gonorrhoeae*. Here, Cahoon and Seifert showed that a G-quadruplex drives antigenic variation by serving as a recombination hotspot. Disruption of the G4 motif by changing only a single nucleotide blocked recombination and subsequent antigenic variation (Cahoon and Seifert, 2009).

More recently, G-quadruplexes have been found to be important in defining the origins of replication (Besnard et al., 2012; Hoshina et al., 2013; Valton et al., 2014). Furthermore G-quadruplexes have been implicated in transcription, RNA localization, translation, telomere protection and meiosis (see reference (Bochman et al., 2012) and references therein). Another argument that G-quadruplexes have a biological purpose is given by a genome-wide computational analysis, which shows that G4 motifs are evolutionary conserved in yeast (Capra et al., 2010).

G-quadruplexes and disease

The link between G-quadruplexes and disease is growing rapidly. One of the first diseases identified with a clear link to G-quadruplexes was the ATR-X syndrome, which is characterized by mental retardation and α-thalassaemia (Law et al., 2010). ATR-X was shown to bind G-quadruplexes and it was suggested that mutated ATR-X in combination with elevated G-quadruplex formation leads to epigenetic changes at the α-globin locus and subsequently anemia. A different recent study reported that G-quadruplex formation in DNA and RNA-molecules at the C9*orf*72 locus causes the neurodegenerative disease amyotrophic lateral sclerosis (ALS) (Haeusler et al., 2014). Another compelling study suggests that RNA-G-quadruplexes can control the translation of oncoproteins as MYC, NOTCH and BCL2 (Wolfe et al., 2014). Finally, since G-quadruplexes can drive genomic instability, G-quadruplexes have a clear link to cancer. Understanding how G-quadruplex structures can lead to genomic instability is therefore crucial.

MODEL ORGANISMS

Apart from human cell lines, I have used the nematode *Caenorhabditis elegans* and the zebrafish *Danio rerio* as model organisms for the work described in this thesis. Below, I will give a brief introduction about the strengths and limitations of *C. elegans* and *D. rerio*.

C. elegans

C. elegans is a multicellular organism of ± 1mm in size and consists of 959 cells. It has a short life cycle of approximately 3 days and a life span of three weeks. A fertilized egg develops into an adult worm via four larval stages named L1- L4. In the absence of food, larvae can switch to a stage called dauer, which allows the worm to survive up to several months. Worms can also be kept as frozen stock at -80 degrees for years, if not decades. In the laboratory the nematode is usually grown on agar plates or in liquid cultures and uses *E. coli* as a food source and is therefore relatively inexpensive to maintain. In a wildtype population two *C. elegans* sexes are found: the majority are self-fertilizing hermaphrodites and 0.1% are male.

C. elegans is highly appreciated for its genetics. The ability to self-fertilize enables the generation of genetically identical progeny, whereas the use of males allows the combination of mutations via crossings. Forward genetic screens can be easily performed via mutagenesis, whereas RNAi-based reverse genetic screens are applicable since the availability of genome-wide RNAi libraries (Kamath et al., 2003). Furthermore, many mutants are available and well documented at www.wormbase.org. The number of available mutant alleles was recently boosted by a so-called million mutation project in which more than 2000 mutagenized strains were sequenced, resulting in a library containing more than 800,000 unique single

nucleotide variants and 16,000 insertions/deletions (indels) (Thompson et al., 2013). Until recently, a limitation of *C. elegans* was the inability to perform targeted genome editing. However, this hurdle is now overcome by the availability of tools using zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and the clustered regulatory interspaced short palindromic repeat (CRISPR)/Cas9 system (Waaijers et al., 2013; Wood et al., 2011).

In 1998 the entire genome of *C. elegans* was sequenced and it quickly became apparent that many genes and pathways are strongly conserved. Also many genes involved in DNA repair are highly conserved and many human counterparts can be found back in the worm's genome as illustrated in table 1. Another feature that makes *C. elegans* an attractive model for the study of DNA repair is its germline, which has a spatio-temporal organization of mitotic and meiotic cells that can be easily monitored for the presence of damaged DNA (Lemmens and Tijsterman, 2011).

D. rerio

In recent years, the zebrafish has obtained a prominent role in biomedical research. Historically, the zebrafish has mostly been used to study developmental biology, however with evolving techniques the zebrafish has become also an excellent tool for studying disease mechanisms. Since the zebrafish is optically translucent for the first few weeks, the development from a single cell to a swimming fish can be monitored in great detail (see Figure 5a for the first 48 hours of embryonic development of a zebrafish). Recently, also transparent adult fish (Figure 5b) have been created, which makes it relatively easy to perform live *in vivo* imaging in full grown fish (White et al., 2008).

