DISSERTATION

DNA REPAIR PROTEINS METNASE AND PARP1 REGULATE DNA INTEGRATION

Submitted by

Jingyi Nie

Graduate Degree Program in Cell and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2015

Doctoral Committee:

Advisor: Jac A. Nickoloff

Susan M. Bailey
Howard L. Liber
Chaoping Chen

ABSTRACT

DNA REPAIR PROTEINS METNASE AND PARP1 REGULATE DNA INTEGRATION

DNA integration occurs naturally in various formats and plays important roles in evolution. DNA integration also affects human and animal health. Various genome-editing tools have been developed based on site-specific DNA integration. In mammalian cells, DNA integration is largely random. The mechanism of random DNA integration is not fully understood but has close association with repair of double-strand DNA damage. There are two major pathways to repair double-strand breaks (DSBs), homologous recombination (HR) and non-homologous end joining (NHEJ). In mammalian cells, NHEJ occurs more frequently than HR, possibly explains why random integration is more efficient than homology-directed integration or gene targeting. Proteins function in DSB repair pathways often engage in DNA integration.

Metnase is a fusion protein that only expresses in higher primates, including humans. Metnase contains a SET methyltransferase domain and a transposase domain. Metnase promotes efficiency and accuracy of NHEJ and promotes DNA integration. The SET domain dimethylates histone H3K36 at DSB sites, and the transposase domain binds to the human *Mariner* transposon *Hsmar1* terminal inverted repeat (TIR) sequence specifically. Both domains have been shown to be important for the role of Metnase in NHEJ. In this study, we tested the role of Metnase in promoting plasmid integration. We hypothesized that Metnase promotes plasmid integration through its functions in the NHEJ pathway. Metnase enhances the efficiency and accuracy of NHEJ, we predict that overexpression of Metnase can prevent integrating plasmid and genomic

DNA at integration sites from large deletions. Besides, if the specific TIR binding of Metnase can direct more DNA integration into the TIR sequence in the human genome, overexpression of Metnase would increase the ratio of DNA integration found at or nearby TIR region. To test this hypothesis, we mapped plasmid integration in the human cell line HEK293T at low and high levels of Metnase expression. Our results demonstrated that Metnase promotes plasmid DNA integration independently of TIR sequence in the human genome. Overexpression of Metnase suppressed microhomology-mediated DNA integration, supporting our hypothesis that Metnase promotes DNA integration through classical NHEJ (cNHEJ).

In contrast to cNHEJ, alternative NHEJ (aNHEJ) utilizes a different set of core proteins to rejoin broken ends. Compared to cNHEJ, aNHEJ is more error-prone and considered as the major generator of chromosomal translocations. Initiation of aNHEJ requires end resection. PARP1 plays an important role in initiating aNHEJ by recruiting end resection factors to DSBs. PARP1 has also been shown to promote DSB-induced chromosomal translocations. Based on the structural similarity between chromosomal translocations and DNA integration, we hypothesized that PARP1 may promote a sub-set of DNA integration, possibly through aNHEJ. We tested the effects of two PARP1 inhibitors PJ34 and Olaparib on DNA integration. Surprisingly, the two inhibitors showed opposite effects on DNA integration. PJ34 suppressed DNA integration, while Olaparib promoted DNA integration. We then confirmed PARP1 promoted DNA integration in a stable PARP1 knockdown cell line. Future studies are needed to understand the engagement of PARP1 in DNA integration and interpret the result where Olaparib promotes DNA integration.

ACKNOWLEDGEMENTS

I would like to express my special appreciation and thanks to my advisor Dr. Jac Nickoloff, you have been a wonderful mentor for me. I would like to thank you for encouraging my research and for allowing me to grow into a research scientist. Your advice on both research as well as on my career have been priceless. I would also like to thank my committee members, Dr. Susan Bailey, Dr. Chaoping Chen, and Dr. Howard Liber for serving as my committee members. Thank you all for your brilliant comments and suggestions, and for being supportive through the years.

I want to thank present and past members of the Radiation Cancer Biology & Oncology section of the Department of Environmental & Radiological Health Sciences at Colorado State University: Dr. Amanda Ashely for giving me an introduction into the Nickoloff lab and for sharing their knowledge. Thanks to Dr. Lucas Argueso and Dr. Chris Allen for the good advice on research directions. Thanks to all fellow graduate students for sharing a good time. Also thanks to lab managers Chrissy Battaglia and Lynn Taylor for their hard work in maintain a tidy and organized research environment. Special thanks to Dr. Xianan Liu, Chris Nelson and Dr. Déborah Afonso Cornélio for helping to finish various projects. All of you have been there to support me for my Ph.D. research.

A special thanks to my family. Words cannot express how grateful I am to my parents for all of the sacrifices that you have made for my behalf. I would also like to thank all of my friends who supported me in writing, and incented me to strive towards my goal. I would like express appreciation to my beloved boyfriend Yi Ern Cheah who was always my support in the moments of difficulty and frustration.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

DNA INTEGRATION:

ROLES IN HUMAN HEALTH AND APPLICATIONS IN RESEARCH

DNA integration is the process of a foreign fragment of DNA inserting into DNA within a cell, typically a chromosome. Integrated foreign DNA becomes part of the genome and is replicated along with the host DNA and passed down to daughter cells. If the genetic information in the foreign DNA is expressed, the host cell that accepts the foreign DNA may gain new properties (reviewed in Johnsborg et al., 2007). In addition, insertion of foreign DNA can cause modifications to the recipient genome in various ways. In some cases, the integrated foreign DNA interrupts the coding sequences of the host genome, resulting in silencing of certain genes (Covarrubias et al., 1986; Covarrubias et al., 1987). In other cases, integration of foreign DNA causes rearrangement of the recipient chromosome. These rearrangements include DNA deletions, insertion of extra bases, duplication of integration sites, and chromosomal translocations (Covarrubias et al., 1986; Covarrubias et al., 1987; Hamada et al., 1993; Kato et al., 1986; Mahon et al., 1988; Mark et al., 1992; Robins et al., 1981; Wilkie and Palmiter, 1987). Work from Doerfler group has revealed de novo methylation of integrated foreign DNA and the recipient genome at both the integration site and at sites remote from the integration site; and this may underlie altered transcription patterns of the recipient genome due to integration of foreign DNA (reviewed in Doerfler, 2011, 2012). As a result, DNA integration can disrupt genome stability of the recipient cell. Interruption of genome stability by integrated foreign DNA increases mutation rates, and can play an important role in evolution. In natural settings, DNA

integration can be observed in several different situations (Figure 1). Transposable elements, including transposons and retrotransposons compose about 45% of the human genome (Cordaux and Batzer, 2009). Retroviruses, such as human immunodeficiency virus (HIV), reverse transcribe their RNA genome into DNA, which then integrates into the genome of the infected host genome (Craigie and Bushman, 2012). Plasmids are small circular double-strand DNAs that can transfer from host to host. Plasmids are most common in bacteria, and have been extensively engineered to express genes of interest in eukaryotes. Certain plasmids are able to self-replicate independent of the host chromosomal DNA, if they have a functional replication origin, or they may integrate into the host genome and replicate along with the host DNA. Transposable elements, retroviral DNAs, and plasmid DNAs have been extensively investigated and engineered for many purposes, including construction of transgenic cell lines and animals, and to deliver therapeutic DNA in gene therapy, as described in the following sections.

Integration of transposable elements

The first transposable element, or transposon, was identified by Barbara McClintock in maize (McClintock, 1953), which earned her Nobel prize in 1983. Transposons are short DNA fragments that can move from one site of chromosome to another site. Transposable elements are classified into two classes based on the mechanism of transposition. Class I transposable elements move in a "copy and paste" fashion and are referred to as retrotransposons (Figure 1A and 1B). DNA of retrotransposons is first transcribed into RNA, the RNA then anneals to the nicked target DNA and primes reverse transcription and synthesis of the first strand of complementary DNA (cDNA), followed by synthesis of the second strand of cDNA and complete insertion of the newly synthesized double-strand cDNA into the new locus in the genome (reviewed in Boeke and Chapman, 1991; Eickbush and Jamburuthugoda, 2008). Class II

transposable elements move in a "cut and paste" fashion (Figure 1C). They encode genes for transposases, which make staggered cuts at the ends of transposons and the target site. Each end of the transposon then covalently links to a single-strand overhang of the target site by DNA ligase. Gaps left on the overhang sequences are then filled by DNA polymerases, resulting in short repetitive sequences flanking the transposon at the new locus (Hallet and Sherratt, 1997). We will revisit these short repetitive sequences in future discussion, and refer to them as terminal inverted repeats (TIRs). Class II transposons do not require RNA intermediates, therefore are referred as DNA transposons.

Class I transposable elements, or retrotransposons are important in shaping the human genome (Lander et al., 2001), and some are associated with human diseases (reviewed in Jung et al., 2013). There are two categories of retrotransposons, autonomous and non-autonomous. Autonomous retrotransposons encode reverse transcriptases in their sequences, while non-autonomous retrotransposons depend on cellular source of reverse transcriptase to move. Autonomous retrotransposons are further classified as long terminal repeats (LTRs)-containing (Figure 1A) and LTRs-lacking retrotransposons (Figure 1B). LTRs generally serve as the primer for transcription into retrotransposon RNA and reverse transcription into cDNA (reviewed in Havecker et al., 2004). The most abundant LTR-containing members are human endogenous retroviruses (HERVs). HERVs make up about 8% of the human genome but all HERVs have been inactivated due to deletions and other mutations (Belshaw et al., 2005; Belshaw et al., 2004; Jha et al., 2009). The most common LTR-lacking retrotransposons in the human genome are long interspersed elements (LINEs) (Lander et al., 2001). The LINE-1 (L1) element and its associated short interspersed elements (SINE) partner Alu are the only known active retrotransposons in humans, comprising 22% of the human genome (Beck et al., 2011b). SINEs

are a type of non-autonomous retrotransposon that lack reverse transcriptase encoding ability. SINEs are associated with disease as well, and contribute to genome diversity (Batzer and Deininger, 2002; Goodier and Kazazian, 2008; Nystrom-Lahti et al., 1995; Ostertag et al., 2003). Endogenous retrotransposons integrated into the human genome during evolution still have potential to induce chromosomal rearrangement though recombination between different copies. Activities of endogenous retrotransposons have been associated with various genetic diseases (reviewed in Belancio et al., 2008; Belancio et al., 2010; Carreira et al., 2014; Kassiotis, 2014; Katoh and Kurata, 2013; Thomas et al., 2012). In the human genome, SINEs are moved by L1-encoded protein. L1 elements encode an endonuclease, which is critical for targeting new loci by L1. The L1 endonuclease cleaves at 5'-TTAAAA-3' sites on chromosomes and produces 3' overhangs (Feng et al., 1996). This small fragment of single-strand DNA (ssDNA) serves as primer to synthesize cDNA using L1 transcript as a template (Christensen and Eickbush, 2005; Cost et al., 2002; Luan et al., 1993). The complete mechanisms of cDNA synthesis and incorporation of the newly synthesized double-strand cDNA into chromosomes remain to be elucidated. Both L1 endonuclease-dependent and -independent transpositions are known to occur (reviewed in Beck et al., 2011b).

Class II transposable elements, or DNA transposons move around genomes without an RNA intermediate (Figure 1C). DNA transposons make up about 2-3% of human genome (Lander et al., 2001). *Mariner*-like elements, or *Mariner* transposons have attracted a lot attention because they have the ability for horizontal transmission among species (Lawrence and Hartl, 1992; Overbeek et al., 1991; Robertson, 1993; Robertson and MacLeod, 1993). *Mariner* was first identified in Drosophila (Jacobson et al., 1986) and later found in many insects and vertebrates species including human (Lampe et al., 2003; Lohe et al., 1995; Mandal and

Kazazian, 2008; Plasterk et al., 1999). The typical structure of *Mariner* transposons includes sequences that encode transposase, framed by TIRs of <100 bp (Lampe et al., 1996; Vos and Plasterk, 1994). *Mariner* transposase is responsible for full excision of a *Mariner* transposon from its original locus and integration at new locus (Lampe et al., 1996; Vos et al., 1996). The TIR contains a binding site of transposase (Vos and Plasterk, 1994). *Mariner* transposons belong to IS630-Tc1-*mariner* superfamily of transposable elements whose transposases share a common D(Asp)DE(Glu) motif (Doak et al., 1994; Plasterk et al., 1999). These transposases are related to the integrase of retroviruses such as HIV (Capy et al., 1996). In *Mariner* transposons, the actual structure is DDD and converting this DDD motif into DDE inactivates the transposase (Lohe et al., 1997). Proteins with a DDE motif are able to make a single-strand scission on double-strand DNA (dsDNA) molecules, exposing a reactive 3' hydroxyl (Craig, 1995), allowing scission of the entire transposable element from the host chromosome. Most transposons of the IS630-Tc1 family reintegrate into TA dinucleotide (Doak et al., 1994), although the precise integration mechanism remains unclear. To date, two types of *Mariner* transposons have been identified and characterized in the human genome, named *Hsmar1* and *Hsmar2* (Oosumi et al., 1995; Robertson and Martos, 1997; Robertson and Zumpano, 1997). Both are inactive pseudogenes that are no longer moving around the human genome. *Hsmar2* transposons are inactive due to accumulated mutations, copies of *Hsmar2* are present near homologous recombination hotspots in various regions where large deletions or repeated sequences have been linked to genetic disorders such as Prader-Willi and Angelman syndromes, Williams Syndrome, and Charcot-Marie-Tooth syndrome (Christian et al., 1995; Kiyosawa and Chance, 1996; Urban et al., 1996). Thus, even though these repeated elements are no longer able to move around the human genome, they continue to contribute to genome instability, because repeated sequences are

unstable and could cause chromosome rearrangement through replication slippage, sister-chromosome exchange-associated slippage, single-strand anneal and other mechanisms (Bzymek and Lovett, 2001). For example, instability of simple repeated sequences has been linked to colorectal tumorigenesis (Ionov et al., 1993; Perucho et al., 1994; Shibata et al., 1994; Thibodeau et al., 1993), ovarian carcinoma (Dodson et al., 1993), and Barrett's esophageal adenocarcinoma (reviewed in Shammas, 2011).

Integration of viral DNA

Viruses are small infectious agents that infect all types of living organisms, from archaea, bacteria, to plants and animals (Koonin et al., 2006). The structure of a virus is very simple. Most viruses only contain genetic material made of RNA or DNA, wrapped in a protein coat. Viruses only propagate utilizing the transcription and translation machinery of the host cell they infect. Infection of a virus in the host cell causes mild to severe effect on the host, including death. In addition to directly inducing diseases in humans and animals, several types of virus infection have been associated with carcinogenesis. Human papillomavirus is present in more than 95% of cervical cancer tissues and is well accepted as a causal agent of cervical cancer (zur Hausen, 2009). Heptitis B virus and Heptitis C virus contribute to up to 80% of hepatocellular carcinoma (Perz et al., 2006). Integration of Merkel cell polyomavirus sequence in the host genome is associated with Merkel cell carcinoma (Feng et al., 2008). Overall, it is estimated that viral infection accounts for about 20% of malignancies (reviewed in Damania, 2007). Some tumor-associated viruses are present in 100% of the tumor tissue, and have been determined to be the causative agent of that tumor, such as human papillomavirus in cervical cancer, Epstein-Barr virus in lymphoproliferative disease, hepatitis B virus in hepatocellular carcinoma and others (reviewed in Pagano et al., 2004). Virus-induced carcinogenesis share common general features.

In one way, expression of oncogenes encoded in viral sequence interrupts the regulation of cell division and leads to uncontrolled proliferation. In another way, integration of viral sequence and expression of some viral proteins disrupt the infected host genome and cause accumulation of mutations and genomic instability in the host genome, possibly result in activation of prooncogenes or disruption of tumor suppressor genes (reviewed in Chen et al., 2014).

Although not required for all viruses, many viruses are able to integrate their genetic materials into the host genome. Retroviruses are a family of RNA viruses where integration of their genetic material into the host genome is part of their life cycles. The genome of a retrovirus is a single-strand RNA. After retrovirus enters a host cell, retrovirus RNA is reversed transcribed into double-strand DNA, which is then integrated into the genome of the host by virus-coded integrase. The host will transcribe and translate the integrated genomic information of the retrovirus along with its own genome. Expression of genes within the integrated DNA produces viral components that are assembled into new virus particles, which are released from the host cell to initiate another round of infection. Therefore, integration of retroviral DNA into the host genome is a critical step for retroviruses to propagate.

Retroviruses infect a wide range of animals. In humans, retrovirus infection is associated with various diseases, including malignancies, immunodeficiencies, and neurological disorders (reveiwed in Rosenberg and Jolicoeur, 1997). For example, human immunodeficiency virus (HIV) is a member of the retrovirus family. The integration process of HIV proviral DNA was investigated extensively (reviewed in Krishnan and Engelman, 2012). Integrase plays a key role in integration of retroviral DNA into host genome (Figure 1D). Integrase is related to transposase of DNA transposons as it catalyzes attack of the phosphodiester bond of target genomic DNA by the 3'-OH groups of viral DNA (reviewed in Asante-Appiah and Skalka, 1997). Similar to LTR-

containing retrotransposons, HIV genome also has LTR sequences at each end. HIV integrase processes the LTR sequences and joins it to target genomic DNA. Integration of the proviral DNA into the infected genome is a critical step in the life cycle of HIV. Therefore, integrase is a proven target for acquired immunodeficiency syndrome (AIDS) drug development (Gu, 2014). Integrated retroviruses are domesticated to become part of the recipient genome and passed on to the daughter cells. It is estimated that about 4.7% of human genome is made up of human endogenous retroviruses (Lander et al., 2001).

Integration of plasmid DNA

Plasmids are small circular double-strand DNAs capable of replication independent of chromosomal DNA. Plasmids are most commonly found in bacteria, where they usually confer a new phenotype, such as drug resistance. Plasmids can easily transfer from one bacterium cell to another, and are therefore, major factor in horizontal gene transfer. Plasmids are very important tools in molecular biology. They are widely used to introduce foreign genes or modified genes into cells through integration of plasmid in the recipient genome. Although plasmids are generally capable of epichromosomal replication in bacteria, when transferred to other cells such as yeast or mammalian cells, their replication often requires integration into the host chromosome, allowing stable expression of the foreign or modified genes. For example, yeast integrative plasmids are designed in a way that the plasmids rely on integration into the yeast chromosome to replicate. Including replication origin that function in host eukaryotic cells permits "shuttle" plasmid replication independent of integration into the host genome. In mammalian cells, such autonomous plasmids are called "episomes". Integration of plasmid DNA into yeast or mammalian genome can be random or targeted (Figure 1E). Targeted plasmid integration to a specific locus on the host chromosome can be achieved by including homologous

sequence of the target chromosome in the plasmid, using techniques developed for both yeast and mammalian cells during the 1970s and 1980s (Boone et al., 1986; Scherer and Davis, 1979; Smithies et al., 1985; Thomas et al., 1986). In general, targeted plasmid integration in yeast is much more efficient than in mammalian cells (Ruff et al., 2014). The mechanism of plasmid integration in yeast and mammalian cells is not well understood. It is generally accepted that integration of plasmid into chromosomes occurs as part of DNA damage repair. This will be further discussed in later sections.

Other types of natural DNA integration

DNA integration can also occur through several less common mechanisms. When tumor cells undergo apoptosis induced by chemotherapy or radiation therapy, DNA of tumor cells is fragmented and recycled by surrounding cells through apoptotic bodies. Integration and amplification of tumor oncogenes in normal cells may transform the recipient cells into new tumor cells (Bergsmedh et al., 2001; Ehnfors et al., 2009; Holmgren, 2010). In an effort to sequence the breakpoint junctions of reciprocal chromosomal translocations, Willett-Brozick *et al*. found that mitochondrial DNA was inserted at some translocation junctions (Willett-Brozick et al., 2001). Analysis of the rice genome revealed that ancient mitochondrial DNA has been integrated into chromosomes, with interruptions by foreign DNA (Ueda et al., 2005). Conversely, it is found that extrachromosomal plasmid DNA captured pieces from chromosomes as the plasmid is repaired (Little and Chartrand, 2004). Thus, DNA integration plays an important role in evolution and continues to alter the genomic landscapes of modern organisms.

In summary, DNA integration into a recipient genome occurs frequently under natural situations via various mechanisms. Most sources of integrative DNA are foreign DNAs that previous not reside inside the recipient cell, while internal sources of integrative DNA are less

common. The integrated DNA could disrupt the intact structure of host genome, mutate or completely shut down essential genes as well as tumor suppressor genes and promote oncogene activation. Copy number and location of integrated expression vector potentially affect protein expression level (Fukushige and Sauer, 1992; Lacy et al., 1983), and cause non-targeted effects in the recipient cell, such as alternation of the DNA methylation and transcription patterns of the host genome (Lichtenberg et al., 1988; Liu et al., 2004; Muller et al., 2001; Remus et al., 1999). Therefore, integration of viral DNA and transposons in the human genome are closely associated with human health. On another hand, natural DNA integration events create mutations and provide new opportunities for organisms to adapt to environmental changes. Integration of engineered transposons, retroviruses, and plasmids with special sequences are important tools for manipulating genome for research and therapeutic purposes.

Application of DNA integration

Introduction of foreign genes with functional promoters into cells provide new features to the recipient cell. Foreign genes are typically delivered to the host cell via a plasmid or viral vectors. Integration of the vectors into the cell genome allows stable and long-term expression of foreign genes, and therefore has wide application in the fields of life science research, biotechnology and biomedical science. For example, extra copies of genes can be integrated into a cell to achieve overexpression of the corresponding gene products. This strategy is commonly used to study the function of the gene products. In gene therapy, therapeutic vectors that integrate (lentivirus, adenovirus, *etc.*) into the recipient genome can provide long-term therapeutic effects.

In many cases, integration site and copy number of integrated genes are not examined as long as the gene is stably expressed. However, in some cases targeted integration of foreign

DNA into a specific sequence is preferred. In mammalian cells, the efficiency of gene targeting varies. For example, wild type Adeno-associated virus integrates specifically into human chromosome 19 at the 19q13.3-qter (AAVS1) locus (Giraud et al., 1994). Proviral cDNA of HIV integrates preferentially into active transcription units and favors regions with histone modifications that facilitate transcription (Schroder et al., 2002; Wang et al., 2007). Plasmid DNAs tend to integrate randomly without site preferences (Bushman et al., 2005; Narezkina et al., 2004).

Homologous sequences are sometimes included in plasmid- or viral- based vectors to target specific loci via homologous recombination. However, the efficiency of homology-directed DNA integration is very low with just homologous sequences as the guide, and this has led to the development of several methods to enhance targeted DNA integration in mammalian cells (reviewed in Garrels et al., 2012; Wang, 2015). Various DNA modifying enzymes are employed to enhance precise integration of foreign DNA into target loci. Transposases encoded by transposon facilitates transposon-based vectors to integrate into specific sites. For example, site-specific recombinases Cre and FLP are commonly used system to create gene knockin and knockout to modify specific genome loci (O'Gorman et al., 1991; Sauer and Henderson, 1990). Zinc-finger nucleases (ZFN) and Transcription-activator like endonucleases (TALENs) are artificial restriction enzymes engineered to contain a DNA-binding domain fused with a DNA-cleavage domain. The DNA-binding domain can be engineered to recognize a specific sequence and guide the DNA-cleavage domain to cleave at a specific sequence targeted by the homologous sequence carried by the integrating plasmid to facilitate homologous recombination between the plasmid and the target site (Cermak et al., 2011; Li et al., 2011a; Li et al., 2011b; Mani et al., 2005; Miller et al., 2007; Miller et al., 2011; Morton et al., 2006; Porteus, 2008;

Wright et al., 2005). The recent development of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 system is an exciting advantage in precise editing of eukaryotic genomes (Cong and Zhang, 2015). The CRISPR-Cas9 system uses guide RNAs to guide site-specific cleavage of the target chromosome. Repair of the cleavage results in mutagenic deletions or insertion at the target site, or homology-directed replacement of the target site. These technologies are important for production of genetically modified cell lines or organisms for research purposes, and for production of cells with special properties in the biotech industry.

Various types of natural DNA integration (transposons, viruses, and others) have contributed to evolution and continue to impact stability and integrity of the human genome. As a commonly used genome engineering technology, DNA integration is wildly applied in biological research, and biotechnology and biopharmaceutics industry to product transgenic organisms that produce medicines, biofuels, biomaterials, and other products. For these reasons, it is important to understand the mechanisms of DNA integration, both for random and targeted DNA integrations, to allow more control in genome engineering.

DNA integration is closely related to DNA damage repair. Accumulated data indicates that various types of DNA damaging agents (UV, ionizing radiation (IR), restriction enzyme, Topoisomerase inhibitors, and others) promote the efficiency of DNA integration (reviewed in Wurtele et al., 2003). Therefore, it is reasonable to assume that the process of DNA integration is closely related to DNA damage and DNA repair. It has been shown that proteins involved in DNA repair affect the efficiency of foreign DNA integration. For example, transient knockdown of KU70 or X-ray repair cross-complementing protein 4 (XRCC4) suppressed random integration and increased targeted DNA integration in human cells (Bertolini et al., 2009). In

Neurospora, homology-directed foreign DNA integration requires the yeast Rad51 homolog MEI3, while non-homologous integration requires MUS-52 (a yeast KU80 homolog) and MUS-53 (a human Ligase IV homolog) (Ishibashi et al., 2006). The low gene targeting efficiency in mammalian cells is attributed to the relatively low efficiency of homologous recombination (HR) and relatively high efficiency of non-homologous end joining (NHEJ) pathways of double-strand break (DSB) repair. Therefore, it is necessary to discuss the mechanisms of DNA repair.

DNA DOUBLE-STRAND BREAK REPAIR

DNA is constantly subjected to damage caused by extracellular and intracellular DNA damaging agents. UV exposure from sunlight induces the formation of pyrimidine dimmers. Exposure to carcinogenic chemicals prevails everywhere in the modern industrial world. Exposure to large dose of IR in rare events, such as the Fukushima nuclear power plants explosion, can put millions of people at risk of DNA damage. Reactive oxygen species (ROS) produced during oxygen metabolism causes single-strand breaks (SSBs). When replication forks encounter damaged DNA, progression of is stalled and will not proceed until damage is repaired. Sustained damage results in collapse of replication forks, producing DSBs. It was estimated that about 1% of SSBs can be converted to DSBs per cell per cycle in human cells (Vilenchik and Knudson, 2003). Cells possess several types of DNA damage repair mechanisms to deal with different types of DNA damage. These mechanisms are highly conserved from bacteria to mammals.

Base excision repair (BER) corrects base damage arising through oxidation, alkylation, or deamination by replacing the nucleotide with damaged base and filling in the gap with the correct nucleotide (Figure 2A) (reviewed in Dianov and Hubscher, 2013; Robertson et al., 2009). Nucleotide excision repair (NER) repairs bulky DNA adducts and crosslinked DNA that result in

distortion of DNA double helix structure by removing the damaged strand and resynthesizing it using the complementary strand as a template (Figure 2B) (reviewed in Nouspikel, 2009). Mismatch repair (MMR) fixes non-Watson-Crick base pairing caused by polymerase error or base damage in the process of DNA synthesis by removing the error-containing newly synthesized strand and filling in the gap with the correct sequence (Figure 2C) (reviewed in Fukui, 2010; Kunz et al., 2009).

Base damage, base mismatches, and stabilized topoisomerase decatenation intermediates inhibit progression of replication forks. When not resolved in a timely manner, stalled replication forks collapse to produce DSBs. Direct attack on DNA deoxyribose backbone by IR is another source of DSBs. Among all types of DNA damage, DSBs are the most lethal. Failure to repair DSBs leads to chromosomal abnormality, genome instability, and even cell death. Pre-existing DNA damage is not a prerequisite for DNA integration, but DNA damage, such as DSBs induced by topoisomerase II inhibitors and endonucleases have been shown to increase the efficiency of foreign DNA integration (Fujimaki et al., 1996; Jin et al., 2005), suggesting that DNA integration and DSB repair share similar machinery or pathways. HR (Figure 2D) and NHEJ (Figure 2E) are the two major DSB repair mechanisms. In mammalian cells, HR is mainly active in S and G2 cell cycle phases where sister chromatids serve as template for repair, NHEJ is active throughout the cell cycle.

Homologous recombination

Homologous recombination repairs broken ends using homologous sequence as a template, the basic steps in HR are 1) resection of 5'- ended DNA; 2) formation of RAD51 nucleoprotein filament, homolog search, and strand invasion; 3) synthesis of new strand using homologous sequence as a template; 4) resolution of recombination intermediate and completion

of repair DNA by filling small gaps or nicks. Depending on the availability of homologous sequence and whether both ends of the break sites are present, repair through HR can occur in different ways: synthesis-dependent strand-annealing (SDSA), classical double-strand break repair (DSBR), single-strand annealing (SSA) and break-induced replication (BIR) (reviewed in Pardo et al., 2009). In mammalian cells, broken ends are recognized and bound immediately by the Mre11-Rad50-NBS1 (MRN) complex (reviewed in Paull and Lee, 2005). MRN initiates limited end resection and recruits C-terminal-binding protein interacting protein (CtIP) and exonuclease 1 (EXO1) that are responsible for extensive end resection (Nimonkar et al., 2011; Peterson et al., 2013). Resection exposes stretches of ssDNA, which are immediately bound and protected by replication protein A (RPA). Recombinase Rad51 displaces RPA to form a nucleoprotein filament on ssDNA that conducts the search for homologous template. Rad51 catalyzes invasion of the nucleoprotein filament into homologous template and forms heteroduplex structures (Baumann and West, 1998). The heteroduplex can expand and form a Holiday junction (HJ). New DNA sequence is synthesized using homologous sequence as template. HJs are resolved by coordinated action of helicases, endonucleases, polymerases and topoisomerases. Sister chromatids, homologous chromosomes, and repetitive sequences can be used as templates in homologous recombination. Repair by HR can restore the original genetic information with the possibility of gene conversion and crossover.

Foreign DNA integration via HR requires a stretch of homology between the integrating foreign DNA and the recipient genome. To a certain extent, efficiency of HR and homology-directed foreign DNA integration positively correlate with the length of homology (Fujitani and Kobayashi, 1995; Fujitani et al., 1995; Hasty et al., 1991; Rubnitz and Subramani, 1984; Shen and Huang, 1986). In the commonly used model organism yeast *Saccharomyces cerevisiae*, HR

is the dominant pathway for DSB repair. Therefore, homology-directed DNA integration in yeast is fairly efficient. However, in mammalian cells, efficiency of homology-directed DNA integration is low. In gram-negative *Acinetobacter*, it has been reported that efficiency of DNA integration can be significantly increased when one end of the integrating DNA contains homologous sequence with the recipient genome, compared to integration in the absence of any homology. The homologous sequence serves as an anchor to facilitate the illegitimate integration on the other end. This type of DNA integration is termed homology-facilitated illegitimate recombination (HFIR) (de Vries and Wackernagel, 2002). In *Acinetobacter*, the efficiency of DNA integration via HFIR is 10,000 times less frequent than fully homologous recombination, but 100,000-fold more frequent than integration with no homology (de Vries and Wackernagel, 2002; Hulter and Wackernagel, 2008). Foreign DNA acquisition through HFIR is suppressed by various nucleases including RecBCD, SbcCD, and RecJ, particularly when the homology is short, suggesting that longer homology is more efficient in stabilizing the integration intermediate and escaping degradation by nuclease, thus promoting DNA integration (Harms and Wackernagel, 2008). After the integrating DNA is anchored to the recipient genome via the homologous end, the other end integrates in a homology-independent manner. This last step in the process should be same as fully homology-independent integration, which shares certain features with DSB repair by NHEJ.

Non-homologous end joining

Non-homologous end joining pathway repairs DSBs independent of homologous sequence (reviewed in Waters et al., 2014). DSBs are recognized by MRN complex, which recruits DNA damage signaling protein Ataxia Telangiectasia Mutated (ATM). ATM phosphorylates histone variant H2AX at serine 139 at a restricted region of the damaged DNA.

The phosphorylated H2AX, or gamma (γ)-H2AX recruits more ATM, which then phosphorylates more H2AX flanking both sides of the DSB and newly phosphorylated H2AX recruits more ATM. This positive feedback loop extends γ-H2AX signal to megabases on both sides of the DSB (reviewed in Bekker-Jensen and Mailand, 2010; Kinner et al., 2008; Kuo and Yang, 2008). KU70 and KU80 heterodimers bind to broken ends, protecting the ends and serving as the docking site of DNA-dependent protein kinase (DNA-PK) complex (reviewed in Grundy et al., 2014). DNA-PK binds to each end and tethers the ends to form a bridge domain. Formation of bridge domain triggers phosphorylation and activation of DNA-PK. Activated DNA-PK phosphorylates downstream repair proteins (reviewed in Davis et al., 2014; Pawelczak et al., 2011; Weterings and Chen, 2007). The scaffold protein XRCC4 facilitates recruitment of Ligase IV to rejoin the two ends (Chen et al., 2000; Critchlow et al., 1997). Often, ends are processed by exonucleases Artemis and meiotic recombination 11 (MRE11) to prepare ends for efficient ligation, particularly for chemically diverse ends at DSBs created by IR (reviewed in Kurosawa and Adachi, 2010; Sung et al., 2014). NHEJ repairs DSBs independent of homologous sequences. However, ends may be resected before rejoining and additional bases are sometimes added to ends prior to rejoining. KU, DNA-PKcs, Ligase IV dependent rejoining of blunt ends is referred as the classical non-homologous end joining (cNHEJ), to differentiate from alternative non-homologous end joining (aNHEJ) (Bennardo et al., 2008).

Alternative NHEJ is less efficient and more error prone than cNHEJ. Alternative NHEJ is inhibited by Ligase IV and has been considered as the main generator for chromosomal translocations (Byrne et al., 2014; Simsek et al., 2011; Wray et al., 2013). However, some argue that different from murine cell, in human genome, the main generator of translocations is cNHEJ (Ghezraoui et al., 2014). In aNHEJ, a subset of HR proteins, including MRE11 and CtIP are used

to initiate end resection (Badie et al., 2015; Della-Maria et al., 2011; Lee-Theilen et al., 2011; Zhang and Jasin, 2011). The purpose of end resection is to exposure pre-existing microhomology (5–25 nucleotides) between two ends. Ligase III anneals ends at microhomology. As a backup, Ligase I can anneal ends independent of pre-existing microhomology (Simsek et al., 2011). Because cNHEJ does not utilize microhomology to anneal ends, microhomology is often considered an indication of aNHEJ (Pannunzio et al., 2014).

In contrast to targeted integration, either homology-directed or site-specific integration such as Cre-Lox and FLP, most DNA integration is random as it is not sequence or locus specific. Sometimes, random integration is referred to as illegitimate DNA integration. Analysis of random DNA integration sites in various species reveal several common features typically found at foreign DNA-recipient genome junctions, including deletions of various lengths on both the foreign DNA and the recipient genome, extra nucleotides inserted at the junction, and rearrangement of the recipient genome. These features resemble the outcomes of NHEJ (reviewed in Wurtele et al., 2003). Deficiency in key cNHEJ proteins, KU, XRCC4, Ligase IV have been shown to suppress foreign DNA integration (Bertolini et al., 2009; Ishibashi et al., 2006). Transient knockdown of key cNHEJ proteins KU70 and XRCC4 suppressed random integration and promoted homology-directed gene targeting (Bertolini et al., 2009; Iiizumi et al., 2008; Tanaka et al., 2010). In mammalian cells, cNHEJ dominates DSB repair, as it is active across all phases of cell cycle. This explains why random integration is more efficient than targeted integration in mammalian cells. Classical NHEJ does not require microhomology to anneal broken ends, but microhomology is often found at integration junctions (de Vries and Wackernagel, 2002; Dubose et al., 2013; Hamada et al., 1993; Merrihew et al., 1996; Rohan et al., 1990), raising the possibility that aNHEJ has a role in a significant fraction of integration

events. However, little work has been done to investigate DNA integration via aNHEJ. It is possible that the large amount of free ends upon entering of foreign DNA into cytoplasm may exceed the capacity of cNHEJ, and if true, aNHEJ proteins may be responsible for a portion of DNA integration.

We hypothesized that both cNHEJ and aNHEJ pathways are utilized in the process of foreign DNA integration. To test this hypothesis, we studied the roles two DNA repair proteins Metnase and PARP1 in DNA integration. Metnase functions to promote Ligase IV-mediated cNHEJ and may suppress aNHEJ. PARP1 has important roles in BER, SSB repair (SSBR), HR, and aNHEJ. Metnase is known to promote both NHEJ and DNA integration (Hromas et al., 2008; Lee et al., 2005; Williamson et al., 2008a). PARP1 is important for DSB-induced chromosomal translocations (Wray et al., 2013). Considering structural similarities between chromosomal translocation and DNA integration, we hypothesized that PARP1 promotes DNA integration through aNHEJ pathway. The overall purpose of this thesis research is to generate a more comprehensive understanding of cellular machineries that contribute to DNA integration.

METNASE

Metnase was first identified as an expressed sequence tag (EST) and originally named SETMAR, for its SET (Su(var)3-9, Enhancer of Zeste, Trithorax) domain and *Mariner* transposase domain. The Metnase gene is 13 kb long and the protein has 671 amino acids. The Metnase gene is located on chromosome 3 at 3p26.1. This region is frequently abnormal in non-Hodgkin's lymphoma, acute and chronic lymphocytic leukemia, myeloma, myelodysplasia, hereditary prostate cancer and breast cancer (Lee et al., 2005). Metnase is only expressed in higher primates (Cordaux et al., 2006). In humans, Metnase is universally expressed in all tissues, with higher expression in liver, ovary and placenta tissues, and lower expression in skeletal

muscle (Lee et al., 2005). Intriguingly for unknown reasons, cells transformed by T-antigen do not express Metnase at a detectable level by western blotting and real-time polymerase chain reaction (Williamson et al., 2008b). Metnase derived from fusion of a *Mariner* transposable element, with its transposase/nuclease domain, to a SET protein during primate evolution about 40 million years ago. After fusion, the SET and nuclease domain each retain methyltransferase activity and nuclease activity respectively, and contribute to the overall function of Metnase in DNA dynamic processes.

Metnase promotes DNA integration and NHEJ

Metnase has multiple cellular functions; one of the first functions discovered was that it promotes plasmid and viral DNA integration (Lee et al., 2005; Williamson et al., 2008a). Metnase was also shown to promote NHEJ (Hromas et al., 2008; Lee et al., 2005). Because NHEJ is a critical mammalian DSB repair pathway, it is not surprising that Metnase overexpression increases cellular resistance to IR, and promotes removal of IR-induced γ-H2AX foci, suggesting that Metnase also promotes repair of chromosomal DSBs (Hromas et al., 2008; Lee et al., 2005). HR repairs DSBs as well, but Metnase appears to not function in HR (De Haro et al., 2010; Hromas et al., 2008). The enhancement of NHEJ by Metnase may be in part due to its interaction with key NHEJ proteins (LigIV, XRCC4) and its ability to recruit NHEJ proteins to DSBs (Fnu et al., 2011; Hromas et al., 2008). Efficient NHEJ is important for rejoining broken chromosomes and preventing chromosomal translocations. Although mouse cells do not express Metnase, overexpression of Metnase in mouse cells suppressed DSB-induced chromosomal translocations, indicating that Metnase can operate within the evolutionarily conserved cNHEJ pathway (Wray et al., 2010). Chromosomal translocations are common in acute myeloid leukemia (AML) (Gauwerky and Croce, 1993). A recent study of Metnase expression in healthy

individuals and AML patients revealed that full length Metnase is overexpressed in chromosomal translocation negative AML patients compared to chromosomal translocation positive AML patients (Jeyaratnam et al., 2014), which raised the possibility that Metnase has a role in suppressing chromosomal translocations.

