University of South Carolina
Scholar Commons

Theses and Dissertations

2017

Mismatch Tolerance during Homologous Recombination in Mammalian Cells

Shen Li
University of South Carolina

Follow this and additional works at: http://scholarcommons.sc.edu/etd

 Part of the Biology Commons

Recommended Citation

Li, S.(2017). *Mismatch Tolerance during Homologous Recombination in Mammalian Cells.* (Doctoral dissertation). Retrieved from http://scholarcommons.sc.edu/etd/4133

This Open Access Dissertation is brought to you for free and open access by Scholar Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact SCHOLARC@mailbox.sc.edu.

MISMATCH TOLERANCE DURING HOMOLOGOUS RECOMBINATION IN MAMMALIAN CELLS

by

Shen Li

Bachelor of Science
Beijing Normal University, 2003

Master of Science
Beijing Normal University, 2006

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Biological Sciences

College of Arts and Sciences

University of South Carolina

2017

Accepted by:

Alan Waldman, Major Professor

Deanna Smith, Committee Member

Douglas Pittman, Committee Member

David Reisman, Committee Member

Hexin Chen, Committee Member

Barbara Waldman, Committee Member

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

© Copyright by Shen Li, 2017
All Rights Reserved.

ACKNOWLEDGEMENTS

I greatly appreciate the continued encouragement, support, and guidance from Dr. Alan Waldman throughout my graduate study. And I thank all other members in my doctoral committee, Dr. David Reisman, Dr. Deanna Smith, Dr. Douglas Pittman, Dr. Hexin Chen and Dr. Barbara Waldman for their scientific advice, encouragement and guidance. I need to thank past members I had chance to work with in Dr. Waldman Lab including Yibin Wang, Jason Smith and Andrew Patrick for their advice and encouragement. I need to thank undergraduate students Bryan Wehrenberg, James Elliot Copper and Joshua Mercadel, graduate students Jake Massey for their contribution to the presented research in this dissertation. I thank my parents Zhonghe Li and Sulan Shen, my parents in law Xingyi Duan and Aiying Chu, and my brother Xin Li for their unconditional support. And at last, I thank my wife Qing Duan and son Alex Duan Li for their love, support and companion.

ABSTRACT

Homologous recombination (HR) serves critical roles in DNA repair to maintain genome stability, and the malfunction of HR contributes to carcinogenesis and cancer development. Current research focuses on the regulation against deleterious recombination between imperfectly matching sequences, which has been documented in certain myeloid leukemias, hereditary nonpolyposis colorectal cancers and other genetic diseases. Homology dependency of recombination was examined in cultured thymidine kinase-deficient mouse fibroblasts. Cells have chromosomal integration of DNA constructs harboring a herpes tk gene (the "recipient") and a closely linked truncated "donor" tk sequence. The recipient was rendered non-functional by insertion of the recognition site for endonuclease I-SceI, and the donor sequence could restore the function of the recipient through spontaneous gene conversion or via recombinational repair provoked by a double-strand break (DSB) at the I-SceI site. Recombination events were recoverable by HAT selection for tk-positive clones. Three different donor sequences contained 16, 25, or 33 mismatches relative to the recipient, and these mismatches were clustered within these "homeologous" sequences surrounded by region of high homology. Previous work indicated that mammalian cells fastidiously avoid recombination between homeologous sequences, while our results revealed that when homeologous sequences are surrounded by high homology, mismatches are frequently included in gene conversion events. Knock-down of DNA mismatch repair provided evidence that incorporation of mismatches into gene conversion tracts involved repair

of mismatched heteroduplex intermediates. Our results demonstrate that mismatch repair of multiple mispaired bases does not function to impede exchange between homeologous sequences. Moreover, gene conversion tracts from spontaneous recombination showed that either all or none of the mismatches were transferred from donor to recipient, suggesting that recombination must begin and end in high homology. But this requirement was somewhat relaxed in DSB-induced events with recombination ending in homeology. Further experiments with a rearranged construct were attempted to investigate the relaxed homology requirement during DSB repair. In addition to the study on homology requirement of recombination, research works were also carried out to characterize the roles of RecQ4 helicases in DSB repair. It is the first demonstration that RecQ4 deficiency reduces the fraction of crossover events in DSB-induced recombination. Moreover, BLM deficiency failed to boost crossover events in RecQ4 deficient cells. It is postulated that these two helicases act agonistically to determine the generation of crossover events.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION TO HOMOLOGOUS RECOMBINATION AND DNA DOUBLE STRAND BREAK REPAIR IN MAMMALIAN CELLS

Cells are facing numerous attacks on chromosomal DNA throughout their life cycles, such as reactive oxygen species, genotoxic chemicals or radiations. These attacks from endogenous or exogenous sources may damage nitrogenous base, break deoxyribose, introduce single-strand nick, interstrand crosslink or even double-strand break into the genomic DNA. To counteract these DNA lesions, cells maintain several repair pathways to protect genome integrity (1, 2). Base excision repair (BER) and nucleotide excision repair (NER) recognize the damaged nucleotides on one strand of DNA, cleave and resynthesize them according to the remaining, complementary DNA strand. But when DNA damages affect both strands of a DNA sequence, for example, DNA double-strand break (DSB) or interstrand crosslink (ICL), homologous recombination (HR) is needed for faithful repair (3, 4). HR relies on RecA helicase to perform homology search and strand invasion in bacteria while in eukaryotes, Rad51 (homolog of *E.coli* RecA) serves the same function of RecA (5–7). Besides HR, there are other homology independent, error prone pathways to repair DSB, for example, nonhomologous end joining (NHEJ), single strand annealing (SSA) and alternative NHEJ (8, 9).

The pivotal roles of HR in maintaining genome stability are well demonstrated in several rare genetic diseases characterized with genome instability and cancer

predispositions. Patients with Fanconi anemia may have congenital birth defects, short stature, high incidence of acute myeloid leukemia, and bone marrow failure at an early age (10). These syndromes are caused by mutants within a group of genes, the Fanconi anemia complementation group (FANC) gene family. Cells deficient in FANC genes barely respond to replication stress caused by DNA ICLs and fail to repair them through HR, therefore these cells are not able to maintain the genome stability especially when facing crosslinking agents like cisplatin and mitomycin C. RecQ helicase BLM participates in HR and inhibits Holliday junction resolution toward crossover (CO) events. BLM deficient cells have high rate of sister chromatid exchange (SCE) as well as elevated recombination between homologous chromosomes (11–13). As a result, the patients with BLM deficiency (Bloom syndrome) are characterized with genome instability, short stature and predisposition to cancer within multiple tissues. Genetic studies also revealed that deficiencies of other HR mediators, such as BRCA1, BRCA2, RAD54B, RAD51B, RAD51C and RAD51D, endanger the genome stability and contribute to the susceptibility to several cancers including breast cancer, ovarian cancer and colon cancer (14). Specifically, mutations in BRCA1 and BRCA2 significantly increase the carrier's risk in breast cancer and ovarian cancer, and also contribute to other types of cancer. These two genes mainly assist in RAD51 nucleoprotein filament formation, homology search and strand invasion during HR (4). As described above, HR pathway is so tightly connected to genome stability that discoveries of remaining HR components, functions of these participating proteins, and regulation within this complicate process would broaden our knowledge of the genome maintenance in somatic

cells, and help elucidating the genomic alterations in carcinogenesis and cancer development.

Two working models of HR are briefly illustrated in Figure 1.1. The recombination starts from a DNA damage event breaking both strands of a DNA molecule. Then, 5' exonuclease resects the broken DNA ends to produce 3’ single-strand overhangs for homologous pairing (4, 15). Rad51 coats the single-strand overhang, directs it to invade a DNA sequence with high homology, and pairs it with the complementary strand (4). Following the strand invasion, DNA synthesis elongates the invading strand under the guide of the homologous template, therefore restoring the broken region as well as nearby nucleotide sequences. After strand invasion and DNA synthesis, HR may diverge into two ways.

In synthesis dependent strand annealing (SDSA) pathway, the nascent DNA strand aborts DNA synthesis, detaches from the homologous template and anneals to the other resected broken end. Through cooperation of DNA polymerase, exonuclease and DNA ligase, the broken DNA molecule is finally rejoined through base pairing of the complementary DNA ends. SDSA will repair the DSB-damaged DNA sequence according to the chosen homologous template, and leave gene conversion tract in the restored DNA molecule (16–19) .

In the double strand break repair (DSBR) model, the initial strand invasion and DNA synthesis peels off the non-template strand of the template to form a bubble shaped D-loop, and then the extended D-loop anneals to the other resected broken end starting the DNA synthesis of the second strand (7, 20). Later on, the temporary “synaptonemal” intermediate matures into two stable Holliday junctions (double Holliday junctions,

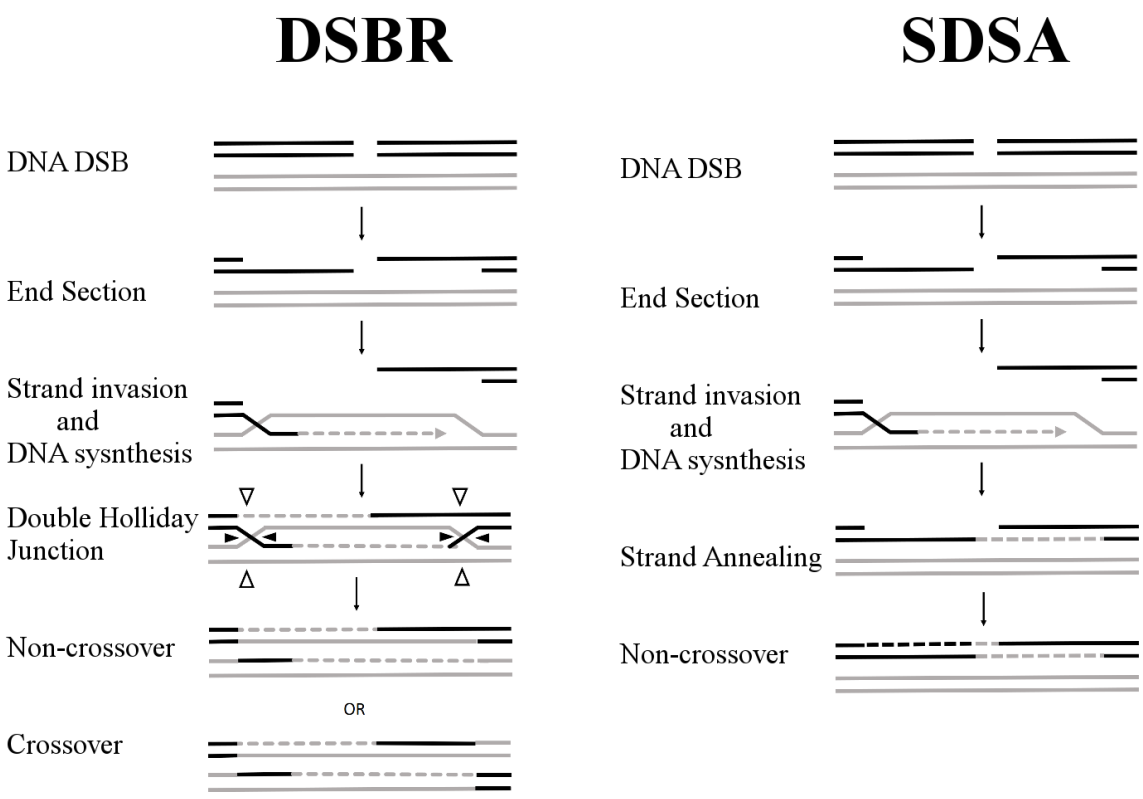

Figure 1.1 Double strand break repair model (DSBR) and synthesis dependent strand annealing model (SDSA).
Two models describing the procedure to repair DNA DSB through HR are demonstrated in above Figure. Black lines represent the strands of the broken DNA molecule while the gray lines represent the strands of the template DNA molecule. The grey arrowheads show the 3' end of the DNA strand and the direction of DNA synthesis. The dashed lines indicate they are the newly synthesized DNA strands. Black or white arrowheads point to the sites for cleavage and re-ligation in Holliday junction resolution. Adapted from San Filippo et al. 2008 (4).

dHJs) bridging these two participating DNA molecules. Finally, the Holliday junctions (HJs) are resolved by coordinated endonuclease cleavage and strand ligation. Holliday junction resolution has two competing outcomes: gene conversion (GC), which restores the broken DNA molecule using genetic information from its homologous template, or crossover (CO), which not only restores the broken DNA molecule, but also splices these two DNA molecules around the repair region.

Supplementary to the DSBR model, Holliday junction dissolution describes a sub-pathway to separate the dHJs connected DNA molecules. RecQ helicase BLM, Topoisomerase IIIα and RMI1 (RecQ mediated genome instability 1 or BLAP75)-RMI2 (BLAP18) propel the convergent branch migration of dHJs to form hemicatenated DNA molecules, then cut and re-ligate one DNA strand to eliminate the topological link between these two molecules (21, 22). The Holliday junction dissolution only restores the broken DNA molecule with gene conversion product indistinguishable from that of SDSA pathway.

