Natural and Synthetic Strategies for DNA Maintenance and Recording

Natural and synthetic strategies for DNA maintenance and recording

A dissertation presented

by

Nava Gharaei

To

The Department of Molecular and Cellular Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biochemistry

Harvard University

Cambridge, Massachusetts

December 2018

Table of Contents

List of figures v
Abstract vii
Chapter 1. DNA maintenance 1
Introduction 2
Genome maintenance 3
Core DNA repair pathways 7
Base excision repair 7
Nucleotide excision repair 12
Mismatch repair 16
Double strand DNA break repair 20
Complex DNA repair pathways 27
DNA Interstrand crosslink repair 28
DNA-protein crosslink repair 52
Clustered DNA damage 57
Discussion 59
Chapter 2. Bio-molecular recording and bio-computation 63
Introduction 64
Writing on the DNA of living cells and its applications 67
Template-dependent molecular recording 73
Retroelement-based DNA writers 73
SCRIBE 74
CRISPR-based DNA writers 77
CRISPR spacer acquisition 77
Cas9-based systems 83
Template-independent molecular recording 86
Site-specific recombinases as molecular recorders 86
CRISPR/Cas9-based molecular recorders 89
CRISPR/Cas9-based evolving barcodes 92
Base editors 95
CRISPR/Cas9-based base editors 95
RNA Polymerase-base editor fusions 99

Polymerase-based DNA writers 102
CRISPR/Cas9 - DNA polymerase fusion 102
Polymerase-based ticker tapes 104
Discussion 106
Bibliography 109

List of figures

Figure 1.1 Different types of DNA damage and their repair pathways 5

Figure 1.2 Base excision repair (BER) mechanism 10

Figure 1.3 Nucleotide excision repair (NER) mechanism 14

Figure 1.4 Mismatch repair (MMR) mechanism 17

Figure 1.5 The general strategy for single strand DNA damage repair 19

Figure 1.6 Strategies for double stand break repair 21

Figure 1.7 Double strand break repair mechanisms 25

Figure 1.8 DNA interstrand crosslinks (ICLs) 29

Figure 1.9 DNA interstrand crosslinks (ICLs) are detected and repaired by different mechanisms during different phases of cell cycle 30

Figure 1.10 Replication-dependent interstrand crosslink (ICL) repair mechanism 35

Figure 1.11 Replication-independent interstrand crosslink (ICL) repair mechanism 50

Figure 1.12 DNA-protein crosslink (DPC) repair mechanism 55

Figure 1.13 Other functions and applications of DNA repair pathways 60

Figure 2.1 The interplay between DNA writers and DNA repair mechanisms 66

Figure 2.2 Using genomic DNA as a medium for storing biological information 69

Figure 2.3 Using genomic DNA for information processing and computation in living cells 70

Figure 2.4 Classes of DNA writers based on their mode of action 72

Figure 2.5 Mechanism of DNA writing by SCRIBE (an example of retroelement-based DNA writers) 76

Figure 2.6 DNA writing by Cas1-Cas2 spacer acquisition system 79

Figure 2.7 DNA writing by RT-Cas1-Cas2 spacer acquisition system 81

Figure 2.8 Retroelement- and Cas9-mediated DNA writing by CRISPEY 85

Figure 2.9 Template-independent DNA writing by site-specific recombinases 87

Figure 2.10 Cas9-mediated DNA writing 90
Figure 2.11 Tracing cellular lineage via stochastic and template-independent CRISPR/Cas9 DNA writers 91
Figure 2.12 Extended template-independent molecular recording by CRISPR/Cas9 mediated evolving barcodes 94
Figure 2.13 CRSISPR/Cas9-based base editors 96
Figure 2.14 DNA writing by RNA polymerase-base editor fusion 101
Figure 2.15 Template-independent DNA writing by DNA polymerases via EvolvR system 103
Figure 2.16 DNA polymerase-mediated molecular ticker tape 105

Natural and synthetic strategies for DNA maintenance and recording

Abstract

As the sole guardian of the necessary knowledge for the creation of life (as we know), the genomic DNA is carefully protected by extremely diligent repair pathways. Highlighting the importance of these repair processes are thousands of genetic diseases that are caused by just a single nucleotide mutation (in genomes that are made of billions of nucleotides). On the other hand, adaptation, which is at the very core of the concept of life, relies on the ability of living cells to incorporate value-added information to the genome. Thus, mutations (which provide substrates for adaptation) are also essential and permitted to occur naturally. The dynamic interplay between the DNA damaging/modifying agents/enzymes and the DNA repair machinery determines the editing outcome of the book of life. Therefore, genomic DNA can be considered as a biological hard drive in living cells, where information can be added, saved or erased by two competing processes. The continued quest to better understand the mechanisms underlying cellular DNA repair processes on one hand, and the technological advances that allow us to dynamically write arbitrary information into the genome, on the other hand, are the keys to understanding and controlling the flow of information into this biological hard drive. Here, I first present our current outlook to the genome maintenance and DNA repair processes. I then discuss various DNA writing strategies. In addition, I will outline how the interplay between DNA repair machinery and these technologies can be leveraged to both dynamically and precisely control the flow of information into the genomic DNA, as well as better understand the underlying mechanisms of DNA repair pathways. The synergic advances in these areas could ultimately let us gain control over the book of life, with applications ranging from the study of human biology and diseases to programming cellular phenotypes.

Chapter 1. DNA maintenance

Introduction

DNA, the genetic material of almost every living system, holds the information necessary for making the living from mere atoms. Having the most difficult task of guarding the essential biological information, DNA was once assumed to be very stable. However, in the 1960s, the natural vulnerability of DNA was discovered (Lindahl 2016). Later it became clear that DNA can be damaged by many exogenous (e.g., chemicals, radiation, certain drugs, etc.), as well as endogenous (e.g., normal byproducts of cellular processes, natural decay of DNA, errors in DNA processes, etc.) factors; even DNA replication, itself, can cause DNA damage. To prevent these inevitable damages from leading to loss or alteration of the genetic information, intricate strategies for sensing and repairing them have been evolved. Without these strategies life as we know it would not have been possible.

DNA repair strategies require the activation and recruitment of a complex system, where the exact timing of activation and recruitment of the players are critical in determining the outcome of the repair. Any defects in this system can lead to devastating diseases such as Alzheimer's, Parkinson's and cancer. Despite their crucial role, the details of many of these repair pathways have remained largely unknown. Moreover, the variety of DNA lesions and the sophistication of their dedicated repair pathways suggest that there are many repair systems that remain uncovered. Indeed, more and more complicated DNA lesions with their own dedicated repair strategies are being discovered. Uncovering the details of these strategies can pave the way to new therapeutics that can help improve or boost the repair system in cases where they become defective.

In this chapter, I will discuss our current knowledge of the DNA repair pathways, the implications of our discoveries and the remaining questions in the field, beginning with the molecular mechanism of the five core DNA repair pathways: base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), non-homologous end joining (NHEJ)

and homologous recombination (HR). Then, I will discuss how these core repair pathways work together to repair more complex DNA lesions, such as DNA interstrand crosslinks (ICLs), DNA protein crosslinks (DPCs) and clustered DNA damage (CDD). While many aspects of these repair pathways are conserved, prokaryotic and eukaryotic repair systems can be distinct. As the eukaryotic versions are usually more comprehensive and at the same time more relevant to human health, in this chapter, I will focus on the eukaryotic DNA repair systems.

Genome maintenance

Our DNA is continuously under attack by various endogenous and exogenous factors with the estimated rate of 200,000 events per cell per day (Barnes and Lindahl 2004; Ciccia and Elledge 2010). Even DNA replication can introduce DNA lesions (Swenberg et al. 2011; Hoeijmakers 2001). This is mostly due to the random incorporations of mismatched bases or ribonucleotides (instead of deoxy-ribonucleotides) by the DNA polymerases. In addition, DNA polymerases become more error-prone when replicating repeats, especially short tandem repeats. These errors can expand or shorten the repeats with some leading to disorders such as Huntington's and ALS (Amyotrophic Lateral Sclerosis)[1] (Jones, Houlden, and Tabrizi 2017). Another endogenous source of DNA damage is the natural decay of DNA (Lindahl 1993). For example, spontaneous depurination/depyrimidination and cytosine deamination happen with estimated rates of 10,000 and 100 - 500 per cell per day, respectively (Ciccia and Elledge 2010). On the other hand, methylation and oxidation of DNA by byproducts of cellular reactions lead to approximately, 40,000 and 400-1500 lesions per cell per day, respectively (Ciccia and Elledge 2010). If not faithfully repaired, these lesions can lead to mutations which in turn can cause, cell death, cancer, or any of the other thousands of genetic disorders.

[1] As these disorders are caused by trinucleotide repeat expansions, they have been termed "the trinucleotide repeat disorders" (Jones, Houlden, and Tabrizi 2017).

Besides endogenous factors, DNA lesions can be caused by exogenous factors such as UV radiation, ionizing radiation, and chemicals (Wang and Lindahl 2016). With the estimated rate of more than 4000 lesions per cell per hour in peak sunlight, UV radiation is the most common mutagen and the major risk factor for skin cancer (Ciccia and Elledge 2010). The most prominent DNA damage by UV radiation is the covalent attachment of adjacent pyrimidine bases which can create 6-4 pyrimidine pyrimidone photoproducts (6-4PPs) and cyclobutane-pyrimidine dimers (CPDs). By distorting the DNA helix and preventing proper base pairing, these intrastrand crosslinks can interfere with various DNA functions, as well as introduce mutations during replication. On the other hand, ionizing radiations[2], such as X rays and gamma rays, have much higher energy levels and can introduce various types of DNA damage, including single and double strand breaks. Among these lesions, double strand breaks (DSBs) are the most deleterious. If not repaired in a timely fashion, DSBs can result in deletions, insertions, duplications, inversions, and translocations of small or large portions of the DNA.[3] Finally, chemical factors, such as base analogs, intercalating agents, reactive oxygen and nitrogen species, and alkylating factors usually act by modifying bases.[4] These modifications can then lead to base changes, formation of abasic sites or attachments of bulky groups to the DNA.

Importantly, DNA lesions cause structural and chemical alterations to the DNA with different levels of severity. If they persist in the DNA, they can interfere with essential cellular functions (e.g., DNA replication and transcription) and lead to mutations. To avoid these harmful effects,

[2] They can be from a variety of sources such as nuclear reactions, radioactive material, solar flares, medical procedures and consumer products.

[3] The estimated number of DSBs (per cell per day) created by the Hiroshima and Nagasaki atomic bombs, space missions, airline travels and CT scans (computed tomography scan) are 0.2 – 160, 2, 0.0028, 0.28 respectively (Ciccia and Elledge 2010).

[4] They can be from a variety of sources including laboratory reagents, chemical weapons, peanuts and food borne fungi. In addition, reactive oxygen and nitrogen species and alkylating factors can also be produced as byproducts of cellular metabolism or stress.

cells have evolved different strategies for maintaining genome integrity (Figure 1.1). For example, DNA polymerases have an estimated error rate of 1 in 10^6 bases. However, due to their incorrect base pairing, these misincorporations are often recognized and repaired quickly by the mismatch repair pathway (discussed in detail below). This improves the error rate of DNA replication by more than 1000 fold. Besides this pathway, there are many other pathways that have evolved to repair DNA lesions. In the cells, they work together to maintain the genome integrity.

Figure 1.1 Different types of DNA damage and their repair pathways

Different mutagens (from exogenous or endogenous sources) can cause different types of damage in the DNA. These damages can be identified and repaired by different repair pathways. The more common DNA lesions have dedicated repair pathways which constitute the core repair pathways. The less common DNA lesions need more complex repair strategies which employ several core repair pathways. NTP (ribonucleotide triphosphate), DSB (double strand break), SSB (single strand break), BER (base excision repair), NER (nucleotide excision repair), MMR (mismatch repair), NHEJ (non-homologous end joining), HR (homologous recombination), TLS (translesion synthesis), FA (Fanconi anemia pathway).

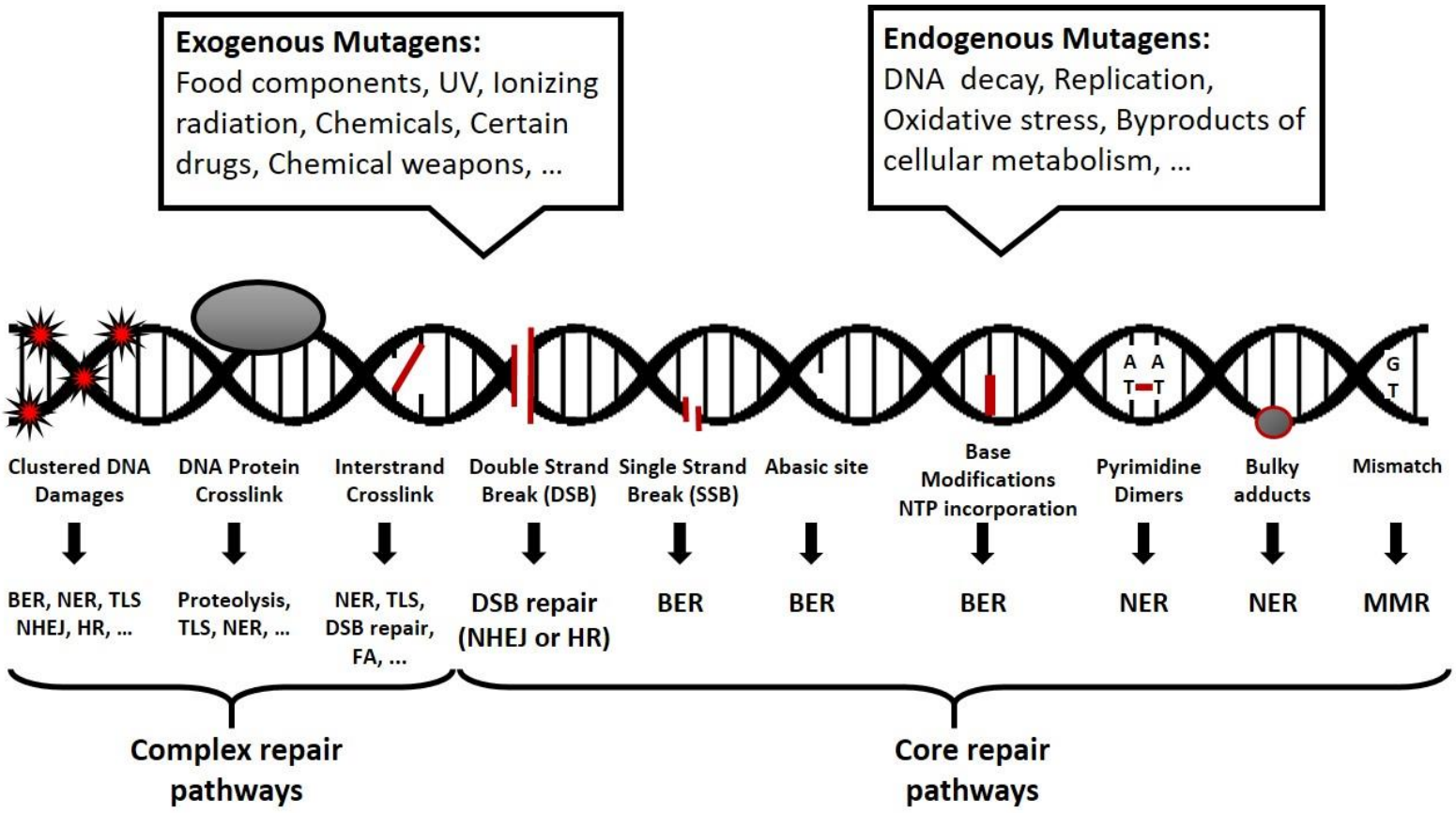

Figure 1.1 (Continued)

Core DNA repair pathways

Although there are many different types of DNA damage, the most common forms are ribonucleotide incorporations, base mismatches, base modifications (including attachments of small or bulky adducts to the base), abasic sites and DNA breaks (including single stranded and double stranded breaks). These common DNA lesions have specialized repair mechanisms which constitute the core DNA repair pathways. The core repair pathways have been studied extensively and their molecular mechanisms are very well understood. In this section, I will discuss our current knowledge of these repair pathways.

Base excision repair

First identified by Tomas Lindahl in 1974, base excision repair (BER), which is thought to be essential, protects us against premature aging, cancer and neurodegeneration (Bernstein and Bernstein 2015; Do et al. 2014; Farkas et al. 2014; Bayraktar and Kreutz 2018). BER is a well-conserved mechanism (found in all three domains of life) for the repair of small non-distorting base damages usually caused by oxidation, deamination or alkylation of a base (Schermerhorn and Delaney 2014). These damages comprise the majority of endogenous DNA damage (caused for example by natural DNA decay) (Lindahl 1993). However, they can also be caused by exogenous factors such as chemicals, drugs, and radiation (Wallace 2014).

As it can be inferred from its name, base excision repair removes damaged bases from the DNA.[5] In BER, substrate-specific DNA glycosylases search the DNA for these hidden damaged bases and cleave them from the backbone (Friedman and Stivers 2010). In humans, there are 11 known glycosylases each identifying a few related damages (and some have overlapping specificities). Based on their substrate, glycosylases can be divided into three groups: 1)

[5] The downstream steps of BER can also repair single strand DNA breaks.

glycosylases that recognize oxidative damages (six glycosylases), 2) glycosylases that identify mispaired uracils and thymines (four glycosylases), and 3) glycosylases that recognize alkylated bases (one glycosylase) (Brooks et al. 2013; Balliano and Hayes 2015). After finding a specific damage, it is thought that the dedicated glycosylase flips the base out of the DNA double helix[6] and cuts the N-glycosidic bond that attaches it to the backbone (Figure 1.2.1). This results in the production of an apurinic/apyrimidinic (AP) site (Slupphaug et al. 1996; Huffman, Sundheim, and Tainer 2005).

Glycosylases have also been categorized based on their mechanism of action into two groups of monofunctional (which includes the glycosylases that identify alkylated bases or mispaired uracils and thymines) and bifunctional (which includes the glycosylases that are dedicated to identifying oxidative damages) (Dalhus et al. 2009). While monofunctional glycosylases leave the DNA sugar-phosphate backbone intact, bifunctional glycosylases cut the backbone at the AP site (by their AP lyase activity, leaving behind the remainder of the AP site at the nick) (Stivers and Jiang 2003). After the creation of an AP site by monofunctional glycosylases, AP endonuclease 1 (APE1) hydrolyses the DNA backbone at the AP site, creating an entry point for the DNA polymerases (Figure 1.2.2). In the case of bifunctional glycosylases, β-elimination or βδ-elimination of the AP site (by the glycosylase β-lyase or βδ-lyase activities) result in a nick in the backbone. While β-elimination leaves a 3' unsaturated hydroxyaldehyde, which has to be further processed by APE1, βδ-elimination only leaves a 3' phosphate, which can be removed by polynucleotide kinase circumventing a need for APE1[7] (Fromme, Banerjee, and Verdine 2004; Svilar et al. 2011; Dianov and Hubscher 2013).

[6] It has also been suggested that rather than actively flipping the damaged bases, glycosylases catch them when they stochastically flip (Stivers 2008; Cao et al. 2004).

[7] Although APE1 is not needed, it can also remove this remaining phosphate.

The subsequent steps, which include synthesis and ligation, have been compared to passing of a baton and could be either through short patch repair (which only replaces the damaged nucleotide) or long patch repair (which leads to the replacement of 2 to 12 nucleotides) (Figure 1.2.3) (Krokan and Bjoras 2013; Jacobs and Schar 2012; Dianov and Hubscher 2013). In short patch repair, polymerase β (Pol β), which is a repair DNA polymerase, both removes the leftover sugar from the AP nucleotide[8] (still attached to the 5' of the nick), and fills in the one nucleotide gap that is created (Figure 1.2.4a) (Robertson et al. 2009; Svilar et al. 2011; Hegde, Hazra, and Mitra 2008). The short patch repair can also be redirected to the long patch repair, usually as a result of Pol β not being able to efficiently remove the remaining sugar. In long patch repair, replicative polymerases δ/ε take over and displace the DNA strand that contains the damage (Figure 1.2.4b). This flap is then removed by the flap endonuclease 1 (FEN1) (Figure 1.2.5b) (Liu et al. 2005; Sun et al. 2017). The remaining nick in the DNA is then ligated by DNA ligase I in long patch repair, or DNA ligase IIIα/XRCC1 (X-ray repair cross-complementing protein 1) in both short and long patch repair pathways (Figure 1.2.5a and 1.2.6b) (Parsons et al. 2005; Fan and Wilson 2005).

Although the short patch repair is usually the dominant pathway, the decision between short and long patch repair pathways can depend on the type of damage, the organism, cell type, the cell cycle phase and the accessibility of BER factors (Fortini and Dogliotti 2007; Gellon et al. 2008). For example, some DNA damage such as oxidized or reduced AP sites cannot be removed by Pol β, and therefore, have to be redirected to the long patch repair pathway. Moreover, while the short patch repair can be efficiently performed in both proliferating and nonproliferating cells, since long patch repair depends on replicative factors such as pol δ/ε and FEN1, it can only be performed in proliferating cells (Akbari et al. 2009).

[8] In the case of bifunctional glycosylases, this lyase activity is not required.

Figure 1.2 Base excision repair (BER) mechanism

1) BER, which removes small non-distorting base damages, is initiated when a substrate specific glycosylase finds the damage in the DNA. After detection, the glycosylase cleaves out the damaged base creating an apurinic/apyrimidinic (AP) site. 2) The APE1 nuclease cleaves the DNA backbone at the AP site. 3) This incision creates an entry point for DNA polymerases. The remainder of the repair could be performed either by: a) short patch repair (which only replaces a single nucleotide), or b) long patch repair (which replaces 2-12 nucleotides). 4a) In short patch repair, the repair polymerase, Pol β, loads from the nick, and 5a) both removes the leftover sugar from the AP site and fills in the one nucleotide gap created by this removal. 6a) Ligation of the remaining nick completes the repair process. 4b) In long patch repair pathway, the replicative polymerases, pol δ/ε, and their sliding clamp PCNA are loaded onto the DNA from the nick that APE1 has created. 5b) Pol δ/ε both displace the damaged strand and resynthesize the DNA. 6b) The displaced strand is then removed by FEN1. 7b) As in short patch repair, ligation of the remaining nick completes the repair process.

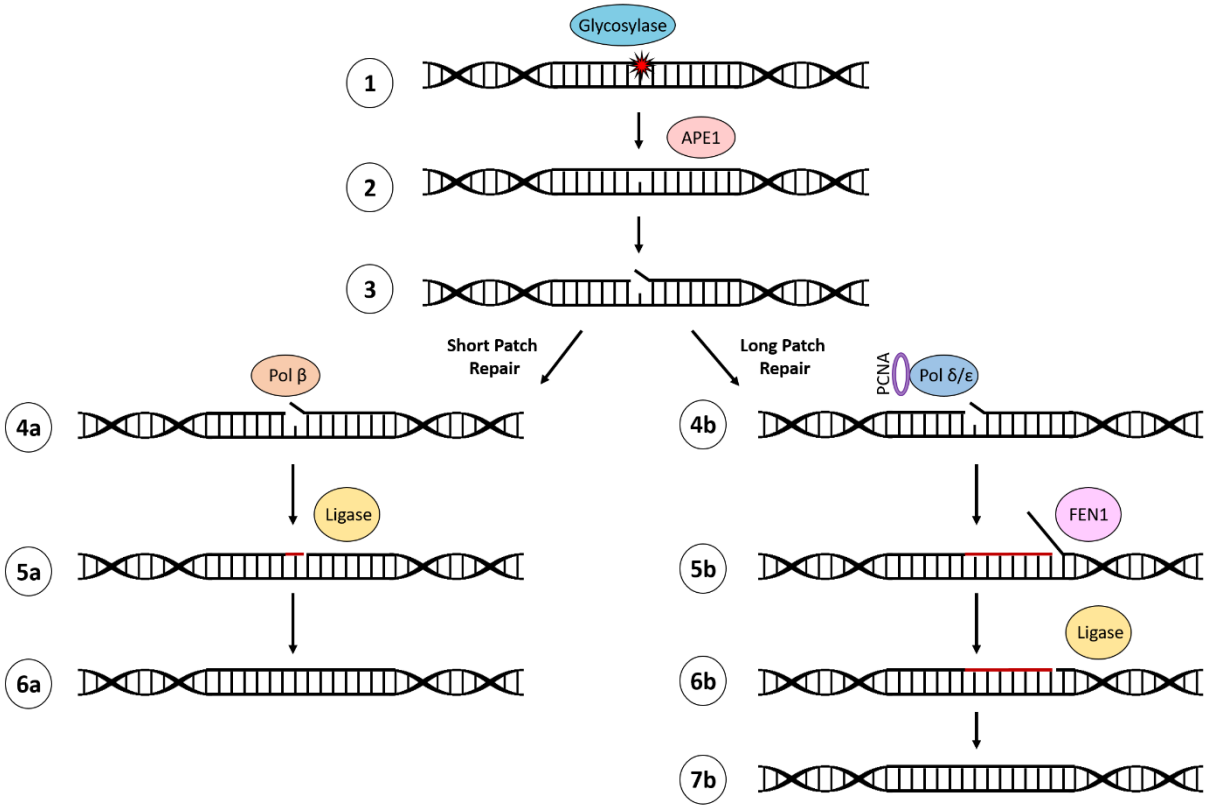

Figure 1.2 (Continued)

Nucleotide excision repair

Discovered in 1964 by the Setlow, Howard-Flanders, Painter and Hanawalt groups, nucleotide excision repair (NER) mends distorting DNA damage (such as those generated by UV radiation) (Setlow and Carrier 1964; Boyce and Howard-Flanders 1964; Pettijohn and Hanawalt 1964; Rasmussen and Painter 1964). As can be inferred from the severe conditions caused by the defects in NER (e.g., Xeroderma Pigmentosum, Cockayne's Syndrome and Trichothiodystrophy[9]), this pathway is extremely important in genome maintenance. NER is a complex pathway employing more than 30 proteins among which are 9 major proteins: XPA, XPB, XPC, XPD, XPE, XPF, XPG (all of which cause Xeroderma Pigmentosum, if mutated, and are named after this disorder), CSA and CSB (mutations of the CS proteins cause Cockayne's Syndrome, and are named after this disorder) (DiGiovanna and Kraemer 2012; Lehmann 2003; Gillet and Scharer 2006). After the discovery that NER repairs the damages on the transcribed strands with a higher rate, NER was divided into two subpathways (which only differ in their damage detection capabilities): global genome NER (GG-NER), which can identify lesions anywhere in the genome, and transcription-coupled NER (TC-NER), which can identify the lesions in the transcribed strands (Figure 1.3.1) (Spivak and Ganesan 2014; Gillet and Scharer 2006; Spivak 2015).

In GG-NER, the DNA distortion caused by damage is recognized by the XPC complex (XPC, RAD23B, and centrin2) which binds to the opposite strand of the damage (Figure 1.3.2). By avoiding to bind directly to the damage, XPC significantly increases its capacity for recognizing diverse DNA lesions (Lee et al. 2014; Maillard, Solyom, and Naegeli 2007; Min and Pavletich 2007). After binding, the XPC complex melts the DNA around the damage and recruits the

[9] These severe conditions include an extremely high risk of cancer and developmental and neurological abnormalities.

downstream factors (i.e. TFIIH complex) (Figure 1.3.3). In cases where the lesion does not induce a significant DNA distortion, the DNA damage binding complex (DDB complex, which comprise of DDB1 and DDB2 (XPE)) binds to the lesion and creates a kink in the DNA that can be identified by XPC (Tang and Chu 2002; Zhu and Wani 2017).

In TC-NER, blockage of a transcribing RNA polymerase due to a lesion initiates the repair (Figure 1.3.2). Upon RNA polymerase stalling, CSB (Cockayne's syndrome group B protein), which is a transcription elongation factor that travels with the RNA polymerase, recruits CSA (Cockayne's syndrome group A protein) and the NER repair factors (i.e. TFIIH complex) (Figure 1.3.3). In this pathway, after completion of the repair, CSA mediated degradation of CSB is necessary for the resumption of the RNA synthesis (Marteijn et al. 2014).

After the initiation step, recruitment of the TFIIH complex (which is comprised of ten proteins) converges the two subpathways (i.e. GG-NER and TC-NER). XPB and XPD, two components of the TFIIH complex, are helicases that in addition to creating a bubble around the lesion, act as a secondary verification step for the presence of the lesion (Ziani et al. 2014; Compe and Egly 2012). They move along the DNA in the 5' – 3' direction (Mathieu et al. 2013; Coin, Oksenych, and Egly 2007; Winkler et al. 2000). If they stall due to a lesion, they recruit RPA, XPA, and XPG, which form the pre-incision complex (Figure 1.3.4) (Mathieu et al. 2013; Marteijn, Hoeijmakers, and Vermeulen 2015; Araujo, Nigg, and Wood 2001). In this complex RPA, which is the single stranded DNA binding protein, forms a polymer on the undamaged strand in the bubble and protects it from nuclease activity (de Laat et al. 1998). XPA recruits XPF-ERCC1 endonuclease which cleaves the 5' of the bubble (Figure 1.3.5) (Sijbers et al. 1996; Orelli et al. 2010; Li, Peterson, et al. 1995). After the 5' incision, XPG makes a cut at the 3' of the bubble[10] (Fagbemi, Orelli, and Scharer 2011; Araujo, Nigg, and Wood 2001; Dunand-Sauthier et al. 2005). These

[10] This is in addition to XPG's role in providing structural support for the TFIIH complex

dual incisions release the 24 - 32 nucleotide oligo that contains the damage and generate an entry point for DNA polymerases (Figure 1.3.6). The replicative DNA polymerases δ/ε are then loaded onto the DNA and fill in the created gap (Figure 1.3.7). Finally, similar to BER, DNA ligase I (only in replicating cells) or DNA ligase IIIα/XRCC1 (in both replicating and nonreplicating cells) ligate the remaining nick, restoring the DNA double helix structure (Figure 1.3.8) (Shivji et al. 1995; Koch et al. 2016).

Figure 1.3 Nucleotide excision repair (NER) mechanism

1) NER removes DNA lesions that either block the transcription machinery (transcription coupled NER) or are helix distorting (global genome NER). 2) In transcription coupled NER, after an RNA polymerase is stalled by the damage, transcription factors CSA and CSB initiate the repair. In global genome NER, the XPC complex (composed of XPC, Centrin2 and RAD23B) recognizes the DNA helix distortion due to the damage and initiates the repair process. 3) After detection of the damage, the TFIIH complex is recruited. Two components of the TFIIH complex, XPB and XPD, are helicases that create a bubble around the damage. 4) XPB and XPD also move along the DNA and by stalling at the site of the damage confirm its presence. This allows the continuation of the repair process by recruiting RPA, XPA and XPG. RPA (small blue circle) is the single stranded binding protein that covers the undamaged strand and protects it from nuclease activity. 5) XPA further recruits the XPF-ERCC1 endonuclease, which creates a cut at the 5’ side of the bubble. XPG completes the excision of the damaged region by cutting the 3’ of the bubble. 6) The replicative polymerases, pol δ/ε, and their sliding clamp, PCNA, are then loaded onto the DNA from the 3’ end and synthesize the excited region using the undamaged strand as template. 7) Ligation of the remaining nick completes the repair process.

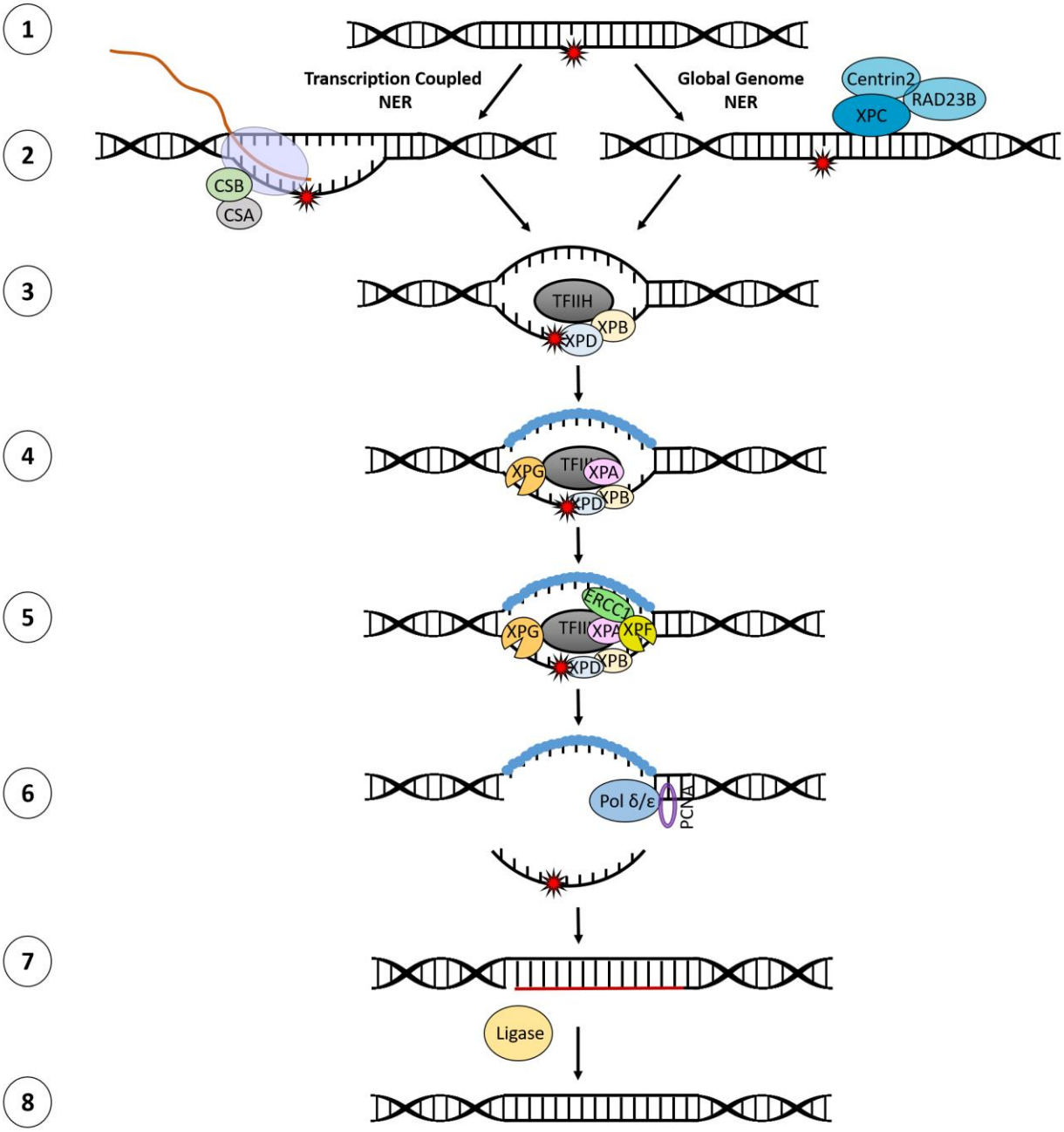

Figure 1.3 (Continued)

Mismatch repair

The third and final pathway for repairing single stranded DNA damage is the highly conserved mismatch repair (MMR) pathway (Groothuizen and Sixma 2016). This pathway, which has a crucial role in the repair of mismatches (as well as small insertions and deletions) introduced by replication and recombination, was proposed as a notion by Robin Holliday in 1964 and discovered by the Meselson group in 1975 (Holliday 1964; Wildenberg and Meselson 1975; Wagner and Meselson 1976).

MMR is especially important for repairing the errors that scape proofreading during DNA replication. In this pathway, MutSα (MSH2-MSH6) or MutSβ (MSH2-MSH3) detect the lesions based on the weakened base pair bond in the mismatch and the slight helix distortion caused by it (Figure 1.4.2) (Dalhus et al. 2009). While MutSα detects 1-2 nucleotide damages (mismatches and small loops), MutSβ can recognize the larger insertion and deletion loops (Kunz, Saito, and Schar 2009). Detection of the lesion recruits MutLα (which is a heterodimer of MLH1 and PMS2) to the site of the damage (Figure 1.4.3).

As in BER and NER, after detection of the damage, MMR degrades and resynthesizes the region that contains the lesion. However, since the damage is a mismatched base or region, rather than a chemical modification, MMR has to decide which one of the strands contains the mutation. Since the mutations are mostly due to replication errors, MMR has been programmed to assume that the nascent strand is the damaged one. MMR distinguishes the nascent strands based on the discontinuities (nicks and gaps) that are present in the newly synthesized DNA (e.g., gaps between the Okazaki fragments or nicks created during the excision of ribonucleotides) (Pavlov, Mian, and Kunkel 2003; Ghodgaonkar et al. 2013). To find these discontinuities, MutSα/β and MutLα translocate along the DNA. If the nick (or gap) is located 5’ of the damage, Exo1 (which is a 5’-3’ exonuclease) degrades the region containing the damage (Figure 1.4.5). DNA polymerase δ/ε then load from the 3’ end and synthesize the DNA using the other strand (Figure

1.4.6) (Jiricny 2006; Tishkoff et al. 1997; Genschel, Bazemore, and Modrich 2002)[11]. However, if the nick (or gap) is located 3' of the damage, DNA polymerase δ/ε are loaded from the nick and degrade the damaged region (through their 3'→5' exonuclease activity), as well as resynthesize it (Tran, Gordenin, and Resnick 1999). If there are not any pre-existing nicks or gaps, MutLα cuts both sides of the damage on the nascent strand based on the position and orientation of RFC, PCNA, and RPA. Then, Exo1 degrades the damaged region and polymerase δ/ε resynthesize it (Jiricny 2006).

Figure 1.4 Mismatch repair (MMR) mechanism

1) MMR removes base mismatches, as well as insertions and deletions. 2) MutSα (composed of MSH2 and MSH6) detects the damage based on the helix distortion and weakened base pairing at the site. 3) After detection of the damage MutLα (composed of MLH1 and PMS2) is recruited. 4) If there are not any pre-existing nicks or gaps in the DNA, MutLα nicks the DNA creating an entry point for Exo1. 5) While Exo1 degrades the damaged region, RPA (small blue circle) protects the single stranded DNA that has been created. 6) The replicative polymerases, pol δ/ε, and their sliding clamp PCNA are loaded onto the DNA from the 3' end and synthesize the excited region. 7) Ligation of the remaining nick completes the repair process.

[11] It is also possible for the DNA polymerase δ/ε to load from the 5' nick and displace, as well as resynthesize the damaged region (Kadyrov et al. 2009).

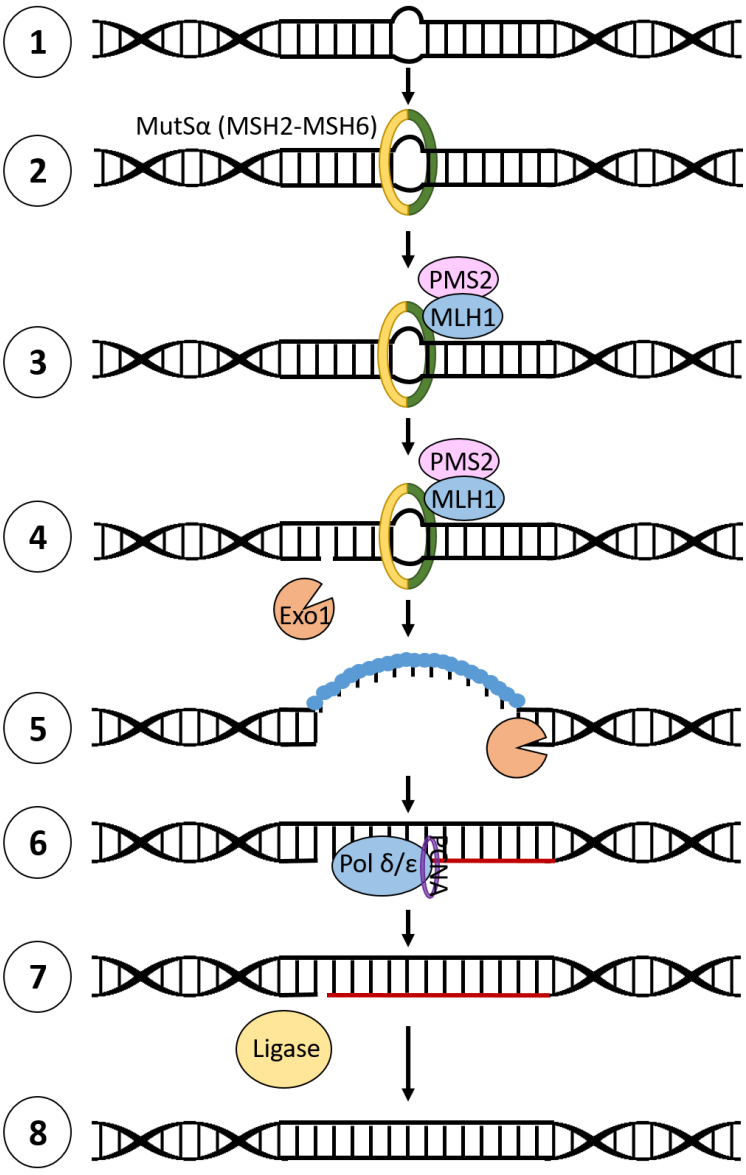

Figure 1.4 (Continue4)

Although for the most part different players are involved in detection and repair of single strand damages, the general repair strategy is shared between BER, NER and MMR. In this strategy, the damaged region, which is recognized based on helix distortion, is excised from the DNA. The undamaged strand is then used as template for resynthesis of the excised region. The final step, which is the ligation of the remaining nick by DNA ligases restores the DNA double helix structure (Figure 1.5).