Metnase promotes replication fork restart

Treatment with the DNA replication inhibitor hydroxyurea (HU), the topoisomerase inhibitor camptothecin (CPT) or UV-B exposure stops replication fork from progression. Metnase knockdown sensitizes cells to these three reagents at different levels, with highest sensitivity to HU. The DNA fiber assay indicated that Metnase promotes restart of stalled replication forks after HU treatment (De Haro et al., 2010). In response to DSB, Metnase is phosphorylated by checkpoint kinase 1 (Chk1) at Serine 495; and phosphorylation of Metnase suppresses its ability to promote restart of stalled replication forks and enhances its ability to enhance DSB repair (Hromas et al., 2012). The role of Metnase in DNA replication has another aspect. Decatenation of supercoiled DNA by topoisomerase I and II ensures progression of replication forks. Metnase interacts with Topoisomerase IIα (Topo IIα), enhances Topo IIα-mediated decatenation (Williamson et al., 2008b; Wray et al., 2009b), and confers resistance to Topo II inhibitors in breast cancer cells (Wray et al., 2009a). This ability is suppressed by automethylation of Metnase (Williamson et al., 2008b). The mechanisms underlying the multiple functions of Metnase are mostly unknown. However, understanding the activities of its SET and transposase domains provide clues to this mystery.

Functions of the SET and nuclease domain of Metnase

The SET domain of Metnase is encoded by the first two exons and belongs to a large family of evolutionarily conserved family of SET methyltransferases (Cordaux et al., 2006; Qian

and Zhou, 2006). In Metnase, a short deletion at the end of the second exon of the Metnase-specific-SET gene removed its stop codon, allowing extension of the second exon and fusion with the *Hsmar1* element that integrated downstream of this SET gene around 40-58 million years ago (Cordaux et al., 2006). The SET domain of Metnase retains methyltransferase activity. Four residues (K9, K27, K36, K79) of histone H3 and two residues (K20, K79) of histone H4 are targeted by Metnase, among these, H3K36 appears to be targeted most efficiently (Lee et al., 2005). Metnase dimethylates histone H3K36 at DSBs (Fnu et al., 2011). Activity of Metnase SET domain is important for Metnase-enhanced foreign DNA integration, NHEJ and suppression of chromosomal translocations (Fnu et al., 2011; Lee et al., 2005; Wray et al., 2010). The SET domain of Metnase is also important for interaction with human psoralen 4 (hPso4) protein, which recruits Metnase to DSBs and is critical for Metnase-mediated plasmid end joining and integration (Beck et al., 2008). Through interaction with hPso4, the SET domain contributes to Metnase binding of DNA in a sequence-independent manner. The methylation of histones may contribute to functions of Metnase by modifying chromosome configuration, providing better access for the Metnase transposase domain and other repair factors to broken ends.

The third exon of the Metnase gene encodes the transposase domain, which derived from the *Mariner* transposase (Cordaux et al., 2006). Two distinct features of this domain, the helix-turn-helix (HTH) motif and the D(Asp)DE(Glu)-like (DD(Asn)) motif, are responsible for its DNA binding and DNA cleavage activities respectively. The HTH motif of Metnase binds to dsDNA and is responsible for specific binding to the 19-mer core sequence of human transposon *Hsmar1* TIR (Kim et al., 2014; Liu et al., 2007; Roman et al., 2007). The family of *Mariner* transposases shares a conserved DDD domain, which is responsible for DNA cleavage and transposition. In Metnase, the third conserved aspartic acid (D) residue is mutated to asparagine

(N). As a result, the transposase domain no longer catalyzes transposition, but it does retain DNA nicking and cleavage activities (Liu et al., 2007). Despite the specific TIR binding conferred by the HTH motif, DNA cleavage is sequence independent (Roman et al., 2007). Biochemical characterization reveals that the transposase domain of Metnase cleaves certain DNA structures preferentially. For partial duplex DNA with a 5' overhang, Metnase tends to cut the ssDNA, while partial duplex DNA with a 3' overhang, is cleaved at both dsDNA and ssDNA regions (Beck et al., 2011a). The endonuclease activity of Metnase appears to be important for supporting NHEJ (Beck et al., 2011a; Rath et al., 2014). Efficient NHEJ reduces the time that broken DNA ends are subjected to degradation, limiting base deletions and mis-joining of ends. Indeed, overexpression of Metnase prevents large end deletions during NHEJ and foreign DNA integration (Hromas et al., 2008; Williamson et al., 2008a). It is possible that limited end trimming by Metnase prepares the ends for ligation by Ligase IV. Besides, methylation of histones (for example, H3K36me2) by the SET domain of Metnase also promotes NHEJ by recruitment of other NHEJ components, including Nibrin (NBS1) and KU70 (Fnu et al., 2011; Lee et al., 2005). As a result, Metnase promotes NHEJ to prevent extensive degradation of broken chromosomes, at the cost of small deletions due to its end endonuclease trimming activity.

In summary, Metnase acts as a general enhancer of multiple cellular functions related to DNA metabolism. Metnase enhances efficiency of NHEJ through its DNA binding and processing activity, DSB-induced histone methylation activity, and its interaction with KU70, XRCC4, and Ligase IV. Overexpression of Metnase suppresses chromosomal translocations in murine cells and in AML patients. Foreign DNA that enters a cell may be linear or circular, but even circular DNA is often damaged during transit into the nucleus and in either case the broken ends are detected by DNA repair system (Nickoloff et al., 1998). If foreign DNA integrates into

the cellular genome through NHEJ, it is not surprising that Metnase promotes both NHEJ and DNA integration, and in both situations, overexpression of Metnase protects ends from large deletions.

We hypothesized that Metnase promotes foreign DNA integration through its ability to interact with DNA and with components of the cNHEJ system. The DNA binding activity of Metnase displays at two levels: specific TIR binding through its transposase domain and non-specific DNA binding activity through its interaction with hPso4 (Beck et al., 2008; Roman et al., 2007). The actual DNA binding pattern of Metnase in a cell may differ from what has been observed *in vitro*. If Metnase has the same DNA binding activity *in vivo*, could Metnase preferentially stimulate DNA integration into TIR remnants in the recipient genome via its TIR binding ability? In the human genome, it is estimated that about 7000 *Mariner* remnants (with TIR sequence) remain following millennia of *Mariner* transposition (Liu et al., 2007). By estimation, the sequences of all TIR-containing *Mariner* remnants account about 0.02% of the entire human genome (3.3 billion base pairs). If Metnase preferentially stimulates DNA integration into TIR or TIR-adjacent region, a significant portion of TIRs should be found nearby integration junctions. Metnase enhances assembly of NHEJ machinery through its interaction with both DNA and NHEJ enzymes, including KU, XRCC4, and Ligase IV. As a result, broken ends are held together rather than drifting apart so they can be rejoined quickly. Hence, under Metnase overexpression, broken ends are rejoined with less deletion on average (Hromas et al., 2008). If Metnase promotes DNA integration through cNHEJ pathway, we predicted that overexpression of Metnase would prevent large deletions of both the integrating foreign DNA and the genome sequence at the integration sites. However, it is important to keep in mind that DNA integration differs from NHEJ in at least two aspects. In NHEJ, both ends are

chromosomal and hence occur in the context of histones (chromatin), while with DNA integration, the integrating DNA enters the cell in a pure state, although foreign DNA usually interacts with histones and form nucleoprotein complexes prior to integration (Hebbar and Archer, 2008; Reeves et al., 1985). In fact, proper chromatinization of transfected plasmid enhances expression of transgenes in mammalian cells (Fukunaga et al., 2012; Kamiya et al., 2007; Kamiya et al., 2013; Nishikawa et al., 2003; Sumida et al., 2006). In addition, NHEJ rejoins two broken ends, while in DNA integration, for example, linear plasmid integration, the two ends are present on the foreign DNA but these somehow join to genomic DNA. Attempt to investigate the mechanism of Metnase-stimulated DNA integration will be further described in Chapter 2.

PARP1

PARP1 is a member of the Poly (ADP-ribose) polymerase (PARP) family. PARP1 participates in the DNA damage response and plays important roles in maintaining genomic stability. PARP1 interacts with a large profile of cellular proteins. PARP1 catalyzes addition of ADP- ribose polymers (PAR) to its targets. Proteins that interact with PARP1 are involved in multiple cellular functions, including DNA damage signaling and repair, chromatin modification, transcription and RNA metabolism, replication, cell death, cell cycle regulation and mitosis (Gagne et al., 2008). PARP1 has confirmed functions in HR and aNHEJ, and may also be involved in cNHEJ (reviewed in Beck et al., 2014). Using three different chromosomal translocation assays, Wray *et.al* demonstrated that PARP1 inhibition with Olaparib or rucaparib, and knockdown of PARP1 via small-interfering RNA (siRNA) significantly reduced chromosomal translocations in murine cells and human cells (Wray et al., 2013). Considering the

structural similarity between translocations and DNA integration, we predicted that PARP1 may also participate in DNA integration.

In PARP1, the three zinc finger domains (ZnI, ZnII, ZnIII) constitute the DNA binding domain (Figure 3B). ZnI and ZnII are essential for DNA damage recognition. Binding of ZnI and ZnII to DNA induces recruitment and dimerization of the DNA binding domain to DNA (Gagne et al., 2008; Langelier et al., 2011). This is required for recruitment and activation of PARP1 at DNA damage sites. PARP1 is able to recognize both single-strand and double-strand damage (Eustermann et al., 2011). PARP1 interacts with a subset of DNA repair proteins in both BER/SSBR and DSB repair pathways to promote repair of DNA damage and maintain genome stability.

PARP1 in BER and SSBR

The engagement of PARP1 in BER and SSBR has long been established by evidence showing the PARP1 knockdown mice and cells are hypersensitive to ssDNA damage generating agents, such as alkylating agent, γ-irradiation, methyl methanesulfonate (MMS), in company with slow growth, cell cycle arrest and chromosome instability due to impaired DNA damage repair (Dantzer et al., 2000; de Murcia et al., 1997; Trucco et al., 1998). Further investigations attribute the roles of PARP1 to interaction with BER/SSBR components at both early and late phases of repair (Dantzer et al., 2000; Frouin et al., 2003; Harris et al., 2009; Heale et al., 2006; Langelier et al., 2011; Noren Hooten et al., 2011). PARP1 may modify these proteins and facilitate recruitment and activations of these key components of BER and SSBR. However, some have argued that engagement of PARP1 is not essential for BER (Strom et al., 2011; Vodenicharov et al., 2000). Therefore, more investigation is required to clarify the roles of PARP1 in BER/SSBR.

PARP1 in DSB repair and chromosomal translocations

PARP1 interacts with many proteins that contribute to DSB repair, ATM, DNA-PK, KU, MRE11, NBS1, and others (Gagne et al., 2008). PARP1 is involved in resolving HU-induced stalled replication forks through HR (Bryant et al., 2009; Hochegger et al., 2006; Yang et al., 2004; Ying et al., 2012). Current understanding states PARP1 is recruited to the stalled fork, initially to protect fork degradation by MRE11; if the stalled fork cannot be reactivated, the fork collapses to form a DSB, then PARP1 promotes restart of fork through HR (Bryant et al., 2009; Ying et al., 2012). In addition, PARP1 has to be released from DSBs to allow RPA binding and efficient HR (Illuzzi et al., 2014; Ray Chaudhuri et al., 2012; Yang et al., 2004).

PARP1 is known to interact with KU70 and DNA-PK in the absence of DNA (Gagne et al., 2008; Ray Chaudhuri et al., 2012). Both KU70 and DNA-PK are key factors for cNHEJ. *In vitro,* DNA-PK phosphorylates PARP1, which in turn adds PAR units to DNA-PK and stimulates the activities of DNA-PK. The interactions with key cNHEJ proteins suggest that PARP1 may be involved in cNHEJ as well, but more work is needed to elucidate the mechanism. The involvement of PARP1 in NHEJ is mostly in the aNHEJ pathway (Audebert et al., 2004; Paddock et al., 2011; Wang et al., 2006; Wray et al., 2013). Different from the cNHEJ pathway that relies on KU, DNA-PK, and XRCC4-Ligase IV complex, the aNHEJ pathway is operated by the PARP1, XRCC1-Ligase III or Ligase I core (Audebert et al., 2004; Simsek et al., 2011). It is likely that cNHEJ and aNHEJ are capable of operating in parallel and the decision between the two pathways in controlled by the competition between PARP1 and KU70 for DSB binding (Boboila et al., 2010; Cortizas et al., 2013; Paddock et al., 2011; Wang et al., 2006; Wray et al., 2010). *In vitro* experiments showed that KU70 has higher affinity for DSBs than PARP1 (Wang et al., 2006). It is possible that the same reason partially leads to the *in vivo* dominance of cNHEJ.

Plus, proteins such as Metnase function in a way to promote assembly of cNHEJ complex and enhance efficiency of cNHEJ, as a result, aNHEJ and aNHEJ-generated chromosomal translocations appear to be suppressed by cNHEJ components (Bennardo et al., 2008; Boboila et al., 2010; Hromas et al., 2008). PARP1 is required by aNHEJ (Mansour et al., 2010). PARP1 binds to both ssDNA and dsDNA break, and recruits MRE11 and NBS1 to DSBs (Haince et al., 2008). MRE11 catalyzes 3' to 5' end resection and promotes recruitment of CtIP to DSBs (Limbo et al., 2007; Williams et al., 2008). Both MRE11 and CtIP are important for aNHEJ, very likely for their ability of end processing (Rass et al., 2009; Zhang and Jasin, 2011). Therefore, PARP1 is an initiator of aNHEJ and contributes to DSB-induced chromosomal translocations (reviewed in Byrne et al., 2014).

Both linear and circular plasmid DNA transfected into cells is often damaged during transit into the nucleus and appears as free ends to the cell (Nickoloff et al., 1998). Linear DNA is toxic to cells as the ends of linear DNA fragments appear to be DSBs to the cells. Therefore, linear DNA must be removed either by degradation or by integration into the chromosomes to prevent cell death. It is reasonable to assume that both KU and PARP1 are recruited to bind to these double-strand free ends and engage them into DSB repair processes. We hypothesized that the plasmid DNA ends bound by PARP1 would integrate into the recipient genome through the aNHEJ pathway. To test this, we examined the effects of PARP1 inhibitors on DNA integration, and we created stable PARP1 knockdown cell line and compared DNA integration efficiency in both wild type and PARP1 negative cells. Our findings will be discussed in Chapter 3.

SUMMARY

The process of foreign DNA integration is closely related to DSB repair. The mechanism of homology-independent foreign DNA integration remains mysterious. Does foreign DNA

integrate randomly into genomic DNA because of a requirement for existing spontaneous DNA lesions? Do the abundance of cellular DNA binding proteins capture foreign DNA and bring it to existing DNA lesions through protein-protein interaction? In the absence of homology between the foreign DNA and recipient genome, does foreign DNA integrate completely randomly? Or do sequence-specific DNA binding proteins stimulate foreign DNA integration into specific sites in the genome? Previous studies have established a strong connection between homology-independent DNA integration and DSB repair through NHEJ. There are two branches of NHEJ: cNHEJ depends on Ku, DNA-PK, XRCC4, and Ligase IV and other proteins; while aNHEJ involve PARP1, XRCC1, Ligase III or Ligase I and others.

Metnase is a higher-primate only protein that promotes both DNA integration and NHEJ. Metnase retains histone modification activity and DNA processing activity. Does Metnase promote DNA integration partially due to its ability to nick the chromosome DNA and its histone modification ability? Inherited from its transposase ancestor, Metnase has specific TIR binding activity. Despite the fact that Metnase no longer catalyzes complete transposition (Roman et al., 2007), would its TIR binding direct more foreign DNA integrating into the >7000 remnants of *Mariner* transposon remnants in human genome (Liu et al., 2007)? Metnase interacts with cNHEJ proteins and promotes efficiency of end joining, would Metnase promote foreign DNA integration via the cNHEJ pathway and its DNA binding activities? If Metnase promotes DNA integration via the cNHEJ pathway, Metnase overexpression will protect the integrating DNA and recipient genome from large deletions; if the TIR-binding activity of Metnase stimulates integration into the *Mariner* TIR remnants in the human genome, overexpression of Metnase will result in increased portion of integration site near a TIR sequence. To test this hypothesis, we transfected a linear plasmid into cultured human cell lines with different Metnase expression

levels, isolated integration products and sequenced plasmid/genome junctions. Isolated junctions were identified by BLAST analysis to identify individual integration sites. Integration site patterns and modifications of the recipient genome were analyzed. The results will be discussed in Chapter 2.

PARP1 is important for initiating end resection that leads to aNHEJ, which is considered an important pathway for generating chromosomal translocations. Since chromosomal translocations share structural similarity with DNA integration, it is possible that aNHEJ is also involved in DNA integration. To test if PARP1-mediated aNHEJ is responsible for a portion of DNA integration, we tested plasmid integration efficiency under PARP1 inhibitors and in PARP1 knockdown cells. The results will be discussed in Chapter 3.

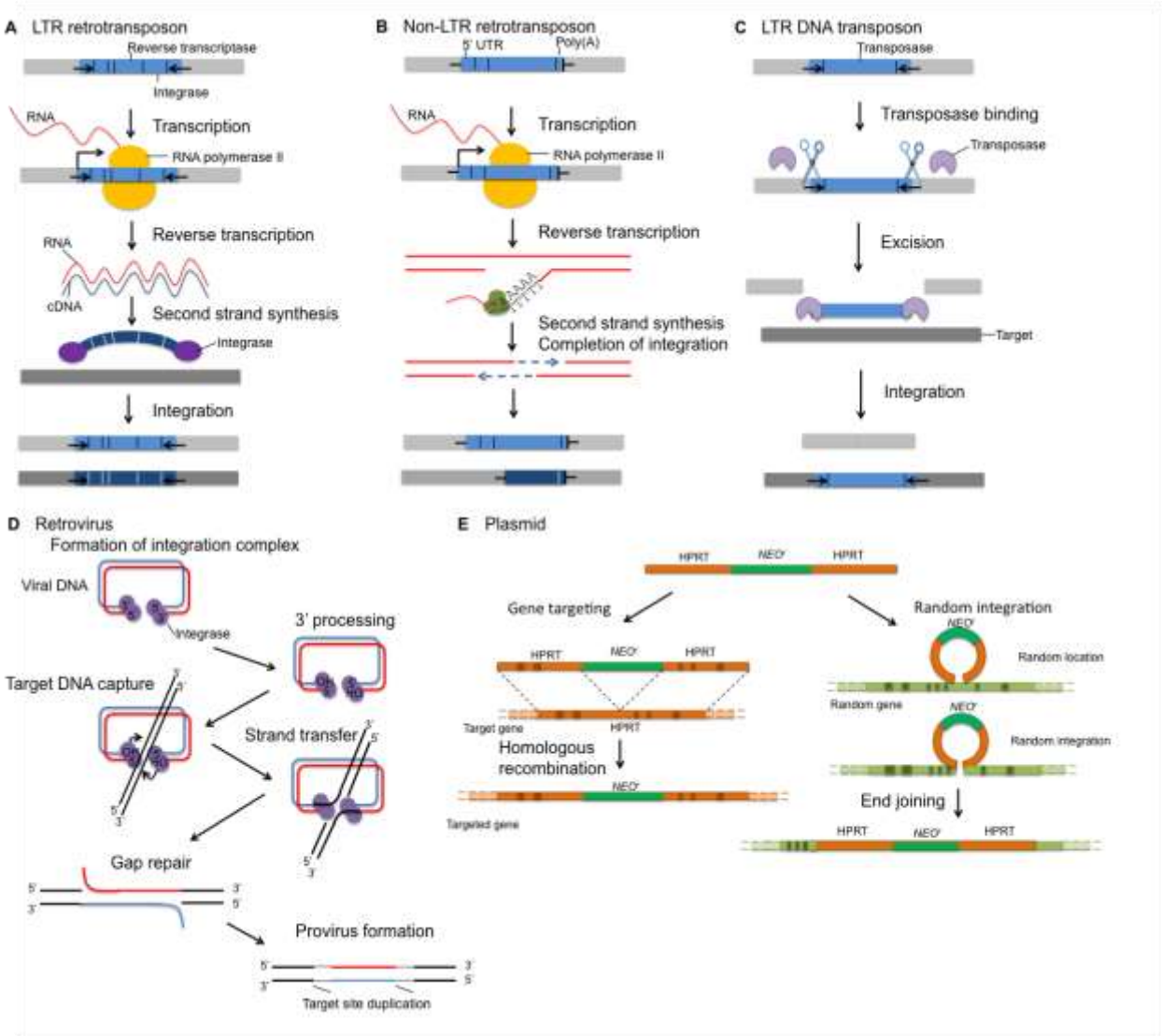

Figure 1. Examples of natural DNA integration (adopted from Hansson, 2014; Levin and Moran, 2011; Suzuki and Chew, 2012)
(A) LTR retrotransposons encode reverse transcriptase, integrase and other enzymes, flanked by two LTRs (black arrows). The 5′ LTR contains a promoter for transcription by the host RNA polymerase II. Reverse transcription produces double-strand cDNA, which is inserted into new target site by integrase (purple circles).
(B) Non-LTR retrotransposons lack LTRs and transposite by target-site-primed reverse transcription.
(C) DNA transposons encode a transposase (purple circles), flanked by TIRs (black arrows), and mobilize via a 'cut and paste' mechanism
(D) Retrovirus DNA integration is catalyzed by integrase. Integration of virus DNA results in duplication of target site
(E) Plasmid (HPRT targeting vector) has a neo^r gene flanked by HPRT homology. In gene targeting (left) results in interruption of cellular HPRT gene via homologous recombination. In random integration (right), the vector integrated in non-targeted site.

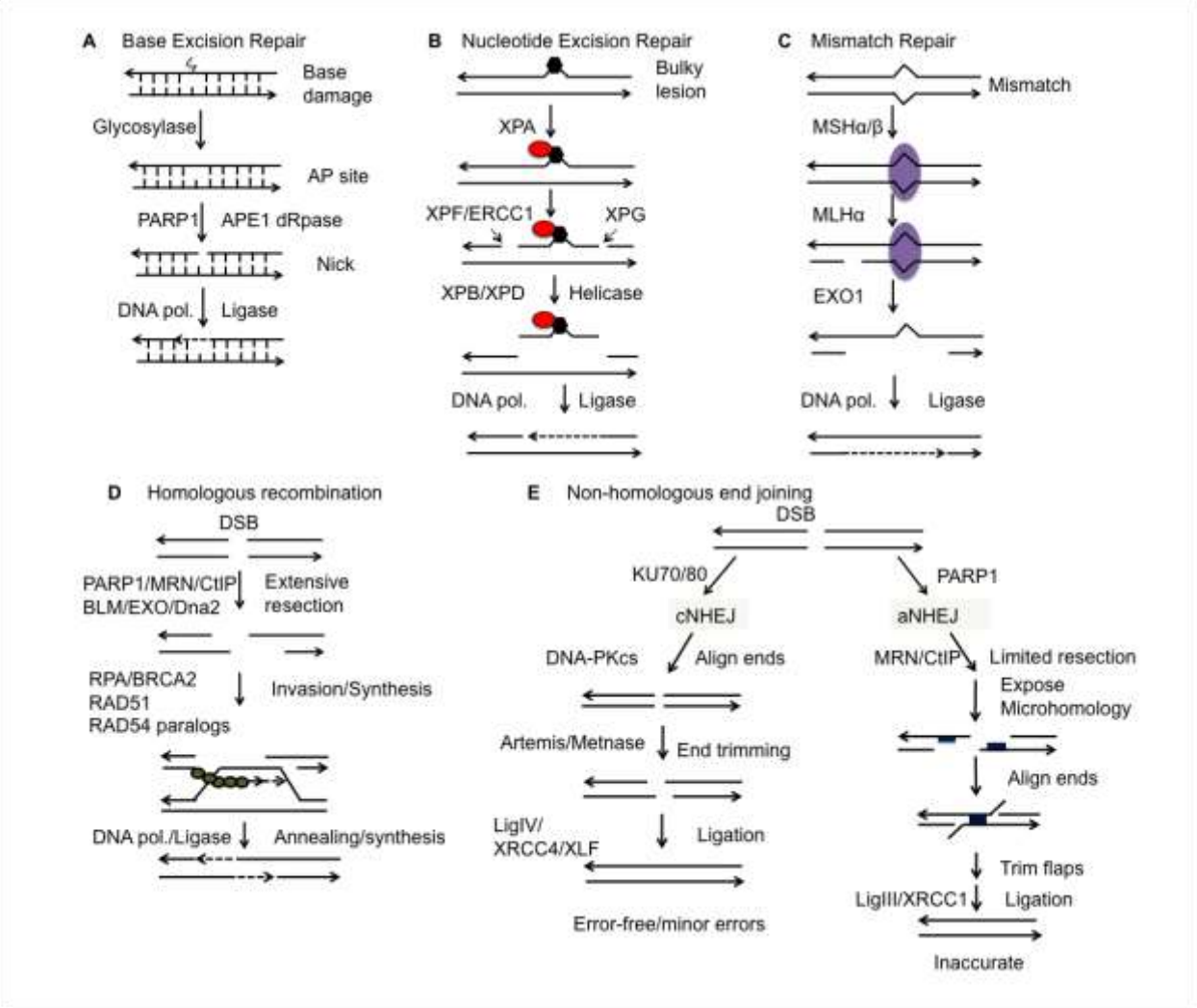

Figure 2. DNA repair pathways (adopted from Nickoloff, 2015)
(A) BER repairs damaged base and results in short patch repair.
(B) NER repairs bulky lesion by removing a ~30 nucleotides single-strand oligonucleotide carrying the lesion.
(C) MMR involves long-patch excision and resynthesis initiated at nicks distant from the mismatch.
(D) HR catalyzed by RAD51 (green oval) repairs DSBs using homologous templates, which is generally accurate.
(E) NHEJ includes relatively accurate cNHEJ, and inaccurate aNHEJ pathways distinguished by the extent of end resection, requirement for microhomology (blue boxes). KU70/80 and PARP1 compete for DSB binding and initiate cNHEJ and aNHEJ respectively.

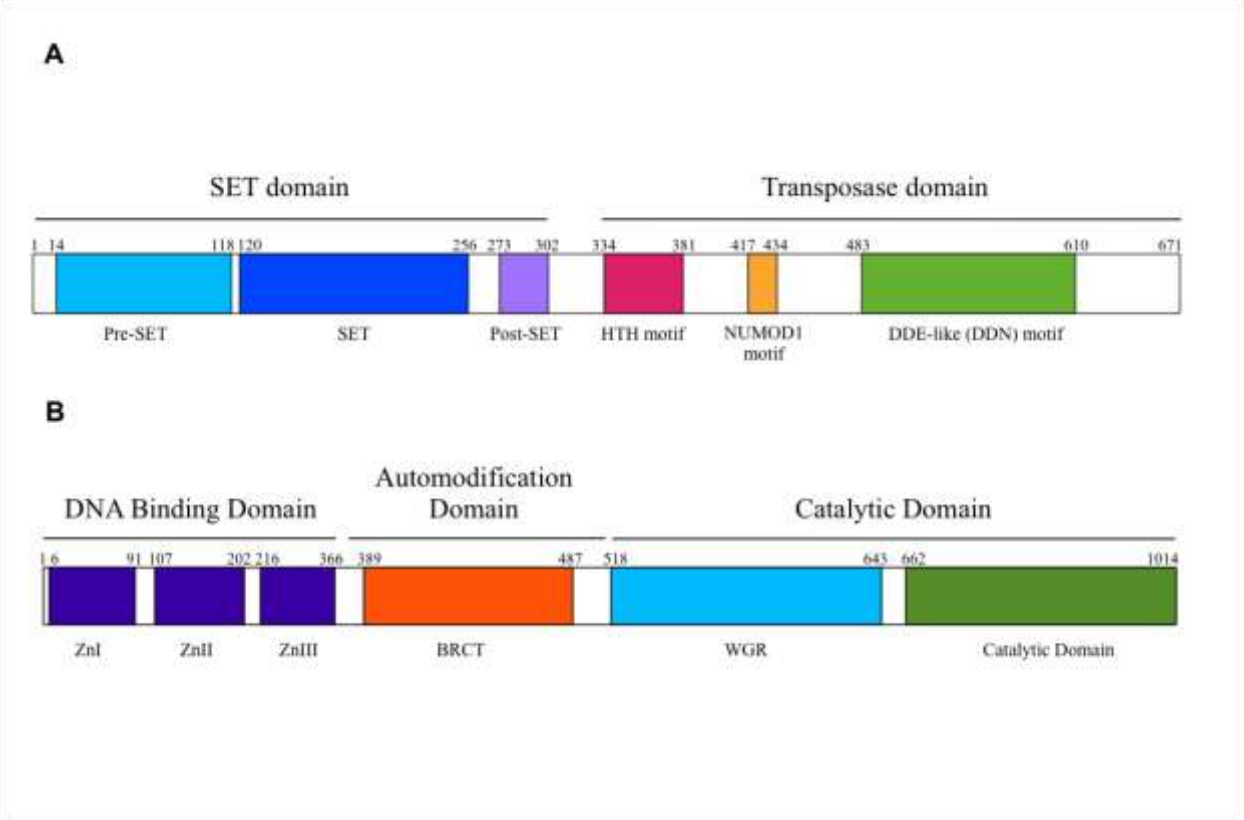

Figure 3. Schematic diagram of human Metnase and PARP1
(A) Metnase is composed of SET domain and Transposase domain. The SET domain methylates histones at DSBs. The transposase domain retains DNA nicking and cleavage activity. The DDN motif of transposase domain is important for the DNA binding activity of Metnase.
(B) PARP1 contains DNA binding domain, automodification domain, and catalytic domain. The DNA binding domain is composed of three zinc-finger domains (ZnI, ZnII, ZnIII). The automodification domain contains a BRCT domain that is important for protein-protein interactions. The catalytic domain has a WGR domain with unknown function, and a conserved catalytic domain for PAR synthesis

REFERENCES

Asante-Appiah, E., and Skalka, A.M. (1997). Molecular mechanisms in retrovirus DNA integration. Antiviral Res *36*, 139-156.

Audebert, M., Salles, B., and Calsou, P. (2004). Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. J Biol Chem *279*, 55117-55126.

Badie, S., Carlos, A.R., Folio, C., Okamoto, K., Bouwman, P., Jonkers, J., and Tarsounas, M. (2015). BRCA1 and CtIP promote alternative non-homologous end-joining at uncapped telomeres. Embo J.

Batzer, M.A., and Deininger, P.L. (2002). Alu repeats and human genomic diversity. Nat Rev Genet *3*, 370-379.

Baumann, P., and West, S.C. (1998). Role of the human RAD51 protein in homologous recombination and double-stranded-break repair. Trends Biochem Sci *23*, 247-251.

Beck, B.D., Lee, S.S., Williamson, E., Hromas, R.A., and Lee, S.H. (2011a). Biochemical characterization of metnase's endonuclease activity and its role in NHEJ repair. Biochemistry *50*, 4360-4370.

Beck, B.D., Park, S.J., Lee, Y.J., Roman, Y., Hromas, R.A., and Lee, S.H. (2008). Human Pso4 is a metnase (SETMAR)-binding partner that regulates metnase function in DNA repair. J Biol Chem *283*, 9023-9030.

Beck, C., Robert, I., Reina-San-Martin, B., Schreiber, V., and Dantzer, F. (2014). Poly(ADP-ribose) polymerases in double-strand break repair: Focus on PARP1, PARP2 and PARP3. Exp Cell Res *329*, 18-25.

Beck, C.R., Garcia-Perez, J.L., Badge, R.M., and Moran, J.V. (2011b). LINE-1 elements in structural variation and disease. Annu Rev Genomics Hum Genet *12*, 187-215.

Bekker-Jensen, S., and Mailand, N. (2010). Assembly and function of DNA double-strand break repair foci in mammalian cells. DNA Repair (Amst) *9*, 1219-1228.

Belancio, V.P., Hedges, D.J., and Deininger, P. (2008). Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res *18*, 343-358.

Belancio, V.P., Roy-Engel, A.M., and Deininger, P.L. (2010). All y'all need to know 'bout retroelements in cancer. Semin Cancer Biol *20*, 200-210.

Belshaw, R., Dawson, A.L., Woolven-Allen, J., Redding, J., Burt, A., and Tristem, M. (2005). Genomewide screening reveals high levels of insertional polymorphism in the human

endogenous retrovirus family HERV-K(HML2): implications for present-day activity. J Virol *79*, 12507-12514.

Belshaw, R., Pereira, V., Katzourakis, A., Talbot, G., Paces, J., Burt, A., and Tristem, M. (2004). Long-term reinfection of the human genome by endogenous retroviruses. Proc Natl Acad Sci U S A *101*, 4894-4899.

Bennardo, N., Cheng, A., Huang, N., and Stark, J.M. (2008). Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. PLoS Genet *4*, e1000110.

Bergsmedh, A., Szeles, A., Henriksson, M., Bratt, A., Folkman, M.J., Spetz, A.L., and Holmgren, L. (2001). Horizontal transfer of oncogenes by uptake of apoptotic bodies. Proc Natl Acad Sci U S A *98*, 6407-6411.

Bertolini, L.R., Bertolini, M., Maga, E.A., Madden, K.R., and Murray, J.D. (2009). Increased gene targeting in Ku70 and Xrcc4 transiently deficient human somatic cells. Mol Biotechnol *41*, 106-114.

Boboila, C., Jankovic, M., Yan, C.T., Wang, J.H., Wesemann, D.R., Zhang, T., Fazeli, A., Feldman, L., Nussenzweig, A., Nussenzweig, M., *et al.* (2010). Alternative end-joining catalyzes robust IgH locus deletions and translocations in the combined absence of ligase 4 and Ku70. Proc Natl Acad Sci U S A *107*, 3034-3039.

Boeke, J.D., and Chapman, K.B. (1991). Retrotransposition mechanisms. Current opinion in cell biology *3*, 502-507.

Boone, C., Bussey, H., Greene, D., Thomas, D.Y., and Vernet, T. (1986). Yeast killer toxin: site-directed mutations implicate the precursor protein as the immunity component. Cell *46*, 105-113.

Bryant, H.E., Petermann, E., Schultz, N., Jemth, A.S., Loseva, O., Issaeva, N., Johansson, F., Fernandez, S., McGlynn, P., and Helleday, T. (2009). PARP is activated at stalled forks to mediate Mre11-dependent replication restart and recombination. Embo J *28*, 2601-2615.

Bushman, F., Lewinski, M., Ciuffi, A., Barr, S., Leipzig, J., Hannenhalli, S., and Hoffmann, C. (2005). Genome-wide analysis of retroviral DNA integration. Nat Rev Microbiol *3*, 848-858.

Byrne, M., Wray, J., Reinert, B., Wu, Y., Nickoloff, J., Lee, S.H., Hromas, R., and Williamson, E. (2014). Mechanisms of oncogenic chromosomal translocations. Ann N Y Acad Sci *1310*, 89-97.

Bzymek, M., and Lovett, S.T. (2001). Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. Proc Natl Acad Sci U S A *98*, 8319-8325.

Capy, P., Vitalis, R., Langin, T., Higuet, D., and Bazin, C. (1996). Relationships between transposable elements based upon the integrase-transposase domains: is there a common ancestor? J Mol Evol *42*, 359-368.

Carreira, P.E., Richardson, S.R., and Faulkner, G.J. (2014). L1 retrotransposons, cancer stem cells and oncogenesis. The FEBS journal *281*, 63-73.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J., and Voytas, D.F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res *39*, e82.

Chen, L., Trujillo, K., Sung, P., and Tomkinson, A.E. (2000). Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent protein kinase. J Biol Chem *275*, 26196-26205.

Chen, Y., Williams, V., Filippova, M., Filippov, V., and Duerksen-Hughes, P. (2014). Viral carcinogenesis: factors inducing DNA damage and virus integration. Cancers (Basel) *6*, 2155-2186.

Christensen, S.M., and Eickbush, T.H. (2005). R2 target-primed reverse transcription: ordered cleavage and polymerization steps by protein subunits asymmetrically bound to the target DNA. Mol Cell Biol *25*, 6617-6628.

Christian, S.L., Robinson, W.P., Huang, B., Mutirangura, A., Line, M.R., Nakao, M., Surti, U., Chakravarti, A., and Ledbetter, D.H. (1995). Molecular characterization of two proximal deletion breakpoint regions in both Prader-Willi and Angelman syndrome patients. Am J Hum Genet *57*, 40-48.

Cong, L., and Zhang, F. (2015). Genome engineering using CRISPR-Cas9 system. Methods Mol Biol *1239*, 197-217.

Cordaux, R., and Batzer, M.A. (2009). The impact of retrotransposons on human genome evolution. Nat Rev Genet *10*, 691-703.

Cordaux, R., Udit, S., Batzer, M.A., and Feschotte, C. (2006). Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. Proc Natl Acad Sci U S A *103*, 8101-8106.

Cortizas, E.M., Zahn, A., Hajjar, M.E., Patenaude, A.M., Di Noia, J.M., and Verdun, R.E. (2013). Alternative end-joining and classical nonhomologous end-joining pathways repair different types of double-strand breaks during class-switch recombination. J Immunol *191*, 5751-5763.

Cost, G.J., Feng, Q., Jacquier, A., and Boeke, J.D. (2002). Human L1 element target-primed reverse transcription in vitro. Embo J *21*, 5899-5910.

Covarrubias, L., Nishida, Y., and Mintz, B. (1986). Early postimplantation embryo lethality due to DNA rearrangements in a transgenic mouse strain. Proc Natl Acad Sci U S A *83*, 6020-6024.

Covarrubias, L., Nishida, Y., Terao, M., D'Eustachio, P., and Mintz, B. (1987). Cellular DNA rearrangements and early developmental arrest caused by DNA insertion in transgenic mouse embryos. Mol Cell Biol *7*, 2243-2247.

Craig, N.L. (1995). Unity in transposition reactions. Science *270*, 253-254.

Craigie, R., and Bushman, F.D. (2012). HIV DNA integration. Cold Spring Harb Perspect Med *2*, a006890.

Critchlow, S.E., Bowater, R.P., and Jackson, S.P. (1997). Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. Current biology : CB *7*, 588-598.

Damania, B. (2007). DNA tumor viruses and human cancer. Trends Microbiol *15*, 38-44.

Dantzer, F., de La Rubia, G., Menissier-De Murcia, J., Hostomsky, Z., de Murcia, G., and Schreiber, V. (2000). Base excision repair is impaired in mammalian cells lacking Poly(ADP-ribose) polymerase-1. Biochemistry *39*, 7559-7569.

Davis, A.J., Chen, B.P., and Chen, D.J. (2014). DNA-PK: a dynamic enzyme in a versatile DSB repair pathway. DNA Repair (Amst) *17*, 21-29.

De Haro, L.P., Wray, J., Williamson, E.A., Durant, S.T., Corwin, L., Gentry, A.C., Osheroff, N., Lee, S.H., Hromas, R., and Nickoloff, J.A. (2010). Metnase promotes restart and repair of stalled and collapsed replication forks. Nucleic Acids Res *38*, 5681-5691.

de Murcia, J.M., Niedergang, C., Trucco, C., Ricoul, M., Dutrillaux, B., Mark, M., Oliver, F.J., Masson, M., Dierich, A., LeMeur, M., *et al.* (1997). Requirement of poly(ADP-ribose) polymerase in recovery from DNA damage in mice and in cells. Proc Natl Acad Sci U S A *94*, 7303-7307.

de Vries, J., and Wackernagel, W. (2002). Integration of foreign DNA during natural transformation of Acinetobacter sp. by homology-facilitated illegitimate recombination. Proc Natl Acad Sci U S A *99*, 2094-2099.

Della-Maria, J., Zhou, Y., Tsai, M.S., Kuhnlein, J., Carney, J.P., Paull, T.T., and Tomkinson, A.E. (2011). Human Mre11/human Rad50/Nbs1 and DNA ligase IIIalpha/XRCC1 protein complexes act together in an alternative nonhomologous end joining pathway. J Biol Chem *286*, 33845-33853.

Dianov, G.L., and Hubscher, U. (2013). Mammalian Base Excision Repair: the Forgotten Archangel. Nucleic acids research.

Doak, T.G., Doerder, F.P., Jahn, C.L., and Herrick, G. (1994). A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. Proc Natl Acad Sci U S A *91*, 942-946.

Dodson, M.K., Thibodeau, S.N., Halling, K.C., Cliby, W.A., Delacey, K.A., Hartmann, L.C., Podratz, K.C., and Jenkins, R.B. (1993). Pcr Microsatellite Instability in Sporadic Epithelial Ovarian-Carcinoma. American Journal of Human Genetics *53*, 292-292.

Doerfler, W. (2011). Epigenetic consequences of foreign DNA insertions: de novo methylation and global alterations of methylation patterns in recipient genomes. Rev Med Virol *21*, 336-346.

Doerfler, W. (2012). Impact of foreign DNA integration on tumor biology and on evolution via epigenetic alterations. Epigenomics *4*, 41-49.