Without exogenous stress or disturbance, HR happens at an extremely low rate during cell proliferation, and it is called spontaneous recombination. The origin of spontaneous recombination is debatable and probably different causes mentioned below all contribute to the spontaneous recombination events. HR may start from a spontaneously occurring DSB, and generate gene conversion or crossover events as described above (Figure 1.1). Secondly, a stalled replication fork may cause spontaneous recombination as well (23). When a replication fork meets obstacle, and pauses, it may regress, unwinding the newly synthesized DNA strands from the parental DNA (Figure 1.2, right panel). These two complementary strands then anneal and turn the replication

fork into a chicken foot structure. The newly synthesized leading strand within the structure would invade the parental DNA to reestablish the replication fork. DNA recombination restoring stalled replication fork usually does not cause strand splicing between sister chromatids. Thirdly, collapsed replication forks spark spontaneous recombination (24). When leading strand encounters a nick on its complementary, parental DNA strand, it will run off the ongoing replication fork as a one-ended DSB. The broken end would invade the parental DNA molecule and reconstruct a new replication fork with the help of RAD51 (Figure 1.2, left panel). Unlike stalled replication fork, restoration of the collapsed replication fork tends to generate SCE. Spontaneous recombination is needed to resolve other DNA lesions during replication as well, which involves HJs and associated resolution, similar to previous two described pathways (25). Lastly, the newly synthesized DNA strand may simply unwind from the original parental strand, and anneal to a nearby homologous strand to continue DNA replication (4). The template switching allows daughter DNA molecule to copy from a homologous partner and overcome the lesions on its parental DNA molecule.

In recombination studies, fluctuation test has been used to collect spontaneous recombination events as well as to estimate the rate of the recombination events (26). The fluctuation test was first introduced to delineate the origin of the bacteria's immunity against virus, from spontaneous mutation or acquired immunity (27). Proportion of resistant bacteria were monitored in a growing culture and subsequent cultures, and the great variation across the subcultures confirmed the mutation hypothesis. The occurrence of HR events fits the mathematical description in aforementioned paper, which implies

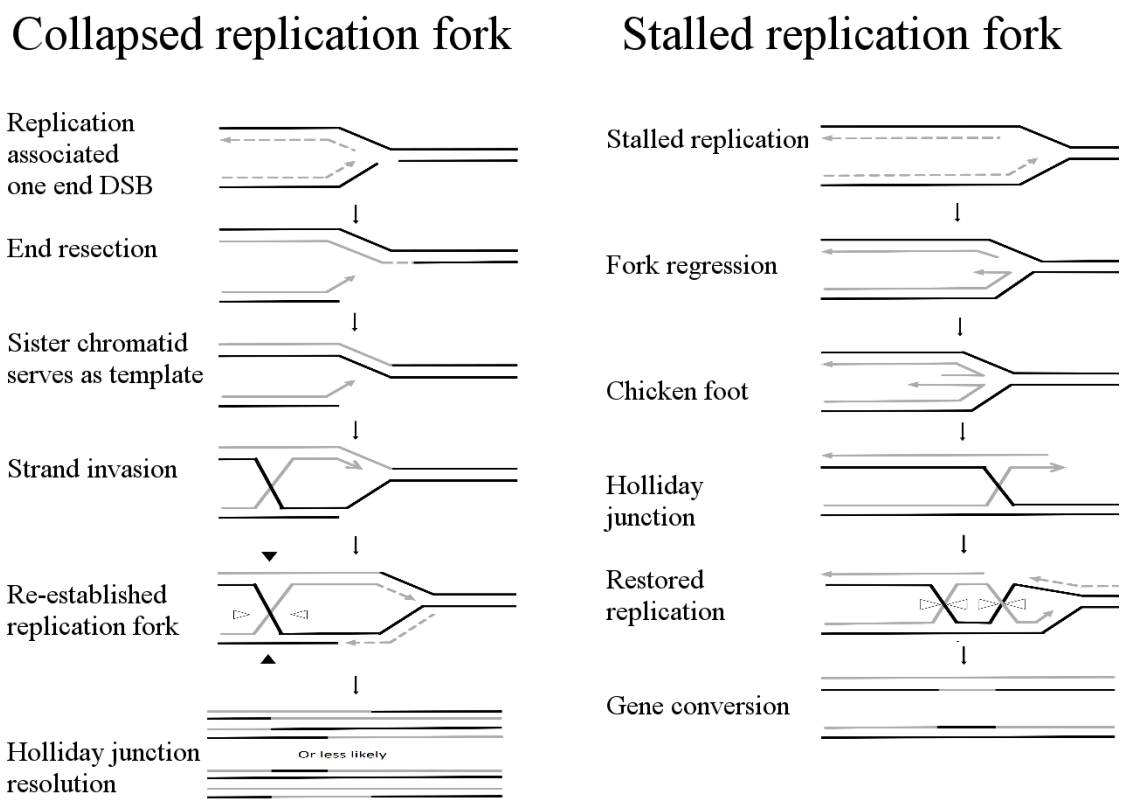

Figure 1.2 DNA recombination restores problematic replication forks.
The collapsed and stalled replication fork are rescued by HR, as shown above. Parental DNA strands are drawn with black line while the daughter strands are drawn with gray line. Newly synthesized DNA strands are drawn with dash line and the arrowheads indicate the 3' end of the DNA strand. Adapted from Helleday 2003 (28).

that the recombination events arise during cell proliferation similar to the spontaneous mutations.

HR participates in DSB repair to maintain genome stability, as stated earlier in this chapter and the recombination frequency is dramatically boosted by DNA DSB. Ionizing radiation, reactive oxygen species, and other chemicals produce DNA DSBs directly or indirectly (29). Approximately 50 endogenous DNA DSBs happen per cell cycle, and the number will double if the cells are facing 1 Gy_t ionizing radiation (9). Timely repair of DSBs is vital for cells to overcome these intracellular or extracellular stresses, while the failure would lead to cell cycle arrest and even cell death (8). As illustrated in Figure 1.1, during HR, broken DNA ends are resected to generate 3' single-strand overhangs and then directed to a homologous template for restorative DNA synthesis. HR faithfully repairs the broken DNA molecule, while sometimes produces crossover events splicing the broken DNA molecule and the template DNA molecule. HR is also critical to repair other DNA lesions. DNA ICLs may be cleaved to produce DSB, and then HR takes the turn to complete the late stage of repair (10). Conceivably HR can fix DSBs generated by other repair pathways, and accurately restore the broken sequence based on a homologous template.

HR provides faithful repair even if the broken DNA molecule loses a significant amount of DNA sequence after DNA damage. The benefits of the HR are obvious unless the chosen homologous template has a mutation within region needed for repair. In this rare circumstance, loss of heterozygosity (LOH) will severely affect the normal gene function in the host cell (9, 13).

Cells also utilize other pathways to repair DSBs, especially non-dividing cells lacking in HR activity. Non-homologous end joining (NHEJ), single strand annealing (SSA), precise ligation (PL) and other alternative pathways can repair DNA DSBs (Figure 1.3). NHEJ is a prominent pathway to repair DSB, especially when the damaged DNA does not have homologous sequence available nearby. NHEJ shares protein components with HR initially: MRN (MRE11-RAD50-NBS1) complex trims broken DNA ends and cleaves obstructive adducts for both pathways. NHEJ employs Ku70 (XRCC6, Lupus Ku autoantigen protein p70 or X-ray repair cross complementing 6), Ku80 to tether the broken ends and uses DNA ligase IV to reestablish the linkage (30–32).

NHEJ events keep the structure and continuity of the damaged chromosome, though produce sequence deletions ranging from several base pairs to thousands of base pairs. These events would affect normal gene function if the coding sequence or regulatory element get changed after repair. Available evidence also shows that NHEJ events are capable of efficient and accurate DSB repair when compatible DNA ends exist for direct ligation, which was called PL previously.

When the two-ended DSBs somehow still have complementary ends, PL directly ligates them, and preserves the original DNA sequence without any alteration (33, 34). Limited number of studies in DNA repair addressed PL since most recombination systems are designed to select events with sequence changes after repair. However, PL is likely the most time-and energy-efficient way of accurate repair if compatible ends remain after DSB. Using mutant I-SceI sites to recover re-ligation events, a recent study revealed that PL is the predominant way to repair DSBs with complementary ends, while

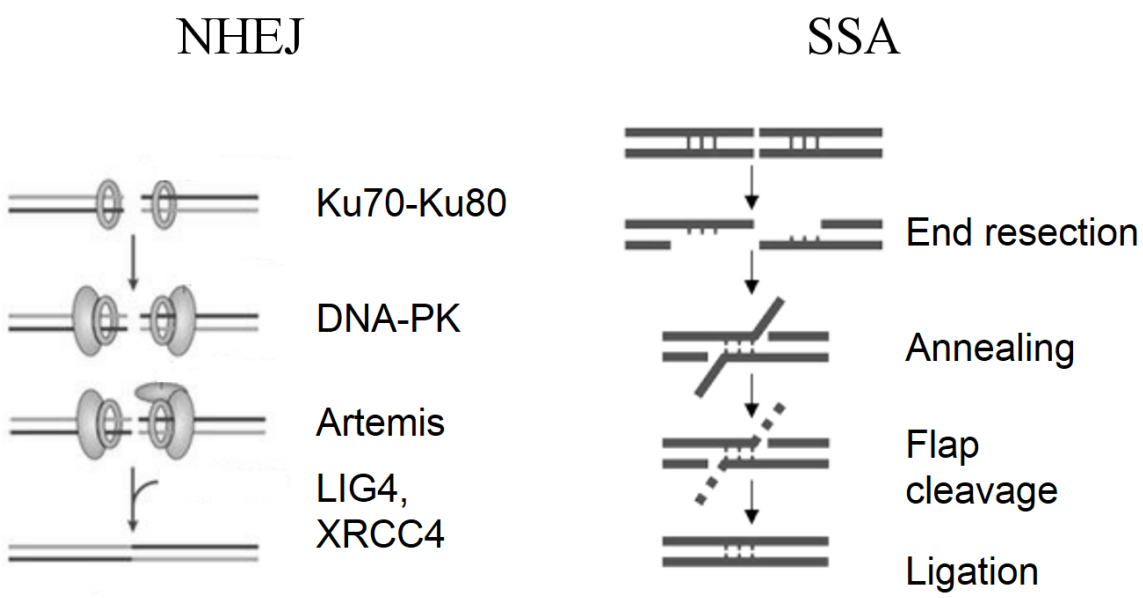

Figure 1.3 NHEJ and SSA pathways to fix DNA DSBs.
Overview of NHEJ and SSA pathway. Critical proteins in NHEJ are drawn on the left, while re-ligation procedure of SSA is drawn on the right. Adapted from Lazzerini-Denchi et al. 2016 (35) and Schubert et al. 2011 (36).

only 5% of the repair events are NHEJ with nucleotide deletions (37). Moreover, PL relies on NHEJ components, for example, Ku70, XRCC3 and ligase IV. It is conceivable that before the generation of NHEJ events with sequence deletion, the compatible ends were tethered together, and PL ligated them successfully multiple times (37, 38).

NHEJ efficiently repairs DSB without using a DNA template across cell cycles, and it may bring minor or moderate deletions to the DNA sequence. When NHEJ is delayed or damaged, cells may take an alternative pathway: alternative NHEJ (A-NHEJ) (31, 34, 38, 39). For example, NHEJ relies on ligase IV, Ku70 and DNA-PKcs. If cells have deficiency in these proteins, the broken ends may get extensive resection before ligation, therefore, repair events with long deletions become commonly observed. Sometimes, the broken end may be ligated with a broken ends from other breakages, producing chromosomal translocation or fusion (31, 40). In A-NHEJ, re-ligation of the broken ends relies more on microhomology, and utilizes a set of protein components including PARP1, CtBP-interacting protein (CtIP), ligase I and ligase III. Some of these protein components are also used by HR during strand resection (40, 41). Also, monitoring of the DNA repair process shows that A-NHEJ happens late in the DSB repair, and serves as a backup pathway when NHEJ failed in initial attempts.

If DNA DSB happens in one of the tandem repeats on a chromosome, the complementary sequences on either side of the breakage could be revealed though extensive unwinding or end resection (34, 42, 43). After strand annealing between these two tandem repeat sequences, the single strand overhangs/flaps are cleaved off, and then the strand ligation mediates the fusion of the two tandem repeats. SSA also uses

homology to repair DSB, but unlike HR, SSA mediated repair and sequence deletion are independent of Rad51.

Break induced repair (BIR) rescues a DSB event even when one side of a broken DNA molecule is lost after DNA damage, which is not cued by aforementioned pathways (44). BIR may use a nearby sister chromatid as template to accurately restore the broken chromosome: RAD51 directs the DNA single strand overhang to invade and establish a replication fork on the sister chromatid (Figure 1.4). Unlike semi-conservative DNA replication, BIR moves the replication fork through sliding of the D-loop, while the synthesis of lagging strand is carried out using the newly synthesized leading strand as template (45).Though BIR saves cells from deleterious loss of chromosome, it will produce chromosomal fusion or translocation if an improper template was chosen.

SSA, A-NHEJ and BIR are error prone pathways bringing profound changes to local DNA sequence and chromosomal structure. LOH, sequence deletion, chromosomal fusion and translocation are known outcomes if cells adapt these pathways. Since aforementioned repair pathways bring different risks to the genome, choice of pathway to counteract DSB is critical to the benefits of the host cells. These pathways are likely attempted in an order for the best of the cells, which may differ because of different cell lines, cell cycle or genetic background.