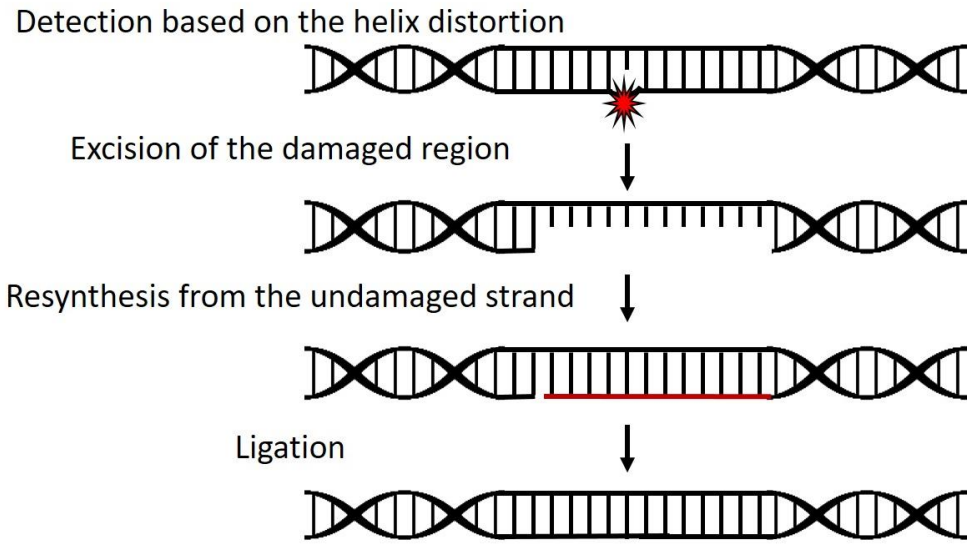

Figure 1.5 The general strategy for single strand DNA damage repair

The three core DNA repair pathways dedicated to repairing the single strand damages share the same strategy. In this strategy, the damage is detected based on its effect on DNA structure. After detection the damaged region is excised and resynthesized from the undamaged strand. Ligation of the remaining nick completes the repair process

Double strand DNA break repair

A complete break in the DNA helix, known as the double strand break (DBS) is amongst the most dangerous forms of DNA damage. DSBs, which can be due to exogenous (e.g., ionizing radiation) or endogenous (e.g., replication fork collapse and V(D)J recombination) sources, are either repaired by non-homologous end joining (NHEJ) or homologous recombination (HR) (Figure 1.6). To choose between these two pathways, cells take into account a variety of factors including the structure of DNA at the site of the break and the presence of a homologous template. As this decision is critical in determining the outcome of the repair, it is highly regulated (Aylon, Liefshitz, and Kupiec 2004; Ira et al. 2004; Mahaney et al. 2013). Although HR is the dominant DSB repair pathway in some organisms (e.g. *Saccharomyces cerevisiae*), NHEJ is the predominant DSB repair pathway in mammals. The preference for NHEJ in the latter group is thought to be due to the deleterious effects of loss of heterozygosity (especially in recessive mutations of antioncogenes) which can result from mitotic HR (Knudson 1993).

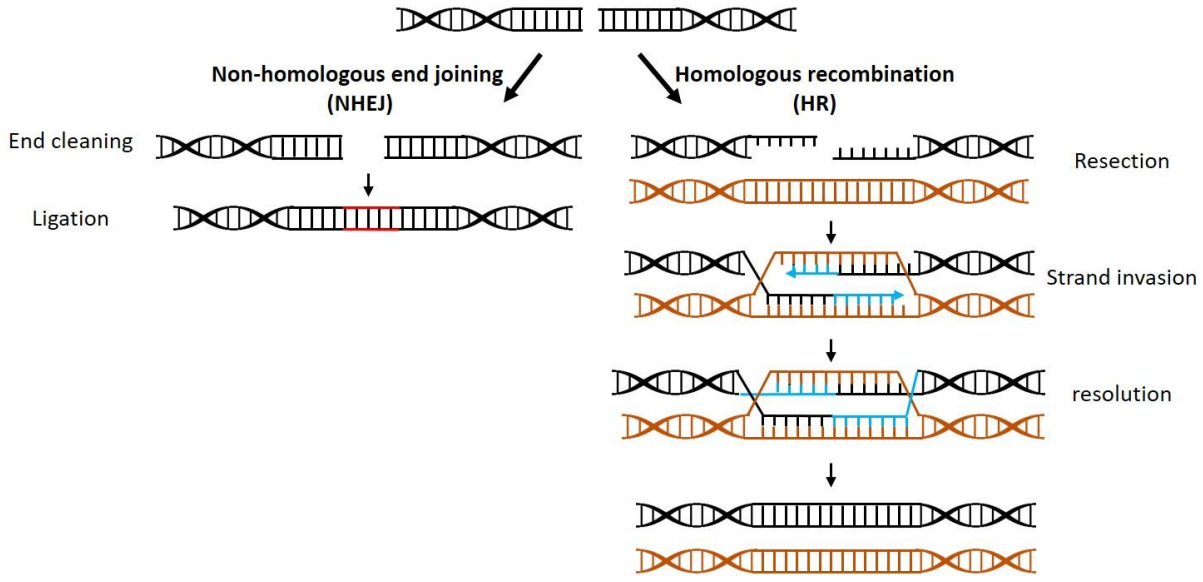

Figure 1.6 Strategies for double stand break repair

Double strand breaks can be repaired by two core DNA repair pathways: non-homologous end joining (NHEJ) and homologous recombination (HR). While HR uses a homologous sequence as template for repairing the damage, NHEJ repairs the break by ligating the ends together. As such, NHEJ is highly prone to introducing mutations at the site (shown in red). Many factors including the structure of the break and the presence of a homologous template factor in the decision between the two pathways. However, resection of the ends commits the repair to HR.

Non homologous end joining

The predominant outcome of the decision between the DSB repair pathways is the evolutionary conserved non-homologous end joining (NHEJ). NHEJ is a fast and crude mechanism for ligation of the broken ends of the DNA together (Chiruvella, Liang, and Wilson 2013; Radhakrishnan, Jette, and Lees-Miller 2014; Guirouilh-Barbat et al. 2004). Moore and Haber coined the term “non-homologous end joining”, since in contrast to homologous recombination (discussed in detail in the next section), this pathway does not require a homologous DNA sequence as a template for guiding the repair (Moore and Haber 1996). Instead, this pathway uses microhomologies (i.e. short homologous DNA sequences) near the broken ends (usually in the single stranded overhangs) as its guide for repair.

If the overhangs of the broken ends are compatible, NHEJ can repair the DSB accurately (Moore and Haber 1996; Boulton and Jackson 1996; Budman and Chu 2005). However, if the overhangs are not compatible or the broken ends are damaged, NHEJ employs a variety of strategies for cleaning the ends (Ma et al. 2005; Waters et al. 2014). This can introduce different types of mutations (including insertions and deletions) to the genome or even ligate the wrong DNA ends, causing DNA translocation or telomere fusion (Espejel et al. 2002). Indeed, the broken ends of DSBs located in the same topological domain have been shown to frequently ligate together (Alt et al. 2013; Zhang et al. 2012; Zarrin et al. 2007). To avoid losing the broken ends and ligating the wrong DNA ends, the broken ends are tethered together by one of the early responders to DSB, the MRN complex (which consist of MRE11, RAD50 and NBS1), as well as DNA dependent protein kinase (DNA-PK) (Zha, Boboila, and Alt 2009; Hammel et al. 2010; Cottarel et al. 2013).

NHEJ is initiated by sliding of Ku (composed of Ku70 and Ku80) onto the DNA ends (Figure 1.7.2a) (Walker, Corpina, and Goldberg 2001). Ku then recruits the DNA-PK catalytic subunit (DNA-PKcs) and together they form the DNA-PK holoenzyme (Figure 1.7.3a). DNA-PK then

phosphorylates itself and several other NHEJ factors, and protects the DNA ends from excessive resection (Jiang et al. 2015).

Based on the physical and chemical condition of the broken ends, different enzymes, such as DNA polymerases β, λ and μ, PNK (polynucleotide kinase), aprataxin, MRN, terminal deoxynucleotidyl transferase and SNM1C/artemis (which is a nuclease), are employed to make the incompatible, and damaged ends compatible (Figure 1.7.4a) (Menon and Povirk 2016; Waters et al. 2014). This processing, however, can cause mutations in the genome. The DNA ends are then ligated by ligase IV-XRCC4 (X-ray repair cross-complementing protein 4) (Figure 1.7.5a) (Critchlow, Bowater, and Jackson 1997; Grawunder et al. 1997; Li, Otevrel, et al. 1995). In addition, XLF (XRCC4 like factor), a homolog of XRCC4, binds to XRCC4 and Ku and stimulates the ligation process (Yano et al. 2011; Mahaney et al. 2013).

Homologous recombination

By using undamaged homologous sequences as template, homologous recombination (HR) is capable of repairing the DSB without introducing mutations. To this end, the broken ends are resected in the 5'-3' direction, creating 3' overhangs which can then invade the homologous region on the sister chromatid. Therefore, resection plays a critical role in the decision between NHEJ and HR, as it commits the cells to HR and inhibits NHEJ (Shibata et al. 2014). BRCA1 (breast cancer type 1 susceptibility protein) promotes the initiation of resection by the MRN complex (which as mentioned above is one of the first responders to the site of DSB and a key regulator of the DNA end resection) and its binding partner, CtIP (Figure 1.7.2b) (Stracker and Petrini 2011; Limbo et al. 2007; Mimitou and Symington 2008; Prakash et al. 2015).

After the initial resection by MRN-CtIP, the exonucleases EXO1 and DNA2 further resect the 5' ends of the DSB to produce more extended 3' overhangs (Figure 1.7.3b) (Mimitou and Symington 2011). While EXO1 does not require the help of a helicase for resection, DNA2

exonuclease activity is dependent on the RecQ family of helicases, including the Bloom complex (which consists of Bloom helicase, Rmi1, Rmi2, and Topoisomerase IIIα) or the Werner syndrome helicases (WRN) (Mimitou and Symington 2008). The created ssDNA 3' overhangs are stabilized by the replication protein A (RPA). BRCA1 also recruits PALB2 (partner and localizer of BRCA2), which in turn recruits BRCA2 (breast cancer type 2 susceptibility protein) (Prakash et al. 2015). BRCA2, then, mediates the replacement of RPA on the 3' overhangs by the recombinase RAD51 (Figure 1.7.4b) (Holloman 2011). Interestingly, Rad51 paralogs, RAD51B, RAD51C, RAD51D and XRCC2 also form a complex (BCDX2) that plays a role in RAD51 recruitment, assembly and stabilization on the 3' overhangs, through a mechanism that is not well understood (Chun, Buechelmaier, and Powell 2013; Suwaki, Klare, and Tarsounas 2011).

After replacing RPA on the 3' overhangs, Rad51 enables homologous pairing and strand invasion of the 3' overhangs which forms the displacement loop (D-loop) (Figure 1.7.5b). The invading ssDNA can now get extended by DNA polymerases copying the information from the undamaged homologous template (Figure 1.7.6b). Capturing the second 3' overhang by the D-loop creates a double Holiday junction (Figure 1.5.7b) (West 2009). The Holliday junction can then get resolved by the assistance of BLM, GEN1 and the SLX4-MUS81-EME1-SLX1 nucleases (Figure 1.7.8b) (Tacconi and Tarsounas 2015). However, if the incisions are made on the wrong strands, this step can lead to sister chromatid exchange (Tacconi and Tarsounas 2015).

Figure 1.7 Double strand break repair mechanisms

1) Double strand breaks can be repaired either by: a) non-homologous end joining (NHEJ) or b) homologous recombination (HR). 2a) NHEJ is initiated by sliding of the Ku complex onto the DNA ends. 3a) Ku then recruits DNA-PKc to the ends. 4a) Based on the condition of the ends (e.g., compatibility of the ends or presence of other damages), different DNA processing enzymes such as DNA polymerases, PNK (polynucleotide kinase), aprataxin, APLF (aprataxin and PNK like factor), MRN (complex of MRE11, RAD50 and NBS1) and its binding partner CtIP, TdT (terminal deoxynucleotidyl transferase) and SNM1C/artemis (which is a nuclease), get recruited to the site. 5a) The compatible DNA ends are ligated by Ligase IV-XRCC4, while XLF stimulates the ligation. 6a) Due to the processing of the ends, NHEJ will likely introduce mutations at the site of damage.

2b) By promoting the resection of the broken ends by MRN and its binding partner CtIP, BRCA1 commits the repair pathway to HR. 3b) After this initial resection, EXO1 and DNA2 with the help of RecQ family of helicases (e.g., Bloom complex and WRN) further resect the ends producing more extended 3' overhangs. RPA (small blue circle) protects the 3' overhangs from damage. 4b) BRCA2 facilitates the replacement of RPA with RAD51, which 5b) promotes invasion of the 3' overhangs into a homologous sequence and formation of the displacement loop (D-loop). 6b) DNA polymerases now can extend the 3' overhangs by copying the information from the undamaged homologous sequence. 7b) Capturing the 5' ends of the broken DNA, leads to the formation of a double Holliday junction which is resolved by resolvases. 8b) This completes the repair pathway and restores the DNA structure. Unlike NHEJ, HR does not introduce mutations at the site of damage.

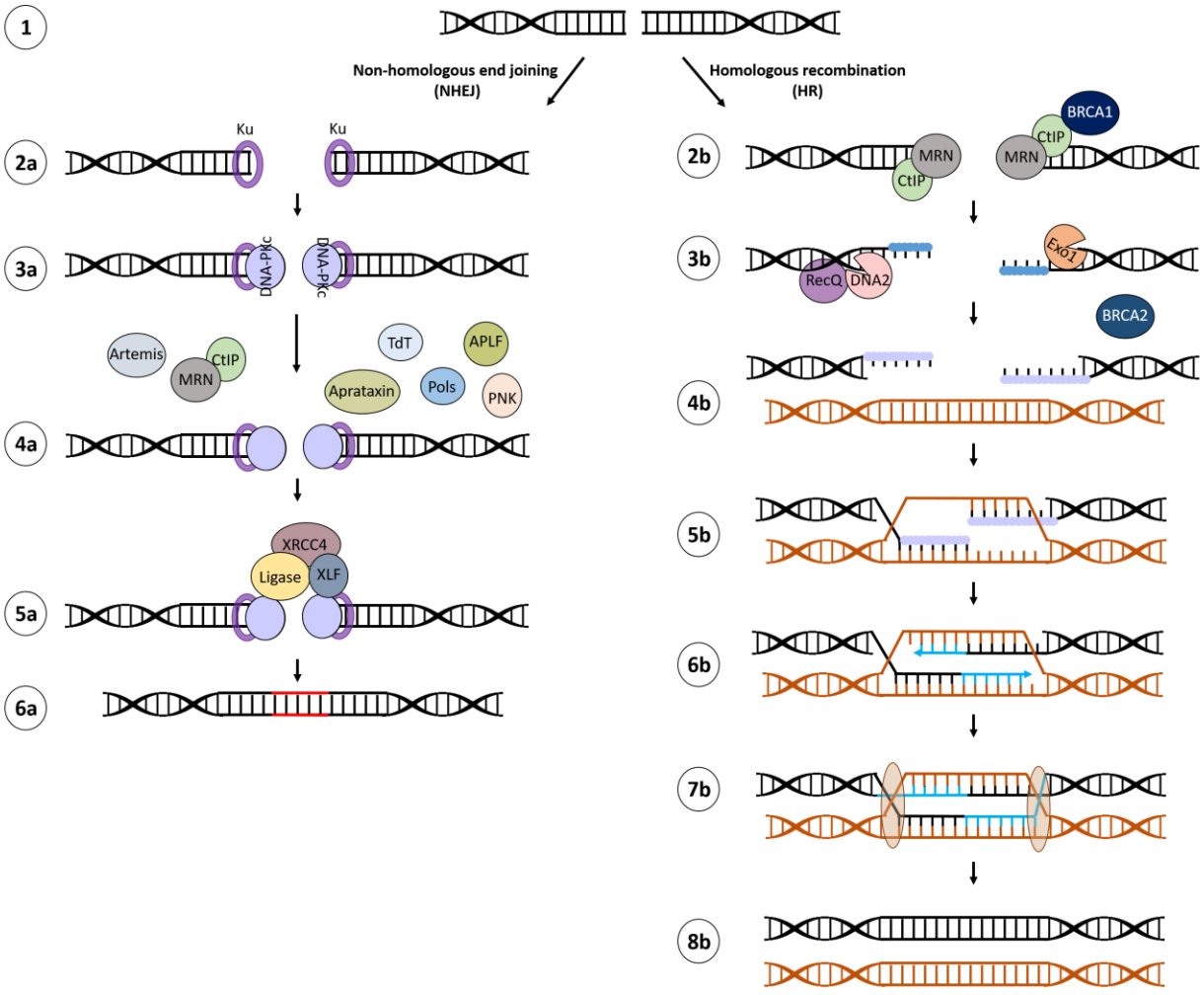

Figure 1.7 (Continued)

Translesion DNA synthesis

Unrepaired DNA damage can hinder replication forks, which in turn can start a series of signaling that initiate DNA repair. If the damage is not repaired in a timely manner, the replication fork may collapse. Fork collapse can have extremely deleterious effects and may lead to cell death (Waters et al. 2009). Considering these deleterious effects, tolerating the damage by replicating past it (termed translesion synthesis (TLS)) can help the cells survive in these situations. As such, TLS is a coping mechanism rather than a repair mechanism. In fact, TLS leaves the damage to be repaired after the passage of the replication fork, when the time constraint is less severe (Friedberg 2005).

During TLS, the replicative polymerases (pol δ/ε) are replaced with translesion polymerases which have a more accommodating active site, can synthesize the DNA across a damaged template, lack proofreading and as a result have lower fidelity (Yang and Woodgate 2007). So far five major translesion polymerases: Rev1, pol κ, pol ι, pol η and pol ζ (Rev3/Rev7) have been identified in eukaryotes. All the known translesion polymerases, except pol ζ, belong to the Y family of DNA polymerases. Interestingly, Pol ζ is from the B family of DNA polymerases which contains the replicative polymerases (Goodman and Woodgate 2013; Lawrence 2004; Ohmori et al. 2001). Translesion synthesis usually requires two translesion polymerases, the first polymerase inserts a nucleotide across from the damage and the second polymerase further extends the DNA until it dissociates, and is replaced with the replicative polymerases (Shachar et al. 2009).

Complex DNA repair pathways

In addition to the common DNA lesions, there are other less frequent forms of DNA damage, including interstrand crosslinks (which covalently attach the two strands of DNA together), DNA-protein crosslinks (which are caused by formation of a covalent bond between a protein and DNA),

and clustered damages (which consists of multiple DNA lesions occurring within one or two turns in the DNA helix). These damages require more complex repair strategies necessitating the collaboration of different core repair pathways. In this section, I will discuss how different repair pathways work together to repair these complex damages. Although the detail of these pathways are still under investigation and many of their aspects remain enigmatic, here I will discuss our current knowledge of the field and the remaining questions.

DNA Interstrand crosslink repair

By preventing unwinding of DNA strands, interstrand cross-links (ICLs) introduce a barrier to transcription and replication and therefore are one of the most deleterious forms of DNA damage. It has been shown that, if they stay unrepaired, only one ICL can kill bacterial and yeast cells, and approximately 20-40 ICLs can kill a mammalian cell (Grossmann et al. 2001; Lawley and Phillips 1996; Muniandy et al. 2010; McVey 2010). In addition, the repair of ICLs can be very error-prone which can lead to genome instability. These lesions can be produced as a result of endogenous events such as lipid peroxidation or exogenous agents such as anticancer drugs (Figure 1.8) (Noll, Mason, and Miller 2006).

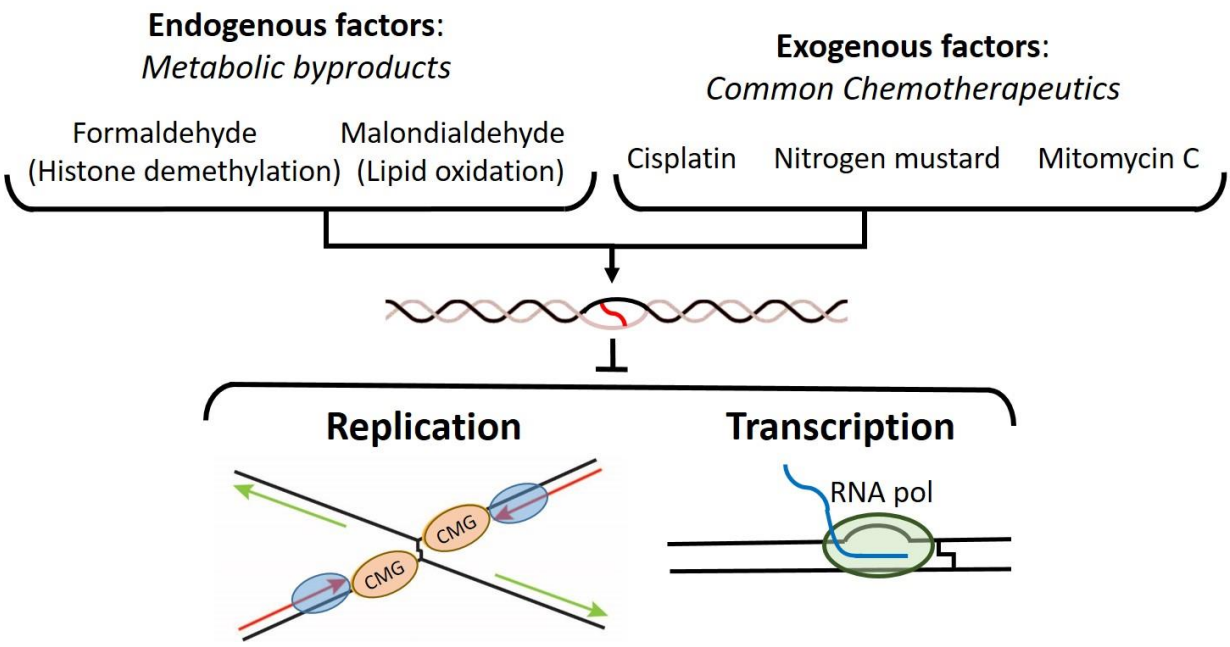

Figure 1.8 DNA interstrand crosslinks (ICLs)

DNA interstrand crosslinks (ICLs), which covalently bind the two strands of DNA together, can be produced by a variety of endogenous (e.g., metabolic byproducts) and exogenous (e.g., common chemotherapeutics) factors. By preventing the unwinding of DNA, they act as a barrier for essential cellular processes such as replication and transcription.

During different phases of the cell cycle, different pathways are activated for the repair of ICLs (Williams, Gottesman, and Gautier 2013). The main mechanism of ICL repair is thought to be dependent on DNA replication which occurs during the S-phase (Raschle et al. 2008). However, ICLs can also be repaired outside of the S-phase by a replication-independent pathway. This mode of repair is critical for the repair of ICLs in non-dividing or terminally differentiated cells (Figure 1.9) (Sarkar et al. 2006; Hlavin et al. 2010; Muniandy et al. 2009; Shen et al. 2006; Zheng et al. 2003; Williams, Gottesman, and Gautier 2012; Ben-Yehoyada et al. 2009; Kato et al. 2017).

Figure 1.9 DNA interstrand crosslinks (ICLs) are detected and repaired by different mechanisms during different phases of cell cycle

Cells have developed different strategies for repairing ICLs during the different phases of cell cycle. During the S phase, ICLs are predominantly detected by converging replication forks, accordingly the repair pathway employed in this phase has been termed the replication dependent pathway. Since outside of the S phase, ICLs are detected either through collision of transcription machinery or by factors detecting helix distortion (and do not involve replication forks), this pathway has been termed replication independent pathway. After detection, in both pathways, separation of the two strands (i.e., unhooking) happens by incisions on either side of the ICL. While in the replication independent pathway, this unhooking creates a gap that can be filled by trasnlesion polymerases, in replication dependent pathway, unhooking creates a DSB that requires both HR and translesion synthesis for repair. Finally, removal of the remaining adduct by NER restores the DNA integrity. It is important to note that many of these steps and factors involved in them are under investigation and future studies are required to establish or amend these pathways.

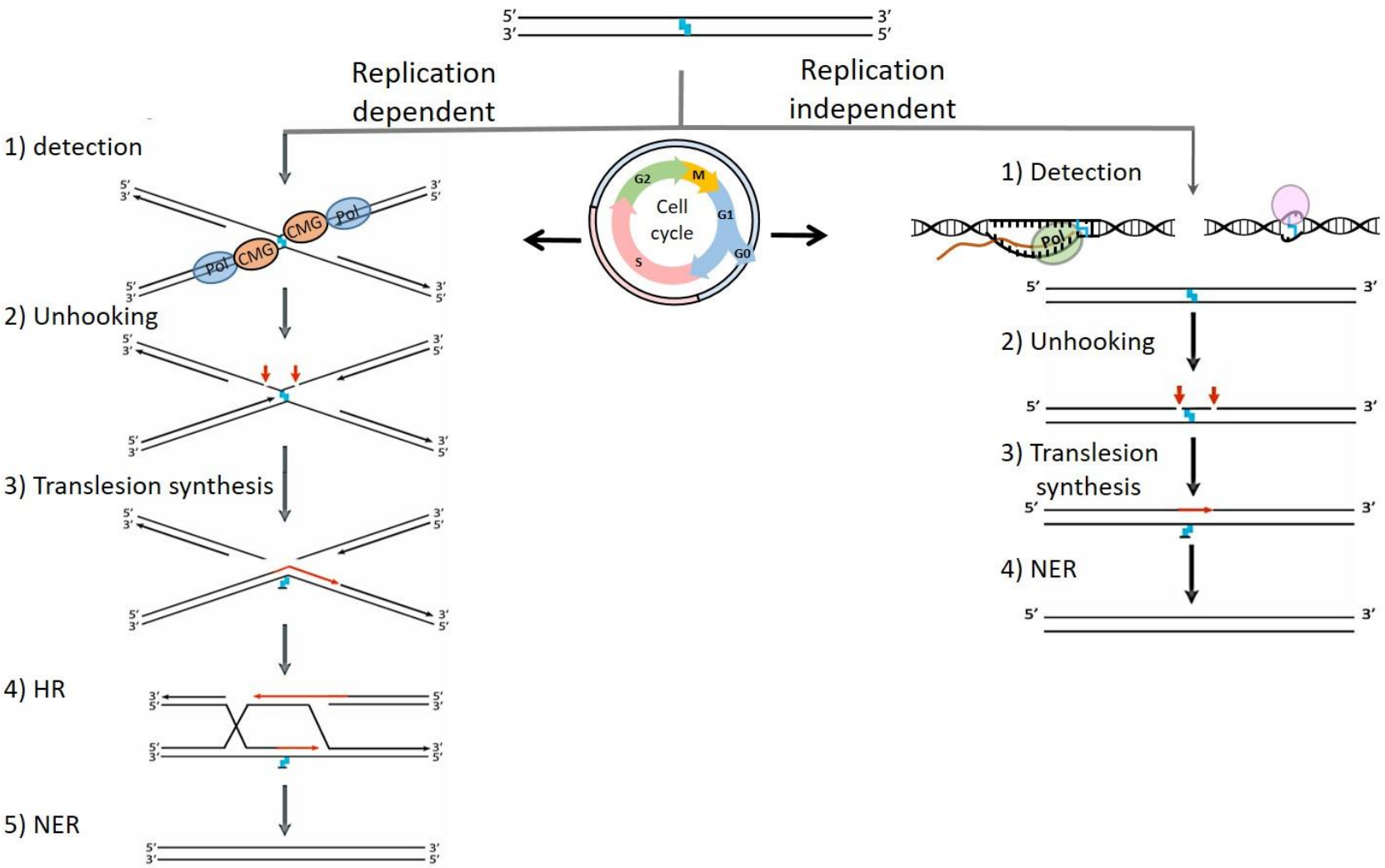

Figure 1.9 (Continued)

Sources of DNA interstrand crosslinks

Developed by Fritz Haber, nitrogen mustard, which is an ICL inducing agent, was originally used as a chemical weapon during World War 1. In an effort to find the antidote, researchers discovered that nitrogen mustard had reduced the number of immune cells in the people exposed to this chemical (Goodman, Wintrobe, and et al. 1946; Einhorn 1985). This led to the idea of using it as a drug for the treatment of leukemia. Therefore, a derivative of nitrogen mustard, mechlorethamine, was developed as the first chemotherapeutic drug (Goodman, Wintrobe, and et al. 1946). However, it was not until the 1940s that the mechanism of action of nitrogen mustard was attributed to the generation of ICLs (Goldacre, Loveless, and Ross 1949). Due to their effectiveness in preventing DNA replication (and thus cellular proliferation), even today ICL inducing agents are widely used for cancer treatment (Deans and West 2011). However, many cancer cells become resistant to these treatments by mechanisms that are still elusive. By better understanding the repair pathways that allow resistance to ICLs, we might be able to design more effective cancer therapies.

In addition to Nitrogen mustard, a variety of bifunctional crosslinking agents such as cisplatin, psoralens, mitomycin C, deoxybutane and their derivatives can also induce ICLs. However, the structure of the ICLs produced by these agents, as well as the degree to which they distort the DNA helix can be very different. For example, mitomycin C, psoralen and nitrogen mustard do not distort the helix greatly, while cisplatin causes a more pronounced distortion in the DNA. These differences depend on several factors such as the sequence preference of the crosslinking agents and the distance between the crosslinked bases. In fact, these physical differences can greatly influence the detection and repair of ICLs. In addition to ICLs, these crosslinking agents can cause DNA protein crosslinks, DNA monoadducts and intrastrand crosslinks, although to varying degrees. For example, mitomycin C produces approximately 15% ICLs, 50% monoadducts and 35% intrastrand crosslinks (Lopez-Martinez, Liang, and Cohn 2016; Dronkert and Kanaar 2001).

ICLs can also be generated endogenously by the byproducts of metabolic reactions. For example, reactive aldehydes (e.g., formaldehyde and acetaldehyde, which are byproducts of carbohydrates metabolism, histone demethylation and biosynthesis of purines and amino acids), malondialdehyde and unsaturated aldehydes (e.g., acrolein and crotonaldehyde, which are byproducts of lipid peroxidation during oxidative stress), and nitric oxide can also produce ICLs (Garaycoechea et al. 2012). In addition, ICLs can form spontaneously between abasic site (which are present in cells at a steady state level of 50,000-200,000, due to the natural decay of DNA or as repair intermediates) (Price et al. 2014; Dianov and Hubscher 2013). Finally, environmental factors such as psoralens and furocoumarins (found in plants such as bergamot, celery, and parsley), as well as Mitomycin C (produced by *Streptomyces caespitosus*) can also cause ICLs (Deans and West 2009). Therefore, ICLs are formed regularly in natural conditions and their repair deficiency can lead to several genetic disorders, including Fanconi anemia and karyomegalic interstitial nephritis. As such, understanding this repair pathway may help develop treatments for these patients.

Replication-dependent interstrand crosslink repair

In the replication-dependent ICL repair pathway, the collision of the replication forks with the ICL[12] activates a series of repair processes mediated by Fanconi anemia proteins (see below) (Wang 2007; Raschle et al. 2008; Knipscheer et al. 2009). The replication forks stall behind the CMG replicative helicase (comprised of Cdc45, MCM2-7, and GINS), which unwinds the DNA

[12] Although several studies, in Xenopus egg extract and mammalian cells, have shown that a single fork can initiate repair (Le Breton et al. 2011; Nakanishi et al. 2011; Mutreja et al. 2018; Rohleder et al. 2016; Huang et al. 2013), Zhang et al. show that if only one fork hits the ICL, CMG is not unloaded and repair fails (Zhang et al. 2015). The preference between these two models has been debated extensively and remains controversial (Deans and West 2011; Kottemann and Smogorzewska 2013; Muniandy et al. 2010; Williams, Gottesman, and Gautier 2013; Bunting and Nussenzweig 2010; Clauson, Scharer, and Niedernhofer 2013; Legerski 2010; Walden and Deans 2014; Zhang and Walter 2014).

ahead of the DNA polymerase (Figure 1.10.2) (Fu et al. 2011). Soon after forks stall at the ICL, the nascent lagging strands undergo 5' → 3' resection (Figure 1.10.3). This likely leads to the activation of the ATR checkpoint signaling pathway, whose activity depends on the presence of ssDNA (Zou and Elledge 2003). Dissociation of the helicases[13] allows one of the leading strands to resume synthesis until it reaches the ICL[14] (Figure 1.10.3) (Fu et al. 2011; Long et al. 2014)[15]. Next, endonucleases cut both sides of the ICL on the non-template parental strand of the approached fork (Figure 1.10.4). Incisions lead to a double stranded break (DSB)[16] in one chromatid and unhooking of the ICL (which remains attached to the other strand) (Figure 1.10.5). Now, translesion DNA polymerases can bypass the crosslink by inserting one nucleotide across from the ICL adduct. Further extension of this bypassed strand by TLS polymerases completes the replication of one chromatid (Figure 1.10.5). The DSB in the other chromatid undergoes resection, preparing it for repair by homologous recombination between the two sister chromatids (Figure 1.10.6). Finally, the remaining adduct is removed by NER, restoring the correct form of the DNA double helix (Figure 1.10.7).

[13] The ubiquitin mediated unloading of CMG, which is essential for repair, is regulated by BRCA1-BARD1 complex, and requires the assistance of an active DNA polymerase (Long et al. 2014).

[14] Based on the lack of errors in this synthesis, it has been suggested that the replicative polymerases are able to approach the ICL till -1 position (Budzowska et al. 2015). However, it is possible that based on the structure of the ICL other polymerases would be needed for this step (Ho et al. 2011).

[15] It is not known how the decision between the two replication forks is made.

[16] DSBs have been one of the early observations of the replication dependent ICL repair (De Silva et al. 2000) and depend on the ubiquitination of a Fanconi anemia protein FANCD2 (see below) (Knipscheer et al. 2009).

Figure 1.10 Replication-dependent interstrand crosslink (ICL) repair mechanism

1) By crosslinking the two strands of DNA together, ICLs create a barrier for essential DNA functions, such as transcription and replication. 2) The replication dependent ICL repair starts when replication forks collide with the ICL. The stalled replication forks recruit FANCM, which in turn recruits the rest of the Fanconi anemia (FA) core complex. 3) Fork stalling also causes the resection of lagging strands by exonucleases. This leads to creation of single stranded DNA and activation of the ATR checkpoint kinase. Unloading of the replicative helicases (CMG), by an unknown mechanism, allows for the approach of one of the DNA polymerases to the ICL. 4) After its recruitment, the FA core complex activates the FANCI-FANCD2 heterodimer by mono-ubiquitinating them. This is further facilitated by phosphorylation of FANCI by the activated ATR kinase. The activated FANCI-FANCD2 control the rest of the pathway including the recruitment of the endonucleases that unhook the ICL by making incisions at either side of the ICL. SLX4 is a large scaffold protein that is thought to both help recruit the endonucleases, as well as orient them for making the correct incisions. It has been shown that XPF performs at least one of the incisions, while the endonuclease responsible for the second incision is still unknown. 5) Although the incisions liberate the two strands from each other, they leave an oligo adduct attached to one of the strands and create a DSB in the other one. Since replicative polymerases are not able to use the adducted template for synthesis, translesion polymerases are recruited to complete the replication of one of the strands. Further resection of the DSB ends in the sister chromatid creates longer 3' overhangs. Rad51 (small light blue circle) replaces the RPA (small dark blue circles) on the overhangs. 6) Rad51 facilitates strand invasion to the sister chromatid. Homologous recombination (HR) repairs the DSB. 7) Finally the remaining adduct is thought to be removed by nucleotide excision repair (NER) pathway.

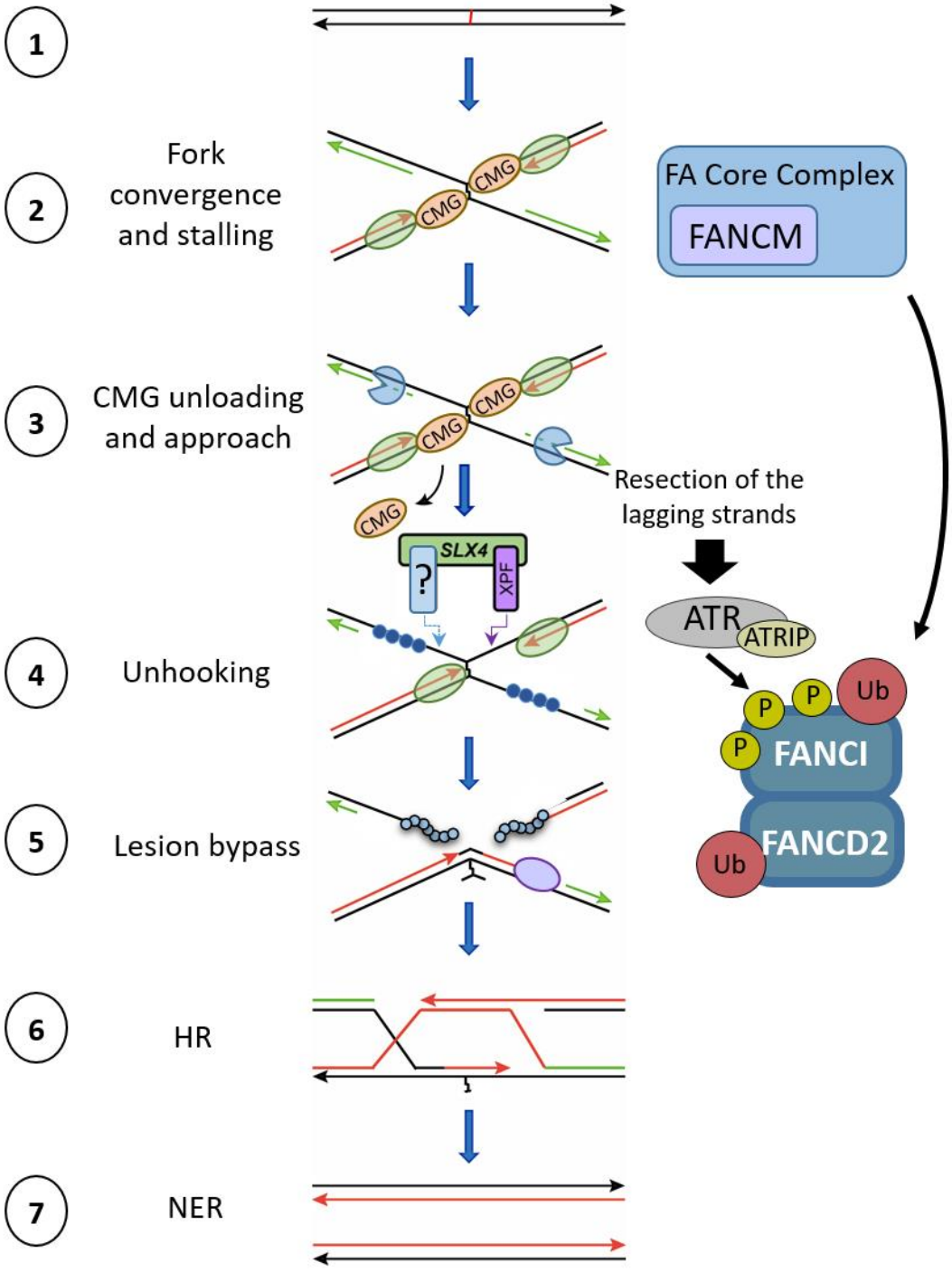

Figure 1.10 (Continued)

Factors involved in replication-dependent ICL repair

Genetic studies have identified four general groups of factors involved in replication-dependent ICL repair: 1) the Fanconi anemia factors which are the master regulators of the pathway, 2) nucleases, which can be divided into two groups a) endonucleases responsible for unhooking the ICL and resolving the HR intermediates, and b) exonucleases responsible for resection of the lagging strands as well as the DSB ends, 3) translesion polymerases which are able to synthesize DNA across from the remaining ICL adduct, and 4) factors involved in homologous recombination[17].