Within 3 months an embryo develops into a fertile adult, and its total life span usually varies between 2 to 3 years. On a weekly basis, females can spawn up to hundreds of eggs per clutch. These eggs can be easily collected and injected with transgenes or morpholinos (oligos that inhibit the expression of a protein). Until recent, targeted genome engineering was not possible and only forward genetic screens (mutagenesis or transposon-based) and target-selected mutagenesis (requiring extensive sequencing) (Wienholds et al., 2003) led to a relatively small library of mutants. However, in the last four years ZFN-, TALEN and CRISPR/Cas9-technology (Hwang et al., 2013) have proven to be exquisite tools for targeted genome editing in the zebrafish.

In the last decade, the zebrafish has been increasingly used as a model in cancer research (Amatruda and Patton, 2008). Many genetic cancer models and tools have been developed, but also xenotransplantations, in which human tumorigenic cells are injected, are becoming prevalent.

A drawback of working with zebrafish as a model organism is that it takes approximately five to six months before a homozygous animal is established, and that creating a homozygous mutant can sometimes be complicated because some segments of the zebrafish genome have been duplicated and thus for numerous genes there are duplicated copies.

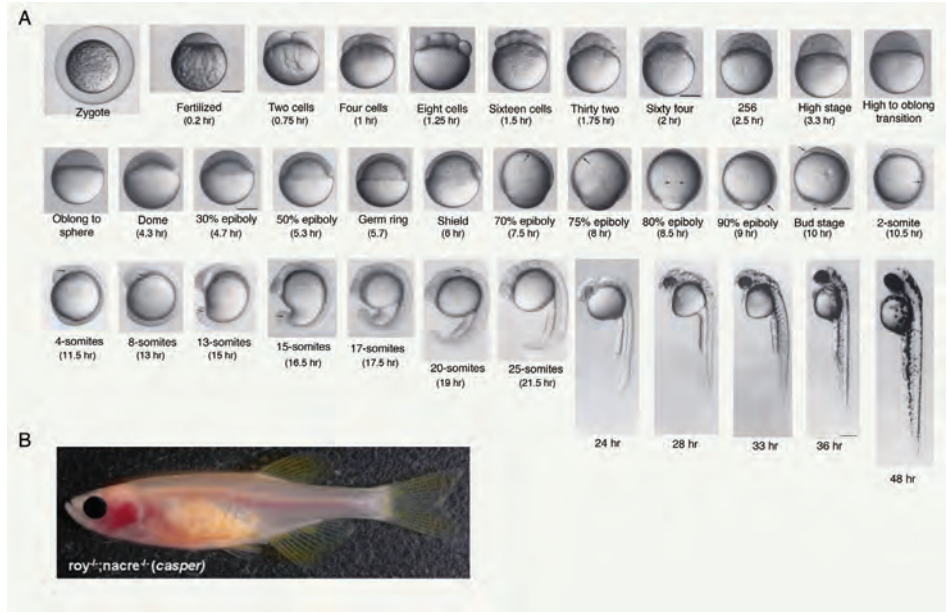

Figure 5 | The zebrafish *Danio rerio*. (A) Embryonic stages during zebrafish development. Figure adapted from Kimmel *et al*,1995. (B) Picture of an adult zebrafish of which the body is largely transparent due to the loss of melanocyte and iridophores. Figure adapted from White *et al*, 2008.

AIM OF THIS THESIS

Microsatellites and G-quadruplexes are sequences that are abundant in the genome and are linked to diseases such as cancer and neurodegenerative diseases. Unstable microsatellites and G-quadruplexes are thought to be an important underlying cause of these devastating diseases. However, many aspects about G4 DNA and microsatellite instability are incompletely understood. For example, what determines why some microsatellites and G-quadruplexes are more prone to induce mutations than others? Which genes and pathways prevent microsatellite and G4 DNA instability? What are the genetic consequences of microsatellite and G-quadruplex instability and which molecular mechanisms act to produce genomic changes at these sequences? The answers to these questions will be of great importance in the development of new and better treatments of microsatellite- and G4 DNA-related diseases. In this thesis I aim to provide new insights into the biology concerning microsatellite and G-quadruplex instability.

OUTLINE OF THIS THESIS

Chapter 2 concerns microsatellite instability. We investigate whether factors such as tract length, orientation, nucleotide composition and transcription influence the stability of a microsatellite. To this end, we make use of newly-developed reporters that read out microsatellite instability in human cells. We furthermore test whether these reporter systems can be used to screen for microsatellite instability-inducing compounds as well as for genes that protect the genome against microsatellite instability.

Chapter 3 presents a new genetic tool that enables mosaic analysis in the zebrafish. We developed technology that employs MSI, which we show to be a powerful tool to trace single cells and also to study tumor induction in a living animal.