NHEJ is a prominent pathway to repair DSBs in all phases of cell cycle, and usually acts swiftly to fix the lethal DNA lesion. NHEJ does not need extensive resection and prolonged DNA synthesis; nucleotide cleavage and/or addition is sufficient to rejoin the broken DNA. If complementary ends still exist after DSBs, precise ligation of the broken fragments would dominate the ligation events. When the cells failed in this

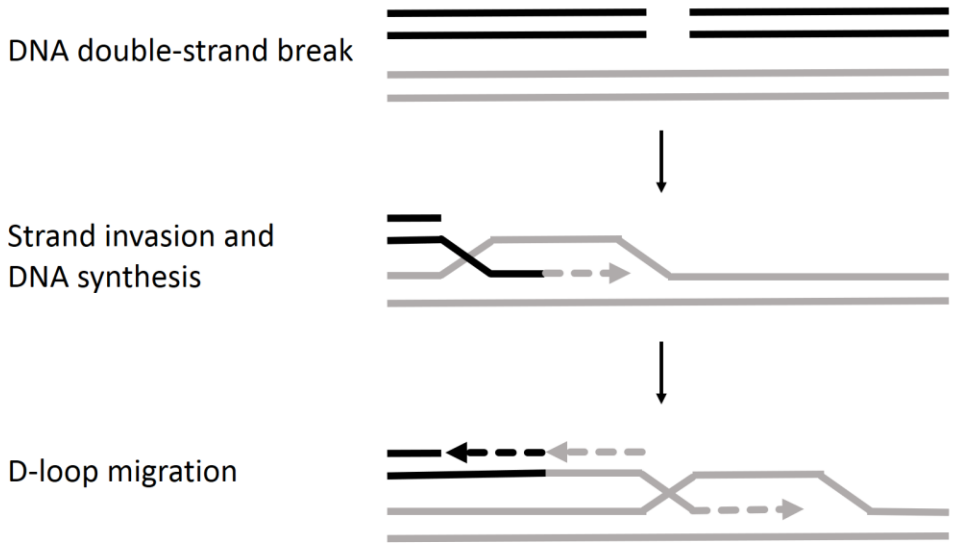

Figure 1.4 Break induced replication model (BIR).
BIR fixes DSB by re-synthesizing the lost DNA fragment. The DNA sequence of broken molecule is in black while sequence of template DNA is in gray. Newly synthesized DNA strand is drawn with dashed line, while the grey or black color represents the identity of the sequence. The arrowheads indicate the 3' end and the direction of DNA synthesis. Adapted from Donnianni et al. 2013 (45).

particular attempt, they would turn to error-prone NHEJ, HR or other alternative events for repair (32).

HR in somatic cells relies on recombinase Rad51, which is only detectable in G2/S phase, due to cell cycle dependent transcription (46, 47). In the meantime, sister chromatids provide the nearby homologous template for efficient recombination (19). Apparently, without the critical recombinase Rad51 for strand invasion and exchange, and without a nearby template, homologous paring and exchange between two sequences are not the preferred pathway for repair. When both NHEJ and HR are available to compete, the CtIP, recruited by MRE11, stimulates end resection and channels the repair toward HR. CtIP's activity relies on CDK-dependent phosphorylation (48, 49), and interestingly, RecQ4, one of the RecQ helicase facilitates CtIP's recruitment at the DSB site (15) while another RecQ helicase, WRN, suppresses the recruitment of CtIP and MRE11, propelling DSB repair through NHEJ without extensive end resection (41).

If initial attempts through NHEJ or HR failed, extensive resection and degradation may reveal remote homology or microhomology for re-ligation, so SSA or A-NHEJ would follow afterward. In NHEJ deficient cells, the repair pathway turns to A-NHEJ, which uses microhomology and ligase III to mediate the end rejoining (50, 51).

As stated in previous paragraphs, recombination mediators and interacting partners not only participate in DSB repair but also shift the repair toward different pathways and recombination products. RecQ4, one of the RecQ helicase, has been associated with DSB repair. Patients with defective RecQ4 have Rothmund Thomson syndrome, which was characterized with poikiloderma, cataracts, predisposition to osteosarcoma, and accelerated aging. It is of great importance to explore RecQ4's

function in DSB repair since the RecQ4 deficient cells show several typical signs of genome instability. Experiments were conducted on RecQ4 helicases by analyzing DSB-induced recombinants after siRNA knockdown in human fibroblast cells. Increase of gene conversion events was observed in RecQ4 deficient cells, opposite to the impact BLM deficiency brought. Moreover, BLM deficiency failed to increase crossover events in RecQ4 deficient cells while it dramatically boosted crossover events in wild type cells. Our results showed that RecQ4 and BLM shift the final products of HR toward opposite directions, indicating agonistic roles they may play in HR. Since crossover is potentially deleterious outcome in DSB repair, current genetic study provides preliminary data and evidence to understand RecQ4's impact on the genome stability, and these results warrant further mechanistic studies.

Another potential risk of HR lies in template selection. HR maintains the genome stability in somatic cells by cuing DNA breakage from radiation, genome toxic chemicals or endogenous cell stress (8, 29). However, if the Rad51 coated DNA overhang invades a non-allelic template sequence, and retrieves the incorrect DNA sequence, mutation and malfunction are imaginable (52). Sporadically, recombination between non-allelic sequences causes genetic diseases and contributes to cancer development (30, 52). When DNA recombination happens between imperfectly matching sequences (synonymous with diverged sequences and homeologous sequences), the restored DNA molecule would have certain DNA sequence replaced by that of the DNA template. These recombination events with a homeologous conversion tract are thus called homeologous recombination events (HeR events) (53–56). Due to HeR's deleterious consequences to gene function and genome stability, a lot of research have been done to address the mechanism how

cells avoid DNA recombination between diverged sequences. It has been confirmed in multiple model organisms that functional DNA mismatch repair (MMR) is needed to prevent recombination events between diverged sequence, however, at which stage the MMR intervenes the homeologous recombination and the detailed mechanism is not well known and still in debate (54, 57–59).

Pursuing the previous findings in our laboratory (53, 55), a systematic research has been done to explore the mechanism how mammalian cells exclude highly diverged sequences from DNA recombination, and the results are presented in current dissertation. New substrates were constructed to have different lengths of diverged sequences surrounded by perfect homology. Both spontaneous recombinants and DSB recombinants from new substrates were collected and analyzed. The obtained data clearly demonstrates that HeR events are common when sufficient homology for recombination locates on both sides of the homeologous sequences. Additionally, strand exchange between homeologous sequences is supported by evidence of heteroduplex DNA intermediate after MSH2 knockdown. These results argue that DNA recombination does not target and reject genetic exchange between highly diverged sequences in these new substrates. Though MMR prevents recombination from happening between imperfectly matching sequences, our results added to current paradigm that in certain circumstances, MMR does not impede the exchange between diverged sequences, but facilitates it by converting mismatched nucleotides to the recombination template. Moreover, our results revealed the relaxed homology requirement of DSB-induced recombination, which demonstrates the higher risk of HeR events during DNA damage and repair.

CHAPTER 2 RECOVERY AND ANALYSIS OF SPONTANEOUS AND DSB-INDUCED RECOMBINANTS IN MOUSE LTK- CELLS

HR is able to faithfully repair a damaged DNA molecule. It directs the DSB ends of the broken DNA to its homologous template, synthesizes the lost DNA sequence and restores the broken region seamlessly. However, improper choice of recombination template may bring HeR events: If the broken DNA sequence pairs with another non-allelic DNA sequence and gets synthesized depending on a wrong template, it will likely bring mutations and malfunction to the involved genes.

Our research focuses on the risk of homologous recombination between non-allelic, evolutionarily related fragments with high sequence divergence. These events may convert the DNA sequence of one fragment to the other, and sometimes they even recombine these two DNA molecules within the diverged sequence creating reciprocal crossovers. These events are collectively called homeologous recombination (HeR), since the HR took diverged sequences (homeologous sequences) as substrates. Many reports confirmed HeR events between non-allelic sequences in germ-line cells or somatic cells (52, 60). HeR events did happen between members of a gene family since they both share substantial homology and reside in nearby locations. HeR events also happened between highly diverged and physically separated repeat sequences like Alu sequences. One type of hereditary nonpolyposis colorectal cancer (HNPCC) was caused by chromosomal rearrangement between two Alu sequences (60). In addition, tens of thousands of degenerative pseudogenes in mammalian cells still have adequate homology to

recombine with their parent genes (61–63). All these HeR events between non-allelic substrates may bring fatal mutations to functional genes, produce chromosomal deletions or duplications, or even generate severe gross rearrangement by splicing non-homologous chromosomes. Studying the HeR events between these substrates allows us to understand the tight regulation cells maintain to achieve reparation rather than impairment.

Previous studies demonstrated that HR scrutinizes potential templates and only allows efficient exchange between perfectly matching sequences (64). The strict exclusion of diverged sequences during HR relies on DNA mismatch repair (MMR). Horizontal gene transfer (HGT) between closely related bacteria species elevates significantly when the recipient cells are deficient in MMR (54, 65). In yeast, active MMR prevents HR from happening between diverged sequences as well (59, 66–68). Systematic studies have been carried out in *Saccharomyces cerevisiae* comparing HeR events in wild type and MMR deficient cells: MMR reduces HeR events up to 30-fold as the sequence divergence increases from 0% to 9%, while beyond 9%, the rarity of HeR events is likely a combination of the 30-fold reduction enforced by MMR and weak interaction between highly diverged sequences (66). In mouse cells, a single nucleotide mismatching could reduce the spontaneous recombination 30-fold (26). In mammalian cells, HeR events between 1% to 1.5% diverged sequences increase 7 to 10-fold after ablation of MSH2 (3, 69). Beside MMR, other protein factors also contribute to the prevention of HeR events, probably by unwinding the potential mismatched heteroduplex DNA (hDNA) intermediate in the HeR process. Yeast Sgs1 helicase is vital to the rejection of HeR events (68), and recent results in our laboratory show that RecQ helicase BLM, the mammalian homolog of Sgs1, suppresses HeR events in human cells as well

(70). *In vitro* experiments show that BLM and MSH2 cooperate in detecting strand annealing between diverged DNA sequences and peeling them apart, therefore, aborting the strand invasion (56). The end resection and strand invasion toward homeologous sequence may be attempted multiple times before the DNA repair is redirected to alternative pathways.

A series of research projects have been carried out in our laboratory to investigate HeR events between highly diverged sequences in mammalian cells, and some representative substrates are illustrated in Figure 2.1. These recombination substrates were designed to collect recombination events involving diverged DNA sequences originally from herpes simplex virus: the DNA sequences of type 1 and type 2 herpes simplex virus thymidine kinase (HSV-1 TK and HSV-2 TK) share 80% homology, similar to that between highly diverged Alu sequences or other repetitive elements (53, 55). Spontaneous recombination between HSV-1 TK and HSV-2 TK sequences was not recovered in our previous experiments using the pHOME substrate, and the recombination frequency was estimated to be lower than 10^{-9} (event per cell division) (26, 55). When it comes to the recombination between HSV-1 TK recipient sequence and hybrid donor sequence of HSV-1 TK and HSV-2 TK (pHYB21A or pHYB12-8), the spontaneous recombination frequencies are at 10^{-8}, similar that with HSV-1 TK region of aforementioned donors (53). In these experiments on pHYB21A and pHYB12-8, only 1 out of 81 recombinant clones shows a conversion tract in the HSV-1 TK recipient toward HSV-2 TK sequence, a signature of HeR events. Exchange between homeologous sequence became common only in recombination substrate pHYB121, which carries HSV-1 TK recipient and hybrid donor composed of HSV-1 TK, HSV-2 TK and HSV-1

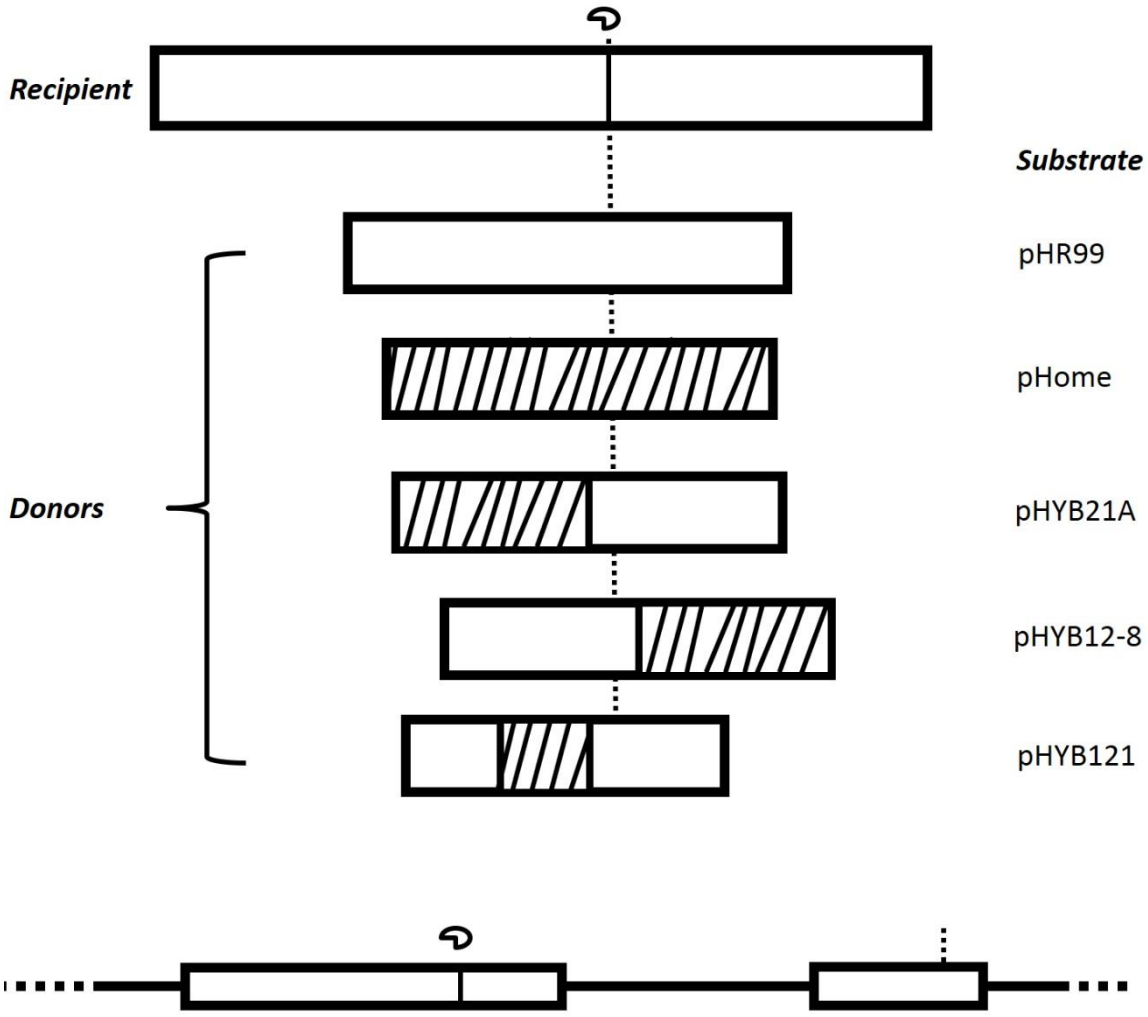

Figure 2.1 Alignment of recipients and donors from selected recombination substrates. The structures of donors and recipients used in previous experiments are shown above, and the alignment reveals the position of endonuclease recognition sites (XhoI or I-SceI) within recipients relative to the donor's sequence.
The white rectangles represent DNA sequences of HSV-1 TK and the striped rectangles represent HSV-2 TK sequences. The I-SceI or XhoI site in recipient is marked by a tilted partial circle and solid line. And the dash line runs across these donors to show the nucleotide position where recipients hold the oligonucleotides insertion harboring the recognition site of either XhoI or I-SceI. Arrangement of recipient and donor after chromosomal integration is shown below the alignment.