The Fanconi anemia pathway

Fanconi anemia (FA), which is a rare hereditary human disease[18] causing bone marrow failure, cancer predisposition, developmental defects, and genomic instability, was first described by a Swiss physician, Guido Fanconi, in 1927 (Fanconi 1927; Lobitz and Velleuer 2006; Alter 2014). However, it was not until the 1970s that the cause of FA was suggested to be the increased sensitivity of FA cells to ICL inducing agents, signifying a deficiency in the repair of ICLs (specially ICLs made by endogenous factors such as aldehydes) as the underlying cause of this condition (Sasaki and Tonomura 1973; Auerbach and Wolman 1976)[19]. We now know that FA is caused by biallelic mutations in the Fanconi proteins, 21 of which have been identified so far (FANCA, B, C,

[17] The HR factors have been discussed earlier in the chapter.

[18] In the US, the estimated occurrence of FA is 1 out of 360000 live births, with 1 out of 181 people being a carrier (Rosenberg, Tamary, and Alter 2011).

[19] This discovery is still being used for diagnosis of FA, based on the rate of accumulation of chromosomal abnormalities after treatment of cells with an ICL inducing agent, diepoxybutane. This is especially important since diagnosis of this heterogeneous disease based on the symptoms alone is very difficult (Auerbach 1988). However, the deficiency in ICL repair still has not been established as the sole cause of all FA symptoms (Bagby and Olson 2003)

D1, D2, E, F, G, I, J, L, M, N, O, P, Q, R, S, T, U and V[20]) (Wang 2007; Mamrak, Shimamura, and Howlett 2017; Kottemann and Smogorzewska 2013; Ceccaldi, Sarangi, and D'Andrea 2016; Walden and Deans 2014). The FA proteins are only activated during the S-phase[21] in the presence of active replication and as such, they are only involved in the replication-dependent repair of ICLs (Rothfuss and Grompe 2004; Taniguchi et al. 2002).

The Fanconi anemia pathway, which is activated upon stalling of the replication forks, employs the FA proteins for regulation of different steps of ICL repair (such as damage recognition and recruitment of nucleases and repair polymerases to the site of damage for ICL unhooking and HR). The FA proteins can be divided into three main groups. The first group includes FANCA, B, C, E, F, G, L, and M, and the Fanconi associated proteins: FAAP20, FAAP24 and FAAP100, and MHF1 and MHF2. Together, they make up the FA core complex, which is an E3 ubiquitin ligase (Kim et al. 2008; Qiao, Moss, and Kupfer 2001). From this group, the FANCM subcomplex (including FANCM, FAAP24, MHF1, and MHF2) is one of the first responders to the site of damage and recruits the rest of the core complex by binding to FANCF (Huang et al. 2010; Huang et al. 2013; Singh et al. 2010). In addition to the FANCM subcomplex, the core complex includes three other subcomplexes: a catalytic subcomplex (FANCB, FANCL and FAAP100) and two subcomplexes that assist with the binding of the catalytic subcomplex to the DNA (one consisting of FANCA, FANCG, and FAAP20, and the other made up of FANCC, FANCE and FANCF) (Huang et al. 2014; Yuan et al. 2012; Leung et al. 2012).

Upon recruitment to the ICL, FANCL from the catalytic subcomplex (with the help of UBE2T/FANCT which is the E2 ubiquitin-conjugating enzyme in this reaction but not part of the

[20] FANCH complementation group was found to be analogous to FANCA complementation group and therefore was removed (Joenje et al. 2000; Joenje et al. 1997).

[21] FANCM, the first link of the FA pathway, is hyperphosphorylated during mitosis. This has been suggested to inhibit FA pathway outside of the S-phase (Kim et al. 2008).

core complex (Rickman et al. 2015; Virts et al. 2015; Alpi et al. 2008)) mono-ubiquitinates its only known substrates, the heterodimer of FANCI and FANCD2 (ID2)[22](Garcia-Higuera et al. 2001; Smogorzewska et al. 2007; Huang et al. 2014; Longerich et al. 2014; Sato et al. 2012; Alpi et al. 2008). FANCI and FANCD2, which constitute the second group of the FA proteins, are highly conserved (Joo et al. 2011). Ubiquitination of ID2 stabilizes its interaction with DNA (specifically at sites with stalled replication fork structure (Liang et al. 2016)) allowing it to coordinate the downstream repair steps[23] through the third group of FA proteins.

The third group of FA proteins can be subdivided into two groups: 1) factors involved in incisions and unhooking of the ICL: XPF/FANCQ (an endonuclease that works with its regulatory subunit ERCC1) and SLX4/FANCP (a large scaffolding protein that binds several endonucleases such as SLX1, MUS81 and XPF, and is thought to interact with FANCD2 through its UBZ (ubiquitin binding zinc finger domain), and 2) the homologous recombination factors: RAD51/FANCR (a strand transfer catalyst), RAD51C/FANCO (a RAD51 paralog), XRCC2/FANCU (a RAD51 paralog), BRCA1/FANCS (a homologous recombination enhancer), BRIP/FANCJ (a BRCA1-interacting helicase), BRCA2/FANCD1 (which enhances RAD51 mediated strand invasion), PALB2/FANCN (a BRCA2 regulator), and REV7/FANCV (a translesion DNA polymerase) (Klein

[22] While the replication forks are stalled, the lagging strands are resected generating ssDNA. Since ssDNA is a signal for DNA damage, it activates the ATR (ataxia telangiectasia and Rad3-related protein) kinase which is one the most upstream kinases in the DNA Damage Response (Zou and Elledge 2003). ATR phosphorylates ID2 heterodimer, further facilitating its ubiquitination and activation by the Fanconi core complex (Andreassen, D'Andrea, and Taniguchi 2004; Pichierri and Rosselli 2004).

[23] Ubiquitination of FANCD2 seems to have a more important role in ICL repair, since a mutant version of FANCI lacking the ubiquitination site can partially rescue FANCI knock out cells (Garcia-Higuera et al. 2001; Matsushita et al. 2005; Timmers et al. 2001; Ishiai et al. 2008; Smogorzewska et al. 2007). The main role of FANCI seems to be in promoting FANCD2 ubiquitination (Sato et al. 2012; Longerich et al. 2014; Smogorzewska et al. 2007).

Douwel et al. 2014; Bogliolo et al. 2013; Stoepker et al. 2011; Sawyer et al. 2015; Park et al. 2016)[24].

Nucleases

Nucleases are one of the major classes of factors that have been shown to be involved in various repair pathways (e.g., DSB repair) including different stages of ICL repair: resection of the stalled lagging strands, incisions that lead to ICL unhooking, trimming of the unhooked ICL adduct, resection of the DSB, resolution of the Holliday junctions created during HR, and removal of the ICL adduct. As such, many nucleases have been implicated in ICL repair including XPF-ERCC1, SLX4-SLX1, MUS81-EME1, FAN1, CTIP, MRE11, EXO1, DNA2, SNM1A, and SNM1B.

Endonucleases

Unhooking of the ICL, which liberates the two strands from each other, is a crucial step in all ICL repair pathways and is performed by structure-specific endonucleases. Several nucleases have been suggested to be involved in incisions including XPF-ERCC1, SLX4-SLX1, MUS81-EME1, and FAN1 (Hanada et al. 2006; Trujillo et al. 2012; Svendsen et al. 2009; Kratz et al. 2010; Kuraoka et al. 2000).

Recently, it was shown that at least one of the incisions during unhooking is made by XPF-ERCC1 (Klein Douwel et al. 2014; Klein Douwel et al. 2017). The XPF-ERCC1 heterodimer can cleave the dsDNA on the 3' arm/flap side, near its junction with ssDNA in DNA structures (Tsodikov et al. 2005; de Laat et al. 1998; Bowles et al. 2012)[25]. SLX4/FANCP which is a scaffold

[24] Deubiquitination of ID2 after the completion of repair by a complex of USP1 and UAF1 (USP1 associated factor 1) is important for regulation of ICL repair (Cohn et al. 2007; Nijman et al. 2005; Oestergaard et al. 2007; Smogorzewska et al. 2007; Kim et al. 2009).

[25] The XPF-ERCC1 heterodimer is composed of XPF/FANCQ, the catalytic subunit (which contains a ERCC4 (excision repair cross-complementation group 4) domain), and ERCC1, (excision repair cross-complementation group 1) the regulatory subunit required for substrate specificity in which the nuclease domain is inactive. XPF-ERCC1 has been studied extensively in

protein for nucleases (including SLX1, XPF-ERCC1, and MUS81-EME1) is thought to recruit XPF-ERCC1 (Klein Douwel et al. 2014)[26]. It is also possible that when XPF-ERCC1 or SLX4 are not recruited to the ICL, MUS81-EME1, FAN1 or other endonucleases perform the incisions necessary for unhooking. Interestingly, although FAN1 deficiency leads to ICL sensitivity, FAN1 and the FA pathway are not epistatic and mutations in FAN1 lead to karyomegalic interstitial nephritis (KIN) rather than Fanconi anemia (Yoshikiyo et al. 2010; Zhou et al. 2012; Trujillo et al. 2012). This suggests that FAN1 is involved in a FA-independent ICL repair pathway. However, its role in ICL repair remains enigmatic.

Although the exact mechanism of recruitment of endonucleases is unknown, they are thought to be recruited and controlled by the ubiquitinylated ID2 complex, which is localized at the site of damage (Taniguchi et al. 2002; Klein Douwel et al. 2014; Klein Douwel et al. 2017)[27]. For example, FAN1 and SLX4 interaction with ubiquitinylated ID2 complex through their UBZ domain has been shown to be important for their recruitment (Kratz et al. 2010; Liu et al. 2010; MacKay et al. 2010; Smogorzewska et al. 2010). ID2 is also believed to direct the incisions during unhooking. This is perhaps the most important step during the repair process as aberrant incisions (e.g., incisions on both parental strands) not only would lead to repair failure, it could cause extreme genome rearrangement.

Exonucleases

The main function of exonucleases in DNA repair is thought to be through resection and processing of the DNA ends at the site of damage. In replication-dependent ICL repair, they are

---

the context of NER (see section on NER) (Tsodikov et al. 2005; de Laat et al. 1998; Bowles et al. 2012).

[26] SLX4 also enhances XPF activity (Hodskinson et al. 2014).

[27] FANCI-D2 has been shown to have a preference for binding branched DNA structures (with the ability to accommodate an ICL), allowing it to locally control the repair process (Joo et al. 2011).

involved in resection of the lagging strands (of stalled replication forks), as well as the new 5' ends of the DSB formed after ICL unhooking (Long et al. 2011). These resections are required for both activation of ATR[28], one of the early stages of the ICL repair pathway, as well as the initiation of homologous recombination, one of the final steps of ICL repair. Therefore, it is not surprising that several nucleases including MRE11, CTIP, EXO1, and DNA2 have been implicated in ICL repair (Duquette et al. 2012; Karanja et al. 2012; Suhasini et al. 2013).

Although the role of exonucleases in DSB repair has been studied extensively, their role in ICL repair and the mechanism by which they are recruited and regulated during ICL repair are still largely unknown. For example, it is not clear if they have redundant roles or each has a distinct role in the repair process (e.g., if the exonucleases involved in the early resection of the lagging strands are different from those involved in the later resection of the DSB). In addition, since overproduction of ssDNA can lead to genome rearrangement and clustered mutagenesis (Karanja et al. 2014), over-resection can be extremely toxic. As such, regulation of the activity of exonucleases has to be strictly regulated during the repair process. However, the factors that activate and regulate the nucleases are largely unknown.

The molecular mechanism by which exonucleases are recruited to ICL is also still largely unknown. Two recent publications suggest that ubiquitinated FANCD2 interacts with and recruits CtIP to ICLs (Murina et al. 2014; Unno et al. 2014). However, these results are mainly based on low-resolution co-localization experiments and the exact mechanism by which FANCD2 recruits CtIP is not clear. In addition, it is not clear how MRE11, EXO1, and DNA2 are recruited to the ICL. Since different exonucleases might be involved in different stages of ICL repair, their recruitment might be regulated through different mechanisms. Especially since lagging strand

[28] ATR is activated by ssDNA and promotes ID2 ubiquitination and activation.

resection is assumed to happen before FANCD2 activation, the exonucleases involved in this process are not likely recruited by FANCD2. A potential candidate for this role is FANCM.

An additional role for exonucleases during ICL repair is processing of the unhooked ICL adduct to a one nucleotide adduct (monoadduct). This digestion is thought to be required for filling of the gap by TLS polymerases (Ho et al. 2011; Minko et al. 2008). Two members of the β-CASP subfamily of metallo-β-lactamase related nucleases (which have 5'-3' exonuclease activity), SNM1A and SNM1B, have been implicated in this step (Allerston et al. 2015; Sengerova et al. 2012; Cattell, Sengerova, and McHugh 2010)[29]. Depletion of either of the nucleases causes cellular sensitivity to ICL inducing agents, although SNM1A depletion causes greater sensitivity (Hemphill et al. 2008; Bae et al. 2008; Mason and Sekiguchi 2011). In addition, both have been shown to digest past an ICL *in vitro*, with SNM1A showing higher activity, suggesting that these nucleases are involved in trimming the remaining adduct (Wang et al. 2011; Sengerova et al. 2012). Another possibility is that after the 5' incision during unhooking, SNM1A or SNM1B load from the nick and digest past the ICL removing the requirement for a second endonuclease to perform the 3' incision (Sengerova et al. 2012). In another study, depletion of SNM1B inhibited FANCD2 localization on the ICLs, suggesting an earlier role for SNM1B such as remodeling of the replication fork (Mason and Sekiguchi 2011; Mason et al. 2013).

Translesion polymerases

Interestingly, cells from patients with Fanconi anemia have fewer point mutations than healthy cells (Niedernhofer, Lalai, and Hoeijmakers 2005; Niedzwiedz et al. 2004). This has been suggested to be due to the role of the translesion (TLS) polymerases in several steps of the FA dependent ICL repair including the insertion of a nucleotide across from the adduct and extension beyond the adduct both after unhooking and the HR repair of the DSB in the sister chromatid. In

[29] The third member of the SNM1 family, SNM1C/Artemis, is involved in NHEJ but not ICL repair.

comparison to replicative polymerases, TLS polymerases have a higher error rate. Therefore, there is a higher chance of mutations during synthesis by TLS polymerases, which decreases the fidelity of the ICL repair. These error-prone polymerases have been shown to be recruited to the ICL by the FA core complex (Budzowska et al. 2015; Niedzwiedz et al. 2004; Mirchandani, McCaffrey, and D'Andrea 2008; Kim et al. 2012).

Several TLS polymerases including Rev1, pol ζ, η, ν and κ have been implicated in ICL repair (Enoiu, Jiricny, and Scharer 2012; Williams, Gottesman, and Gautier 2012; Sarkar et al. 2006; Niedzwiedz et al. 2004; Roy and Scharer 2016). Rev1[30] can interact with PCNA, as well as pol ζ, η, ι, and κ. As such, it is thought to have the role of a landing pad for TLS polymerases, enhancing the switch from replicative polymerases to TLS polymerases during ICL repair (Budzowska et al. 2015). Although, it is believed that different TLS polymerases perform the insertion (adding one nucleotide across from the adduct) and extension (synthesizing the strand after insertion) steps, it not clear which TLS polymerases are involved in these steps (Raschle et al. 2008). It is highly likely that the choice between the TLS polymerases is highly dependent on the structure of the ICL and the unhooked adduct (Raschle et al. 2008; Yamanaka et al. 2010; Minko et al. 2008; Klug et al. 2012; Ho et al. 2011). It is also possible that replicative polymerases have the ablity to perform the insertion step (maybe with lower efficiencies) but require a TLS polymerase for the extension since they will not be able to extend from a mispaired primer.

Alternative replication-dependent interstrand crosslink repair pathways

ICLs can be formed by diverse crosslinking agents between different nucleotides that are separated by various distances or are located in different DNA structures such as major or minor grooves. As such, ICLs are very heterogeneous in their chemical and physical characteristics.

[30] Rev1 is a dCMP transferase and is able to insert a dCMP across from abasic sites or modified guanine bases.

Therefore, cells may require different strategies for detecting and repairing different types of ICLs (Deans and West 2011; Ho et al. 2011). In addition, based on the state of the cells and the availability of the factors involved, different factors or pathways may substitute the steps explained above. For example, the DSB formed after unhooking may get repaired by NHEJ instead of HR (McHugh, Sones, and Hartley 2000; Aggarwal et al. 2013). Finally, although the replication-dependent pathway is thought to be the main mechanism of ICL repair, during the other phases of cell cycle, ICLs still can be detected and repaired by other mechanisms[31] (Muniandy et al. 2010; Williams, Gottesman, and Gautier 2013).

Interstrand crosslink traverse model

Recent observations in mammalian systems have challenged the belief that ICLs completely block the replication (and their repair requires the convergence of two replication forks to the site of damage). By using single-molecule studies in Chinese hamster ovary cells, it was demonstrated that 60% of the replication forks can rapidly bypass a 4,5',8 trimethylpsoralen ICL without unhooking it, albeit with some short pausing (Huang et al. 2013; Rohleder et al. 2016). This mechanism has been termed the replication traverse model (Huang et al. 2013). Although this mechanism is still largely uncharacterized, it has been shown to depend on the DNA binding and translocase activities of FANCM (which is able to remodel stalled replication forks) (Huang et al. 2013), interaction of FANCM and PCNA (which is the DNA replication processivity factor) (Rohleder et al. 2016), Bloom syndrome complex (which consists of Bloom helicase, DNA topoisomerase IIIα, RMI1 and RMI2) (Ling et al. 2016), RAD51(which is an HR factor) and ATR (which is the DNA replication checkpoint kinase) (Mutreja et al. 2018), and it is not dependent on

[31] For example during the G0/1 phase, ICLs can be detected by stalled transcription machinery or by the NER, BER or MMR pathways due to DNA distortion. After detection they are repaired by replication-independent mechanisms that are dependent on unhooking of the ICL (by incisions at 3' and 5' of ICL) followed by translesion DNA synthesis across the unhooked adduct.

other FA pathway proteins (Rohleder et al. 2016). In this model, upon encountering an ICL, the stalled replication fork is reversed (through a mechanism that depends on FANCM, Bloom, RAD51, and ATR) creating a four-way junction (Huang et al. 2013; Rohleder et al. 2016; Mutreja et al. 2018). Replication fork restart results in a quick bypass of the ICL (without the need for ICL repair)[32]. However, it is not known whether the CMG bypasses the ICL or a different helicase is loaded. This mechanism helps to maintain genome integrity by avoiding stalled replication forks and fork-associated repair. The subsequent repair steps (which happen after the replication) are thought to be the same and are initiated by FANCM (Huang et al. 2013).

Glycosylase dependent interstrand crosslink repair

In addition to nucleotide excision repair and homologous recombination, base excision repair has also been implicated in ICL repair (Berquist and Wilson 2012; Wilson and Seidman 2010). For example, glycosylases: AAG (Maor-Shoshani et al. 2008) and NEIL1 (Mace-Aime et al. 2010) were shown to protect cells against psoralen ICLs. More recently, it was shown that psoralen and abasic site ICLs, but not cisplatin ICLs, can undergo replication-dependent repair independently of the FA pathway (Semlow et al. 2016). In this pathway, glycosylase NEIL3[33] cleaves one of the glycosidic bonds in the crosslink, liberating the two DNA strands and allowing the resumption of replication by translesion DNA polymerases. Interestingly this pathway does not require CMG unloading or incisions and thus avoids DSB and HR. However, if this pathway is inhibited, the repair is rerouted to the FA pathway. Therefore, although the replication machinery detects diverse ICLs, the choice of repair pathway depends on the chemical and physical characteristics of the ICL and the availability of different repair factors.

[32] FANCM recruits the ssDNA binding protein RPA, which might help melt the DNA around the damage (Huang et al. 2010).

[33] Interestingly, although NEIL3 is a bifunctional glycosylase, in this context it seems to acts as a monofunctional glycosylase.

Replication-independent interstrand crosslink repair

Although the replication-dependent ICL repair mechanism has attracted a lot of attention, largely due to its importance in cancer therapy, ICLs can also be repaired in the absence of replication (e.g., during the G1 phase of cell cycle) (Sarkar et al. 2006; Hlavin et al. 2010; Muniandy et al. 2009; Shen et al. 2006; Zheng et al. 2003; Williams, Gottesman, and Gautier 2012; Smeaton et al. 2008; Ben-Yehoyada et al. 2009; Kato et al. 2017). In fact, this mode of repair can be crucial for non-dividing or slowly dividing cells (e.g., neurons) during treatment with ICL inducing anticancer agents, or may even help cancerous cells acquire drug resistance.

Replication-independent ICL repair (RIR) was originally thought to be achieved by two rounds of excision repair (Cole 1973). In line with this belief, genetic studies have revealed two major groups of proteins involved in RIR: excision repair factors (involved in BER, NER or MMR) and TLS polymerases (Muniandy et al. 2009; Sarkar et al. 2006; McHugh, Sones, and Hartley 2000; Thoma et al. 2005; Wood 2010; Zhao et al. 2009; Noll, Mason, and Miller 2006; Huang et al. 2011). However, their exact mechanism of action remains unknown.

In the absence of replication, ICLs could be detected either by stalled transcription machinery, or factors that detect distortion of the DNA double helix such as NER and MMR damage sensors (Figure 1.11.1) (Enoiu, Jiricny, and Scharer 2012; Wang et al. 2001; Ben-Yehoyada et al. 2009). Interestingly, as in NER, the efficiency of ICL repair is higher if they are located in actively transcribing regions (Zheng et al. 2003; Islas, Vos, and Hanawalt 1991). In these regions, the stalled transcription machinery recruits the NER factors through CSA or CSB (with the same mechanism as TC-NER) (Hlavin et al. 2010; Zheng et al. 2003; Enoiu, Jiricny, and Scharer 2012).

In the absence of replication or transcription, ICLs can still be detected by DNA damage surveillance factors that can sense the distortion in DNA structure. XPC, which is the GG-NER surveillance factor, have been shown to detect the less distorting psoralen and mitomycin C

induced ICLs (Thoma et al. 2005; Muniandy et al. 2009; Zheng et al. 2003; Wood 2010). While, the MMR surveillance factors, MutSα and MutSβ, can recognize the more distorting ICLs such as cisplatin, cisplatin-D, O6-methylguanine, and O4-methylthymine, as well as psoralen ICLs (Duckett et al. 1996; Yamada et al. 1997; Vasquez 2010; Zhang et al. 2002; Kato et al. 2017).

Upon recognition, the ICL is thought to be unhooked by incisions on both sides of the lesion, leading to the formation of a gap in one of the DNA strands (Figure 1.11.2). It has been shown that in mammalian cell extracts, in the absence of replication, the NER endonucleases, XPF and XPG, incise 5' to the ICL (Smeaton et al. 2009; Wang et al. 2001; Sarkar et al. 2006). In order to unhook the ICL, an incision has to also be made 3' of the ICL. Several endonucleases including CtIP, XPF, SLX4-SLX1, FAN1, and SNM1A have been implicated in this step (Klein Douwel et al. 2014; Sengerova et al. 2012). However, the nuclease performing this step is still unknown.

After unhooking, the ICL adduct (likely attached to an oligonucleotide) remains connected to one of the strands. Digestion of this remaining adduct to a one-nucleotide adduct (monoadduct) is likely required to enable DNA synthesis across from it (Figure 1.11.3). It has been shown that SNM1A is able to digest an oligonucleotide attached by an ICL, leaving behind a monoadduct (Sengerova et al. 2012; Roy et al. 2016; Wang et al. 2011)[34]. This suggests that SNM1A is involved in further processing of the adduct, which is left after unhooking of the ICL. Now, a DNA polymerase (either a TLS or a replicative polymerase) would be able to insert a nucleotide across from the monoadduct (Figure 1.11.4). However, extension beyond the monoadduct, which requires the use of a distorted primer-terminus (caused by the incompatibility between the 3'-terminal nucleotide and the template base carrying the monoadduct), most likely will not be

[34] Although FAN1 is also able to process an ICL-attached oligonucleotide, the remaining adduct is 3 nucleotides in size (Takahashi et al. 2015). Since FAN1 leaves a 3-nucleotide adduct which probably requires further processing, FAN1 is likely involved in the unhooking step and not in the processing of the remaining adduct.

possible by a replicative polymerase. This can result in stalling of the replicative polymerase and ubiquitylation of the PCNA which recruits TLS polymerases to the site of damage (Plosky et al. 2006). TLS polymerases are then able to extend the DNA from a terminal mismatch and fill the gap (Figure 1.11.4). Although several TLS polymerases, including rev1, pol ζ, pol η, and pol κ, have been implicated in RIR, it is not clear which one performs this gap-filling step (Enoiu, Jiricny, and Scharer 2012; Williams, Gottesman, and Gautier 2012; Sarkar et al. 2006). It is also possible that several TLS polymerases have redundant functions and are able to perform the extension step. Finally, after gap filling likely by a TLS polymerase, ligation to the other side finishes the repair. The remaining monoadduct is believed to be removed by NER (Figure 1.11.5).

Despite its potentially crucial role in postmitotic cells, such as heart muscle cells and neurons, as well as cells with deficient replication-dependent ICL repair pathways (e.g., Fanconi anemia cells), the detailed molecular mechanism of replication-independent ICL repair pathway(s) has remained largely enigmatic. In addition, structurally different ICLs can be detected and repaired by different replication independent mechanisms (and different efficiencies) adding to the complexity of studying this mode of repair (Hlavin et al. 2010; Yamada et al. 1997; Smeaton et al. 2009; Roy et al. 2016).

Figure 1.11 Replication-independent interstrand crosslink (ICL) repair mechanism

1) In the absence of replication, interstrand crosslinks (ICLs) can be detected either by stalling of the transcription machinery or by factors that detect DNA distortion. 2) After detection, incisions at either side of the ICL unhook the crosslink. This leads to the creation of a gap in one strand and the liberation of the ICL, which stays attached to the other strand as an adduct. 3) Digestion of the ICL adduct to a mono-adduct is likely necessary for allowing the DNA polymerases to use this strand as template. 4) DNA polymerases are loaded onto the 3' of the gap and fill in the region using the strand carrying the mono-adduct as template. This step likely requires translesion polymerases, as proper base pairing with the nucleotide carrying the adduct is not possible. 5) The remaining adduct is thought to be removed by the nucleotide excision repair (NER) pathway.

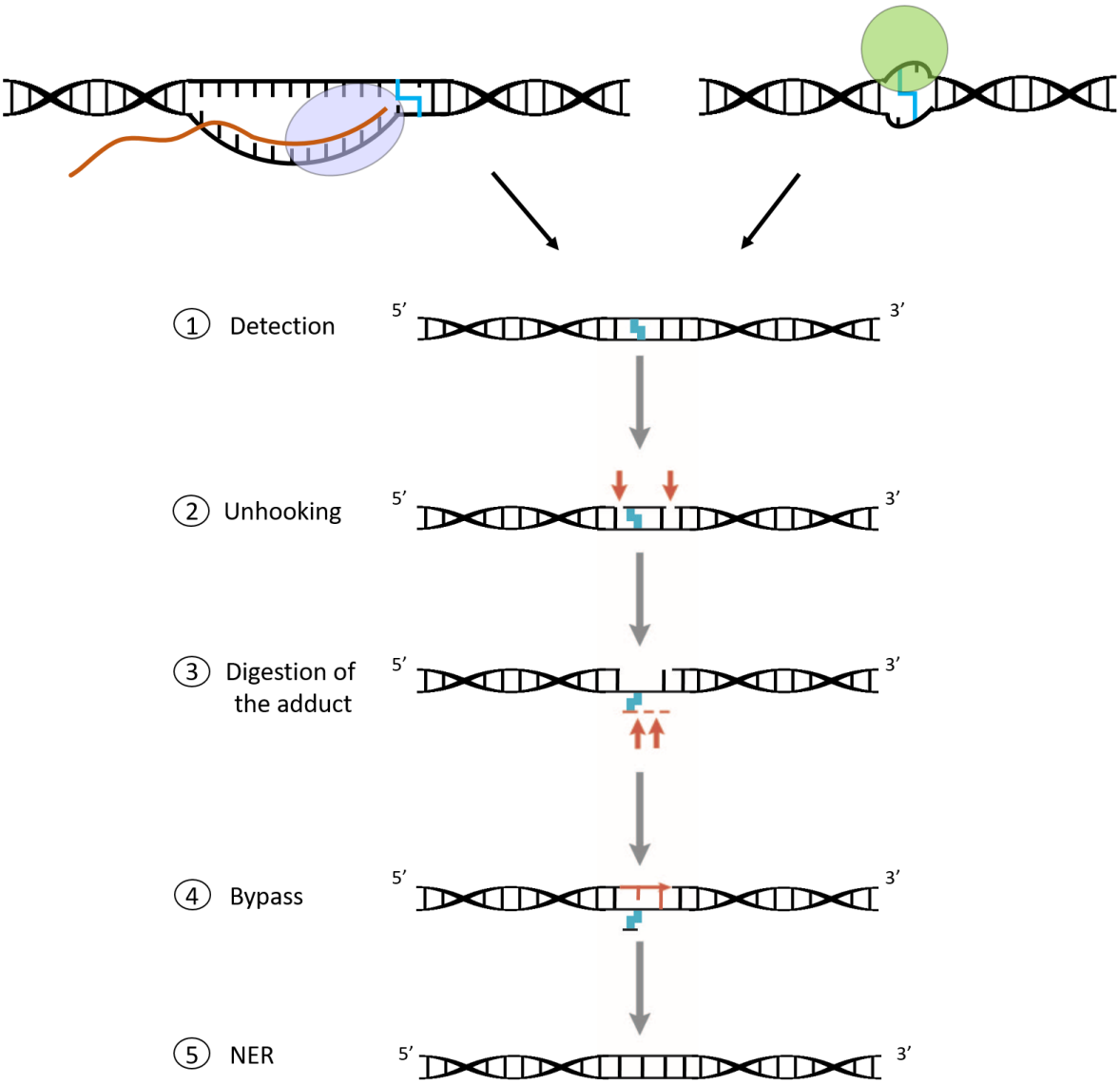

Figure 1.11 (Continued)

DNA-protein crosslink repair

Almost all DNA processes (e.g., replication, transcription, regulation and packaging) are performed by proteins. Exogenous and endogenous crosslinking factors, such as ionizing radiation (e.g., ultraviolet light, X-rays and γ-rays) and chemical crosslinking factors (e.g., aldehydes, heavy metal ions and chemotherapeutics), can covalently attach these DNA associated proteins to the DNA (Nakano et al. 2017; Xie et al. 2016; Ewig and Kohn 1977). In addition, DNA processing enzymes that form covalent bonds with DNA during intermediate reaction steps (such as topoisomerases, DNA methyltransferases, DNA polymerases and glycosylases) can get trapped in these intermediate steps and create DPCs (Pommier et al. 2014; Liu et al. 2003; Prasad et al. 2014; Quinones and Demple 2016). The steric hindrance of these large DNA-protein crosslinks (DPCs) prevents essential cellular functions such as DNA replication, transcription and even DNA repair. Therefore, if not repaired these frequent DNA lesions can be extremely toxic.

DPCs inhibit different DNA processes with different efficiencies, which depend on the structure and strength of the DNA processing machinery, as well as the size and orientation of the DPC. For example, replicative helicases are inhibited by DPCs larger than 14.1 kD *in vitro,* suggesting that their central channel is large enough to allow the passage of small DPCs (Nakano et al. 2013). In addition, while replicative DNA polymerases are completely inhibited by DPCs *in vitro*, translesion DNA polymerases η, κ, ν and ι can replicate over small 10 amino acids (but not larger) DPCs (Novakova et al. 2003; Nakano et al. 2012; Yeo et al. 2014; Yamanaka et al. 2010)[35]. Interestingly, the T7 RNA polymerase does not get completely blocked by DPCs no matter the size of the DPC but becomes very error prone after bypassing the DPC (with 40–75% of the

[35] The complete inhibition of plasmid replication by a DPC containing a DNA methyl transferase, histone fragments, or a T4 endonuclease V in *E. coli* cells suggest that DPCs completely inhibit replication forks *in vivo* (Nakano et al. 2007; Kuo, Griffith, and Kreuzer 2007; Kumari et al. 2010).

transcripts containing mutations) (Nakano et al. 2012). It is thought that the trailing RNA polymerases help the leading RNA polymerase bypass the DPC by a factor of 5.2 – 17 (Nakano et al. 2012). Finally, in all the above-mentioned scenarios, the DPC is located on the translocating strand. DPCs located on the non-translocating strands have not been observed to cause any hindrance to the DNA processing machinery.

Due to the heterogeneous nature of DPCs, various mechanisms have been evolved for their repair. For example, common DPCs caused by enzymatic reactions such as trapped topoisomerases have dedicated repair mechanisms. In the case of topoisomerases, tyrosyl DNA phosphodiesterase (TDP) relieves the trapped topoisomerase by cleaving the bond between the tyrosine residue in its active site and the nicked DNA (Pommier et al. 2014). Moreover, enzymatic DPCs located at the DSBs (e.g., meiotic recombinases) are removed during the resection of DSB ends by MRE11-CtIP (Hartsuiker et al. 2009). In addition to DSB repair, NER has been shown to remove the majority DPCs, especially DPCs located on the template strand of transcribed genes (Nakano et al. 2007; Fornace 1982; Stingele et al. 2014). However, NER is only able to remove small DPCs (< 11kDa) (Nakano et al. 2009; Baker et al. 2007). Therefore, it is thought that first either proteasome or Spartan (SPRTN, which is a DPC dedicated protease), processes larger DPCs that have been marked by ubiquitin[36], making them suitable for removal by NER (Baker et al. 2007; Maskey et al. 2017; Morocz et al. 2017). This proteolysis processing of DPCs is part of the replication-coupled DPC repair pathway that is initiated by the collision of the replication fork with the DPC (Stingele and Jentsch 2015).

Similar to ICLs, DPCs can be detected and processed after their collision with a replication fork (Figure 1.12.1). This collision, initiates the proteolysis of the DPC either by proteasome or

[36] Marking the DPCs by ubiquitin is likely a necessary strategy for avoiding the unwanted proteolysis of DNA associated proteins (Stingele and Jentsch 2015).

SPRTN, which is recruited to the DPC by ubiquitinylated PCNA (Figure 1.12.2). Proteolysis of the DPC to a small peptide allows the bypass of the CMG helicase (Figure 1.12.3). However, the replicative DNA polymerases still are not able to pass the DPC (Figure 1.12.4) and require the assistance of TLS polymerases (Figure 1.12.5) (Duxin et al. 2014; Centore et al. 2012). After bypassing the peptide, replicative polymerases replace the TLS and complete the DNA replication (Figure 1.12.6). The remaining peptide is then thought to be removed by NER (Figure 1.12.7). It is thought that if the proteolysis fails, cleavage or collapse of the stalled replication fork redirects the repair to HR (Stingele and Jentsch 2015). Interestingly, Fanconi anemia pathway seems to be involved in DPC tolerance and repair especially in cases where the DPC prevents DNA unwinding (e.g., DPCs that encircle the DNA or are crosslinked to both strands) (Rosado et al. 2011; Langevin et al. 2011; Garaycoechea et al. 2012). However, the details of this pathway remain enigmatic and deserve further characterization, especially since many chemotherapeutics induce DPCs.

Figure 1.12 DNA-protein crosslink (DPC) repair mechanism

1) Similar to ICLs, DPCs can be detected upon collision with a replication fork. 2) After detection, the DPC is degraded by a protease (proteasome or SPRTN). 3) This proteolysis leaves a small peptide attached to the DNA. 4) Although, the replicative helicase (CMG) is capable of bypassing the adduct, the replicative polymerases (pol δ/ε) are not. 5) To prevent fork stalling and collapse, translesion polymerases (TLS), which are capable of synthesizing past the adduct, are recruited. 6) After bypassing the lesion, pol δ/ε replace the TLS polymerases, and 7) complete the DNA replication. 8) The leftover peptide adduct is later removed by NER.

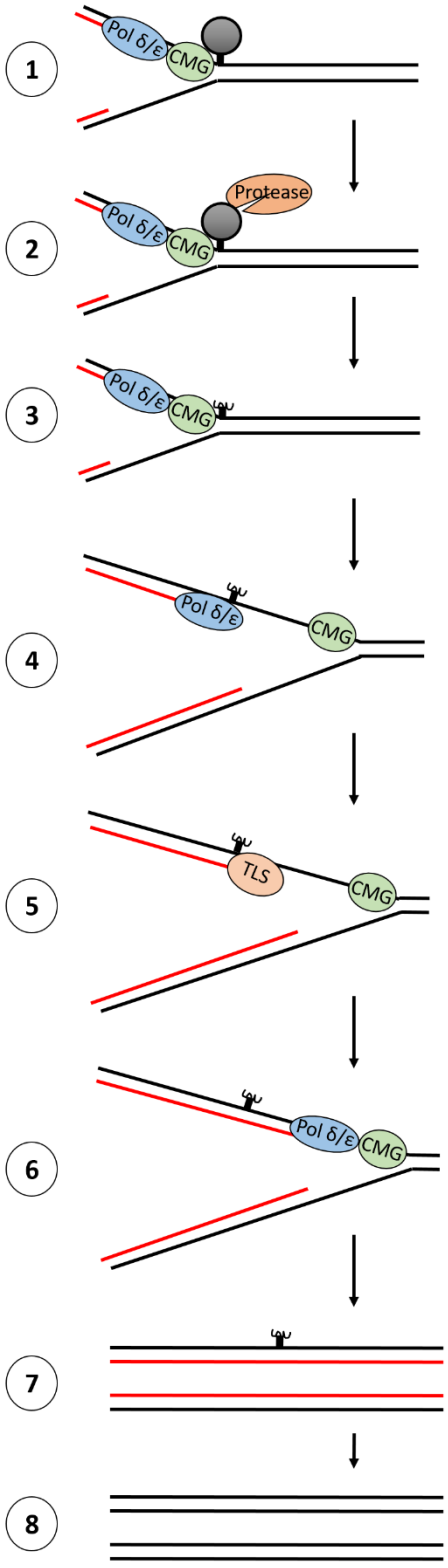

Figure 1.12 (Continued)

Clustered DNA damage

Clustered DNA damage are defined as multiple DNA lesions (mostly abasic sites, breaks, oxidized bases or their combination) located within one or two turns of the DNA helix (Ward 1994; Goodhead 1994). In the 1980s, separate studies by Ward and Goodhead brought the importance and deleterious effects of these complex DNA damage to light (Ward, Blakely, and Joner 1985; Goodhead et al. 1980; Ward 1981). Although these damages can have endogenous causes such as oxidative stress, they have a very low likelihood of arising endogenously and are considered a signature of ionizing radiation (Eccles, O'Neill, and Lomax 2011; Goodhead 1994; Georgakilas et al. 2004; Bennett et al. 2005).