Chapter 4 focuses on G4 DNA instability in *C. elegans*. Previous studies have shown that G-quadruplexes can induce deletions in the genome typically 50-300bp in size. In this chapter, we reveal the molecular mechanism that explains the formation of these deletions.

In **Chapter 5** we examine G4 DNA instability in human cells. We address the question whether G-quadruplexes are fragile in human cells as well and whether they induce deletions through a similar mechanism as witnessed in *C. elegans*.

In **Chapter 6** I provide a summarizing discussion and include a number of future perspectives related to microsatellite and G-quadruplex instability and their link to disease.

REFERENCES

ACS, A.C.S. (2013). Cancer Facts & Figures 2013. 1–64.

Amatruda, J.F., and Patton, E.E. (2008). Genetic models of cancer in zebrafish. International Review of Cell and Molecular Biology *271*, 1–34.

Arana, M.E., Seki, M., Wood, R.D., Rogozin, I.B., and Kunkel, T.A. (2008). Low-fidelity DNA synthesis by human DNA polymerase theta. Nucleic Acids Res. *36*, 3847–3856.

Asagoshi, K., Lehmann, W., Braithwaite, E.K., Santana-Santos, L., Prasad, R., Freedman, J.H., Van Houten, B., and Wilson, S.H. (2012). Single-nucleotide base excision repair DNA polymerase activity in C. elegans in the absence of DNA polymerase β. Nucleic Acids Res. *40*, 670–681.

Ballew, B.J., Yeager, M., Jacobs, K., Giri, N., Boland, J., Burdett, L., Alter, B.P., and Savage, S.A. (2013). Germline mutations of regulator of telomere elongation helicase 1, RTEL1, in Dyskeratosis congenita. Hum. Genet. *132*, 473–480.

Bang, I. (1910). Untersuchungen über die Guanylsäure. Biochemische Zeitschrift.

Barber, L.J., Youds, J.L., Ward, J.D., McIlwraith, M.J., O'Neil, N.J., Petalcorin, M.I.R., Martin, J.S., Collis, S.J., Cantor, S.B., Auclair, M., et al. (2008). RTEL1 maintains genomic stability by suppressing homologous recombination. Cell *135*, 261–271.

Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., Marin, J.-M., and Lemaitre, J.-M. (2012). Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. Nature Structural & Molecular Biology *19*, 837–844.

Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., Vitale, L., Pelleri, M.C., Tassani, S., Piva, F., et al. (2013). An estimation of the number of cells in the human body. Ann. Hum. Biol. *40*, 463–471.

Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013). Quantitative visualization of DNA G-quadruplex structures in human cells. Nature Chemistry *5*, 182–186.

Biffi, G., Tannahill, D., Miller, J., Howat, W.J., and Balasubramanian, S. (2014). Elevated Levels of G-Quadruplex Formation in Human Stomach and Liver Cancer Tissues. PLoS ONE *9*, e102711.

Bochman, M.L., Paeschke, K., and Zakian, V.A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. Nat. Rev. Genet. 1–11.

Bosch, P.C., Segura-Bayona, S., Koole, W., van Heteren, J.T., Dewar, J.M., Tijsterman, M., and Knipscheer, P. (2014). FANCJ promotes DNA synthesis through G-quadruplex structures. Embo J *33*, 2521–2533.

Boulton, J.N.C.M.G.T.S., Taylor, M.R.G., and Boulton, S.J. (2012). Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. Molecular Cell *47*, 497–510.

Brosh, R.M. (2013). DNA helicases involved in DNA repair and their roles in cancer. Nature Publishing Group *13*, 542–558.

Cahoon, L.A., and Seifert, H.S. (2009). An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae. Science *325*, 764–767.

Cannavo, E., and Cejka, P. (2014). Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks. Nature *514*, 122–125.

Cantor, S.B., Bell, D.W., Ganesan, S., Kass, E.M., Drapkin, R., Grossman, S., Wahrer, D.C., Sgroi, D.C., Lane, W.S., Haber, D.A., et al. (2001). BACH1, a Novel Helicase-like Protein, Interacts Directly with BRCA1 and Contributes to Its DNA Repair Function. Cell *105*, 149–160.

Capra, J.A., Paeschke, K., Singh, M., and Zakian, V.A. (2010). G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. PLoS Comput. Biol. *6*, e1000861.

Chan, S.H., Yu, A.M., and McVey, M. (2010). Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila. PLoS Genet. *6*, e1001005.

Chen, X., Wilson, J.B., McChesney, P., Williams, S.A., Kwon, Y., Longerich, S., Marriott, A.S., Sung, P., Jones, N.J., and Kupfer, G.M. (2014). The Fanconi anemia proteins FANCD2 and FANCJ interact and regulate each other's chromatin localization. J. Biol. Chem. *289*, 25774–25782.