Dubose, A.J., Lichtenstein, S.T., Narisu, N., Bonnycastle, L.L., Swift, A.J., Chines, P.S., and Collins, F.S. (2013). Use of microarray hybrid capture and next-generation sequencing to identify the anatomy of a transgene. Nucleic Acids Res *41*, e70.

Ehnfors, J., Kost-Alimova, M., Persson, N.L., Bergsmedh, A., Castro, J., Levchenko-Tegnebratt, T., Yang, L., Panaretakis, T., and Holmgren, L. (2009). Horizontal transfer of tumor DNA to endothelial cells in vivo. Cell Death Differ *16*, 749-757.

Eickbush, T.H., and Jamburuthugoda, V.K. (2008). The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Res *134*, 221-234.

Eustermann, S., Videler, H., Yang, J.C., Cole, P.T., Gruszka, D., Veprintsev, D., and Neuhaus, D. (2011). The DNA-binding domain of human PARP-1 interacts with DNA single-strand breaks as a monomer through its second zinc finger. J Mol Biol *407*, 149-170.

Feng, H., Shuda, M., Chang, Y., and Moore, P.S. (2008). Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science *319*, 1096-1100.

Feng, Q., Moran, J.V., Kazazian, H.H., Jr., and Boeke, J.D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell *87*, 905-916.

Fnu, S., Williamson, E.A., De Haro, L.P., Brenneman, M., Wray, J., Shaheen, M., Radhakrishnan, K., Lee, S.H., Nickoloff, J.A., and Hromas, R. (2011). Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. Proc Natl Acad Sci U S A *108*, 540-545.

Frouin, I., Maga, G., Denegri, M., Riva, F., Savio, M., Spadari, S., Prosperi, E., and Scovassi, A.I. (2003). Human proliferating cell nuclear antigen, poly(ADP-ribose) polymerase-1, and p21waf1/cip1. A dynamic exchange of partners. J Biol Chem *278*, 39265-39268.

Fujimaki, K., Aratani, Y., Fujisawa, S., Motomura, S., Okubo, T., and Koyama, H. (1996). DNA topoisomerase II inhibitors enhance random integration of transfected vectors into human chromosomes. Somat Cell Mol Genet *22*, 279-290.

Fujitani, Y., and Kobayashi, I.I. (1995). Random-walk model of homologous recombination. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics *52*, 6607-6622.

Fujitani, Y., Yamamoto, K., and Kobayashi, I. (1995). Dependence of frequency of homologous recombination on the homology length. Genetics *140*, 797-809.

Fukui, K. (2010). DNA mismatch repair in eukaryotes and bacteria. Journal of nucleic acids *2010*.

Fukunaga, S., Kanda, G., Tanase, J., Harashima, H., Ohyama, T., and Kamiya, H. (2012). A designed curved DNA sequence remarkably enhances transgene expression from plasmid DNA in mouse liver. Gene Ther *19*, 828-835.

Fukushige, S., and Sauer, B. (1992). Genomic targeting with a positive-selection lox integration vector allows highly reproducible gene expression in mammalian cells. Proc Natl Acad Sci U S A *89*, 7905-7909.

Gagne, J.P., Isabelle, M., Lo, K.S., Bourassa, S., Hendzel, M.J., Dawson, V.L., Dawson, T.M., and Poirier, G.G. (2008). Proteome-wide identification of poly(ADP-ribose) binding proteins and poly(ADP-ribose)-associated protein complexes. Nucleic Acids Res *36*, 6959-6976.

Garrels, W., Ivics, Z., and Kues, W.A. (2012). Precision genetic engineering in large mammals. Trends Biotechnol *30*, 386-393.

Gauwerky, C.E., and Croce, C.M. (1993). Chromosomal translocations in leukaemia. Semin Cancer Biol *4*, 333-340.

Ghezraoui, H., Piganeau, M., Renouf, B., Renaud, J.B., Sallmyr, A., Ruis, B., Oh, S., Tomkinson, A.E., Hendrickson, E.A., Giovannangeli, C., *et al.* (2014). Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. Mol Cell *55*, 829-842.

Giraud, C., Winocour, E., and Berns, K.I. (1994). Site-specific integration by adeno-associated virus is directed by a cellular DNA sequence. Proc Natl Acad Sci U S A *91*, 10039-10043.

Goodier, J.L., and Kazazian, H.H., Jr. (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell *135*, 23-35.

Grundy, G.J., Moulding, H.A., Caldecott, K.W., and Rulten, S.L. (2014). One ring to bring them all--the role of Ku in mammalian non-homologous end joining. DNA Repair (Amst) *17*, 30-38.

Gu, W.G. (2014). Newly approved integrase inhibitors for clinical treatment of AIDS. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie *68*, 917-921.

Haince, J.F., McDonald, D., Rodrigue, A., Dery, U., Masson, J.Y., Hendzel, M.J., and Poirier, G.G. (2008). PARP1-dependent kinetics of recruitment of MRE11 and NBS1 proteins to multiple DNA damage sites. J Biol Chem *283*, 1197-1208.

Hallet, B., and Sherratt, D.J. (1997). Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. FEMS microbiology reviews *21*, 157-178.

Hamada, T., Sasaki, H., Seki, R., and Sakaki, Y. (1993). Mechanism of chromosomal integration of transgenes in microinjected mouse eggs: sequence analysis of genome-transgene and transgene-transgene junctions at two loci. Gene *128*, 197-202.

Hansson, G.r.K. (2014). The 2007 Nobel Prize In Physiology or Medicine - Advanced Information" (Nobelprize.org.).

Harms, K., and Wackernagel, W. (2008). The RecBCD and SbcCD DNases suppress homology-facilitated illegitimate recombination during natural transformation of Acinetobacter baylyi. Microbiology *154*, 2437-2445.

Harris, J.L., Jakob, B., Taucher-Scholz, G., Dianov, G.L., Becherel, O.J., and Lavin, M.F. (2009). Aprataxin, poly-ADP ribose polymerase 1 (PARP-1) and apurinic endonuclease 1 (APE1) function together to protect the genome against oxidative damage. Hum Mol Genet *18*, 4102-4117.

Hasty, P., Rivera-Perez, J., and Bradley, A. (1991). The length of homology required for gene targeting in embryonic stem cells. Mol Cell Biol *11*, 5586-5591.

Havecker, E.R., Gao, X., and Voytas, D.F. (2004). The diversity of LTR retrotransposons. Genome biology *5*, 225.

Heale, J.T., Ball, A.R., Jr., Schmiesing, J.A., Kim, J.S., Kong, X., Zhou, S., Hudson, D.F., Earnshaw, W.C., and Yokomori, K. (2006). Condensin I interacts with the PARP-1-XRCC1 complex and functions in DNA single-strand break repair. Mol Cell *21*, 837-848.

Hebbar, P.B., and Archer, T.K. (2008). Altered histone H1 stoichiometry and an absence of nucleosome positioning on transfected DNA. J Biol Chem *283*, 4595-4601.

Hochegger, H., Dejsuphong, D., Fukushima, T., Morrison, C., Sonoda, E., Schreiber, V., Zhao, G.Y., Saberi, A., Masutani, M., Adachi, N., *et al.* (2006). Parp-1 protects homologous recombination from interference by Ku and Ligase IV in vertebrate cells. Embo J *25*, 1305-1314.

Holmgren, L. (2010). Horizontal gene transfer: you are what you eat. Biochem Biophys Res Commun *396*, 147-151.

Hromas, R., Williamson, E.A., Fnu, S., Lee, Y.J., Park, S.J., Beck, B.D., You, J.S., Leitao, A., Nickoloff, J.A., and Lee, S.H. (2012). Chk1 phosphorylation of Metnase enhances DNA repair but inhibits replication fork restart. Oncogene *31*, 4245-4254.

Hromas, R., Wray, J., Lee, S.H., Martinez, L., Farrington, J., Corwin, L.K., Ramsey, H., Nickoloff, J.A., and Williamson, E.A. (2008). The human set and transposase domain protein Metnase interacts with DNA Ligase IV and enhances the efficiency and accuracy of non-homologous end-joining. DNA Repair (Amst) *7*, 1927-1937.

Hulter, N., and Wackernagel, W. (2008). Double illegitimate recombination events integrate DNA segments through two different mechanisms during natural transformation of Acinetobacter baylyi. Mol Microbiol *67*, 984-995.

Iiizumi, S., Kurosawa, A., So, S., Ishii, Y., Chikaraishi, Y., Ishii, A., Koyama, H., and Adachi, N. (2008). Impact of non-homologous end-joining deficiency on random and targeted DNA integration: implications for gene targeting. Nucleic Acids Res *36*, 6333-6342.

Illuzzi, G., Fouquerel, E., Ame, J.C., Noll, A., Rehmet, K., Nasheuer, H.P., Dantzer, F., and Schreiber, V. (2014). PARG is dispensable for recovery from transient replicative stress but required to prevent detrimental accumulation of poly(ADP-ribose) upon prolonged replicative stress. Nucleic Acids Res *42*, 7776-7792.

Ionov, Y., Peinado, M.A., Malkhosyan, S., Shibata, D., and Perucho, M. (1993). Ubiquitous Somatic Mutations in Simple Repeated Sequences Reveal a New Mechanism for Colonic Carcinogenesis. Nature *363*, 558-561.

Ishibashi, K., Suzuki, K., Ando, Y., Takakura, C., and Inoue, H. (2006). Nonhomologous chromosomal integration of foreign DNA is completely dependent on MUS-53 (human Lig4 homolog) in Neurospora. Proc Natl Acad Sci U S A *103*, 14871-14876.

Jacobson, J.W., Medhora, M.M., and Hartl, D.L. (1986). Molecular structure of a somatically unstable transposable element in Drosophila. Proc Natl Acad Sci U S A *83*, 8684-8688.

Jeyaratnam, D.C., Baduin, B.S., Hansen, M.C., Hansen, M., Jorgensen, J.M., Aggerholm, A., Ommen, H.B., Hokland, P., and Nyvold, C.G. (2014). Delineation of known and new transcript variants of the SETMAR (Metnase) gene and the expression profile in hematologic neoplasms. Exp Hematol *42*, 448-456 e444.

Jha, A.R., Pillai, S.K., York, V.A., Sharp, E.R., Storm, E.C., Wachter, D.J., Martin, J.N., Deeks, S.G., Rosenberg, M.G., Nixon, D.F., *et al.* (2009). Cross-sectional dating of novel haplotypes of HERV-K 113 and HERV-K 115 indicate these proviruses originated in Africa before Homo sapiens. Mol Biol Evol *26*, 2617-2626.

Jin, X., Mo, M.H., Wei, Z., Huang, X.W., and Zhang, K.Q. (2005). Transformation and mutagenesis of the nematode-trapping fungus Monacrosporium sphaeroides by restriction enzyme-mediated integration (REMI). J Microbiol *43*, 417-423.

Johnsborg, O., Eldholm, V., and Havarstein, L.S. (2007). Natural genetic transformation: prevalence, mechanisms and function. Res Microbiol *158*, 767-778.

Jung, Y.D., Ahn, K., Kim, Y.J., Bae, J.H., Lee, J.R., and Kim, H.S. (2013). Retroelements: molecular features and implications for disease. Genes Genet Syst *88*, 31-43.

Kamiya, H., Fukunaga, S., Ohyama, T., and Harashima, H. (2007). The location of the left-handedly curved DNA sequence affects exogenous DNA expression in vivo. Archives of biochemistry and biophysics *461*, 7-12.

Kamiya, H., Miyamoto, S., Goto, H., Kanda, G.N., Kobayashi, M., Matsuoka, I., and Harashima, H. (2013). Enhanced transgene expression from chromatinized plasmid DNA in mouse liver. International journal of pharmaceutics *441*, 146-150.

Kassiotis, G. (2014). Endogenous retroviruses and the development of cancer. J Immunol *192*, 1343-1349.

Kato, S., Anderson, R.A., and Camerini-Otero, R.D. (1986). Foreign DNA introduced by calcium phosphate is integrated into repetitive DNA elements of the mouse L cell genome. Mol Cell Biol *6*, 1787-1795.

Katoh, I., and Kurata, S. (2013). Association of endogenous retroviruses and long terminal repeats with human disorders. Frontiers in oncology *3*, 234.

Kim, H.S., Chen, Q., Kim, S.K., Nickoloff, J.A., Hromas, R., Georgiadis, M.M., and Lee, S.H. (2014). The DDN catalytic motif is required for Metnase functions in non-homologous end joining (NHEJ) repair and replication restart. J Biol Chem *289*, 10930-10938.

Kinner, A., Wu, W., Staudt, C., and Iliakis, G. (2008). Gamma-H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin. Nucleic Acids Res *36*, 5678-5694.

Kiyosawa, H., and Chance, P.F. (1996). Primate origin of the CMT1A-REP repeat and analysis of a putative transposon-associated recombinational hotspot. Hum Mol Genet *5*, 745-753.

Koonin, E.V., Senkevich, T.G., and Dolja, V.V. (2006). The ancient Virus World and evolution of cells. Biology direct *1*, 29.

Krishnan, L., and Engelman, A. (2012). Retroviral integrase proteins and HIV-1 DNA integration. J Biol Chem *287*, 40858-40866.

Kunz, C., Saito, Y., and Schar, P. (2009). DNA Repair in mammalian cells: Mismatched repair: variations on a theme. Cell Mol Life Sci *66*, 1021-1038.

Kuo, L.J., and Yang, L.X. (2008). Gamma-H2AX - a novel biomarker for DNA double-strand breaks. In vivo *22*, 305-309.

Kurosawa, A., and Adachi, N. (2010). Functions and regulation of Artemis: a goddess in the maintenance of genome integrity. Journal of radiation research *51*, 503-509.

Lacy, E., Roberts, S., Evans, E.P., Burtenshaw, M.D., and Costantini, F.D. (1983). A foreign beta-globin gene in transgenic mice: integration at abnormal chromosomal positions and expression in inappropriate tissues. Cell *34*, 343-358.

Lampe, D.J., Churchill, M.E., and Robertson, H.M. (1996). A purified mariner transposase is sufficient to mediate transposition in vitro. Embo J *15*, 5470-5479.

Lampe, D.J., Witherspoon, D.J., Soto-Adames, F.N., and Robertson, H.M. (2003). Recent horizontal transfer of mellifera subfamily mariner transposons into insect lineages representing four different orders shows that selection acts only during horizontal transfer. Mol Biol Evol *20*, 554-562.

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

Langelier, M.F., Planck, J.L., Roy, S., and Pascal, J.M. (2011). Crystal structures of poly(ADP-ribose) polymerase-1 (PARP-1) zinc fingers bound to DNA: structural and functional insights into DNA-dependent PARP-1 activity. J Biol Chem *286*, 10690-10701.

Lawrence, J.G., and Hartl, D.L. (1992). Inference of horizontal genetic transfer from molecular data: an approach using the bootstrap. Genetics *131*, 753-760.

Lee-Theilen, M., Matthews, A.J., Kelly, D., Zheng, S., and Chaudhuri, J. (2011). CtIP promotes microhomology-mediated alternative end joining during class-switch recombination. Nat Struct Mol Biol *18*, 75-79.

Lee, S.H., Oshige, M., Durant, S.T., Rasila, K.K., Williamson, E.A., Ramsey, H., Kwan, L., Nickoloff, J.A., and Hromas, R. (2005). The SET domain protein Metnase mediates foreign DNA integration and links integration to nonhomologous end-joining repair. Proc Natl Acad Sci U S A *102*, 18075-18080.

Levin, H.L., and Moran, J.V. (2011). Dynamic interactions between transposable elements and their hosts. Nat Rev Genet *12*, 615-627.

Li, T., Huang, S., Jiang, W.Z., Wright, D., Spalding, M.H., Weeks, D.P., and Yang, B. (2011a). TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Res *39*, 359-372.

Li, T., Huang, S., Zhao, X., Wright, D.A., Carpenter, S., Spalding, M.H., Weeks, D.P., and Yang, B. (2011b). Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Res *39*, 6315-6325.

Lichtenberg, U., Zock, C., and Doerfler, W. (1988). Integration of foreign DNA into mammalian genome can be associated with hypomethylation at site of insertion. Virus Res *11*, 335-342.

Limbo, O., Chahwan, C., Yamada, Y., de Bruin, R.A., Wittenberg, C., and Russell, P. (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*, 134-146.

Little, K.C., and Chartrand, P. (2004). Genomic DNA is captured and amplified during double-strand break (DSB) repair in human cells. Oncogene *23*, 4166-4172.

Liu, D., Bischerour, J., Siddique, A., Buisine, N., Bigot, Y., and Chalmers, R. (2007). The human SETMAR protein preserves most of the activities of the ancestral Hsmar1 transposase. Mol Cell Biol *27*, 1125-1132.

Liu, Z., Wang, Y., Shen, Y., Guo, W., Hao, S., and Liu, B. (2004). Extensive alterations in DNA methylation and transcription in rice caused by introgression from Zizania latifolia. Plant Mol Biol *54*, 571-582.

Lohe, A.R., De Aguiar, D., and Hartl, D.L. (1997). Mutations in the mariner transposase: the D,D(35)E consensus sequence is nonfunctional. Proc Natl Acad Sci U S A *94*, 1293-1297.

Lohe, A.R., Moriyama, E.N., Lidholm, D.A., and Hartl, D.L. (1995). Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. Mol Biol Evol *12*, 62-72.

Luan, D.D., Korman, M.H., Jakubczak, J.L., and Eickbush, T.H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell *72*, 595-605.

Mahon, K.A., Overbeek, P.A., and Westphal, H. (1988). Prenatal lethality in a transgenic mouse line is the result of a chromosomal translocation. Proc Natl Acad Sci U S A *85*, 1165-1168.

Mandal, P.K., and Kazazian, H.H., Jr. (2008). SnapShot: Vertebrate transposons. Cell *135*, 192-192 e191.

Mani, M., Kandavelou, K., Dy, F.J., Durai, S., and Chandrasegaran, S. (2005). Design, engineering, and characterization of zinc finger nucleases. Biochem Biophys Res Commun *335*, 447-457.

Mansour, W.Y., Rhein, T., and Dahm-Daphi, J. (2010). The alternative end-joining pathway for repair of DNA double-strand breaks requires PARP1 but is not dependent upon microhomologies. Nucleic Acids Res *38*, 6065-6077.

Mark, W.H., Signorelli, K., Blum, M., Kwee, L., and Lacy, E. (1992). Genomic structure of the locus associated with an insertional mutation in line 4 transgenic mice. Genomics *13*, 159-166.

McClintock, B. (1953). Induction of Instability at Selected Loci in Maize. Genetics *38*, 579-599.

Merrihew, R.V., Marburger, K., Pennington, S.L., Roth, D.B., and Wilson, J.H. (1996). High-frequency illegitimate integration of transfected DNA at preintegrated target sites in a mammalian genome. Mol Cell Biol *16*, 10-18.

Miller, J.C., Holmes, M.C., Wang, J., Guschin, D.Y., Lee, Y.L., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K.A., *et al.* (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol *25*, 778-785.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., *et al.* (2011). A TALE nuclease architecture for efficient genome editing. Nat Biotechnol *29*, 143-148.

Morton, J., Davis, M.W., Jorgensen, E.M., and Carroll, D. (2006). Induction and repair of zinc-finger nuclease-targeted double-strand breaks in Caenorhabditis elegans somatic cells. Proc Natl Acad Sci U S A *103*, 16370-16375.

Muller, K., Heller, H., and Doerfler, W. (2001). Foreign DNA integration. Genome-wide perturbations of methylation and transcription in the recipient genomes. J Biol Chem *276*, 14271-14278.

Narezkina, A., Taganov, K.D., Litwin, S., Stoyanova, R., Hayashi, J., Seeger, C., Skalka, A.M., and Katz, R.A. (2004). Genome-wide analyses of avian sarcoma virus integration sites. J Virol *78*, 11656-11663.

Nickoloff, J.A. (2015). DNA Repair Dysregulation in Cancer: From Molecular Mechanisms to Synthetic Lethal Opportunities. In *Stress Response Pathways in Cancer*, G.T. Wondrak, ed. (Springer Science+Business Media), pp. 7-28.

Nickoloff, J.A., Spirio, L.N., and Reynolds, R.J. (1998). A comparison of calcium phosphate coprecipitation and electroporation. Implications for studies on the genetic effects of DNA damage. Mol Biotechnol *10*, 93-101.

Nimonkar, A.V., Genschel, J., Kinoshita, E., Polaczek, P., Campbell, J.L., Wyman, C., Modrich, P., and Kowalczykowski, S.C. (2011). BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. Genes Dev *25*, 350-362.

Nishikawa, J., Amano, M., Fukue, Y., Tanaka, S., Kishi, H., Hirota, Y., Yoda, K., and Ohyama, T. (2003). Left-handedly curved DNA regulates accessibility to cis-DNA elements in chromatin. Nucleic Acids Res *31*, 6651-6662.

Noren Hooten, N., Kompaniez, K., Barnes, J., Lohani, A., and Evans, M.K. (2011). Poly(ADP-ribose) polymerase 1 (PARP-1) binds to 8-oxoguanine-DNA glycosylase (OGG1). J Biol Chem *286*, 44679-44690.

Nouspikel, T. (2009). DNA repair in mammalian cells : Nucleotide excision repair: variations on versatility. Cell Mol Life Sci *66*, 994-1009.

Nystrom-Lahti, M., Kristo, P., Nicolaides, N.C., Chang, S.Y., Aaltonen, L.A., Moisio, A.L., Jarvinen, H.J., Mecklin, J.P., Kinzler, K.W., Vogelstein, B., *et al.* (1995). Founding mutations and Alu-mediated recombination in hereditary colon cancer. Nat Med *1*, 1203-1206.

O'Gorman, S., Fox, D.T., and Wahl, G.M. (1991). Recombinase-mediated gene activation and site-specific integration in mammalian cells. Science *251*, 1351-1355.

Oosumi, T., Belknap, W.R., and Garlick, B. (1995). Mariner transposons in humans. Nature *378*, 672.

Ostertag, E.M., Goodier, J.L., Zhang, Y., and Kazazian, H.H., Jr. (2003). SVA elements are nonautonomous retrotransposons that cause disease in humans. Am J Hum Genet *73*, 1444-1451.

Overbeek, P.A., Aguilar-Cordova, E., Hanten, G., Schaffner, D.L., Patel, P., Lebovitz, R.M., and Lieberman, M.W. (1991). Coinjection strategy for visual identification of transgenic mice. Transgenic Res *1*, 31-37.

Paddock, M.N., Bauman, A.T., Higdon, R., Kolker, E., Takeda, S., and Scharenberg, A.M. (2011). Competition between PARP-1 and Ku70 control the decision between high-fidelity and mutagenic DNA repair. DNA Repair (Amst) *10*, 338-343.

Pagano, J.S., Blaser, M., Buendia, M.A., Damania, B., Khalili, K., Raab-Traub, N., and Roizman, B. (2004). Infectious agents and cancer: criteria for a causal relation. Semin Cancer Biol *14*, 453-471.

Pannunzio, N.R., Li, S., Watanabe, G., and Lieber, M.R. (2014). Non-homologous end joining often uses microhomology: implications for alternative end joining. DNA Repair (Amst) *17*, 74-80.

Pardo, B., Gomez-Gonzalez, B., and Aguilera, A. (2009). DNA repair in mammalian cells: DNA double-strand break repair: how to fix a broken relationship. Cell Mol Life Sci *66*, 1039-1056.

Paull, T.T., and Lee, J.H. (2005). The Mre11/Rad50/Nbs1 complex and its role as a DNA double-strand break sensor for ATM. Cell Cycle *4*, 737-740.

Pawelczak, K.S., Bennett, S.M., and Turchi, J.J. (2011). Coordination of DNA-PK activation and nuclease processing of DNA termini in NHEJ. Antioxidants & redox signaling *14*, 2531-2543.

Perucho, M., Ionov, Y., Peinado, M.A., Malkhosyan, S., Velazquez, A., Casares, S., Quintero, A., and Shibata, D. (1994). Ubiquitous Somatic Mutations in Simple Repeats Define a Novel Mutator Mechanism for Oncogenesis. J Cell Biochem, 165-165.

Perz, J.F., Armstrong, G.L., Farrington, L.A., Hutin, Y.J., and Bell, B.P. (2006). The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. J Hepatol *45*, 529-538.

Peterson, S.E., Li, Y., Wu-Baer, F., Chait, B.T., Baer, R., Yan, H., Gottesman, M.E., and Gautier, J. (2013). Activation of DSB processing requires phosphorylation of CtIP by ATR. Mol Cell *49*, 657-667.

Plasterk, R.H., Izsvak, Z., and Ivics, Z. (1999). Resident aliens: the Tc1/mariner superfamily of transposable elements. Trends Genet *15*, 326-332.

Porteus, M. (2008). Design and testing of zinc finger nucleases for use in mammalian cells. Methods Mol Biol *435*, 47-61.

Qian, C., and Zhou, M.M. (2006). SET domain protein lysine methyltransferases: Structure, specificity and catalysis. Cellular and molecular life sciences : CMLS *63*, 2755-2763.

Rass, E., Grabarz, A., Plo, I., Gautier, J., Bertrand, P., and Lopez, B.S. (2009). Role of Mre11 in chromosomal nonhomologous end joining in mammalian cells. Nat Struct Mol Biol *16*, 819-824.

Rath, A., Hromas, R., and De Benedetti, A. (2014). Fidelity of end joining in mammalian episomes and the impact of Metnase on joint processing. BMC Mol Biol *15*, 6.

Ray Chaudhuri, A., Hashimoto, Y., Herrador, R., Neelsen, K.J., Fachinetti, D., Bermejo, R., Cocito, A., Costanzo, V., and Lopes, M. (2012). Topoisomerase I poisoning results in PARP-mediated replication fork reversal. Nat Struct Mol Biol *19*, 417-423.

Reeves, R., Gorman, C.M., and Howard, B. (1985). Minichromosome assembly of non-integrated plasmid DNA transfected into mammalian cells. Nucleic Acids Res *13*, 3599-3615.

Remus, R., Kammer, C., Heller, H., Schmitz, B., Schell, G., and Doerfler, W. (1999). Insertion of foreign DNA into an established mammalian genome can alter the methylation of cellular DNA sequences. J Virol *73*, 1010-1022.

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

Robertson, H.M. (1993). The mariner transposable element is widespread in insects. Nature *362*, 241-245.

Robertson, H.M., and MacLeod, E.G. (1993). Five major subfamilies of mariner transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. Insect Mol Biol *2*, 125-139.

Robertson, H.M., and Martos, R. (1997). Molecular evolution of the second ancient human mariner transposon, Hsmar2, illustrates patterns of neutral evolution in the human genome lineage. Gene *205*, 219-228.

Robertson, H.M., and Zumpano, K.L. (1997). Molecular evolution of an ancient mariner transposon, Hsmar1, in the human genome. Gene *205*, 203-217.

Robins, D.M., Ripley, S., Henderson, A.S., and Axel, R. (1981). Transforming DNA integrates into the host chromosome. Cell *23*, 29-39.

Rohan, R.M., King, D., and Frels, W.I. (1990). Direct sequencing of PCR-amplified junction fragments from tandemly repeated transgenes. Nucleic Acids Res *18*, 6089-6095.

Roman, Y., Oshige, M., Lee, Y.J., Goodwin, K., Georgiadis, M.M., Hromas, R.A., and Lee, S.H. (2007). Biochemical characterization of a SET and transposase fusion protein, Metnase: its DNA binding and DNA cleavage activity. Biochemistry *46*, 11369-11376.

Rosenberg, N., and Jolicoeur, P. (1997). Retroviral Pathogenesis.

Rubnitz, J., and Subramani, S. (1984). The minimum amount of homology required for homologous recombination in mammalian cells. Mol Cell Biol *4*, 2253-2258.

Ruff, P., Koh, K.D., Keskin, H., Pai, R.B., and Storici, F. (2014). Aptamer-guided gene targeting in yeast and human cells. Nucleic Acids Res *42*, e61.

Sauer, B., and Henderson, N. (1990). Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. The New biologist *2*, 441-449.

Scherer, S., and Davis, R.W. (1979). Replacement of chromosome segments with altered DNA sequences constructed in vitro. Proc Natl Acad Sci U S A *76*, 4951-4955.

Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. Cell *110*, 521-529.

Shammas, M.A. (2011). Repetitive sequences, genomic instability and Barrett's esophageal adenocarcinoma. Mobile genetic elements *1*, 208-212.

Shen, P., and Huang, H.V. (1986). Homologous recombination in Escherichia coli: dependence on substrate length and homology. Genetics *112*, 441-457.

Shibata, D., Peinado, M.A., Ionov, Y., Malkhosyan, S., and Perucho, M. (1994). Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. Nat Genet *6*, 273-281.

Simsek, D., Brunet, E., Wong, S.Y., Katyal, S., Gao, Y., McKinnon, P.J., Lou, J., Zhang, L., Li, J., Rebar, E.J., *et al.* (2011). DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. PLoS Genet *7*, e1002080.

Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A., and Kucherlapati, R.S. (1985). Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature *317*, 230-234.

Strom, C.E., Johansson, F., Uhlen, M., Szigyarto, C.A., Erixon, K., and Helleday, T. (2011). Poly (ADP-ribose) polymerase (PARP) is not involved in base excision repair but PARP inhibition traps a single-strand intermediate. Nucleic Acids Res *39*, 3166-3175.

Sumida, N., Nishikawa, J., Kishi, H., Amano, M., Furuya, T., Sonobe, H., and Ohyama, T. (2006). A designed curved DNA segment that is a remarkable activator of eukaryotic transcription. The FEBS journal *273*, 5691-5702.

Sung, S., Li, F., Park, Y.B., Kim, J.S., Kim, A.K., Song, O.K., Kim, J., Che, J., Lee, S.E., and Cho, Y. (2014). DNA end recognition by the Mre11 nuclease dimer: insights into resection and repair of damaged DNA. Embo J *33*, 2422-2435.

Suzuki, Y., and Chew, M.L. (2012). Role of host-encoded proteins in restriction of retroviral integration. Front Microbiol *3*, 227.

Tanaka, S., Ishii, C., Hatakeyama, S., and Inoue, H. (2010). High efficient gene targeting on the AGAMOUS gene in an ArabidopsisAtLIG4 mutant. Biochem Biophys Res Commun *396*, 289-293.

Thibodeau, S.N., Bren, G., and Schaid, D. (1993). Microsatellite Instability in Cancer of the Proximal Colon. Science *260*, 816-819.

Thomas, C.A., Paquola, A.C., and Muotri, A.R. (2012). LINE-1 retrotransposition in the nervous system. Annual review of cell and developmental biology *28*, 555-573.

Thomas, K.R., Folger, K.R., and Capecchi, M.R. (1986). High frequency targeting of genes to specific sites in the mammalian genome. Cell *44*, 419-428.

Trucco, C., Oliver, F.J., de Murcia, G., and Menissier-de Murcia, J. (1998). DNA repair defect in poly(ADP-ribose) polymerase-deficient cell lines. Nucleic Acids Res *26*, 2644-2649.

Ueda, M., Tsutsumi, N., and Kadowaki, K. (2005). Translocation of a 190-kb mitochondrial fragment into rice chromosome 12 followed by the integration of four retrotransposons. Int J Biol Sci *1*, 110-113.

Urban, Z., Helms, C., Fekete, G., Csiszar, K., Bonnet, D., Munnich, A., Donis-Keller, H., and Boyd, C.D. (1996). 7q11.23 deletions in Williams syndrome arise as a consequence of unequal meiotic crossover. Am J Hum Genet *59*, 958-962.

Vilenchik, M.M., and Knudson, A.G. (2003). Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. Proc Natl Acad Sci U S A *100*, 12871-12876.

Vodenicharov, M.D., Sallmann, F.R., Satoh, M.S., and Poirier, G.G. (2000). Base excision repair is efficient in cells lacking poly(ADP-ribose) polymerase 1. Nucleic Acids Res *28*, 3887-3896.

Vos, J.C., De Baere, I., and Plasterk, R.H. (1996). Transposase is the only nematode protein required for in vitro transposition of Tc1. Genes Dev *10*, 755-761.

Vos, J.C., and Plasterk, R.H. (1994). Tc1 transposase of Caenorhabditis elegans is an endonuclease with a bipartite DNA binding domain. Embo J *13*, 6125-6132.

Wang, G.P., Ciuffi, A., Leipzig, J., Berry, C.C., and Bushman, F.D. (2007). HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res *17*, 1186-1194.

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

Wang, Z. (2015). Genome engineering in cattle: recent technological advancements. Chromosome Res.

Waters, C.A., Strande, N.T., Wyatt, D.W., Pryor, J.M., and Ramsden, D.A. (2014). Nonhomologous end joining: a good solution for bad ends. DNA Repair (Amst) *17*, 39-51.

Weterings, E., and Chen, D.J. (2007). DNA-dependent protein kinase in nonhomologous end joining: a lock with multiple keys? The Journal of cell biology *179*, 183-186.

Wilkie, T.M., and Palmiter, R.D. (1987). Analysis of the integrant in MyK-103 transgenic mice in which males fail to transmit the integrant. Mol Cell Biol *7*, 1646-1655.

Willett-Brozick, J.E., Savul, S.A., Richey, L.E., and Baysal, B.E. (2001). Germ line insertion of mtDNA at the breakpoint junction of a reciprocal constitutional translocation. Hum Genet *109*, 216-223.

Williams, R.S., Moncalian, G., Williams, J.S., Yamada, Y., Limbo, O., Shin, D.S., Groocock, L.M., Cahill, D., Hitomi, C., Guenther, G., *et al.* (2008). Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair. Cell *135*, 97-109.

Williamson, E.A., Farrington, J., Martinez, L., Ness, S., O'Rourke, J., Lee, S.H., Nickoloff, J., and Hromas, R. (2008a). Expression levels of the human DNA repair protein metnase influence lentiviral genomic integration. Biochimie *90*, 1422-1426.

Williamson, E.A., Rasila, K.K., Corwin, L.K., Wray, J., Beck, B.D., Severns, V., Mobarak, C., Lee, S.H., Nickoloff, J.A., and Hromas, R. (2008b). The SET and transposase domain protein Metnase enhances chromosome decatenation: regulation by automethylation. Nucleic Acids Res *36*, 5822-5831.

Wray, J., Williamson, E.A., Chester, S., Farrington, J., Sterk, R., Weinstock, D.M., Jasin, M., Lee, S.H., Nickoloff, J.A., and Hromas, R. (2010). The transposase domain protein Metnase/SETMAR suppresses chromosomal translocations. Cancer Genet Cytogenet *200*, 184-190.

Wray, J., Williamson, E.A., Royce, M., Shaheen, M., Beck, B.D., Lee, S.H., Nickoloff, J.A., and Hromas, R. (2009a). Metnase mediates resistance to topoisomerase II inhibitors in breast cancer cells. Plos One *4*, e5323.

Wray, J., Williamson, E.A., Sheema, S., Lee, S.H., Libby, E., Willman, C.L., Nickoloff, J.A., and Hromas, R. (2009b). Metnase mediates chromosome decatenation in acute leukemia cells. Blood *114*, 1852-1858.

Wray, J., Williamson, E.A., Singh, S.B., Wu, Y., Cogle, C.R., Weinstock, D.M., Zhang, Y., Lee, S.H., Zhou, D., Shao, L., *et al.* (2013). PARP1 is required for chromosomal translocations. Blood *121*, 4359-4365.

Wright, D.A., Townsend, J.A., Winfrey, R.J., Jr., Irwin, P.A., Rajagopal, J., Lonosky, P.M., Hall, B.D., Jondle, M.D., and Voytas, D.F. (2005). High-frequency homologous recombination in plants mediated by zinc-finger nucleases. Plant J *44*, 693-705.

Wurtele, H., Little, K.C., and Chartrand, P. (2003). Illegitimate DNA integration in mammalian cells. Gene Ther *10*, 1791-1799.

Yang, Y.G., Cortes, U., Patnaik, S., Jasin, M., and Wang, Z.Q. (2004). Ablation of PARP-1 does not interfere with the repair of DNA double-strand breaks, but compromises the reactivation of stalled replication forks. Oncogene *23*, 3872-3882.

Ying, S., Hamdy, F.C., and Helleday, T. (2012). Mre11-dependent degradation of stalled DNA replication forks is prevented by BRCA2 and PARP1. Cancer Res *72*, 2814-2821.

Zhang, Y., and Jasin, M. (2011). An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. Nat Struct Mol Biol *18*, 80-84.

zur Hausen, H. (2009). Papillomaviruses in the causation of human cancers - a brief historical account. Virology *384*, 260-265.

CHAPTER 2

METNASE PROMOTES DNA INTEGRATION VIA CLASSICAL NON-HOMOLOGOUS END JOINING

INTRODUCTION

DNA integration is involved in many important biological processes. Integration of viral DNA into the infected host genome allows replication of the viral genome and transcription of viral genes for virus proliferation. Therefore, targeting the integration step is a promising strategy for anti-virus therapy development, including human immunodeficiency virus (HIV) (reviewed in Hazuda, 2012). Inhibition of HIV integrase, the enzyme that catalyzes integration of viral DNA, displayed good result in HIV treatment (reviewed in Gu, 2014; Li et al., 2014). Transposase, which is related to integrase, catalyzes the transposition of transposons from one locus to another. Remnants of transposable elements comprise an estimated two-thirds of typical mammalian genomes. Although most transposons are inactive, the presence of these mobile elements still influences genomic stability (Solyom and Kazazian, 2012). Less commonly, viral DNA from apoptotic infected cells can be transferred to uninfected cells via apoptotic bodies (Holmgren, 2010). Similarly, when tumor cells undergo apoptosis induced by chemotherapy or radiation therapy, DNA of dying tumor cells is fragmented and recycled by surrounding cells through apoptotic bodies. Integration and amplification of tumor oncogenes into normal cells potentially transform the recipient cells into new tumor cells (Bergsmedh et al., 2001; Ehnfors et al., 2009; Holmgren, 2010).

DNA integration also allows stable expression of a particular gene of interest into recipient cells. This is frequently used to make stable transgenic cell lines or organisms.

Integration of therapeutic genes is the preferred result of gene therapy as it provides sustained therapeutic effects. Transposon-based vectors provide an opportunity to stably integrate therapeutic genes while minimizing risk of virus-related responses (Li and Huang, 2000). Although safe delivery of gene therapy has been challenging, major advances have been made (Limberis, 2012). In the failed cases of early gene therapy attempts, integration of therapeutic DNA into a certain locus was associated with subsequent tumor development in the patient. Deep understanding of integration site selection of therapeutic vectors would have helped to avoid this kind of tragedy. For these reasons, it is important to understand the mechanisms of foreign DNA integration into eukaryotic genomes and the global impact on the recipient genomes to better utilize DNA integration in research and clinical settings.

Integration of foreign DNA can occur in a homology-directed or homology-independent manner. Homology-directed integration, often referred to as "gene targeting", involves exchange of homologous sequences between the delivery vector and the targeting site in the recipient genome, as a result, the genomic DNA of the targeting site is replaced by the vector sequence. This process relies on homologous recombination (HR) between the homologous sequence between the vector and the targeting site. Integration of foreign DNA into the yeast genome is efficiently mediated by homologous sequence between the foreign DNA and targeted integration sites. Therefore, manipulation of yeast genome using plasmids with homologous sequences is relatively easy. By contrast, targeting specific loci in a mammalian genome via homology-directed integration is very inefficient. Increasing the length of homologous sequence between the foreign DNA and the targeted integration locus can improve targeting efficiency to a certain extent (Rubnitz and Subramani, 1984). The inefficiency of homology-mediated gene targeting in

mammalian cells could reflect less efficient HR and/or more efficient homology-independent repair, such as non-homologous end joining (NHEJ).

Before the birth of the CRISPR/Cas9 system, which utilized the clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein-9 nuclease (Cas9) to conduct site-specific genome editing (Cong et al., 2013; Ran et al., 2013), zinc finger nuclease, transposase and other engineered nucleases are promising tools for locus-specific genome editing. Despite the variations of their working mechanisms, they share commonality in that cleavage at the target locus is used to direct incorporation of foreign DNA at a specific site. The close association between repair of double-stranded breaks (DSBs) and integration of foreign DNA inspired us to investigate the mechanism of DNA integration by examining functions of DSB repair proteins.