TK sequence. In this case, 11 out of 39 recombination events have gene conversions in HSV1-TK recipient, and the conversion tracts always include all the nucleotide mismatches between HSV-1 TK recipient and hybrid donor.

Considering the recombination events from pHYB21A and pHYB12-8, the spontaneous recombination events preferentially occur between perfectly matching sequences and avoid nearby homeologous sequences. The exchange between homeologous sequence during spontaneous recombination is barely allowed unless homology lies on both sides of the homeologous sequence, as shown by pHYB121. Therefore, apparently perfect homology is required at both the initiation and resolution site during HR. To further evaluate the supervision against homeologous sequences during HR, increasing lengths of homeologous sequence were engineered in new substrates to reveal how exchange between homeologous sequences is processed during spontaneous recombination. If the "homology" rule is enforced at both initiation and resolution, the recombination events in new substrates would be similar to that of pHYB121, even though they have different lengths of homeologous sequence inside the donor. Moreover, the proposed experiments help to clarify whether the homeologous sequences in the hybrid donors impede HR to achieve the genetic exchange between them, and whether they are actually targeted by aforementioned mechanism for rejection or destruction.

All three new substrates carry a full length HSV-1 TK recipient and a hybrid donor composed of HSV-1 TK, HSV-2 TK and HSV-1 TK sequences. Donors of these substrates differ in the length of the HSV-2 TK sequence, and their upstream and downstream HSV-1 TK sequences exceed the minimal efficient processing segment

(MEPS, approximate 200 bp in mouse cells) required for efficient recombination (26, 71, 72). Spontaneous recombination between donor and recipient may remove the I-SceI insertion from HSV-1 TK recipient, thus restore the function of the viral thymidine kinase. These recombination events enable host cells to survive and form resistant clones in culture medium containing HAT.

Analysis of these recombinants will reveal how clusters of mismatches affect strand exchange between DNA molecules during spontaneous recombination. If HR vigorously rejects homeologous sequences, the HeR recombination frequency or the proportion of HeR recombinants would decrease when the substrate has a longer homeologous region. Meanwhile, nucleotide mismatches between the donor and recipient allow us to examine the border of recombination events, so it is possible to see whether spontaneous recombination starts and ends exclusively in the homologous region, converting all mismatched nucleotides.

Ionizing radiation, reactive oxygen species, and other genome toxic chemicals cause DNA double-strand breaks (DSBs) directly or indirectly (9, 29). In proliferating cells, HR is a vital pathway to repair DNA DSBs and saves cells from cell cycle stalling. However, if not harnessed, HR may conduct genetic exchange between non-allelic DNA sequences, and some even generate chromosomal translocation or fusion. DNA recombination between highly diverged sequences, so called homeologous recombination (HeR), was also observed in mouse Ltk- cells during DSB repair (55). These HeR events happened between thymidine kinase sequences of type 1 and type 2 herpes simplex virus, and their frequencies are 600-fold less than HR events (53). The sequence exchange between homeologous region during DSB repair occurs more if DSBs provoked HR in a

nearby homology: The HeR frequency between HSV-1 TK recipient and hybrid donor composed of HSV-1 TK and HSV-2 TK sequences is 3 times more than HeR events between pure homeologous sequences.

As described in previous paragraphs and in Figure 2.1, HeR event between homeologous sequences was not recovered during spontaneous recombination (55, 73). Even half of the homeologous donor was replaced with homologous sequence to boost initiation of HR (pHYB21A or pHYB12-8), exchange between homeologous sequences is still a rare event (53). Spontaneous recombination prefers perfect homology to start and end, thus it either manages to incorporates all mismatched nucleotides in conversion tract once encountering homeologous sequence, or simply return to the initiation site for resolution without converting any nucleotide if homology is not available on the other side of the homeologous sequence (53). However, in DSB induced HR between HSV-1 TK and hybrid donor composed of HSV-1 TK and HSV-2 TK sequences, genetic exchange frequently propagates into nearby homeologous region, converting some mismatched nucleotides (55). Therefore, the homology requirement for resolution appeared to be somewhat relaxed in proximity to a DSB during DSB repair. To test this hypothesis, experiments were designed to examine whether DSB induced HR events frequently get resolved in the middle of the homeologous region when homologous region is available on both sides. Transferring of mismatched nucleotides from donor to recipient will reveal the range of homeologous exchange. After comparing the choices of resolution site, difference in homology requirement between DSBs induced HR and spontaneous recombination could be obtained.

To learn how HR processes homeologous sequences during DSB repair, endonuclease I-SceI was transiently expressed in selected cell lines. After the I-SceI introduced DSB gets repaired through HR, NHEJ or other alternative pathways, the host cell may gain a functional thymidine kinase and form a recombinant clone in HAT selection. Experiments were carried out on the same selected cell lines examined for spontaneous recombination, which have different lengths of homeology in the donor (55). Therefore, the comparison between DSB repair recombinants and spontaneous recombinants may reveal whether exchange between homeologous sequences might be processed differently during DSB repair. The HeR events and conversion tracts in these cell lines may answer the following questions: whether the homology requirement in resolution is indeed relaxed during DSB induced recombination, whether the homeologous region poses an obstacle to HR in conducting sequence exchange between them, and whether strand exchange with substantial mismatches is targeted for rejection or destruction. For example, if there is vigorous rejection of the homeologous sequence, clone frequency or proportion of HeR recombinants should decrease in cell lines carrying a longer homeologous region in their substrates.

Materials and Methods:

Construction of substrate: The backbones of new substrates are from vector pJS-1(72). Three substrates, pLD1, pHYB121, pJSBam8 were developed from pJS-1 in previous experiments (53, 55), and they are used directly to build new pBWW substrates (Figure 2.2 to Figure 2.5). pJSBam8 has a 2.5 Kb HSV1-TK gene inserted at the unique BamHI site of pJS-1, and the gene sequence has been rendered nonfunctional by 8bp oligonucleotides insertion containing XhoI I site at nucleotide position 1215. The

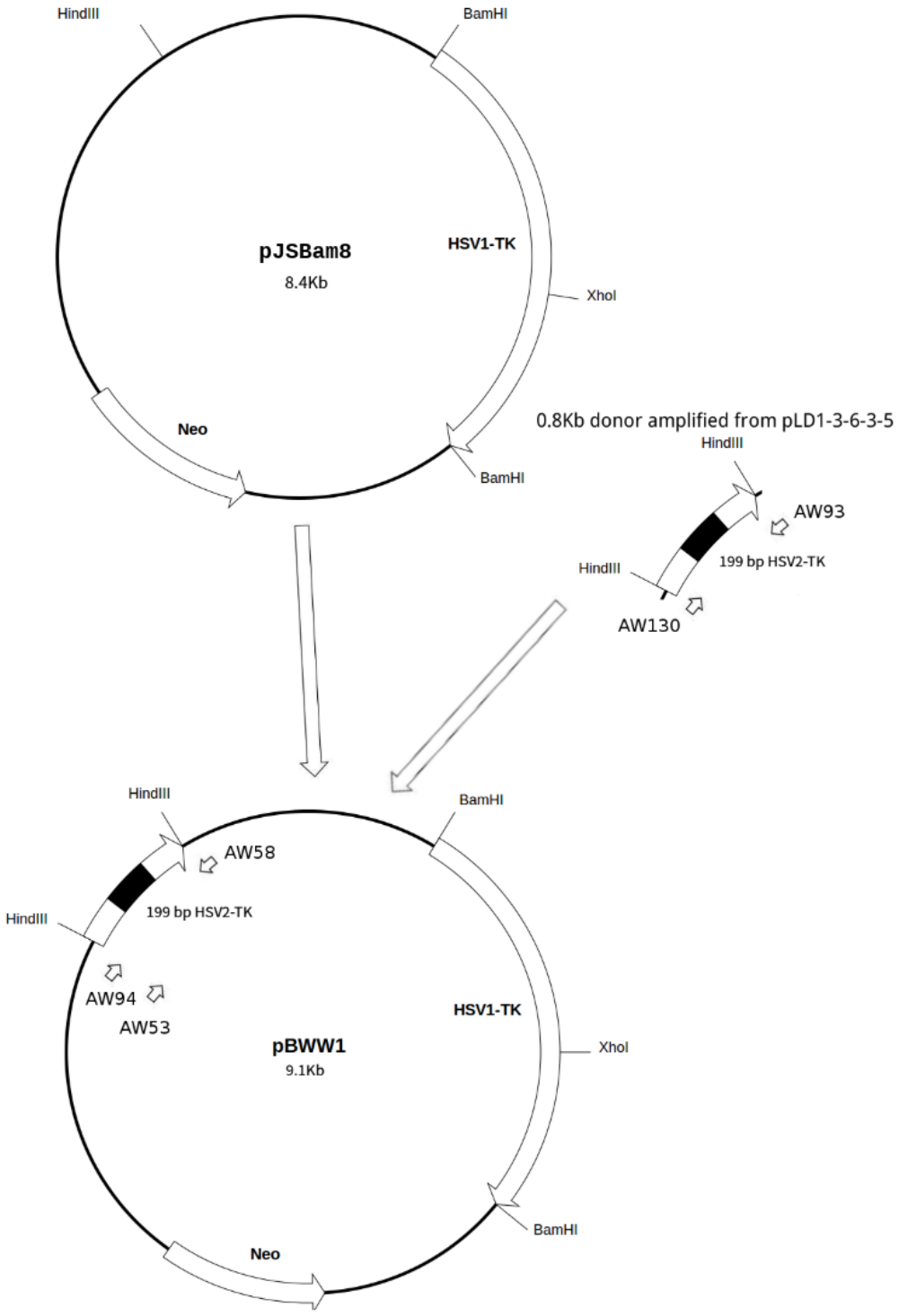

Figure 2.2 Workflow to construct substrate pBWW1.
Plasmids, primers and restriction sites in use are drawn and labeled in figure. The white bars in donors and recipient are HSV-1 TK sequences, while the solid bars in donors are HSV-2 TK sequences. pBWW1 has 33 mismatches between donor and recipient, and it was used to build 3 other recombination substrates examined in current studies.