Cheung, I., Schertzer, M., Rose, A., and Lansdorp, P.M. (2002). Disruption of dog-1 in Caenorhabditis elegans triggers deletions upstream of guanine-rich DNA. Nat. Genet. *31*, 405–409.

Chu, W.K., and Hickson, I.D. (2009). RecQ helicases: multifunctional genome caretakers. Nature Publishing Group *9*, 644–654.

Ciccia, A., and Elledge, S.J. (2010). The DNA Damage Response: Making It Safe to Play with Knives. Molecular Cell *40*, 179–204.

Cimprich, K.A., and Cortez, D. (2008). ATR: an essential regulator of genome integrity. Nat. Rev. Mol. Cell Biol. *9*, 616–627.

Cleaver, J.E., Lam, E.T., and Revet, I. (2009). Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. Nat. Rev. Genet. *10*, 756–768.

De, S., and Michor, F. (2011). DNA secondary structures and epigenetic determinants of cancer genome evolution. Nature Structural & Molecular Biology *18*, 950–955.

Deans, A.J., and West, S.C. (2011). DNA interstrand crosslink repair and cancer. Nature Publishing Group *11*, 467–480.

Decottignies, A. (2012). Alternative end-joining mechanisms: a historical perspective. Front Genet *4*, 48–48.

Deriano, L., and Roth, D.B. (2012). Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. Genetics *47*, 433–455.

Fernandez-Vidal, A., Guitton-Sert, L., Cadoret, J.-C., Drac, M., Schwob, E., Baldacci, G., Cazaux, C., and Hoffmann, J.-S. (2014). A role for DNA polymerase θ in the timing of DNA replication. Nature Communications *5*, 4285.

Fry, M., and Loeb, L.A. (1999). Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. J. Biol. Chem. *274*, 12797–12802.

Garaycoechea, J.I., and Patel, K.J. (2014). Why does the bone marrow fail in Fanconi anemia? Blood *123*, 26–34.

Garaycoechea, J.I.J., Crossan, G.P.G., Langevin, F.F., Daly, M.M., Arends, M.J.M., and Patel, K.J.K. (2012). Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. Nature *489*, 571–575.

García-Gómez, S., Reyes, A., Martínez-Jiménez, M.I., Chocrón, E.S., Mourón, S., Terrados, G., Powell, C., Salido, E., Méndez, J., Holt, I.J., et al. (2013). PrimPol, an Archaic Primase/Polymerase Operating in Human Cells. Molecular Cell *52*, 541–553.

Gellert, M., Lipsett, M.N., and Davies, R.D. (1962). Helix formation by guanylic acid. Proc. Natl. Acad. Sci. U.S.a. *48*, 2013–2018.

Gemayel, R., Vinces, M.D., Legendre, M., and Verstrepen, K.J. (2010). Variable Tandem Repeats Accelerate Evolution of Coding and Regulatory Sequences. Genetics *44*, 445–477.

Ghodgaonkar, M.M., Lazzaro, F., Olivera-Pimentel, M., Artola-Borán, M., Cejka, P., Reijns, M.A., Jackson, A.P., Plevani, P., Muzi-Falconi, M., and Jiricny, J. (2013). Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair. Molecular Cell *50*, 323–332.

Ghosal, G., and Chen, J. (2013). DNA damage tolerance: a double-edged sword guarding the genome. Transl Cancer Res *2*, 107–129.

Gray, L.T., Vallur, A.C., Eddy, J., and Maizels, N. (2014). G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. Nat. Chem. Biol. *10*, 313–318.

Guillemette, S., Branagan, A., Peng, M., Dhruva, A., Schärer, O.D., and Cantor, S.B. (2014). FANCJ localization by mismatch repair is vital to maintain genomic integrity after UV irradiation. Cancer Res. *74*, 932–944.

Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.A.J., Shaw, P.G., Kim, M.-S., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R., et al. (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature *507*, 195–200.

Harris, P.V., Mazina, O.M., Leonhardt, E.A., Case, R.B., Boyd, J.B., and Burtis, K.C. (1996). Molecular cloning of Drosophila mus308, a gene involved in DNA cross-link repair with homology to prokaryotic DNA polymerase I genes. Mol. Cell. Biol. *16*, 5764–5771.

Henderson, A., Wu, Y., Huang, Y.C., Chavez, E.A., Platt, J., Johnson, F.B., Brosh, R.M., Sen, D., and Lansdorp, P.M. (2013). Detection of G-quadruplex DNA in mammalian cells. Nucleic Acids Res. *42*, 860–869.

Higgins, G.S., Harris, A.L., Prevo, R., Helleday, T., McKenna, W.G., and Buffa, F.M. (2010). Overexpression of POLQ confers a poor prognosis in early breast cancer patients. Oncotarget *1*, 175–184.