Metnase is a DSB repair protein only expressed in higher primates (Cordaux et al., 2006). Metnase promotes both NHEJ and DNA integration (Saleh-Gohari et al., 2005; Williamson et al., 2008a), linking DNA integration to DSB repair through NHEJ. Metnase is a fusion protein with multiple cellular functions. Metnase comprises a SET (Su(var)3-9, Enhancer of Zeste, Trithorax) methyltransferase domain and a transposase (nuclease) domain. The SET methyltransferase domain catalyzes methylation of histones, including H3K36 dimethylation at DSB sites, which enhances recruitment of early NHEJ proteins KU and NBS1 to the DSB, as a result, dimethylation of H3K36 at DSB site by Metnase enhances repair through NHEJ (Fnu et al., 2011). The transposase domain is a lineage of *Mariner* transposase, and shares homology with human transposon *Hsmar1* and HIV integrase (Liu et al., 2007a). The transposase domain retains *Hsmar1* terminal inverted repeat (TIR) specific binding and DNA cleavage activities, but no longer catalyzes complete transposition of transposons. Formation of Metnase foci at DSB sites

depend on the human psoralen 4 (hPso4) protein, which forms complex with Metnase and allows Metnase to interact with non-TIR DNA (Beck et al., 2008). Aside from recruiting Metnase to DSB sites, hPso4 protein engages in multiple cellular functions, including RNA splicing, repair of interstrand cross-linking DNA damage and hydroxyurea-induced stalled replication fork (Abbas et al., 2014; Mahajan and Mitchell, 2003; Zhang et al., 2005). Metnase protects plasmid DNA from large deletions during NHEJ (Hromas et al., 2008) and prevents large deletions during viral DNA integration into human cells (Williamson et al., 2008a). Metnase also interacts with the key NHEJ protein Ligase IV. This interaction has been shown to enhance the efficiency and accuracy of NHEJ (Hromas et al., 2008). Therefore, the current model for Metnase promotion of NHEJ is as follows: Metnase is recruited to DSB sites, likely through interaction with hPso4 protein (Beck et al., 2008). Metnase dimethylates H3K36 to enhance recruitment of NHEJ factors KU70 and MRN to DSBs (Fnu et al., 2011). In addition, Metnase facilitates recruitment of Ligase IV and its DNA binding partner X-ray repair cross-complementing protein 4 (XRCC4), probably through direct binding to Ligase IV (Hromas et al., 2008). Overall, Metnase facilitates holding broken ends together and speeding up the kinetics of end joining. As a result, the time that broken ends are exposed to cellular nucleases that degrade DNA is reduced; hence ends are protected from large deletions.

DNA integration is closely related to NHEJ. Metnase may promote DNA integration through a similar mechanism, by increasing the efficiency of end-joining between the integrating DNA and the recipient chromosome. We hypothesized that Metnase promotes foreign DNA integration through its functions in NHEJ. As Metnase protects ends from large resection during NHEJ, Metnase may similarly protect the ends of integrating DNA and the recipient genome during integration. DNA damage sites are slightly preferred during retroviral and non-homology

directed integration in general (Desfarges and Ciuffi, 2010; Wurtele et al., 2003). Metnase retains *Mariner* TIR-specific binding activity and DNA cleavage activity. If these activities result in sequence-specific cleavage by Metnase, increased integration into TIR sequences in *Mariner* remnants would be observed under Metnase overexpression.

In this study, we isolated integration sites of a plasmid in HEK293T cells under low or high level of Metnase expression. We found that transient overexpression of Metnase did not alter the distribution of plasmid integration sites in the human genome. To be more specific, Metnase did not promote integration into or nearby TIR sequences. Microhomologies and insertions are common features found at the junctions between plasmid and chromosome sequences. Overexpression of Metnase did not change the frequency of DNA integration via microhomology, but reduced the length of microhomology. Longer microhomology is a feature of DSB repair through alternative non-homologous end joining (aNHEJ), implying that DNA integration mediated by longer microhomologies may occur via aNHEJ. Our data suggests that Metnase promotes DNA integration through classical non-homologous end joining (cNHEJ) and suppresses DNA integration through aNHEJ.

EXPERIMENTAL PROCEDURES

Cell cultures

HEK293T cells were chosen for this study because the Metnase expression level in these cells is below detectable level via western blotting and reverse transcription polymerase chain reaction (RT-PCR) (Williamson et al., 2008b). HEK293T cells are generated by integration of the SV40 large T- antigen into HEK293 cell (Lebkowski et al., 1985), which was generated by transformation of human embryonic kidney cells with sheared adenovirus 5 DNA. Later,

HT1080 cells were used because it is a male human fibrosarcoma cell line that have realtively stable karyotype and only 1 copy of X-chromosome (Wei et al., 1998).

HEK293T cells were cultured in DMEM (Life Technologies) supplemented with 10% fetal bovine serum (Sigma) and Penicillin-Streptomycin at 100 units/ml, and incubated at 37 ℃ with 5% CO_2. HEK293 and HT1080 cells were cultured in the same conditions except growth medium was α-MEM (Life Technologies) supplemented with 10% fetal bovine serum (Atlas Biologicals).

Isolation of integration products

Puromycin resistance expressing plasmid pPUR (Figure 1A) was linearized at AlwNI site and concentrated by ethanol precipitation. Metnase overexpressing pcDNA vector (pcDNA-Metnase) as well as the empty pcDNA vector were purified from *E.coli* using a QIAGEN plasmid maxi kit. AlwNI-digested pPUR (0.033 μg) and circular pcDNA or pcDNA-Metnase (0.33 μg) were cotransfected into HEK293T cell seeded into each well of 6-Well dishes using Lipofectamine2000 (Life Technologies). Transfected cells in each well were split into three 10 cm dishes 24 h after transfection, with 10 ml DMEM. Cells were incubated for 2 h before puromycin was added to a final concentration of 1 μg/ml. Puromycin resistant colonies were detached with 0.25% trypsin and transferred to 96-well plate 11~12 days later. Isolated single colonies were expanded in growth medium containing puromycin in 10 cm dishes. Genomic DNA of puromycin resistant integration products was extracted with a QIAGEN DNeasy Blood & Tissue Kit.

Rapid Amplification of Genomic Ends (RAGE)

Genomic DNA (30 μl) of individual integration product was digested with NlaIII, and purified with QIAEX II Gel Extraction Kit (QIAGEN). Purified DNA was eluted into 30 μl of

dH_2O. Poly-cytosine (PolyC) tails were added to genomic DNA fragments using terminal deoxynucleotidyl transferase (TdT) reaction (1X terminal transferase buffer, 0.25 mM CoCl2, 0.05 mM dCTP, 10 U terminal transferase, 30 μl of fragmented genomic DNA in 50 μl reaction, incubated at 37 ℃ for 30 min, heat inactivated at 75 ℃ for 20 min). Two rounds of PCR were conducted to amplify the integration junction. In the first round PCR, 5μl of TdT reaction was used as the template, selective anchor primer (SAP) was paired with two pPUR specific primers (pPURNlaIIILeft or pPURNlaIIIRight, sequences of primers are list below) complementary to sequences approximately 300 bp flanking the AlwNI site on pPUR (Figure 1B). The first PCR was diluted 1:5 and 1 μl of the dilution was used as the template in a second round PCR. Universal amplification primer (UAP) was paired with nested pPUR specific primers (pPURNlaIIILeft2 and pPURNlaIIIRight2, sequences of primers are list below) in a second PCR (Figure 1B). Products from the second round PCR were analyzed on 0.7% agarose gels. A DNA band is selected if the band only appeared in integration product but not in non-transfected parent HEK293T. Selected bands were extracted with Corning Costar Spin-X plastic centrifuge tube filter and concentrated with DNA clean & concentrator-5 (Zymo Research) for sequencing or using QIAEX II Gel Extraction Kit and prepared for sequencing using pPURNlaIIILeft2 or pPURNlaIIIRight2. PCR primers and conditions are listed as follows:

SAP: 5'- GGAGACTGACATGGACTGAAGGAGTAAAGGGIIGGGIIGGGIIGGGCATG -3'

UAP: 5'- GGAGACTGACATGGACTGAAGGAGT -3' (Liu and Baird, 2001)

pPURNlaIIILeft： 5'- TTTCGTTCCACTGAGCGTCAGACC -3'

pPURNlaIIIRight： 5'- ACGCTCAAGTCAGAGGTGGCGAAA -3

pPURNlaIIILeft2： 5'- TTTCTGCGCGTAATCTGCTGCTTG -3'

pPURNlaIIIRight2： 5'- CCCGACAGGACTATAAAGATACCAGGCGT -3

First PCR: 95 ℃ for 1 min; 20 cycles of 95 ℃ for 30 s, 70 ℃ (decrease 0.3 ℃ for each cycle) for 30 s, 72 ℃ for 1 min; 20 cycles of 95 ℃ for 30 s, 63 ℃ for 30 s, 72 ℃ for 1 min; 72 ℃ for 10 min. Second PCR: 95 ℃ for 1 min, 35 cycles of 95 ℃ for 30 s, 63 ℃ for 30 s, 72 ℃ for 1 min; 72 ℃ for 10 min.

Identification of genomic integration sites by Basic Local Alignment Search Tool (BLAST)

Sequencing results were compared to the pPUR plasmid sequence using the BLAST program of the National Center for Biotechnology Information (NCBI) to identify the plasmid component in the junction. The rest of the sequence (including the last 15 bases of the plasmid component) was then BLAST to the reference human genome to locate the genomic component. In cases of multiple hits, the hit with the lowest expect threshold was chosen.

Karyotyping by G-banding

HEK293T cells were cultured in DMEM medium to 70% confluent. Colcemid solution (Life Technologies) was added to a final concentration of 0.05 μg/ml for 30 min. Cells were then harvested into the 15 ml conical tube, centrifuged at 200 xg for 10 min, supernatant was removed and cell pellet was resuspended in 1 ml of 0.075M KCl. Cells are fixed by adding 1 ml of Carnoy's Fixative (3:1 ratio of methanol : glacial acetic acid) hypotonic solution and incubate for 30 min at room temperature. Cells were then centrifuged at 300 xg for 5 min and supernatant was removed and cell pellet was resuspended in 300-500 μl of fixative. To prepare metaphase spread, 25-35 μl of cell suspension was evenly distributed onto a clean glass slide placed above the hot water bath (60-70 ℃) for 10 s then face the slide down in the steam from the hot water for 1~3 s, and dry on a hot metal plate. Slides were incubated in a 60 ℃ oven for 16 h, then treated with Trypsin from bovine pancrease Type I (Sigma) for 5 s and Giemsa staining for 5 min. A total of 20 metaphase spreads were analyzed for karyotype. Karyotype analysis was done in accordance

with recommendations of the International System for Human Cytogenetic Nomenclature (ISCN 2013).

RESULTS

To determine if Metnase stimulates DNA integration into TIR sequences of the *Mariner* remnants in the human genome, and if Metnase protects the integration DNA and recipient genome from large deletions, we mapped integration sites of a linear plasmid in HEK293T cells. AlwNI-digested pPUR plasmid was co-transfected with circular pcDNA or pcDNA-Metnase into HEK293T cells and puromycin-resistant colonies were scored. Metnase overexpression induced by pcDNA-Metnase transfection has been verified via western blotting and RT-PCR (Lee et al., 2005). Consistent with previous findings (Lee et al., 2005; Williamson et al., 2008a), cotransfection of Metnase expression vector pcDNA-Metnase led to 7-fold increase of pPUR integration compared to empty pcDNA vector (Figure 2).

Single puromycin resistant colonies were isolated as integration products. Integration junctions were amplified and sequences using RAGE. For convenience, integration products from pPUR/pcDNA co-transfection are referred to as PCD products and integration products from pPUR/pcDNA-Metnase co-transfection are referred to as MET products. A total of 105 PCD products and 108 MET products were analyzed using RAGE, and integration junctions were successfully obtained from 35 PCD products and 31 MET products. This is primarily due to the technical limitations of RAGE. For example, non-specific amplification of genomic fragment outcompetes amplification of the integration junction; integrated plasmid may be resected beyond the primer sites; plasmid concatemers were common despite we reduced the amount of pPUR DNA to 0.033 μg per well, amplification of plasmid-plasmid junctions sometimes outcompetes amplification of plasmid-chromosome junctions. To locate the integration site on

the human genome, the DNA sequences were first compared to pPUR by BLAST. Sequences not related to pPUR were then analyzed by BLAST to the reference human genome. Junctions that returned unique BLAST results to the human genome are shown in Figure 9, and the results are summarized in Figure 4-8.

Each integrated pPUR plasmid produces two plasmid-chromosome junctions, referred to as left or right junction respective to the AlwNI cut site. In about 64% of the products, only one junction was obtained. In the rest, both left and right junctions were identified. However, the two junctions did not always map to the same chromosome. If the two junctions mapped on two different chromosomes, they were classified as separately integrated plasmids. If the two junctions mapped on the same chromosome, the start and end nucleotides of the two junctions were lined up to determine whether they come from the same integrated plasmid (Figure 3). If the two junctions belonged to the same plasmid, deletion of the genomic DNA was calculated. The products (35 PCD products and 31 MET products) were sorted according the number and type of junctions obtained and shown in Table 1. Together, 47 PCD junctions (21 left junctions and 26 right junctions) and 43 MET junctions (16 left junctions and 28 right junctions) were included for further analysis (Figure 9). The average length of sequenced junctions was 416 base pairs. If the left end of pPUR was involved, the sequenced junction typically contained 150-180 bp of pPUR sequence; this number was typically 200-250 bp if the right end was involved (Figure 1B). The length of genomic DNA amplified ranges from 17 to 448 base pairs. Analysis of these 90 plasmid-genome junctions provided information about potential integration “hot spots”, deletions on plasmid ends and the host genome, microhomologies and insertions (non-templated or captured) between plasmid and the integration sites.

Metnase overexpression did not alter distribution of plasmid integration sites

The distribution of integrated plasmid in HEK293T cells did not appear to be random. A total of 46 integrated plasmids were scored from PCD products and 41 from MET products. In both sets, a high frequency of X-chromosome integration was observed: 37.0% of control PCD products and 34.1% of MET products integrated in the X-chromosome (Figure 4A). According to the current human genome assembly, the X-chromosome accounts for 5% of the human genome. Integration into X-chromosome is significantly higher (p-value<0.0001) than the calculated frequency assuming random integration. No significant difference was observed between with and without Metnase overexpression. This indicates that Metnase does not influence the chromosomal distribution of plasmid integration.

No homology was found between pPUR and any human chromosome, including the X-chromosome. Besides, integration sites on X-chromosome distributed sparsely along the chromosome without displaying any preferred region. Therefore, the high frequency of X-chromosome integration should not be due to HR between pPUR and X-chromosome. Possible biases caused by PCR primers are also eliminated, as none of the RAGE primers had significant alignment with these X-chromosome. Using the same method, a small number of pPUR integration junctions were isolated from HEK293 cells and HT1080 cells. A total of 10 junctions were isolated from HEK293 cells and HT1080. If the high frequency (35.6% on average) of X-chromosome integration is also true in these two cell lines, 3 or 4 X-chromosome integration was expected. But none of their integration sites were on the X- chromosome. The high frequency of X-chromosome integration in the HEK293T data set may be specific to this cell line.

We recognized the fact that HEK293T cells do not have a normal karyotype. Transforming kidney embryonic cells with sheared adenoviral DNA generated HEK293 cells.

HEK293T cells were generated by expressing SV40 large T-antigen in HEK293 cells (Lebkowski et al., 1985). HEK293 cells are hypotriploid and generally contain 3 copies of X-chromosome, the expression of T-antigen allows more than one round of replication in each cell cycle, very likely to cause more aneuploidy (Friedrich et al., 1992). To test if the high frequency of integration into X-chromosome is because 293T cells gain extra copies of X-chromosomes, we then karyotyped the HEK293T cells by G-banding. According the recommended standard by the International System for Human Cytogenetic Nomenclature, A total of 20 metaphase spreads were analyzed. The karyotype is summarized as 65~70, XX, add(X)(q28), der(1)t(1;15)(q44;q21)x2, +der(1)t(1;?)(p12;?), +2, del(3)(q21), -4, der(4)t(4;?)(q33;?), +der(5)t(5;?), -8, -9, i(9)(q10), -13, der(13)t(13;?)(q34;?), -15, del(17)(p12), -18, +21, -22, +3~6mar [cp20]. Typically, HEK293T cell are also hypotriploid, with number of chromosomes ranging from 65 to 70. All hypotriploid HEK293T cells have 3 copies of X-chromosome, and several derivative chromosomes (Figure 4B). Besides, HEK293T cells typically have 4 copies of chromosome 1, 2, 5, and 21. HEK293T cells are generated from female human embryonic kidney cells, which normally have 2 copies of X-chromosome. Since the ratio of X-chromosome in HEK293T is not higher than normal diploid, the high frequency of X-chromosome integration should not related to copies numbers of X-chromosome.

It is estimated that 25~30% of the human genome comprises protein-coding genes (Strachan and Read, 1999), while only less than 2% of the human genome is coding sequences (exons) (Lander et al., 2001). Transcription induces local decondensation of chromosomes and this may enhance access for foreign DNA and proteins contributing to DNA integration. Integration of plasmid DNA into mouse embryonic stem cells showed preference into genes (Suzuki et al., 2010). In our dataset, 52.2% (24 of 46) of integrated plasmid in PCD products and

65.1% (27 of 42) of integrated plasmid in MET products were inside gene sequences. The percentage of integration into gene sequences is slightly higher in MET products, but difference is not statistically significant (Fisher's exact test, P=0.28).

Metnase promotes DNA integration independently of its TIR binding activity

The nuclease domain of Metnase derived from *Mariner* transposase (Cordaux et al., 2006). Metnase cannot catalyze full transposition, but it retains specific binding to TIR elements of human transposon *Hsmar1* (Roman et al., 2007). To examine whether Metnase-TIR interaction promotes DNA integration into TIR sequences or related sequences in the ~7,000 *Hsmar1* remnants in the human genome (Liu et al., 2007b), we analyzed 50,000 bp on each side of each integration junction by BLAST for sequences related to full length TIR (TATTAGGTTGGTGCAAAAGTAATTGCGGTTTT), as well as the 19-mer Metnase TIR binding site (GGTGCAAAAGNNNTTGCGG). None of the PCD products displayed significant homology to TIR; only in one MET product the integration site was near TIR sequence. And interestingly, that integration site was in the *Metnase* gene. In another two MET products, integration into the *Metnase* gene was also observed but later verified to be the sequence of pcDNA-Metnase (which contains the cDNA sequence of Metnase) rather than the genomic copy of *Metnase* gene. These two junctions were excluded from the dataset. We cannot determine if the one MET product kept in the dataset also represents integrated pcDNA-Metnase, so it is hard to conclude this was an event where Metnase stimulated pPUR integration into the *Metnase* gene or TIR-related sequence. Another feature of transposition mediated by the *Hsmar1*transposase, which is the ancestor of Metnase-specific transposase, is that the transposon integrates into either side of TA dinucleotide sequences (Liu et al., 2007b). However, preference of integration next to TA dinucleotide was not observed in with overexpression of Metnase (Figure 4C). We conclude

that TIR-specific binding activity of Metnase does not direct plasmid integration into TIR-related sequences in the human genome. Therefore, sites of DNA integration remain random with overexpression of Metnase.

Metnase did not protect plasmid ends from large deletions

Previous work demonstrated that overexpression of Metnase protects lentiviral DNA from large deletions during integration into mammalian genome (Williamson et al., 2008a). This was attributed to its ability to promote NHEJ and therefore limit end processing. In the present PCD and MET integration product sets, more than half of the integrated plasmids exhibited deletions (Figure 5B). The maximum deletion was 288 bp among PCD products, and 192 bp among MET products. However, the average length of plasmid deletions was not significantly different between PCD and MET products (Figure 5A) and the distribution of deletion sizes were similar between PCD and MET products (Figure 5B). It is not clear why results with viral and plasmid DNA integration differ (see Discussion).

The left and right integration junctions in each integration products were amplified in separate PCR reactions and to determine if they belong to the same integrated plasmid (Figure 3). Genomic deletion at integration was calculated by subtracting the positions of the first base of its two junctions (Figure 3A). Due to technical limitations, we were only able to uncover both left and right junctions from three plasmids, one from PCD products, two from MET products. The one plasmid in PCD product caused a 16 bp deletion at the integration site. The two plasmids in MET products caused 28 bp and 4,564 bp deletions. With this limited data set, definitive conclusions cannot be drawn about whether Metnase affects deletions of genomic DNA during plasmid integration. In addition, karyotyping of HEK293T cells suggested that chromosomal translocations are common in these cells. Therefore, it will be helpful to further

explore this question using a cell line with a defined and more stable karyotype; or use an integration site mapping method that captures both junctions of the same plasmid without separation, for example, by digesting genomic DNA containing integrated plasmid DNA with restriction enzymes such that the plasmid origin of replication and selectable marker can be recovered along with segments of flanking genomic DNA on either side of the integrated plasmid, and amplifying these fragments by recircularizing and rescuing these new plasmids via transformation of *E.coli*, and then sequencing the plasmid-genomic DNA junctions.

Metnase suppresses use of microhomology during plasmid integration

To analyze the features of integration junctions, we defined junctions as one of three types. Tight junctions were those with precise joining of plasmid end to the chromosome; insertions had additional sequences between the plasmid and chromosome; and microhomology refers to short sequences shared between the plasmid and chromosome at the integration junction (identified by BLAST analysis) (Figure 6A). Insertions were the most common. The next most abundant type was microhomology-mediated junctions. Tight junctions only comprised less than 10% of mapped junctions (Figure 6B). These distributions were approximately the same among the PCD and MET products, indicating that Metnase does not significantly alter the ratio of tight, microhomology, and insertion junctions (Figure 6B).

Microhomology is a feature of DSB repair through aNHEJ. Microhomology-mediated DNA integration may be related mechanistically to the aNHEJ pathway. Metnase promotes cNHEJ and suppresses aNHEJ (Fnu et al., 2011; Wray et al., 2010). For the purpose of our project, we define microhomology as 1-25 nucleotides. In our dataset, although the fractions of products exhibiting microhomology at integration junctions were similar (Figure 6B), MET products displayed significantly smaller microhomologies, with the majority having only 1 base

of microhomology (Figure 7), and these junctions with this minimal level of microhomology may arise by cNHEJ rather than aNHEJ (see Discussion).

Insertions can arise by three distinct mechanisms (Figure 8C). First, if ends are resected (either in the plasmid or at DSBs in chromosomes into which plasmids integrate), ends may align imperfectly via short microhomologies, leaving gaps to be filled; and gap filling can create insertions (Lee and Lee, 2007; Ma et al., 2003; Simsek et al., 2011; Yu and McVey, 2010). Second, insertions can arise by sequence capture, for example, fragments of plasmid or other DNA may join to plasmid or chromosomal DNA prior to the final joining of plasmid and chromosomal DNA (Lin and Waldman, 2001; Little and Chartrand, 2004). The third mechanism is via non-templated addition of bases to plasmid or chromosomal ends by enzymes like TdT (Cabaniols et al., 2001; Landau et al., 1987; Purugganan et al., 2001). In our dataset, the length of insertions varied from 1 base to 91 bases. Overall, no significant difference was observed between PCD and MET products regarding the average length of insertions (Figure 8A). The longest insertion in PCD products was 91 bases and the longest insertion in MET products was 56 bases. Suspecting that these long insertions may result from DNA fragment capturing at the integration junctions, we then BLAST these two insertions against the National Center for Biotechnology Information (NCBI) Nucleotide collection. The 91 bases long insertion could not be aligned to any known sequence in the NCBI nucleotide collection database. But the 56 bases long insertion was BLAST to *Bos mutus* (Yak) SMAD3 mRNA, but not any target in human genome. Besides, one of the insertions from PCD product came from capture of pcDNA sequence. Since most insertions were shorter than 32 bases, these junctions were plotted separated and fitted with linear line (Figure 8B). Both fitting from PCD and MET products showed that frequency of appearance decreased as the length of insertion increased. But the

slopes of the two fitted lines were not significantly different (P= 0.2862). In conclusion, Metnase overexpression did not have significant effect on the size or composition of insertions.

DISCUSSION

Metnase promotes plasmid integration independent of genomic sequence

Illegitimate DNA integration in mammalian cells constantly challenges the integrity and stability of the genome. DNA repair systems play important roles in maintaining genomic stability and preventing cancer development. It has been shown that both exogenous and genomic DNA fragments can be captured and patched at DSB sites during repair (Little and Chartrand, 2004). Metnase is a multi-functional fusion protein that promotes both DNA integration and DSB repair through NHEJ. To investigate the role of Metnase in DNA integration, we collected integration products of a plasmid in HEK293T cells with low and high levels of Metnase expression and mapped the integration sites. In our dataset, pPUR integration significantly preferred the X-chromosome (p-value<0.0001). Generation of HEK293T cells included two steps. Transforming embryonic kidney cells with sheared adenoviral DNA generated HEK293 cells. Then transforming HEK293 cells by SV40 large T- antigen generated HEK293T cells. (Lebkowski et al., 1985). Both transformation steps could introduce aneuploidy to the cell (Friedrich et al., 1992). Karyotyping of HEK293T cells showed that HEK293T cells are typically hypotriploid with 3 copies of the X-chromosome. The ratio of X-chromosome in HEK293T cells is not higher than its ancestor human embryonic kidney cells, which are female cells. Therefore, the high frequency of the X-chromosome integration was not due to abnormal copy number. Besides, we did not see any biased plasmid integration into X-chromosome in our HEK293 integration products. The high frequency of X-chromosome integration may be specific to HEK293T cells.

Previous work in mouse embryonic stem cells showed preference of illegitimate DNA integration into genes (Suzuki et al., 2010). Our data displayed the same pattern. The high frequency of integration into genes suggests a connection between plasmid integration and active transcription, where chromatin structure is "loosened", allowing better access of transcription factors. This is consistent with previous studies showing that transcription units are favored by foreign DNA integration (Alonso et al., 2003; Devine and Boeke, 1996), although some evidence argues that accumulation of transcription proteins hinders integration of retroviral DNA (Maxfield et al., 2005). Therefore, more work is needed to explain the preference of plasmid integration into genes. In our data, there was no significant difference between PCD and MET products in terms of integration into genes, consistently supporting the conclusion that Metnase enhances plasmid integration without site or sequence specificity.

In vitro, the nuclease domain of Metnase displays strong binding specificity toward TIR sequence. We predicted an increased ratio of DNA integration at *Mariner* remnants in the human genome with Metnase overexpression, mimicking the transposition process. However, search for TIR sequences nearby integration junctions did not yield significant alignment. The TIR specific binding of Metnase may be tempered due to interactions with other proteins. For example, Beck *et al.* (Beck et al., 2010) showed that hPso4 protein forms stable complex with Metnase, together, the Metnase-hPso4 complex binds to DNA in a sequence-independent manner, probably reflecting the non-selective binding activity of hPso4. Further studies using chromatin immunoprecipitation may reveal Metnase binding sites that could be compared with the integration sites described herein.

Furthermore, the transposase domain of Metnase is derived from human *Hsmar1* transposase, which belongs to the *Mariner* transposase family. Using an *in vitro* transposition

assay, it has been shown that Metnase transposase domain can mediate integration of a precleaved transposon, and the integration sites were flanked by TA dinucleotide, a shared feature of the *Mariner* family transposases (Liu et al., 2007b). In our data, overexpression of Metnase did not increase integration into TA dinucleotide (Figure 4C). This further supports the idea that Metnase promotes DNA integration independently of its TIR binding activity.

It is likely that DNA integration occurs at endogenous DSB sites. Endogenous DNA damage occurs due to attack of the DNA backbone by DNA damaging agents generated during normal cellular metabolism, such as reactive oxygen species, lipid peroxides, reactive chemicals and others, or by DNA instability such as depurination and deamination, or as an intermediate during repair of base or nucleotide damages, as part of the DNA backbone excised prior to repair of the damaged site (reviewed in De Bont and van Larebeke, 2004). Single-strand DNA damage may be converted to DSBs if two single-strand breaks (SSBs) happen to occur near each other on opposite strands, or SSBs block progression of replication forks leading to collapse of stalled forks. It was estimated that about 50 endogenous DSBs are generated from conversion of SSBs during each cell per cycle in human cells (Vilenchik and Knudson, 2003). Integration of foreign DNA could be coordinated with repair of these endogenous DSBs. It has been shown that the transposase domain is important for the activity of Metnase in promoting NHEJ by trimming the single-strand overhangs from DSBs, as blunt ends are more suitable substrates for Ligase IV (Beck et al., 2011; Lee et al., 2005). At present, there is no direct evidence that DNA nicking and cleavage activity of Metnase creates damage on chromosomes to promote DNA integration. Despite its TIR-specific binding, cleavage of DNA by Metnase-specific transposase is sequence-independent (Roman et al., 2007). Therefore, Metnase appears to promote DNA integration by its ability to enhance overall efficiency of DSB repair.

Metnase promotes microhomology-independent DNA integration

Deletions of DNA ends during integration are common. Previous work showed that Metnase reduced end deletions during NHEJ and viral DNA integration (Hromas et al., 2008; Williamson et al., 2008a). We expected that Metnase overexpression would prevent large deletions of plasmid DNA during plasmid integration but this was not observed. Deletions over 100 bp were found at low frequency among integration products both with and without Metnase overexpression, and the deletion length was not affected by Metnase overexpression (Figure 5). The difference in deletion outcome between the prior viral DNA integration experiments (Williamson et al., 2008a) and plasmid DNA integration in the present study may reflect any of several differences in the experimental systems, including mode of DNA entry (infection vs. lipofection), cell type (HEK293 vs HEK293T), and stable vs transient Metnase overexpression. With lipofection, plasmid DNA is incorporated into liposomes that traverse plasma and nuclear membranes, and the DNA may be damaged and/or processed by lysosomal or other cellular factors prior to its interaction with chromosomal DNA; in a co-transfection system as used here, such processing may occur before Metnase is expressed, hence if the goal is to use Metnase to protect plasmids from deletions during integration, it is probably best to express Metnase prior to plasmid transfection, preferably via stable overexpression.

Microhomology is commonly found at illegitimate integration junctions (Suzuki et al., 2010; Yan et al., 2010). Microhomology was found in more than 30% of the integration junctions in the present study. When Metnase was overexpressed, the average length of microhomology was reduced (Figure 7). End joining through Ligase IV-mediated cNHEJ does not require microhomology, but Ligase III-mediated aNHEJ requires end processing to expose microhomology (Pannunzio et al., 2014). DNA integration may occur by mechanisms that

employ cNHEJ and aNHEJ proteins. Metnase has been shown to promote end joining through cNHEJ and suppresses aNHEJ-mediated chromosomal translocations (Hromas et al., 2008; Wray et al., 2010). The shorter microhomologies (1-3 bases) found in integration products when Metnase was overexpressed in consistent with the function of Metnase in promoting cNHEJ. Ligase III catalyzes microhomology-dependent chromosomal translocation (Simsek et al., 2011). It is possible that Ligase III also contributes to microhomology-dependent DNA integration, and if this is the case, Metnase overexpression may shift events toward the cNHEJ pathway, thus reducing the dependence on microhomology.

The last feature of integration junctions examined in this project was non-templated insertions between plasmid and chromosomes. Junctions with insertions were the most common outcome in our dataset (Figure 6B). Most insertions were around 30 bases or shorter. There was a trend that Metnase overexpression resulted in shift towards shorter insertions. About 77.2% of insertions with Metnase overexpression were no longer than 10 bases, compared to 56.7% without Metnase overexpression. However, this difference was not statistically significant, but with a larger dataset this trend may become clearer. Most of the insertions appeared to be non-templated, but one insertion appeared to come from the *Bos mutus* (Yak) genome, which may have occurred from contamination, for example of the fetal bovine serum supplement in the cell culture medium.

Overall, data collected from HEK293T cells partially support our hypothesis that Metnase promotes plasmid integration by promoting NHEJ. Overexpression of Metnase decreased length of microhomology between integrated plasmid and the integration site. Extensive deletions and usage of microhomology are features of DSB repair through aNHEJ. While Metnase mainly functions in cNHEJ, our data are consistent with overexpression of

Metnase promoting cNHEJ-mediated DNA integration, therefore reducing microhomology usage during integration. Our effort to examine the involvement of aNHEJ in DNA integration will be discussed in Chapter 3. We predicted that overexpression of Metnase will protect plasmid end and genomic DNA from large deletions during plasmid integration. However, our data showed that large deletions were not avoided in MET products. Some of the deletions may be a result of liposome-mediated transfection. Plasmid DNA may be damaged during this process. Therefore, it is possible that large deletions occurred before the plasmid entered the nucleus when Metnase could exert its effects. On the other hand, the genomic DNA is not subject to enzymatic digestions therefore deletions at the integration sites should only due to insertion of the plasmid. So far, we have not collect sufficient data to draw confirmative conclusions regarding if Metnase protects genomic DNA from large deletions during plasmid integration. A different method such as plasmid rescue could improve the frequency of capturing paired integration junctions.

Table 1. Overview of integration products

Type of integration	PCD products		MET products	
	# of products	%	# of products	%
1 end of 1 plasmid	23	65.7	19	61.3
1 end of 2 plasmids on different Chr.	7	20	5	16.1
1 end of 2 plasmids on the same Chr.	4	11.4	5	16.1
2 ends of 1 plasmid	1	2.9	2	6.5
Total	35	100	31	100

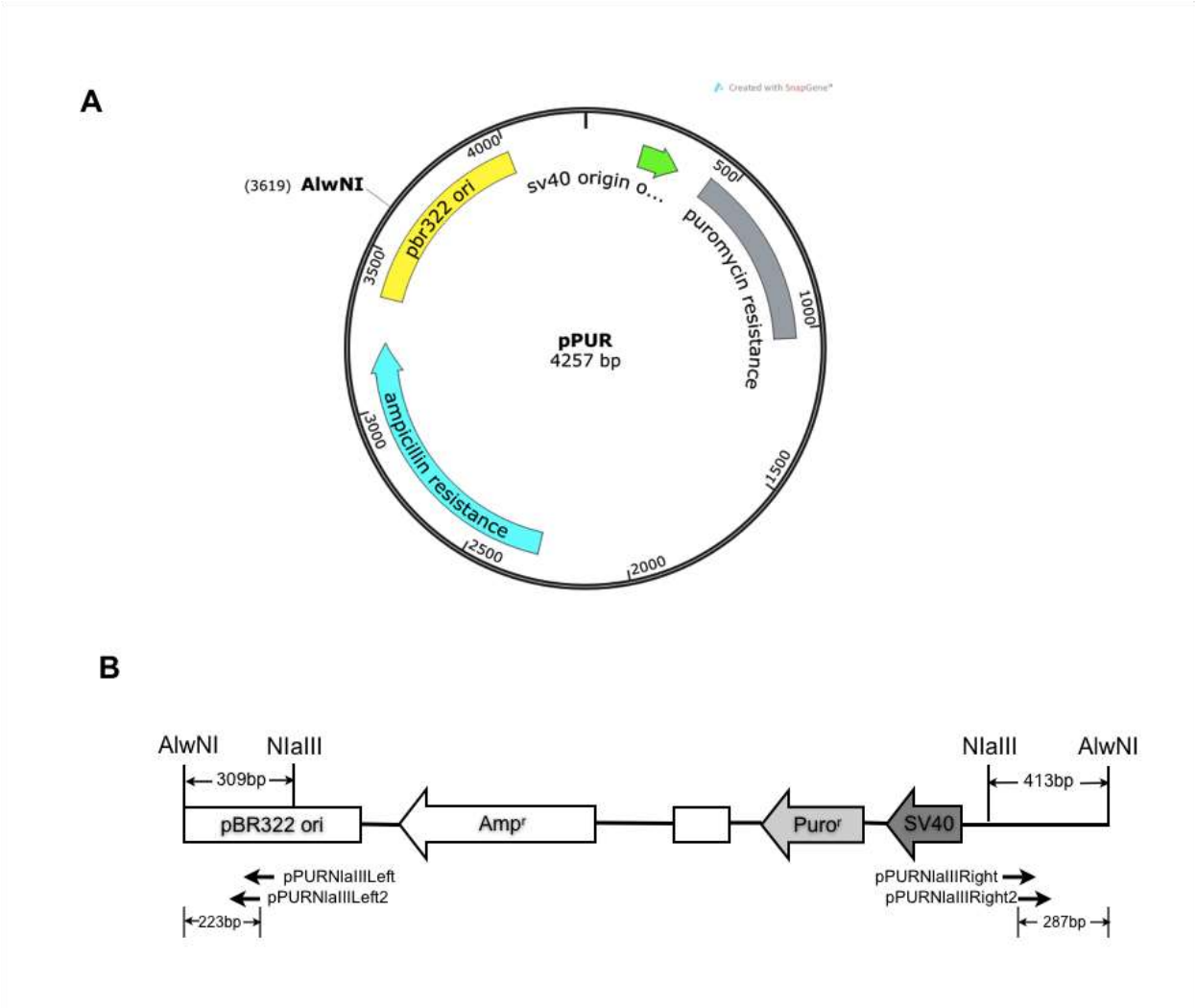

Figure 1. Maps of pPUR plasmid
(A) Structure of circular pPUR plasmid. The AlwNI site indicates where the plasmid was linearized before transfection.
(B) Schematic diagram of linear pPUR plasmid. Important features of pPUR plasmid are laid out as boxes or arrow shaped boxes, the arrows indicate the direction of the ORFs on the plasmid. Upstream side of AlwNI site is defined as the left end and downstream side is defined as the right side. Plasmid specific primers and the distances from the start of second set specific primers to the AlwNI cut site are indicated at the bottom.