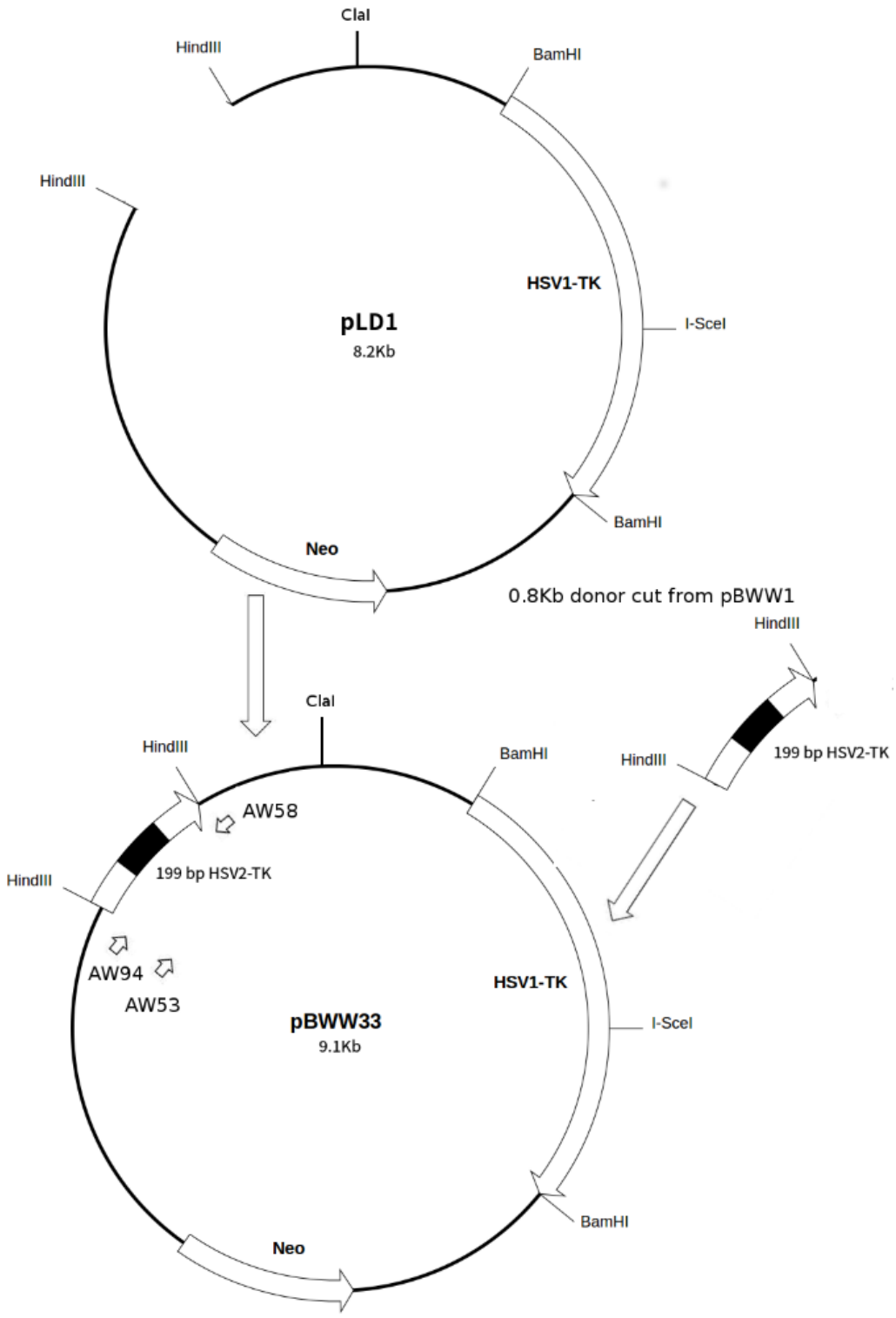

Figure 2.3 Workflow to construct substrate pBWW33.
Plasmids, primers and restriction sites in use are drawn and labeled in figure. The white bars in donors and recipient are HSV-1 TK sequences, while the solid bars in donors are HSV-2 TK sequences. pBWW33 has 33 mismatches between donor and I-SceI site containing recipient, and it was used in subsequent recombination experiments.

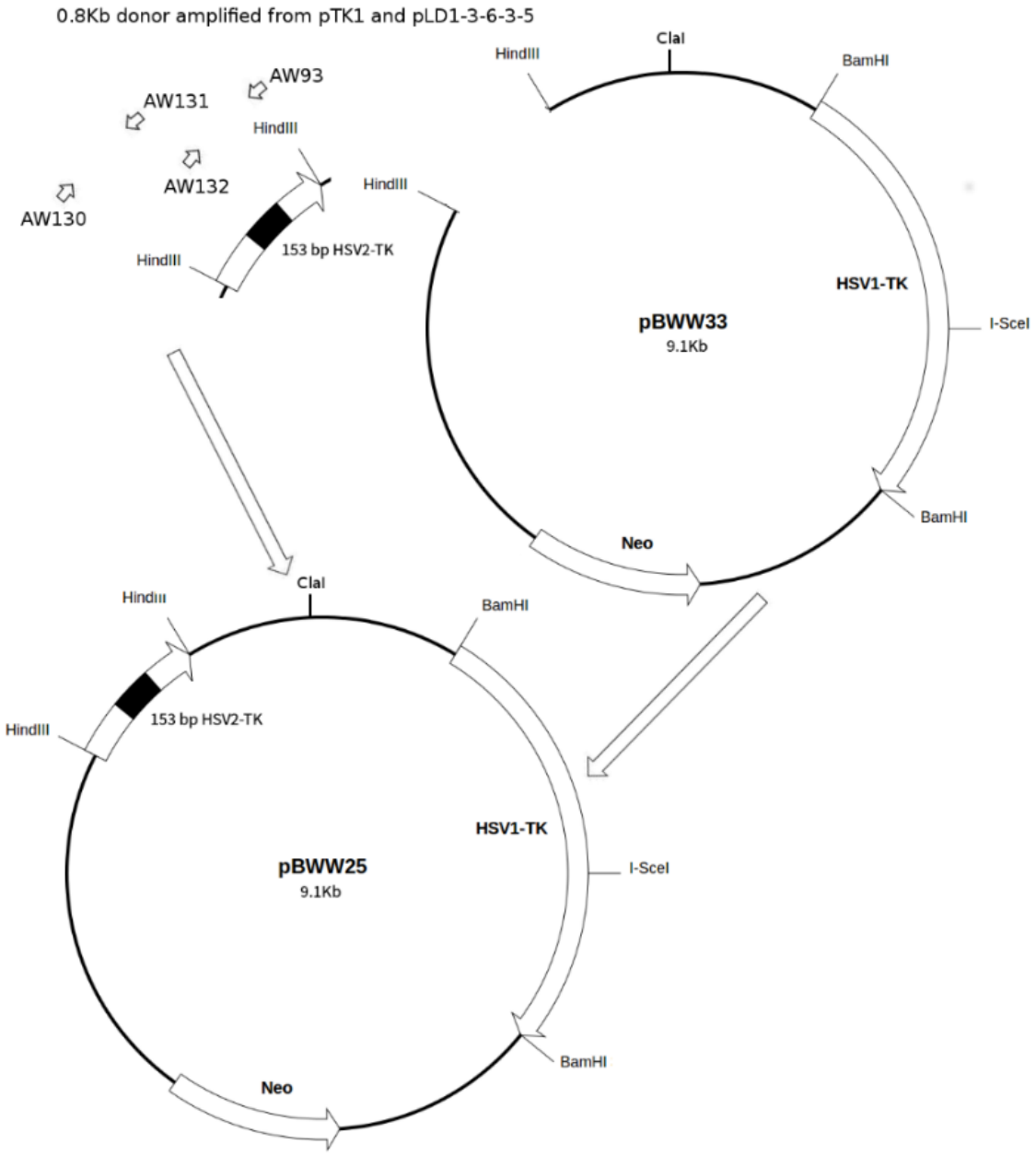

Figure 2.4 Workflow to construct substrate pBWW25.
Plasmids, primers and restriction sites in use are drawn and labeled in figure. The white bars in donors and recipient are HSV-1 TK sequences, while the solid bars in donors are HSV-2 TK sequences. pBWW25 has 25 mismatches between donor and recipient, and it was used in subsequent recombination experiments.

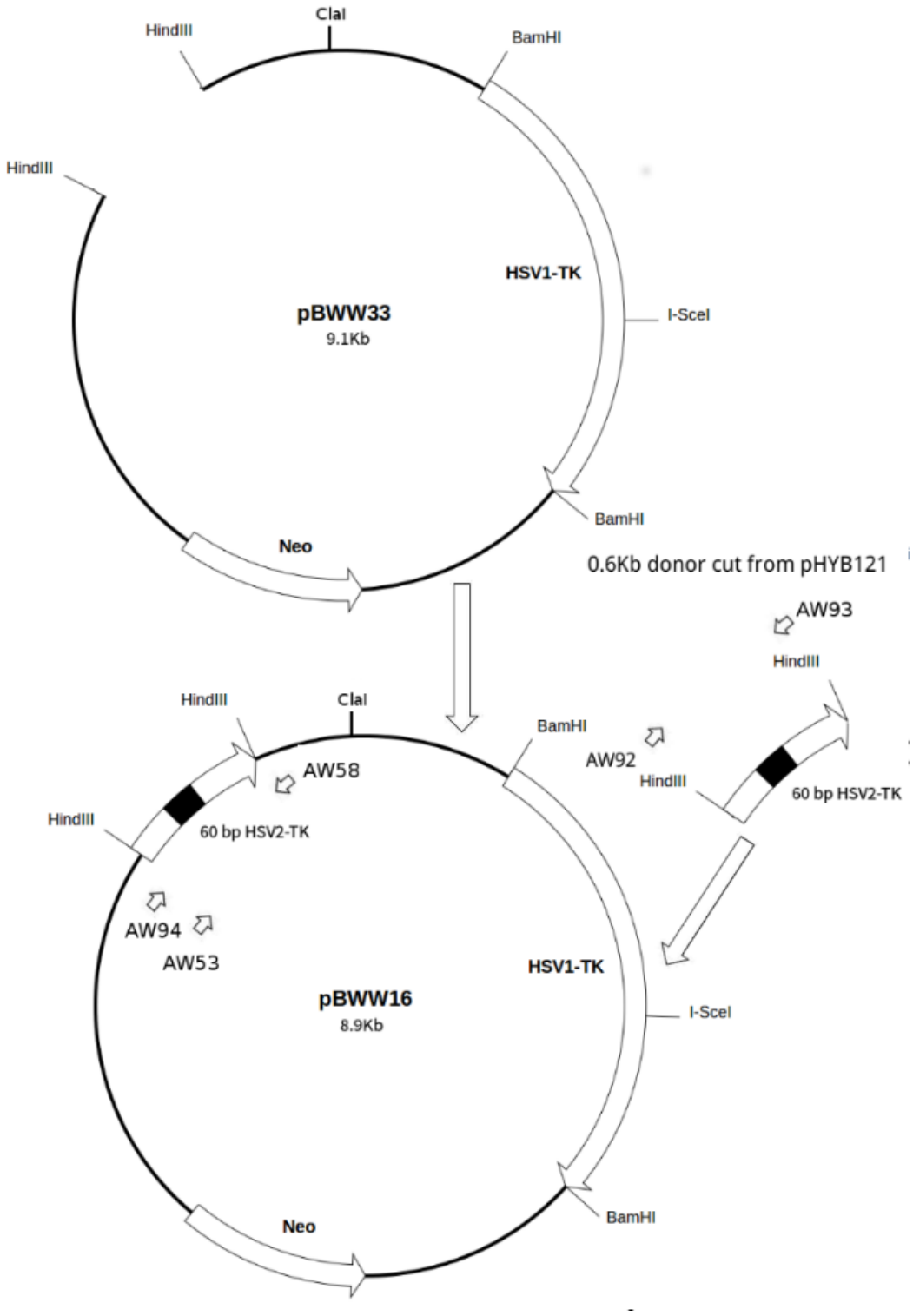

Figure 2.5 Workflow to construct substrate pBWW16.
Plasmids, primers and restriction sites in use are drawn and labeled in figure. The white bars in donors and recipient are HSV-1 TK sequences, while the solid bars in donors are HSV-2 TK sequences. pBWW16 has 16 mismatches between donor and recipient, and it was used in subsequent recombination experiments.

nucleotides in HSV1-TK are numbered the same as in the Wagner's paper hereafter (74). pHYB121 has the same nonfunctional HSV-1 TK recipient inserted at BamHI site as substrate pJSBam8, but it also has a hybrid donor inserted at the unique HindIII site. The hybrid donor of pHYB121 is composed of continuous HSV-1 TK, HSV-2 TK and HSV-1 TK sequences: nucleotides 848-1096 of HSV-1 TK sequence followed by nucleotides 1097-1167 of HSV-2 TK sequence followed by nucleotides 1168-1459 of HSV-1 TK sequence. pLD1 also has a 2.5 Kb HSV-1 TK sequence at the same BamHI site, but unlike pHYB121 and pJSBam8, a 30bp oligonucleotide was inserted at position 1215 of the tk gene introducing an I-SceI site. pLD1 has a hybrid donor of HSV-1 TK and HSV-2 TK sequence inserted at the unique HindIII site, which was excised and discarded during substrate construction. pTK1 has the original full length HSV-1 TK sequence at its unique BamHI site and it was used as a PCR template. One recombinant from substrate pLD1, pLD1-3-6-3-5, has a recombined HSV-1 TK recipient with nucleotides 968-1167 replaced by HSV-2 Tk sequence, so its recipient has 33 different nucleotides compared with the original HSV-1 TK sequence.

At the beginning, the hybrid donor of pBWW33 was amplified from cell line pLD1-3-6-3-5's genomic DNA using AW130 (5'ATT ATC TAA GCT TAG CCA CGG AAG TCC GCC TGG 3' which contains HindIII recognition site and matches HSV1-TK nucleotides 623-642) and AW93 (5'TGA TCA TAA GCT TAA GAC GTC CAA GGC CCA GG 3'which contains HindIII recognition site and matches HSV1-TK nucleotides 1459-1441). The 0.8 Kb PCR product was cut by HindIII, purified using a MinElute Gel Purification Kit (Qiagen) and ligated with the HindIII linearized pJSBam8 to produce substrate pBWW1. Therefore, the pBWW1 has the hybrid donor, 0.8 Kb HSV1-TK

sequence with 33 nucleotides changed to HSV-2 TK sequence, in addition to a XhoI site disrupted HSV-1 TK recipient. The donor in pBWW1 was then excised and ligated into the HindIII site of pLD1 to generate pBWW33.