Hiom, K. (2009). FANCJ: Solving problems in DNA replication. DNA Repair *9*, 250–256.

Hira, A., Yabe, H., Yoshida, K., Okuno, Y., Shiraishi, Y., Chiba, K., Tanaka, H., Miyano, S., Nakamura, J., Kojima, S., et al. (2013). Variant ALDH2 is associated with accelerated progression of bone marrow failure in Japanese Fanconi anemia patients. Blood *122*, 3206–3209.

Hoeijmakers, J.H.J. (2009). DNA damage, aging, and cancer. N. Engl. J. Med. *361*, 1475–1485.

Hoshina, S., Yura, K., Teranishi, H., Kiyasu, N., Tominaga, A., Kadoma, H., Nakatsuka, A., Kunichika, T., Obuse, C., and Waga, S. (2013). Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. J. Biol. Chem. *288*, 30161–30171.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.-R.J., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. *31*, 227–229.

Jackson, S.P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. Nature *461*, 1071–1078.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., et al. (2003). Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature *421*, 231–237.

Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A., et al. (2013). Mutational landscape and significance across 12 major cancer types. Nature *502*, 333–339.

Katti, M.V., Ranjekar, P.K., and Gupta, V.S. (2001). Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol. Biol. Evol. *18*, 1161–1167.

Kawamura, K., Bahar, R., Seimiya, M., Chiyo, M., Wada, A., Okada, S., Hatano, M., Tokuhisa, T., Kimura, H., Watanabe, S., et al. (2004). DNA polymerase theta is preferentially expressed in lymphoid tissues and upregulated in human cancers. Int. J. Cancer *109*, 9–16.

Klein Douwel, D., Boonen, R.A.C.M., Long, D.T., Szypowska, A.A., Räschle, M., Walter, J.C., and Knipscheer, P. (2014). XPF-ERCC1 acts in Unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4. Molecular Cell *54*, 460–471.

Knipscheer, P., Räschle, M., Smogorzewska, A., Enoiu, M., Ho, T.V., Schärer, O.D., Elledge, S.J., and Walter, J.C. (2009). The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. Science *326*, 1698–1701.

Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., et al. (2012). Rate of de novo mutations and the importance of father's age to disease risk. Nature *488*, 471–475.

Kottemann, M.C., and Smogorzewska, A. (2013). Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. Nature *493*, 356–363.

Kruisselbrink, E., Guryev, V., Brouwer, K., Pontier, D.B., Cuppen, E., and Tijsterman, M. (2008). Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCJ-defective C. elegans. Curr. Biol. *18*, 900–905.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Langevin, F.F., Crossan, G.P.G., Rosado, I.V.I., Arends, M.J.M., and Patel, K.J.K. (2011). Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. Nature *475*, 53–58.

Law, M.J., Lower, K.M., Voon, H.P.J., Hughes, J.R., Garrick, D., Viprakasit, V., Mitson, M., De Gobbi, M., Marra, M., Morris, A., et al. (2010). ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. Cell *143*, 367–378.

Lemée, F., Bergoglio, V., Fernandez-Vidal, A., Machado-Silva, A., Pillaire, M.-J., Bieth, A., Gentil, C., Baker, L., Martin, A.-L., Leduc, C., et al. (2010). DNA polymerase theta up-regulation is associated with poor survival in breast cancer, perturbs DNA replication, and promotes genetic instability. Proc. Natl. Acad. Sci. U.S.a. *107*, 13390–13395.

Lemmens, B.B.L.G., and Tijsterman, M. (2011). DNA double-strand break repair in Caenorhabditis elegans. Chromosoma *120*, 1–21.

Levitus, M., Waisfisz, Q., Godthelp, B.C., de Vries, Y., Hussain, S., Wiegant, W.W., Elghalbzouri-Maghrani, E., Steltenpool, J., Rooimans, M.A., Pals, G., et al. (2005). The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J. Nat. Genet. *37*, 934–935.

Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W., and Li, G.-M. (2013). The Histone Mark H3K36me3 Regulates Human DNA Mismatch Repair through Its Interaction with MutS&alpha. Cell *153*, 590–600.

Li, Y.-C., Korol, A.B., Fahima, T., Beiles, A., and Nevo, E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol. Ecol. *11*, 2453–2465.

Lieber, M.R. (2010). The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. Annu. Rev. Biochem. *79*, 181–211.

Lin, W., Sampathi, S., Dai, H., Liu, C., Zhou, M., Hu, J., Huang, Q., Campbell, J., Shin-ya, K., Zheng, L., et al. (2013). Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity. Embo J *32*, 1425–1439.

Lindahl, T. (1974). An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues. Proc. Natl. Acad. Sci. U.S.a. *71*, 3649–3653.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature *362*, 709–715.