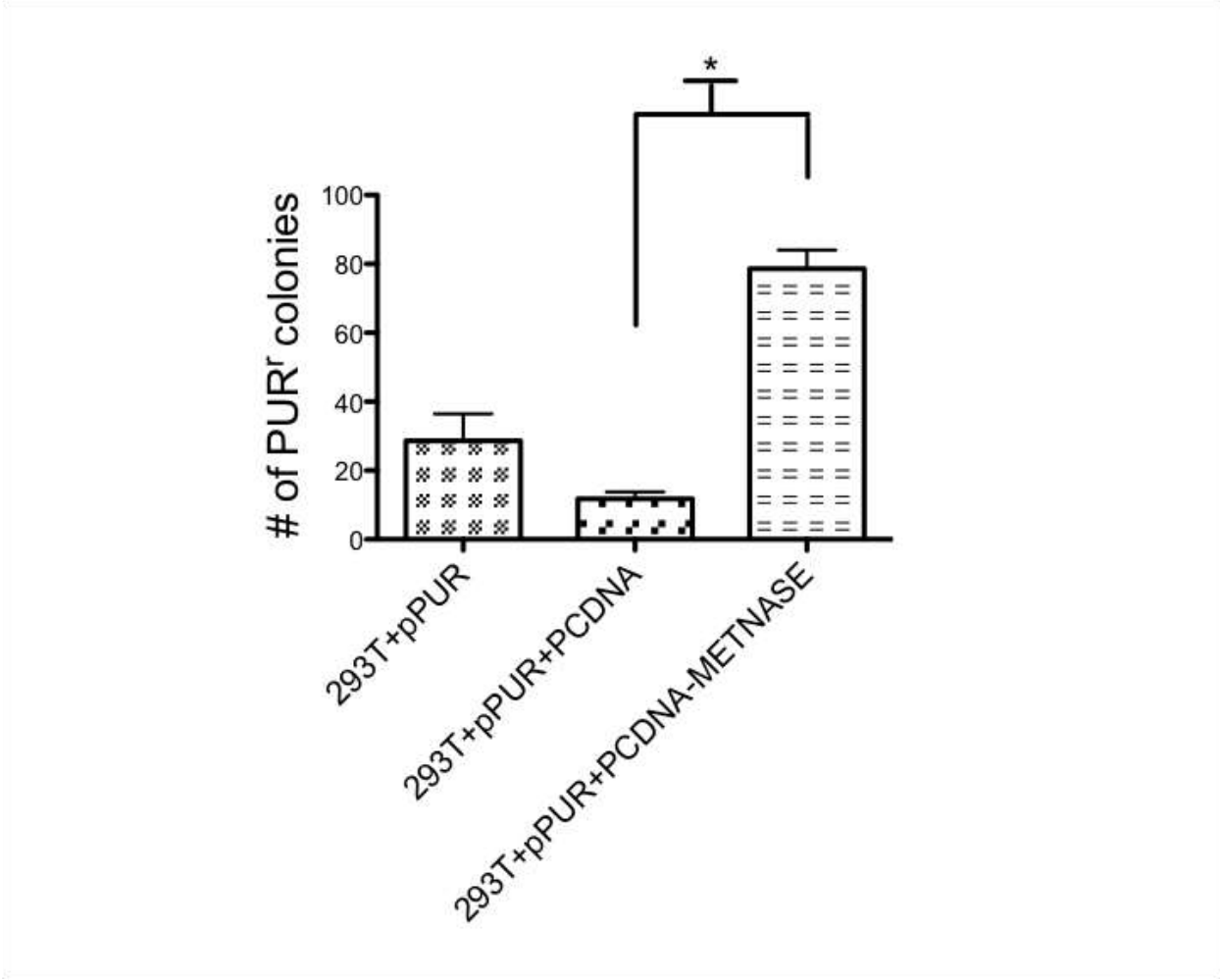

Figure 2. Metnase promotes integration of plasmid
Same number of HEK293T cells were transfected with linear pPUR, linear pPUR and circular pcDNA, or linear pPUR and circular pcDNA-Metnase. Integration efficiency of pPUR was measured as number of puromycin resistant colonies. * P=0.0007

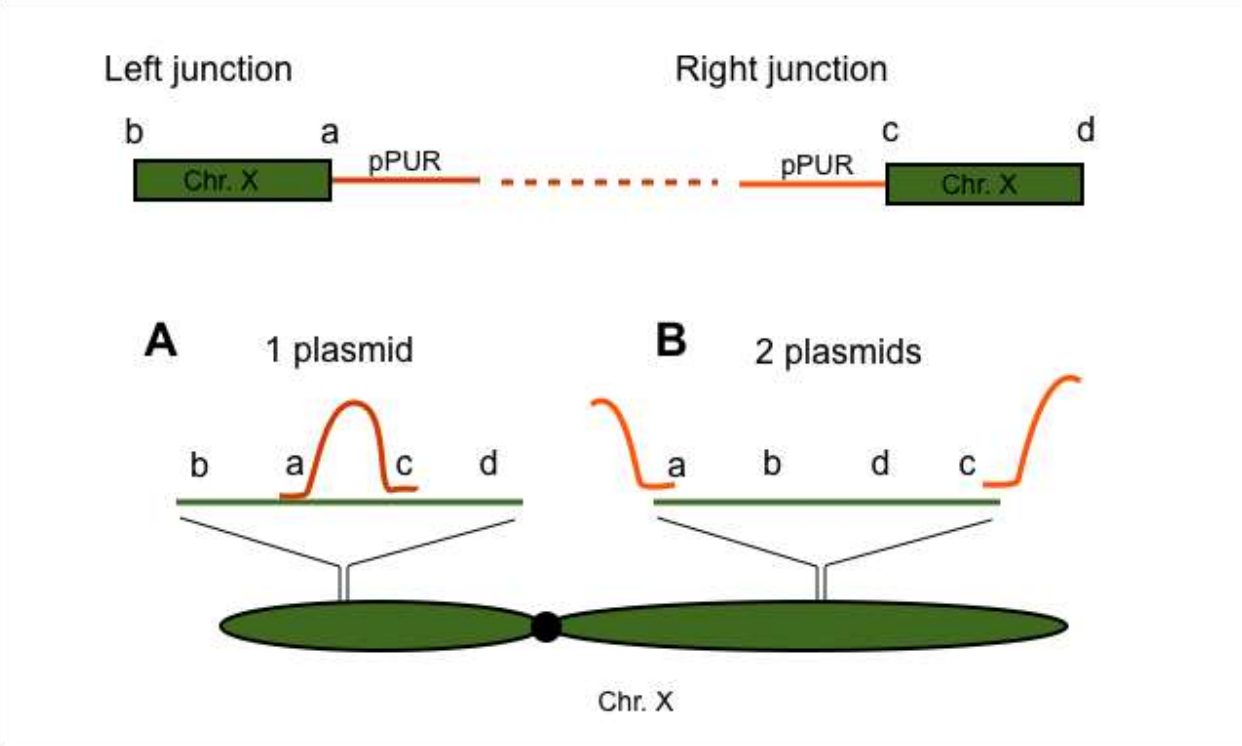

Figure 3. Standard to determine whether two junctions on the same chromosome (eg. Chr. X) belong to the same plasmid

Sequence of pPUR is shown in orange line, genomic DNA is shown in green rectangle boxes. Lower case letter a, b, c, d represents the position of the start and end nucleotide of genomic DNA in the reference human genome. If the genomic DNA of the two junctions line up on the chromosome as in (A), then they are the two junctions from one single integrated plasmid, number of base deletion on genomic DNA can be calculated by subtracting a from c; If they line up as in (B), then they are from two plasmids.

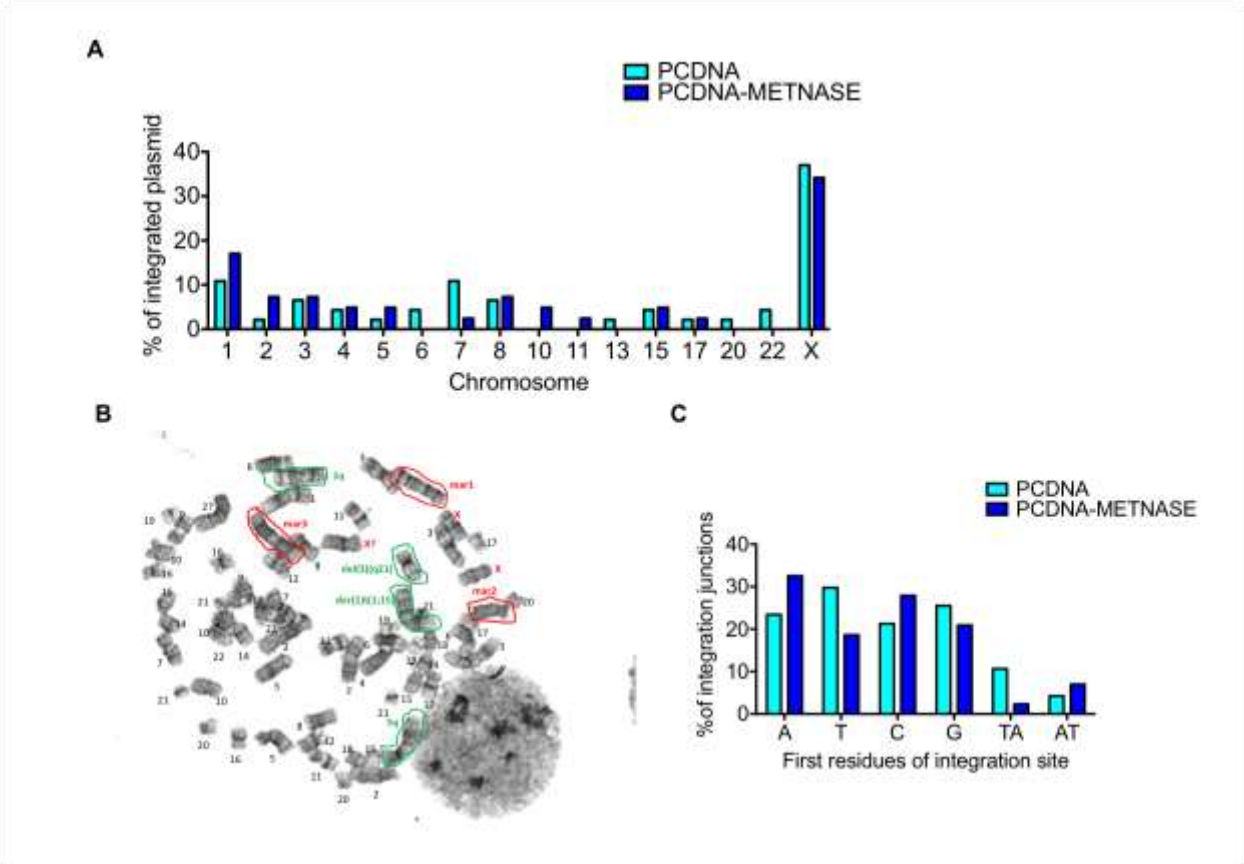

Figure 4. Landscape of integration sites

(A) Percentage of integration junctions found on each chromosome was plotted for transfection without Metnase expression (cyan bars) and with Metnase expression (blue bars).

(B) Karyotype of HEK293T cell by G-banding. In this cell, the karyotype is 65<3n>, XXX, der(1)t(1;15)(q44;q21), +der(1)t(1;?)(p12;?), +2,del(3)(q21), -4, -4, der(5)t(5;?), -7, -8, -13, -13, -17, -18, +21, -22, +3mar. A total of 7 derivative chromosomes exist in this cell, 4 of them can be identified (marked with green) and the other 3 are marker chromosomes (marked with red) whose components cannot be identified.

(C) Percentage of the first residue of integration sites as A, T, C, G was plotted for transfection without Metnase expression (cyan bar) and with Metnase expression (blue bar).

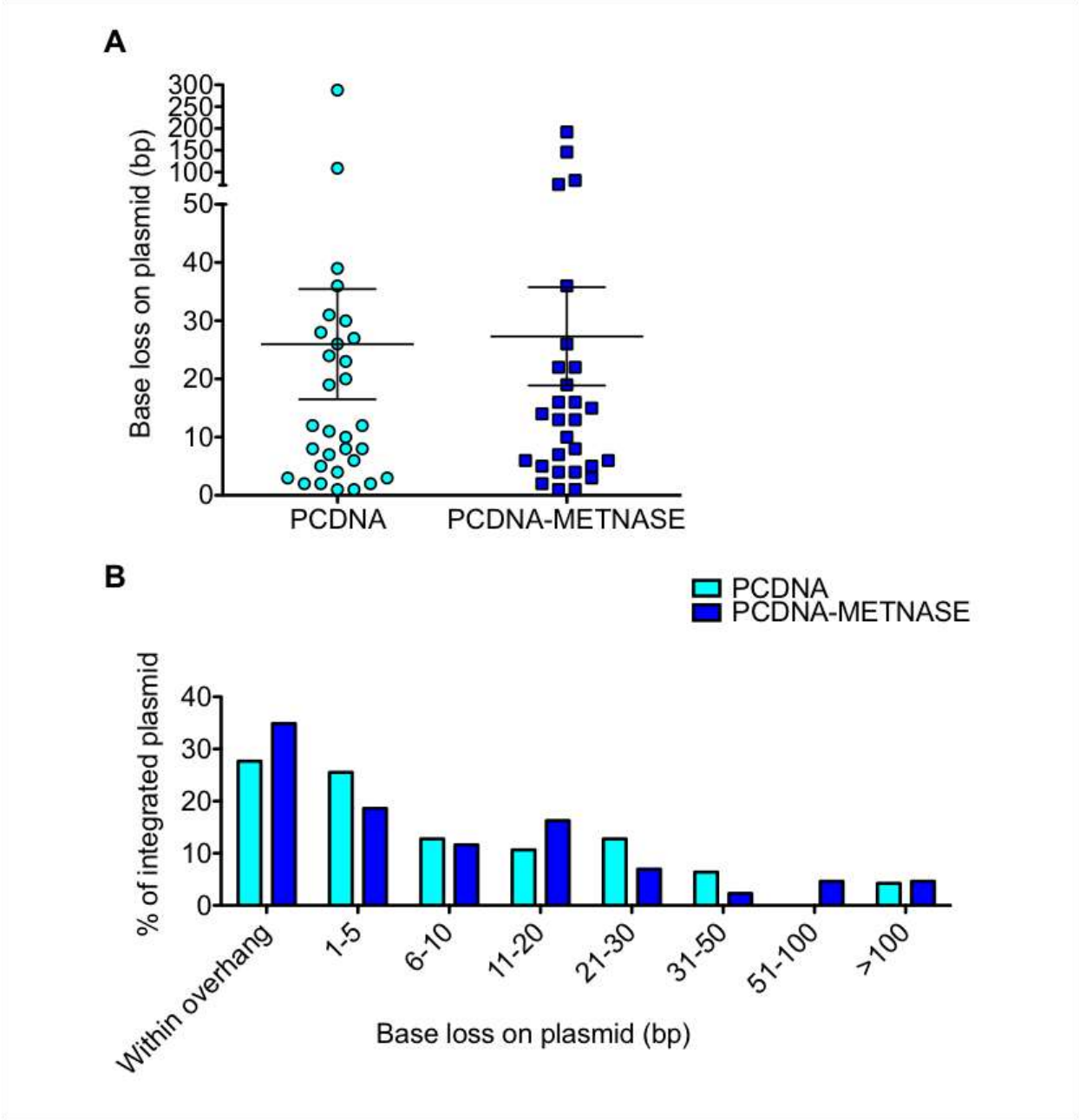

Figure 5. Modifications of plasmid ends

(A) Base loss at the end of pPUR plasmid was plotted. P= 0.3057, t-test.

(B) The size of deletions of the end of the plasmid was sorted into bins, and the percentage in each deletion size range was plotted.

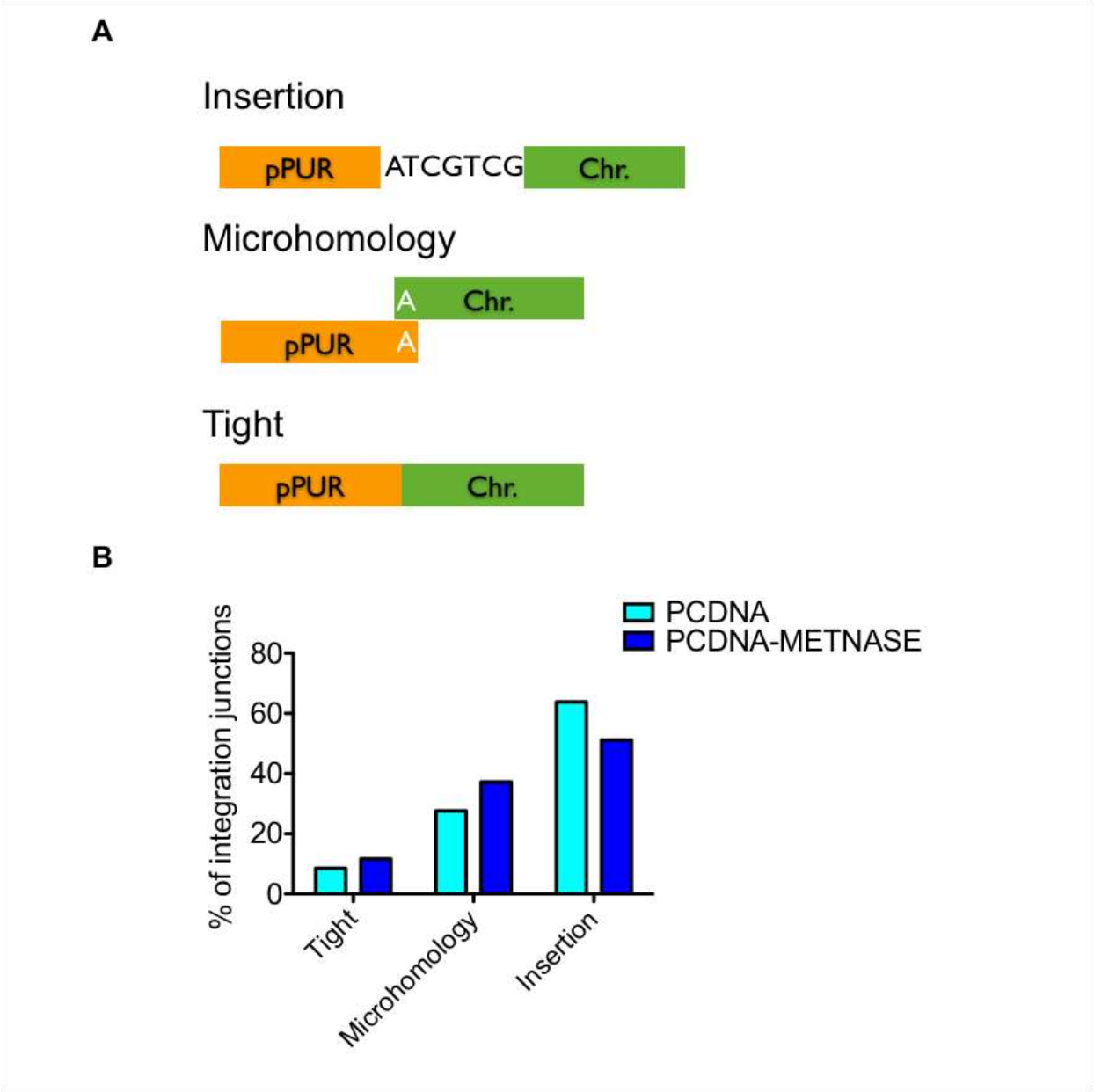

Figure 6. Categories of integration junctions
(A) Schematic diagram of three categories of integration junctions.
(B) Percentage of each type of integration junctions of transfection without Metnase (cyan bar) and with Metnase (blue) was plotted.

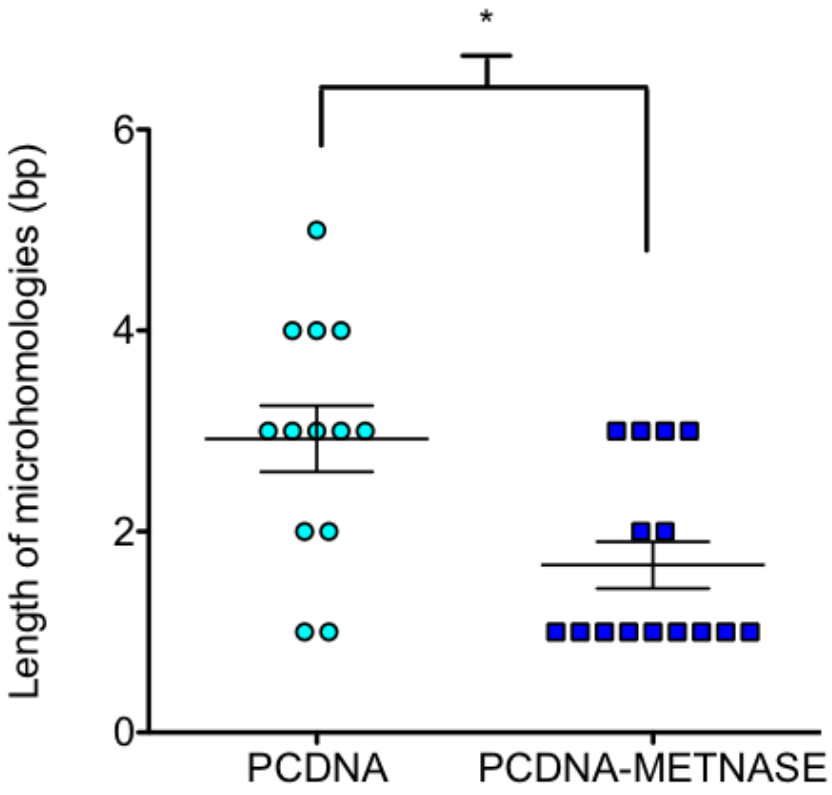

Figure 7. Microhomology between plasmid and integration site
Length of microhomology junctions with microhomology was plotted as circles for transfection without Metnase and squares for transfection with Metnase. * P=0.0075, t-test.

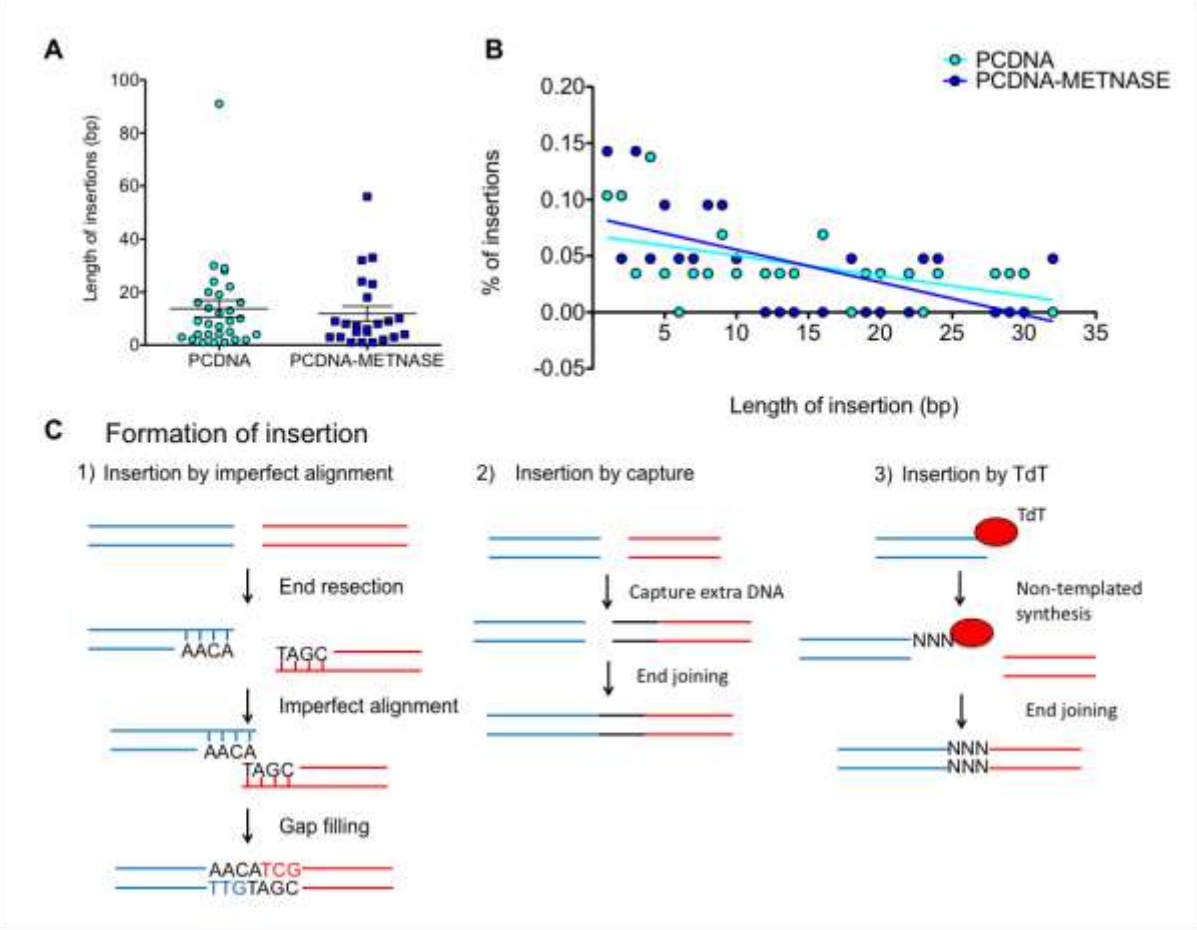

Figure 8. Insertion between plasmid and integration site

(A) Length of insertions with insertions at plasmid-chromosome junctions was plotted as circles for transfection without Metnase and squares for transfection with Metnase.

(B) Junctions with insertions less than 30bp were plotted with length of insertion on the X- axis and percentage on the Y-axis. Junctions from PCD and MET products were fitted with linear line.

(C) Mechanisms of insertion at end joining or integration junction. 1) Imperfect alignment of two ends is followed by gap filling, resulting in extra bases at the integration junction. 2) Cellular DNA or plasmid DNA could be captured that the junction. 3) Terminal transferase (TdT) catalyzes non-templated addition of nucleotides to ends.

Plasmid sequence:
Left boundary- TCTGTAGCAC CGCCTACATA CCTCGCTCTG CTAATCCTGT TACCAGTGG 3′
Right boundary- TCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA 3′

PCD1L GCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGG 3′
CCTTT
TCAAAAAGTGGGCAAAGGATATGAGCAGACACTTCTCAAAAGAAGACATTTATGTTGCCAA (X)

PCD3L CACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGG 3′
CCCACCTGGCTTATGATGCTAGACCCTCTACCACTGTTGCTGTGCTGAAAGCCACACAGCAC (1)

PCD4L CAGTGGCTG....GCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACG
AAATATCTGGAGCT
GGGGAGAGGCAGTAAACTTTTTTAAGACACTCAATATCTATTGGCCCTGGGAAAaaAATTAT (3)

PCD4R ACTTATCGCCACTGGCAGCAGCCA 3′
CTAGCTGCT
AGGTAATTTTGAAGCTTATGGCATCAAAATGCCGCTGGTGTTTATACATCTAGCTTTCCTGA (X)

PCD6L CACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGG 3′
CCCACCTGGCTTATGATGCTAGACCCTCTACCACTGTTGCTGTGCTGAAAGCCACACAGCAC (1)

PCD11R ACTTATCGCCACTGGCAGCAGCCA 3′
CTGTAACA
CTTGAGCCTGGGAGGCAAAGGTTGCAGTGACCTGAGATTGTGCCTCTGCACTCCAACCTGG (13)

PCD13R CGACTTATCGCCACTGGCAGCAGCCA 3′
GTCGCCACTGGGACTTAA
AAAATTAGCCCAGCGTGGTAATGAAGGCCCATAGTCCTTTCTACTCAGGAGGCTGAGGTGGG (X)

PCD14L CACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGG 3′
CCCACCTGGCTTATGATGCTAGACCCTCTACCACTGTTGCTGTGCTGAAAGCCACACAGCAC (1)

PCD14R AGAGGTGGCGAA....TTACCAGTGG 3′
GGCGAAGAGTCCTAGCTTGGGATT
GGTGGCGAAATCCTGAGGGGCTTCTCCTCCTAGACCGACTGTCCTCCCTTTTCTTTGAAATC (4)

PCD16L CACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGG 3′
CCCACCTGGCTTATGATGCTAGACCCTCTACCACTGTTGCTGTGCTGAAAGCCACACAGCAC (1)

PCD19L CACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGG 3′
CCCACCTGGCTTATGATGCTAGACCCTCTACCACTGTTGCTGTGCTGAAAGCCACACAGCAC (1)

PCD31L CACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGG3′
TGGTGCTCTCCATCCCAGGAAGAAATTAGAACTCTGTTGGCTATAGGACACAAGTGGGGGT (X)

PCD31R TGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA 3′
CATATTA
TGTTTTTTTTCTTTAAAGCTTCTAATTTCCATATTGCTATCTGATTTCAACACTGCAGCACT (7)


PCD33L TGCTAATCCTGTTACCAGTGG 3′
TATG
AGCTGCAGAGAGCCACAGCTCCCAATCAGCCACGTGATCAG (4)
PCD34R ATCGCCACTGGCAGCAGCCA 3′
TATGTATATGTATATGAAA
TATTCCAGTGCTAGTGGTAAATAATCTATATCCTTTTGGTCAATAACACCTGATTTCATGTA (X)
PCD35R ACTTATCGCCACTGGCAGCAGCCA 3′
CTAGCTGCT
AGGTAATTTTGAAGCTTATGGCATCAAAATGCCGCTGGTGTTTATACATCTAGCTTTCCTGA (X)
PCD41R TCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA 3′
ATGATAAGGGCTTCTAGAAATAAAACCTCAGCAGACAGGAAACTGCCTTCAAACTTCAGCAA (1)
PCD45R CACGACTTATCGCCACTGGCAGCAGCCA 3′
TATGCTTGAACT
ACCTCAGGTGATCTGCTTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGACATGccCgcCCc (8)
PCD49L ACATACCTCGCTCTGCTAATCCTGTTACCAGTGG 3′
TTAATAATAGCTACTGTAATGCTATAATGACACA (6)
PCD50R TAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA 3′
AAACCCAAAACCAGAAGCAATCAAATATTTACCTTTTCCTAGTCACTCACAAATGAAAGGGT (X)
PCD52R GACTTATCGCCACTGGCAGCAGCCA 3′
TG
CCGCTAGCTTGTCCCCGGCCCAGCGATGCCTGCTTGCCG (22)
PCD55L CATACCTCGCTCTGCTAATCCTGTTACCAGTGG 3′
GCCAAAGAAGTGAGATACTCACTATGAATTTCCAAGCCTTAAAAATCCAGGTGGAGAAATTA (X)
PCD55R GGTAAGACACGACTTATCGCCACTGGCAGCAGCCA 3′
A
TAGTGAGTATCTCACTTCTTTGGCTTAAAGTAGTGATGGCAAATATATTTCAACTTCTGAGT (X)
PCD58L AGGTATCCGGTAAGCGGCAG
CTCCAGAATTAATGATTCAGTGCCCCAGACAAAGGTGGCATAGTGTGCAGCCACAGGAAGCA (7)
PCD58R AGACACGACTTATCGCCACTGGCAGCAGCCA 3′
TGTT
TGGAGCTGGCATAGGTCCCTCTGCACCAGGCACAAGGTGGCCTTTCCCATG (7)
PCD62R TGTAGGTCGTTCGCTCCAAGCTG....GCAGCAGCCA 3′
GCTGCTGCTTTAAACATAAGTTGATTTCTTCCAAAGACCTGAGGCAGCAGCAGTGGGCTATC (1)
PCD64R ATCGCCACTGGCAGCAGCCA 3′
GTGATTTCTTATGCTATAAA
CAGATTAGGGTTAGATCTCTGTGCATGCCACTGGTGACATTTATCTATTAGGCAACATTCTT (X)


PCD66R ATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA 3′
ATAAGTCGTGTCTTACTTACTCTGAGTTT
AGTTTCTTAAAGTACTTTGATATCACTCAATAAACTACAGTTAAACTCAGAGTTTTACATTT (X)
PCD71L TGTTACCAGTGG 3′
GTCATTGACGTCAATAGGGGGCGTACTTGG
TTTCTTTCTTTTTTTTTTTTTTTTTTAAAGAGTCTCACTCTGTTGCCCAGGCTGGAGTGCA (15)
PCD71R GGTAAGACACGACTTATCGCCACTGGCAGCAGCCA 3′
AAGATGCAAAGGAGTCGGAAAACTCCTTTTCTGGTCTGgAGGCAAgAtCTTCCCTGGAGAA (15)
PCD74L CAGTGGCTG....GCAGGGTCGGAACAGGAGAG
A
AATATCTATTGCTCCTGGGAAAAAACTTA (3)
PCD83L TACCTCGCTCTGCTAATCCTGTTACCAGTGG 3′
CAATTAAAAGGCTCTGCTAAGC
AATGTCTTTCAAGTTTATTTAGAGCCCCAGAACACTCCAACACTCAATtAAAAGGCTTACCA (3)
PCD83R AGACACGACCTATCGCCACTGGCAGCAGCCA 3′
TGTT
ACCAGGCACAAGGTGGCCTTTCCCATG (7)
PCD87R TtGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA 3′
GTAAACCCTTTCCAGGACGGATCGACTGGGGAACAGGCATCCGGCTTCCAGGGGTAGCCCCC (7)
PCD93L CTGCTAATCCTGTTACCAGTGG 3′
AATGTGATCTGAGAGGCAGTTTGTTTACATCTATACCCTTTTTAAAGTCACAAGTACAATTA (X)
PCD93R CGACTTATCGCCACTGGCAGCAGCCA 3′
TGGCGCCATATATATA
AGAAATACCATTTGACCCAGCCATCCCATTACTGGGTATATACCCAAATGAGTATAAATCAT (5)
PCD94L TCGCTCTGCTAATCCTGTTACCAGTGG 3′
CATCCTCTTCCTGACTTTTATTCCTTACACAAATTTTACTTTGTCATAATCTATAATGTTTA (1)
PCD94R CTTATCGCCACTGGCAGCAGCCA 3′
AAGGAATTCAGTGTCAGCAAGGGCCGCCAGACTGGCCAGAGTCCCAAAGTGTTTACACTCCG (1)
PCD97L CTAATCCTGTTACCAGTGG 3′
AAGTGGCGCCAAG
CCTGTAATCCCAGCACTTCGGGAGGCTGAGGCGGGCGGATCACCTGAGGTCAGGAGTTCAA (17)
PCD97R ACTTATCGCCACTGGCAGCAGCCA 3′
AGT
GCCACTACAGACACAAGCTTAAAGATTGCCCCAGTGTACTCAGGTCACAGGAACCCTCAAGT (X)


PCD99L GCTCTGCTAATCCTGTTACCAGTGG 3′
ATGAGTCTTCTTGGGTTACCTTTTACTTGACATG (X)
PCD99R CTTATCGCCACTGGCAGCAGCCA 3′
AG
GTAAAAATATGGTTGAACTCAACCACTAGTTGCtCATTAGAACAGACTTTCCCAGTGTACTG (X)
PCD102L AGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGG 3′
CTGAGAGGGCGGGCCTGCCCTCTGCGCCCCCCTGCCCTCCGA
GGCCTGCCAGCCTCCAGGTGAGGTCGGCCGAGGCCCGCCTCTTTGCATC
GGCCGCAGCTGATCCCAGCCCAGAGGAGGCCGGGCCCAGCACCCTGGGAAAcgACTTGGCT
GCCgGCCAGGGTGTGGCCCC (22)
PCD102R AGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA 3′
TAAAACTGTTTTTTAT
AGGCAACCCCTCCCACAAAACAAAGGAGAAAAAAAAAAACAAAAAACTGTTTTTTAAAAAAT (2)
PCD108L CCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGG 3′
G
TGTCCGTTCCCATTTCCGGACTGTGAGATGAAGCTTCTGGAAGCTCCCTCTGATTTCAGCTC (8)
PCD116R ACGACTTATCGCCACTGGCAGCAGCCA 3′
TA
TCTGTGTATAATAAGATACATTTAGGTCTGAATCTAACCTAAACTGGCAATTTTGGCATTGA (8)
PCD119L TACATACCTCGCTCTGCTAATCCTGTTACCAGTGG 3′
CACGGTTAGC
GGGTTGGGGAAGTGGGCAGGAGTGGGTGGAAGGGGGTGAAGAGGTGAGCCATAATGGGTGGA (X)
PCD127L CTCGCTCTGCTAATCCTGTTACCAGTGG 3′
AGGCAGTGGAATTAGCTGTTGTTTTCTCCTTCCTTGGAGCAGGGTTATTATGTCATG (6)
PCD127R CTTATCGCCACTGGCAGCAGCCA 3′
ATTTTGACAGTAAGAATAAATGAAGTGGTCCTTTTAG (20)
PCD160L TCTGCTAATCCTGTTACCAGTGG 3′
CAAAGCAGGCTTCACTTAAAAAAATAAGATTTTCTCTTTACATTTAAAATATATCCCACTAC (X)
PCD175L CCTCGCTCTGCTAATCCTGTTACCAGTGG 3′
GGAC
ACTATAGAAAAAGTAAGATTTCTATATGTTTTATAATCAGTAAAAAGTTATTCCTTATGATT (X)


MET1R AGACACGACTTATCGCCACTGGCAGCAGCCA 3'end
CTCATGAGGCACCCACTTATCGAGCTTTTTCACCTTTCCAATTTGCTTCAAATGcCGAACGA (3)
MET18L TCGCTCTGCTAATCCTGTTACCAGTGG 3'end
GTTACCAAAGGCACACAATAGTCTTTAGGTTTTTCTGGTTAATGGTGATTGAAAACGGAGCT (X)
MET18R CACGACTTATCGCCACTGGCAGCAGCCA 3'end
TAAGGATGATCTCATGATGTTTGGTTATTCATTCATTTTCAAGAGTGAGGTCCTTAAGCCTA (10)
MET23L CAGTGGCTG....GGAACAGGAG
CGA
GCAGTAAACTTTTTTAAGACACTCAATATCTATTGGCCCTGGGAAAGTAATTATTGGATCCC (3)
MET23R AGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA 3'end
AATGAACAAAGTATTTGAGAAATATGGGATTATATATAGCAACCAAAATGTGAATTTTTGGC (X)
MET26R CTTATCGCCACTGGCAGCAGCCA 3'
TTATTTCTCACTGGCAAAAATGTGTTGATTGTAATGGTTCCTATTTTGATAA (17)
MET28L GCTCTGCTAATCCTGTTACCAGTGG 3'end
TGTCTTCTGTCATAATCACCTATGGAAGTGCCTGCCAGGTTCCAGACTATTTTATCTTCCAT (1)
MET28R TTATCGCCACTGGCAGCAGCCA 3'
CCTGAGGCAGGTGGATCACCTGAGGTCAGGCGTTCGAGACCAGCCTGGCTAACATG (1)
MET30R TAAGACACGACTTATCGCCACTGGCAGCAGCCA 3'end
AGAGCAAGAGTGTCTCAAAAAAATAAAATAAAATAAATAATAACTGGGCACAGTGGCTCACG (X)
MET32L TAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGG 3'end
TCAAGTA
GGTGCGAATCCCTGCTGTAGAAAACAGGCTGAGGAAAGAAAGCGACtCAGGATGTTGTCTGG (2)
MET32R TATCGCCACTGGCAGCAGCCA 3'end
TTGCTCAGGCTGGAGTGCACTGGTGCGATCATGGCTCACTGCAGCCTCTATCTCCTGGCTGA (2)
MET33L CGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGG 3'end
AAAGTCAAGTGTAAAGAAGCACTGTAAATATCATGAATAAAcGcCcCCTCCTCACAGTATTA (8)
MET33R CTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA 3'end
TGATA
CACACCAGTTAGAATGGCAATCATTAAAAAGTCAGGAAACAACAGGTGCTGGAGAGGATGCG (8)*
MET36L TCGCTCTGCTAATCCTGTTACCAGTGG 3'
ACAGCCTGTTACA
CTATGGGGTTAATCCAGCAGCAGCATGTACCCTGGAAAAAGGAGGAGGTGACAACAGGAAAA (X)
MET36R TTATCGCCACTGGCAGCAGCCA 3'
TAATTAACTGCTAATCCTGTACAGCTAAGCCT
GTAATCCCAGCACTTTAGGAGGCCAAGGCGGGTGGACCACCTGAAGTCAGGAGTTTGATGAA (X)

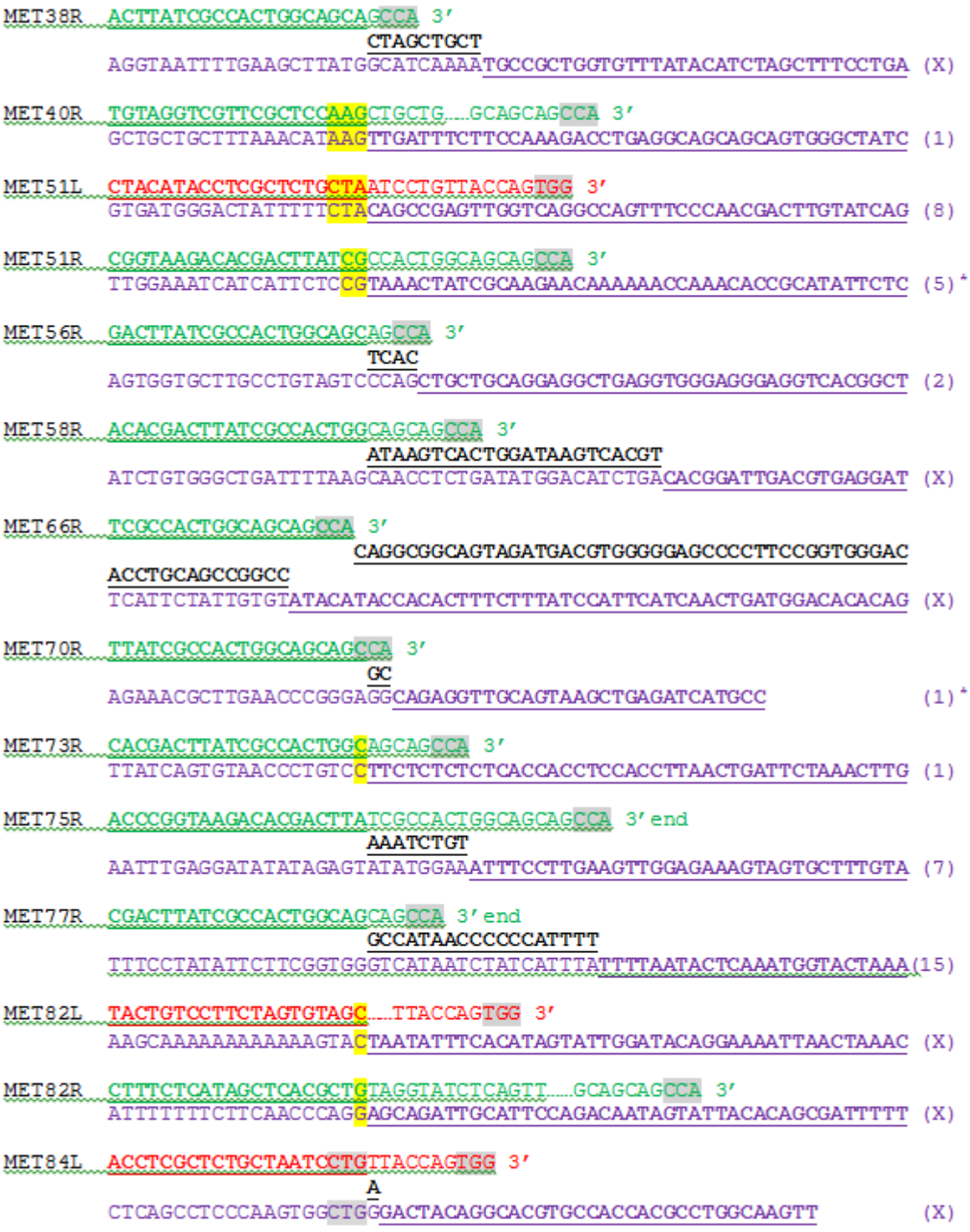
MET38R ACTTATCGCCACTGGCAGCAGCCA 3′
CTAGCTGCT
AGGTAATTTTGAAGCTTATGGCATCAAAATGCCGCTGGTGTTTATACATCTAGCTTTCCTGA (X)
MET40R TGTAGGTCGTTCGCTCCAAGCTGCTG....GCAGCAGCCA 3′
GCTGCTGCTTTAAACATAAGTTGATTTCTTCCAAAGACCTGAGGCAGCAGCAGTGGGCTATC (1)
MET51L CTACATACCTCGCTCTGCTAATCCTGTTACCAGTGG 3′
GTGATGGGACTATTTTTCTACAGCCGAGTTGGTCAGGCCAGTTTCCCAACGACTTGTATCAG (8)
MET51R CGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA 3′
TTGGAAATCATCATTCTCCGTAAACTATCGCAAGAACAAAAAACCAAACACCGCATATTCTC (5)*
MET56R GACTTATCGCCACTGGCAGCAGCCA 3′
TCAC
AGTGGTGCTTGCCTGTAGTCCCAGCTGCTGCAGGAGGCTGAGGTGGGAGGGAGGTCACGGCT (2)
MET58R ACACGACTTATCGCCACTGGCAGCAGCCA 3′
ATAAGTCACTGGATAAGTCACGT
ATCTGTGGGCTGATTTTAAGCAACCTCTGATATGGACATCTGACACGGATTGACGTGAGGAT (X)
MET66R TCGCCACTGGCAGCAGCCA 3′
CAGGCGGCAGTAGATGACGTGGGGGAGCCCCTTCCGGTGGGAC
ACCTGCAGCCGGCC
TCATTCTATTGTGTATACATACCACACTTTCTTTATCCATTCATCAACTGATGGACACACAG (X)
MET70R TTATCGCCACTGGCAGCAGCCA 3′
GC
AGAAACGCTTGAACCCGGGAGGCAGAGGTTGCAGTAAGCTGAGATCATGCC (1)*
MET73R CACGACTTATCGCCACTGGCAGCAGCCA 3′
TTATCAGTGTAACCCTGTCCTTCTCTCTCTCACCACCTCCACCTTAACTGATTCTAAACTTG (1)
MET75R ACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA 3′end
AAATCTGT
AATTTGAGGATATATAGAGTATATGGAAATTTCCTTGAAGTTGGAGAAAGTAGTGCTTTGTA (7)
MET77R CGACTTATCGCCACTGGCAGCAGCCA 3′end
GCCATAACCCCCCATTTT
TTTCCTATATTCTTCGGTGGGTCATAATCTATCATTTATTTTAATACTCAAATGGTACTAAA(15)
MET82L TACTGTCCTTCTAGTGTAGC....TTACCAGTGG 3′
AAGCAAAAAAAAAAAAAGTACTAATATTTCACATAGTATTGGATACAGGAAAATTAACTAAAC (X)
MET82R CTTTCTCATAGCTCACGCTGTAGGTATCTCAGTT.....GCAGCAGCCA 3′
ATTTTTTTCTTCAACCCAGGAGCAGATTGCATTCCAGACAATAGTATTACACAGCGATTTTT (X)
MET84L ACCTCGCTCTGCTAATCCTGTTACCAGTGG 3′
A
CTCAGCCTCCCAAGTGGCTGGGACTACAGGCACGTGCCACCACGCCTGGCAAGTT (X)

Figure 9. Complete list of integration junctions

Sequenced junctions are shown with bold underlined letters. Plasmid sequences are shown in red for the left boundary and green for the right boundary. Genomic sequences are shown in purple. Insertions between plasmid and chromosome are shown in black. Overhangs of AlwNI cut site are highlighted in gray. Microhomologies are highlighted in yellow. The * indicates the target genomic sequence belongs to repetitive sequence that exist in multiple chromosomes.