The hybrid donor of pBWW25 was amplified from cell line pLD1-3-6-3-5‘s genomic DNA and substrate pTK1 through three steps of PCR. Firstly, AW130 and AW131(5’ GGC GAT AGG GTG CCG GTC GAA GAT GAG GG3’ which matches HSV1-TK nucleotides 1016-988) were used to amplify HSV1-TK nucleotides 623 to 1016 from pTK1. Secondly, AW132 (5’CCC TCA TCT TCG ACC GGC ACC CTA TCG CC3’ which matches HSV1-TK nucleotides 988-1016) and AW93 were used to amplify HSV1-TK nucleotides 988-1459 from pLD1-3-6-3-5. Lastly, these two PCR fragments were annealed using the overlapping HSV-1 TK nucleotides 988-1016 sequence, and then extended before the last PCR amplification by AW130 and AW93. After final PCR amplification, the complete donor fragment was cut by HindIII and purified as described above. The donor of pBWW16 was excised from pHYB121 using HindIII, and it was purified the same way as done to other donors. The previously made pBWW33 has the I-SceI disrupted HSV-1 TK recipient and hybrid donor carrying 33 nucleotides from HSV-2 TK sequence. The donor of pBWW33 was cut out by HindIII, and then the plasmid backbone was ligated with these two newly prepared donor fragments to generate pBWW25 and pBWW16 respectively.

Digested DNA fragments were separated on agarose gel, and then the target bands were excised and purified using a MinElute Gel Purification Kit (Qiagen). Or, the digested DNA fragments were separated by low melting point agarose gels and harvested

through phenol extractions. PCR products were subjected to additional gel purification before HindIII digestion if there were additional, nonspecific bands.

For each ligation reaction, 15ng donor fragment and 45 ng vector backbone were mixed with 1 uL T4 ligase (NEB) in 20 uL 1X ligation buffer (NEB), and the ligation reaction was kept at 16 ºC overnight. On the second day, 5 uL ligation solution was introduced into alpha-Select chemically competent bacteria (Bioline) as instructed by the supplier. Transformants were plated at different dilutions on LB media containing 40 ug/mL neomycin, and they were kept overnight at 37 ºC in incubator. Viable clones were picked and cultured overnight in 5 mL LB medium, and 4 to 5 mL bacterial culture was extracted for plasmid using Qiaprep Spin Miniprep Kit (Qiagen).

The donor insertion of pBWW33 was identified by HindIII digestion, and the size of insertion was confirmed by PCR amplification of the whole insertion using AW53 (5'TTA GCT CCT TCG GTC CTC CG 3' which matches vector pJS1 5094-5113) and AW58 (5'CCA ACT TAC TTC TGA CAA CG 3' which matches pJS1 5147-5128). Sequencing of substrate pBWW33 using primers AW94 (5'TAA TAC GAC TCA CTA TAG GGG GCT TCA TTC AGC TCC GGT TCC 3' which matches pJS1 5023-5044) confirmed the orientation of the donor insertion. For substrate pBWW16, AW94 and AW93 will amplify the correct donor insertion, while AW94 and AW92 (5' ATT ATC TAA GCT TGA TAT CGG CCG GGG ACG CGG 3' which matches HSV1-TK 848-867) amplify the reversed donor insertion. The donor of pBWW16 was amplified using AW94 and AW58 (5'CCA ACT TAC TTC TGA CAA CG 3' which matches pJS1 5147-5128) and sequenced before use. Finally, I-SceI digested substrates were checked on agarose gel to make sure they have intact recognition site of I-SceI within HSV1-TK recipient.

Cell culture: Mouse Ltk- cell line and its derivative cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS), minimal essential medium nonessential amino acids and 50 ug of gentamicin. All cells were incubated at 37 ºC with 5% CO_2.

Cell line establishment: Mouse Ltk- cells were transfected with each substrate through electroporation, and the cells with chromosomal integration of the substrate were selected using G418. Five million cells were pelleted in a 15 mL conical tube, washed one time by 5 mL PBS and then suspended in 775 uL PBS. The ClaI linearized substrate (2.5 ug in 25 uL digestion buffer) was added into the cell suspension to make it 800 uL before electroporation. The electroporation was carried out in a cuvette with 0.4 cm gap width at 1000 volts and 25 microfarads, and then the time constant was always 0.4 milliseconds. Right after electroporation, all the cells were transferred into a T175 flask and they were cultured for 2 days before G418 selection. Two days later, the cells were trypsinized, and 100,000 cells were plated in each T75 flask with 200 ug/mL G418. Syringe transfection (75) was attempted one time on pBWW33. Linearized pBWW33 (1.25 ug) was mixed with 0.5 million cells in 1mL culture medium, and then the cell suspension was taken into a 1ml syringe. All cells were forced through 30-gauge needle 4 times (8 strokes including the first transfer) to allow plasmid penetrating cell membrane. After the procedure, all cells were plated into 10 T75 flasks. On the next day, G418 was added into the culture medium at 200 ug/ml. The G418 selection usually took 11 to 14 days before resistant clones grew up.

G418 resistant clones were picked and cultured separately in a 24 well plate. After 2 weeks, all cells in a well were trypsinized and transferred in one T25 flask. Upon

confluence, cells of each flask were separated into two T25 flasks: cells in one flask were frozen and stored at -80°C, while the cells in second flask were used to prepare genomic DNA.

Genomic DNA of these clones were extracted as described before (73), and their concentrations were measured by spectrophotometer. For Southern blot, 8 ug genomic DNA was digested by HindIII or BamHI, separated on agarose gel and transferred onto nitrocellulose. Meanwhile, 10 pg digested substrate was loaded aside to display the intact donor or recipient, and their signals are equivalent to a single copy integration in mice genome. Full length HSV-1 TK sequence was used to synthesize probe, which hybridizes with recipient and donors (76). Cell lines were chosen for further experiments if they have intact donor and recipient, and their signal indicate a single copy integration of the substrate.

Collect spontaneous homeologous and homologous recombinants: Fluctuation test was carried out on each cell line to collect spontaneous recombinants as described before (77). In fluctuation test, 10 subclones were grown up in parallel. Initially, 100 cells were seeded per well in 24 well plate to start a subclone. Once confluent, the cells from each well were trypsinized and transferred into T25 flask. They were subsequently transferred into T75 flask, and then two T175 flasks, where the cells grew up to approximate 40 million. Finally, for each subclone, cells from the 2 T175 flasks were trypsinized and plated in HAT selection. Maximal 10 million cells were plated in a T175 flask with 30 mL 1 X HAT medium to achieve effective selection as well as proper clone density. HAT resistant clones usually became visible 2 to 3 weeks afterward, and then they were picked and cultured for sequence analysis as described before.

Calculation of clone frequency, homologous recombination frequency and homeologous recombination frequency: In fluctuation test, the clone frequency was calculated for each subclone by dividing the number of HAT resistant clones by the number of cells subjected to HAT selection. Then, the clone frequencies of 10 subclones were averaged to generate the clone frequency of the cell line. Similarly, the HR frequencies of the 10 subclones were calculated separately, and then they were averaged for the cell line. For each subclone, the HR frequency was calculated by multiplying clone frequency of the subclone with the proportion of HR clones in analyzed clones. HeR frequencies were calculated for each subclone separately the same as HR events, and they were averaged to generate HeR frequency of the cell line. If there were other events, like the reactivation of mouse thymidine kinase pseudogene, these events were still counted in clone frequency, therefore, the spontaneous recombination frequency would be different from the clone frequency and calculated separately case by case.

PCR amplification and sequencing: All PCR amplifications were performed using GE's illustra Puretaq Ready-To-Go PCR Beads as instructed by the supplier. The dried PCR bead was dissolved in 25 uL reaction solution containing DNA template and 50 ng of each primer. And this 25 uL reconstituted PCR mixture was overlaid with mineral oil to prevent evaporation in following thermal-cycling.

PCR amplification of HSV-1 TK sequence used following "Touchdown" protocol: The initial denaturation starts at 95ºC for 5 minutes. In the first 12 cycles, the 1 minute annealing step starts at 72ºC and then the temperature drops 2ºC after every two cycles until it reaches 62ºC; the elongation takes 3 minutes at 72ºC, and the denaturation takes 1 minute at 95ºC. In next 24 cycles, annealing takes 1 minute at 60ºC, elongation takes 3

minutes at 72ºC and denaturation takes 1 minute at 95ºC. Final elongation step takes 10 minutes at 72ºC.

To analyze collected recombinants, the HSV-1 TK recipient sequence was amplified by primer pair AW85 and AW133 or primer pair AW100 and AW133 from 500 ng genomic DNA using touchdown protocol. These two primer pairs amplify 1.4Kb or 1.1Kb sequences from the HSV1-TK recipient, and both products include the nucleotides 968-1126 of HSV-1 TK sequence, which would be altered by these recombination events. To prepare these PCR products for sequencing, 10 uL PCR solution was treated by EXOI and SAP (NEB) for 15 minutes at 37ºC, and then the mixture was incubated at 80ºC for 15 minutes to deactivate the added enzymes. Samples were sequenced using T7 promoter primer at the core facility in University of South Carolina or at ETON Bioscience Inc.

Analysis of spontaneous homeologous and homologous recombinants: All alignments were carried out using SciEdCentral software to compare recombinant DNA sequences with reference sequences (HSV-2 TK sequence and HSV-1 TK recipient sequence).

Transfection by electroporation: To collect DSB repair recombinants, the plasmid pSce was electroporated into cells to introduce a DSB in the recipient at the I-SceI site. Before electroporation, cells were cultured in T75 flasks until their number exceeded $5X10^6$. On the day of transfection, cells in T75 flasks were trypsinized, re-suspended in culture medium and counted on a hemocytometer. For each transfection, 5 $X10^6$ cells were transferred into a 15 mL conical tube, and centrifuged at 300 x g for 3 minutes. The cells pellet was washed using 5 mL phosphate buffered saline (PBS), and then centrifuged again as described above. The newly pelleted cells were re-suspended in

800 uL PBS before electroporation. For each electroporation, 20 ug pSce plasmid was first added to the cuvette, and then 800 uL cells suspension. The DNA and cells were mixed in the cuvette by pipetting several times. All electroporation experiments were performed using a Bio-Rad Gene Pulser set to 750 volts and 25 uF. The actual time constant of these electroporation experiments was always 0.4 ms.

Immediately after electroporation, cells were transferred into a 15 mL conical tube containing 4.2 mL culture medium to obtain a final concentration of $1X10^6$ cells/mL. A small portion of the electroporated cells was removed to measure plating efficiency as described in next paragraph, while the rest cells were plated into 3 T75 flasks for subsequent HAT selection.

Measurement of plating efficiency: To obtain plating efficiency, 100 electroporated cells were transferred into each one of the two T25 flasks and cultured for two weeks. The clone numbers in these two flasks were averaged and then divided by 100 to give plating efficiency of the experiment.

Selection and culture of recombinant clone: The electroporated cells were separated into three T75 flasks. The cells were cultured in 15 mL media for 2 days before being fed with HAT selection media. About 2 weeks after selection, discrete clones formed in T75 flasks, and they were randomly picked for further analysis (55). The HAT resistant clones in T75 flasks were counted to calculate clone frequency. The total clone number in three flasks was divided by $5X10^6$, and then divided by plating efficiency to generate clone frequency.

Southern blot: The genomic DNA of selected recombinants was digested by BamHI and HindIII, and then separated through electrophoresis for Southern blot.

Moreover, HindIII digestion was used to check the junction fragment in some recombinants that were suspected to carry more than one copy of substrate.

Subcloning of cells: To obtain subclones of an existing cell line, the cells were diluted to 100 cells/mL Then, 0.5 mL diluted cell suspension was transferred into a T75 flask together with 15mL media and incubated for 2 weeks. Discrete cell clones formed in the T75 were picked and cultured separately for later analysis.

Results:

Structure of substrates: Three new substrates all have a full length HSV-1 TK recipient and a hybrid donor made of HSV1-TK and HSV2-TK sequences. The HSV-1 TK recipient in three substrates is the same and it has 30bp insertion at position 1215: the nucleotides are numbered the same as in the Wagner's paper (74). This 30 bp nucleotide insertion contains the recognition site of endonuclease I-SceI, and destroys the thymidine kinase's function encoded by the recipient. Donors in these substrates range from 612 bp to 837 bp, and they are hybrid donors composed of continuous HSV-1 TK, HSV-2 TK and HSV-1 TK sequences. All the donors have upstream HSV-1 TK sequences between 250 bp and 380 bp, then HSV-2 TK sequences between 60 bp and 200 bp, and finally the same downstream HSV-1 TK sequence of around 300 bp. When the HSV-1 TK recipient is aligned to these three hybrid donors (Figure 2.6), the I-SceI insertion site always locates against the downstream HSV-1 TK sequences of these donors, and about 50 bp away from the border of HSV-2 TK sequences. The HSV-2 TK sequences in these substrates are 60 bp, 153 bp and 199 bp respectively, and they carry 16, 25 and 33 nucleotide mismatches compared with the HSV-1 TK recipient (Figure 2.7). These substrates are named pBWW16, pBWW25 and pBWW33 after the number of