Lindahl, T., and Barnes, D.E. (2000). Repair of endogenous DNA damage. Cold Spring Harb Symp Quant Biol *65*, 127–133.

Lindahl, T., and Nyberg, B. (1972). Rate of Depurination of Native Deoxyribonucleic Acid. Biochemistry *11*, 3610–&.

Lujan, S.A., Williams, J.S., Clausen, A.R., Clark, A.B., and Kunkel, T.A. (2013). Ribonucleotides Are Signals for Mismatch Repair of Leading-Strand Replication Errors. Molecular Cell 1–7.

Lynch, H.T., Lynch, P.M., Lanspa, S.J., Snyder, C.L., Lynch, J.F., and Boland, C.R. (2009). Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. Clin. Genet. *76*, 1–18.

Mahaney, B.L., Meek, K., and Lees-Miller, S.P. (2009). Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. Biochem J *417*, 639–650.

Matsuoka, S., Ballif, B.A., Smogorzewska, A., McDonald, E.R., Hurov, K.E., Luo, J., Bakalarski, C.E., Zhao, Z., Solimini, N., Lerenthal, Y., et al. (2007). ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science *316*, 1160–1166.

McElhinny, S.A.N., Kissling, G.E., and Kunkel, T.A. (2010). Differential correction of lagging-strand replication errors made by DNA polymerases {alpha} and {delta}. Proc. Natl. Acad. Sci. U.S.a. *107*, 21070–21075.

Mourón, S., Rodriguez-Acebes, S., Martínez-Jiménez, M.I., García-Gómez, S., Chocrón, S., Blanco, L., and Méndez, J. (2013). Repriming of DNA synthesis at stalled replication forks by human PrimPol. Nature Structural & Molecular Biology *20*, 1383–1389.

Muzi-Falconi, M., Giannattasio, M., Foiani, M., and Plevani, P. (2003). The DNA polymerase alpha-primase complex: multiple functions and interactions. ScientificWorldJournal *3*, 21–33.

Nick McElhinny, S.A., Kumar, D., Clark, A.B., Watt, D.L., Watts, B.E., Lundström, E.-B., Johansson, E., Chabes, A., and Kunkel, T.A. (2010). Genome instability due to ribonucleotide incorporation into DNA. Nat. Chem. Biol. *6*, 774–781.

O'Brien, P.J., Siraki, A.G., and Shangari, N. (2005). Aldehyde sources, metabolism, molecular toxicity mechanisms, and possible effects on human health. Crit Rev Toxicol *35*, 609–662.

Pascucci, B., Stucki, M., Jónsson, Z.O., Dogliotti, E., and Hübscher, U. (1999). Long patch base excision repair with purified human proteins. DNA ligase I as patch size mediator for DNA polymerases delta and epsilon. J. Biol. Chem. *274*, 33696–33702.

Peña-Diaz, J., and Jiricny, J. (2012). Mammalian mismatch repair: error-free or error-prone? Trends in Biochemical Sciences *37*, 206–214.

Piazza, A., Serero, A., Boulé, J.-B., Legoix-Né, P., Lopes, J., and Nicolas, A. (2012). Stimulation of Gross Chromosomal Rearrangements by the Human CEB1 and CEB25 Minisatellites in Saccharomyces cerevisiae Depends on G-Quadruplexes or Cdc13. PLoS Genet. *8*, e1003033.

Pines, A., Mullenders, L.H., van Attikum, H., and Luijsterburg, M.S. (2013). Touching base with PARPs: moonlighting in the repair of UV lesions and double-strand breaks. Trends in Biochemical Sciences *38*, 321–330.

Prasad, R., Longley, M.J., Sharief, F.S., Hou, E.W., Copeland, W.C., and Wilson, S.H. (2009). Human DNA polymerase theta possesses 5'-dRP lyase activity and functions in single-nucleotide base excision repair in vitro. Nucleic Acids Res. *37*, 1868–1877.

Ribeyre, C., Lopes, J., Boulé, J.-B., Piazza, A., Guédin, A., Zakian, V.A., Mergny, J.-L., and Nicolas, A. (2009). The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo. PLoS Genet. *5*, e1000475–e1000475.

Robertson, A.B., Klungland, A., Rognes, T., and Leiros, I. (2009). DNA repair in mammalian cells: Base excision repair: the long and short of it. Cell Mol Life Sci *66*, 981–993.

Rodriguez, R., Miller, K.M., Forment, J.V., Bradshaw, C.R., Nikan, M., Britton, S., Oelschlaegel, T., Xhemalce, B., Balasubramanian, S., and Jackson, S.P. (2012). Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. Nat. Chem. Biol. *8*, 301–310.

Rosado, I.V., Langevin, F., Crossan, G.P., Takata, M., and Patel, K.J. (2011). Formaldehyde catabolism is essential in cells deficient for the Fanconi anemia DNA-repair pathway. Nature Structural & Molecular Biology *18*, 1432–1434.