Based on their composition, clustered DNA damage can be divided into two groups of complex double strand breaks (DSBs) and non-DSB clusters. As discussed the previous sections, the DSBs can be repaired either by NHEJ or HR. However, the presence of other DNA lesions in the vicinity of the DSBs can lead to lower efficiency and accuracy of the repair. On the other hand, non-DSB clusters constitute more than 80% of the CDDs (Box, Dawidzik, and Budzinski 2001; Lomax, Gulston, and O'Neill 2002; Douki, Riviere, and Cadet 2002). Since most of the non-DSB clustered damages affect DNA bases, BER usually plays an important role in their repair. However, its efficiency has been shown to be highly dependent on the spatial organization of damages (Barzilay and Hickson 1995; Nilsen et al. 2012; Lomax, Cunniffe, and O'Neill 2004). For examples, some CDDs are resistant or very slow to repair, because their configuration inhibits glycosylates or nucleases (Tsao et al. 2007; Georgakilas et al. 2004). In these cases, the persistence of the DNA damage can lead to mutations, fork collapse, chromosomal breakage, and cell death. In addition, based on the type and location of these damages other pathways have also been observed to get involved. For example, if the damages are located on opposite strands their repair can lead to double strand breaks which require NHEJ or HR for repair (Mourgues, Lomax, and O'Neill 2007; Singh and Das 2013). Even if the damages on the opposite strands do

not lead to double strand breaks, they will force the DNA polymerases (usually required in these repair pathways to replace the excised damaged region) to use a damaged template. As a result, it has been suggested that the translesion polymerases play a role in the repair of this type of damage. However, in the absence of an intact template, the introduction of mutations at the region is inevitable (Belousova and Lavrik 2015).

In these clusters, each lesion can affect the detection and repair of the nearby lesions by altering the structure and chemistry of DNA. Therefore, in addition to the types of damages, their spatial organization (i.e. their orientation and distance) can greatly affect the outcome (Georgakilas, Bennett, and Sutherland 2002; Semenenko, Stewart, and Ackerman 2005; Sage and Harrison 2011). In addition, the genome level distribution of damages can affect the availability of repair factors, as well as the accuracy of repair especially for DSB (Rydberg 1996; Hlatky et al. 2002; Holley and Chatterjee 1996). CDDs are thought to be repaired simultaneously or sequentially by multiple repair pathways. Since each step can change the structure of the DNA (which will affect the efficiency of the subsequent steps), the repair outcome of the clusters is highly dependent on the first responders that set the path for the repair, as well as the order of recruitment of different factors (Belousova and Lavrik 2015; Singh and Das 2013). Although recently several studies have started to tackle the repair problem of CDDs, many aspects of their repair are still enigmatic. As radiation therapy (e.g., for cancer treatment) is becoming more common, understanding the details of these repair pathways is necessary to ameliorate the adverse effects of this treatment on patients.

Discussion

Despite the significant progress we have made in characterizing different pathways that have evolved for the identification and repair of different forms of DNA damage, there are still many unknowns in the field. As we explore the new frontiers of the field, we are realizing more and more that DNA damage repair pathways are not as rigid as once imagined. In fact, they are much more fluid; they can overlap and collaborate for identifying and repairing different damages in the cell with many factors that were once thought to be involved in one repair pathway, now being realized to play roles in multiple pathways. This is especially apparent in the case of complex DNA damage such as DNA interstrand crosslinks and DNA protein crosslinks.

In addition to their primary role in DNA repair, many repair pathways have additional roles that go beyond genome maintenance in living cells (Figure 1.13). For example, BER is also involved in regulating gene expression, RNA metabolism (e.g., many repair proteins including APE1 have been shown to interact with various RNA species), and immunoglobulin class switch, as well as mediating epigenetic marks (e.g., erasing DNA methyl marks). In addition, HR allows the exchange of information between chromosomes during meiosis, V(D)J recombination during the development of adaptive immunity and immunoglobulin heavy chain class switch in mature B lymphocytes (Neale and Keeney 2006; Alt et al. 2013). NHEJ has also been shown to be important in immune system development, neurodevelopment and memory formation (Ferguson and Alt 2001; Frank et al. 1998; Gao et al. 1998; Sekiguchi et al. 1999; Suberbielle et al. 2015). Therefore, in contrast to their primary role in guarding the information in the genome, they can be leveraged by the cells to manipulate this information.

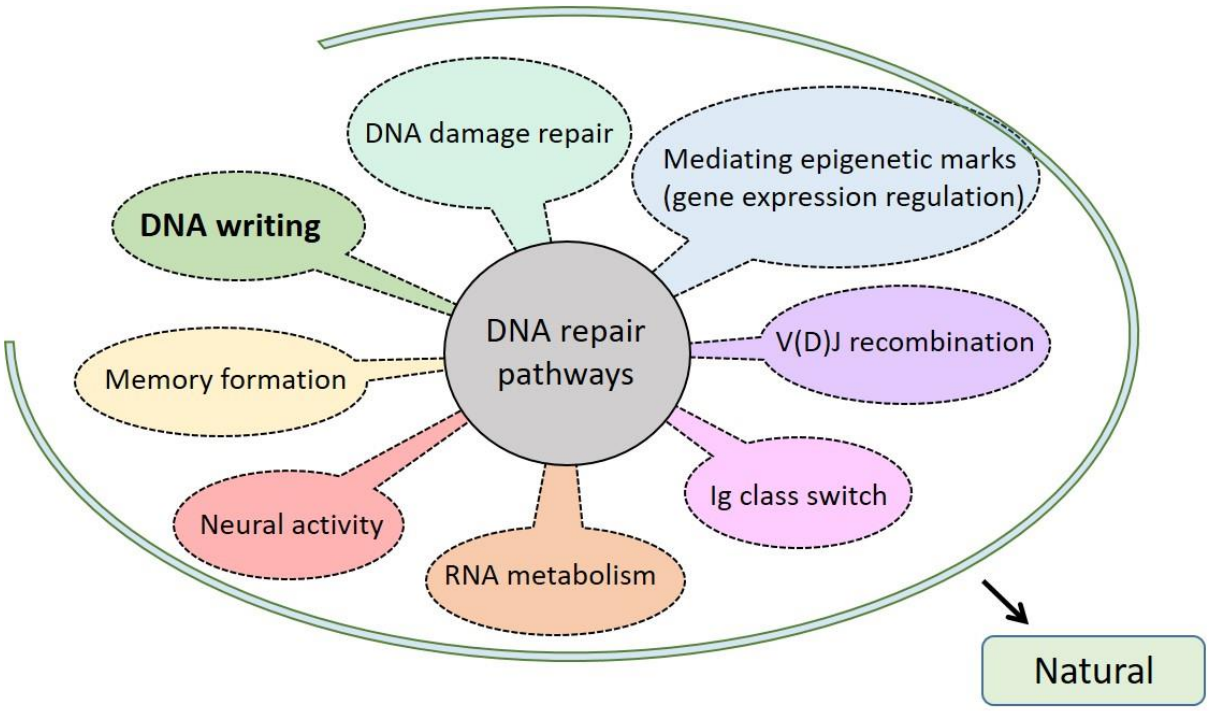

Figure 1.13 Other functions and applications of DNA repair pathways

Although the primary role of DNA repair pathways is guarding the information stored in the DNA, they are also involved in various other cellular processes. Interestingly, cells have developed strategies to convince these pathways to help manipulate the information stored in the genome or epigenome in much faster time scales than allowed by evolution (e.g., during processes such as mediating epigenetic marks, V(D)J recombination, Ig class switch, RNA metabolism, controlling neural activity and memory formation in the brain). Inspired by these natural roles of DNA repair pathways, in recent years we have begun to develop technologies for writing and recording artificial information such as biological and environmental cues – that are not naturally stored in the genome – into the genomic DNA (discussed in the second chapter).

Notably, recent findings point to the DSB formation and repair as an important factor in brain development, physiology and function (Wei et al. 2016; Wei et al. 2018; Madabhushi et al. 2015; Suberbielle et al. 2013; Crowe et al. 2006). In fact, neural diversity of the brain (such as the genomic copy number variation in the frontal cortex) is thought to be, at least partly, due to the repeated DSBs and genetic modifications in mature neurons, as well as during neurodevelopment (McConnell et al. 2013; McConnell et al. 2017; Poduri et al. 2013; Wei et al. 2016).[37] Interestingly, these events happen especially in genes involved in synaptogenesis and neural functions, which have been associated with autism, schizophrenia and brain cancers (McConnell et al. 2013; Northcott et al. 2012; Rausch et al. 2012; Frattini et al. 2013; Wei et al. 2018). For example, activity-induced DSBs have been shown to activate early response genes which are mostly involved in memory formation and learning (Madabhushi et al. 2015; West and Greenberg 2011; Cholewa-Waclaw et al. 2016; Suberbielle et al. 2013). As these DSBs can lead to the accumulation of mutation in these regions over the life span of individuals, they might be the reason behind aging-associated cognitive decline (Alt and Schwer 2018). However, their exact repair mechanism in neurons is still unknown. Discovering how these DSBs are controlled and repaired might hold the key to understanding brain functions and why it malfunctions in neurological diseases (that affect over a billion people worldwide) such as Alzheimer's and autism. This can potentially help design effective diagnosis and treatments tailored to each specific condition.

It is now a century past from the discovery of DNA as the keeper of the blueprint of life. We have come a long way from the time that the existence of DNA repair pathways and the importance of studying DNA damage were faced with skepticism. We now know that we are

[37] These DSBs, which have a variety of sources such as neural activation, oxidative stress, replication and transcription, have been suggested to lead to exon shuffling or alternative splicing (Alt and Schwer 2018).

constantly dealing with the adverse effects of DNA damage, from simple sunburns to devastating diseases such as cancer, Fanconi anemia, Xeroderma pigmentosum, and Alzheimer's. Discoveries in the field of DNA repair has helped us design treatments that can ameliorate the symptoms of different diseases caused by DNA repair defects. However, we still have a long way to go before we can completely understand these repair pathways and use our knowledge to design treatments for curing these diseases.

Chapter 2. Bio-molecular recording and bio-computation

Introduction

Over the course of evolution, genomic DNA has emerged as the unit of information storage in living cells. Since extensive changes could have catastrophic effects on a cell's fitness and survival, the information encoded in this medium is guarded by various repair pathways. Therefore, with a few exceptions (e.g., during B-cell and T-cell receptor maturation in immune cells), natural mutations occur very infrequently during the lifespan of a cell. Over longer timescales, however, significant levels of mutations could accumulate in a genome due to errors in replication and repair, as well as recombination events. These mutations serve as the substrates for evolutionary processes and adaption (with some of them stabilizing in the genome of organisms over many years). Thus, in a broad sense, genomic DNA can be considered as a read-only biological information storage medium in short timescales, and a read-and-write storage medium in evolutionary timescales.

Much like a hard drive, a genomic read-and-write storage medium can be used to capture biological events. Indeed, the mutations that have accumulated over evolutionary timescales in genomic DNA of different species have been traditionally used as molecular records for inferring the evolutionary history of species (Woese and Fox 1977). This further highlights the power of genomic DNA as a medium to capture biological events. Advances in molecular biology and genome engineering technologies have made it possible to introduce targeted mutations (i.e., write) into the genomic DNA. In the traditional genome engineering technologies, DNA writing was coupled with the delivery of genetic materials into the cells, and as a result, it was not possible to couple DNA writing with biological events (e.g., changes in transcription). The more recent generations of DNA writing technologies, however, enable more dynamic control over DNA writing operations (Farzadfard and Lu 2014; Cong et al. 2013; Mali et al. 2013; Farzadfard et al. 2018; Perli, Cui, and Lu 2016; Roquet et al. 2016; Komor et al. 2016). These DNA writing technologies, which are often inspired by and exploit natural DNA writing systems, open up the possibility of

using genome as a dynamic information storage medium in living cells in timescales much shorter than previously was possible (e.g., within one generation). Additionally, DNA writing events enabled by these technologies can be coupled to transient biological events of interest allowing the permanent storage of that information in the form of mutational signatures in the DNA, thus expanding the types of information that can be recorded and the breadth of applications that can be achieved by these technologies.

Dynamic DNA writing technologies rely on various forms of addressable DNA damaging/modifying enzymes that can introduce targeted mutations into the genomic loci of interest. The two processes of DNA writing (adding new information to genomic DNA) and DNA repair (removing erroneous information) often work on opposite directions. Any mutation that is introduced by DNA writers to a genome could potentially be identified by the cellular repair machinery and removed. The dynamic interplay between these DNA writers and the cellular repair machinery determines the mutational outcomes of these processes and the overall performance of DNA-based recording. Thus, the processes of DNA writing and DNA repair are deeply interconnected; on one hand, better understanding and engineering repair pathways could help us to design and implement better DNA writing systems, and on the other hand, by offering to introduce targeted and precise lesions into a genome, DNA writing systems could provide valuable tools to study repair mechanisms in living cells (Figure 2.1).

In this chapter, I will discuss recent advances in the dynamic DNA writing technologies and their applications for information storage and processing in different contexts and present example routines that can be used to retrieve the recorded information. I also discuss various strategies that can be used to better control the outcome of DNA writers and improve their performance by engineering cellular repair machinery. Finally, I discuss the current limitations and future directions of these technologies and postulate on the avenues at the interface of dynamic genome engineering and DNA repair that can be explored to improve the performance

of DNA writers and molecular recorders and pave the way towards building sophisticated genome surgery technologies for biotechnological and biomedical applications.

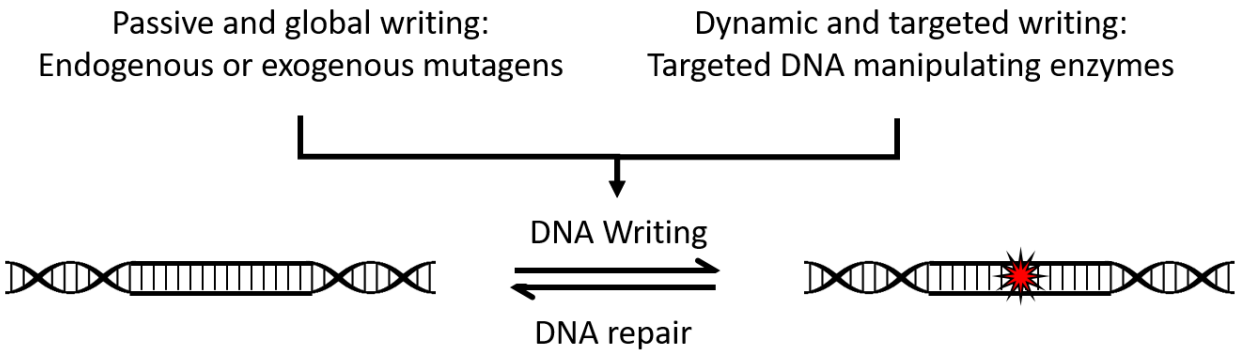

Figure 2.1 The interplay between DNA writers and DNA repair mechanisms

Information is added to the DNA in the form of mutations either passively and globally (though endogenous or exogenous mutagens) or triggered in an active and targeted manner via specialized enzymes. DNA repair machinery tries to minimize these often erroneous information (i.e., mutations) and reverse them back to the original sequence. Thus, encoding information via DNA writing requires careful engineering and tuning of DNA writing and repair processes.

Writing on the DNA of living cells and its applications

While the majority of mutations in living cells occur passively due to unregulated mechanisms (e.g., errors in replication and repair), examples of active and targeted mutagenesis can be found in nature. For example, natural systems that benefit from DNA writing include CRISPR spacer acquisition and phase variation in bacteria, and B-cell and T-cell receptor maturation in the vertebrate immune system (Sternberg et al. 2016; van der Woude and Baumler 2004). In all these cases, a confined genomic locus is actively mutagenized within a very short time period (e.g., less than one cell generation) using dedicated and addressable DNA modifying enzymes and repurposed repair factors. These processes often give an organism the ability to generate large genetic diversity that helps the organism add valuable information to its genome (without compromising the risk of elevated mutations in its entire genome), which in turn helps the organism to adapt faster than possible by natural mutation.

By engineering and repurposing the natural DNA writers, bioengineers have started to utilize genomic DNA as a dynamic medium for information processing and storage in living cells. The ability to dynamically write on the DNA of living cells has profound implications that go beyond what natural DNA writers have been evolved to do and makes many biotechnological and biomedical applications possible. For example, DNA writing can be used to record various types of transient biomolecular information, such as signaling events, interactions, and lineage progression into the DNA. In these technologies, which collectively are referred to as molecular recording, a transient signal (e.g., transcriptional activity) or an event of interest (lineage progression) is linked to the activity of a DNA writer which captures and permanently records the signal in the form of targeted mutations in the DNA. These mutational signatures, generated by the DNA writer as a result of the linked activity, can then be retrieved by sequencing and used to infer the recorded information (Figure 2.2). As such, molecular recording technologies could help to better study and understand how living cells sense and process internal and external

information over time, a task that has remained challenging owing to the transient nature of biological events. The DNA writing technologies could also be used to rationally program cellular behavior and engineer cells that dynamically respond to information that they receive (e.g. concentration of small molecules) in desired and predictable ways (Figure 2.3). Such control over the order and timing of cellular responses could have broad biomedical and biotechnological applications, from the study of human disease conditions to engineering cell-based therapeutics.

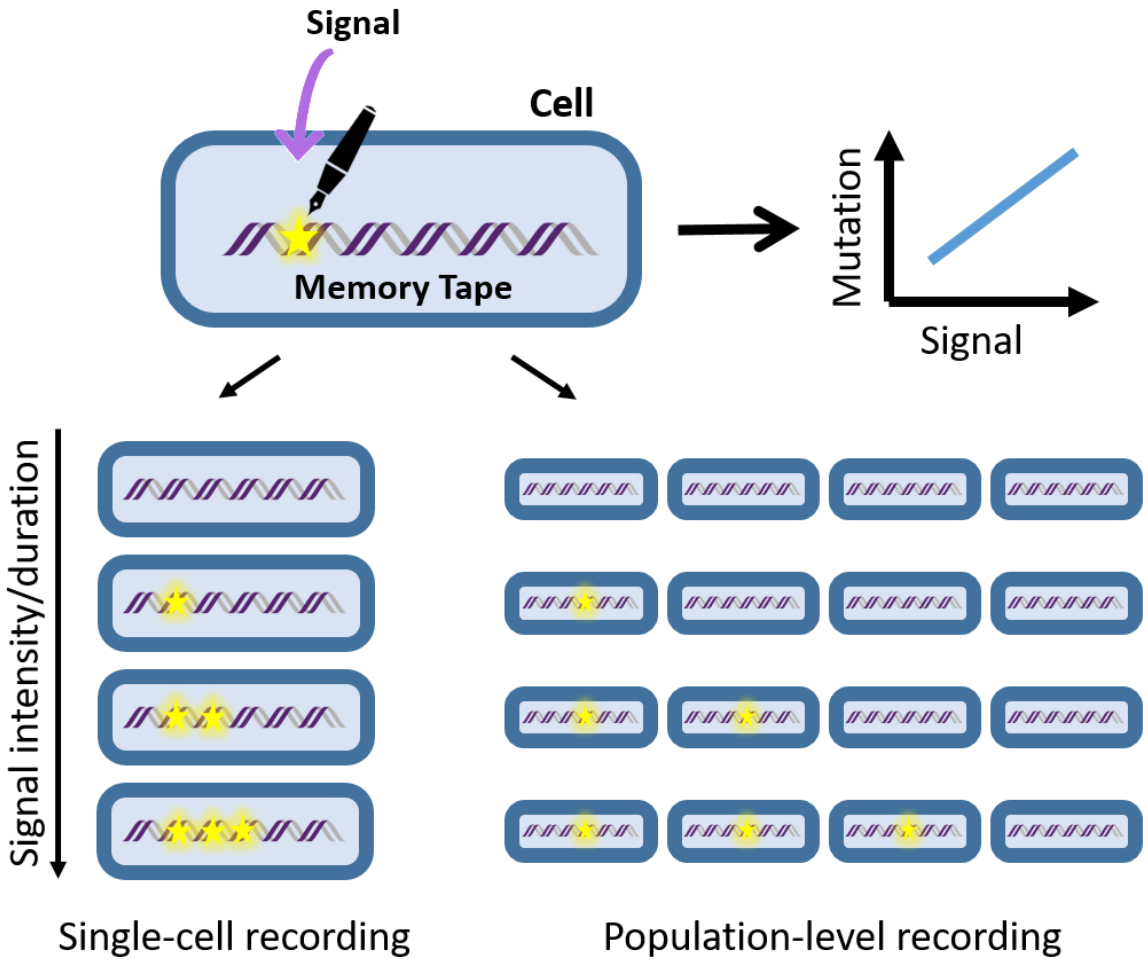

Figure 2.2 Using genomic DNA as a medium for storing biological information

A desired signal can be coupled with a dynamic DNA writer, which introduces mutations into genomic DNA as a function of the input signal. The accumulated mutations in a single-cell (bottom left) or a cell population (bottom-right) can be retrieved by sequencing and used to infer the dynamics of the original signal.

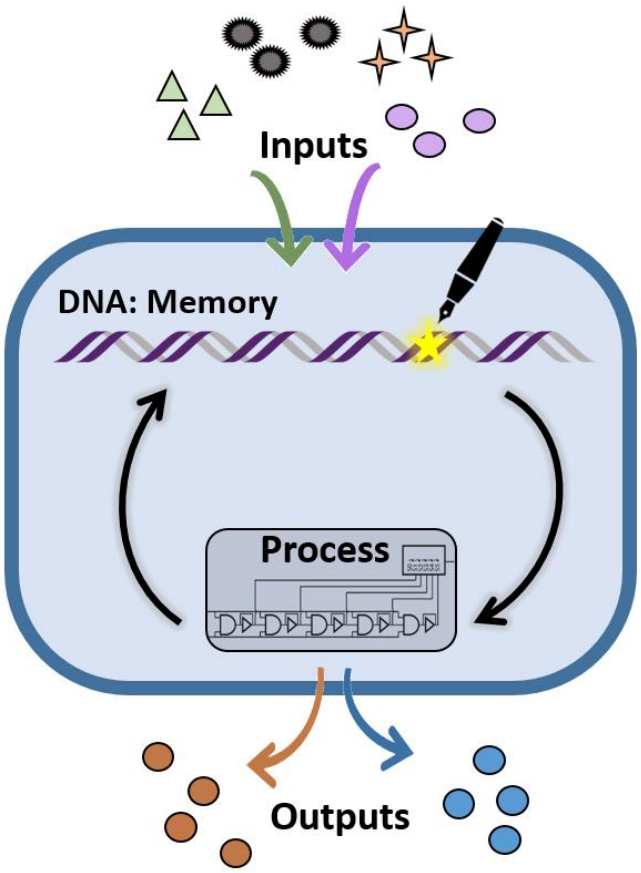

Figure 2.3 Using genomic DNA for information processing and computation in living cells

In addition to molecular recording, genomic DNA can be used as a medium to integrate multiple recorded signals and generate appropriate phenotypic responses (outputs). This capacity enables us to implement various forms of computation and logic in living cells to control the sequence and timing of molecular events (e.g., gene expression) in living cells based on combination and order of the inputs that the cells receive.

Based on their mode of action and dependence on templates for writing, DNA writers can be categorized into two classes: template-dependent and template-independent DNA writers (Figure 2.4). Template-dependent DNA writers are DNA-modifying enzymes that can incorporate (i.e., write) information encoded on templates (e.g., dsDNA, ssDNA, or RNA) into the genomic DNA. On the other hand, template-independent DNA writers are capable of introducing *de novo* targeted mutations into the genomic DNA without the requirement for any templates. Some of these DNA writers heavily rely on cellular DNA replication and repair machinery, while others can achieve DNA writing without the requirement for any of the host-encoded factors. Several technologies based on each of these two classes of DNA writers have been described so far. Each of these technologies offers a distinct set of features and dependencies on host-encoded factors, and thus are suitable for certain applications. Here, I review the current DNA writing technologies for each of these two classes of DNA writers and discuss their modes of action, host-factor dependencies, features, and potential applications.

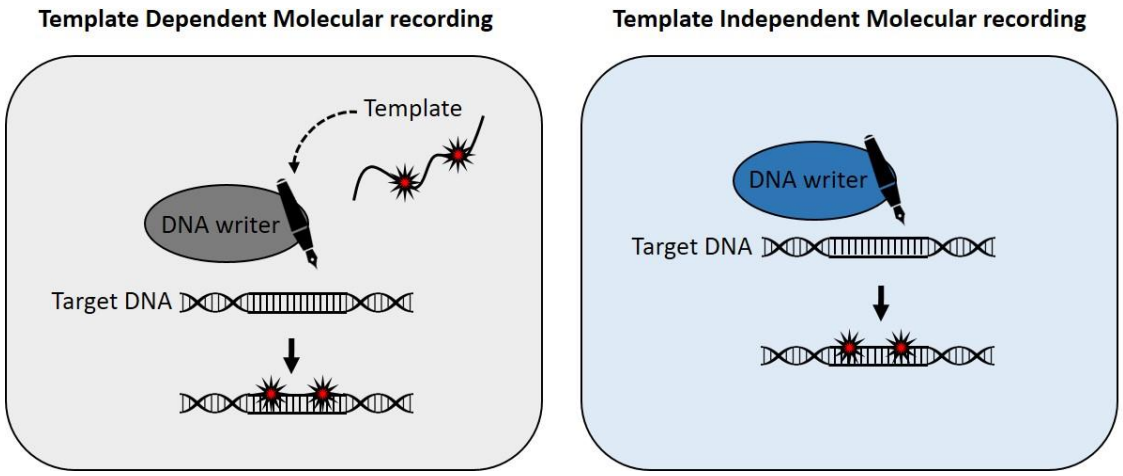

Figure 2.4 Classes of DNA writers based on their mode of action

Template-dependent DNA writers (left) enable writing of information from a trans-encoded oligonucleotide into a target DNA. On the other hand, template-independent writers can initiate *de novo* mutations in DNA without a requirement for templates.

Template-dependent molecular recording

Various strategies for template-dependent DNA writing and molecular recording have been described so far. In these strategies, an (often exogenously-provided) oligonucleotide (i.e., RNA, ssDNA or dsDNA) is provided in-trans to serve as a template for modification of the locus of interest. Although the majority of template-dependent DNA writers employ homologous recombination for site-specific incorporation of the exogenous sequence into the target locus, there are other mechanisms that exploit cis-encoded elements on the target to gain site-specificity. Depending on the type of DNA writer, the writing events could result in predetermined or random mutational outcomes in the target DNA.

Retroelement-based DNA writers

Retroelements are genetic elements that can amplify via reverse transcription of RNA intermediates (Boeke and Stoye 1997). Retroviruses and retrotransposons are examples of retroelements in nature. DNA writing by retroelements often involves three major steps: transcription of a template DNA into RNA, reverse transcription of the RNA into ssDNA, and an optional step involving the writing of ssDNA into the genome either non-specifically or in a targeted fashion which can be based on sequence homology (via homologous recombination) or via target-specific genetic elements (e.g., retrohoming). Various stages of the process can be potentially regulated by placing the components of the system under the control of inducible promoters, thus allowing to perform conditional writing and molecular recording. Several DNA writing systems based on retroelements have been described so far. These DNA writing systems take advantage of a bacterial class of retroelements called retron for intracellular production of ssDNAs. The produced ssDNAs are then incorporated into a specific genomic locus based on homology via recombineering (e.g., SCRIBE (Farzadfard and Lu 2014)) or CRISPR-assisted technologies (e.g. CRISPEY (Sharon et al. 2018)).

SCRIBE

By leveraging a bacterial class of retroelements called retron, Farzadfard et al. developed a synthetic system for DNA writing and molecular recording in bacteria (Farzadfard and Lu 2014). In this system, the retron cassette is composed of three components: a reverse transcriptase (RT), a template for reverse transcription (msd) and a primer RNA (msr). These factors work together to produce a hybrid RNA-ssDNA molecule called multi-copy single-stranded DNA (msDNA) *in vivo* (Lim and Maas 1989). Natural msDNA is expressed as an episome and does not integrate into the bacterial genome. Thus, it cannot be considered a DNA writer. To build a reverse transcriptase-mediated DNA writer, the authors first engineered the msd template sequence and demonstrated that the middle part of msd can be removed to express ssDNAs of interest. The authors then took advantage of a specific recombinase (called beta-recombinase), which was previously demonstrated to promote recombination of synthetic ssDNAs to their homologous genomic sites through a process referred to as recombineering. They demonstrated that the intracellular expression of ssDNA by the retron cassette followed by beta-mediated recombineering can be used to write the information encoded in the retron template into its homologous genomic loci. In this process, the writing of ssDNA into its target occurs solely based on sequence homology without the need for additional target-encoded genetic elements (Figure 2.5).

By placing the expression of the engineered retron and beta-recombinase cassettes into a synthetic operon and coupling the expression of this synthetic operon with an inducible promoters, the authors demonstrated that the activity (i.e., intensity and duration of expression) of that promoter can be recorded in the DNA in the form of precise mutations that accumulate in a target genomic loci in an entire cell population. As such, this system (termed SCRIBE for Synthetic Cellular Recorders Integrating Biological Events) offers a molecular recorder with a wide-dynamic-range. Using this system, the authors demonstrated that analog information such as

signal intensity and duration can be recorded into permanent DNA records within cell populations. Since the ssDNAs can be targeted based on homology, the SCRIBE system can potentially target and use any genomic loci for recording. Thus high levels of storage capacity can be achieved. The written information can then be retrieved by sequencing. Alternatively, DNA writing can be targeted to functional outputs and reporter genes to achieve functional readouts (e.g., different cellular phenotypes).

In a follow-up study, Farzadfard et al. (in preparation) further improved the DNA writing efficiency of the SCRIBE system by knocking out the exonucleases responsible for depleting the cellular ssDNA pool. They demonstrated that this improved system, termed HiSCRIBE (High-efficiency Synthetic Cellular Recorders Integrating Biological Events), enables editing *E. coli* genome with up to ~100% editing efficiency within bacterial communities. In addition, using barcoded libraries of HiSCRIBE, the author demonstrated that cell-cell interactions between bacterial mating pairs (spatial information) can be recorded in the DNA of interacting partners in the form of unique DNA barcodes. This strategy could be potentially used to map the organization of bacterial (and possibly other) cell communities in a high-throughput and high-resolution fashion. Finally, they demonstrated that random libraries of HiSCRIBE can be used to continuously diversify desired genomic loci connected to the phenotypes of interest. By linking this genetic diversity generation system to cellular fitness and an appropriate selective pressure, they demonstrated that this technology can be used to continuously tune a cellular phenotype of interest. HiSCRIBE offers a highly-efficient, scarless, and cis-element independent molecular recorder. However, it is only applicable to systems with highly efficient ssDNA-mediated recombineering such as some bacteria and yeast.

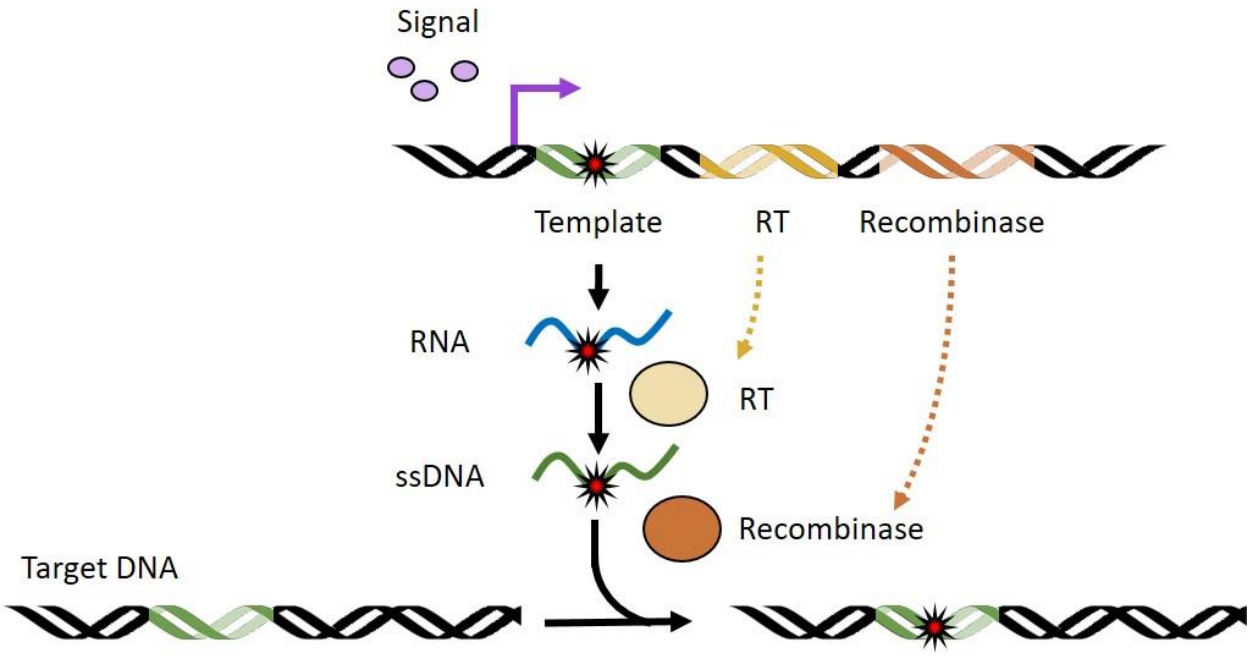

Figure 2.5 Mechanism of DNA writing by SCRIBE (an example of retroelement-based DNA writers)

Expression of the SCRIBE cassette (which is composed of the retron and beta recombinase) is placed under the control of a promoter responding to a signal of interest. At the presence of the signal, the template dsDNA (part of the retron cassette) is first transcribe into RNA which is then reverse transcribed into ssDNA via the retron-encoded reverse transcriptase (RT). The ssDNA is then recombined into its homologous genomic loci through the beta-mediated recombineering process, resulting in writing of the template-encoded mutation into the genome.

CRISPR-based DNA writers

Recent advances in dynamic genome engineering, especially those based on CRISPR technology (Cong et al. 2013; Mali et al. 2013), have ushered a new era in biology. CRISPR (clustered regularly interspaced short palindromic repeat) is an ancient prokaryotic adaptive immune system (that protects the cells against foreign DNA, such as viruses, plasmids, and other mobile genetic elements) and is found in more than 40% of sequenced bacteria and 90% of archaea (Horvath and Barrangou 2010). CRISPR system is mainly composed of two modules: a spacer acquisition module (required for CRISPR adaptation) and an effector module (required for destroying the foreign DNA). Many proteins that comprise these two modules are specialized DNA and RNA processing enzymes. As discussed below, proteins involved in spacer acquisition such as Cas1 and Cas2, and RNA-programmable effector proteins such as Cas9 and Cas12 have been engineered into powerful DNA writers for molecular recording in living cells. Ongoing efforts for characterization of other DNA and RNA processing protein families found in various classes of CRISPR systems hope to expand the possibilities of these systems. Owing to their programmability and other unique features, many of these proteins could offer valuable additions to the genome engineering and DNA (or RNA) writing toolbox.

CRISPR spacer acquisition

Spacer acquisition by the Cas1-Cas2 complex in the CRISPR adaptive immunity system is an example of natural template-dependent DNA writing. In this system, the Cas1-Cas2 complex is responsible for the expansion of the CRISPR array by sampling the intracellular ssDNA pool in bacteria and incorporating short DNA fragments (called spacers) into the CRISPR array (made of acquired spacers interspaced with the repeat elements). The spacer acquisition meditates adaptation through acquiring small pieces of invading DNAs and incorporating them into genomically encoded DNA array. Various sources could contribute to the intracellular ssDNA pool (e.g., DNA, plasmids, phages, and electroporated oligos). However, DNAs originated from

episomal sources (i.e., plasmids, phages, and oligos) are preferred. Thus, spacer acquisition serves as a form of adaptive immunity against invading DNA molecules.

In this system, Cas1-Cas2 complex preferentially and unidirectionally adds (writes) new spacers to the upstream side of the array. As a result, the chronological order of the addition of spacers is preserved within the array. Thus, the information encoded in the CRISPR array can serve as a form of memory and a molecular record of exposure of that cell to various DNA molecules (which are often originated from invading genetic elements). The spacers encoded in the CRISPR array locus can be transcribed and processed into small RNAs which can then guide the CRISPR effector modules (e.g. Cas9 protein) toward the original DNA sequence, thus providing some form of immunity to that invading DNA molecule.

Shipman et al. (Shipman et al. 2016) leveraged the memory formation characteristic of this natural DNA writing system to build a molecular recorder that could record the intensity and duration of a given signal in the length of a CRISPR spacer array. They achieved this, by placing the Cas1-Cas2 expression under an inducible promoter (for the desired signal) and providing the ssDNA pool as transformed oligonucleotides (Figure 2.6). Later, they demonstrated that arbitrary digital information such as a small digital file (here a 2.6 kilobytes animated image) can be encoded in the CRISPR array composition of bacterial populations (Shipman et al. 2017). To achieve this, they first encoded the binary information of the individual frames into pools of oligonucleotides and DNA barcodes. The pools corresponding to each frame were then sequentially introduced into a population of *E. coli* cells overexpressing the Cas1-Cas2 complex. The authors then sequenced the CRISPR array in the cells in bulk and showed that the spacers collectively recorded in the CRISPR locus of the entire population can be used to faithfully reconstruct the original movie (i.e., the individual frames and their chronological order).

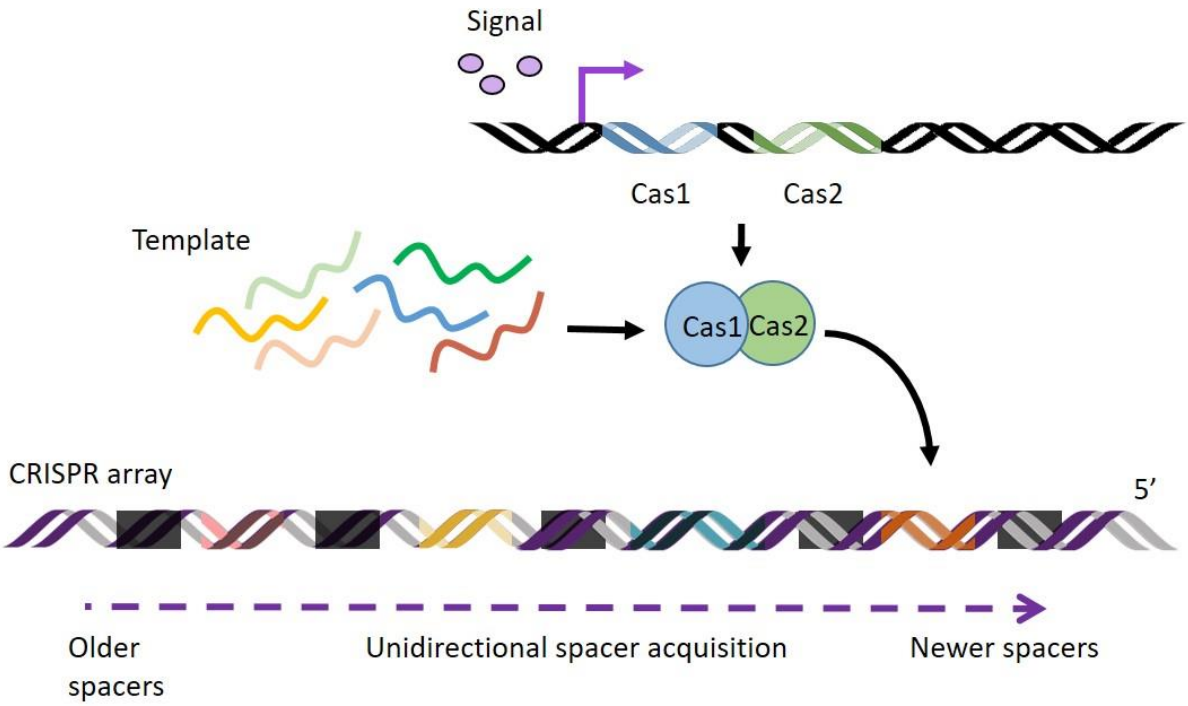

Figure 2.6 DNA writing by Cas1-Cas2 spacer acquisition system

Intracellular ssDNAs are captured by the Cas1-Cas2 complex and preferentially added to the 5'-side of the array, thus resulting in a temporal record of the intracellular ssDNA pool.

The abovementioned studies demonstrated the potential of CRISPR spacer acquisition system for molecular recording and information storage. However, these systems still rely on the exogenously-delivered templates for writing. To eliminate the need for exogenous ssDNA pools, Sheth et al. designed a system that allowed the modulation of the intracellular pool of ssDNAs by transcriptional signals (Sheth et al. 2017). To achieve this, they took advantage of copy-number variable plasmids. By placing the expression of the genetic element responsible for the copy number control of these plasmid under the control of an inducible promoter, they were able to dynamically modulate and control the copy number of the template plasmids and thus the intracellular ssDNA pool (as the increase in the copy number leads to the increase in plasmid degradation and natural plasmid assimilation). Using this system, termed TRACE (temporal recording in arrays by CRISPR expansion), they were able to record the temporal exposure history of up to three signals, as well as bacterial population lineage information in the composition of the CRISPR array.