REFERENCES

Abbas, M., Shanmugam, I., Bsaili, M., Hromas, R., and Shaheen, M. (2014). The role of the human psoralen 4 (hPso4) protein complex in replication stress and homologous recombination. J Biol Chem *289*, 14009-14019.

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., *et al.* (2003). Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science *301*, 653-657.

Beck, B.D., Lee, S.S., Hromas, R., and Lee, S.H. (2010). Regulation of Metnase's TIR binding activity by its binding partner, Pso4. Arch Biochem Biophys *498*, 89-94.

Beck, B.D., Lee, S.S., Williamson, E., Hromas, R.A., and Lee, S.H. (2011). Biochemical characterization of metnase's endonuclease activity and its role in NHEJ repair. Biochemistry *50*, 4360-4370.

Beck, B.D., Park, S.J., Lee, Y.J., Roman, Y., Hromas, R.A., and Lee, S.H. (2008). Human Pso4 is a metnase (SETMAR)-binding partner that regulates metnase function in DNA repair. J Biol Chem *283*, 9023-9030.

Bergsmedh, A., Szeles, A., Henriksson, M., Bratt, A., Folkman, M.J., Spetz, A.L., and Holmgren, L. (2001). Horizontal transfer of oncogenes by uptake of apoptotic bodies. Proc Natl Acad Sci U S A *98*, 6407-6411.

Cabaniols, J.P., Fazilleau, N., Casrouge, A., Kourilsky, P., and Kanellopoulos, J.M. (2001). Most alpha/beta T cell receptor diversity is due to terminal deoxynucleotidyl transferase. J Exp Med *194*, 1385-1390.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., *et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Cordaux, R., Udit, S., Batzer, M.A., and Feschotte, C. (2006). Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. Proc Natl Acad Sci U S A *103*, 8101-8106.

De Bont, R., and van Larebeke, N. (2004). Endogenous DNA damage in humans: a review of quantitative data. Mutagenesis *19*, 169-185.

Desfarges, S., and Ciuffi, A. (2010). Retroviral integration site selection. Viruses *2*, 111-130.

Devine, S.E., and Boeke, J.D. (1996). Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. Genes Dev *10*, 620-633.

Ehnfors, J., Kost-Alimova, M., Persson, N.L., Bergsmedh, A., Castro, J., Levchenko-Tegnebratt, T., Yang, L., Panaretakis, T., and Holmgren, L. (2009). Horizontal transfer of tumor DNA to endothelial cells in vivo. Cell Death Differ *16*, 749-757.

Fnu, S., Williamson, E.A., De Haro, L.P., Brenneman, M., Wray, J., Shaheen, M., Radhakrishnan, K., Lee, S.H., Nickoloff, J.A., and Hromas, R. (2011). Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. Proc Natl Acad Sci U S A *108*, 540-545.

Friedrich, T.D., Laffin, J., and Lehman, J.M. (1992). Simian virus 40 large T-antigen function is required for induction of tetraploid DNA content during lytic infection. J Virol *66*, 4576-4579.

Gu, W.G. (2014). Newly approved integrase inhibitors for clinical treatment of AIDS. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie *68*, 917-921.

Hazuda, D.J. (2012). HIV integrase as a target for antiretroviral therapy. Current opinion in HIV and AIDS *7*, 383-389.

Holmgren, L. (2010). Horizontal gene transfer: you are what you eat. Biochem Biophys Res Commun *396*, 147-151.

Hromas, R., Wray, J., Lee, S.H., Martinez, L., Farrington, J., Corwin, L.K., Ramsey, H., Nickoloff, J.A., and Williamson, E.A. (2008). The human set and transposase domain protein Metnase interacts with DNA Ligase IV and enhances the efficiency and accuracy of non-homologous end-joining. DNA Repair (Amst) *7*, 1927-1937.

Landau, N.R., Schatz, D.G., Rosa, M., and Baltimore, D. (1987). Increased frequency of N-region insertion in a murine pre-B-cell line infected with a terminal deoxynucleotidyl transferase retroviral expression vector. Mol Cell Biol *7*, 3237-3243.

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

Lebkowski, J.S., Clancy, S., and Calos, M.P. (1985). Simian Virus-40 Replication in Adenovirus-Transformed Human-Cells Antagonizes Gene-Expression. Nature *317*, 169-171.

Lee, K., and Lee, S.E. (2007). Saccharomyces cerevisiae Sae2- and Tel1-dependent single-strand DNA formation at DNA break promotes microhomology-mediated end joining. Genetics *176*, 2003-2014.

Lee, S.H., Oshige, M., Durant, S.T., Rasila, K.K., Williamson, E.A., Ramsey, H., Kwan, L., Nickoloff, J.A., and Hromas, R. (2005). The SET domain protein Metnase mediates foreign DNA integration and links integration to nonhomologous end-joining repair. Proc Natl Acad Sci U S A *102*, 18075-18080.

Li, S., and Huang, L. (2000). Nonviral gene therapy: promises and challenges. Gene Ther *7*, 31-34.

Li, Y., Xuan, S., Feng, Y., and Yan, A. (2014). Targeting HIV-1 integrase with strand transfer inhibitors. Drug discovery today.

Limberis, M.P. (2012). Phoenix rising: gene therapy makes a comeback. Acta Biochim Biophys Sin (Shanghai) *44*, 632-640.

Lin, Y., and Waldman, A.S. (2001). Capture of DNA sequences at double-strand breaks in mammalian chromosomes. Genetics *158*, 1665-1674.

Little, K.C., and Chartrand, P. (2004). Genomic DNA is captured and amplified during double-strand break (DSB) repair in human cells. Oncogene *23*, 4166-4172.

Liu, D., Bischerour, J., Siddique, A., Buisine, N., Bigot, Y., and Chalmers, R. (2007a). The human SETMAR protein preserves most of the activities of the ancestral Hsmar1 transposase. Mol Cell Biol *27*, 1125-1132.

Liu, D., Bischerour, J., Siddique, A., Buisine, N., Bigot, Y., and Chalmers, R. (2007b). The human SETMAR protein preserves most of the activities of the ancestral Hsmar1 transposase. Mol Cell Biol *27*, 1125-1132.

Liu, X.N., and Baird, V. (2001). Rapid amplification of genomic DNA ends by Nla III partial digestion and polynucleotide tailing. Plant Mol Biol Rep *19*, 261-267.

Ma, J.L., Kim, E.M., Haber, J.E., and Lee, S.E. (2003). Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. Mol Cell Biol *23*, 8820-8828.

Mahajan, K.N., and Mitchell, B.S. (2003). Role of human Pso4 in mammalian DNA repair and association with terminal deoxynucleotidyl transferase. Proc Natl Acad Sci U S A *100*, 10746-10751.

Maxfield, L.F., Fraize, C.D., and Coffin, J.M. (2005). Relationship between retroviral DNA-integration-site selection and host cell transcription. Proc Natl Acad Sci U S A *102*, 1436-1441.

Pannunzio, N.R., Li, S., Watanabe, G., and Lieber, M.R. (2014). Non-homologous end joining often uses microhomology: implications for alternative end joining. DNA Repair (Amst) *17*, 74-80.

Purugganan, M.M., Shah, S., Kearney, J.F., and Roth, D.B. (2001). Ku80 is required for addition of N nucleotides to V(D)J recombination junctions by terminal deoxynucleotidyl transferase. Nucleic Acids Res *29*, 1638-1646.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nature protocols *8*, 2281-2308.

Roman, Y., Oshige, M., Lee, Y.J., Goodwin, K., Georgiadis, M.M., Hromas, R.A., and Lee, S.H. (2007). Biochemical characterization of a SET and transposase fusion protein, Metnase: its DNA binding and DNA cleavage activity. Biochemistry *46*, 11369-11376.

Rubnitz, J., and Subramani, S. (1984). The minimum amount of homology required for homologous recombination in mammalian cells. Mol Cell Biol *4*, 2253-2258.

Saleh-Gohari, N., Bryant, H.E., Schultz, N., Parker, K.M., Cassel, T.N., and Helleday, T. (2005). Spontaneous homologous recombination is induced by collapsed replication forks that are caused by endogenous DNA single-strand breaks. Mol Cell Biol *25*, 7158-7169.

Simsek, D., Brunet, E., Wong, S.Y., Katyal, S., Gao, Y., McKinnon, P.J., Lou, J., Zhang, L., Li, J., Rebar, E.J., *et al.* (2011). DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. PLoS Genet *7*, e1002080.

Solyom, S., and Kazazian, H.H., Jr. (2012). Mobile elements in the human genome: implications for disease. Genome Med *4*, 12.

Strachan, T., and Read, A.P. (1999). Human Molecular Genetics. In, n. edition, ed. (New York: Wiley-Liss).

Suzuki, K., Ohbayashi, F., Nikaido, I., Okuda, A., Takaki, H., Okazaki, Y., and Mitani, K. (2010). Integration of exogenous DNA into mouse embryonic stem cell chromosomes shows preference into genes and frequent modification at junctions. Chromosome Res *18*, 191-201.

Vilenchik, M.M., and Knudson, A.G. (2003). Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. Proc Natl Acad Sci U S A *100*, 12871-12876.

Wei, K., Kodym, R., and Jin, C. (1998). Radioresistant cell strain of human fibrosarcoma cells obtained after long-term exposure to x-rays. Radiat Environ Biophys *37*, 133-137.

Williamson, E.A., Farrington, J., Martinez, L., Ness, S., O'Rourke, J., Lee, S.H., Nickoloff, J., and Hromas, R. (2008a). Expression levels of the human DNA repair protein metnase influence lentiviral genomic integration. Biochimie *90*, 1422-1426.

Williamson, E.A., Rasila, K.K., Corwin, L.K., Wray, J., Beck, B.D., Severns, V., Mobarak, C., Lee, S.H., Nickoloff, J.A., and Hromas, R. (2008b). The SET and transposase domain protein Metnase enhances chromosome decatenation: regulation by automethylation. Nucleic Acids Res *36*, 5822-5831.

Wray, J., Williamson, E.A., Chester, S., Farrington, J., Sterk, R., Weinstock, D.M., Jasin, M., Lee, S.H., Nickoloff, J.A., and Hromas, R. (2010). The transposase domain protein Metnase/SETMAR suppresses chromosomal translocations. Cancer Genet Cytogenet *200*, 184-190.

Wurtele, H., Little, K.C., and Chartrand, P. (2003). Illegitimate DNA integration in mammalian cells. Gene Ther *10*, 1791-1799.

Yan, B., Li, D., and Gou, K. (2010). Homologous illegitimate random integration of foreign DNA into the X chromosome of a transgenic mouse line. BMC Mol Biol *11*, 58.

Yu, A.M., and McVey, M. (2010). Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. Nucleic Acids Res *38*, 5706-5717.

Zhang, N., Kaur, R., Lu, X., Shen, X., Li, L., and Legerski, R.J. (2005). The Pso4 mRNA splicing and DNA repair complex interacts with WRN for processing of DNA interstrand cross-links. J Biol Chem *280*, 40559-40567.

CHAPTER 3

PARP1 REGULATES DNA INTEGRATION

INTRODUCTION

Integration of foreign DNA into the mammalian genome is an important yet poorly-understood biological process. DNA integration occurs in various natural formats and has been adopted for various genome editing tools (Chapter 1). For the most part, integration of foreign DNA in mammalian cells tends to be random. Various methods have been developed to improve targeted integration (or "gene targeting") in mammalian cells. However, non-targeted integration of genomic editing vectors cannot be completely avoided (reviewed in Ain et al., 2014). Integration of foreign DNA into the recipient genome can inactivate genes at the integration sites, and may disrupt the overall stability of the recipient genome (reviewed in Wurtele et al., 2003). Therefore, it is important to better control the integration process of various vectors. A thorough understanding of the mechanisms of foreign DNA integration will greatly improve our ability to precisely edit genomes.

DNA integration efficiency increases in the presence of DNA damage. In mammalian cells, random DNA integration likely utilizes the cellular non-homologous end joining (NHEJ) machinery (reviewed in Wurtele et al., 2003). The dominant NHEJ pathway repairs double-strand breaks (DSBs) independent of end resection or microhomology (5-25 nucleotides). This pathway initiates with binding of the broken ends by KU70/80, followed by DNA-dependent protein kinase (DNA-PK) binding and formation of a bridge-like structure to keep the ends adjacent to each other, and ligation by Ligase IV (Chapter 1), and is referred to as classical non-homologous end joining (cNHEJ). Deficiency of key cNHEJ proteins, such as Ligase IV or , X-

ray repair cross-complementing protein 4 (XRCC4), does not completely abort rejoining of broken chromosomes, leading to the discovery of the alternative non-homologous end joining (aNHEJ) pathway (Delacote et al., 2002; Deriano and Roth, 2013; Kabotyanski et al., 1998; Liang and Jasin, 1996; Liang et al., 1996). There are two major differences between classical and alternative NHEJ. First, aNHEJ relies on limited end resection to expose microhomologies between the two ends. Microhomologies serve as anchor points for the two ends to anneal together, while cNHEJ can rejoin the ends with minimal or no end processing. Second, Ligase III and Ligase I catalyze end ligation in aNHEJ instead of Ligase IV in cNHEJ (Simsek et al., 2011).

Alternative NHEJ is considered the major pathway of generating chromosomal translocations. Analysis of chromosomal translocations in leukemia cells reveals high frequency of microhomologies at translocation junctions, suggesting a connection between aNHEJ and chromosomal translocations (Jeyaratnam et al., 2014; Mattarucchi et al., 2008; Zhang and Rowley, 2006). Key cNHEJ proteins KU, XRCC4, Ligase IV are not required for chromosome translocations (Simsek and Jasin, 2010), in fact, overexpression of cNHEJ proteins KU or Metnase suppresses translocations in murine cells (Weinstock et al., 2007; Wray et al., 2010). Such evidence leads to the model that cNHEJ repairs broken chromosomes efficiently and therefore prevents broken ends from drifting apart and misjoining with the wrong ends, while aNHEJ repairs the broken chromosomes less efficiently and less accurately (Boboila et al., 2010a; Boboila et al., 2010b), hence aNHEJ mediates chromosomal translocations. However, chromosomal translocation analysis in several cNHEJ-deficient human cells showed that chromosomal translocation efficiency was reduced compared to control cell, suggesting that cNHEJ also play an important role in generating translocations in human cells; the same study also showed that translocation junctions in these cNHEJ-deficient human cells had increased

ratio of microhomologies longer than 3 nucleotides, which indicates increased portion of end joining via aNHEJ that utilizes microhomology to joining broken ends (Ghezraoui et al., 2014). More work is required to resolve the different roles of cNHEJ on generating chromosomal translocations in mice and humans. The importance of aNHEJ in chromosomal translocations is well supported by current publications (reviewed in Byrne et al., 2014).

Alternative NHEJ is distinct from cNHEJ in early stage in terms of requirement of end resection. In cNHEJ, KU binds to DSB to protect ends from resection; in aNHEJ, *meiotic* recombination 11 (MRE11) and *C-terminal-binding protein* interacting protein (CtIP) resect the ends to expose limited microhomology. Endonuclease CtIP plays an essential role in mediating chromosomal translocations through aNHEJ (Lee-Theilen et al., 2011; Zhang and Jasin, 2011). Knockdown of CtIP results in shifting translocation junctions towards shorter deletions and microhomologies, emphasizing the key roles of end resection in initiating aNHEJ. Initiation of end resection for aNHEJ requires PARP1, which competes with KU for DSB binding and promotes recruitment of MRE11-RAD50-NBS1 (MRN) complex and CtIP for end resection (Della-Maria et al., 2011; Haince et al., 2008; Wang et al., 2006).

PARP1 is the founding member of the poly (ADP-ribose) polymerase (PARP) family, a large protein family with 18 members (reviewed in Ame et al., 2004). PARP1 catalyzes addition of ADP-ribose units to target proteins (Vyas et al., 2013). PARP1 has a large profile of targets, including histones and multiple DNA repair proteins (Gagne et al., 2008). PARP1 is activated by various types of DNA damage and has important roles in repair of both single-strand and double-strand DNA damage, and DNA damage response signaling (reviewed in Beck et al., 2014; De Vos et al., 2012; Sousa et al., 2012). In murine cells, inhibition of PARP1 with the small molecule Olaparib significantly suppressed chromosomal translocations following DSB

induction with different nucleases (Wray et al., 2013). Joining of broken ends from different chromosomes is a process similar to DNA integration. Because PARP1 has a significant role in initiating chromosomal translocations, it is very likely that independent from cNHEJ, PARP1also mediates some part of DNA integration, possibly via the aNHEJ pathway. If PARP1-mediated aNHEJ is responsible for part of DNA integration, inhibition of PARP1 would reduce DNA integration efficiency. To test this hypothesis, we tested the effects of PARP1 inhibitors on plasmid integration efficiency. Surprisingly, the two PARP1 inhibitors we used yielded opposite effects on DNA integration. We further examined the involvement of PARP1 in DNA integration in a stable PARP1 knockdown cell line. Our data confirmed the engagement of PARP1 in DNA integration, but more work is required to uncover the underlying mechanisms and to resolve the opposite effects of PARP1 inhibitors on DNA integration.

EXPERIMENTAL PROCEDURES

Cell cultures and PARP1 knockdown

HT1080 is a male human fibrosarcoma cell line with a stable karyotype (Wei et al., 1998). HT1904 is a derivative of HT1080 that contains an integrated a single integrated copy of an I-SceI site which allows induction of a single, specific DSB in the genome (Fnu et al., 2011). These cell lines were chosen for this project to avoid issues with data interpretation due to unstable karyotype and multiple copies of X-chromosome encountered in HEK293T cells (Chapter 2). Unlike HEK293T cells, which do not express Metnase (Williamson et al., 2008) and are resistant to G418 due to the presence of a neomycin gene, HT1080 and HT1904 cells express normal level of Metnase (Fnu et al., 2011), and are not resistant to G418. Therefore, plasmids carrying neomycin or other antibiotic resistance cassettes can be used to study DNA integration efficiency assay in HT1080 and HT1904 cells.

HT1080 and HT1904 cells were maintained in α-MEM (Life Technologies) supplemented with 10% fetal bovine serum (Atlas Biologicals) under 5% CO_2 at 37 ℃. PAPR1 short hairpin RNA (shRNA) (5'CTTCGTTAGAATGTCTGCCTT'3) lentiviral suspension was purchased from Functional Genomics Shared Resource of University of Colorado. Transduction was performed as follows: HT1080 cells were seed in 6-well tissue culture plate at $2.5X10^5$ cells per well with 2 ml culture medium and incubate for 16 h. When ready for transduction, polybrene was added into each well to a final concentration of 8 μg/ml and rocked gently to mix well. Lentiviral suspension of shRNA was thawed in 37 ℃ waterbath and added to cells immediately. At 20 h post transduction, culture medium was aspirated and replaced with fresh culture medium with puromycin at 10 μg/ml. Cells were maintained in puromycin-containing medium and single puromycin resistant clones were isolated.

Western blotting and antibodies

Individual puromycin resistant clones from PARP1 shRNA lentivirus transduction were expanded into 90% confluence in 10 cm culture dishes. Whole cell extracts were extracted with M-PER Mammalian Protein Extraction Reagent (Thermo Pierce) and quantified using BCA Protein assay (Thermo Pierce). Cell extracts from each single clone were fractionated on 10% SDS-PAGE gel and transferred to PVDF (polyvinyl difluoride) membranes in transfer buffer (0.58% Tris base, 0.29% glycine, 0.037% SDS, 20% methanol) at 300 mA for 45 min. After blocking in 5% milk in PBST (1X Dulbeccos's phosphate buffered saline (Fisher Scientific) with 0.1% Tween-20) at room temperature for 1 h, the membrane was incubated overnight with primary antibodies diluted in 5% milk at 4 ℃. The membrane was washed in PBST twice for 15 min and twice for 5 min. Membranes were incubated for 2 h with horseradish peroxidase (HRP)-conjugated secondary antibodies diluted in 5% milk at room temperature, then washed in PBST

twice for 10 min. After developing with Clarity Western ECL substrate (Bio-Rad), membranes were imaged using the ChemiDoc XRS+ System (Bio-Rad). Antibodies and their working dilutions are listed as follows:

Primary:

Rabbit-anti-PARP1 (Abcam Cat. ab6079), 1:400

Mouse-anti-GAPDH (Abcam Cat. Ab9484), 1:1,000

Secondary:

Goat-anti-rabbit IgG-HRP (Santa Cruz Biotechnology Cat. sc-2054), 1:1,000

Donkey-anti-mouse IgG-HRP (Santa Cruz Biotechnology Cat. sc-2314), 1:1,000

Integration efficiency assay

Cells were seeded in 24-well plates at $5X10^4$ cells per well and cultured overnight to reach 80-90% confluence. Linear plasmid was transfected using Lipofectamine 2000 Transfection Reagent (Life Technologies) according to manufacturer protocol. Transfection reagent was removed from cells 4 h after transfection, and 24 h post-transfection incubation, cells were detached using 0.25% typsin-EDTA (Life Technologies) and suspended in 1 ml α-MEM and aliquoted as follows: 1) 900 μl of cell suspension was transfered into 10 cm dishes for drug selection, 2) 10 μl of cell suspension was diluted into 1 ml α-MEM and 100 μl of the dilution were seeded into 10 cm dishes. After incubating for 24 h, antibiotics were added to final concentrations of 1 μg/ml puromycin (Tocris Bioscience), 450 μg/ml G418 (Gold Biotechnology). Dishes were stained using 0.5% crystal violet (Sigma) and colonies with > 50 cells were scored. Integration efficiency was calculated as the number of antibiotic resistant colonies per live cell seeded in selective medium. When PARP1 inhibitors were used, drugs were added 16 h before transfection and maintained in cultures for 24 h after transfection.

PARP1 inhibitors

PJ34 hydrochloride (*N*-(5,6-Dihydro-6-oxo-2-phenanthridinyl)-2-acetamide hydrochloride) inhibits activity of PARP1 with an IC_{50} of 600 nM and PARP2 with IC_{50} of 1000 nM (Pellicciari et al., 2008; Soriano et al., 2001). Olaparib (4-[[3-[4-(cyclopropanecarbonyl)piperazine-1-carbonyl]-4-fluorophenyl]methyl]-2H-phthalazin-1-one) inhibits the activity of PARP1 with IC_{50} of 5 nM, PARP2 with IC_{50} of 1 nM, tankyrase-1 (TNKS1) with IC_{50} of 1,500 nM (Menear et al., 2008). Olaparib is used as a selective PARP1 inhibitor in research and clinical trials, but it actually has 5-fold selectivity for PARP2 compared to PARP1 (Menear et al., 2008).

RESULTS

PARP1 inhibitors PJ34 and Olaparib have different effects on plasmid integration

To test the hypothesis that PARP1 inhibition suppresses DNA integration, HT1080 and its derivative HT1904 cells were mock treated, or treated with 50 μM of PJ34 (gift from K. Luger) or 3 μM of Olaparib (Selleck Chemicals) for 16 h before and 24 h after transfection of linear AlwNI-digested pPUR plasmid (Chapter 2 Figure 1). Integration efficiency was scored as number of puromycin resistant colonies per live cell seeded in selective medium. Olaparib reduced cell viability (measured as cfu after drug treatment and transfection) by ~50% compared to transfection alone, and PJ34 reduced viability by >90% (Figure 3B). This result suggests that the two inhibitors have different toxicity to cells and may work through different mechanisms. PJ34 inhibition of PARP1 significantly reduced integration efficiency in both HT1080 and HT1904 cells (Figure 3A). In contrast, inhibition of PARP1 with Olaparib resulted in a modest (~2-fold) increase in integration efficiency (Figure 3C). Similar increase in integration was also seen with Olaparib treatment of HT1080 cells transfected with BglII-digested pEGFP-N1-yHIS1

(Figure 1) and selected with G418 (Figure 3D). It is surprising that the two inhibitors had opposite effects on DNA integration. This data implies that the mode of action of the two PARP inhibitors is different, and/or one or both inhibitors may have off target effects on other proteins that also affect DNA integration.

PARP1 promotes plasmid integration

To further investigate the role of PARP1 in DNA integration and help resolve the contrasting results between Olaparib and PJ34, we used shRNA to establish PARP1 knockdown cell line in HT1080 cells. Single clones of HT1080shPARP cells were isolated and knockdown of PARP1 was confirmed using western blotting (Figure 4A). Two PARP1 knockdown colonies (HT1080shPARP-C4, C5) were selected because they showed the greatest reduction in PARP1 expression, ~30-40% of parent HT1080 cells (Figure 4B). Linear pEGFP-N1-yHIS1 plasmid (Figure 1) was transfected into HT1080 and PARP1 knockdown cells using Lipofectamine 2000 and selected in 450μg/ml of G418. Integration efficiency was scored as number of G418 resistant colonies per live cell. The integration efficiency of pEGFP-N1-yHIS1 in the two HT1080 shPARP1 clones was similar, ~50% of the parent HT1080 cells (Figure 4C). The reduced DNA integration in PARP1 knockdown cells supports our hypothesis that PARP1 promotes DNA integration.

DISCUSSION

PARP1 mediates generation of DSB-induced chromosomal translocations through aNHEJ pathway in murine cells (Wray et al., 2013). Chromosomal translocation connects two ends from different broken chromosomes. This process is similar the DNA integration where ends of foreign DNA insert into a chromosome, very likely at a site with an existing DNA lesion. We hypothesized that PARP1-mediated aNHEJ is also responsible for a portion of DNA

integration. To test this hypothesis, we first examined if PARP1 inhibitors suppresses DNA integration in human cells as they suppressed chromosomal translocations in murine cells (Wray et al., 2013). Surprisingly, we found that the two PARP1 inhibitors selected had opposite effects on DNA integration: PJ34 suppressed plasmid integration as expected, while Olaparib moderately promoted plasmid integration. To resolve this conflict, we established a stable PARP1 knockdown cell line via PARP1-specific shRNA and tested plasmid integration. The result demonstrated that PARP1 knockdown resulted in decreased plasmid integration compared to parent cells. We conclude that PARP1 promotes DNA integration and that PARP1 inhibitors may have off target effects that result in altered effects on DNA integration.

Both knockdown of PARP1 with shRNA and PARP1 inhibitor PJ34 suppressed plasmid integration (Figure 4C and 3A), supporting our hypothesis that PARP1 promotes DNA integration, possibly via its role in aNHEJ. PARP1 accumulates at DSBs and interacts with (Nibrin) NBS1 and MRE11, and this is required for rapid assembly of the MRN complex at DSB sites (Haince et al., 2008). MRE11 is an endonuclease required for initial end processing in aNHEJ (Rass et al., 2009) and homologous recombination (HR) (Williams et al., 2008). MRN complex regulates recruitment of CtIP and its binding partner BRCA1 (breast cancer 1, early onset) (Chen et al., 2008). MRN-CtIP-BRCA1 forms important end resection machinery to regulate aNHEJ and HR (Badie et al., 2015; Chen et al., 2008; Hu et al., 2014; Li and Yu, 2013; Limbo et al., 2007; Nakamura et al., 2010; Rass et al., 2009; Yu and Chen, 2004; Zhang and Jasin, 2011). PARP1 interacts with X-ray repair cross-complementing protein (XRCC1) and Ligase III and is required for end annealing step of aNHEJ (Audebert et al., 2004; Mansour et al., 2010; Masson et al., 1998). Alternative NHEJ is originally thought to be back up repair pathway when the cNHEJ is defective or compromised, but now considered active even when cNHEJ is

active, albeit less efficient than cNHEJ (Bennardo et al., 2008; Cortizas et al., 2013; Iliakis, 2009; Iliakis et al., 2004; Mladenov and Iliakis, 2011). The competition between KU and PARP1 for DSB binding and the type of DSB are likely to determine repair by cNHEJ or aNHEJ (Bennardo et al., 2008; Cortizas et al., 2013; Wang et al., 2006). It is likely that upon entering the cytoplasm, the linear plasmid appears to be DSBs to the cell. The large amount of "DSBs" activate both classical and alternative NHEJ to "repair" these broken ends. Binding of KU leads to integration through cNHEJ, while binding of PARP1 leads to integration through aNHEJ by initiating end resection (Mansour et al., 2010; Robert et al., 2009). In our data, the shRNA reduced PARP1 expression to 30-40% of the parent cells (Figure 4B). The 50% reduction of DNA integration in HT1080shPARP1 cells may be largely due to the attenuation of alternative NHEJ pathway related to PARP1 knockdown. Treatment of PJ34 resulted in more reduction on DNA integration (Figure 3A), implying that PJ34 may have other effects other than inhibition of poly (ADP-ribose) synthesis by PARP1. For example, shRNA is only specific to PARP1, but PJ34 is equally potent to PARP1 and PARP2. The role of PARP2 in DSB repair is poorly understood, but it has been shown that PARP2 suppresses lgH and c-myc translocation during immunoglobulin class switch recombination (Robert et al., 2009). PJ34 causes mitotic catastrophe in cancer cells with extra centrosomes by preventing the bi-polar clustering of extra centrosomes, which facilitates chromosome segregation of cancer cells (Castiel et al., 2011); attenuates AKT-associated phosphorylation of FOXO3A transcription factor (Wang et al., 2011); and partially inhibits expression of tumor necrosis factor-alpha (TNF-α), interleukin-6 (IL-6) and intercellular adhesion molecule-1 (ICAM-1) (Haddad et al., 2006). Therefore, more work is needed to gain comprehensive understanding of pathways and cellular processed affected by PJ34 treatment.

Unlike PJ34 or PARP1-specific shRNA, Olaparib treatment surprisingly increased plasmid integration (Figure 3C and 3D). Cell survival was not impacted by PARP1 knockdown, but markedly reduced by PJ34 and moderately reduced by Olaparib treatments (Figure 3B). Literature that directly compares dose effect of PJ34 and Olaparib was not found. In our project, we tried both 10 μM and 50 μM of PJ34. The lower concentration (10 μM) did not have effect on cell survival or DNA integration (data not shown). The higher concentration (50 μM) resulted in severe cell death and significant decrease in DNA integration (Figure 3A and 3B). PJ34 inhibits PARP activity in dose-dependent and cell-dependent manner. It has been shown that at 40 μM, PJ34 reduced PARP activity to 10% of its normal level in neuroblastoma cells and to 40% in its normal tissue counterpart (McCluskey et al., 2012). The HT1080 cells used in this project are fibrosarcoma cells. It is rational to reason that at 50 μM PJ34 was able to reduce PARP activity to about 10% of untreated cells, if not lower than that. This may results in lower poly (ADP-ribose) synthesis than in PARP1 knockdown cells (Figure 4B). This is consistent with the lower level of DNA integration observed with PJ34 treatment. The concentration and treatment protocol of Olaparib was chosen based on published work showing 3 μM of Olaparib suppressed PARP1 activity about 3-fold and significantly reduced chromosomal translocations (Wray et al., 2013). According to others, 3 μM of Olaparib is sufficient to bring PARP activity to below 10% (Murai et al., 2014).

With the discussion above, it is reasonable that under our experimental conditions, both PJ34 and Olaparib suppressed PARP1 catalytic activity to less than 10%, the opposite outcomes of DNA integration implies that mode of action and off target effects of the two inhibitors should help us interpret this puzzling result. As an ADP-ribosyltransferases (ART), PARP1 cleaves nicotinamide adenine dinucleotide (NAD^+) to produce ADP-ribose (PAR) units and continuously

transfer PAR units to the acceptor residue of target proteins. The catalytic domain of PARP1 is highly conversed among the PARP family. The catalytic domain contains a conserved ART domain that includes a donor site and an acceptor site; the donor site binds to NAD^+ and the acceptor site is where PAR polymer chains are extended. The nicotinamide-binding pocket (NI site) in the donor site is the major target for most PARP inhibitors, including PJ34 and Olaparib (reviewed in Steffen et al., 2013). Nicotinamide inhibits poly(ADP-ribosyl)ation activity of PARPs by competing for the binding site with NAD^+; for PARP inhibitor development, nicotinamide derivatives are designed to bind sites outside of the NI site or the donor site to improve potency and selectivity of the drug (reviewed in Ferraris, 2010; Steffen et al., 2013). PJ34 is a phenanthridine-based chemical (Figure 2A) and Olaparib is a pthalazinone-based chemical (Figure 2B); both classes of chemicals are bicyclic lactam derivatives that mimick nicotinamide (reviewed in Ferraris, 2010). Both PJ34 and Olaparib inhibit PARP1 and PARP2, as the structures of their donor sites are very similar (Pellicciari et al., 2008; Steffen et al., 2013). PJ34 inhibits PARP1 with an IC_{50} of 600 nM and PARP2 with an IC_{50} of 1000 nM (Pellicciari et al., 2008; Soriano et al., 2001); Olaparib inhibits PARP1 with an IC_{50} of 5 nM, PARP2 with an IC_{50} of 1 nM (Menear et al., 2008). Therefore, PJ34 and Olaparib have different selectivity towards PARP1 and PARP2.

Aside from catalytic inhibition of PARP1 and PARP2, Olaparib traps PARP1 and PARP2 on DSBs, in fact, trapping of the enzymes to DSBs is more cytotoxic to cells than unrepaired single-strand damage caused by catalytic inhibition of PARP1 activity (Murai et al., 2012; Murai et al., 2014). The stabilized DNA-PARP1 complexes have to be repaired for cell survival. Repair of this type of damage requires several repair pathways (Murai et al., 2012). It is possible that the stabilized DNA-PARP1 complexes stimulated DSB repair-mediated DNA integration. We

speculate that both PJ34 and Olaparib treatment cause partial reduction of DNA integration due to loss of PARP1-mediated aNHEJ. But with Olaparib treatment, not PJ34, more DNA integration is stimulated by DNA-PARP1 stabilization-induced damage. The overall outcome represents balanced result between these two opposite effects, hence moderate increase of DNA integration was observed under Olaparib treatment. It is unknown whether PJ34 also traps PARP1 to DSB as does Olaparib. Further experiments need to be performed to determine if this model also applies to PJ34.

In summary, we have presented evidence that PARP1 is involved in DNA integration. However, more work is needed to reveal the mechanisms underneath the different impacts of PARP1 knockdown and different PARP inhibitors on DNA integration. With the newly developed CRISPR technology, it is possible to examine effects of PARP1 knockout on various biological end points, including DNA integration. To better interpret the data, it is necessary to better understand effects of PARP1 inhibitors other than catalytic inhibition of PARP1. This can be achieved by treating PARP1 knockdown or knockout cells with PARP inhibitors of interest. It is also helpful to examine if cellular PARP1 is activated upon entering of foreign DNA. PARP1 is mainly involved in the initial steps of aNHEJ by facilitating end resections. It would be interesting to map integration sites of integration products (methods described in Chapter 2) under PARP1 inhibition or depletion to confirm if the ratio of DNA integration junctions with microhomology, particularly microhomology longer than 3 nucleotides would be reduced or even eliminated. Besides, it would be informative to determine if knockdown of other key aNHEJ proteins MRE11, CtIP, Ligase III would suppress DNA integration and reduce microhomology usage at integration junctions. These experiments would provide more confirmative evidence for aNHEJ-mediated DNA integration.