Sale, J.E., Lehmann, A.R., and Woodgate, R. (2012). Y-family DNA polymerases and their role in tolerance of cellular DNA damage. Nature Publishing Group *13*, 141–152.

Sarkies, P., Murat, P., Phillips, L.G., Patel, K.J., Balasubramanian, S., and Sale, J.E. (2012). FANCJ coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. Nucleic Acids Res. *40*, 1485–1498.

Sarkies, P., Reams, C., Simpson, L.J., and Sale, J.E. (2010). Epigenetic instability due to defective replication of structured DNA. Molecular Cell *40*, 703–713.

Schwab, R.A., Nieminuszczy, J., Shin-ya, K., and Niedzwiedz, W. (2013). FANCJ couples replication past natural fork barriers with maintenance of chromatin structure. J. Cell Biol. *201*, 33–48.

Seki, M., Marini, F., and Wood, R.D. (2003). POLQ (Pol θ), a DNA polymerase and DNA-dependent ATPase in human cells. Nucleic Acids Res. *31*, 6117–6126.

Seki, M., Masutani, C., Yang, L.W., Schuffert, A., Iwai, S., Bahar, I., and Wood, R.D. (2004). High-efficiency bypass of DNA damage by human DNA polymerase Q. Embo J *23*, 4484–4494.

Shiloh, Y. (2003). ATM and related protein kinases: safeguarding genome integrity. Nat Rev Cancer *3*, 155–168.

Shima, N., Munroe, R.J., and Schimenti, J.C. (2004). The mouse genomic instability mutation chaos1 is an allele of Polq that exhibits genetic interaction with Atm. Mol. Cell. Biol. *24*, 10381–10389.

Siegel, E.C., and Bryson, V. (1963). Selection of resistant strains of Escherichia coli by antibiotics and antibacterial agents: role of normal and mutator strains. Antimicrob Agents Chemother (Bethesda) *161*, 629–634.

Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and mechanism of helicases and nucleic acid translocases. Annu. Rev. Biochem. *76*, 23–50.

Sommers, J.A., Banerjee, T., Hinds, T., Wan, B., Wold, M.S., Lei, M., and Brosh, R.M. (2014). Novel Function of the Fanconi Anemia Group J or RECQ1 Helicase to Disrupt Protein-DNA Complexes in a Replication Protein A-stimulated Manner. Journal of Biological Chemistry *289*, 19928–19941.

Suhasini, A.N., Sommers, J.A., Muniandy, P.A., Coulombe, Y., Cantor, S.B., Masson, J.-Y., Seidman, M.M., and Brosh, R.M. (2013). Fanconi Anemia Group J Helicase and MRE11 Nuclease Interact To Facilitate the DNA Damage Response. Mol. Cell. Biol. *33*, 2212–2227.

Sun, H.H., Karow, J.K.J., Hickson, I.D.I., and Maizels, N.N. (1998). The Bloom's syndrome helicase unwinds G4 DNA. J. Biol. Chem. *273*, 27587–27592.

Symington, L.S. (2014). DNA repair: Making the cut. Nature *514*, 39–40.

Takata, K.I., Shimizu, T., Iwai, S., and Wood, R.D. (2006). Human DNA Polymerase N (POLN) Is a Low Fidelity Enzyme Capable of Error-free Bypass of 5S-Thymine Glycol. Journal of Biological Chemistry *281*, 23445–23455.

Thompson, O., Edgley, M., Strasbourger, P., Flibotte, S., Ewing, B., Adair, R., Au, V., Chaudhry, I., Fernando, L., Hutter, H., et al. (2013). The million mutation project: a new approach to genetics in Caenorhabditis elegans. Genome Res. *23*, 1749–1762.

Uchiyama, Y., Takeuchi, R., Kodera, H., and Sakaguchi, K. (2009). Distribution and roles of X-family DNA polymerases in eukaryotes. Biochimie *91*, 165–170.

Umate, P., Tuteja, N., and Tuteja, R. (2011). Genome-wide comprehensive analysis of human helicases. Commun Integr Biol *4*, 118–137.

Valton, A.-L., Hassan-Zadeh, V., Lema, I., Boggetto, N., Alberti, P., Saintomé, C., Riou, J.-F., and Prioleau, M.-N. (2014). G4 motifs affect origin positioning and efficiency in two vertebrate replicators. Embo J *33*, 732–746.

van der Lelij, P., Chrzanowska, K.H., Godthelp, B.C., Rooimans, M.A., Oostra, A.B., Stumm, M., Zdzienicka, M.Z., Joenje, H., and de Winter, J.P. (2010). Warsaw Breakage Syndrome, a Cohesinopathy Associated with Mutations in the XPD Helicase Family Member DDX11/ChlR1. The American Journal of Human Genetics *86*, 262–266.