In addition to the Cas1-Cas2 systems that can incorporate ssDNA into CRISPR array, some variants of Cas1 are fused to reverse transcriptase domains. It has been demonstrated that these RT-Cas1 fusions along with their cognate Cas2 proteins are able to reverse transcribe and incorporate fragments of cellular RNA into CRISPR array (Silas et al. 2016). Very recently, it was demonstrated that these systems could work in orthogonal bacterial hosts and are able to record the temporal dynamics of highly-expressed genes in the CRISPR array (Figure 2.7) (Schmidt, Cherepkova, and Platt 2018).

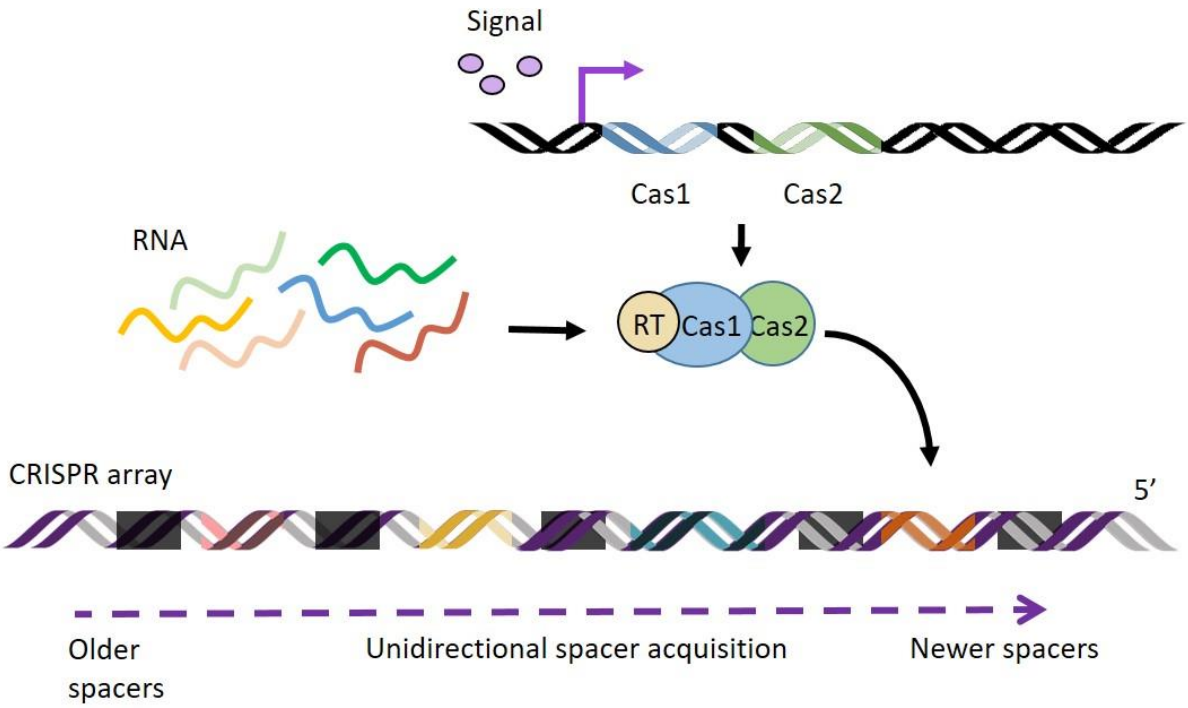

Figure 2.7 DNA writing by RT-Cas1-Cas2 spacer acquisition system

Intracellular RNAs are captured by the RT-Cas1-Cas2 complex, reverse transcribed to ssdNA and preferentially added to the 5'-side of the array, thus resulting in a temporal record of the intracellular RNA pool.

The spacer acquisition systems provide valuable tools for molecular recording, arbitrary information storage, and lineage tracing applications. In addition, the insertional nature of mutations introduced by these DNA writers offers a high recording capacity. Importantly, since spacers are added in a chronical order that can be accurately traced back, they are especially useful for lineage tracing applications, as well as recording temporal information (Sheth et al. 2017; Schmidt, Cherepkova, and Platt 2018). The capacity to record the dynamics of transcription in a highly-multiplexed fashion by the Cas1-RT fusion system could provide a valuable tool for the study of genome-wide transcription. Especially, unlike the current RNA-seq measurements that only provide snapshots in time, this system could be used to record dynamics of transcription in a highly multiplexed manner and provide longitudinal insight into biological processes (e.g., in disease conditions) from single time-point measurements.

While these proof-of-concept studies demonstrate the potential of CRISPR spacer acquisition systems for recording technologies, further development is needed to make these molecular recorders and DNA writers practical and useful for biotechnological and biomedical applications. For example, these systems are currently limited to prokaryotic hosts and their application to other hosts requires a better understanding of the extent to which spacer acquisition relies on the host proteins. Additionally, the stability of the recorded information (repeats) in other hosts (especially eukaryotic systems) needs to be carefully assessed and extended, especially for using these systems for long-term information storage. Furthermore, it remains to be determined whether the temporal recording resolution of these systems can be increased. Finally, to convert the Cas1-RT system into an ideal dynamic and multiplexed transcriptional recording device, the sources of bias of this system toward highly-expressed genes need to be investigated and minimized.

Cas9-based systems

Due to their efficiency, programmability, and ease of use in various organisms, the effector proteins of the CRISPR systems have been engineered into powerful DNA writing tools in molecular biology (Cong et al. 2013; Mali et al. 2013; Jinek et al. 2012). The Cas9 protein from *Streptococcus pyogenes* (SpCas9) is the first characterized and the most widely used member of the effector proteins in the CRISPR systems. Cas9 protein is an RNA-addressable DNA endonuclease where the specificity of the protein can be defined based on homology using a short RNA complementary to a target. In the natural system, the CRISPR array is transcribed and processed into individual spacers called crRNA. The crRNA (that carries homologous sequences to the target DNA) forms a complex with trans-activating crRNA (tracrRNA), which can then bind to Cas9 and guide it to the target DNA. Cas9 first binds to a common sequence in the DNA, called protospacer adjacent motif or PAM (NGG in the case of SpCas9). After binding to the PAM, Cas9 unwinds the DNA and facilitate the formation of an R-loop between the crRNA and the target. If the correct sequence is found, a stable R-loop is formed allowing Cas9 to create a DSB at the site (by nicking both DNA strands with its two nickase domains). By replacing the spacer DNA with a desired DNA sequence, any locus can be targeted and cut with CRISPR/Cas9 (both *in vitro* and *in vivo*). To simplify the system, crRNA and tracrRNA can be linked together to form a chimera, termed single guide RNA (sgRNA, or gRNA), which was shown to be able to guide Cas9 to the target sequence (Jinek et al. 2012).

The DSB generated by Cas9 is predominantly repaired by non-homologous end joining (NHEJ), which introduces various mutations at the site of damage. This characteristic can be used to knock out genes of interest. Alternatively, by providing the system with DNA molecules with homology to the DSB ends, a fraction of DSBs can be redirected to the homologous recombination (HR) pathway (Mali et al. 2013; Cong et al. 2013). Through HR, the provided DNA will be incorporated at the site of damage, adding the provided sequence to the target site, thus enabling

template-dependent DNA writing. Template-dependent DNA writing by Cas9 nuclease has been mainly used for genome engineering applications. However, in principle, signal recording and conditional DNA writing can be achieved by placing the expression of Cas9 or its gRNA under the control of a signal inducible promoters.

Very recently, by taking advantage of the retron mediated ssDNA expression, CRISPR/Cas9 nuclease, and the host proficient homologous recombination machinery, Sharon et al. (Sharon et al. 2018) developed a DNA writing system in *Saccharomyces cerevisiae*. In this system, the retron template (msd) is fused to the gRNA. The gRNA-msd fusion allows localization of the ssDNA (made by the retron reverse transcription) to the site of Cas9-generated DSB. This localized *in vivo* generated ssDNA (which is also complementary to the target but carries a desired mutation) can then be used by the yeast repair machinery as a template for HR-mediated repair of the DSB. This allows template-dependent writing of information encoded in the retron template (msd) into the target genomic loci via HR (Figure 2.8). The authors used this system, termed CRISPEY (Cas9 retron precise parallel editing via homology), for high-throughput fitness screening of thousands of specific variants in yeast. Unlike SCRIBE that relies on beta recombinase for ssDNA recombination, this system takes advantages of the proficient homologous recombination machinery in yeast for ssDNA integration and DNA writing. This makes CRISPEY only applicable to yeast and a few other organisms.

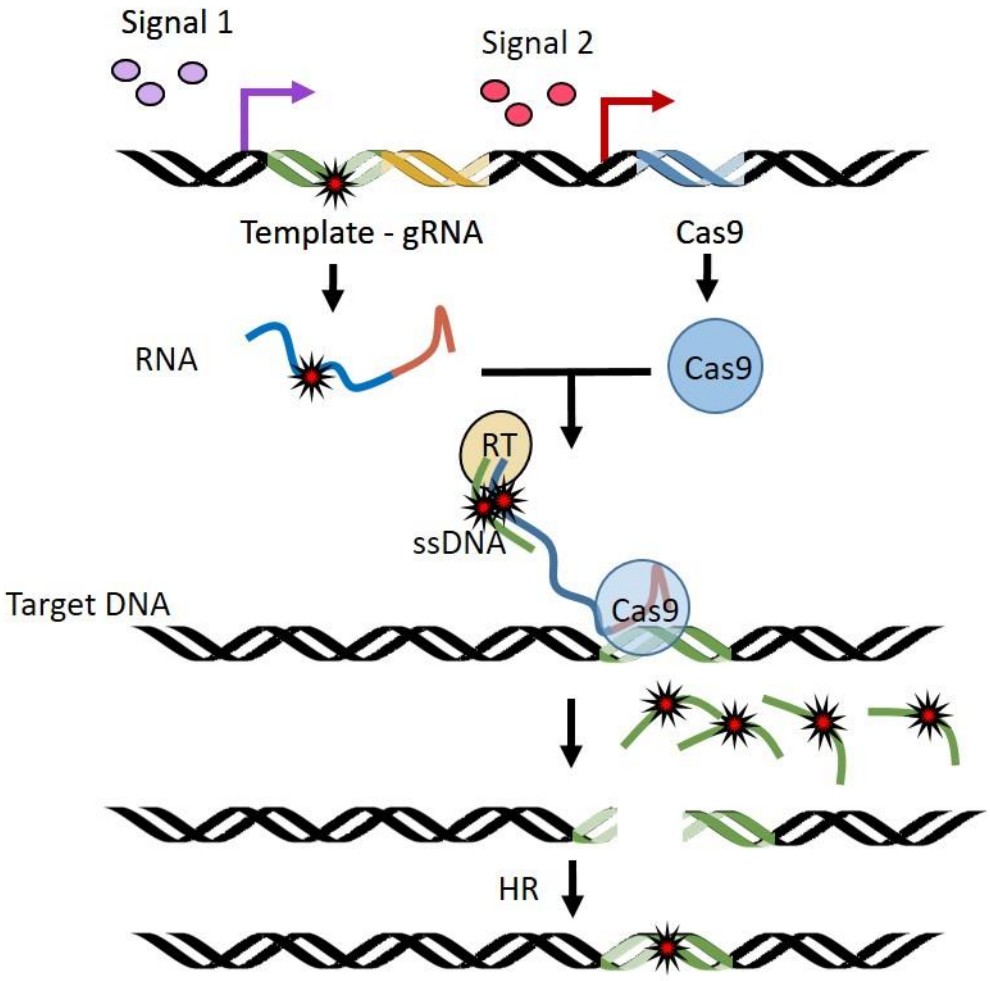

Figure 2.8 Retroelement- and Cas9-mediated DNA writing by CRISPEY

In the CRISPEY system, a gRNA-msd fusion provides both a gRNA for guiding the Cas9 and the template for retron mediated synthesis of ssDNA. The msd is designed to be homologous to the Cas9 target site. As a result, the produced ssDNA can serve as a template for the HR-mediated repair of the DSB generated by Cas9. Upon binding of Cas9 and cleavage of its target site, the DSB is repaired by the yeast homologous recombination machinery, resulting in a template-dependent writing of the msd template into the target genome.

Introducing mutations into one or both of the endonuclease catalytic sites of Cas9, converts this enzyme to an RNA-addressable nickase (nCas9) or DNA binding protein (dCas9), respectively. These non-cutting variants of Cas9 offer attractive options for achieving more precise DNA writing operations that do not require DSBs (especially since DSBs often result in uncontrollable and stochastic mutational outcomes). Other effector domains also can be fused to these Cas9 variants to achieve more functionalities. Some of the examples of such chimeric Cas9-effector proteins include base editors (discussed below), transcriptional activators, repressors and epigenetic modulators (for controlling gene expression), and Cas9-GFP fusions (or other fluorescent proteins, for visualization of the genome architecture or acquiring the topological information of our desired DNA sequence).

Template-independent molecular recording

Template-independent DNA writers are DNA modifying systems capable of generating *de novo* targeted mutations without requiring a template. Depending on the type of the DNA writer, the writing events could result in predetermined or random mutational outcomes in the target DNA.

Site-specific recombinases as molecular recorders

Due to their high precision and efficiency in integrating, deleting or inverting a predefined DNA segment based on the orientation of their recognition sites encompassing the region, site-specific recombinases have been extensively used for creating transgenic model organisms (Stark 2017). However, by viewing each change as a unit of information storage, recombinases can be considered as highly-efficient molecular recorders (i.e., each recombinase is able to act as a binary memory switch that can create permanent memory by inverting, deleting or integrating a DNA segment, analogous to switching between the “0” and “1” memory states) (Figure 2.9). For example, Yang et al. used 11 orthogonal recombinases to build a memory array capable of

recording 2048 distinct memory states in the genome of *E. coli* (Yang et al. 2014). DNA changes made by site-specific recombinases are permanent and unidirectional, however, it has been demonstrated that by coupling the integrase and excisionase functions (in a technology termed RAD for rewritable recombinase addressable data) or by using invertases with opposite directionality, these changes can be reset (Bonnet, Subsoontorn, and Endy 2012; Fernandez-Rodriguez et al. 2015).

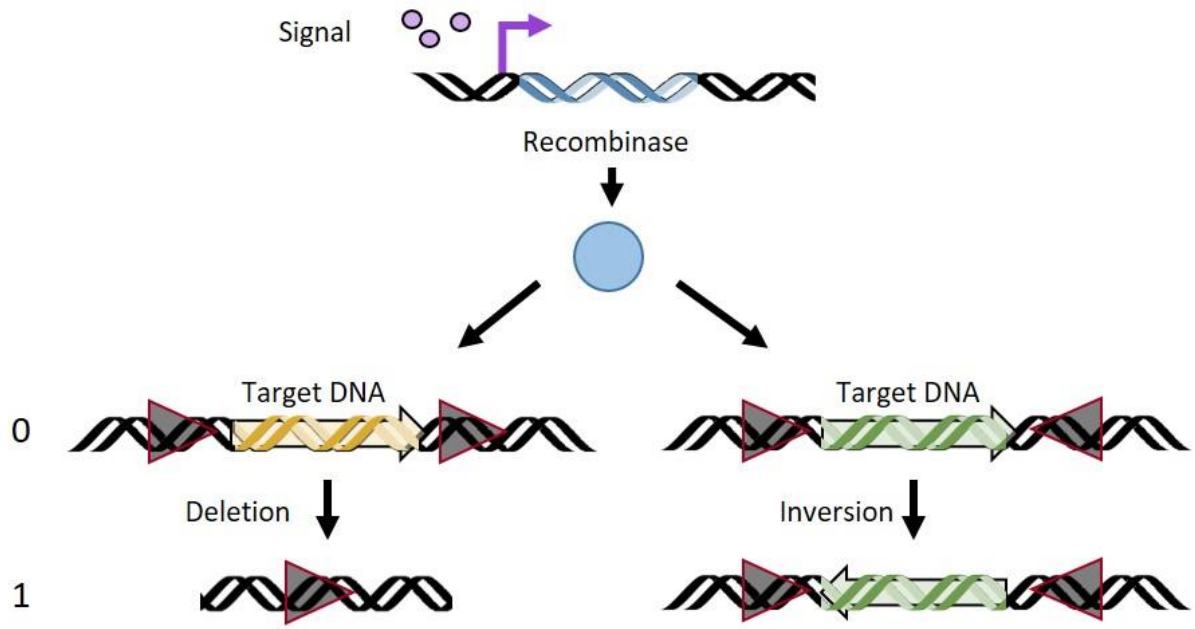

Figure 2.9 Template-independent DNA writing by site-specific recombinases

Site-specific DNA recombinases can efficiently excise (left) or invert (right) pieces of DNA placed between their cognate recognition sites. By coupling the expression of the recombinase protein to a signal of interest, presence or absence of that signal can be recorded as permanent changes in the DNA.

In addition to providing memory, multiple recombinases could be layered to provide data processing and computation in living cells. One of the earliest works in this area is by Friedland et al., who built the DNA invertase cascade (DIC) counter, able to count up to three events in *E. coli* (Friedland et al. 2009). Building on these results, recombinases were used to build synthetic logic circuits and recombinase-based state machines (RSMs) capable of performing Boolean functions in *E. coli* and mammalian cells (Siuti, Yazbek, and Lu 2013; Roquet et al. 2016; Bonnet, Subsoontorn, and Endy 2012; Bonnet et al. 2013). Courbet et al. used these logic circuits for building bacterial biosensors (termed bactosensors) capable of detecting biomarkers in human urine and serum samples (Courbet et al. 2015). Shur et al. built a recombinase-based gene circuit capable of sequential integration of different sequences of DNA depending on the signal that it receives (this concept is similar to the ticker tape molecular recorder concept discussed below) (Shur and Murray 2017).

Applicability of recombinase-based molecular recorders to other bacterial and mammalian systems has also been demonstrated. Mimee et. al. used recombinases for information storage in *Bacteroides thetaiotaomicron*, which is a common bacteria in the human gut microbiome (Mimee et al. 2015). In addition, Weinberg et al. demonstrated that recombinase-based memory and computation is applicable to mammalian systems, by engineering BLADE (Boolean logic and arithmetic through DNA excision) in human HEK 293T cells (Weinberg et al. 2017).

Although recombinases are efficient, precise and applicable to both prokaryotic and eukaryotic systems, they need pre-engineered recognition sites, their storage capacity is low (i.e. each site is limited to one cycle of recording) and their scalability is limited to the number of available recombinases. In addition, expressing multiple orthogonal recombinase systems, highly increases the metabolic burden of the cells, thus limiting the number of memory states that can be encoded in a single cell.

CRISPR/Cas9-based molecular recorders

This class of DNA writers relies on the fact that in the absence of a homologous template, the site-specific DSBs created by CRISPR/Cas9 are repaired by NHEJ. NHEJ can introduce a variety of almost random mutations at the site of damage (see section on NHEJ in chapter 1), which can be used as cellular barcodes (Figure 2.10). McKenna et al. used an array of 10 target sites to develop a method, termed GESTALT (Genome Editing of Synthetic Target Arrays for Lineage Tracing) that enabled them to trace cellular lineages during zebrafish development (McKenna et al. 2016). The authors injected Cas9 and multiple gRNAs to zebrafish embryo and monitored the accumulation of mutations during development. By enabling multiple recording cycles in the same cell, the array accumulates diversity over time and during cell proliferation (Figure 2.11). Analysis of these mutational patterns allows lineage tracing (based on shared mutational signatures). Several other relevant studies have developed similar technologies for lineage tracing in zebrafish (Junker et al. 2017; Alemany et al. 2018). For example, in a technology, termed Scartrace, to avoid adverse effects on viability, the authors targeted the GFP locus in a histone-GFP transgenic zebrafish by injecting them with Cas9 (protein or mRNA) and GFP gRNA during the single cell stage. They observed hundreds of unique mutations in the GFP locus, enabling them to perform lineage tracing during the zebrafish development and fin regeneration.

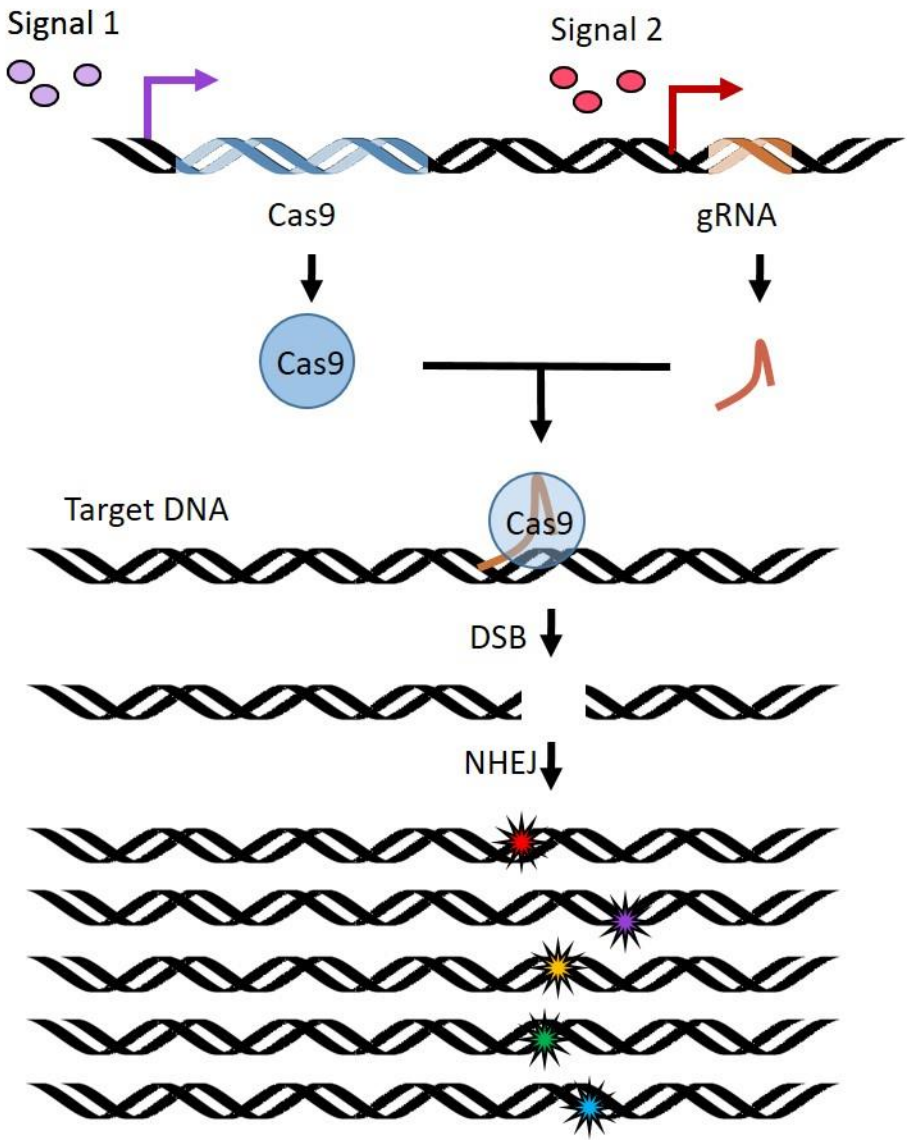

Figure 2.10 Cas9-mediated DNA writing

Targeting Cas9 nuclease to a locus of interest followed by error-prone repair of the generated DSB results in template-independent and stochastic writing into that locus. The DNA writer's components can be optionally expressed from inducible promoters to achieve more control over the writing operations.

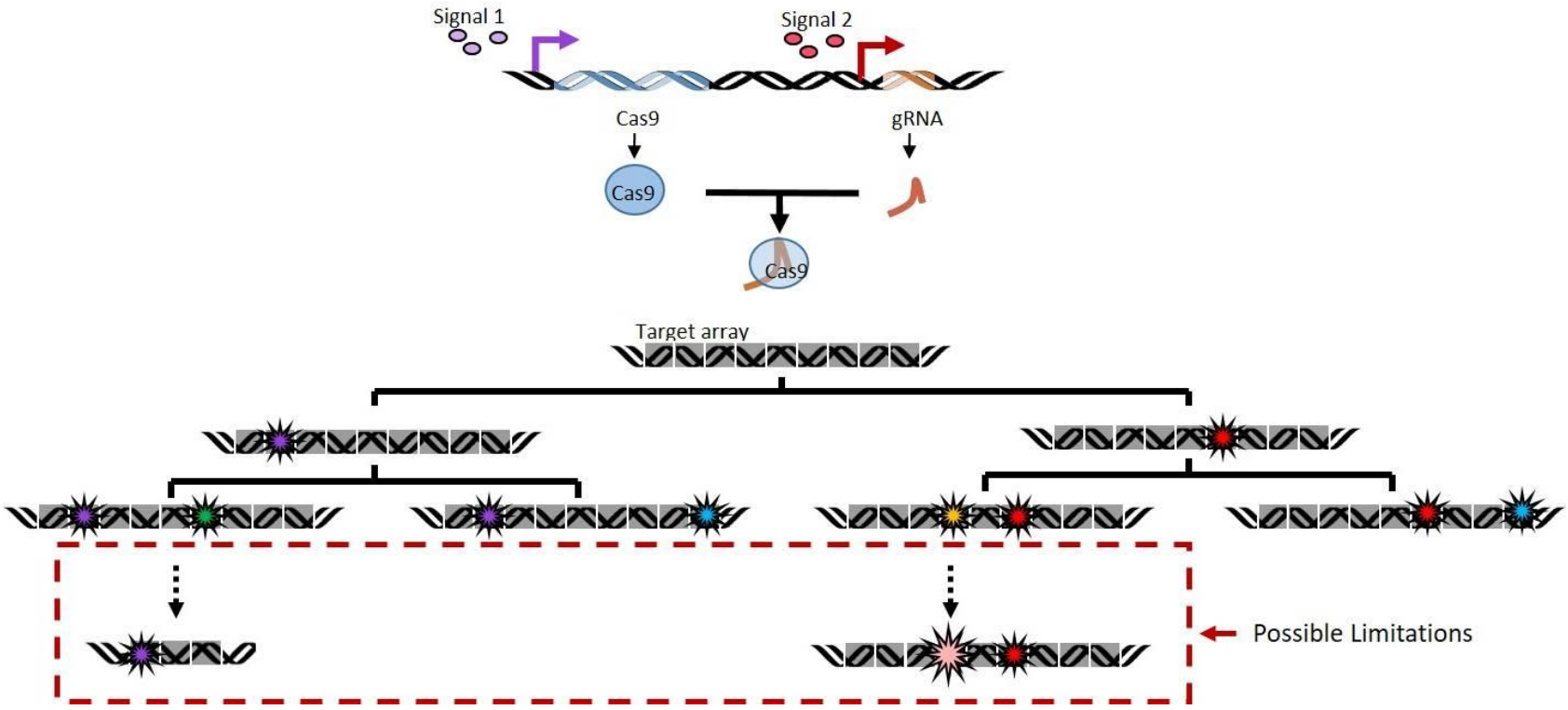

Figure 2.11 Tracing cellular lineage via stochastic and template-independent CRISPR/Cas9 DNA writers

Cas9 and gRNA(s) of interest can be introduced at an early developmental stage of a multicellular organism (e.g., zygote) along with an array of gRNA targets (or alternatively, multiple gRNAs targeting nearby endogenous loci). Successive rounds of template-independent Cas9-mediated DNA writing leads to the formation of multiple mutations in the array over multiple generations. These mutations can then be used as barcodes for tracking the ancestry relationships between the cells. Here, simultaneous DSBs in the repeat array could lead to the deletion of intervening repeats or overwriting on previous records.

Frieda et al. modified the system by placing an array of target sites (to increase the recording capacity) with DNA barcodes (termed scratchpad) downstream of a promoter (Frieda et al. 2016). In this system, termed MEMOIR (Memory by Engineered Mutagenesis with Optical In situ Readout), since the scratchpad is transcribed, the state of the array (after DNA writing) and the barcode can be accessed using smFISH (multiplexed single-molecule RNA hybridization). The authors used this system to map the lineage information of mouse embryonic stem cells. In addition, by placing the gRNA under the control of an inducible promoter (here Wnt-responsive promoter), the authors were able to use the system for recording the signal activity (i.e., the level of gene expression over time). MEMOIR is able to provide lineage and topological information, as well as event histories with single-cell *in situ* readout.

CRISPR/Cas9-based evolving barcodes

In the previous class, although using an array of gRNA target sites increases the recording capacity, simultaneous DSB in a repeat array could lead to the deletion of intervening repeats (and loss of recorded information). Furthermore, in the abovementioned DNA memory architectures after undergoing one successful DNA writing cycle, each target site is destroyed and becomes irresponsive to additional rounds of writing. These limitations were addressed by the development of self-targeting gRNA (stgRNA) (Perli, Cui, and Lu 2016) or homing gRNA (hgRNA)[38] (Kalhor, Mali, and Church 2017) that can undergo multiple cycles of DSBs and NHEJ and thus offer extended recording capacity (Figure 2.12). This continuous mutagenesis of stgRNA is achieved by placing a protospacer adjacent motif (PAM) in the stgRNA encoding locus. Although, in each cycle, the stgRNA sequence changes (which generates a new barcode), it still will be able to target itself and therefore, can go through multiple cycles of recording. This feature allows the creation of a large and diverse set of DNA barcodes over time. By encoding 60 different

[38] here I will be referring to this class of gRNAs as stgRNA

stgRNA in the mouse genome, Kalhor et al. used this memory architecture to trace cellular lineages during the early stages of development in mouse embryos (Kalhor et al. 2018). Besides creating barcodes for lineage tracing, if the expression of Cas9 or stgRNA is placed under a signal-inducible promoter, the system can be used for recording analog information such as signal intensity and duration. Perli et al. used this strategy to build mSCRIBE (mammalian synthetic cellular recorder integrating biological events). They showed that, by placing the expression of Cas9 under the NF-kB inducible promoter, they can record the intensity and duration of lipopolysaccharide-induced inflammation in mice (Perli, Cui, and Lu 2016).

While the stgRNA memory architecture increases the recording capacity over the previous group, and their simplicity makes them readily adaptable, they are only applicable to organisms that have efficient NHEJ. In addition, new mutations created in each cycle may destroy PAM, the previous recordings or the RNA scaffold. Moreover, multiple rounds of deletions (which are the predominant mutation during NHEJ) can shorten the stgRNA limiting the number of recording cycles. This last limitation can be partially overcome by using stgRNAs that are longer than the canonical 20 nucleotide gRNAs (which also could result in a higher level of diversity) (Perli, Cui, and Lu 2016). In addition, undesirable chromosomal rearrangement and cellular toxicity could be a major caveat that may limit the use of DSB-dependent DNA writers for long-term recording.

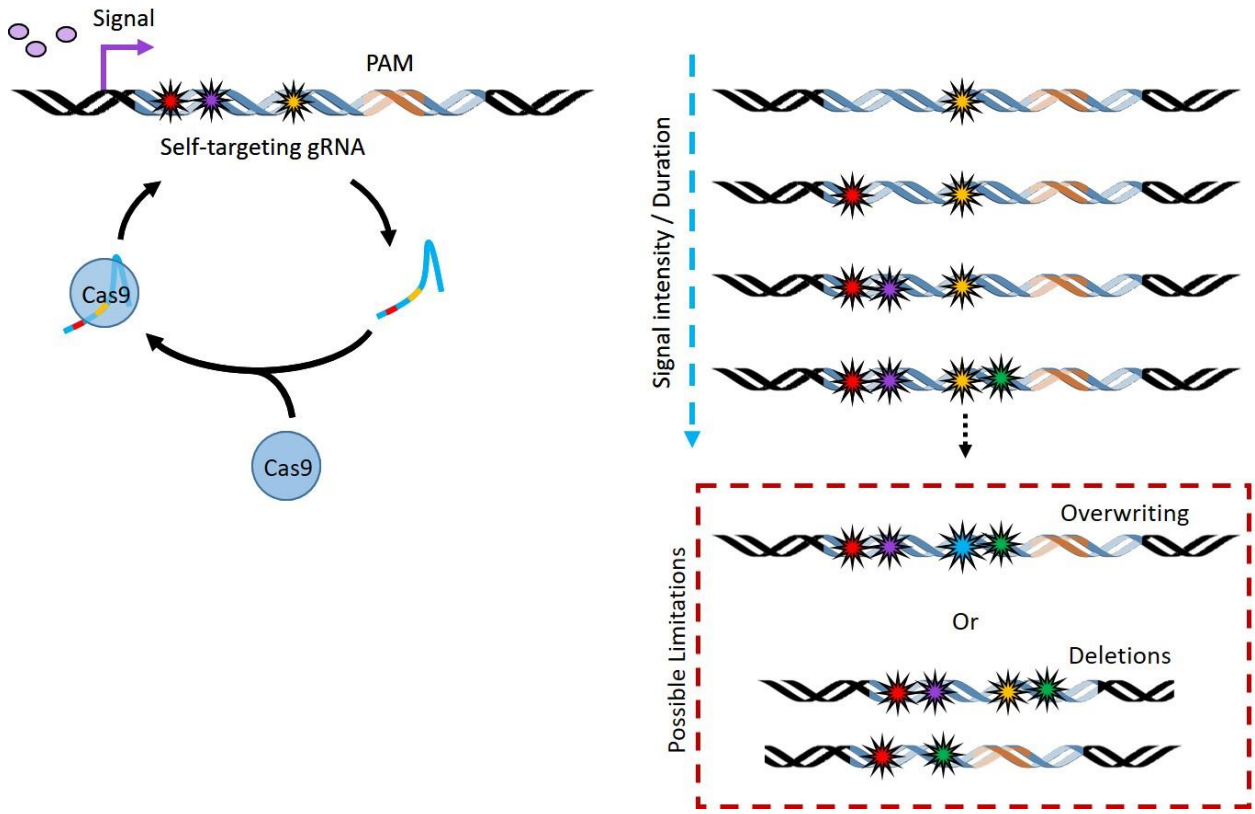

Figure 2.12 Extended template-independent molecular recording by CRISPR/Cas9 mediated evolving barcodes

The number of recording cycles by Cas9-mediated DNA writers can be extended by engineering a PAM inside the gRNA. In this memory architecture, termed self-targeting gRNA (stgRNA), Cas9 is recruited to the stgRNA encoding locus in a successive manner (as the sequence of the stgRNA and its target both change simultaneously). By placing the expression of stgRNA under the control of an inducible promoter, the signal intensity and duration can be recorded in the degree of mutations in the population. Furthermore, the large number of variants that are generated by this mechanism can be used as barcodes for tracing cellular lineages. In this system, however, new mutations may overwrite previous recordings. In addition, deletions, which are the predominant outcome of NHEJ, can shorten stgRNA, erase recordings or remove functional motifs of the stgRNA, limiting the recoding capacity of the system.

Base editors

CRISPR/Cas9-based base editors

In contrast to the Cas9 nuclease DNA writers that generate semi-random mutations, base editors can create precise and minimally-disruptive point mutations (Komor et al. 2016; Nishida et al. 2016). This class of DNA writers were built by fusing a base editor, such as a cytidine deaminase (Komor et al. 2016) or an adenosine deaminase (Gaudelli et al. 2017), to a nuclease dead Cas9 (dCas9), which can be addressed to any location on the DNA using a desired gRNA. Upon binding of the base editor to its target, the base editing process starts by deamination of accessible bases in the write window, thus resulting in the conversion of C to U (in the case of cytidine deaminase) or A to I (in the case of adenosine deaminase). If left unrepaired, the deaminated nucleobases can form base pairs with non-cognate bases and lead to mutations during replication and repair processes. Through this process, the deaminase moiety can thus introduce transition mutations C to T (or G to A on the complementary strand) in the case of a cytidine deaminase, and A to G (or T to C on the complementary strand) in the case of adenosine deaminase. Since these deaminases can only act on ssDNA, they can only function on a small accessible ssDNA window that is generated upon R-loop formation between gRNA and its target strand. This requirement limits the window of activity of base editors to a narrow window in the vicinity of gRNA binding site.

Since cytidine deamination occurs both naturally and by mutagenic agents, cells have evolved efficient mechanisms for detecting and repairing these deamination events. These repair mechanisms limit the efficiency of DNA writing. Therefore, by inhibiting these pathways the efficiency of base editors can be improved. For example, Komor et al showed that by fusing Uracil DNA-Glycosylase inhibitor peptide (UGI, which inhibits the glycosylase responsible for detection of Uracils in the DNA and initiation of the base excision repair pathway) to dCas9, the DNA writing efficiency of base editors can be significantly improved (Komor et al. 2016). Furthermore using

Cas9 nickase (instead of dCas9), which nicks the opposite strand of the target, helps trick cellular mismatch repair machinery to use the edited strand as the template for repair. This helps redirect the outcome of repair towards the desired edited allele (Figure 2.13) (Komor et al. 2016).

Figure 2.13 CRSISPR/Cas9-based base editors

By fusing non-cutting variants of Cas9 (either nCas9 or dCas9) to ssDNA-specific deaminases (either cytidine deaminase or adenosine deaminase) and targeting the fusion protein to the locus of interest by a desired gRNA, nucleotide-resolution DNA writing (in the form of C to T or A to G mutations can be achieved). The efficiency of DNA writing can be improved by using nCas9 instead of dCas9 to trick the mismatch repair system to use the edited strand as template and thus diverting the outcome of the repair towards the edited allele. Furthermore, in the case of cytidine deaminase base editors, addition of uracil glycosylase inhibitor (UGI) peptide to nCas9 helps inhibit the base excision repair pathway responsible for identification and removal of deaminated cytidine from genomic DNA. Conditional DNA writing can be achieved by controlling the expression of the gRNA using inducible promoters. Due to the precise nature of the mutational outcome of these writers, these recorders can be interconnected and placed in cascades where the output of some operators serve as the input for downstream operators. Alternatively, the output (i.e. mutational signature) can target functional elements (e.g., start codon, stop codon, ribosome binding site, promoter, active sites, etc.). These features enables us to build more complex forms of memory and logic operations in living cells.

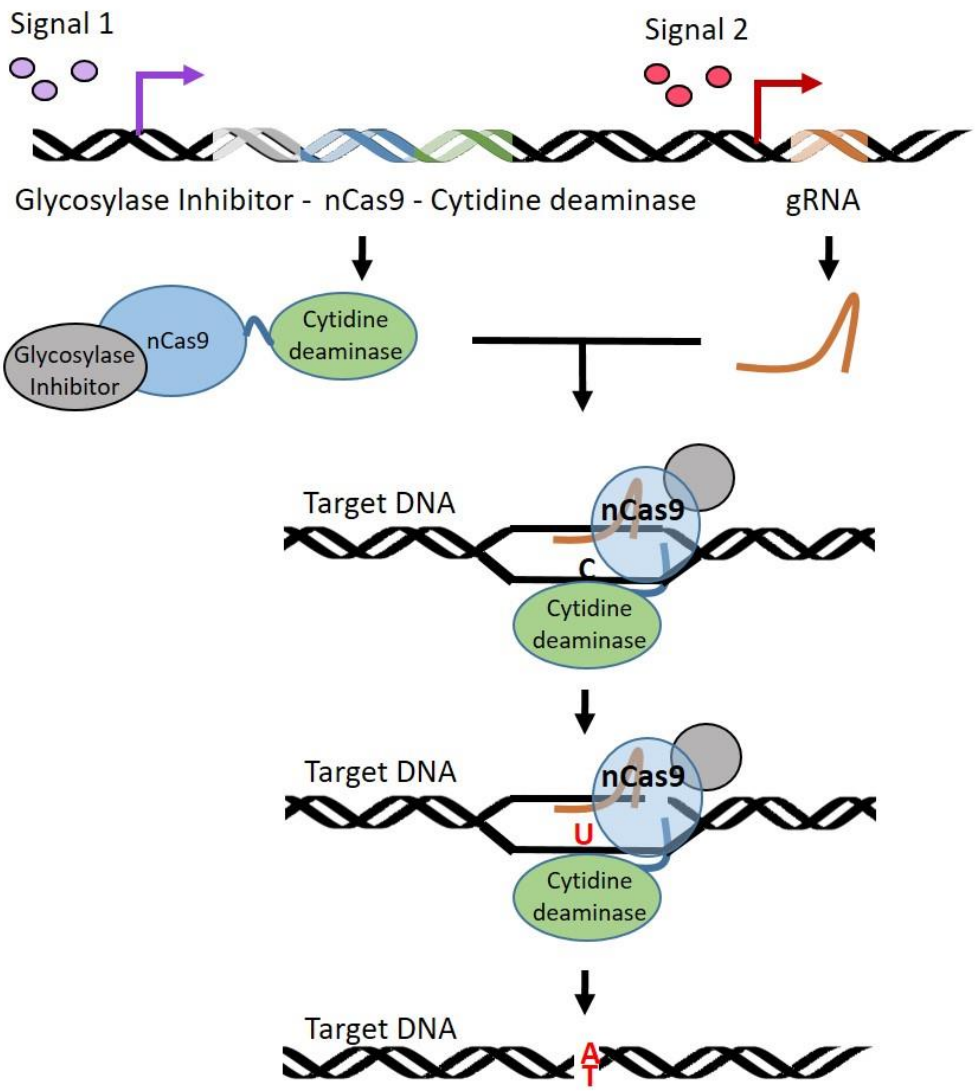

Figure 2.13 (Continued)

By using signal-inducible promoters to control the expression of the components of the base editor system, Tang et al. built a molecular recording device called CAMERA2 (CRISPR-mediated analog multi-event recording apparatus) (Tang and Liu 2018). CAMERA2 is able to record signal presence, as well as its intensity and duration based on the frequency of mutants in the populations of *E. coli* and mammalian cells. Furthermore, they demonstrated that since these recorders are not dependent on the formation of deleterious DSBs, multiple recorders can be used to record multiple signals in the same cell.