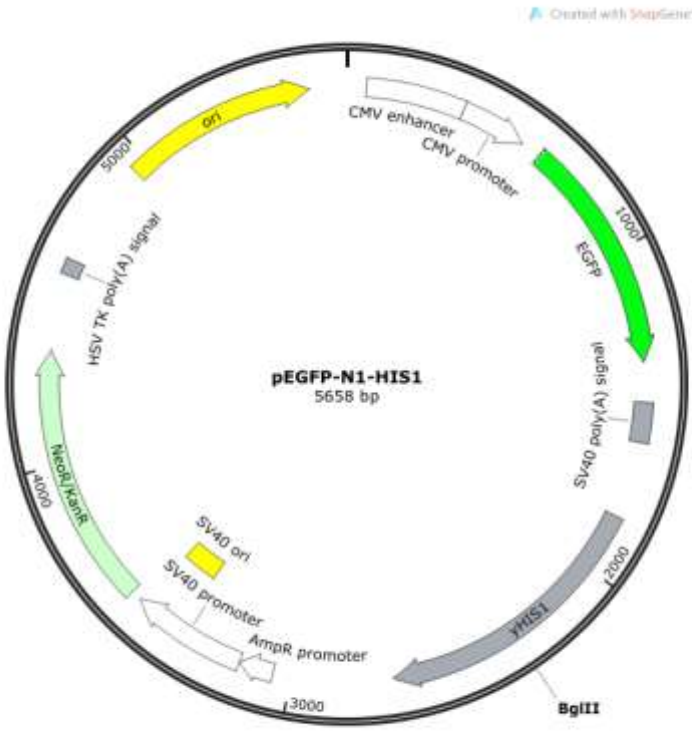

Figure 1. Structure of pEGFP-N1-yHIS1 plasmid
The plasmid was constructed by inserting HIS1 gene from yeast into pEGFP-N1 plasmid. This extra piece of DNA provides a buffering region to protect important features on the plasmid during integration. The BglII site indicates the site where the plasmid was linearized before transfection.

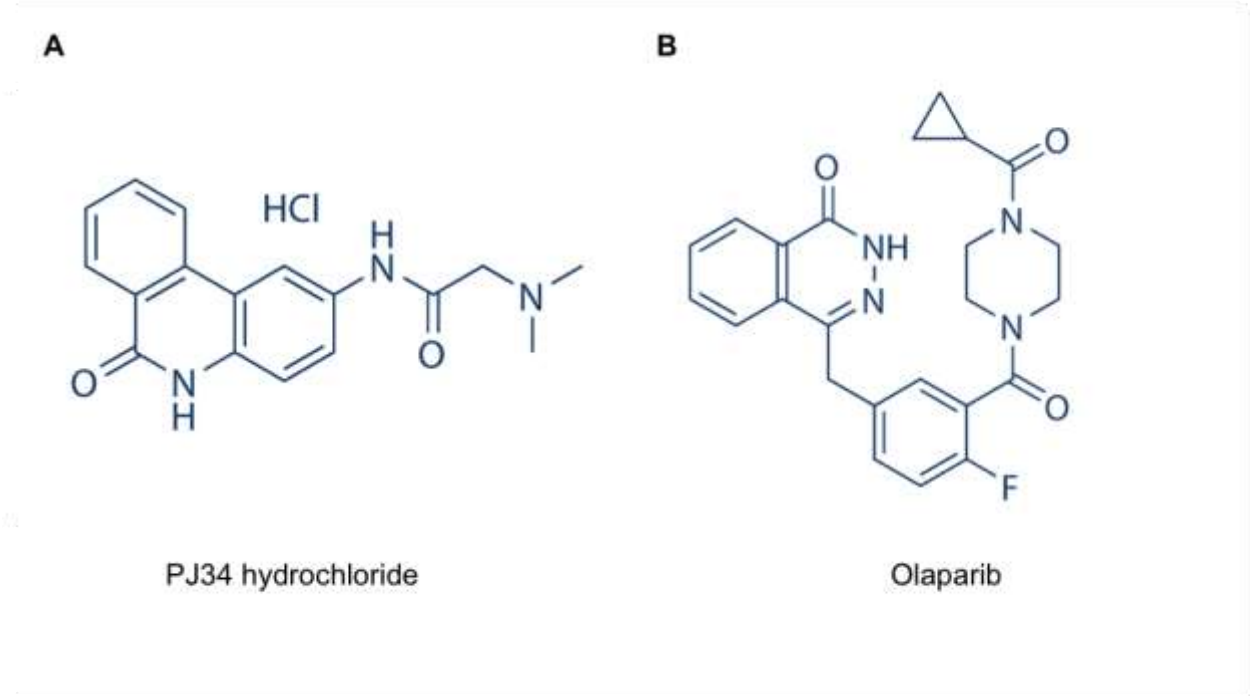

Figure 2. Chemical structure of PJ34 and Olaparib
(A) PJ34 hydorchloride belongs to phenanthridine family. It inhibits PARP1 and PARP2 with undefined selectivity.
(B) Olaparib belongs to pthalazinone family. It inhibits PARP1, PARP2 and tankyrase 1 with 5 fold selectivity for PARP2 vs. PARP1.

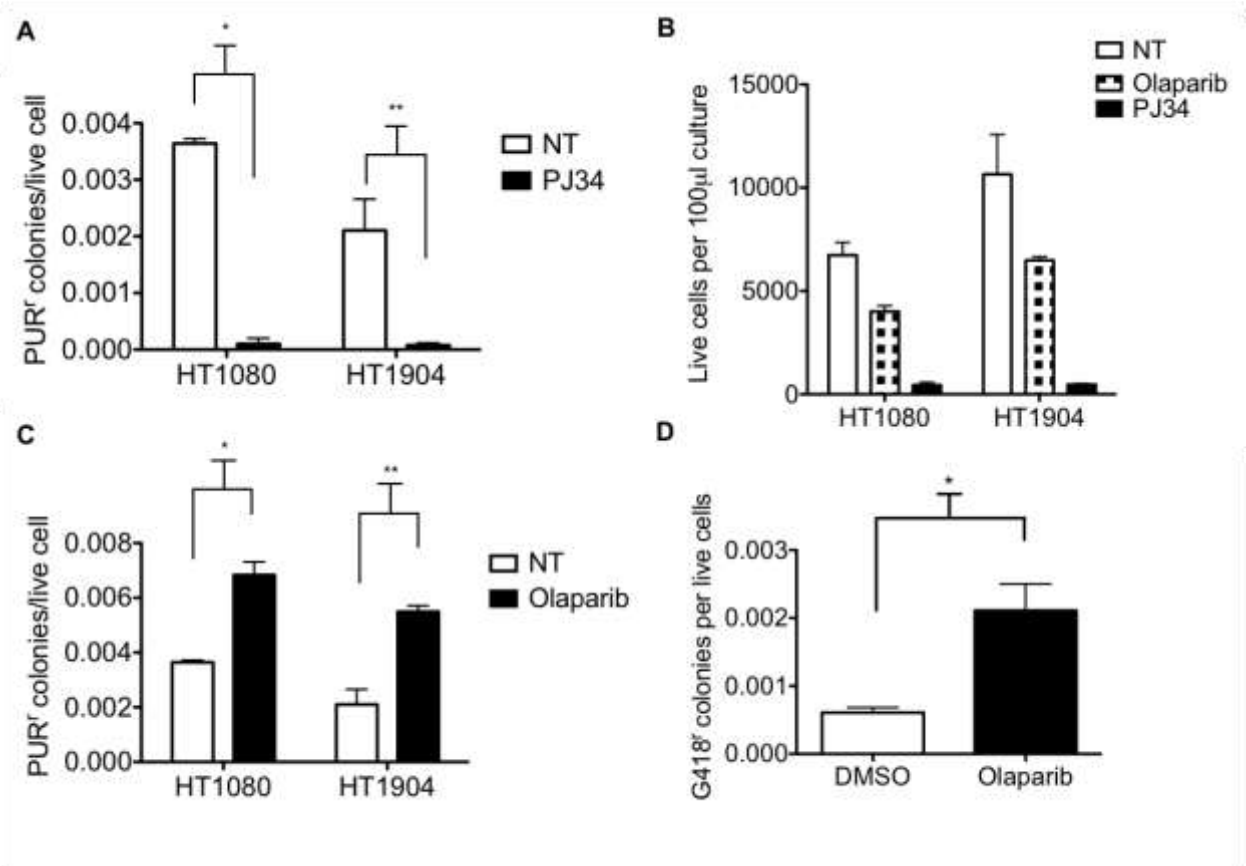

Figure 3. Effects of PARP1 inhibitors on integration efficiency

(A) HT1080 and HT1904 cells were treated with 50µM PJ34 16 hours before and 24 hours after transfection of linear pPUR. Integration efficiency was measured as number of PUR^r colonies per live cell. *P-value<0.0001 for HT1080 cells, **P-value=0.0218 for HT1904 cells (n=3 per group).

(B) HT1080 and HT1904 cells were plated without drug selection 24 hours after transfection. Cell viability (cfu) were calculated.

(C) HT1080 and HT1904 cells were treated with 3µM Olaparib 16 hours before and 24 hours after transfection of linear pPUR. Integration efficiency was measured as in (A). *P-value=0.0027 for HT1080 cells, **P-value=0.0047 for HT1904 cells (n=3 per group).

(D) HT1080 cells were treated with 3µM of Olaparib 16 hours before and 24 hours after transfection of linear pEGFP-N1-yHIS1. Integration efficiency was measured as number of $G418^r$ colonies per live cell. *P-value=0.0198 (n=3 per group).

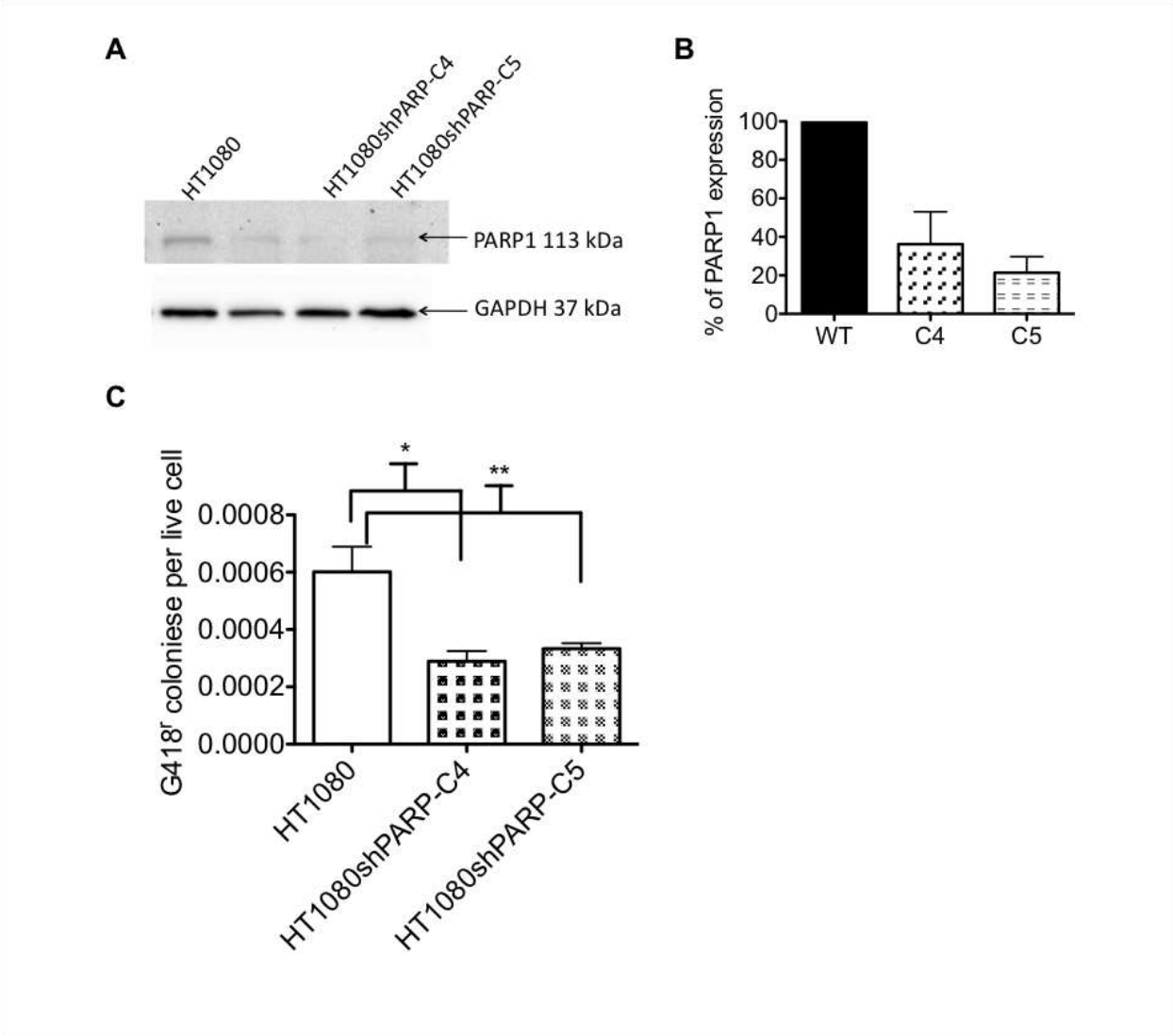

Figure 4. Integration efficiency in PARP1 knockdown cells
(A) and (B) PARP1 knockdown in each clone of HT1080shPARP was confirmed using western blotting (A). Intensity of bands were quantified using ImageLab (B).
(C) Integration efficiency of pEGFP-N1-yHIS1 plasmid was measured as number of $G418^r$ colonies per live cell in HT1080 and HT1080shPARP1. *P-value=0.0310 comparing HT1080shPARP-C4 to HT1080, **P-value=0.0417 comparing HT1080shPARP-C5 to HT1080 (n=3 per group).

REFERENCES

Ain, Q.U., Chung, J.Y., and Kim, Y.H. (2014). Current and future delivery systems for engineered nucleases: ZFN, TALEN and RGEN. J Control Release.

Ame, J.C., Spenlehauer, C., and de Murcia, G. (2004). The PARP superfamily. Bioessays *26*, 882-893.

Audebert, M., Salles, B., and Calsou, P. (2004). Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. J Biol Chem *279*, 55117-55126.

Badie, S., Carlos, A.R., Folio, C., Okamoto, K., Bouwman, P., Jonkers, J., and Tarsounas, M. (2015). BRCA1 and CtIP promote alternative non-homologous end-joining at uncapped telomeres. Embo J.

Beck, C., Robert, I., Reina-San-Martin, B., Schreiber, V., and Dantzer, F. (2014). Poly(ADP-ribose) polymerases in double-strand break repair: Focus on PARP1, PARP2 and PARP3. Exp Cell Res *329*, 18-25.

Bennardo, N., Cheng, A., Huang, N., and Stark, J.M. (2008). Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. PLoS Genet *4*, e1000110.

Boboila, C., Jankovic, M., Yan, C.T., Wang, J.H., Wesemann, D.R., Zhang, T., Fazeli, A., Feldman, L., Nussenzweig, A., Nussenzweig, M., *et al.* (2010a). Alternative end-joining catalyzes robust IgH locus deletions and translocations in the combined absence of ligase 4 and Ku70. Proc Natl Acad Sci U S A *107*, 3034-3039.

Boboila, C., Yan, C., Wesemann, D.R., Jankovic, M., Wang, J.H., Manis, J., Nussenzweig, A., Nussenzweig, M., and Alt, F.W. (2010b). Alternative end-joining catalyzes class switch recombination in the absence of both Ku70 and DNA ligase 4. J Exp Med *207*, 417-427.

Byrne, M., Wray, J., Reinert, B., Wu, Y., Nickoloff, J., Lee, S.H., Hromas, R., and Williamson, E. (2014). Mechanisms of oncogenic chromosomal translocations. Ann N Y Acad Sci *1310*, 89-97.

Castiel, A., Visochek, L., Mittelman, L., Dantzer, F., Izraeli, S., and Cohen-Armon, M. (2011). A phenanthrene derived PARP inhibitor is an extra-centrosomes de-clustering agent exclusively eradicating human cancer cells. BMC Cancer *11*, 412.

Chen, L., Nievera, C.J., Lee, A.Y., and Wu, X. (2008). Cell cycle-dependent complex formation of BRCA1.CtIP.MRN is important for DNA double-strand break repair. J Biol Chem *283*, 7713-7720.

Cortizas, E.M., Zahn, A., Hajjar, M.E., Patenaude, A.M., Di Noia, J.M., and Verdun, R.E. (2013). Alternative end-joining and classical nonhomologous end-joining pathways repair different types of double-strand breaks during class-switch recombination. J Immunol *191*, 5751-5763.

De Vos, M., Schreiber, V., and Dantzer, F. (2012). The diverse roles and clinical relevance of PARPs in DNA damage repair: current state of the art. Biochem Pharmacol *84*, 137-146.

Delacote, F., Han, M., Stamato, T.D., Jasin, M., and Lopez, B.S. (2002). An xrcc4 defect or Wortmannin stimulates homologous recombination specifically induced by double-strand breaks in mammalian cells. Nucleic Acids Res *30*, 3454-3463.

Della-Maria, J., Zhou, Y., Tsai, M.S., Kuhnlein, J., Carney, J.P., Paull, T.T., and Tomkinson, A.E. (2011). Human Mre11/human Rad50/Nbs1 and DNA ligase IIIalpha/XRCC1 protein complexes act together in an alternative nonhomologous end joining pathway. J Biol Chem *286*, 33845-33853.

Deriano, L., and Roth, D.B. (2013). Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. Annu Rev Genet *47*, 433-455.

Ferraris, D.V. (2010). Evolution of poly(ADP-ribose) polymerase-1 (PARP-1) inhibitors. From concept to clinic. Journal of medicinal chemistry *53*, 4561-4584.

Fnu, S., Williamson, E.A., De Haro, L.P., Brenneman, M., Wray, J., Shaheen, M., Radhakrishnan, K., Lee, S.H., Nickoloff, J.A., and Hromas, R. (2011). Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. Proc Natl Acad Sci U S A *108*, 540-545.

Gagne, J.P., Isabelle, M., Lo, K.S., Bourassa, S., Hendzel, M.J., Dawson, V.L., Dawson, T.M., and Poirier, G.G. (2008). Proteome-wide identification of poly(ADP-ribose) binding proteins and poly(ADP-ribose)-associated protein complexes. Nucleic Acids Res *36*, 6959-6976.

Ghezraoui, H., Piganeau, M., Renouf, B., Renaud, J.B., Sallmyr, A., Ruis, B., Oh, S., Tomkinson, A.E., Hendrickson, E.A., Giovannangeli, C., *et al.* (2014). Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. Mol Cell *55*, 829-842.

Haddad, M., Rhinn, H., Bloquel, C., Coqueran, B., Szabo, C., Plotkine, M., Scherman, D., and Margaill, I. (2006). Anti-inflammatory effects of PJ34, a poly(ADP-ribose) polymerase inhibitor, in transient focal cerebral ischemia in mice. Br J Pharmacol *149*, 23-30.

Haince, J.F., McDonald, D., Rodrigue, A., Dery, U., Masson, J.Y., Hendzel, M.J., and Poirier, G.G. (2008). PARP1-dependent kinetics of recruitment of MRE11 and NBS1 proteins to multiple DNA damage sites. J Biol Chem *283*, 1197-1208.

Hu, Y., Petit, S.A., Ficarro, S.B., Toomire, K.J., Xie, A., Lim, E., Cao, S.A., Park, E., Eck, M.J., Scully, R., *et al.* (2014). PARP1-driven poly-ADP-ribosylation regulates BRCA1 function in homologous recombination-mediated DNA repair. Cancer Discov *4*, 1430-1447.

Iliakis, G. (2009). Backup pathways of NHEJ in cells of higher eukaryotes: cell cycle dependence. Radiother Oncol *92*, 310-315.

Iliakis, G., Wang, H., Perrault, A.R., Boecker, W., Rosidi, B., Windhofer, F., Wu, W., Guan, J., Terzoudi, G., and Pantelias, G. (2004). Mechanisms of DNA double strand break repair and chromosome aberration formation. Cytogenet Genome Res *104*, 14-20.

Jeyaratnam, D.C., Baduin, B.S., Hansen, M.C., Hansen, M., Jorgensen, J.M., Aggerholm, A., Ommen, H.B., Hokland, P., and Nyvold, C.G. (2014). Delineation of known and new transcript variants of the SETMAR (Metnase) gene and the expression profile in hematologic neoplasms. Exp Hematol *42*, 448-456 e444.

Kabotyanski, E.B., Gomelsky, L., Han, J.O., Stamato, T.D., and Roth, D.B. (1998). Double-strand break repair in Ku86- and XRCC4-deficient cells. Nucleic Acids Res *26*, 5333-5342.

Lee-Theilen, M., Matthews, A.J., Kelly, D., Zheng, S., and Chaudhuri, J. (2011). CtIP promotes microhomology-mediated alternative end joining during class-switch recombination. Nat Struct Mol Biol *18*, 75-79.

Li, M., and Yu, X. (2013). Function of BRCA1 in the DNA damage response is mediated by ADP-ribosylation. Cancer Cell *23*, 693-704.

Liang, F., and Jasin, M. (1996). Ku80-deficient cells exhibit excess degradation of extrachromosomal DNA. J Biol Chem *271*, 14405-14411.

Liang, F., Romanienko, P.J., Weaver, D.T., Jeggo, P.A., and Jasin, M. (1996). Chromosomal double-strand break repair in Ku80-deficient cells. Proc Natl Acad Sci U S A *93*, 8929-8933.

Limbo, O., Chahwan, C., Yamada, Y., de Bruin, R.A., Wittenberg, C., and Russell, P. (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*, 134-146.

Mansour, W.Y., Rhein, T., and Dahm-Daphi, J. (2010). The alternative end-joining pathway for repair of DNA double-strand breaks requires PARP1 but is not dependent upon microhomologies. Nucleic Acids Res *38*, 6065-6077.

Masson, M., Niedergang, C., Schreiber, V., Muller, S., Menissier-de Murcia, J., and de Murcia, G. (1998). XRCC1 is specifically associated with poly(ADP-ribose) polymerase and negatively regulates its activity following DNA damage. Mol Cell Biol *18*, 3563-3571.

Mattarucchi, E., Guerini, V., Rambaldi, A., Campiotti, L., Venco, A., Pasquali, F., Lo Curto, F., and Porta, G. (2008). Microhomologies and interspersed repeat elements at genomic breakpoints in chronic myeloid leukemia. Genes Chromosomes Cancer *47*, 625-632.

McCluskey, A.G., Mairs, R.J., Tesson, M., Pimlott, S.L., Babich, J.W., Gaze, M.N., Champion, S., and Boyd, M. (2012). Inhibition of poly(ADP-Ribose) polymerase enhances the toxicity of 131I-metaiodobenzylguanidine/topotecan combination therapy to cells and xenografts that express the noradrenaline transporter. J Nucl Med *53*, 1146-1154.

Menear, K.A., Adcock, C., Boulter, R., Cockcroft, X.L., Copsey, L., Cranston, A., Dillon, K.J., Drzewiecki, J., Garman, S., Gomez, S., *et al.* (2008). 4-[3-(4-cyclopropanecarbonylpiperazine-1-carbonyl)-4-fluorobenzyl]-2H-phthalazin- 1-one: a novel bioavailable inhibitor of poly(ADP-ribose) polymerase-1. Journal of medicinal chemistry *51*, 6581-6591.

Mladenov, E., and Iliakis, G. (2011). Induction and repair of DNA double strand breaks: the increasing spectrum of non-homologous end joining pathways. Mutat Res *711*, 61-72.

Murai, J., Huang, S.Y., Das, B.B., Renaud, A., Zhang, Y., Doroshow, J.H., Ji, J., Takeda, S., and Pommier, Y. (2012). Trapping of PARP1 and PARP2 by Clinical PARP Inhibitors. Cancer Res *72*, 5588-5599.

Murai, J., Huang, S.Y., Renaud, A., Zhang, Y., Ji, J., Takeda, S., Morris, J., Teicher, B., Doroshow, J.H., and Pommier, Y. (2014). Stereospecific PARP trapping by BMN 673 and comparison with olaparib and rucaparib. Mol Cancer Ther *13*, 433-443.

Nakamura, K., Kogame, T., Oshiumi, H., Shinohara, A., Sumitomo, Y., Agama, K., Pommier, Y., Tsutsui, K.M., Tsutsui, K., Hartsuiker, E., *et al.* (2010). Collaborative action of Brca1 and CtIP in elimination of covalent modifications from double-strand breaks to facilitate subsequent break repair. PLoS Genet *6*, e1000828.

Pellicciari, R., Camaioni, E., Costantino, G., Formentini, L., Sabbatini, P., Venturoni, F., Eren, G., Bellocchi, D., Chiarugi, A., and Moroni, F. (2008). On the way to selective PARP-2 inhibitors. Design, synthesis, and preliminary evaluation of a series of isoquinolinone derivatives. ChemMedChem *3*, 914-923.

Rass, E., Grabarz, A., Plo, I., Gautier, J., Bertrand, P., and Lopez, B.S. (2009). Role of Mre11 in chromosomal nonhomologous end joining in mammalian cells. Nat Struct Mol Biol *16*, 819-824.

Robert, I., Dantzer, F., and Reina-San-Martin, B. (2009). Parp1 facilitates alternative NHEJ, whereas Parp2 suppresses IgH/c-myc translocations during immunoglobulin class switch recombination. J Exp Med *206*, 1047-1056.

Simsek, D., Brunet, E., Wong, S.Y., Katyal, S., Gao, Y., McKinnon, P.J., Lou, J., Zhang, L., Li, J., Rebar, E.J., *et al.* (2011). DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. PLoS Genet *7*, e1002080.

Simsek, D., and Jasin, M. (2010). Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. Nat Struct Mol Biol *17*, 410-416.

Soriano, F.G., Virag, L., and Szabo, C. (2001). Diabetic endothelial dysfunction: role of reactive oxygen and nitrogen species production and poly(ADP-ribose) polymerase activation. Journal of molecular medicine *79*, 437-448.

Sousa, F.G., Matuo, R., Soares, D.G., Escargueil, A.E., Henriques, J.A., Larsen, A.K., and Saffi, J. (2012). PARPs and the DNA damage response. Carcinogenesis *33*, 1433-1440.

Steffen, J.D., Brody, J.R., Armen, R.S., and Pascal, J.M. (2013). Structural Implications for Selective Targeting of PARPs. Frontiers in oncology *3*, 301.

Vyas, S., Chesarone-Cataldo, M., Todorova, T., Huang, Y.H., and Chang, P. (2013). A systematic analysis of the PARP protein family identifies new functions critical for cell physiology. Nature communications *4*, 2240.

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

Wang, S., Wang, H., Davis, B.C., Liang, J., Cui, R., Chen, S.J., and Xu, Z.X. (2011). PARP1 inhibitors attenuate AKT phosphorylation via the upregulation of PHLPP1. Biochem Biophys Res Commun *412*, 379-384.

Wei, K., Kodym, R., and Jin, C. (1998). Radioresistant cell strain of human fibrosarcoma cells obtained after long-term exposure to x-rays. Radiat Environ Biophys *37*, 133-137.

Weinstock, D.M., Brunet, E., and Jasin, M. (2007). Formation of NHEJ-derived reciprocal chromosomal translocations does not require Ku70. Nature cell biology *9*, 978-981.

Williams, R.S., Moncalian, G., Williams, J.S., Yamada, Y., Limbo, O., Shin, D.S., Groocock, L.M., Cahill, D., Hitomi, C., Guenther, G., *et al.* (2008). Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair. Cell *135*, 97-109.

Williamson, E.A., Rasila, K.K., Corwin, L.K., Wray, J., Beck, B.D., Severns, V., Mobarak, C., Lee, S.H., Nickoloff, J.A., and Hromas, R. (2008). The SET and transposase domain protein Metnase enhances chromosome decatenation: regulation by automethylation. Nucleic Acids Res *36*, 5822-5831.

Wray, J., Williamson, E.A., Chester, S., Farrington, J., Sterk, R., Weinstock, D.M., Jasin, M., Lee, S.H., Nickoloff, J.A., and Hromas, R. (2010). The transposase domain protein Metnase/SETMAR suppresses chromosomal translocations. Cancer Genet Cytogenet *200*, 184-190.

Wray, J., Williamson, E.A., Singh, S.B., Wu, Y., Cogle, C.R., Weinstock, D.M., Zhang, Y., Lee, S.H., Zhou, D., Shao, L., *et al.* (2013). PARP1 is required for chromosomal translocations. Blood *121*, 4359-4365.

Wurtele, H., Little, K.C., and Chartrand, P. (2003). Illegitimate DNA integration in mammalian cells. Gene Ther *10*, 1791-1799.

Yu, X., and Chen, J. (2004). DNA damage-induced cell cycle checkpoint control requires CtIP, a phosphorylation-dependent binding partner of BRCA1 C-terminal domains. Mol Cell Biol *24*, 9478-9486.

Zhang, Y., and Jasin, M. (2011). An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. Nat Struct Mol Biol *18*, 80-84.

Zhang, Y., and Rowley, J.D. (2006). Chromatin structural elements and chromosomal translocations in leukemia. DNA Repair (Amst) *5*, 1282-1297.

CHAPTER 4

DISCUSSION AND FUTURE DIRECTIONS

DNA INTEGRATION IS MEDIATED BY DNA REPAIR

As discussed in Chapter 1, various formats of naturally occurring DNA integration alter the recipient genomes, bring new phenotypes to the recipient organism, facilitate spread of virus, disrupt genomic stability and contribute to evolution of new species. For example, in the human genome, about 45% of bases come from activities of ancient transposable elements (Cordaux and Batzer, 2009). Most of these endogenous transposable elements are inactive, but some remains mobile and have been associated with tumorigenesis (reviewed in Kassiotis, 2014; Mullins and Linnebacher, 2012). Integration of viral DNA into the infected host genome is part of the life cycle of some viruses and retroviruses. Incorporation of some viral DNA has been associated with tumorigenesis (reviewed in Chen et al., 2014). Natural formats of DNA integration have been adapted to edit genomes in lab. Integrative vectors were engineered from plasmid, virus DNA, transposon and other mobile DNAs for various purposes. Applications of integrative vectors have contributed enormously to basic research, biotechnology development, gene therapy and other areas. Despite the commonness and importance of DNA integration, its mechanism, particularly integration of vectors that do not contain extensive homologous sequence to the recipient genome, remains to be elucidated.

Early studies revealed that a functional connection between DNA integration and repair of DNA damage. Integration of plasmid DNA can be enhanced by pretreating the plasmid with DNA damaging reagents (Hsiung et al., 1980; Spivak et al., 1984; Spivak et al., 1988), or by introducing DNA damage to recipient cells (Bodley et al., 1993; Debenham and Webb, 1984;

Fujimaki et al., 1996; Leadon et al., 1987; Manivasakam et al., 2001; Nagata et al., 1984; Nakayama et al., 1998; Perez et al., 1985; Perez and Skarsgard, 1986; Postel, 1985; Rubin, 1988; Shcherbakova and Filatov, 2000; van Duin et al., 1985; Vos and Hanawalt, 1989). These early studies showed that DNA damage introduced by endonucleases, ultraviolet light, X-rays, γ-rays, DNA crosslinking chemicals, Topoisomerase I and II inhibitors and others enhanced integration of plasmid DNA to mammalian cells. Transfection of foreign DNA into cells is similar to introducing DNA damage to cells. In plasmid transfection or virus infection, plasmid or virus DNA traverse the cell membrane and cytoplasm, directly or in vesicles such as lysosomes that are loaded with digestive enzymes. Then DNA has to traverse the nuclear membrane to reach vicinity to DNA lesions by chance or are recruited by DNA binding proteins, and gets inserted into the chromosome as part of the repair process (reviewed in Wurtele et al., 2003). Plasmid or viral DNA is often damaged by digestive enzymes in the lysomsomes or in the cytoplasm before entering the nucleus and integrating into the host chromosomes (Nickoloff et al., 1998). The free ends of foreign DNA could be detected as double-strand breaks (DSBs) and activate the cellular DNA damage response (DDR) pathways. The DDR system senses DNA damage and activates cascades of signaling pathways to activate checkpoint for cell cycle arrest, and recruit repair factors to damaged DNA. The DDR system ensures that cells with severely damaged DNA are eliminated from the population or stay senescence. Therefore, the DDR system plays a critical antitumor role by maintaining genome stability. Virus infection activates DDR signaling. Viruses use some DDR components for their early replication steps, and then inhibit DDR in later parts of their life cycles (reviewed in Chaurushiya and Weitzman, 2009; Lilley et al., 2007). The (MRE11-RAD50-NBS1) MRN complex can recognize the inverted terminal repeats of adeno-associated virus and redistribute to viral replication center and restrict viral DNA replication

(Schwartz et al., 2007). The MRN complex senses infected viral DNA as DSB and activates Ataxia telangiectasia mutated (ATM)-dependent G2/M checkpoint (Carson et al., 2003). DNA repair proteins used in homologous recombination (HR) and non-homologous end joining (NHEJ) have been shown to redistribute to the replication center of herpes simplex virus, among them, KU has been shown to prevent viral replication (Taylor and Knipe, 2004). MRN and KU are likely to promote viral DNA concatenation through NHEJ (Araujo et al., 2005; Schwartz et al., 2007; Stracker et al., 2002; Taylor and Knipe, 2004), as a cellular defense mechanism of virus infection. It is possible that DNA repair proteins interact with transfected or infected DNA in the cytoplasm and escort it to the nuclear target. Requirement of DNA repair system for DNA integration was confirmed in yeast, plant and mammalian cells (Bertolini et al., 2009; Ishibashi et al., 2006; Manivasakam et al., 2001; Rubin, 1988; Tanaka et al., 2010). Key HR or NHEJ proteins have been shown to play important roles in DNA integration. In *Neurospora*, homology-directed integration requires MEI-3 (yeast RAD51 homolog) and homology-independent integration requires MUS-53 (human Ligase IV homolog) (Ishibashi et al., 2006). In yeast, RAD52 and RAD 59 are required for RAD51-independent integration through recombination between direct repeats via single-strand annealing (SSA) mechanism (Mott and Symington, 2011). Deficiency of KU70, X-ray repair cross-complementing protein 4 (XRCC4) or Ligase IV increased ratio of targeted integration in human cells and plant cells (Bertolini et al., 2009; Tanaka et al., 2010). In Chinese hamster ovary cells, restriction enzyme-induced DSBs stimulate DNA integration, a process dependent on functional KU80 proteins (Manivasakam et al., 2001). Therefore, DSB repair systems are the major machinery that mediates DNA integration.

In correspondence to the two types of DSB repair pathways, HR and NHEJ, DNA integration can be divided into homology-directed integration and homology-independent

integration. HR-mediated homology-directed integration is favored for genome editing or gene therapy. Technologies were developed to integrate foreign DNA at specific sites by introducing site-specific DSBs (reviewed in Ain et al., 2014). Until the development of mammalian genome editing with CRISPR/Cas9 system, precise targeting was accessible only in yeast and a few other organisms, because of the dominance of random (NHEJ-like) integration in mammalian cells. CRISPR/Cas9 system goes a long ways toward eliminating that hurdle-about 3 decades after routine use in yeast. Even though CRISPR/Cas9 increases targeting efficiency, it does not suppress NHEJ. So genome editing will always be plagued by random integration. Deep understanding of the mechanisms of DNA integration allows us to improve genome editing by enhancing targeted DNA integration and suppressing random integration.

In this dissertation, we investigated the roles of two DNA repair proteins Metnase and PARP1 in DNA integration. Metnase was first discovered as an NHEJ and DNA integration promoting protein in 2005 (Lee et al., 2005), later was shown to function in multiple cellular functions, including decatenation and replication fork restart (reviewed in Shaheen et al., 2010) and as a downstream effector of checkpoint kinase 1 (Chk1) in response to DNA damage (Hromas et al., 2012). PARP1 is another DNA repair protein that involve in repair of multiple types of DNA damages (reviewed in Javle and Curtin, 2011). PARP1 defects cause single-strand damage to accumulate, increasing fork collisions with lesions/fork collapse to DSBs which are natural substrates for integration of foreign DNA, and it regulates alternative non-homologous end joining (aNHEJ), one of several mechanisms by which DNA might integrate. This work expands our knowledge on roles of Metnase and PARP1 in random plasmid integration, and generates insights on involvement of classical non-homologous end joining (cNHEJ) and aNHEJ

in DNA integration. With further work, it is possible to improve targeted DNA integration or gene targeting by manipulating the activities of proteins such as Metnase and PARP1.

METNASE PROMOTES THE OVERALL EFFICIENCY OF DNA INTEGRATION

Metnase is a fusion protein that is only expressed in higher primates, including human (Cordaux et al., 2006; Lee et al., 2005). Metnase possesses automethylation, histone methylation, human transposon *Hsmar1* terminal inverted repeat (TIR)-specific DNA binding, and DNA cleavage activities (Beck et al., 2008; Cordaux et al., 2006; Lee et al., 2005; Liu et al., 2007; Roman et al., 2007). Metnase interacts with NHEJ proteins and promotes NHEJ and DNA integration (Hromas et al., 2008; Lee et al., 2005; Williamson et al., 2008a). Metnase binds to Topoisomerase II and confers resistance to etoposide inhibition (Ponder et al., 2011; Williamson et al., 2008b; Wray et al., 2009a; Wray et al., 2009b). Metnase is a target of Chk1 and involved in checkpoint regulation in response to DNA replication stress (De Haro et al., 2010; Hromas et al., 2012). The functions of Metnase in DNA integration, DSB repair, DNA replication and checkpoint regulation link multiple DNA dynamic processes within the DDR. DDR proteins play critical roles in DNA dynamic processes, and signaling pathways that determine cell fate, under both non-threatening, and genotoxic conditions. Cell fates include apoptosis, senescence, cell cycle arrest, and mutagenesis broadly defined as genome instability, which takes many forms as it includes any sequence change to DNA, from point mutations to chromosomal translocations and whole chromosome gain/loss. Metnase and PARP1 as important DDR actors, require further study on all fronts because genome instability is a major contributing factor to carcinogenesis, as well as cancer progression to more lethal, metastatic states (reviewed in Hanks et al., 2004; Morgan and Shilatifard, 2015).

With regard to its functions in NHEJ and DNA integration, both the SET (*Su*(*var*)*3-9*, *Enhancer of Zeste*, Trithorax) and the nuclease domains of Metnase are required. The SET domain methylates histones close to DSBs and stimulates recruitment of Nibrin (NBS1) and KU70 (Fnu et al., 2011). The nuclease domain participates in trimming the 3' overhangs of broken ends, as blunt ends are preferred substrates of Ligase IV (Beck et al., 2011; Mohapatra et al., 2013). In addition, as a derivative of *Mainer* transposase, the nuclease domain of Metnase retains specific TIR binding activity (Beck et al., 2008; Roman et al., 2007). TIRs are short repetitive sequences flanking *Mariner* transposons. They are often left behind or get amplified as a result of transposition. There are an estimated 7000 copies of *Mariner* TIR sequences in the human genome due to ancient transposon movement (Liu et al., 2007). We mapped integration sites of a pPUR plasmid in HEK293T cells under low and high levels of Metnase expression. No evidence supported that Metnase employs TIR binding to stimulate integration. *In vitro*, Metnase binds specifically to the TIR sequence of human transposon *Hsmar1* (Roman et al., 2007). Based on this, we predicted Metnase would promote integration in or near *Mariner* remnants. We found instead that Metnase enhances overall plasmid integration independent of TIR sequences. This indicates that Metnase does not display TIR-specific binding *in vivo*, probably because TIR-specific binding is blocked or inhibited by its interaction with hPso4, which binds DNA in a sequence-independent manner (Beck et al., 2008). Metnase binds to hPso4 and forms a stable complex, which binds to DNA nonspecifically largely driven by hPso4 DNA binding (Beck et al., 2010). It is likely that *in vivo*, Metnase does not interact with free ends of plasmid or chromosomes directly, but in the context of one or more protein complexes. Therefore, the TIR-specific DNA binding of Metnase does not contribute to its function in DNA integration. There are other ways Metnase could stimulate integration. Metnase is recruited to DSBs by hPso4 and

dimethylates histone H3K36, which promotes recruitment of KU70 and NBS1 for NHEJ (Beck et al., 2008; Fnu et al., 2011). Recruitment of KU to ends of foreign DNA may help to bring foreign DNA to DSB sites on chromosomes, followed by DNA-PKcs binding and cNHEJ to integrate foreign DNA at DSB site. Metnase binds to blunt DSB and DSB with overhangs (Kim et al., 2014; Mohapatra et al., 2013). Trimming of the 3' overhang of DSB by the transposase domain of Metnase promotes rejoining of broken ends by Ligase IV (Mohapatra et al., 2013). Metnase also interacts with Ligase IV, likely enhancing recruitment of Ligase IV to DSBs (Hromas et al., 2008). Metnase may promote DNA integration via the same or similar mechanisms.