Vannier, J.B., Sandhu, S., Petalcorin, M.I., Wu, X., Nabi, Z., Ding, H., and Boulton, S.J. (2013). RTEL1 Is a Replisome-Associated Helicase That Promotes Telomere and Genome-Wide Replication. Science *342*, 239–242.

Vannier, J.-B., Pavicic-Kaltenbrunner, V., Petalcorin, M.I.R., Ding, H., and Boulton, S.J. (2012). RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. Cell *149*, 795–806.

Waaijers, S., Portegijs, V., Kerver, J., Lemmens, B.B.L.G., Tijsterman, M., van den Heuvel, S., and Boxem, M. (2013). CRISPR/Cas9-Targeted Mutagenesis in Caenorhabditis elegans. Genetics *195*, 1187–1191.

Wang, M., Wu, W., Wu, W., Rosidi, B., Zhang, L., Wang, H., and Iliakis, G. (2006). PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. Nucleic Acids Res. *34*, 6170–6182.

White, R.M., Sessa, A., Burke, C., Bowman, T., LeBlanc, J., Ceol, C., Bourque, C., Dovey, M., Goessling, W., Burns, C.E., et al. (2008). Transparent adult zebrafish as a tool for in vivo transplantation analysis. Cell Stem Cell *2*, 183–189.

Wienholds, E., van Eeden, F., Kosters, M., Mudde, J., Plasterk, R.H.A., and Cuppen, E. (2003). Efficient target-selected mutagenesis in zebrafish. Genome Res. *13*, 2700–2707.

Wilson, M.A., Kwon, Y., Xu, Y., Chung, W.-H., Chi, P., Niu, H., Mayle, R., Chen, X., Malkova, A., Sung, P., et al. (2013). Pif1 helicase and Polδ promote recombination-coupled DNA synthesis via bubble migration. Nature *502*, 393–396.

Wogan, G.N., Hecht, S.S., Felton, J.S., Conney, A.H., and Loeb, L.A. (2004). Environmental and chemical carcinogenesis. Semin Cancer Biol *14*, 473–486.

Wolfe, A.L., Singh, K., Zhong, Y., Drewe, P., Rajasekhar, V.K., Sanghvi, V.R., Mavrakis, K.J., Jiang, M., Roderick, J.E., Van der Meulen, J., et al. (2014). RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. Nature *513*, 65–70.

Wood, A.J., Lo, T.-W., Zeitler, B., Pickle, C.S., Ralston, E.J., Lee, A.H., Amora, R., Miller, J.C., Leung, E., Meng, X., et al. (2011). Targeted genome editing across species using ZFNs and TALENs. Science *333*, 307.

Wu, Y.Y., and Brosh, R.M.R. (2012). DNA helicase and helicase-nuclease enzymes with a conserved iron-sulfur cluster. Nucleic Acids Res. *40*, 4247–4260.

Yamtich, J., and Sweasy, J.B. (2010). DNA polymerase Family X: Function, structure, and cellular roles. Biochimica Et Biophysica Acta (BBA) - Proteins and Proteomics *1804*, 1136–1150.

Yoon, J.-H., Roy Choudhury, J., Park, J., Prakash, S., and Prakash, L. (2014). A role for DNA polymerase θ in promoting replication through oxidative DNA lesion, thymine glycol, in human cells. Journal of Biological Chemistry *289*, 13177–13185.

Yoshimura, M., Kohzaki, M., Nakamura, J., Asagoshi, K., Sonoda, E., Hou, E., Prasad, R., Wilson, S.H., Tano, K., Yasui, A., et al. (2006). Vertebrate POLQ and POLbeta cooperate in base excision repair of oxidative DNA damage. Molecular Cell *24*, 115–125.

Youds, J.L., Barber, L.J., Ward, J.D., Collis, S.J., O'Neil, N.J., Boulton, S.J., and Rose, A.M. (2008). DOG-1 is the Caenorhabditis elegans BRIP1/FANCJ homologue and functions in interstrand cross-link repair. Mol. Cell. Biol. *28*, 1470–1479.

Yousefzadeh, M.J., Wyatt, D.W., Takata, K.-I., Mu, Y., Hensley, S.C., Tomida, J., Bylund, G.O., Doublié, S., Johansson, E., Ramsden, D.A., et al. (2014). Mechanism of suppression of chromosomal instability by DNA polymerase POLQ. PLoS Genet. *10*, e1004654–e1004654.

◀ This photo shows the head of a stained worm (*C. elegans*). The blue staining in a cell reports the presence of a mutation in which a G-quadruplex (an unusually folded piece of DNA) and surrounding DNA was deleted. In this photo the blue cells mark the pharynx (foregut) of the worm.