Using the base editing technology, Farzadfard et al. developed a more sophisticated memory architecture, termed DOMINO (DNA-based Ordered Memory and Iteration Network Operator), for recording transcriptional signals in a programmable fashion, as well as encoding logic and memory in living cells (Farzadfard et al. 2018). They demonstrated that by controlling the expression of the gRNA using inducible promoters, the intensity and duration of exposure to that signal can be recorded over the course of multiple days (in the form of targeted and precise dC to dT mutations that accumulate in the designed target site). Furthermore, they showed that multiple recorders can be used to record multiple signals in the form of mutational signatures in nearby target loci. This feature was further leveraged to device memory architectures for encoding various forms of logic functions, such as sequence-independent, sequential, and temporal logic which enable recording of the combination, order, and timing of the signals that cells receive. They further showed that these DOMINO logic operators can be used to rationally design and build genetic programs to control the sequence and timing of molecular events in living cells in a robust and scalable fashion. Finally, they demonstrated that this system can be used to build artificial learning gene circuits (where the output of the circuit is reinforced gradually upon longer exposure to biological signals of interest). This circuit allowed them to engineer cells with the ability to learn and remember their history of exposure to inputs, and respond accordingly during future encounters. In addition, this genetic circuit can be used as an online state reporter to

continuously monitor signaling dynamics without the need for sequencing as readout (which necessitates stopping the recording).

Base editor DNA writers and molecular recorders can overcome some of the limitations of the other genetic memory architectures, such as limited recording capacity and scalability. In addition, in contrast to the previous classes which due to their disruptive nature and requirement for special DNA repair mechanisms were not suitable for long-term recording or applicable to many organisms, base editor molecular recorders are applicable to both prokaryotic and eukaryotic systems. Combining the base editor DNA writer system with the stgRNA system could pave the way towards high-capacity molecular recording systems that could be especially useful for lineage tracing applications. Furthermore, by combining cytidine and adenine deaminase base editors, bidirectional DNA writing systems (which can perform writing, as well as erasing) could be designed, enabling the implementation of more sophisticated DNA-based computation and memory operations in living cells.

RNA Polymerase-base editor fusions

Due to the fixed position of the base editor in respect to a given gRNA and thus the corresponding R-loop, the Cas9-based base editors are limited to a narrow window in the vicinity of gRNA binding site. By fusing a deaminase to an RNA polymerase, on the other hand, point mutations can be introduced in a much wider window along the entire transcribed region. Moore et al. used this strategy to build a system for *in vivo* continuous site-directed mutagenesis (Moore, Papa, and Shoulders 2018). They fused a cytidine deaminase to a T7 RNA polymerase and demonstrated that mutations continuously accumulate in a region placed under the control of a T7 promoter (Figure 2.14). In the system the gene of interest can be placed under the control of the T7 promoter and continuously mutagenized. This library of variants can then be subjected to a selection to evolve characteristics of interest in that protein. Alternatively, if the RNA polymerase-base editor chimera is placed under the control of a signal-inducible promoter, the

system can be used as a molecular recorder to store the signal intensity and duration in the form of point mutations that accumulate in the transcribed region. However, in order for these strategies to result in efficient DNA writing, cellular repair machinery responsible for the repair of deaminated bases have to be inhibited. Specially since, unlike CRISPR-based base editors, polymerase-deaminase fusions have a very short residence time on their targets.

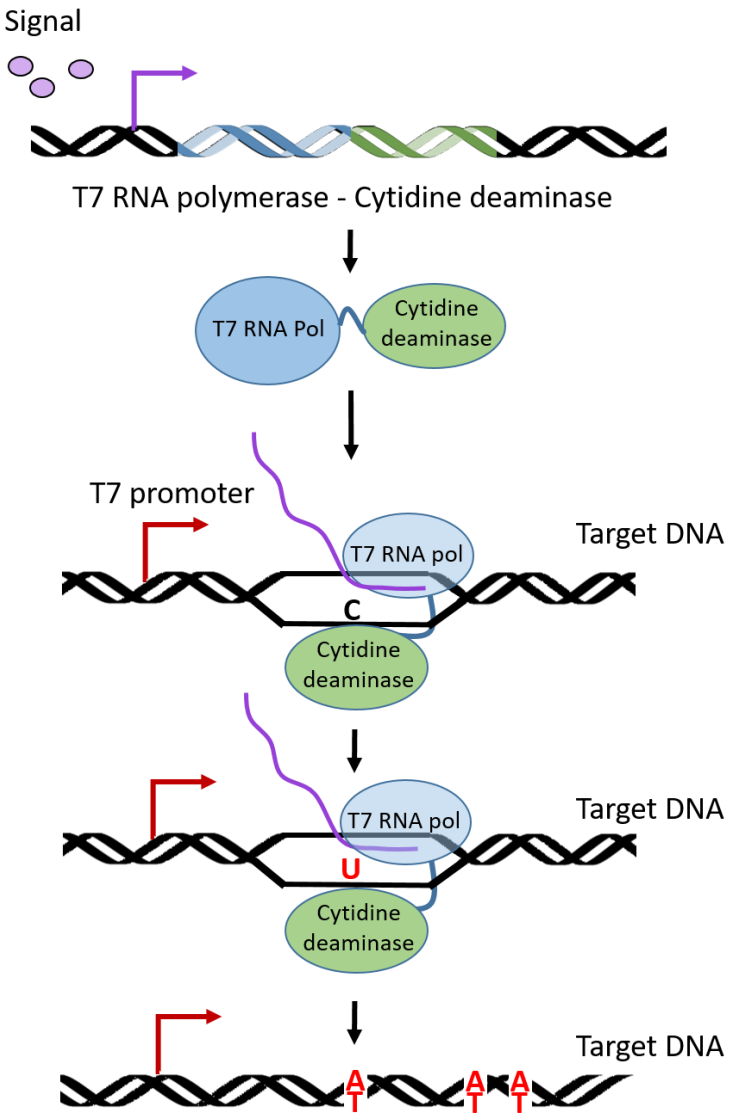

Figure 2.14 DNA writing by RNA polymerase-base editor fusion

The writing window of base editors can be extended by fusing the deaminase protein to an RNA polymerase which can carry over the deaminase over a long distance and thus achieve an extended writing window. To limit the writing to one or a few transcriptional units and minimize non-specific writing, an orthogonal RNA polymerase such as T7 RNA polymerase and its cognate promoter can be used.

Polymerase-based DNA writers

CRISPR/Cas9 - DNA polymerase fusion

Systems that enable targeted mutagenesis within a specified window while avoiding to increase the mutation rate of other parts of the genome could enable powerful strategies for evolutionary engineering applications. Towards this goal, Halprin et al. developed a DNA writing technology, termed EvolvR, which enables mutagenesis of a desired region of DNA by targeting an error-prone DNA polymerase to a target locus while simultaneously generating priming site for the polymerase by the Cas9 nickase (Halperin et al. 2018). The EvolvR DNA writer is composed of nCas9 fused to polI3M, an error-prone version of DNA polymerase I (that is engineered to have higher misincorporation and processivity rates, and lack proofreading activity). After being targeted to a specific site on the DNA by its guide RNA, nCas9 creates a nick on the complementary DNA strand of the guide RNA. After nCas9 dissociation, PolI3M loads onto the DNA from the nick and replaces a ~56 nucleotide region by both DNA synthesis and degradation of the displaced DNA (Figure 2.15). If the introduced mutations do not prevent targeting of nCas9 to the region, EvolvR could repeat this mutagenesis cycle. Using this system, the authors were able to introduce mutations in an editing window of up to 350 nucleotides and increase the mutation rate up to seven orders of magnitude in *E. coli*. Unlike base editors that are only able to generate a certain set of point mutations (i.e., transition mutations), EvolvR strategy enables to generate all the possible forms of point mutations (i.e., both transition and transversion mutations). However, similar to RNA polymerase-base editor fusions, it may be necessary to inhibit cellular repair machinery responsible for the repair of single nucleotide DNA damage (i.e., NER, BER, and MMR) to increase the efficiency of these writers. The performance of EvolvR DNA writer in other organisms and the degree that it relies on host-encoded factors to execute its mutagenesis function remains to be determined. Nevertheless, this strategy could inspire new

generations of DNA writers with broad mutational spectrum and tunable writing window for targeted *in vivo* mutagenesis and continuous evolution applications.

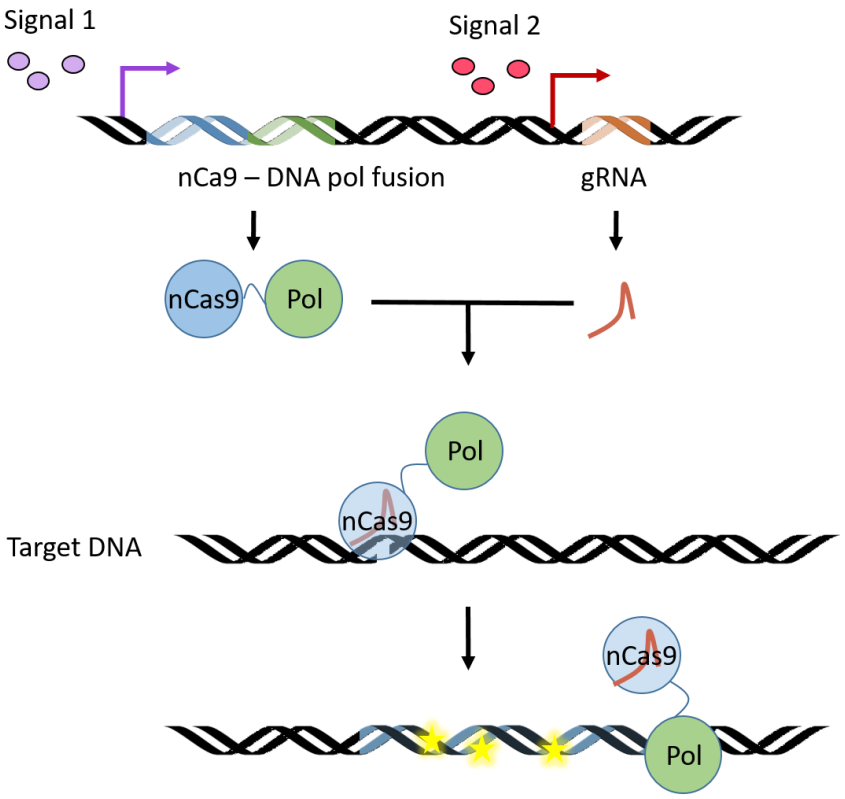

Figure 2.15 Template-independent DNA writing by DNA polymerases via EvolvR system

A fusion of nCas9 and an error-prone polymerase are directed to a desired genomic locus by a gRNA complementary to the target. Nicking the target site by nCas9 initiates the writing process. Loading of the polymerase onto the nick leads to error-prone DNA synthesis at the vicinity of the nick, resulting in localized mutagenesis in that region. The expression of the writer components (i.e., the nCas9-polymerase fusion and the gRNA) can be optionally placed under the control of signals of interest to enable recording of those signals in the DNA.

Polymerase-based ticker tapes

Due to their ability to manipulate information encoded in DNA with fast kinetics, error-prone DNA polymerases offer attractive DNA writing strategies for recording molecular events that occur very fast. It has been suggested that by engineering a DNA polymerase with a fidelity sensitive to changes in ion concentrations (e.g., calcium), temporal dynamics of such changes can be recorded as misincorporation patterns during the synthesis of a template (Figure 2.16). Such a molecular recording system, termed a molecular ticker tape, would be a highly valuable tool for recording the dynamics of highly transient signals such as neural activities (Kording 2011). Several studies have explored the possibility of engineering such a system that allows recording of the concentration of cations (which could be used as a proxy for neural activity) through DNA synthesis *in vitro* (Glaser et al. 2013). However, the performance of these systems are far from practical and none of these systems have been applied to living organisms. Future development of polymerase-based DNA writers and engineering of signal responsive modules (e.g., Calcium responsive peptides) into EvolvR or analogous systems may pave the way towards building molecular recorders with high temporal resolution.

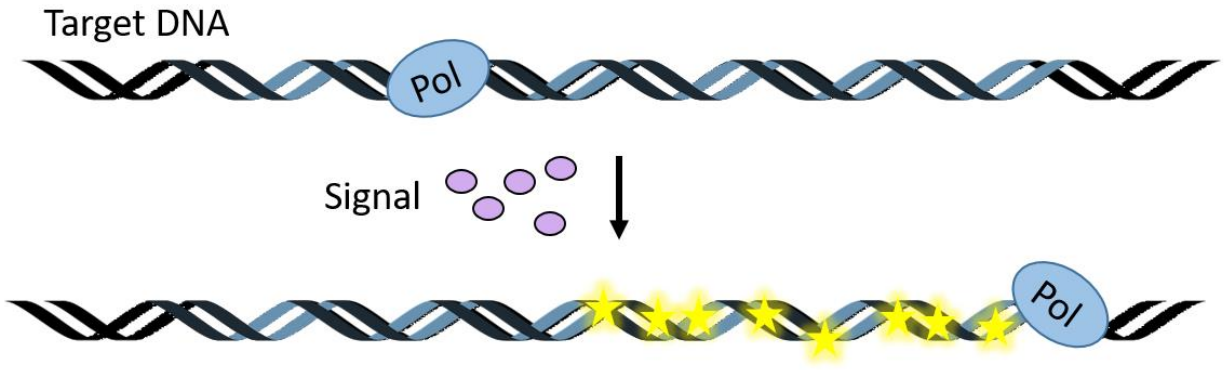

Figure 2.16 DNA polymerase-mediated molecular ticker tape

Engineering a DNA polymerase with a tunable fidelity could pave the way towards building a molecular recorder with high temporal resolution. In such a system, conditional decrease in the fidelity of a replicating DNA polymerase can be recorded in the form of mutational patterns in the DNA. Since DNA synthesis occurs with a fast kinetics in living cells (~1 kb per second), such a system could potentially allow recording of highly transient molecular events such as neural signaling in the DNA.

Discussion

With various dedicated repair pathways protecting the genomic DNA from unwanted mutations, DNA has mainly served as a read-only memory medium for information storage in living cells (at least in short timescales). Recent advances in targeted genome engineering technologies along with our ever-increasing ability to tune and control the outcome of the cellular pathways has made it possible to dynamically write biologically-relevant information into the genomic DNA. Due to these technological advances, genomic DNA is no longer viewed as a read-only information medium and one can now utilize it as a read-write memory for processing and storing various type of information in living cells. Future development of DNA writers with minimal fitness costs and precise and well-controlled mutational outcomes could pave the way toward more advanced memory architectures with more flexible read and write operations. This, in turn, would enable many biomedical and biotechnological applications ranging from the study of basic biology and mechanisms of diseases to engineering logic and computations in living cells.

Due to the interplay between DNA writers and the cellular repair and replications machinery, DNA writing technologies would immensely benefit from technologies that allow to tune and better control cellular repair pathways. Such strategies would allow directing mutational signatures generated by the DNA writers toward desired outcomes, reducing off-target writing, and minimizing fitness cost associated with DNA writing. This, in turn, would allow designing more robust recorders capable of monitoring signaling dynamics over long periods *in vivo*. On the other hand, by enabling to introduce targeted and precise mutations *in vivo*, DNA writers could offer valuable tools for the study of molecular mechanisms of repair pathway. For example, CRISPR/Cas9 DNA writers were recently used to study the outcome of repair of DSBs after NHEJ. By introducing many DSBs across the genome of human cells using CRISPR/Cas9, it was found that the outcome of NHEJ is not entirely random and more deterministic than once thought (Shen et al. 2018). Other types of DNA writer can also be used in analogous ways to better understand

mechanisms underlying various repair pathways involved in the corresponding DNA writing system.

By enabling to dynamically manipulate the blueprint of life, DNA writing technologies allow to reverse the flow of information (from what is expected in central dogma) and directly add new information to this medium on the fly. As such, they can be used to both dynamically record biological information and at the same time, perform computation and a set of inputs, and control and tune cellular phenotypes. The current DNA writing and molecular recording technologies, however, suffer from low temporal resolution and cannot record events that occur in fast time-scales as discrete events. Alternative DNA writing architectures with higher temporal resolution are needed to record highly transient biological events such as neural pulses and protein-protein interactions. Additionally, more robust and expanded biosensor toolbox (e.g., signal-responsive promoters) is needed in order to unleash the utility of molecular recorders in biology as the applicability and performance of these recorders to great extent depends on the performance of the molecular sensor that drives their expression.

While DNA has been mainly used as the main medium for molecular recording applications so far, with the advance of *the in vivo* RNA writing technologies (Cox et al. 2017; Konermann et al. 2018), cellular RNAs can also potentially be used as an alternative medium for dynamic information processing and molecular recording. The dynamic nature of cellular RNAs and independence of RNA writing form cellular repair mechanisms could be advantageous, for example, for building molecular recorders with higher temporal resolution and faster recorder kinetics. However, due to the short half-life of cellular RNAs, this medium can only be used for short-term storage applications and might be better suited for recording information in non-dividing cells. It might be possible to convert information recorded in RNA to long-lasting DNA records (e.g., via reverse transcription-mediated DNA writers such as HiSCRIBE and Cas1-RT/Cas2 systems).

Future advancements of these technologies and other complementary technologies developed in recent years could democratize the use of genomic DNA as a read-and-write medium for processing biological information and interrogation of cellular functions in a high-throughput and resolution fashion. These advances ultimately could provide a clearer picture of the signaling dynamics and factors involved in various biological processes, and revolutionize our ability to rationally program and dynamically control cellular phenotypes.

Bibliography

Aggarwal, M., T. Banerjee, J. A. Sommers, C. Iannascoli, P. Pichierri, R. H. Shoemaker, and R. M. Brosh, Jr. 2013. 'Werner syndrome helicase has a critical role in DNA damage responses in the absence of a functional fanconi anemia pathway', *Cancer Res*, 73: 5497-507.

Akbari, M., J. Pena-Diaz, S. Andersen, N. B. Liabakk, M. Otterlei, and H. E. Krokan. 2009. 'Extracts of proliferating and non-proliferating human cells display different base excision pathways and repair fidelity', *DNA Repair (Amst)*, 8: 834-43.

Alemany, Anna, Maria Florescu, Chloé S. Baron, Josi Peterson-Maduro, and Alexander van Oudenaarden. 2018. 'Whole-organism clone tracing using single-cell sequencing', *Nature*, 556: 108.

Allerston, C. K., S. Y. Lee, J. A. Newman, C. J. Schofield, P. J. McHugh, and O. Gileadi. 2015. 'The structures of the SNM1A and SNM1B/Apollo nuclease domains reveal a potential basis for their distinct DNA processing activities', *Nucleic Acids Res*, 43: 11047-60.

Alpi, A. F., P. E. Pace, M. M. Babu, and K. J. Patel. 2008. 'Mechanistic insight into site-restricted monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI', *Mol Cell*, 32: 767-77.

Alt, F. W., and B. Schwer. 2018. 'DNA double-strand breaks as drivers of neural genomic change, function, and disease', *DNA Repair (Amst)*.

Alt, F. W., Y. Zhang, F. L. Meng, C. Guo, and B. Schwer. 2013. 'Mechanisms of programmed DNA lesions and genomic instability in the immune system', *Cell*, 152: 417-29.

Alter, B. P. 2014. 'Fanconi anemia and the development of leukemia', *Best Pract Res Clin Haematol*, 27: 214-21.

Andreassen, P. R., A. D. D'Andrea, and T. Taniguchi. 2004. 'ATR couples FANCD2 monoubiquitination to the DNA-damage response', *Genes Dev*, 18: 1958-63.

Araujo, S. J., E. A. Nigg, and R. D. Wood. 2001. 'Strong functional interactions of TFIIH with XPC and XPG in human DNA nucleotide excision repair, without a preassembled repairosome', *Mol Cell Biol*, 21: 2281-91.

Auerbach, A. D. 1988. 'A test for Fanconi's anemia', *Blood*, 72: 366-7.

Auerbach, A. D., and S. R. Wolman. 1976. 'Susceptibility of Fanconi's anaemia fibroblasts to chromosome damage by carcinogens', *Nature*, 261: 494-6.

Aylon, Y., B. Liefshitz, and M. Kupiec. 2004. 'The CDK regulates repair of double-strand breaks by homologous recombination during the cell cycle', *EMBO J*, 23: 4868-75.

Bae, J. B., S. S. Mukhopadhyay, L. Liu, N. Zhang, J. Tan, S. Akhter, X. Liu, X. Shen, L. Li, and R. J. Legerski. 2008. 'Snm1B/Apollo mediates replication fork collapse and S Phase checkpoint activation in response to DNA interstrand cross-links', *Oncogene*, 27: 5045-56.

Bagby, G. C., and S. B. Olson. 2003. 'Cisplatin and the sensitive cell', *Nat Med*, 9: 513-4.

Baker, D. J., G. Wuenschell, L. Xia, J. Termini, S. E. Bates, A. D. Riggs, and T. R. O'Connor. 2007. 'Nucleotide excision repair eliminates unique DNA-protein cross-links from mammalian cells', *J Biol Chem*, 282: 22592-604.

Balliano, A. J., and J. J. Hayes. 2015. 'Base excision repair in chromatin: Insights from reconstituted systems', *DNA Repair (Amst)*, 36: 77-85.

Barnes, D. E., and T. Lindahl. 2004. 'Repair and genetic consequences of endogenous DNA base damage in mammalian cells', *Annu Rev Genet*, 38: 445-76.

Barzilay, G., and I. D. Hickson. 1995. 'Structure and function of apurinic/apyrimidinic endonucleases', *Bioessays*, 17: 713-9.

Bastin-Shanower, S. A., W. M. Fricke, J. R. Mullen, and S. J. Brill. 2003. 'The mechanism of Mus81-Mms4 cleavage site selection distinguishes it from the homologous endonuclease Rad1-Rad10', *Mol Cell Biol*, 23: 3487-96.

Bayraktar, G., and M. R. Kreutz. 2018. 'The Role of Activity-Dependent DNA Demethylation in the Adult Brain and in Neurological Disorders', *Front Mol Neurosci*, 11: 169.

Belousova, E. A., and O. I. Lavrik. 2015. 'Repair of Clustered Damage and DNA Polymerase Iota', *Biochemistry (Mosc)*, 80: 1010-8.

Ben-Yehoyada, M., L. C. Wang, I. D. Kozekov, C. J. Rizzo, M. E. Gottesman, and J. Gautier. 2009. 'Checkpoint signaling from a single DNA interstrand crosslink', *Mol Cell*, 35: 704-15.

Bennett, P. V., N. L. Cuomo, S. Paul, S. T. Tafrov, and B. M. Sutherland. 2005. 'Endogenous DNA damage clusters in human skin, 3-D model, and cultured skin cells', *Free Radic Biol Med*, 39: 832-9.

Bergstralh, D. T., and J. Sekelsky. 2008. 'Interstrand crosslink repair: can XPF-ERCC1 be let off the hook?', *Trends Genet*, 24: 70-6.

Bernstein, C., and H. Bernstein. 2015. 'Epigenetic reduction of DNA repair in progression to gastrointestinal cancer', *World J Gastrointest Oncol*, 7: 30-46.

Berquist, B. R., and D. M. Wilson, 3rd. 2012. 'Pathways for repairing and tolerating the spectrum of oxidative DNA lesions', *Cancer Lett*, 327: 61-72.

Boeke, J. D., and J. P. Stoye. 1997. 'Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements.' in J. M. Coffin, S. H. Hughes and H. E. Varmus (eds.), *Retroviruses* (Cold Spring Harbor (NY)).

Bogliolo, M., B. Schuster, C. Stoepker, B. Derkunt, Y. Su, A. Raams, J. P. Trujillo, J. Minguillon, M. J. Ramirez, R. Pujol, J. A. Casado, R. Banos, P. Rio, K. Knies, S. Zuniga, J. Benitez, J. A. Bueren, N. G. Jaspers, O. D. Scharer, J. P. de Winter, D. Schindler, and J. Surralles. 2013. 'Mutations in ERCC4, encoding the DNA-repair endonuclease XPF, cause Fanconi anemia', *Am J Hum Genet*, 92: 800-6.

Bonnet, J., P. Subsoontorn, and D. Endy. 2012. 'Rewritable digital data storage in live cells via engineered control of recombination directionality', *Proc Natl Acad Sci U S A*, 109: 8884-9.

Bonnet, J., P. Yin, M. E. Ortiz, P. Subsoontorn, and D. Endy. 2013. 'Amplifying genetic logic gates', *Science*, 340: 599-603.

Boulton, S. J., and S. P. Jackson. 1996. 'Saccharomyces cerevisiae Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways', *EMBO J*, 15: 5093-103.

Bowles, M., J. Lally, A. J. Fadden, S. Mouilleron, T. Hammonds, and N. Q. McDonald. 2012. 'Fluorescence-based incision assay for human XPF-ERCC1 activity identifies important elements of DNA junction recognition', *Nucleic Acids Res*, 40: e101.

Box, H. C., J. B. Dawidzik, and E. E. Budzinski. 2001. 'Free radical-induced double lesions in DNA', *Free Radic Biol Med*, 31: 856-68.

Boyce, R. P., and P. Howard-Flanders. 1964. 'Release of Ultraviolet Light-Induced Thymine Dimers from DNA in E. Coli K-12', *Proc Natl Acad Sci U S A*, 51: 293-300.

Brooks, S. C., S. Adhikary, E. H. Rubinson, and B. F. Eichman. 2013. 'Recent advances in the structural mechanisms of DNA glycosylases', *Biochim Biophys Acta*, 1834: 247-71.

Budman, J., and G. Chu. 2005. 'Processing of DNA for nonhomologous end-joining by cell-free extract', *EMBO J*, 24: 849-60.

Budzowska, M., T. G. Graham, A. Sobeck, S. Waga, and J. C. Walter. 2015. 'Regulation of the Rev1-pol zeta complex during bypass of a DNA interstrand cross-link', *EMBO J*, 34: 1971-85.

Bunting, S. F., and A. Nussenzweig. 2010. 'Dangerous liaisons: Fanconi anemia and toxic nonhomologous end joining in DNA crosslink repair', *Mol Cell*, 39: 164-6.

Cao, C., Y. L. Jiang, J. T. Stivers, and F. Song. 2004. 'Dynamic opening of DNA during the enzymatic search for a damaged base', *Nat Struct Mol Biol*, 11: 1230-6.

Castor, D., N. Nair, A. C. Declais, C. Lachaud, R. Toth, T. J. Macartney, D. M. Lilley, J. S. Arthur, and J. Rouse. 2013. 'Cooperative control of holliday junction resolution and DNA repair by the SLX1 and MUS81-EME1 nucleases', *Mol Cell*, 52: 221-33.

Cattell, E., B. Sengerova, and P. J. McHugh. 2010. 'The SNM1/Pso2 family of ICL repair nucleases: from yeast to man', *Environ Mol Mutagen*, 51: 635-45.

Ceccaldi, R., P. Sarangi, and A. D. D'Andrea. 2016. 'The Fanconi anaemia pathway: new players and new functions', *Nat Rev Mol Cell Biol*, 17: 337-49.

Centore, R. C., S. A. Yazinski, A. Tse, and L. Zou. 2012. 'Spartan/C1orf124, a reader of PCNA ubiquitylation and a regulator of UV-induced DNA damage response', *Mol Cell*, 46: 625-35.

Chaudhury, I., D. R. Stroik, and A. Sobeck. 2014. 'FANCD2-controlled chromatin access of the Fanconi-associated nuclease FAN1 is crucial for the recovery of stalled replication forks', *Mol Cell Biol*, 34: 3939-54.

Chiruvella, K. K., Z. Liang, and T. E. Wilson. 2013. 'Repair of double-strand breaks by end joining', *Cold Spring Harb Perspect Biol*, 5: a012757.

Cholewa-Waclaw, J., A. Bird, M. von Schimmelmann, A. Schaefer, H. Yu, H. Song, R. Madabhushi, and L. H. Tsai. 2016. 'The Role of Epigenetic Mechanisms in the Regulation of Gene Expression in the Nervous System', *J Neurosci*, 36: 11427-34.

Chun, J., E. S. Buechelmaier, and S. N. Powell. 2013. 'Rad51 paralog complexes BCDX2 and CX3 act at different stages in the BRCA1-BRCA2-dependent homologous recombination pathway', *Mol Cell Biol*, 33: 387-95.

Ciccia, A., and S. J. Elledge. 2010. 'The DNA damage response: making it safe to play with knives', *Mol Cell*, 40: 179-204.

Ciccia, A., N. McDonald, and S. C. West. 2008. 'Structural and functional relationships of the XPF/MUS81 family of proteins', *Annu Rev Biochem*, 77: 259-87.

Clauson, C., O. D. Scharer, and L. Niedernhofer. 2013. 'Advances in understanding the complex mechanisms of DNA interstrand cross-link repair', *Cold Spring Harb Perspect Biol*, 5: a012732.

Cohn, M. A., P. Kowal, K. Yang, W. Haas, T. T. Huang, S. P. Gygi, and A. D. D'Andrea. 2007. 'A UAF1-containing multisubunit protein complex regulates the Fanconi anemia pathway', *Mol Cell*, 28: 786-97.

Coin, F., V. Oksenych, and J. M. Egly. 2007. 'Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair', *Mol Cell*, 26: 245-56.

Cole, R. S. 1973. 'Repair of DNA containing interstrand crosslinks in Escherichia coli: sequential excision and recombination', *Proc Natl Acad Sci U S A*, 70: 1064-8.

Compe, E., and J. M. Egly. 2012. 'TFIIH: when transcription met DNA repair', *Nat Rev Mol Cell Biol*, 13: 343-54.

Cong, Le, F. Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D. Hsu, Xuebing Wu, Wenyan Jiang, Luciano A. Marraffini, and Feng Zhang. 2013. 'Multiplex genome engineering using CRISPR/Cas systems', *Science (New York, N.Y.)*, 339: 819-23.

Cottarel, J., P. Frit, O. Bombarde, B. Salles, A. Negrel, S. Bernard, P. A. Jeggo, M. R. Lieber, M. Modesti, and P. Calsou. 2013. 'A noncatalytic function of the ligation complex during nonhomologous end joining', *J Cell Biol*, 200: 173-86.

Courbet, A., D. Endy, E. Renard, F. Molina, and J. Bonnet. 2015. 'Detection of pathological biomarkers in human clinical samples via amplifying genetic switches and logic gates', *Sci Transl Med*, 7: 289ra83.

Cox, David B. T., Jonathan S. Gootenberg, Omar O. Abudayyeh, Brian Franklin, Max J. Kellner, Julia Joung, and Feng Zhang. 2017. 'RNA editing with CRISPR-Cas13', *Science*.

Critchlow, S. E., R. P. Bowater, and S. P. Jackson. 1997. 'Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV', *Curr Biol*, 7: 588-98.

Crossan, G. P., and K. J. Patel. 2012. 'The Fanconi anaemia pathway orchestrates incisions at sites of crosslinked DNA', *J Pathol*, 226: 326-37.

Crowe, M. L. 2005. 'SeqDoC: rapid SNP and mutation detection by direct comparison of DNA sequence chromatograms', *BMC Bioinformatics*, 6: 133.

Crowe, S. L., V. A. Movsesyan, T. J. Jorgensen, and A. Kondratyev. 2006. 'Rapid phosphorylation of histone H2A.X following ionotropic glutamate receptor activation', *Eur J Neurosci*, 23: 2351-61.

Dalhus, B., J. K. Laerdahl, P. H. Backe, and M. Bjoras. 2009. 'DNA base repair--recognition and initiation of catalysis', *FEMS Microbiol Rev*, 33: 1044-78.

de Laat, W. L., E. Appeldoorn, N. G. Jaspers, and J. H. Hoeijmakers. 1998. 'DNA structural elements required for ERCC1-XPF endonuclease activity', *J Biol Chem*, 273: 7835-42.

De Silva, I. U., P. J. McHugh, P. H. Clingen, and J. A. Hartley. 2000. 'Defining the roles of nucleotide excision repair and recombination in the repair of DNA interstrand cross-links in mammalian cells', *Mol Cell Biol*, 20: 7980-90.

Deans, A. J., and S. C. West. 2009. 'FANCM connects the genome instability disorders Bloom's Syndrome and Fanconi Anemia', *Mol Cell*, 36: 943-53.

———. 2011. 'DNA interstrand crosslink repair and cancer', *Nat Rev Cancer*, 11: 467-80.

Dianov, G. L., and U. Hubscher. 2013. 'Mammalian base excision repair: the forgotten archangel', *Nucleic Acids Res*, 41: 3483-90.

DiGiovanna, J. J., and K. H. Kraemer. 2012. 'Shining a light on xeroderma pigmentosum', *J Invest Dermatol*, 132: 785-96.

Do, H., N. C. Wong, C. Murone, T. John, B. Solomon, P. L. Mitchell, and A. Dobrovic. 2014. 'A critical re-assessment of DNA repair gene promoter methylation in non-small cell lung carcinoma', *Sci Rep*, 4: 4186.

Douki, T., J. Riviere, and J. Cadet. 2002. 'DNA tandem lesions containing 8-oxo-7,8-dihydroguanine and formamido residues arise from intramolecular addition of thymine peroxyl radical to guanine', *Chem Res Toxicol*, 15: 445-54.

Dronkert, M. L., and R. Kanaar. 2001. 'Repair of DNA interstrand cross-links', *Mutat Res*, 486: 217-47.

Duckett, D. R., J. T. Drummond, A. I. Murchie, J. T. Reardon, A. Sancar, D. M. Lilley, and P. Modrich. 1996. 'Human MutSalpha recognizes damaged DNA base pairs containing O6-methylguanine, O4-methylthymine, or the cisplatin-d(GpG) adduct', *Proc Natl Acad Sci U S A*, 93: 6443-7.

Dunand-Sauthier, I., M. Hohl, F. Thorel, P. Jaquier-Gubler, S. G. Clarkson, and O. D. Scharer. 2005. 'The spacer region of XPG mediates recruitment to nucleotide excision repair complexes and determines substrate specificity', *J Biol Chem*, 280: 7030-7.

Duquette, M. L., Q. Zhu, E. R. Taylor, A. J. Tsay, L. Z. Shi, M. W. Berns, and C. H. McGowan. 2012. 'CtIP is required to initiate replication-dependent interstrand crosslink repair', *PLoS Genet*, 8: e1003050.

Duxin, J. P., J. M. Dewar, H. Yardimci, and J. C. Walter. 2014. 'Repair of a DNA-protein crosslink by replication-coupled proteolysis', *Cell*, 159: 346-57.

Eccles, L. J., P. O'Neill, and M. E. Lomax. 2011. 'Delayed repair of radiation induced clustered DNA damage: friend or foe?', *Mutat Res*, 711: 134-41.

Einhorn, J. 1985. 'Nitrogen mustard: the origin of chemotherapy for cancer', *Int J Radiat Oncol Biol Phys*, 11: 1375-8.

Enoiu, M., J. Jiricny, and O. D. Scharer. 2012. 'Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis', *Nucleic Acids Res*, 40: 8953-64.

Espejel, S., S. Franco, S. Rodriguez-Perales, S. D. Bouffler, J. C. Cigudosa, and M. A. Blasco. 2002. 'Mammalian Ku86 mediates chromosomal fusions and apoptosis caused by critically short telomeres', *EMBO J*, 21: 2207-19.

Ewig, R. A., and K. W. Kohn. 1977. 'DNA damage and repair in mouse leukemia L1210 cells treated with nitrogen mustard, 1,3-bis(2-chloroethyl)-1-nitrosourea, and other nitrosoureas', *Cancer Res*, 37: 2114-22.

Fagbemi, A. F., B. Orelli, and O. D. Scharer. 2011. 'Regulation of endonuclease activity in human nucleotide excision repair', *DNA Repair (Amst)*, 10: 722-9.

Fan, J., and D. M. Wilson, 3rd. 2005. 'Protein-protein interactions and posttranslational modifications in mammalian base excision repair', *Free Radic Biol Med*, 38: 1121-38.

Fanconi, Guido. 1927. 'Familiare infantile perniziosaartige Anamie (pernizioses Blutbild und Konstitution)', *Jahrbuch Kinderheilk*, 117: 257-80.

Farkas, S. A., V. Vymetalkova, L. Vodickova, P. Vodicka, and T. K. Nilsson. 2014. 'DNA methylation changes in genes frequently mutated in sporadic colorectal cancer and in the DNA repair and Wnt/beta-catenin signaling pathway genes', *Epigenomics*, 6: 179-91.

Farzadfard, F., and T. K. Lu. 2014. 'Synthetic biology. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations', *Science*, 346: 1256272.

Farzadfard, Fahim, Nava Gharaei, Yasutomi Higashikuni, Giyoung Jung, Jicong Cao, and Timothy K. Lu. 2018. 'Single-Nucleotide-Resolution Computing and Memory in Living Cells', *bioRxiv*.

Ferguson, D. O., and F. W. Alt. 2001. 'DNA double strand break repair and chromosomal translocation: lessons from animal models', *Oncogene*, 20: 5572-9.

Fernandez-Rodriguez, J., L. Yang, T. E. Gorochowski, D. B. Gordon, and C. A. Voigt. 2015. 'Memory and Combinatorial Logic Based on DNA Inversions: Dynamics and Evolutionary Stability', *ACS Synth Biol*, 4: 1361-72.

Fornace, A. J., Jr. 1982. 'Detection of DNA single-strand breaks produced during the repair of damage by DNA-protein cross-linking agents', *Cancer Res*, 42: 145-9.

Fortini, P., and E. Dogliotti. 2007. 'Base damage and single-strand break repair: mechanisms and functional significance of short- and long-patch repair subpathways', *DNA Repair (Amst)*, 6: 398-409.

Frank, K. M., J. M. Sekiguchi, K. J. Seidl, W. Swat, G. A. Rathbun, H. L. Cheng, L. Davidson, L. Kangaloo, and F. W. Alt. 1998. 'Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV', *Nature*, 396: 173-7.

Frattini, V., V. Trifonov, J. M. Chan, A. Castano, M. Lia, F. Abate, S. T. Keir, A. X. Ji, P. Zoppoli, F. Niola, C. Danussi, I. Dolgalev, P. Porrati, S. Pellegatta, A. Heguy, G. Gupta, D. J. Pisapia, P. Canoll, J. N. Bruce, R. E. McLendon, H. Yan, K. Aldape, G. Finocchiaro, T. Mikkelsen, G. G. Prive, D. D. Bigner, A. Lasorella, R. Rabadan, and A. Iavarone. 2013. 'The integrated landscape of driver genomic alterations in glioblastoma', *Nat Genet*, 45: 1141-9.

Frieda, Kirsten L., James M. Linton, Sahand Hormoz, Joonhyuk Choi, Ke-Huan K. Chow, Zakary S. Singer, Mark W. Budde, Michael B. Elowitz, and Long Cai. 2016. 'Synthetic recording and in situ readout of lineage information in single cells', *Nature*, 541: 107.

Friedberg, E. C. 2005. 'Suffering in silence: the tolerance of DNA damage', *Nat Rev Mol Cell Biol*, 6: 943-53.