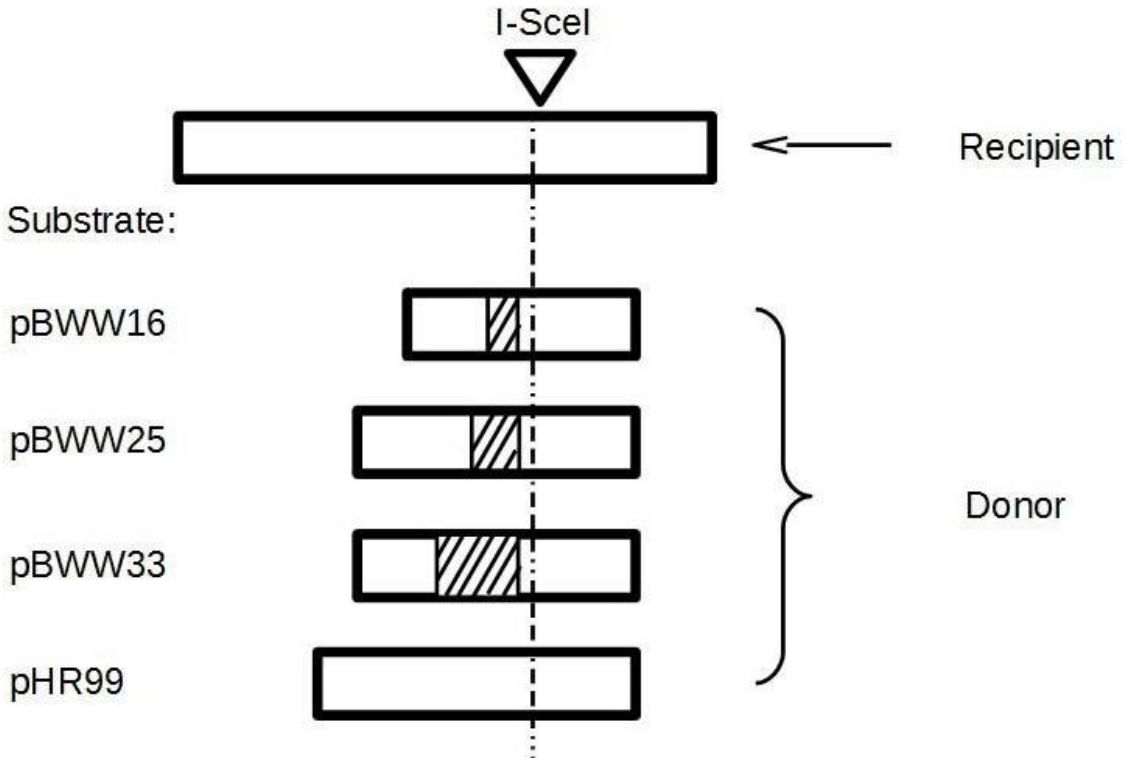

Figure 2.6 Alignment of donors and recipients from pBWW plasmids.
The structures of donors and recipients used in current experiments are shown above, and the alignment reveals the location of recipient's I-SceI recognition site, relative to the donor's sequence. The white rectangles represent HSV-1 TK sequences and the striped rectangles represent HSV-2 TK sequences. Previously constructed substrate pHR99 has similar structure, while it carries a HSV1-TK donor slightly longer than the rest donors.

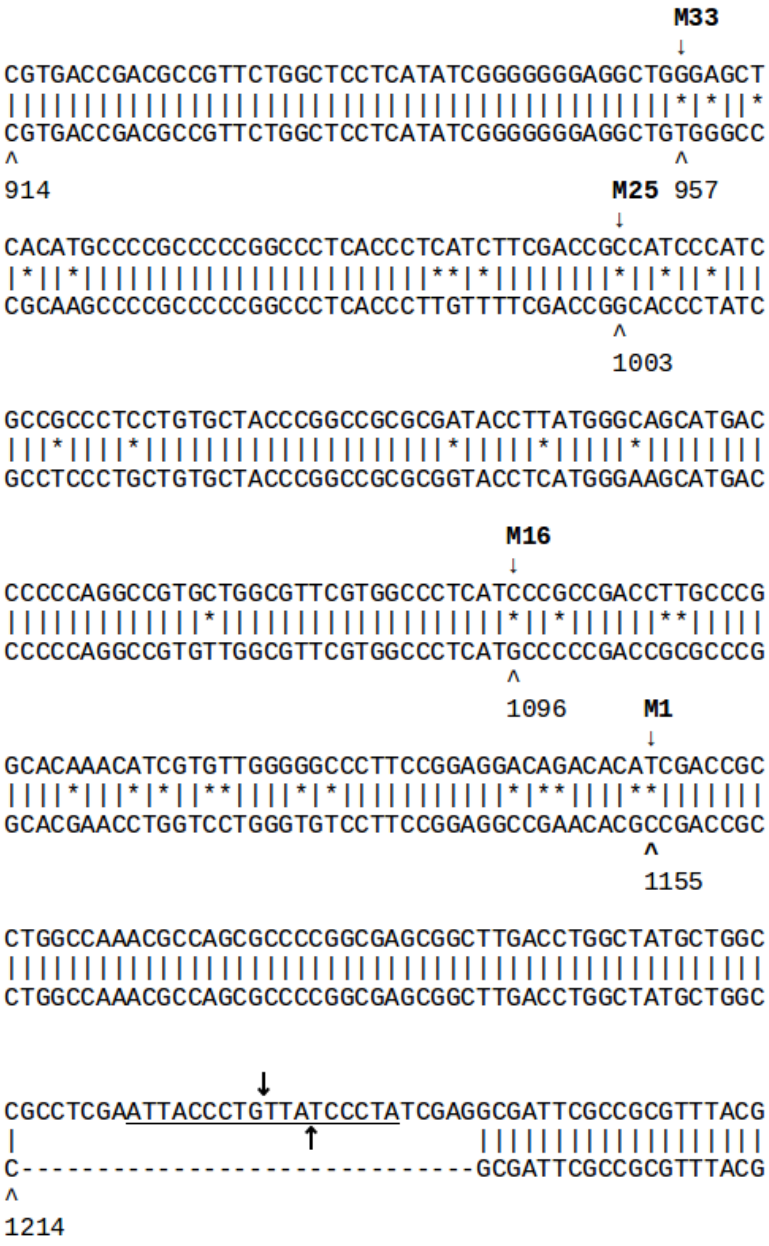

Figure 2.7 Alignment of recipient and donor sequences in substrate pBWW33 to show mismatched nucleotides.
Compared with donor, the recipient has 30 bp oligonucleotide insertion including an 18 bp recognition site of I-SceI (underlined). Totally 33 mismatched nucleotides (indicated by the *) exist between donor and recipient, and they are numbered from M1 to M33 depending on their closeness to the I-SceI site. The labeling of mismatched nucleotides is opposite to the nucleotide sequence of tk: the first mismatched nucleotide is the remotest from I-SceI, thus labeled as M33 while the last mismatched nucleotide is labeled as M1. Mismatched nucleotides M16 and M25 are labeled here since they are first mismatched nucleotides in substrate pBWW16 and pBWW25. Below the DNA sequence, the nucleotides are numbered with "^" the same as in the Wagner's (74).

mismatched nucleotides in their hybrid donors. These substrates were linearized by endonuclease ClaI before transfection, and Figure 2.8 shows the restriction map of these substrates. To check the completeness of these tandem repeats after genome integration, endonuclease BamHI and HindIII were used separately to digest genomic DNA for Southern blot.

Construction of plasmids: pBWW1 and pBWW33 were constructed by Bryan Wehrenberg and Barbara Waldman. pBWW1 plasmid with correct insertion and sequence were confirmed through PCR (Figure 2.9) and DNA sequencing. AW53 and AW58 amplified the whole insertion at HindIII site of the substrate, and these 0.9 Kb PCR products indicate one copy of donor insertion. Plasmids of 7 bacterial clones pBWW1-6 to pBWW1-12 were checked by PCR and DNA sequencing. Donor in pBWW1-8 has the correct direction and sequence, so the plasmid was renamed as pBWW1 and used in following experiments.

Substrate pBWW33 and pBWW25 were constructed by Bryan Wehrenberg and Barbara Waldman, and their donors were examined and sequenced as described above.

To make substrate pBWW16, donor fragment and plasmid backbone were excised by HindIII from pHYB121 and pBWW33, respectively. They were gel purified and checked before ligation (Figure 2.10).

Correct pBWW16 plasmids were found through enzyme digestion (Figure 2.11) and PCR amplification using primer pair AW53 and AW58 or AW53 and AW39 (Figure 2.12). The direction and sequence of donors were further examined through DNA sequencing. Donor of pBWW16-2 has correct direction and sequence, so the plasmid was renamed as pBWW16 and used in following experiments.

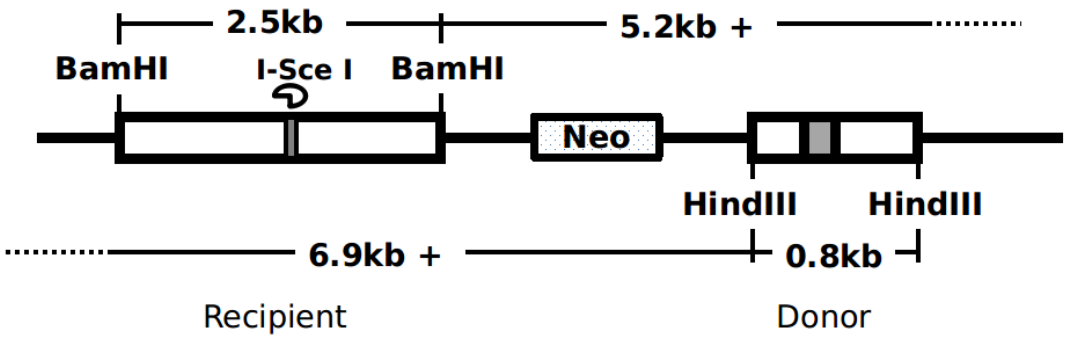

Figure 2.8 Structure and restriction map of pBWW substrates.
Restriction map of pBWW substrates shows their structure after chromosomal integration, and it also demonstrates the size of DNA fragments after HindIII or BamHI digestion. The linearized substrate is about 9.1 Kb. They have a 0.6 Kb or a 0.8 Kb hybrid donor flanked by HindIII sites, and a 2.5 Kb HSV-1 TK recipient flanked by BamHI sites. The Neo cassette between donor and recipient is used for selection in prokaryotic or eukaryotic cells.

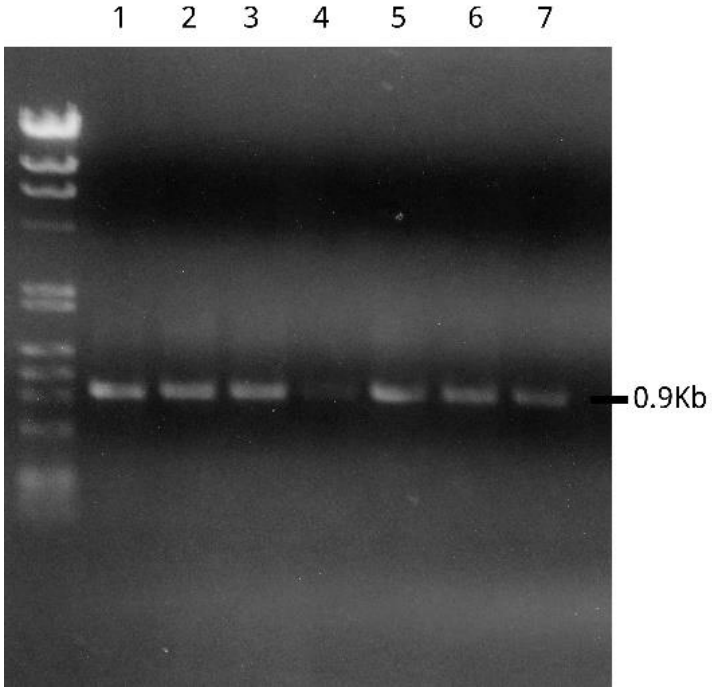

Figure 2.9 Identification of plasmid pBWW1 with 0.9 Kb donor insertion using primers AW53 and AW58.
Lane 1 to 7, PCR product of pBWW1-6, 7, 8, 9, 10, 11 and 12.

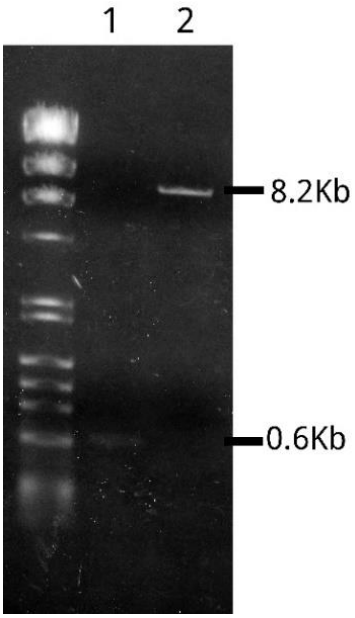

Figure 2.10 Fragments of donor and backbone were examined before ligation for plasmid pBWW16.
Lane 1, 0.6 Kb donor. Lane 2, 8.2 Kb plasmid backbone of pBWW33.

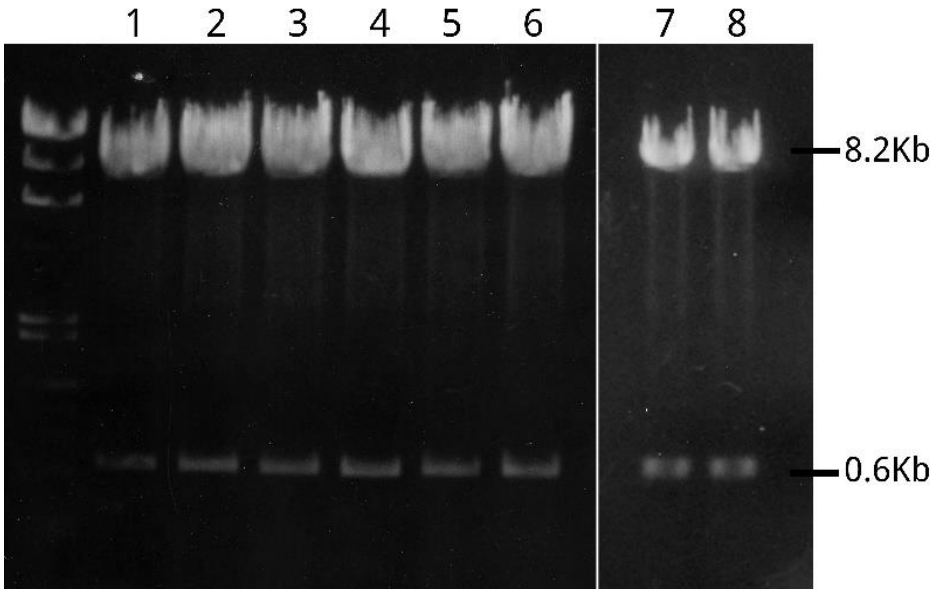

Figure 2.11 Identification of plasmid pBWW16 with 0.6 Kb donor insertion.
Lane 1 to 8, HindIII digestion of plasmids from clone pBWW16-1 to pBWW16-8.