To further explore the mechanism of integration, we analyzed plasmid-chromosome junctions, defining junction features such as microhomology use, size of plasmid and/or chromosome deletion, and size identity, and source of insertions. In general, the integration junctions in our dataset demonstrated features that have been reported by others (Dubose et al., 2013; Merrihew et al., 1996; Uh et al., 2006; Wang et al., 2004; Yan et al., 2013; Zheng et al., 2001). Most ends of pPUR were resected during integration. We predicted that Metnase overexpression would protect the plasmid end from large deletion, similar to previous observations that Metnase promotes NHEJ and prevents large deletions of viral DNA during integration (Hromas et al., 2008). However, we did not observe this effect (Chapter 2). It is possible that end resection or other end-processing occurred during plasmid-transit to the nucleus, before it reached and interacted with chromosomes. Only about 10% of the junctions featured precise joining of plasmid and chromosome DNA. Approximately 40% of the junctions featured microhomology between plasmid and chromosome, and about 50% of junctions captured extra pieces of DNA, either from other sources, or non-templated. Metnase overexpression resulted in

shorter microhomologies (≤3 bp) at integration junctions (Chapter 2). Repair through cNHEJ utilizes Ligase IV, and does not require microhomology. Repair through aNHEJ mainly utilizes Ligase III, which acts on ends that anneal via microhomologies. However, this does not mean that all microhomology-mediated end joining occurs through aNHEJ. V(D)J recombination is a well accepted model of cNHEJ. During V(D)J recombination in human cells, about 60% of joining sites have 1 or 2 bp of microhomology (Gauss and Lieber, 1996). Ends joined independent of microhomology annealing could appear to be microhomology-mediated by chance. Computational simulation of 1000 end joining events showed that microhomologies of 1-3 bp appeared in about 40% of the junctions by chance (Newman et al., 2015). Therefore, microhomologies less than 3 bp are common outcomes of cNHEJ. It is challenging to estimate the ratio of repair through cNHEJ versus aNHEJ, since aNHEJ is usually measured in cNHEJ deficient systems. One proposal is that aNHEJ often uses extent microhomology (≥4 bp) (reviewed in Pannunzio et al., 2014). In our data, microhomology over 4 bases was only found in transfection without Metnase overexpression. This observation supports our hypothesis that Metnase promotes DNA integration via cNHEJ and agrees with the previous report that Metnase suppresses aNHEJ-mediated chromosomal translocations (Wray et al., 2010).

PARP1 MEDIATES DNA INTEGRATION VIA ALTERNATIVE NHEJ

We showed that Metnase overexpression reduced the use of microhomologies longer than 4 bp at integration junctions. Microhomologies over 4 bp are likely to arise by aNHEJ, which is initiated by limited end resection, a process mediated by *C-terminal-binding protein interacting protein* (*CtIP*), the MRN complex, PARP1 and other factors (Deriano and Roth, 2013; Zhang and Jasin, 2011). PARP1 was shown to have an important role in chromosomal translocations, due to its role in promoting aNHEJ (Wray et al., 2013), which is the major pathway of

chromosomal translocation, whereas cNHEJ prevents translocations (reviewed in Byrne et al., 2014; Deriano and Roth, 2013). However, one recent study suggested that cNHEJ is the major generator of chromosomal translocation in human cells (Ghezraoui et al., 2014). Clearly, more work is required to define the mechanism(s) of chromosomal translocation. Nonetheless, chromosomal translocation does share structural similarity with DNA integration, as both processes involve joining of two ends previously distant from each other. Our data and other integration junction studies are consistent with integration resulting from both cNHEJ and aNHEJ, with Metnase promoting cNHEJ (Chapter 2), and PARP1 promoting aNHEJ. Based on the idea that aNHEJ is an important pathway for DNA integration, and that PARP1 depletion or inhibition suppresses DSB-induced chromosomal translocations in three different translocation assays (Wray et al., 2013), we hypothesized that depletion or inhibition of PARP1 would suppress DNA integration.

To test this hypothesis, we employed two PARP inhibitors, PJ34 and Olaparib, and assessed DNA integration. Consistent with our expectation, PJ34 treatment greatly reduced integration efficiency. Surprisingly, Olaparib treatment increased integration by about 2-fold, a result verified with two different plasmids. To better interpret this result, we examined DNA integration efficiency in stable PARP1 knockdown HT1080 cells and found plasmid integration reduced about 50% compared to parent HT1080 cells (Chapter 3).

To account for these disparate PARP1 inhibitor results, we considered various features of cNHEJ and aNHEJ pathways. PARP1 is known to compete with KU for binding to DSB ends, and these proteins promote distinct repair pathways (Mansour et al., 2010; Paddock et al., 2011; Wang et al., 2006). KU and PARP1 are known to have opposite effects on DNA end resection: PARP1 promotes recruitment of MRN and CtIP which initiate limited end resection, while KU

protects ends from resection. As a result, KU binding to ends leads to cNHEJ and PARP1 binding to ends leads to aNHEJ (reviewed in Byrne et al., 2014; Oh et al., 2014). Our findings that microhomology is frequently seen at integration junctions, and that PARP1 knockdown and PARP1 inhibition with PJ34 significantly reduce DNA integration supports a model in which aNHEJ plays an important role in DNA integration. It is interesting that PJ34 treatment suppressed DNA integration much more than knockdown by shRNA. This could be explained by two not mutually exclusive ways. In one, because the shRNA did not achieve100% knockdown of PARP1, aNHEJ-mediated DNA integration was only partially compromised. Alternatively, PJ34 may have other targets in cells besides PARP1 that contribute to DNA integration; or PJ34 could suppress integration by affecting other unknown pathways.

When tested in chromosomal translocation assays, Olaparib showed similar suppression as PARP1 knockdown (Wray et al., 2013). PARP1 is very important for initiating translocation as translocation is mainly mediated by aNHEJ, which depends on PARP1. DNA integration is fundamentally similar to chromosomal translocation. But cNHEJ instead of aNHEJ should account for a large portion of DNA integration. Thus, we predicted that Olaparib would suppress DNA integration, but not as much as chromosomal translocation. Surprisingly, our data showed that Olaparib promoted DNA integration. As a PARP1inhibitor, Olaparib is effective in treating cancers with HR deficiency, but it is relatively non-toxic to normal tissues (Chen et al., 2013; Kaufman et al., 2015; Ledermann et al., 2012). PARP1 is important for repair of single-strand breaks (SSBs) and other types of single-strand damage. Inhibition of PARP1 results in accumulation of SSBs. When not resolved efficiently, SSBs hinder progression of replication forks. HR proficient cells are able to survive by resolving the stalled forks through HR, while HR deficient cells die via apoptosis. For more information of the synthetic lethality between PARP

and HR pathway, see these recent reviews (Hosoya and Miyagawa, 2014; Lee et al., 2014; Lupo and Trusolino, 2014; O'Neil et al., 2013). In addition, alternative models were proposed to explain the synthetic lethality of PARP inhibitor in HR-deficient cells. Patel *et al*. presented evidence that PARP inhibitors enhance phosphorylation of DNA-PK targets, as a result, PARP1 inhibitors induce genome instability and sensitize cells to DNA damage by promoting error-prone NHEJ in HR-deficient cells (Patel et al., 2011). Olaparib does not prevent DNA binding of PARP but traps PARP proteins on DNA (Murai et al., 2012; Murai et al., 2014). We propose the following interpretation for the enhanced DNA integration under Olaparib treatment. On one hand, inhibition of catalytic domain of PARP1 suppresses aNHEJ, which should result in decreased DNA integration; on the other hand, Olaparib induces more DNA lesions by trapping PARP1 and PARP2 on DNA, inhibiting BER, and blocking replication forks. At the same time, DNA-PK activity and cNHEJ are enhanced. As a result, DNA integration was enhanced both by accumulated DNA lesions, which provides more sites for transfected DNA to integrate into, mediated by enhanced cNHEJ. In this view, the effects of increased lesions and enhanced cNHEJ overcome the effect of diminished aNHEJ, giving a net increase in DNA integration with Olaparib treatment.

It is also possible that the two PARP inhibitors, PJ34 and Olaparib inhibit PARP by different mode of action. It is not surprising different chemicals inhibit the same protein through different mechanisms. For example, BMN673, Olaparib, and rucaparib showed similar inhibitory effect to the catalytic domain of PARP1 and PARP2, but BMN673 has much higher ability in trapping PARP on DNA (Murai et al., 2014). This is analogous to the situation with topoisomerase IIα and Metnase. The topoisomerase IIα inhibitor etoposide binds to the same pocket where Metnase binds, but neoamphimedine binds to a different site. As a result, Metnase

confers resistance to etoposide but not to neoamphimedine (Ponder et al., 2011; Williamson et al., 2008b; Wray et al., 2009b). The difference between PJ34 and Olaparib should not be that distinct as both are nicotinamide-mimicking derivatives that bind to the nicotinamide-binding site in the donor pocket of PARP1 and PARP2 (Steffen et al., 2013). They probably bind to distinct residues outside of the donor pocket and result in different potency and selectivity.

PROPOSED MODEL OF PLASMID INTEGRATION VIA NHEJ

Transfected plasmid DNA typically enters the cell through endocytosis, except when foreign DNA is delivered directly to nucleus by electroporation (Nickoloff et al., 1998). If DNA traverses the cell membrane through the endocytic route, DNA is subjected to degradation by lysosomal nucleases (Bai et al., 2013; Howell et al., 2003). Due to degradation, even circular DNA may be released from lysosomes as linear DNA. This could also explain that even with overexpression of Metnase, longer deletions from plasmid was not prevented. Upon entering the nucleus, the free ends of linear plasmid DNA the activate DSB repair pathways. Spontaneous DNA damage may be preferred sites for integration. DNA binding proteins bind to these ends and attempt to "repair" these ends. MRN, KU, PARP1 engage in early steps of DNA integration and likely determine if the integration occurs through KU/DNA-PK-initiated cNHEJ or PARP1/CtIP-initiated aNHEJ by regulating end resection (Figure 1). KU and PARP1 compete for DSB binding (Paddock et al., 2011; Wang et al., 2006), this may play an important role in determining the pathway of random plasmid DNA integration. If KU binds to the ends, DNA integration then occurs via cNHEJ. In the cNHEJ pathway, Metnase enhances the overall integration efficiency, likely by enhancing assembly of cNHEJ complex through its interactions with KU and Ligase IV, and by trimming the ends to favor ligation by Ligase IV. If PARP1 binds to the ends, DNA integration occurs via aNHEJ. In the aNHEJ pathway, PARP1 is

important in promoting recruitment of MRN complex and CtIP for end resection, which is essential for exposing microhomologies to favor Ligase III ligation. Deficiency in either pathway results in decreased DNA integration, but does not eliminate integration as long as the other pathway still functions normally. In terms of the effects of PARP inhibitors, inhibition of PARP1 suppresses the aNHEJ-mediated DNA integration. But it could also result in accumulation of DNA damage due to defect in base excision repair (BER) as a result of PARP1 inhibition. The resulting increase of DNA damage may promote DNA integration through other repair pathways, such as cNHEJ. Through unknown mechanisms, PARP inhibitors enhance the activity of DNA-PK and promote phosphorylation of DNA-PK targets (Patel et al., 2011). This could result in increased DNA integration through cNHEJ.

SIGNIFICANCE

In this dissertation research, we examined the functions of two DNA repair proteins, Metnase and PARP1, in DNA integration. The results demonstrated that Metnase promotes DNA integration via the cNHEJ pathway, which does not depend on microhomologies; PARP1 also promotes DNA integration, and more work is required to determine if PARP1 promotes DNA integration via the aNHEJ pathways. Overexpression of Metnase promotes cNHEJ-mediated DNA integration and suppresses use of microhomology at integration junctions, suggesting that manipulation of Metnase expression level can shift DNA integration pathway choice. It has been demonstrated that knockdown of two other cNHEJ proteins, KU and XRCC4, enhance efficiency of targeted (homology-directed) DNA integration by suppressing NHEJ-mediated random DNA integration (Bertolini et al., 2009). Because Metnase and PARP1 knockdown and inhibition are not lethal (Apostolou et al., 2014; De Haro et al., 2010; de Murcia et al., 1997), manipulating the

activity of these proteins may provide new opportunities to improve the efficiency of gene targeting.

RECOMMENDATIONS FOR FUTURE EXPERIMENTATION

For future study, it would be interesting to pursue the opposite effect on integration of PARP1 knockdown vs Olaparib treatment and investigate if Metnase and PARP1 promote DNA integration independently through distinct NHEJ pathways. We speculated that the opposite effects of PARP1 knockdown and Olaparib result from increased DNA damages induced by Olaparib. This can be tested by monitoring formation of γ-H2AX foci under both conditions. We suspected that PJ34 has off target effects that impact efficiency of DNA integration, therefore, PJ34 treatment resulted in higher level of suppression of DNA integration than PARP1 knockdown. Applying PJ34 and other PARP inhibitors to PARP1 knockdown will help to clarify off target effects that influence DNA integration and DNA repair. If PARP1 promotes DNA integration through microhomology-dependent aNHEJ, we predict that DNA integration in PARP1 knockdown or knockout cells would result in reduced extent microhomologies (≥4 bp) at the integration junctions. This can be tested by comparing integration junctions in PARP1 proficient and deficient cells. It would also be interesting to test if knockdown of other key aNHEJ proteins would suppress DNA integration similar to PARP1 knockdown. Metnase is thought to prevent aNHEJ by promoting the less error-prone cNHEJ. As a result, Metnase suppresses chromosomal translocations (Wray et al., 2010). In contrast, PARP1-initiated aNHEJ leads to chromosomal translocations (Wray et al., 2013). Some argue that cNHEJ instead is the major generator of translocations (Ghezraoui et al., 2014). It is possible that both cNHEJ and aNHEJ contribute to generation of translocations, or there is early phase crosstalk between the two pathways that remains to be defined. As one future direction following our project, it would

be interesting to test DNA integration under Metnase/PARP1 double knockdown or double inhibition conditions. If DNA integration occurs through cNHEJ and aNHEJ independently, double knockdown or double inhibition of Metnase and PARP1 should suppress DNA integration more than each single knockdown.

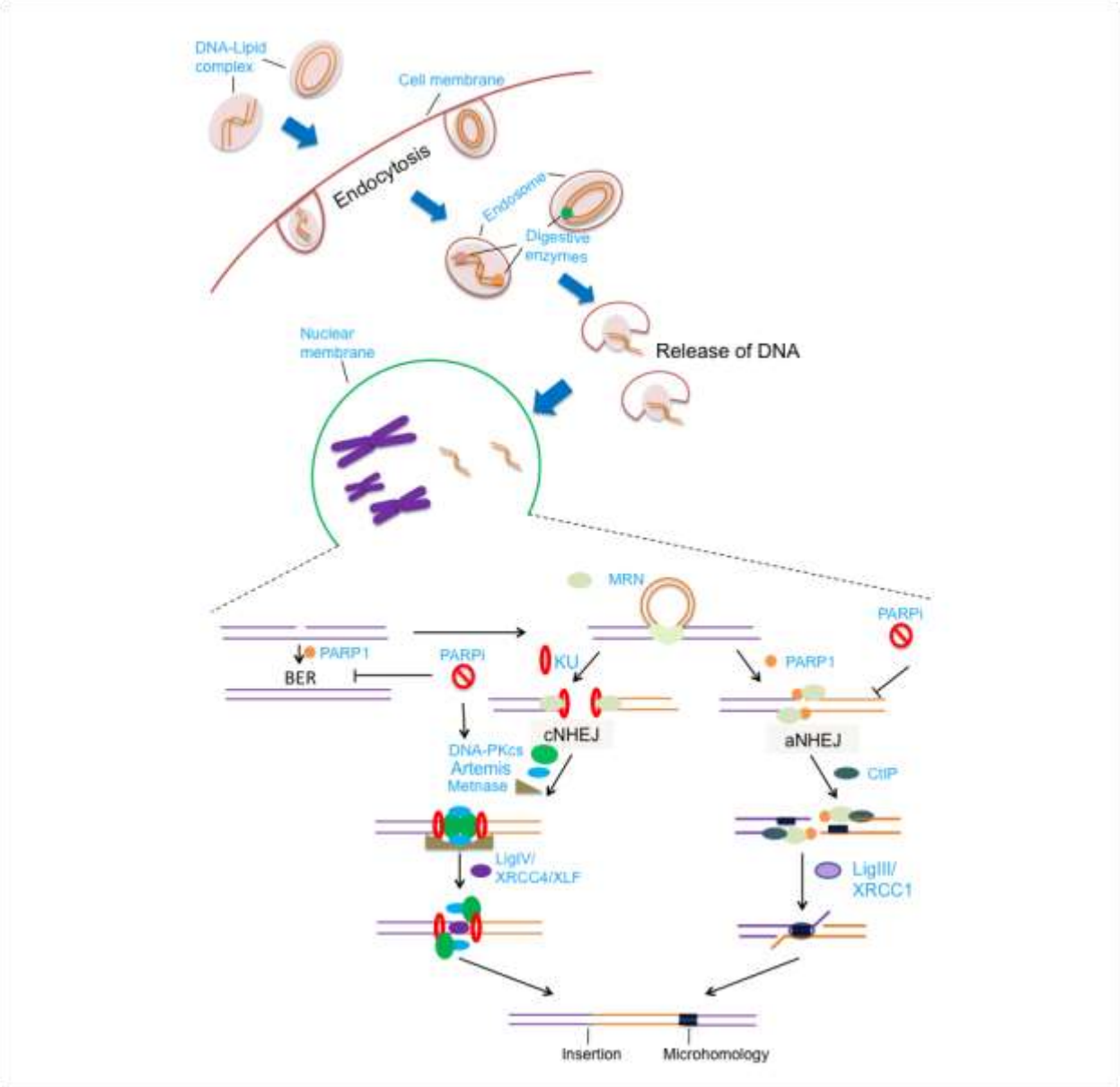

Figure 1. Proposed model for random plasmid integration through NHEJ
Plasmid DNA (linear or circular, orange lines) forms complex with Lipofectamine reagent (pink) and enters the cell membrane (burgundy line) through endocytosis. Cellular digestive enzymes cause deletions to foreign DNA, which is likely to be released as linear DNA. In the nucleus (green line), DNA is bound by DNA repair proteins. Binding of DNA by KU leads to integration through cNHEJ pathway, binding by PARP1 leads to integration through aNHEJ. Base deletions could occur during end processing by digestive enzyme in the cytoplasm or during integration. Insertion of extra bases at the junction could result from both cNHEJ and aNHEJ mediated integration. Microhomologies between the plasmid DNA and integration sites help the end-joining step.

REFERENCES

Ain, Q.U., Chung, J.Y., and Kim, Y.H. (2014). Current and future delivery systems for engineered nucleases: ZFN, TALEN and RGEN. J Control Release.

Apostolou, P., Toloudi, M., Kourtidou, E., Mimikakou, G., Vlachou, I., Chatziioannou, M., Kipourou, V., and Papasotiriou, I. (2014). Potential role for the Metnase transposase fusion gene in colon cancer through the regulation of key genes. Plos One *9*, e109741.

Araujo, F.D., Stracker, T.H., Carson, C.T., Lee, D.V., and Weitzman, M.D. (2005). Adenovirus type 5 E4orf3 protein targets the Mre11 complex to cytoplasmic aggresomes. J Virol *79*, 11382-11391.

Bai, J., Liu, Y., Sun, W., Chen, J., Miller, A.D., and Xu, Y. (2013). Down-regulated lysosomal processing improved pegylated lipopolyplex-mediated gene transfection. J Gene Med *15*, 182-192.

Beck, B.D., Lee, S.S., Hromas, R., and Lee, S.H. (2010). Regulation of Metnase's TIR binding activity by its binding partner, Pso4. Arch Biochem Biophys *498*, 89-94.

Beck, B.D., Lee, S.S., Williamson, E., Hromas, R.A., and Lee, S.H. (2011). Biochemical characterization of metnase's endonuclease activity and its role in NHEJ repair. Biochemistry *50*, 4360-4370.

Beck, B.D., Park, S.J., Lee, Y.J., Roman, Y., Hromas, R.A., and Lee, S.H. (2008). Human Pso4 is a metnase (SETMAR)-binding partner that regulates metnase function in DNA repair. J Biol Chem *283*, 9023-9030.

Bertolini, L.R., Bertolini, M., Maga, E.A., Madden, K.R., and Murray, J.D. (2009). Increased gene targeting in Ku70 and Xrcc4 transiently deficient human somatic cells. Mol Biotechnol *41*, 106-114.

Bodley, A.L., Huang, H.C., Yu, C., and Liu, L.F. (1993). Integration of simian virus 40 into cellular DNA occurs at or near topoisomerase II cleavage hot spots induced by VM-26 (teniposide). Mol Cell Biol *13*, 6190-6200.

Byrne, M., Wray, J., Reinert, B., Wu, Y., Nickoloff, J., Lee, S.H., Hromas, R., and Williamson, E. (2014). Mechanisms of oncogenic chromosomal translocations. Ann N Y Acad Sci *1310*, 89-97.

Carson, C.T., Schwartz, R.A., Stracker, T.H., Lilley, C.E., Lee, D.V., and Weitzman, M.D. (2003). The Mre11 complex is required for ATM activation and the G2/M checkpoint. Embo J *22*, 6610-6620.

Chaurushiya, M.S., and Weitzman, M.D. (2009). Viral manipulation of DNA repair and cell cycle checkpoints. DNA Repair (Amst) *8*, 1166-1176.

Chen, Y., Williams, V., Filippova, M., Filippov, V., and Duerksen-Hughes, P. (2014). Viral carcinogenesis: factors inducing DNA damage and virus integration. Cancers (Basel) *6*, 2155-2186.

Chen, Y., Zhang, L., and Hao, Q. (2013). Olaparib: a promising PARP inhibitor in ovarian cancer therapy. Archives of gynecology and obstetrics *288*, 367-374.

Cordaux, R., and Batzer, M.A. (2009). The impact of retrotransposons on human genome evolution. Nat Rev Genet *10*, 691-703.

Cordaux, R., Udit, S., Batzer, M.A., and Feschotte, C. (2006). Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. Proc Natl Acad Sci U S A *103*, 8101-8106.

De Haro, L.P., Wray, J., Williamson, E.A., Durant, S.T., Corwin, L., Gentry, A.C., Osheroff, N., Lee, S.H., Hromas, R., and Nickoloff, J.A. (2010). Metnase promotes restart and repair of stalled and collapsed replication forks. Nucleic Acids Res *38*, 5681-5691.

de Murcia, J.M., Niedergang, C., Trucco, C., Ricoul, M., Dutrillaux, B., Mark, M., Oliver, F.J., Masson, M., Dierich, A., LeMeur, M., *et al.* (1997). Requirement of poly(ADP-ribose) polymerase in recovery from DNA damage in mice and in cells. Proc Natl Acad Sci U S A *94*, 7303-7307.

Debenham, P.G., and Webb, M.B. (1984). The effect of X-rays and ultraviolet light on DNA-mediated gene transfer in mammalian cells. Int J Radiat Biol Relat Stud Phys Chem Med *46*, 555-568.

Deriano, L., and Roth, D.B. (2013). Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. Annu Rev Genet *47*, 433-455.

Dubose, A.J., Lichtenstein, S.T., Narisu, N., Bonnycastle, L.L., Swift, A.J., Chines, P.S., and Collins, F.S. (2013). Use of microarray hybrid capture and next-generation sequencing to identify the anatomy of a transgene. Nucleic Acids Res *41*, e70.

Fnu, S., Williamson, E.A., De Haro, L.P., Brenneman, M., Wray, J., Shaheen, M., Radhakrishnan, K., Lee, S.H., Nickoloff, J.A., and Hromas, R. (2011). Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. Proc Natl Acad Sci U S A *108*, 540-545.

Fujimaki, K., Aratani, Y., Fujisawa, S., Motomura, S., Okubo, T., and Koyama, H. (1996). DNA topoisomerase II inhibitors enhance random integration of transfected vectors into human chromosomes. Somat Cell Mol Genet *22*, 279-290.

Gauss, G.H., and Lieber, M.R. (1996). Mechanistic constraints on diversity in human V(D)J recombination. Mol Cell Biol *16*, 258-269.

Ghezraoui, H., Piganeau, M., Renouf, B., Renaud, J.B., Sallmyr, A., Ruis, B., Oh, S., Tomkinson, A.E., Hendrickson, E.A., Giovannangeli, C., *et al.* (2014). Chromosomal

translocations in human cells are generated by canonical nonhomologous end-joining. Mol Cell *55*, 829-842.

Hanks, S., Coleman, K., Reid, S., Plaja, A., Firth, H., Fitzpatrick, D., Kidd, A., Mehes, K., Nash, R., Robin, N., *et al.* (2004). Constitutional aneuploidy and cancer predisposition caused by biallelic mutations in BUB1B. Nat Genet *36*, 1159-1161.

Hosoya, N., and Miyagawa, K. (2014). Targeting DNA damage response in cancer therapy. Cancer Sci *105*, 370-388.

Howell, D.P., Krieser, R.J., Eastman, A., and Barry, M.A. (2003). Deoxyribonuclease II is a lysosomal barrier to transfection. Mol Ther *8*, 957-963.

Hromas, R., Williamson, E.A., Fnu, S., Lee, Y.J., Park, S.J., Beck, B.D., You, J.S., Leitao, A., Nickoloff, J.A., and Lee, S.H. (2012). Chk1 phosphorylation of Metnase enhances DNA repair but inhibits replication fork restart. Oncogene *31*, 4245-4254.

Hromas, R., Wray, J., Lee, S.H., Martinez, L., Farrington, J., Corwin, L.K., Ramsey, H., Nickoloff, J.A., and Williamson, E.A. (2008). The human set and transposase domain protein Metnase interacts with DNA Ligase IV and enhances the efficiency and accuracy of non-homologous end-joining. DNA Repair (Amst) *7*, 1927-1937.

Hsiung, N., Warrick, H., deRiel, J.K., Tuan, D., Forget, B.G., Skoultchi, A., and Kucherlapati, R. (1980). Cotransfer of circular and linear prokaryotic and eukaryotic DNA sequences into mouse cells. Proc Natl Acad Sci U S A *77*, 4852-4856.

Ishibashi, K., Suzuki, K., Ando, Y., Takakura, C., and Inoue, H. (2006). Nonhomologous chromosomal integration of foreign DNA is completely dependent on MUS-53 (human Lig4 homolog) in Neurospora. Proc Natl Acad Sci U S A *103*, 14871-14876.

Javle, M., and Curtin, N.J. (2011). The role of PARP in DNA repair and its therapeutic exploitation. Br J Cancer *105*, 1114-1122.

Kassiotis, G. (2014). Endogenous retroviruses and the development of cancer. J Immunol *192*, 1343-1349.

Kaufman, B., Shapira-Frommer, R., Schmutzler, R.K., Audeh, M.W., Friedlander, M., Balmana, J., Mitchell, G., Fried, G., Stemmer, S.M., Hubert, A., *et al.* (2015). Olaparib Monotherapy in Patients With Advanced Cancer and a Germline BRCA1/2 Mutation. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *33*, 244-250.

Kim, H.S., Chen, Q., Kim, S.K., Nickoloff, J.A., Hromas, R., Georgiadis, M.M., and Lee, S.H. (2014). The DDN catalytic motif is required for Metnase functions in non-homologous end joining (NHEJ) repair and replication restart. J Biol Chem *289*, 10930-10938.

Leadon, S.A., Ganesan, A.K., and Hanawalt, P.C. (1987). Enhanced transforming activity of ultraviolet-irradiated pSV2-gpt is due to damage outside the gpt transcription unit. Plasmid *18*, 135-141.

Ledermann, J., Harter, P., Gourley, C., Friedlander, M., Vergote, I., Rustin, G., Scott, C., Meier, W., Shapira-Frommer, R., Safra, T., *et al.* (2012). Olaparib maintenance therapy in platinum-sensitive relapsed ovarian cancer. The New England journal of medicine *366*, 1382-1392.

Lee, J.M., Ledermann, J.A., and Kohn, E.C. (2014). PARP Inhibitors for BRCA1/2 mutation-associated and BRCA-like malignancies. Ann Oncol *25*, 32-40.

Lee, S.H., Oshige, M., Durant, S.T., Rasila, K.K., Williamson, E.A., Ramsey, H., Kwan, L., Nickoloff, J.A., and Hromas, R. (2005). The SET domain protein Metnase mediates foreign DNA integration and links integration to nonhomologous end-joining repair. Proc Natl Acad Sci U S A *102*, 18075-18080.

Lilley, C.E., Schwartz, R.A., and Weitzman, M.D. (2007). Using or abusing: viruses and the cellular DNA damage response. Trends Microbiol *15*, 119-126.

Liu, D., Bischerour, J., Siddique, A., Buisine, N., Bigot, Y., and Chalmers, R. (2007). The human SETMAR protein preserves most of the activities of the ancestral Hsmar1 transposase. Mol Cell Biol *27*, 1125-1132.

Lupo, B., and Trusolino, L. (2014). Inhibition of poly(ADP-ribosyl)ation in cancer: old and new paradigms revisited. Biochim Biophys Acta *1846*, 201-215.

Manivasakam, P., Aubrecht, J., Sidhom, S., and Schiestl, R.H. (2001). Restriction enzymes increase efficiencies of illegitimate DNA integration but decrease homologous integration in mammalian cells. Nucleic Acids Res *29*, 4826-4833.

Mansour, W.Y., Rhein, T., and Dahm-Daphi, J. (2010). The alternative end-joining pathway for repair of DNA double-strand breaks requires PARP1 but is not dependent upon microhomologies. Nucleic Acids Res *38*, 6065-6077.

Merrihew, R.V., Marburger, K., Pennington, S.L., Roth, D.B., and Wilson, J.H. (1996). High-frequency illegitimate integration of transfected DNA at preintegrated target sites in a mammalian genome. Mol Cell Biol *16*, 10-18.

Mohapatra, S., Yannone, S.M., Lee, S.H., Hromas, R.A., Akopiants, K., Menon, V., Ramsden, D.A., and Povirk, L.F. (2013). Trimming of damaged 3' overhangs of DNA double-strand breaks by the Metnase and Artemis endonucleases. DNA Repair (Amst) *12*, 422-432.

Morgan, M.A., and Shilatifard, A. (2015). Chromatin signatures of cancer. Genes Dev *29*, 238-249.

Mott, C., and Symington, L.S. (2011). RAD51-independent inverted-repeat recombination by a strand-annealing mechanism. DNA Repair (Amst) *10*, 408-415.

Mullins, C.S., and Linnebacher, M. (2012). Human endogenous retroviruses and cancer: Causality and therapeutic possibilities. World J Gastroenterol *18*, 6027-6035.

Murai, J., Huang, S.Y., Das, B.B., Renaud, A., Zhang, Y., Doroshow, J.H., Ji, J., Takeda, S., and Pommier, Y. (2012). Trapping of PARP1 and PARP2 by Clinical PARP Inhibitors. Cancer Res *72*, 5588-5599.

Murai, J., Huang, S.Y., Renaud, A., Zhang, Y., Ji, J., Takeda, S., Morris, J., Teicher, B., Doroshow, J.H., and Pommier, Y. (2014). Stereospecific PARP trapping by BMN 673 and comparison with olaparib and rucaparib. Mol Cancer Ther *13*, 433-443.

Nagata, Y., Takagi, H., Morita, T., and Oishi, M. (1984). Stimulation of DNA-mediated transformation by UV irradiation of recipient (mouse FM3A) cells. J Cell Physiol *121*, 453-457.

Nakayama, C., Adachi, N., and Koyama, H. (1998). Bleomycin enhances random integration of transfected DNA into a human genome. Mutat Res *409*, 1-10.

Newman, S., Hermetz, K.E., Weckselblatt, B., and Rudd, M.K. (2015). Next-Generation Sequencing of Duplication CNVs Reveals that Most Are Tandem and Some Create Fusion Genes at Breakpoints. Am J Hum Genet *96*, 208-220.

Nickoloff, J.A., Spirio, L.N., and Reynolds, R.J. (1998). A comparison of calcium phosphate coprecipitation and electroporation. Implications for studies on the genetic effects of DNA damage. Mol Biotechnol *10*, 93-101.

O'Neil, N.J., van Pel, D.M., and Hieter, P. (2013). Synthetic lethality and cancer: cohesin and PARP at the replication fork. Trends Genet *29*, 290-297.

Oh, S., Harvey, A., Zimbric, J., Wang, Y., Nguyen, T., Jackson, P.J., and Hendrickson, E.A. (2014). DNA ligase III and DNA ligase IV carry out genetically distinct forms of end joining in human somatic cells. DNA Repair (Amst) *21*, 97-110.

Paddock, M.N., Bauman, A.T., Higdon, R., Kolker, E., Takeda, S., and Scharenberg, A.M. (2011). Competition between PARP-1 and Ku70 control the decision between high-fidelity and mutagenic DNA repair. DNA Repair (Amst) *10*, 338-343.

Pannunzio, N.R., Li, S., Watanabe, G., and Lieber, M.R. (2014). Non-homologous end joining often uses microhomology: implications for alternative end joining. DNA Repair (Amst) *17*, 74-80.

Patel, A.G., Sarkaria, J.N., and Kaufmann, S.H. (2011). Nonhomologous end joining drives poly(ADP-ribose) polymerase (PARP) inhibitor lethality in homologous recombination-deficient cells. Proc Natl Acad Sci U S A *108*, 3406-3411.

Perez, C.F., Botchan, M.R., and Tobias, C.A. (1985). DNA-mediated gene transfer efficiency is enhanced by ionizing and ultraviolet irradiation of rodent cells in vitro. I. Kinetics of enhancement. Radiat Res *104*, 200-213.

Perez, C.F., and Skarsgard, L.D. (1986). Radiation enhancement of the efficiency of DNA-mediated gene transfer in CHO UV-sensitive mutants. Radiat Res *106*, 401-407.

Ponder, J., Yoo, B.H., Abraham, A.D., Li, Q., Ashley, A.K., Amerin, C.L., Zhou, Q., Reid, B.G., Reigan, P., Hromas, R., *et al.* (2011). Neoamphimedine circumvents metnase-enhanced DNA topoisomerase IIalpha activity through ATP-competitive inhibition. Mar Drugs *9*, 2397-2408.

Postel, E.H. (1985). Enhancement of genetic transformation frequencies of mammalian cell cultures by damage to the cell DNA. Mol Gen Genet *201*, 136-139.

Roman, Y., Oshige, M., Lee, Y.J., Goodwin, K., Georgiadis, M.M., Hromas, R.A., and Lee, S.H. (2007). Biochemical characterization of a SET and transposase fusion protein, Metnase: its DNA binding and DNA cleavage activity. Biochemistry *46*, 11369-11376.

Rubin, J.S. (1988). Effect of gamma rays on efficiency of gene transfer in DNA repair-proficient and -deficient cell lines. Somat Cell Mol Genet *14*, 613-621.

Schwartz, R.A., Palacios, J.A., Cassell, G.D., Adam, S., Giacca, M., and Weitzman, M.D. (2007). The Mre11/Rad50/Nbs1 complex limits adeno-associated virus transduction and replication. J Virol *81*, 12936-12945.

Shaheen, M., Williamson, E., Nickoloff, J., Lee, S.H., and Hromas, R. (2010). Metnase/SETMAR: a domesticated primate transposase that enhances DNA repair, replication, and decatenation. Genetica *138*, 559-566.

Shcherbakova, O.G., and Filatov, M.V. (2000). Camptothecin enhances random integration of transfected DNA into the genome of mammalian cells. Biochim Biophys Acta *1495*, 1-3.

Spivak, G., Ganesan, A.K., and Hanawalt, P.C. (1984). Enhanced transformation of human cells by UV-irradiated pSV2 plasmids. Mol Cell Biol *4*, 1169-1171.

Spivak, G., Leadon, S.A., Vos, J.M., Meade, S., Hanawalt, P.C., and Ganesan, A.K. (1988). Enhanced transforming activity of pSV2 plasmids in human cells depends upon the type of damage introduced into the plasmid. Mutat Res *193*, 97-108.

Steffen, J.D., Brody, J.R., Armen, R.S., and Pascal, J.M. (2013). Structural Implications for Selective Targeting of PARPs. Frontiers in oncology *3*, 301.

Stracker, T.H., Carson, C.T., and Weitzman, M.D. (2002). Adenovirus oncoproteins inactivate the Mre11-Rad50-NBS1 DNA repair complex. Nature *418*, 348-352.

Tanaka, S., Ishii, C., Hatakeyama, S., and Inoue, H. (2010). High efficient gene targeting on the AGAMOUS gene in an ArabidopsisAtLIG4 mutant. Biochem Biophys Res Commun *396*, 289-293.

Taylor, T.J., and Knipe, D.M. (2004). Proteomics of herpes simplex virus replication compartments: association of cellular DNA replication, repair, recombination, and chromatin remodeling proteins with ICP8. J Virol *78*, 5856-5866.

Uh, M., Khattra, J., and Devlin, R.H. (2006). Transgene constructs in coho salmon (Oncorhynchus kisutch) are repeated in a head-to-tail fashion and can be integrated adjacent to horizontally-transmitted parasite DNA. Transgenic Res *15*, 711-727.

van Duin, M., Westerveld, A., and Hoeijmakers, J.H. (1985). UV stimulation of DNA-mediated transformation of human cells. Mol Cell Biol *5*, 734-741.

Vos, J.M., and Hanawalt, P.C. (1989). DNA interstrand cross-links promote chromosomal integration of a selected gene in human cells. Mol Cell Biol *9*, 2897-2905.

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

Wang, Z., Troilo, P.J., Wang, X., Griffiths, T.G., Pacchione, S.J., Barnum, A.B., Harper, L.B., Pauley, C.J., Niu, Z., Denisova, L., *et al.* (2004). Detection of integration of plasmid DNA into host genomic DNA following intramuscular injection and electroporation. Gene Ther *11*, 711-721.

Williamson, E.A., Farrington, J., Martinez, L., Ness, S., O'Rourke, J., Lee, S.H., Nickoloff, J., and Hromas, R. (2008a). Expression levels of the human DNA repair protein metnase influence lentiviral genomic integration. Biochimie *90*, 1422-1426.

Williamson, E.A., Rasila, K.K., Corwin, L.K., Wray, J., Beck, B.D., Severns, V., Mobarak, C., Lee, S.H., Nickoloff, J.A., and Hromas, R. (2008b). The SET and transposase domain protein Metnase enhances chromosome decatenation: regulation by automethylation. Nucleic Acids Res *36*, 5822-5831.

Wray, J., Williamson, E.A., Chester, S., Farrington, J., Sterk, R., Weinstock, D.M., Jasin, M., Lee, S.H., Nickoloff, J.A., and Hromas, R. (2010). The transposase domain protein Metnase/SETMAR suppresses chromosomal translocations. Cancer Genet Cytogenet *200*, 184-190.

Wray, J., Williamson, E.A., Royce, M., Shaheen, M., Beck, B.D., Lee, S.H., Nickoloff, J.A., and Hromas, R. (2009a). Metnase mediates resistance to topoisomerase II inhibitors in breast cancer cells. Plos One *4*, e5323.

Wray, J., Williamson, E.A., Sheema, S., Lee, S.H., Libby, E., Willman, C.L., Nickoloff, J.A., and Hromas, R. (2009b). Metnase mediates chromosome decatenation in acute leukemia cells. Blood *114*, 1852-1858.

Wray, J., Williamson, E.A., Singh, S.B., Wu, Y., Cogle, C.R., Weinstock, D.M., Zhang, Y., Lee, S.H., Zhou, D., Shao, L., *et al.* (2013). PARP1 is required for chromosomal translocations. Blood *121*, 4359-4365.

Wurtele, H., Little, K.C., and Chartrand, P. (2003). Illegitimate DNA integration in mammalian cells. Gene Ther *10*, 1791-1799.

Yan, B.W., Zhao, Y.F., Cao, W.G., Li, N., and Gou, K.M. (2013). Mechanism of random integration of foreign DNA in transgenic mice. Transgenic Res *22*, 983-992.

Zhang, Y., and Jasin, M. (2011). An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. Nat Struct Mol Biol *18*, 80-84.

Zheng, S.J., Henken, B., Sofiari, E., Jacobsen, E., Krens, F.A., and Kik, C. (2001). Molecular characterization of transgenic shallots (Allium cepa L.) by adaptor ligation PCR (AL-PCR) and sequencing of genomic DNA flanking T-DNA borders. Transgenic Res *10*, 237-245.