Friedland, A. E., T. K. Lu, X. Wang, D. Shi, G. Church, and J. J. Collins. 2009. 'Synthetic gene networks that count', *Science*, 324: 1199-202.

Friedman, J. I., and J. T. Stivers. 2010. 'Detection of damaged DNA bases by DNA glycosylase enzymes', *Biochemistry*, 49: 4957-67.

Fromme, J. C., A. Banerjee, and G. L. Verdine. 2004. 'DNA glycosylase recognition and catalysis', *Curr Opin Struct Biol*, 14: 43-9.

Fu, Y. V., H. Yardimci, D. T. Long, T. V. Ho, A. Guainazzi, V. P. Bermudez, J. Hurwitz, A. van Oijen, O. D. Scharer, and J. C. Walter. 2011. 'Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase', *Cell*, 146: 931-41.

Gao, Y., Y. Sun, K. M. Frank, P. Dikkes, Y. Fujiwara, K. J. Seidl, J. M. Sekiguchi, G. A. Rathbun, W. Swat, J. Wang, R. T. Bronson, B. A. Malynn, M. Bryans, C. Zhu, J. Chaudhuri, L. Davidson, R. Ferrini, T. Stamato, S. H. Orkin, M. E. Greenberg, and F. W. Alt. 1998. 'A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis', *Cell*, 95: 891-902.

Garaycoechea, J. I., G. P. Crossan, F. Langevin, M. Daly, M. J. Arends, and K. J. Patel. 2012. 'Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function', *Nature*, 489: 571-5.

Garcia-Higuera, I., T. Taniguchi, S. Ganesan, M. S. Meyn, C. Timmers, J. Hejna, M. Grompe, and A. D. D'Andrea. 2001. 'Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway', *Mol Cell*, 7: 249-62.

Gaudelli, Nicole M., Alexis C. Komor, Holly A. Rees, Michael S. Packer, Ahmed H. Badran, David I. Bryson, and David R. Liu. 2017. 'Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage', *Nature*, 551: 464.

Gellon, L., D. R. Carson, J. P. Carson, and B. Demple. 2008. 'Intrinsic 5'-deoxyribose-5-phosphate lyase activity in Saccharomyces cerevisiae Trf4 protein with a possible role in base excision DNA repair', *DNA Repair (Amst)*, 7: 187-98.

Genschel, J., L. R. Bazemore, and P. Modrich. 2002. 'Human exonuclease I is required for 5' and 3' mismatch repair', *J Biol Chem*, 277: 13302-11.

Georgakilas, A. G., P. V. Bennett, and B. M. Sutherland. 2002. 'High efficiency detection of bi-stranded abasic clusters in gamma-irradiated DNA by putrescine', *Nucleic Acids Res*, 30: 2800-8.

Georgakilas, A. G., P. V. Bennett, D. M. Wilson, 3rd, and B. M. Sutherland. 2004. 'Processing of bistranded abasic DNA clusters in gamma-irradiated human hematopoietic cells', *Nucleic Acids Res*, 32: 5609-20.

Ghodgaonkar, M. M., F. Lazzaro, M. Olivera-Pimentel, M. Artola-Boran, P. Cejka, M. A. Reijns, A. P. Jackson, P. Plevani, M. Muzi-Falconi, and J. Jiricny. 2013. 'Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair', *Mol Cell*, 50: 323-32.

Gillet, L. C., and O. D. Scharer. 2006. 'Molecular mechanisms of mammalian global genome nucleotide excision repair', *Chem Rev*, 106: 253-76.

Glaser, Joshua I., Bradley M. Zamft, Adam H. Marblestone, Jeffrey R. Moffitt, Keith Tyo, Edward S. Boyden, George Church, and Konrad P. Kording. 2013. 'Statistical Analysis of Molecular Signal Recording', *PLoS Computational Biology*, 9: e1003145.

Goldacre, R. J., A. Loveless, and W. C. Ross. 1949. 'Mode of production of chromosome abnormalities by the nitrogen mustards; the possible role of cross-linking', *Nature*, 163: 667-9.

Goodhead, D. T. 1994. 'Initial events in the cellular effects of ionizing radiations: clustered damage in DNA', *Int J Radiat Biol*, 65: 7-17.

Goodhead, D. T., R. J. Munson, J. Thacker, and R. Cox. 1980. 'Mutation and inactivation of cultured mammalian cells exposed to beams of accelerated heavy ions. IV. Biophysical interpretation', *Int J Radiat Biol Relat Stud Phys Chem Med*, 37: 135-67.

Goodman, L. S., M. M. Wintrobe, and et al. 1946. 'Nitrogen mustard therapy; use of methyl-bis (beta-chloroethyl) amine hydrochloride and tris (beta-chloroethyl) amine hydrochloride for Hodgkin's disease, lymphosarcoma, leukemia and certain allied and miscellaneous disorders', *J Am Med Assoc*, 132: 126-32.

Goodman, M. F., and R. Woodgate. 2013. 'Translesion DNA polymerases', *Cold Spring Harb Perspect Biol*, 5: a010363.

Grawunder, U., M. Wilm, X. Wu, P. Kulesza, T. E. Wilson, M. Mann, and M. R. Lieber. 1997. 'Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells', *Nature*, 388: 492-5.

Groothuizen, F. S., and T. K. Sixma. 2016. 'The conserved molecular machinery in DNA mismatch repair enzyme structures', *DNA Repair (Amst)*, 38: 14-23.

Grossmann, K. F., A. M. Ward, M. E. Matkovic, A. E. Folias, and R. E. Moses. 2001. 'S. cerevisiae has three pathways for DNA interstrand crosslink repair', *Mutat Res*, 487: 73-83.

Guirouilh-Barbat, J., S. Huck, P. Bertrand, L. Pirzio, C. Desmaze, L. Sabatier, and B. S. Lopez. 2004. 'Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells', *Mol Cell*, 14: 611-23.

Halperin, S. O., C. J. Tou, E. B. Wong, C. Modavi, D. V. Schaffer, and J. E. Dueber. 2018. 'CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window', *Nature*, 560: 248-52.

Hammel, M., Y. Yu, B. L. Mahaney, B. Cai, R. Ye, B. M. Phipps, R. P. Rambo, G. L. Hura, M. Pelikan, S. So, R. M. Abolfath, D. J. Chen, S. P. Lees-Miller, and J. A. Tainer. 2010. 'Ku and DNA-dependent protein kinase dynamic conformations and assembly regulate DNA binding and the initial non-homologous end joining complex', *J Biol Chem*, 285: 1414-23.

Hanada, K., M. Budzowska, M. Modesti, A. Maas, C. Wyman, J. Essers, and R. Kanaar. 2006. 'The structure-specific endonuclease Mus81-Eme1 promotes conversion of interstrand DNA crosslinks into double-strands breaks', *EMBO J*, 25: 4921-32.

Hartsuiker, E., K. Mizuno, M. Molnar, J. Kohli, K. Ohta, and A. M. Carr. 2009. 'Ctp1CtIP and Rad32Mre11 nuclease activity are required for Rec12Spo11 removal, but Rec12Spo11 removal is dispensable for other MRN-dependent meiotic functions', *Mol Cell Biol*, 29: 1671-81.

Hegde, M. L., T. K. Hazra, and S. Mitra. 2008. 'Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells', *Cell Res*, 18: 27-47.

Hemphill, A. W., D. Bruun, L. Thrun, Y. Akkari, Y. Torimaru, K. Hejna, P. M. Jakobs, J. Hejna, S. Jones, S. B. Olson, and R. E. Moses. 2008. 'Mammalian SNM1 is required for genome stability', *Mol Genet Metab*, 94: 38-45.

Hlatky, L., R. K. Sachs, M. Vazquez, and M. N. Cornforth. 2002. 'Radiation-induced chromosome aberrations: insights gained from biophysical modeling', *Bioessays*, 24: 714-23.

Hlavin, E. M., M. B. Smeaton, A. M. Noronha, C. J. Wilds, and P. S. Miller. 2010. 'Cross-link structure affects replication-independent DNA interstrand cross-link repair in mammalian cells', *Biochemistry*, 49: 3977-88.

Ho, T. V., A. Guainazzi, S. B. Derkunt, M. Enoiu, and O. D. Scharer. 2011. 'Structure-dependent bypass of DNA interstrand crosslinks by translesion synthesis polymerases', *Nucleic Acids Res*, 39: 7455-64.

Hodskinson, M. R., J. Silhan, G. P. Crossan, J. I. Garaycoechea, S. Mukherjee, C. M. Johnson, O. D. Scharer, and K. J. Patel. 2014. 'Mouse SLX4 is a tumor suppressor that stimulates the activity of the nuclease XPF-ERCC1 in DNA crosslink repair', *Mol Cell*, 54: 472-84.

Hoeijmakers, J. H. 2001. 'Genome maintenance mechanisms for preventing cancer', *Nature*, 411: 366-74.

Holley, W. R., and A. Chatterjee. 1996. 'Clusters of DNA induced by ionizing radiation: formation of short DNA fragments. I. Theoretical modeling', *Radiat Res*, 145: 188-99.

Holliday, Robin. 1964. 'A mechanism for gene conversion in fungi', *Genetics Research*, 5: 282-304.

Holloman, W. K. 2011. 'Unraveling the mechanism of BRCA2 in homologous recombination', *Nat Struct Mol Biol*, 18: 748-54.

Horvath, P., and R. Barrangou. 2010. 'CRISPR/Cas, the immune system of bacteria and archaea', *Science*, 327: 167-70.

Huang, J., S. Liu, M. A. Bellani, A. K. Thazhathveetil, C. Ling, J. P. de Winter, Y. Wang, W. Wang, and M. M. Seidman. 2013. 'The DNA translocase FANCM/MHF promotes replication traverse of DNA interstrand crosslinks', *Mol Cell*, 52: 434-46.

Huang, M., J. M. Kim, B. Shiotani, K. Yang, L. Zou, and A. D. D'Andrea. 2010. 'The FANCM/FAAP24 complex is required for the DNA interstrand crosslink-induced checkpoint response', *Mol Cell*, 39: 259-68.

Huang, S. C., J. K. Lee, E. J. Smith, R. T. Doctolero, A. Tajima, S. E. Beck, N. Weidner, and J. M. Carethers. 2011. 'Evidence for an hMSH3 defect in familial hamartomatous polyps', *Cancer*, 117: 492-500.

Huang, Y., J. W. Leung, M. Lowery, N. Matsushita, Y. Wang, X. Shen, D. Huong, M. Takata, J. Chen, and L. Li. 2014. 'Modularized functions of the Fanconi anemia core complex', *Cell Rep*, 7: 1849-57.

Huffman, J. L., O. Sundheim, and J. A. Tainer. 2005. 'DNA base damage recognition and removal: new twists and grooves', *Mutat Res*, 577: 55-76.

Ira, G., A. Pellicioli, A. Balijja, X. Wang, S. Fiorani, W. Carotenuto, G. Liberi, D. Bressan, L. Wan, N. M. Hollingsworth, J. E. Haber, and M. Foiani. 2004. 'DNA end resection, homologous recombination and DNA damage checkpoint activation require CDK1', *Nature*, 431: 1011-7.

Ishiai, M., H. Kitao, A. Smogorzewska, J. Tomida, A. Kinomura, E. Uchida, A. Saberi, E. Kinoshita, E. Kinoshita-Kikuta, T. Koike, S. Tashiro, S. J. Elledge, and M. Takata. 2008. 'FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anemia pathway', *Nat Struct Mol Biol*, 15: 1138-46.

Islas, A. L., J. M. Vos, and P. C. Hanawalt. 1991. 'Differential introduction and repair of psoralen photoadducts to DNA in specific human genes', *Cancer Res*, 51: 2867-73.

Jacobs, A. L., and P. Schar. 2012. 'DNA glycosylases: in DNA repair and beyond', *Chromosoma*, 121: 1-20.

Jiang, W., J. L. Crowe, X. Liu, S. Nakajima, Y. Wang, C. Li, B. J. Lee, R. L. Dubois, C. Liu, X. Yu, L. Lan, and S. Zha. 2015. 'Differential phosphorylation of DNA-PKcs regulates the interplay between end-processing and end-ligation during nonhomologous end-joining', *Mol Cell*, 58: 172-85.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier. 2012. 'A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity', *Science*, 337: 816-21.

Jiricny, J. 2006. 'MutLalpha: at the cutting edge of mismatch repair', *Cell*, 126: 239-41.

Joenje, H., M. Levitus, Q. Waisfisz, A. D'Andrea, I. Garcia-Higuera, T. Pearson, C. G. van Berkel, M. A. Rooimans, N. Morgan, C. G. Mathew, and F. Arwert. 2000. 'Complementation analysis in Fanconi anemia: assignment of the reference FA-H patient to group A', *Am J Hum Genet*, 67: 759-62.

Joenje, H., A. B. Oostra, M. Wijker, F. M. di Summa, C. G. van Berkel, M. A. Rooimans, W. Ebell, M. van Weel, J. C. Pronk, M. Buchwald, and F. Arwert. 1997. 'Evidence for at least eight Fanconi anemia genes', *Am J Hum Genet*, 61: 940-4.

Jones, L., H. Houlden, and S. J. Tabrizi. 2017. 'DNA repair in the trinucleotide repeat disorders', *Lancet Neurol*, 16: 88-96.

Joo, W., G. Xu, N. S. Persky, A. Smogorzewska, D. G. Rudge, O. Buzovetsky, S. J. Elledge, and N. P. Pavletich. 2011. 'Structure of the FANCI-FANCD2 complex: insights into the Fanconi anemia DNA repair pathway', *Science*, 333: 312-6.

Junker, Jan Philipp, Bastiaan Spanjaard, Josi Peterson-Maduro, Anna Alemany, Bo Hu, Maria Florescu, and Alexander van Oudenaarden. 2017. 'Massively parallel clonal analysis using CRISPR/Cas9 induced genetic scars', *bioRxiv*.

Kadyrov, F. A., J. Genschel, Y. Fang, E. Penland, W. Edelmann, and P. Modrich. 2009. 'A possible mechanism for exonuclease 1-independent eukaryotic mismatch repair', *Proc Natl Acad Sci U S A*, 106: 8495-500.

Kalhor, R., P. Mali, and G. M. Church. 2017. 'Rapidly evolving homing CRISPR barcodes', *Nat Methods*, 14: 195-200.

Kalhor, Reza, Kian Kalhor, Leo Mejia, Kathleen Leeper, Amanda Graveline, Prashant Mali, and George M. Church. 2018. 'Developmental barcoding of whole mouse via homing CRISPR', *Science*.

Karanja, K. K., S. W. Cox, J. P. Duxin, S. A. Stewart, and J. L. Campbell. 2012. 'DNA2 and EXO1 in replication-coupled, homology-directed repair and in the interplay between HDR and the FA/BRCA network', *Cell Cycle*, 11: 3983-96.

Karanja, K. K., E. H. Lee, E. A. Hendrickson, and J. L. Campbell. 2014. 'Preventing over-resection by DNA2 helicase/nuclease suppresses repair defects in Fanconi anemia cells', *Cell Cycle*, 13: 1540-50.

Kashiyama, K., Y. Nakazawa, D. T. Pilz, C. Guo, M. Shimada, K. Sasaki, H. Fawcett, J. F. Wing, S. O. Lewin, L. Carr, T. S. Li, K. Yoshiura, A. Utani, A. Hirano, S. Yamashita, D. Greenblatt, T. Nardo, M. Stefanini, D. McGibbon, R. Sarkany, H. Fassihi, Y. Takahashi, Y. Nagayama, N. Mitsutake, A. R. Lehmann, and T. Ogi. 2013. 'Malfunction of nuclease ERCC1-XPF results in diverse clinical manifestations and causes Cockayne syndrome, xeroderma pigmentosum, and Fanconi anemia', *Am J Hum Genet*, 92: 807-19.

Kato, N., Y. Kawasoe, H. Williams, E. Coates, U. Roy, Y. Shi, L. S. Beese, O. D. Scharer, H. Yan, M. E. Gottesman, T. S. Takahashi, and J. Gautier. 2017. 'Sensing and Processing of DNA Interstrand Crosslinks by the Mismatch Repair Pathway', *Cell Rep*, 21: 1375-85.

Kim, H., K. Yang, D. Dejsuphong, and A. D. D'Andrea. 2012. 'Regulation of Rev1 by the Fanconi anemia core complex', *Nat Struct Mol Biol*, 19: 164-70.

Kim, J. M., Y. Kee, A. Gurtan, and A. D. D'Andrea. 2008. 'Cell cycle-dependent chromatin loading of the Fanconi anemia core complex by FANCM/FAAP24', *Blood*, 111: 5215-22.

Kim, J. M., K. Parmar, M. Huang, D. M. Weinstock, C. A. Ruit, J. L. Kutok, and A. D. D'Andrea. 2009. 'Inactivation of murine Usp1 results in genomic instability and a Fanconi anemia phenotype', *Dev Cell*, 16: 314-20.

Klein Douwel, D., R. A. Boonen, D. T. Long, A. A. Szypowska, M. Raschle, J. C. Walter, and P. Knipscheer. 2014. 'XPF-ERCC1 acts in Unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4', *Mol Cell*, 54: 460-71.

Klein Douwel, D., W. S. Hoogenboom, R. A. Boonen, and P. Knipscheer. 2017. 'Recruitment and positioning determine the specific role of the XPF-ERCC1 endonuclease in interstrand crosslink repair', *EMBO J*, 36: 2034-46.

Klug, A. R., M. B. Harbut, R. S. Lloyd, and I. G. Minko. 2012. 'Replication bypass of N2-deoxyguanosine interstrand cross-links by human DNA polymerases eta and iota', *Chem Res Toxicol*, 25: 755-62.

Knipscheer, P., M. Raschle, A. Smogorzewska, M. Enoiu, T. V. Ho, O. D. Scharer, S. J. Elledge, and J. C. Walter. 2009. 'The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair', *Science*, 326: 1698-701.

Koch, S. C., N. Simon, C. Ebert, and T. Carell. 2016. 'Molecular mechanisms of xeroderma pigmentosum (XP) proteins', *Q Rev Biophys*, 49: e5.

Komor, A. C., Y. B. Kim, M. S. Packer, J. A. Zuris, and D. R. Liu. 2016. 'Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage', *Nature*, 533: 420-4.

Konermann, Silvana, Peter Lotfy, Nicholas J Brideau, Jennifer Oki, Maxim N Shokhirev, and Patrick D Hsu. 2018. 'Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors', *Cell*, 173: 665-76. e14.

Kording, Konrad P. 2011. 'Of Toasters and Molecular Ticker Tapes', *PLoS Computational Biology*, 7: e1002291.

Kottemann, M. C., and A. Smogorzewska. 2013. 'Fanconi anaemia and the repair of Watson and Crick DNA crosslinks', *Nature*, 493: 356-63.

Kratz, K., B. Schopf, S. Kaden, A. Sendoel, R. Eberhard, C. Lademann, E. Cannavo, A. A. Sartori, M. O. Hengartner, and J. Jiricny. 2010. 'Deficiency of FANCD2-associated nuclease KIAA1018/FAN1 sensitizes cells to interstrand crosslinking agents', *Cell*, 142: 77-88.

Krokan, H. E., and M. Bjoras. 2013. 'Base excision repair', *Cold Spring Harb Perspect Biol*, 5: a012583.

Kumari, A., I. G. Minko, R. L. Smith, R. S. Lloyd, and A. K. McCullough. 2010. 'Modulation of UvrD helicase activity by covalent DNA-protein cross-links', *J Biol Chem*, 285: 21313-22.

Kunz, C., Y. Saito, and P. Schar. 2009. 'DNA Repair in mammalian cells: Mismatched repair: variations on a theme', *Cell Mol Life Sci*, 66: 1021-38.

Kuo, H. K., J. D. Griffith, and K. N. Kreuzer. 2007. '5-Azacytidine induced methyltransferase-DNA adducts block DNA replication in vivo', *Cancer Res*, 67: 8248-54.

Kuraoka, I., W. R. Kobertz, R. R. Ariza, M. Biggerstaff, J. M. Essigmann, and R. D. Wood. 2000. 'Repair of an interstrand DNA cross-link initiated by ERCC1-XPF repair/recombination nuclease', *J Biol Chem*, 275: 26632-6.

Langevin, F., G. P. Crossan, I. V. Rosado, M. J. Arends, and K. J. Patel. 2011. 'Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice', *Nature*, 475: 53-8.

Lawley, P. D., and D. H. Phillips. 1996. 'DNA adducts from chemotherapeutic agents', *Mutat Res*, 355: 13-40.

Lawrence, C. W. 2004. 'Cellular functions of DNA polymerase zeta and Rev1 protein', *Adv Protein Chem*, 69: 167-203.

Le Breton, C., M. Hennion, P. B. Arimondo, and O. Hyrien. 2011. 'Replication-fork stalling and processing at a single psoralen interstrand crosslink in Xenopus egg extracts', *PLoS One*, 6: e18554.

Lee, Y. C., Y. Cai, H. Mu, S. Broyde, S. Amin, X. Chen, J. H. Min, and N. E. Geacintov. 2014. 'The relationships between XPC binding to conformationally diverse DNA adducts and their excision by the human NER system: is there a correlation?', *DNA Repair (Amst)*, 19: 55-63.

Legerski, R. J. 2010. 'Repair of DNA interstrand cross-links during S phase of the mammalian cell cycle', *Environ Mol Mutagen*, 51: 540-51.

Lehmann, A. R. 2003. 'DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy', *Biochimie*, 85: 1101-11.

Leung, J. W., Y. Wang, K. W. Fong, M. S. Huen, L. Li, and J. Chen. 2012. 'Fanconi anemia (FA) binding protein FAAP20 stabilizes FA complementation group A (FANCA) and participates in interstrand cross-link repair', *Proc Natl Acad Sci U S A*, 109: 4491-6.

Li, L., C. A. Peterson, X. Lu, and R. J. Legerski. 1995. 'Mutations in XPA that prevent association with ERCC1 are defective in nucleotide excision repair', *Mol Cell Biol*, 15: 1993-8.

Li, Z., T. Otevrel, Y. Gao, H. L. Cheng, B. Seed, T. D. Stamato, G. E. Taccioli, and F. W. Alt. 1995. 'The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination', *Cell*, 83: 1079-89.

Liang, C. C., Z. Li, D. Lopez-Martinez, W. V. Nicholson, C. Venien-Bryan, and M. A. Cohn. 2016. 'The FANCD2-FANCI complex is recruited to DNA interstrand crosslinks before monoubiquitination of FANCD2', *Nat Commun*, 7: 12124.

Lim, D., and W. K. Maas. 1989. 'Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in E. coli B', *Cell*, 56: 891-904.

Limbo, O., C. Chahwan, Y. Yamada, R. A. de Bruin, C. Wittenberg, and P. Russell. 2007. 'Ctp1 is a cell-cycle-regulated protein that functions with Mre11 complex to control double-strand break repair by homologous recombination', *Mol Cell*, 28: 134-46.

Lindahl, T. 1993. 'Instability and decay of the primary structure of DNA', *Nature*, 362: 709-15.

———. 2016. 'The Intrinsic Fragility of DNA (Nobel Lecture)', *Angew Chem Int Ed Engl*, 55: 8528-34.

Ling, C., J. Huang, Z. Yan, Y. Li, M. Ohzeki, M. Ishiai, D. Xu, M. Takata, M. Seidman, and W. Wang. 2016. 'Bloom syndrome complex promotes FANCM recruitment to stalled replication forks and facilitates both repair and traverse of DNA interstrand crosslinks', *Cell Discov*, 2: 16047.

Liu, K., Y. F. Wang, C. Cantemir, and M. T. Muller. 2003. 'Endogenous assays of DNA methyltransferases: Evidence for differential activities of DNMT1, DNMT2, and DNMT3 in mammalian cells in vivo', *Mol Cell Biol*, 23: 2709-19.

Liu, T., G. Ghosal, J. Yuan, J. Chen, and J. Huang. 2010. 'FAN1 acts with FANCI-FANCD2 to promote DNA interstrand cross-link repair', *Science*, 329: 693-6.

Liu, Y., W. A. Beard, D. D. Shock, R. Prasad, E. W. Hou, and S. H. Wilson. 2005. 'DNA polymerase beta and flap endonuclease 1 enzymatic specificities sustain DNA synthesis for long patch base excision repair', *J Biol Chem*, 280: 3665-74.

Lobitz, Stephan, and Eunike Velleuer. 2006. 'Guido Fanconi (1892–1979): a jack of all trades', *Nature Reviews Cancer*, 6: 893.

Lomax, M. E., S. Cunniffe, and P. O'Neill. 2004. 'Efficiency of repair of an abasic site within DNA clustered damage sites by mammalian cell nuclear extracts', *Biochemistry*, 43: 11017-26.

Lomax, M. E., M. K. Gulston, and P. O'Neill. 2002. 'Chemical aspects of clustered DNA damage induction by ionising radiation', *Radiat Prot Dosimetry*, 99: 63-8.

Long, D. T., V. Joukov, M. Budzowska, and J. C. Walter. 2014. 'BRCA1 promotes unloading of the CMG helicase from a stalled DNA replication fork', *Mol Cell*, 56: 174-85.

Long, D. T., M. Raschle, V. Joukov, and J. C. Walter. 2011. 'Mechanism of RAD51-dependent DNA interstrand cross-link repair', *Science*, 333: 84-7.

Longerich, S., Y. Kwon, M. S. Tsai, A. S. Hlaing, G. M. Kupfer, and P. Sung. 2014. 'Regulation of FANCD2 and FANCI monoubiquitination by their interaction and by DNA', *Nucleic Acids Res*, 42: 5657-70.

Lopez-Martinez, D., C. C. Liang, and M. A. Cohn. 2016. 'Cellular response to DNA interstrand crosslinks: the Fanconi anemia pathway', *Cell Mol Life Sci*, 73: 3097-114.

Ma, Y., H. Lu, K. Schwarz, and M. R. Lieber. 2005. 'Repair of double-strand DNA breaks by the human nonhomologous DNA end joining pathway: the iterative processing model', *Cell Cycle*, 4: 1193-200.

Mace-Aime, G., S. Couve, B. Khassenov, F. Rosselli, and M. K. Saparbaev. 2010. 'The Fanconi anemia pathway promotes DNA glycosylase-dependent excision of interstrand DNA crosslinks', *Environ Mol Mutagen*, 51: 508-19.

MacKay, C., A. C. Declais, C. Lundin, A. Agostinho, A. J. Deans, T. J. MacArtney, K. Hofmann, A. Gartner, S. C. West, T. Helleday, D. M. Lilley, and J. Rouse. 2010. 'Identification of KIAA1018/FAN1, a DNA repair nuclease recruited to DNA damage by monoubiquitinated FANCD2', *Cell*, 142: 65-76.

Madabhushi, R., F. Gao, A. R. Pfenning, L. Pan, S. Yamakawa, J. Seo, R. Rueda, T. X. Phan, H. Yamakawa, P. C. Pao, R. T. Stott, E. Gjoneska, A. Nott, S. Cho, M. Kellis, and L. H. Tsai. 2015. 'Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes', *Cell*, 161: 1592-605.

Mahaney, B. L., M. Hammel, K. Meek, J. A. Tainer, and S. P. Lees-Miller. 2013. 'XRCC4 and XLF form long helical protein filaments suitable for DNA end protection and alignment to facilitate DNA double strand break repair', *Biochem Cell Biol*, 91: 31-41.

Maillard, O., S. Solyom, and H. Naegeli. 2007. 'An aromatic sensor with aversion to damaged strands confers versatility to DNA repair', *PLoS Biol*, 5: e79.

Mali, Prashant, Luhan Yang, Kevin M. Esvelt, John Aach, Marc Guell, James E. DiCarlo, Julie E. Norville, and George M. Church. 2013. 'RNA-guided human genome engineering via Cas9', *Science (New York, N.Y.)*, 339: 823-26.

Mamrak, Nicholas E, Akiko Shimamura, and Niall G Howlett. 2017. 'Recent discoveries in the molecular pathogenesis of the inherited bone marrow failure syndrome Fanconi anemia', *Blood reviews*, 31: 93-99.

Maor-Shoshani, A., L. B. Meira, X. Yang, and L. D. Samson. 2008. '3-Methyladenine DNA glycosylase is important for cellular resistance to psoralen interstrand cross-links', *DNA Repair (Amst)*, 7: 1399-406.

Marteijn, J. A., J. H. Hoeijmakers, and W. Vermeulen. 2015. 'Check, Check ...Triple Check: Multi-Step DNA Lesion Identification by Nucleotide Excision Repair', *Mol Cell*, 59: 885-6.

Marteijn, J. A., H. Lans, W. Vermeulen, and J. H. Hoeijmakers. 2014. 'Understanding nucleotide excision repair and its roles in cancer and ageing', *Nat Rev Mol Cell Biol*, 15: 465-81.

Maskey, R. S., K. S. Flatten, C. J. Sieben, K. L. Peterson, D. J. Baker, H. J. Nam, M. S. Kim, T. C. Smyrk, Y. Kojima, Y. Machida, A. Santiago, J. M. van Deursen, S. H. Kaufmann, and

Y. J. Machida. 2017. 'Spartan deficiency causes accumulation of Topoisomerase 1 cleavage complexes and tumorigenesis', *Nucleic Acids Res*, 45: 4564-76.

Mason, J. M., I. Das, M. Arlt, N. Patel, S. Kraftson, T. W. Glover, and J. M. Sekiguchi. 2013. 'The SNM1B/APOLLO DNA nuclease functions in resolution of replication stress and maintenance of common fragile site stability', *Hum Mol Genet*, 22: 4901-13.

Mason, J. M., and J. M. Sekiguchi. 2011. 'Snm1B/Apollo functions in the Fanconi anemia pathway in response to DNA interstrand crosslinks', *Hum Mol Genet*, 20: 2549-59.

Mathieu, N., N. Kaczmarek, P. Ruthemann, A. Luch, and H. Naegeli. 2013. 'DNA quality control by a lesion sensor pocket of the xeroderma pigmentosum group D helicase subunit of TFIIH', *Curr Biol*, 23: 204-12.

Matsushita, N., H. Kitao, M. Ishiai, N. Nagashima, S. Hirano, K. Okawa, T. Ohta, D. S. Yu, P. J. McHugh, I. D. Hickson, A. R. Venkitaraman, H. Kurumizaka, and M. Takata. 2005. 'A FancD2-monoubiquitin fusion reveals hidden functions of Fanconi anemia core complex in DNA repair', *Mol Cell*, 19: 841-7.

McConnell, M. J., M. R. Lindberg, K. J. Brennand, J. C. Piper, T. Voet, C. Cowing-Zitron, S. Shumilina, R. S. Lasken, J. R. Vermeesch, I. M. Hall, and F. H. Gage. 2013. 'Mosaic copy number variation in human neurons', *Science*, 342: 632-7.

McConnell, M. J., J. V. Moran, A. Abyzov, S. Akbarian, T. Bae, I. Cortes-Ciriano, J. A. Erwin, L. Fasching, D. A. Flasch, D. Freed, J. Ganz, A. E. Jaffe, K. Y. Kwan, M. Kwon, M. A. Lodato, R. E. Mills, A. C. M. Paquola, R. E. Rodin, C. Rosenbluh, N. Sestan, M. A. Sherman, J. H. Shin, S. Song, R. E. Straub, J. Thorpe, D. R. Weinberger, A. E. Urban, B. Zhou, F. H. Gage, T. Lehner, G. Senthil, C. A. Walsh, A. Chess, E. Courchesne, J. G. Gleeson, J. M. Kidd, P. J. Park, J. Pevsner, F. M. Vaccarino, and Network Brain Somatic Mosaicism. 2017. 'Intersection of diverse neuronal genomes and neuropsychiatric disease: The Brain Somatic Mosaicism Network', *Science*, 356.

McHugh, P. J., W. R. Sones, and J. A. Hartley. 2000. 'Repair of intermediate structures produced at DNA interstrand cross-links in Saccharomyces cerevisiae', *Mol Cell Biol*, 20: 3425-33.

McKenna, A., G. M. Findlay, J. A. Gagnon, M. S. Horwitz, A. F. Schier, and J. Shendure. 2016. 'Whole-organism lineage tracing by combinatorial and cumulative genome editing', *Science*, 353: aaf7907.

McVey, M. 2010. 'Strategies for DNA interstrand crosslink repair: insights from worms, flies, frogs, and slime molds', *Environ Mol Mutagen*, 51: 646-58.

Menon, V., and L. F. Povirk. 2016. 'End-processing nucleases and phosphodiesterases: An elite supporting cast for the non-homologous end joining pathway of DNA double-strand break repair', *DNA Repair (Amst)*, 43: 57-68.

Mimee, M., A. C. Tucker, C. A. Voigt, and T. K. Lu. 2015. 'Programming a Human Commensal Bacterium, Bacteroides thetaiotaomicron, to Sense and Respond to Stimuli in the Murine Gut Microbiota', *Cell Syst*, 1: 62-71.

Mimitou, E. P., and L. S. Symington. 2008. 'Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing', *Nature*, 455: 770-4.

———. 2011. 'DNA end resection--unraveling the tail', *DNA Repair (Amst)*, 10: 344-8.

Min, J. H., and N. P. Pavletich. 2007. 'Recognition of DNA damage by the Rad4 nucleotide excision repair protein', *Nature*, 449: 570-5.

Minko, I. G., M. B. Harbut, I. D. Kozekov, A. Kozekova, P. M. Jakobs, S. B. Olson, R. E. Moses, T. M. Harris, C. J. Rizzo, and R. S. Lloyd. 2008. 'Role for DNA polymerase kappa in the processing of N2-N2-guanine interstrand cross-links', *J Biol Chem*, 283: 17075-82.

Mirchandani, K. D., R. M. McCaffrey, and A. D. D'Andrea. 2008. 'The Fanconi anemia core complex is required for efficient point mutagenesis and Rev1 foci assembly', *DNA Repair (Amst)*, 7: 902-11.

Moore, Christopher L., Louis J. Papa, and Matthew D. Shoulders. 2018. 'A Processive Protein Chimera Introduces Mutations across Defined DNA Regions In Vivo', *Journal of the American Chemical Society*, 140: 11560-64.

Moore, J. K., and J. E. Haber. 1996. 'Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae', *Mol Cell Biol*, 16: 2164-73.

Morocz, M., E. Zsigmond, R. Toth, M. Z. Enyedi, L. Pinter, and L. Haracska. 2017. 'DNA-dependent protease activity of human Spartan facilitates replication of DNA-protein crosslink-containing DNA', *Nucleic Acids Res*, 45: 3172-88.

Mourgues, S., M. E. Lomax, and P. O'Neill. 2007. 'Base excision repair processing of abasic site/single-strand break lesions within clustered damage sites associated with XRCC1 deficiency', *Nucleic Acids Res*, 35: 7676-87.

Muniandy, P. A., J. Liu, A. Majumdar, S. T. Liu, and M. M. Seidman. 2010. 'DNA interstrand crosslink repair in mammalian cells: step by step', *Crit Rev Biochem Mol Biol*, 45: 23-49.

Muniandy, P. A., D. Thapa, A. K. Thazhathveetil, S. T. Liu, and M. M. Seidman. 2009. 'Repair of laser-localized DNA interstrand cross-links in G1 phase mammalian cells', *J Biol Chem*, 284: 27908-17.

Murina, O., C. von Aesch, U. Karakus, L. P. Ferretti, H. A. Bolck, K. Hanggi, and A. A. Sartori. 2014. 'FANCD2 and CtIP cooperate to repair DNA interstrand crosslinks', *Cell Rep*, 7: 1030-8.

Mutreja, K., J. Krietsch, J. Hess, S. Ursich, M. Berti, F. K. Roessler, R. Zellweger, M. Patra, G. Gasser, and M. Lopes. 2018. 'ATR-Mediated Global Fork Slowing and Reversal Assist Fork Traverse and Prevent Chromosomal Breakage at DNA Interstrand Cross-Links', *Cell Rep*, 24: 2629-42 e5.

Nakanishi, K., F. Cavallo, L. Perrouault, C. Giovannangeli, M. E. Moynahan, M. Barchi, E. Brunet, and M. Jasin. 2011. 'Homology-directed Fanconi anemia pathway cross-link repair is dependent on DNA replication', *Nat Struct Mol Biol*, 18: 500-3.

Nakano, T., A. Katafuchi, M. Matsubara, H. Terato, T. Tsuboi, T. Masuda, T. Tatsumoto, S. P. Pack, K. Makino, D. L. Croteau, B. Van Houten, K. Iijima, H. Tauchi, and H. Ide. 2009. 'Homologous recombination but not nucleotide excision repair plays a pivotal role in tolerance of DNA-protein cross-links in mammalian cells', *J Biol Chem*, 284: 27065-76.

Nakano, T., M. Miyamoto-Matsubara, M. I. Shoulkamy, A. M. Salem, S. P. Pack, Y. Ishimi, and H. Ide. 2013. 'Translocation and stability of replicative DNA helicases upon encountering DNA-protein cross-links', *J Biol Chem*, 288: 4649-58.

Nakano, T., S. Morishita, A. Katafuchi, M. Matsubara, Y. Horikawa, H. Terato, A. M. Salem, S. Izumi, S. P. Pack, K. Makino, and H. Ide. 2007. 'Nucleotide excision repair and homologous recombination systems commit differentially to the repair of DNA-protein crosslinks', *Mol Cell*, 28: 147-58.

Nakano, T., R. Ouchi, J. Kawazoe, S. P. Pack, K. Makino, and H. Ide. 2012. 'T7 RNA polymerases backed up by covalently trapped proteins catalyze highly error prone transcription', *J Biol Chem*, 287: 6562-72.

Nakano, T., X. Xu, A. M. H. Salem, M. I. Shoulkamy, and H. Ide. 2017. 'Radiation-induced DNA-protein cross-links: Mechanisms and biological significance', *Free Radic Biol Med*, 107: 136-45.

Neale, M. J., and S. Keeney. 2006. 'Clarifying the mechanics of DNA strand exchange in meiotic recombination', *Nature*, 442: 153-8.

Niedernhofer, L. J., A. S. Lalai, and J. H. Hoeijmakers. 2005. 'Fanconi anemia (cross)linked to DNA repair', *Cell*, 123: 1191-8.

Niedernhofer, L. J., H. Odijk, M. Budzowska, E. van Drunen, A. Maas, A. F. Theil, J. de Wit, N. G. Jaspers, H. B. Beverloo, J. H. Hoeijmakers, and R. Kanaar. 2004. 'The structure-specific endonuclease Ercc1-Xpf is required to resolve DNA interstrand cross-link-induced double-strand breaks', *Mol Cell Biol*, 24: 5776-87.

Niedzwiedz, W., G. Mosedale, M. Johnson, C. Y. Ong, P. Pace, and K. J. Patel. 2004. 'The Fanconi anaemia gene FANCC promotes homologous recombination and error-prone DNA repair', *Mol Cell*, 15: 607-20.

Nijman, S. M., T. T. Huang, A. M. Dirac, T. R. Brummelkamp, R. M. Kerkhoven, A. D. D'Andrea, and R. Bernards. 2005. 'The deubiquitinating enzyme USP1 regulates the Fanconi anemia pathway', *Mol Cell*, 17: 331-9.

Nilsen, L., R. J. Forstrom, M. Bjoras, and I. Alseth. 2012. 'AP endonuclease independent repair of abasic sites in Schizosaccharomyces pombe', *Nucleic Acids Res*, 40: 2000-9.

Nishida, K., T. Arazoe, N. Yachie, S. Banno, M. Kakimoto, M. Tabata, M. Mochizuki, A. Miyabe, M. Araki, K. Y. Hara, Z. Shimatani, and A. Kondo. 2016. 'Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems', *Science*, 353.

Noll, D. M., T. M. Mason, and P. S. Miller. 2006. 'Formation and repair of interstrand cross-links in DNA', *Chem Rev*, 106: 277-301.

Northcott, P. A., D. J. Shih, J. Peacock, L. Garzia, A. S. Morrissy, T. Zichner, A. M. Stutz, A. Korshunov, J. Reimand, S. E. Schumacher, R. Beroukhim, D. W. Ellison, C. R. Marshall, A. C. Lionel, S. Mack, A. Dubuc, Y. Yao, V. Ramaswamy, B. Luu, A. Rolider, F. M. Cavalli, X. Wang, M. Remke, X. Wu, R. Y. Chiu, A. Chu, E. Chuah, R. D. Corbett, G. R. Hoad, S. D. Jackman, Y. Li, A. Lo, K. L. Mungall, K. M. Nip, J. Q. Qian, A. G. Raymond, N. T. Thiessen, R. J. Varhol, I. Birol, R. A. Moore, A. J. Mungall, R. Holt, D. Kawauchi, M. F. Roussel, M. Kool, D. T. Jones, H. Witt, L. A. Fernandez, A. M. Kenney, R. J. Wechsler-Reya, P. Dirks, T. Aviv, W. A. Grajkowska, M. Perek-Polnik, C. C. Haberler, O. Delattre, S. S. Reynaud, F. F. Doz, S. S. Pernet-Fattet, B. K. Cho, S. K. Kim, K. C. Wang, W. Scheurlen, C. G. Eberhart, M. Fevre-Montange, A. Jouvet, I. F. Pollack, X. Fan, K. M. Muraszko, G. Y. Gillespie, C. Di Rocco, L. Massimi, E. M. Michiels, N. K. Kloosterhof, P. J. French, J. M. Kros, J. M. Olson, R. G. Ellenbogen, K. Zitterbart, L. Kren, R. C. Thompson, M. K. Cooper, B. Lach, R. E. McLendon, D. D. Bigner, A.

Fontebasso, S. Albrecht, N. Jabado, J. C. Lindsey, S. Bailey, N. Gupta, W. A. Weiss, L. Bognar, A. Klekner, T. E. Van Meter, T. Kumabe, T. Tominaga, S. K. Elbabaa, J. R. Leonard, J. B. Rubin, L. M. Liau, E. G. Van Meir, M. Fouladi, H. Nakamura, G. Cinalli, M. Garami, P. Hauser, A. G. Saad, A. Iolascon, S. Jung, C. G. Carlotti, R. Vibhakar, Y. S. Ra, S. Robinson, M. Zollo, C. C. Faria, J. A. Chan, M. L. Levy, P. H. Sorensen, M. Meyerson, S. L. Pomeroy, Y. J. Cho, G. D. Bader, U. Tabori, C. E. Hawkins, E. Bouffet, S. W. Scherer, J. T. Rutka, D. Malkin, S. C. Clifford, S. J. Jones, J. O. Korbel, S. M. Pfister, M. A. Marra, and M. D. Taylor. 2012. 'Subgroup-specific structural variation across 1,000 medulloblastoma genomes', *Nature*, 488: 49-56.

Novakova, O., J. Kasparkova, J. Malina, G. Natile, and V. Brabec. 2003. 'DNA-protein cross-linking by trans-[PtCl(2)(E-iminoether)(2)]. A concept for activation of the trans geometry in platinum antitumor complexes', *Nucleic Acids Res*, 31: 6450-60.

Nowotny, M., and V. Gaur. 2016. 'Structure and mechanism of nucleases regulated by SLX4', *Curr Opin Struct Biol*, 36: 97-105.

Oestergaard, V. H., F. Langevin, H. J. Kuiken, P. Pace, W. Niedzwiedz, L. J. Simpson, M. Ohzeki, M. Takata, J. E. Sale, and K. J. Patel. 2007. 'Deubiquitination of FANCD2 is required for DNA crosslink repair', *Mol Cell*, 28: 798-809.

Ohmori, H., E. C. Friedberg, R. P. Fuchs, M. F. Goodman, F. Hanaoka, D. Hinkle, T. A. Kunkel, C. W. Lawrence, Z. Livneh, T. Nohmi, L. Prakash, S. Prakash, T. Todo, G. C. Walker, Z. Wang, and R. Woodgate. 2001. 'The Y-family of DNA polymerases', *Mol Cell*, 8: 7-8.

Orelli, B., T. B. McClendon, O. V. Tsodikov, T. Ellenberger, L. J. Niedernhofer, and O. D. Scharer. 2010. 'The XPA-binding domain of ERCC1 is required for nucleotide excision repair but not other DNA repair pathways', *J Biol Chem*, 285: 3705-12.

Park, J. Y., E. L. Virts, A. Jankowska, C. Wiek, M. Othman, S. C. Chakraborty, G. H. Vance, F. S. Alkuraya, H. Hanenberg, and P. R. Andreassen. 2016. 'Complementation of hypersensitivity to DNA interstrand crosslinking agents demonstrates that XRCC2 is a Fanconi anaemia gene', *J Med Genet*, 53: 672-80.

Parsons, J. L., Dianova, II, S. L. Allinson, and G. L. Dianov. 2005. 'DNA polymerase beta promotes recruitment of DNA ligase III alpha-XRCC1 to sites of base excision repair', *Biochemistry*, 44: 10613-9.

Pavlov, Y. I., I. M. Mian, and T. A. Kunkel. 2003. 'Evidence for preferential mismatch repair of lagging strand DNA replication errors in yeast', *Curr Biol*, 13: 744-8.

Perli, S. D., C. H. Cui, and T. K. Lu. 2016. 'Continuous genetic recording with self-targeting CRISPR-Cas in human cells', *Science*, 353.

Pettijohn, D., and P. Hanawalt. 1964. 'Evidence for Repair-Replication of Ultraviolet Damaged DNA in Bacteria', *J Mol Biol*, 9: 395-410.

Pichierri, P., and F. Rosselli. 2004. 'The DNA crosslink-induced S-phase checkpoint depends on ATR-CHK1 and ATR-NBS1-FANCD2 pathways', *EMBO J*, 23: 1178-87.

Plosky, B. S., A. E. Vidal, A. R. Fernandez de Henestrosa, M. P. McLenigan, J. P. McDonald, S. Mead, and R. Woodgate. 2006. 'Controlling the subcellular localization of DNA polymerases iota and eta via interactions with ubiquitin', *EMBO J*, 25: 2847-55.

Poduri, A., G. D. Evrony, X. Cai, and C. A. Walsh. 2013. 'Somatic mutation, genomic variation, and neurological disease', *Science*, 341: 1237758.

Pommier, Y., S. Y. Huang, R. Gao, B. B. Das, J. Murai, and C. Marchand. 2014. 'Tyrosyl-DNA-phosphodiesterases (TDP1 and TDP2)', *DNA Repair (Amst)*, 19: 114-29.

Prakash, R., Y. Zhang, W. Feng, and M. Jasin. 2015. 'Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins', *Cold Spring Harb Perspect Biol*, 7: a016600.

Prasad, R., J. K. Horton, P. D. Chastain, 2nd, N. R. Gassman, B. D. Freudenthal, E. W. Hou, and S. H. Wilson. 2014. 'Suicidal cross-linking of PARP-1 to AP site intermediates in cells undergoing base excision repair', *Nucleic Acids Res*, 42: 6337-51.

Price, N. E., K. M. Johnson, J. Wang, M. I. Fekry, Y. Wang, and K. S. Gates. 2014. 'Interstrand DNA-DNA cross-link formation between adenine residues and abasic sites in duplex DNA', *J Am Chem Soc*, 136: 3483-90.

Qiao, F., A. Moss, and G. M. Kupfer. 2001. 'Fanconi anemia proteins localize to chromatin and the nuclear matrix in a DNA damage- and cell cycle-regulated manner', *J Biol Chem*, 276: 23391-6.

Quinones, J. L., and B. Demple. 2016. 'When DNA repair goes wrong: BER-generated DNA-protein crosslinks to oxidative lesions', *DNA Repair (Amst)*, 44: 103-09.

Radhakrishnan, S. K., N. Jette, and S. P. Lees-Miller. 2014. 'Non-homologous end joining: emerging themes and unanswered questions', *DNA Repair (Amst)*, 17: 2-8.

Raschle, M., P. Knipscheer, M. Enoiu, T. Angelov, J. Sun, J. D. Griffith, T. E. Ellenberger, O. D. Scharer, and J. C. Walter. 2008. 'Mechanism of replication-coupled DNA interstrand crosslink repair', *Cell*, 134: 969-80.

Rasmussen, R. E., and R. B. Painter. 1964. 'Evidence for Repair of Ultra-Violet Damaged Deoxyribonucleic Acid in Cultured Mammalian Cells', *Nature*, 203: 1360-2.

Rausch, T., D. T. Jones, M. Zapatka, A. M. Stutz, T. Zichner, J. Weischenfeldt, N. Jager, M. Remke, D. Shih, P. A. Northcott, E. Pfaff, J. Tica, Q. Wang, L. Massimi, H. Witt, S. Bender, S. Pleier, H. Cin, C. Hawkins, C. Beck, A. von Deimling, V. Hans, B. Brors, R. Eils, W. Scheurlen, J. Blake, V. Benes, A. E. Kulozik, O. Witt, D. Martin, C. Zhang, R. Porat, D. M. Merino, J. Wasserman, N. Jabado, A. Fontebasso, L. Bullinger, F. G. Rucker, K. Dohner, H. Dohner, J. Koster, J. J. Molenaar, R. Versteeg, M. Kool, U. Tabori, D. Malkin, A. Korshunov, M. D. Taylor, P. Lichter, S. M. Pfister, and J. O. Korbel. 2012. 'Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations', *Cell*, 148: 59-71.

Rickman, K. A., F. P. Lach, A. Abhyankar, F. X. Donovan, E. M. Sanborn, J. A. Kennedy, C. Sougnez, S. B. Gabriel, O. Elemento, S. C. Chandrasekharappa, D. Schindler, A. D. Auerbach, and A. Smogorzewska. 2015. 'Deficiency of UBE2T, the E2 Ubiquitin Ligase Necessary for FANCD2 and FANCI Ubiquitination, Causes FA-T Subtype of Fanconi Anemia', *Cell Rep*, 12: 35-41.

Robertson, A. B., A. Klungland, T. Rognes, and I. Leiros. 2009. 'DNA repair in mammalian cells: Base excision repair: the long and short of it', *Cell Mol Life Sci*, 66: 981-93.

Rohleder, F., J. Huang, Y. Xue, J. Kuper, A. Round, M. Seidman, W. Wang, and C. Kisker. 2016. 'FANCM interacts with PCNA to promote replication traverse of DNA interstrand crosslinks', *Nucleic Acids Res*, 44: 3219-32.

Roquet, N., A. P. Soleimany, A. C. Ferris, S. Aaronson, and T. K. Lu. 2016. 'Synthetic recombinase-based state machines in living cells', *Science*, 353: aad8559.

Rosado, I. V., F. Langevin, G. P. Crossan, M. Takata, and K. J. Patel. 2011. 'Formaldehyde catabolism is essential in cells deficient for the Fanconi anemia DNA-repair pathway', *Nat Struct Mol Biol*, 18: 1432-4.

Rosenberg, P. S., H. Tamary, and B. P. Alter. 2011. 'How high are carrier frequencies of rare recessive syndromes? Contemporary estimates for Fanconi Anemia in the United States and Israel', *Am J Med Genet A*, 155A: 1877-83.

Rothfuss, A., and M. Grompe. 2004. 'Repair kinetics of genomic interstrand DNA cross-links: evidence for DNA double-strand break-dependent activation of the Fanconi anemia/BRCA pathway', *Mol Cell Biol*, 24: 123-34.

Roy, U., S. Mukherjee, A. Sharma, E. G. Frank, and O. D. Scharer. 2016. 'The structure and duplex context of DNA interstrand crosslinks affects the activity of DNA polymerase eta', *Nucleic Acids Res*, 44: 7281-91.

Roy, U., and O. D. Scharer. 2016. 'Involvement of translesion synthesis DNA polymerases in DNA interstrand crosslink repair', *DNA Repair (Amst)*, 44: 33-41.

Rydberg, B. 1996. 'Clusters of DNA damage induced by ionizing radiation: formation of short DNA fragments. II. Experimental detection', *Radiat Res*, 145: 200-9.

Sage, E., and L. Harrison. 2011. 'Clustered DNA lesion repair in eukaryotes: relevance to mutagenesis and cell survival', *Mutat Res*, 711: 123-33.

Sarkar, S., A. A. Davies, H. D. Ulrich, and P. J. McHugh. 2006. 'DNA interstrand crosslink repair during G1 involves nucleotide excision repair and DNA polymerase zeta', *EMBO J*, 25: 1285-94.

Sasaki, M. S., and A. Tonomura. 1973. 'A high susceptibility of Fanconi's anemia to chromosome breakage by DNA cross-linking agents', *Cancer Res*, 33: 1829-36.

Sato, K., K. Toda, M. Ishiai, M. Takata, and H. Kurumizaka. 2012. 'DNA robustly stimulates FANCD2 monoubiquitylation in the complex with FANCI', *Nucleic Acids Res*, 40: 4553-61.

Sawyer, S. L., L. Tian, M. Kahkonen, J. Schwartzentruber, M. Kircher, Genomics University of Washington Centre for Mendelian, Forge Canada Consortium, J. Majewski, D. A. Dyment, A. M. Innes, K. M. Boycott, L. A. Moreau, J. S. Moilanen, and R. A. Greenberg. 2015. 'Biallelic mutations in BRCA1 cause a new Fanconi anemia subtype', *Cancer Discov*, 5: 135-42.

Schermerhorn, K. M., and S. Delaney. 2014. 'A chemical and kinetic perspective on base excision repair of DNA', *Acc Chem Res*, 47: 1238-46.

Schlacher, K., H. Wu, and M. Jasin. 2012. 'A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2', *Cancer Cell*, 22: 106-16.

Schmidt, F., M. Y. Cherepkova, and R. J. Platt. 2018. 'Transcriptional recording by CRISPR spacer acquisition from RNA', *Nature*, 562: 380-85.

Sekiguchi, J. M., Y. Gao, Y. Gu, K. Frank, Y. Sun, J. Chaudhuri, C. Zhu, H. L. Cheng, J. Manis, D. Ferguson, L. Davidson, M. E. Greenberg, and F. W. Alt. 1999. 'Nonhomologous end-joining proteins are required for V(D)J recombination, normal growth, and neurogenesis', *Cold Spring Harb Symp Quant Biol*, 64: 169-81.

Semenenko, V. A., R. D. Stewart, and E. J. Ackerman. 2005. 'Monte Carlo simulation of base and nucleotide excision repair of clustered DNA damage sites. I. Model properties and predicted trends', *Radiat Res*, 164: 180-93.

Semlow, D. R., J. Zhang, M. Budzowska, A. C. Drohat, and J. C. Walter. 2016. 'Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase', *Cell*, 167: 498-511 e14.

Sengerova, B., C. K. Allerston, M. Abu, S. Y. Lee, J. Hartley, K. Kiakos, C. J. Schofield, J. A. Hartley, O. Gileadi, and P. J. McHugh. 2012. 'Characterization of the human SNM1A and SNM1B/Apollo DNA repair exonucleases', *J Biol Chem*, 287: 26254-67.

Setlow, R. B., and W. L. Carrier. 1964. 'The Disappearance of Thymine Dimers from DNA: An Error-Correcting Mechanism', *Proc Natl Acad Sci U S A*, 51: 226-31.

Shachar, S., O. Ziv, S. Avkin, S. Adar, J. Wittschieben, T. Reissner, S. Chaney, E. C. Friedberg, Z. Wang, T. Carell, N. Geacintov, and Z. Livneh. 2009. 'Two-polymerase mechanisms dictate error-free and error-prone translesion DNA synthesis in mammals', *EMBO J*, 28: 383-93.

Sharon, E., S. A. Chen, N. M. Khosla, J. D. Smith, J. K. Pritchard, and H. B. Fraser. 2018. 'Functional Genetic Variants Revealed by Massively Parallel Precise Genome Editing', *Cell*, 175: 544-57 e16.

Shen, M. W., M. Arbab, J. Y. Hsu, D. Worstell, S. J. Culbertson, O. Krabbe, C. A. Cassa, D. R. Liu, D. K. Gifford, and R. I. Sherwood. 2018. 'Predictable and precise template-free CRISPR editing of pathogenic variants', *Nature*, 563: 646-51.

Shen, X., S. Jun, L. E. O'Neal, E. Sonoda, M. Bemark, J. E. Sale, and L. Li. 2006. 'REV3 and REV1 play major roles in recombination-independent repair of DNA interstrand cross-links mediated by monoubiquitinated proliferating cell nuclear antigen (PCNA)', *J Biol Chem*, 281: 13869-72.

Shereda, R. D., Y. Machida, and Y. J. Machida. 2010. 'Human KIAA1018/FAN1 localizes to stalled replication forks via its ubiquitin-binding domain', *Cell Cycle*, 9: 3977-83.

Sheth, R. U., S. S. Yim, F. L. Wu, and H. H. Wang. 2017. 'Multiplex recording of cellular events over time on CRISPR biological tape', *Science*, 358: 1457-61.

Shibata, A., D. Moiani, A. S. Arvai, J. Perry, S. M. Harding, M. M. Genois, R. Maity, S. van Rossum-Fikkert, A. Kertokalio, F. Romoli, A. Ismail, E. Ismalaj, E. Petricci, M. J. Neale, R. G. Bristow, J. Y. Masson, C. Wyman, P. A. Jeggo, and J. A. Tainer. 2014. 'DNA double-strand break repair pathway choice is directed by distinct MRE11 nuclease activities', *Mol Cell*, 53: 7-18.

Shipman, S. L., J. Nivala, J. D. Macklis, and G. M. Church. 2016. 'Molecular recordings by directed CRISPR spacer acquisition', *Science*, 353: aaf1175.

———. 2017. 'CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria', *Nature*, 547: 345-49.

Shivji, M. K., V. N. Podust, U. Hubscher, and R. D. Wood. 1995. 'Nucleotide excision repair DNA synthesis by DNA polymerase epsilon in the presence of PCNA, RFC, and RPA', *Biochemistry*, 34: 5011-7.

Shur, Andrey, and Richard M Murray. 2017. 'Repressing Integrase attachment site operation with CRISPR-Cas9 in E. coli', *bioRxiv*.

Sijbers, A. M., W. L. de Laat, R. R. Ariza, M. Biggerstaff, Y. F. Wei, J. G. Moggs, K. C. Carter, B. K. Shell, E. Evans, M. C. de Jong, S. Rademakers, J. de Rooij, N. G. Jaspers, J. H. Hoeijmakers, and R. D. Wood. 1996. 'Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease', *Cell*, 86: 811-22.

Silas, Sukrit, Georg Mohr, David J. Sidote, Laura M. Markham, Antonio Sanchez-Amat, Devaki Bhaya, Alan M. Lambowitz, and Andrew Z. Fire. 2016. 'Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein', *Science (New York, N.Y.)*, 351: aad4234-aad34.

Singh, T. R., D. Saro, A. M. Ali, X. F. Zheng, C. H. Du, M. W. Killen, A. Sachpatzidis, K. Wahengbam, A. J. Pierce, Y. Xiong, P. Sung, and A. R. Meetei. 2010. 'MHF1-MHF2, a histone-fold-containing protein complex, participates in the Fanconi anemia pathway via FANCM', *Mol Cell*, 37: 879-86.

Singh, V., and P. Das. 2013. 'Condensation of DNA--a putative obstruction for repair process in abasic clustered DNA damage', *DNA Repair (Amst)*, 12: 450-7.

Siuti, P., J. Yazbek, and T. K. Lu. 2013. 'Synthetic circuits integrating logic and memory in living cells', *Nat Biotechnol*, 31: 448-52.

Slupphaug, G., C. D. Mol, B. Kavli, A. S. Arvai, H. E. Krokan, and J. A. Tainer. 1996. 'A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA', *Nature*, 384: 87-92.

Smeaton, M. B., E. M. Hlavin, T. McGregor Mason, A. M. Noronha, C. J. Wilds, and P. S. Miller. 2008. 'Distortion-dependent unhooking of interstrand cross-links in mammalian cell extracts', *Biochemistry*, 47: 9920-30.

Smeaton, M. B., E. M. Hlavin, A. M. Noronha, S. P. Murphy, C. J. Wilds, and P. S. Miller. 2009. 'Effect of cross-link structure on DNA interstrand cross-link repair synthesis', *Chem Res Toxicol*, 22: 1285-97.

Smogorzewska, A., R. Desetty, T. T. Saito, M. Schlabach, F. P. Lach, M. E. Sowa, A. B. Clark, T. A. Kunkel, J. W. Harper, M. P. Colaiacovo, and S. J. Elledge. 2010. 'A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair', *Mol Cell*, 39: 36-47.

Smogorzewska, A., S. Matsuoka, P. Vinciguerra, E. R. McDonald, 3rd, K. E. Hurov, J. Luo, B. A. Ballif, S. P. Gygi, K. Hofmann, A. D. D'Andrea, and S. J. Elledge. 2007. 'Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair', *Cell*, 129: 289-301.

Spivak, G. 2015. 'Nucleotide excision repair in humans', *DNA Repair (Amst)*, 36: 13-8.

Spivak, G., and A. K. Ganesan. 2014. 'The complex choreography of transcription-coupled repair', *DNA Repair (Amst)*, 19: 64-70.

Stark, W. M. 2017. 'Making serine integrases work for us', *Curr Opin Microbiol*, 38: 130-36.

Sternberg, S. H., H. Richter, E. Charpentier, and U. Qimron. 2016. 'Adaptation in CRISPR-Cas Systems', *Mol Cell*, 61: 797-808.

Stingele, J., and S. Jentsch. 2015. 'DNA-protein crosslink repair', *Nat Rev Mol Cell Biol*, 16: 455-60.

Stingele, J., M. S. Schwarz, N. Bloemeke, P. G. Wolf, and S. Jentsch. 2014. 'A DNA-dependent protease involved in DNA-protein crosslink repair', *Cell*, 158: 327-38.

Stivers, J. T. 2008. 'Extrahelical damaged base recognition by DNA glycosylase enzymes', *Chemistry*, 14: 786-93.

Stivers, J. T., and Y. L. Jiang. 2003. 'A mechanistic perspective on the chemistry of DNA repair glycosylases', *Chem Rev*, 103: 2729-59.

Stoepker, C., K. Hain, B. Schuster, Y. Hilhorst-Hofstee, M. A. Rooimans, J. Steltenpool, A. B. Oostra, K. Eirich, E. T. Korthof, A. W. Nieuwint, N. G. Jaspers, T. Bettecken, H. Joenje, D. Schindler, J. Rouse, and J. P. de Winter. 2011. 'SLX4, a coordinator of structure-specific endonucleases, is mutated in a new Fanconi anemia subtype', *Nat Genet*, 43: 138-41.

Stracker, T. H., and J. H. Petrini. 2011. 'The MRE11 complex: starting from the ends', *Nat Rev Mol Cell Biol*, 12: 90-103.

Suberbielle, E., B. Djukic, M. Evans, D. H. Kim, P. Taneja, X. Wang, M. Finucane, J. Knox, K. Ho, N. Devidze, E. Masliah, and L. Mucke. 2015. 'DNA repair factor BRCA1 depletion occurs in Alzheimer brains and impairs cognitive function in mice', *Nat Commun*, 6: 8897.

Suberbielle, E., P. E. Sanchez, A. V. Kravitz, X. Wang, K. Ho, K. Eilertson, N. Devidze, A. C. Kreitzer, and L. Mucke. 2013. 'Physiologic brain activity causes DNA double-strand breaks in neurons, with exacerbation by amyloid-beta', *Nat Neurosci*, 16: 613-21.

Suhasini, A. N., J. A. Sommers, P. A. Muniandy, Y. Coulombe, S. B. Cantor, J. Y. Masson, M. M. Seidman, and R. M. Brosh, Jr. 2013. 'Fanconi anemia group J helicase and MRE11 nuclease interact to facilitate the DNA damage response', *Mol Cell Biol*, 33: 2212-27.

Sun, H., L. He, H. Wu, F. Pan, X. Wu, J. Zhao, Z. Hu, C. Sekhar, H. Li, L. Zheng, H. Chen, B. H. Shen, and Z. Guo. 2017. 'The FEN1 L209P mutation interferes with long-patch base excision repair and induces cellular transformation', *Oncogene*, 36: 194-207.

Suwaki, N., K. Klare, and M. Tarsounas. 2011. 'RAD51 paralogs: roles in DNA damage signalling, recombinational repair and tumorigenesis', *Semin Cell Dev Biol*, 22: 898-905.

Svendsen, J. M., A. Smogorzewska, M. E. Sowa, B. C. O'Connell, S. P. Gygi, S. J. Elledge, and J. W. Harper. 2009. 'Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair', *Cell*, 138: 63-77.

Svilar, D., E. M. Goellner, K. H. Almeida, and R. W. Sobol. 2011. 'Base excision repair and lesion-dependent subpathways for repair of oxidative DNA damage', *Antioxid Redox Signal*, 14: 2491-507.

Swenberg, J. A., K. Lu, B. C. Moeller, L. Gao, P. B. Upton, J. Nakamura, and T. B. Starr. 2011. 'Endogenous versus exogenous DNA adducts: their role in carcinogenesis, epidemiology, and risk assessment', *Toxicol Sci*, 120 Suppl 1: S130-45.

Tacconi, E. M., and M. Tarsounas. 2015. 'How homologous recombination maintains telomere integrity', *Chromosoma*, 124: 119-30.

Takahashi, D., K. Sato, E. Hirayama, M. Takata, and H. Kurumizaka. 2015. 'Human FAN1 promotes strand incision in 5'-flapped DNA complexed with RPA', *J Biochem*, 158: 263-70.

Tang, J., and G. Chu. 2002. 'Xeroderma pigmentosum complementation group E and UV-damaged DNA-binding protein', *DNA Repair (Amst)*, 1: 601-16.

Tang, Weixin, and David R. Liu. 2018. 'Rewritable multi-event analog recording in bacterial and mammalian cells', *Science*.

Taniguchi, T., I. Garcia-Higuera, P. R. Andreassen, R. C. Gregory, M. Grompe, and A. D. D'Andrea. 2002. 'S-phase-specific interaction of the Fanconi anemia protein, FANCD2, with BRCA1 and RAD51', *Blood*, 100: 2414-20.

Taylor, E. R., and C. H. McGowan. 2008. 'Cleavage mechanism of human Mus81-Eme1 acting on Holliday-junction structures', *Proc Natl Acad Sci U S A*, 105: 3757-62.

Thoma, B. S., M. Wakasugi, J. Christensen, M. C. Reddy, and K. M. Vasquez. 2005. 'Human XPC-hHR23B interacts with XPA-RPA in the recognition of triplex-directed psoralen DNA interstrand crosslinks', *Nucleic Acids Res*, 33: 2993-3001.

Timmers, C., T. Taniguchi, J. Hejna, C. Reifsteck, L. Lucas, D. Bruun, M. Thayer, B. Cox, S. Olson, A. D. D'Andrea, R. Moses, and M. Grompe. 2001. 'Positional cloning of a novel Fanconi anemia gene, FANCD2', *Mol Cell*, 7: 241-8.

Tishkoff, D. X., A. L. Boerger, P. Bertrand, N. Filosi, G. M. Gaida, M. F. Kane, and R. D. Kolodner. 1997. 'Identification and characterization of Saccharomyces cerevisiae EXO1, a gene encoding an exonuclease that interacts with MSH2', *Proc Natl Acad Sci U S A*, 94: 7487-92.

Tran, H. T., D. A. Gordenin, and M. A. Resnick. 1999. 'The 3'-->5' exonucleases of DNA polymerases delta and epsilon and the 5'-->3' exonuclease Exo1 have major roles in postreplication mutation avoidance in Saccharomyces cerevisiae', *Mol Cell Biol*, 19: 2000-7.

Trujillo, J. P., L. B. Mina, R. Pujol, M. Bogliolo, J. Andrieux, M. Holder, B. Schuster, D. Schindler, and J. Surralles. 2012. 'On the role of FAN1 in Fanconi anemia', *Blood*, 120: 86-9.

Tsao, D., P. Kalogerinis, I. Tabrizi, M. Dingfelder, R. D. Stewart, and A. G. Georgakilas. 2007. 'Induction and processing of oxidative clustered DNA lesions in 56Fe-ion-irradiated human monocytes', *Radiat Res*, 168: 87-97.

Tsodikov, O. V., J. H. Enzlin, O. D. Scharer, and T. Ellenberger. 2005. 'Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1', *Proc Natl Acad Sci U S A*, 102: 11236-41.

Unno, J., A. Itaya, M. Taoka, K. Sato, J. Tomida, W. Sakai, K. Sugasawa, M. Ishiai, T. Ikura, T. Isobe, H. Kurumizaka, and M. Takata. 2014. 'FANCD2 binds CtIP and regulates DNA-end resection during DNA interstrand crosslink repair', *Cell Rep*, 7: 1039-47.

van der Woude, M. W., and A. J. Baumler. 2004. 'Phase and antigenic variation in bacteria', *Clin Microbiol Rev*, 17: 581-611, table of contents.

Vasquez, K. M. 2010. 'Targeting and processing of site-specific DNA interstrand crosslinks', *Environ Mol Mutagen*, 51: 527-39.

Virts, E. L., A. Jankowska, C. Mackay, M. F. Glaas, C. Wiek, S. L. Kelich, N. Lottmann, F. M. Kennedy, C. Marchal, E. Lehnert, R. E. Scharf, C. Dufour, M. Lanciotti, P. Farruggia, A. Santoro, S. Savasan, K. Scheckenbach, J. Schipper, M. Wagenmann, T. Lewis, M. Leffak, J. L. Farlow, T. M. Foroud, E. Honisch, D. Niederacher, S. C. Chakraborty, G. H. Vance, D. Pruss, K. M. Timms, J. S. Lanchbury, A. F. Alpi, and H. Hanenberg. 2015. 'AluY-mediated germline deletion, duplication and somatic stem cell reversion in UBE2T defines a new subtype of Fanconi anemia', *Hum Mol Genet*, 24: 5093-108.

Wagner, R., Jr., and M. Meselson. 1976. 'Repair tracts in mismatched DNA heteroduplexes', *Proc Natl Acad Sci U S A*, 73: 4135-9.

Walden, H., and A. J. Deans. 2014. 'The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder', *Annu Rev Biophys*, 43: 257-78.

Walker, J. R., R. A. Corpina, and J. Goldberg. 2001. 'Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair', *Nature*, 412: 607-14.

Wallace, S. S. 2014. 'Base excision repair: a critical player in many games', *DNA Repair (Amst)*, 19: 14-26.

Wang, A. T., B. Sengerova, E. Cattell, T. Inagawa, J. M. Hartley, K. Kiakos, N. A. Burgess-Brown, L. P. Swift, J. H. Enzlin, C. J. Schofield, O. Gileadi, J. A. Hartley, and P. J. McHugh. 2011. 'Human SNM1A and XPF-ERCC1 collaborate to initiate DNA interstrand cross-link repair', *Genes Dev*, 25: 1859-70.

Wang, J., and T. Lindahl. 2016. 'Maintenance of Genome Stability', *Genomics Proteomics Bioinformatics*, 14: 119-21.

Wang, W. 2007. 'Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins', *Nat Rev Genet*, 8: 735-48.

Wang, X., C. A. Peterson, H. Zheng, R. S. Nairn, R. J. Legerski, and L. Li. 2001. 'Involvement of nucleotide excision repair in a recombination-independent and error-prone pathway of DNA interstrand cross-link repair', *Mol Cell Biol*, 21: 713-20.

Ward, J. F. 1981. 'Some biochemical consequences of the spatial distribution of ionizing radiation-produced free radicals', *Radiat Res*, 86: 185-95.

———. 1994. 'The complexity of DNA damage: relevance to biological consequences', *Int J Radiat Biol*, 66: 427-32.

Ward, J. F., W. F. Blakely, and E. I. Joner. 1985. 'Mammalian cells are not killed by DNA single-strand breaks caused by hydroxyl radicals from hydrogen peroxide', *Radiat Res*, 103: 383-92.

Waters, C. A., N. T. Strande, D. W. Wyatt, J. M. Pryor, and D. A. Ramsden. 2014. 'Nonhomologous end joining: a good solution for bad ends', *DNA Repair (Amst)*, 17: 39-51.

Waters, L. S., B. K. Minesinger, M. E. Wiltrout, S. D'Souza, R. V. Woodruff, and G. C. Walker. 2009. 'Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance', *Microbiol Mol Biol Rev*, 73: 134-54.

Wei, P. C., A. N. Chang, J. Kao, Z. Du, R. M. Meyers, F. W. Alt, and B. Schwer. 2016. 'Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural Stem/Progenitor Cells', *Cell*, 164: 644-55.

Wei, P. C., C. S. Lee, Z. Du, B. Schwer, Y. Zhang, J. Kao, J. Zurita, and F. W. Alt. 2018. 'Three classes of recurrent DNA break clusters in brain progenitors identified by 3D proximity-based break joining assay', *Proc Natl Acad Sci U S A*, 115: 1919-24.

Weinberg, Benjamin H., N. T. Hang Pham, Leidy D. Caraballo, Thomas Lozanoski, Adrien Engel, Swapnil Bhatia, and Wilson W. Wong. 2017. 'Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells', *Nature biotechnology*, 35: 453-62.

West, A. E., and M. E. Greenberg. 2011. 'Neuronal activity-regulated gene transcription in synapse development and cognitive function', *Cold Spring Harb Perspect Biol*, 3.

West, S. C. 2009. 'The search for a human Holliday junction resolvase', *Biochem Soc Trans*, 37: 519-26.

Wildenberg, J., and M. Meselson. 1975. 'Mismatch repair in heteroduplex DNA', *Proc Natl Acad Sci U S A*, 72: 2202-6.

Williams, H. L., M. E. Gottesman, and J. Gautier. 2012. 'Replication-independent repair of DNA interstrand crosslinks', *Mol Cell*, 47: 140-7.

———. 2013. 'The differences between ICL repair during and outside of S phase', *Trends Biochem Sci*, 38: 386-93.

Wilson, D. M., 3rd, and M. M. Seidman. 2010. 'A novel link to base excision repair?', *Trends Biochem Sci*, 35: 247-52.

Winkler, G. S., S. J. Araujo, U. Fiedler, W. Vermeulen, F. Coin, J. M. Egly, J. H. Hoeijmakers, R. D. Wood, H. T. Timmers, and G. Weeda. 2000. 'TFIIH with inactive XPD helicase functions in transcription initiation but is defective in DNA repair', *J Biol Chem*, 275: 4258-66.

Woese, C. R., and G. E. Fox. 1977. 'Phylogenetic structure of the prokaryotic domain: the primary kingdoms', *Proc Natl Acad Sci U S A*, 74: 5088-90.

Wood, R. D. 2010. 'Mammalian nucleotide excision repair proteins and interstrand crosslink repair', *Environ Mol Mutagen*, 51: 520-6.

Wyatt, H. D., S. Sarbajna, J. Matos, and S. C. West. 2013. 'Coordinated actions of SLX1-SLX4 and MUS81-EME1 for Holliday junction resolution in human cells', *Mol Cell*, 52: 234-47.

Xie, M. Z., M. I. Shoulkamy, A. M. Salem, S. Oba, M. Goda, T. Nakano, and H. Ide. 2016. 'Aldehydes with high and low toxicities inactivate cells by damaging distinct cellular targets', *Mutat Res*, 786: 41-51.

Yamada, M., E. O'Regan, R. Brown, and P. Karran. 1997. 'Selective recognition of a cisplatin-DNA adduct by human mismatch repair proteins', *Nucleic Acids Res*, 25: 491-6.

Yamanaka, K., I. G. Minko, K. Takata, A. Kolbanovskiy, I. D. Kozekov, R. D. Wood, C. J. Rizzo, and R. S. Lloyd. 2010. 'Novel enzymatic function of DNA polymerase nu in translesion DNA synthesis past major groove DNA-peptide and DNA-DNA cross-links', *Chem Res Toxicol*, 23: 689-95.

Yang, L., A. A. Nielsen, J. Fernandez-Rodriguez, C. J. McClune, M. T. Laub, T. K. Lu, and C. A. Voigt. 2014. 'Permanent genetic memory with >1-byte capacity', *Nat Methods*, 11: 1261-6.

Yang, W., and R. Woodgate. 2007. 'What a difference a decade makes: insights into translesion DNA synthesis', *Proc Natl Acad Sci U S A*, 104: 15591-8.

Yano, K., K. Morotomi-Yano, K. J. Lee, and D. J. Chen. 2011. 'Functional significance of the interaction with Ku in DNA double-strand break recognition of XLF', *FEBS Lett*, 585: 841-6.

Yeo, J. E., S. Wickramaratne, S. Khatwani, Y. C. Wang, J. Vervacke, M. D. Distefano, and N. Y. Tretyakova. 2014. 'Synthesis of site-specific DNA-protein conjugates and their effects on DNA replication', *ACS Chem Biol*, 9: 1860-8.

Yoshikiyo, K., K. Kratz, K. Hirota, K. Nishihara, M. Takata, H. Kurumizaka, S. Horimoto, S. Takeda, and J. Jiricny. 2010. 'KIAA1018/FAN1 nuclease protects cells against genomic instability induced by interstrand cross-linking agents', *Proc Natl Acad Sci U S A*, 107: 21553-7.

Yuan, F., L. Qian, X. Zhao, J. Y. Liu, L. Song, G. D'Urso, C. Jain, and Y. Zhang. 2012. 'Fanconi anemia complementation group A (FANCA) protein has intrinsic affinity for nucleic acids with preference for single-stranded forms', *J Biol Chem*, 287: 4800-7.

Zarrin, A. A., C. Del Vecchio, E. Tseng, M. Gleason, P. Zarin, M. Tian, and F. W. Alt. 2007. 'Antibody class switching mediated by yeast endonuclease-generated DNA breaks', *Science*, 315: 377-81.

Zha, S., C. Boboila, and F. W. Alt. 2009. 'Mre11: roles in DNA repair beyond homologous recombination', *Nat Struct Mol Biol*, 16: 798-800.

Zhang, J., J. M. Dewar, M. Budzowska, A. Motnenko, M. A. Cohn, and J. C. Walter. 2015. 'DNA interstrand cross-link repair requires replication-fork convergence', *Nat Struct Mol Biol*, 22: 242-7.

Zhang, J., and J. C. Walter. 2014. 'Mechanism and regulation of incisions during DNA interstrand cross-link repair', *DNA Repair (Amst)*, 19: 135-42.

Zhang, N., X. Lu, X. Zhang, C. A. Peterson, and R. J. Legerski. 2002. 'hMutSbeta is required for the recognition and uncoupling of psoralen interstrand cross-links in vitro', *Mol Cell Biol*, 22: 2388-97.

Zhang, Y., R. P. McCord, Y. J. Ho, B. R. Lajoie, D. G. Hildebrand, A. C. Simon, M. S. Becker, F. W. Alt, and J. Dekker. 2012. 'Spatial organization of the mouse genome and its role in recurrent chromosomal translocations', *Cell*, 148: 908-21.

Zhao, J., A. Jain, R. R. Iyer, P. L. Modrich, and K. M. Vasquez. 2009. 'Mismatch repair and nucleotide excision repair proteins cooperate in the recognition of DNA interstrand crosslinks', *Nucleic Acids Res*, 37: 4420-9.

Zheng, H., X. Wang, A. J. Warren, R. J. Legerski, R. S. Nairn, J. W. Hamilton, and L. Li. 2003. 'Nucleotide excision repair- and polymerase eta-mediated error-prone removal of mitomycin C interstrand cross-links', *Mol Cell Biol*, 23: 754-61.

Zhou, W., E. A. Otto, A. Cluckey, R. Airik, T. W. Hurd, M. Chaki, K. Diaz, F. P. Lach, G. R. Bennett, H. Y. Gee, A. K. Ghosh, S. Natarajan, S. Thongthip, U. Veturi, S. J. Allen, S.

Janssen, G. Ramaswami, J. Dixon, F. Burkhalter, M. Spoendlin, H. Moch, M. J. Mihatsch, J. Verine, R. Reade, H. Soliman, M. Godin, D. Kiss, G. Monga, G. Mazzucco, K. Amann, F. Artunc, R. C. Newland, T. Wiech, S. Zschiedrich, T. B. Huber, A. Friedl, G. G. Slaats, J. A. Joles, R. Goldschmeding, J. Washburn, R. H. Giles, S. Levy, A. Smogorzewska, and F. Hildebrandt. 2012. 'FAN1 mutations cause karyomegalic interstitial nephritis, linking chronic kidney failure to defective DNA damage repair', *Nat Genet*, 44: 910-5.

Zhu, Q., and A. A. Wani. 2017. 'Nucleotide Excision Repair: Finely Tuned Molecular Orchestra of Early Pre-incision Events', *Photochem Photobiol*, 93: 166-77.

Ziani, S., Z. Nagy, S. Alekseev, E. Soutoglou, J. M. Egly, and F. Coin. 2014. 'Sequential and ordered assembly of a large DNA repair complex on undamaged chromatin', *J Cell Biol*, 206: 589-98.

Zou, L., and S. J. Elledge. 2003. 'Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes', *Science*, 300: 1542-8.