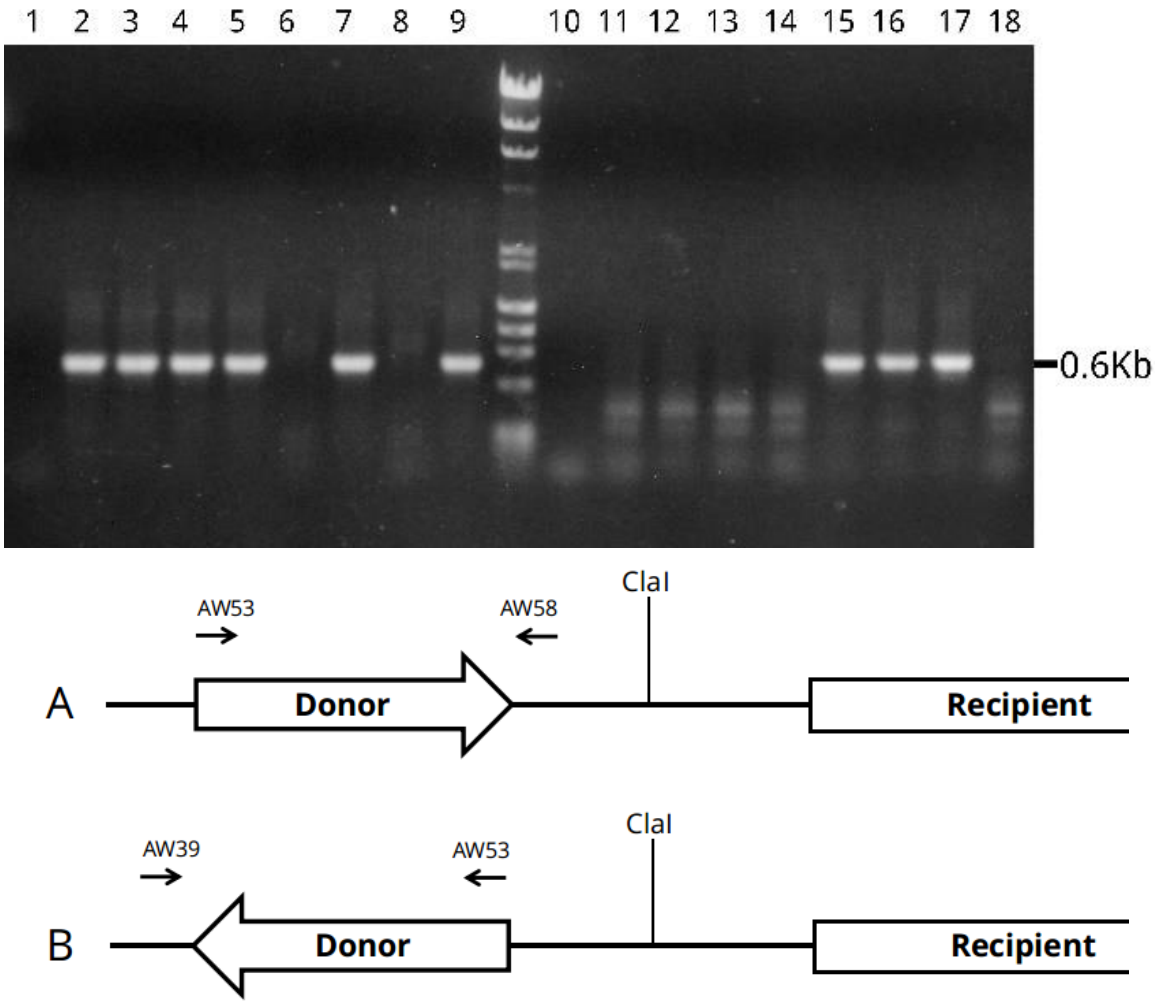

Figure 2.12 Direction of donor insertions in plasmid pBWW16 was confirmed by PCR. Top: Lane 1 and 10, PCR product of negative control; lane 2-9, PCR product from clones pBWW16-1 to pBWW16-8 using primer pair AW53 and AW58; Lane 11 to 18, PCR product from clones pBWW16-1 to pBWW16-8 using AW53 and AW39. Bottom: schema of the PCR strategy. Plasmids of pBWW16-1,2,3,4,8 have the insertion in correct direction and they were further sequenced for accuracy.

Substrates pBWW33, pBWW25 and pBWW16 were cut by I-SceI and examined on gel. The digestion of pBWW16 shows it is linearized by endonuclease I-SceI (Figure 2.13).

Establishing cell lines: Substrates pBWW33, pBWW25 and pBWW16 were linearized by digestion with ClaI (Figure 2.2 to 2.5) before transfection. The mouse Ltk- cells were transfected with each substrate separately through electroporation, and then the treated cells were selected by G418 for stable integration of the substrate. Syringe transfection was only carried out on substrate pBWW33 once, and the treated cells were subjected to the same G418 selection. G418 resistant clones from both methods were screened by Southern blot to find cell lines with a single copy integration of each substrate.

Screening cell lines with single copy integration of substrate: Genomic DNA from each clone was digested using HindIII or BamHI, and loaded together with the corresponding digestion of the substrate for Southern blotting. In HindIII digestion, the cell line with intact substrate should show a DNA fragment lining up with the donor of the substrate, and a junction fragment larger than the non-donor fragment of the digested substrate. Similarly, in BamHI digestion, the cell line should have a 2.5Kb fragment lining up with the recipient of the substrate, and a junction fragment larger than the non-recipient fragment of the digested substrate. All cell lines in use were checked by both digestions to make sure they have intact donor and recipient as well as only one junction fragment. The copy number in cell lines was also estimated by comparing signals of genomic DNA and the plasmid control.

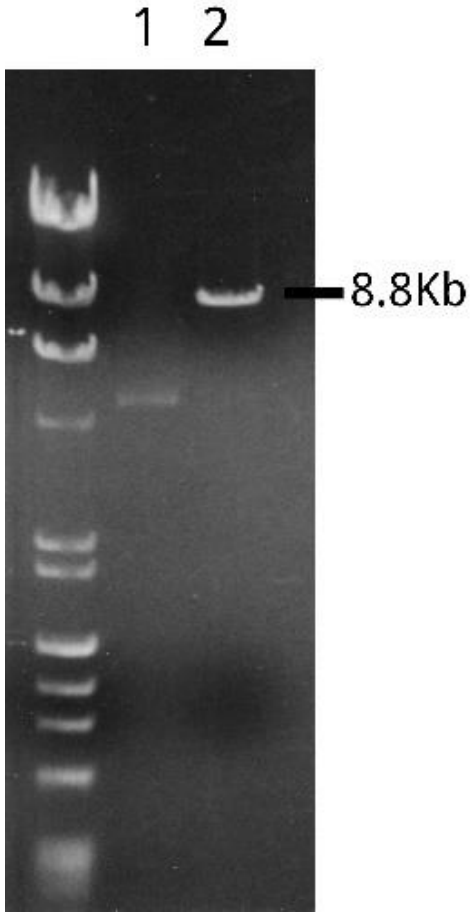

Figure 2.13 I-SceI successfully linearized pBWW16.
Lane 1, pBWW16 plasmid. Lane 2, I-SceI digested pBWW16 plasmid

Southern blotting was carried out on pBWW25 stably transfected cell lines by Bryan Wehrenberg and Barbara Waldman, and the results were recorded in Bryan Wehrenberg's undergraduate Honors thesis entitled "Homeologous Recombination in Mammalian Cells". Southern blot results of pBWW33 and pBWW16 were described in this chapter, and the selected Southern blot images of pBWW33 and pBWW16 stably transfected cell lines are shown in Figure 2.14, Figure 2.15, Figure 2.16 and Figure 2.17. The HSV-1 TK probe recognizes the HSV-1 TK sequence in the recipient as well as that in the hybrid donors. Cell lines were chosen for future experiments only when they had intact donor and recipient bands, and one junction fragment of proper size. These donor or recipient bands in selected cell lines are generally lighter than peers, so they have better chance to have single copy integration of the substrate.

For each substrate, more than 2 cell lines with single copy integration were found. The Southern blot results of selected cell lines pBWW33-48, pBWW33-67, pBWW16-2 and pBWW16-6 were labeled in these figures (Figure 2.14 to Figure 2.17).

Spontaneous recombination frequency and HeR frequency of examined cell lines: Fluctuation tests were performed on selected cell lines, and the statistics and clone frequencies are shown in Table 2.1 to Table 2.5. One additional fluctuation test was carried out on cell lines pBWW33-67 and pBWW33-48, which have 33 mismatched nucleotides between donor and recipient sequences.

The obtained spontaneous recombination frequencies from examined cell lines range from $2.33X10^{-8}$ to $5.38X10^{-7}$, while their HeR frequency ranges from $1.54X10^{-8}$ to $1.83X10^{-7}$. In contrast, the homologous substrate pHR99 has recombination frequency at $5.72X10^{-7}$ with no HeR events.

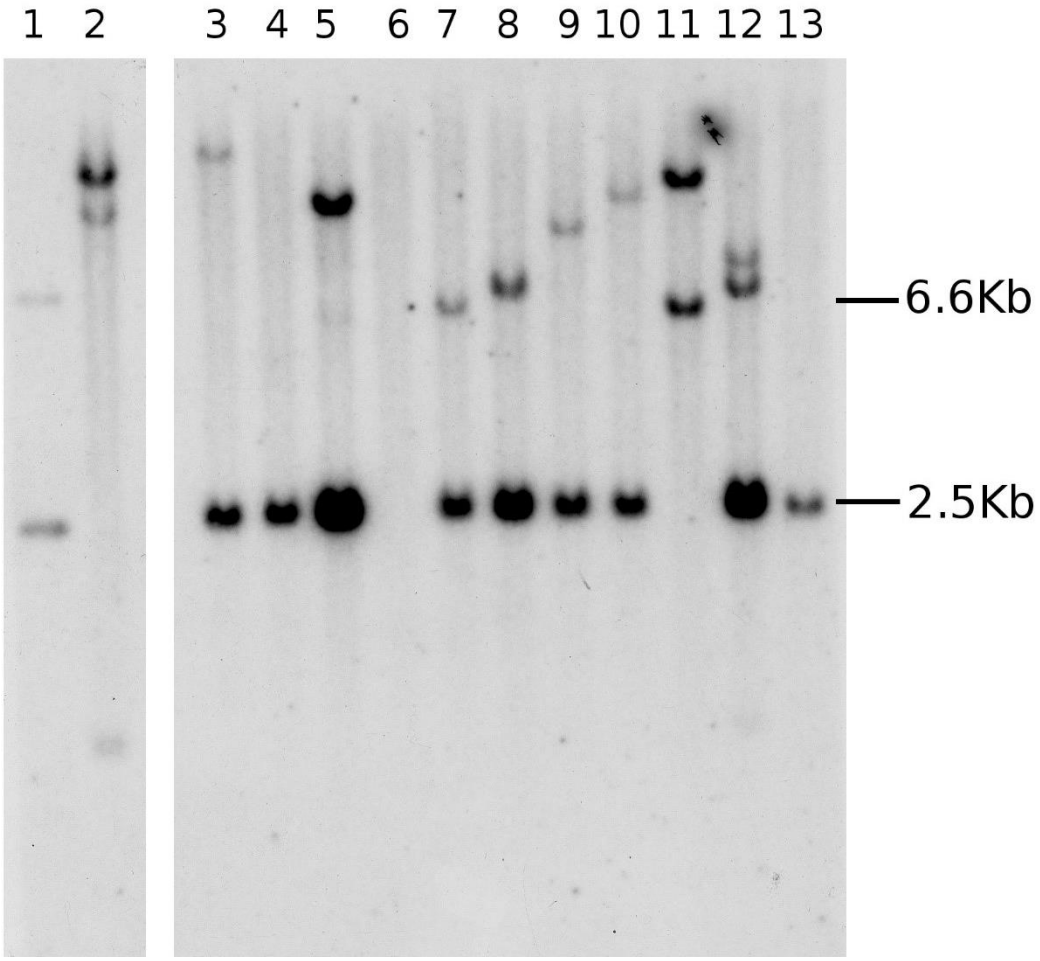

Figure 2.14 Identification of pBWW33 cell lines with correct plasmid integration through BamHI digestion.
The size and copy number of recipient were screened as described above. Lane 1, BamHI digested plasmid pBWW33. Lane 2 to 13, BamHI digested genomic DNA from pBWW33 stably integrated cell lines. pBWW33-67 (lane 3) and pBWW33-48 (lane 9) were chosen for following experiments. The sizes of selected DNA fragments are labeled on the right side: 6.6 Kb lines up with one fragment from HindIII digested lambda DNA, while 2.5 Kb lines up with the fragment of recipient.

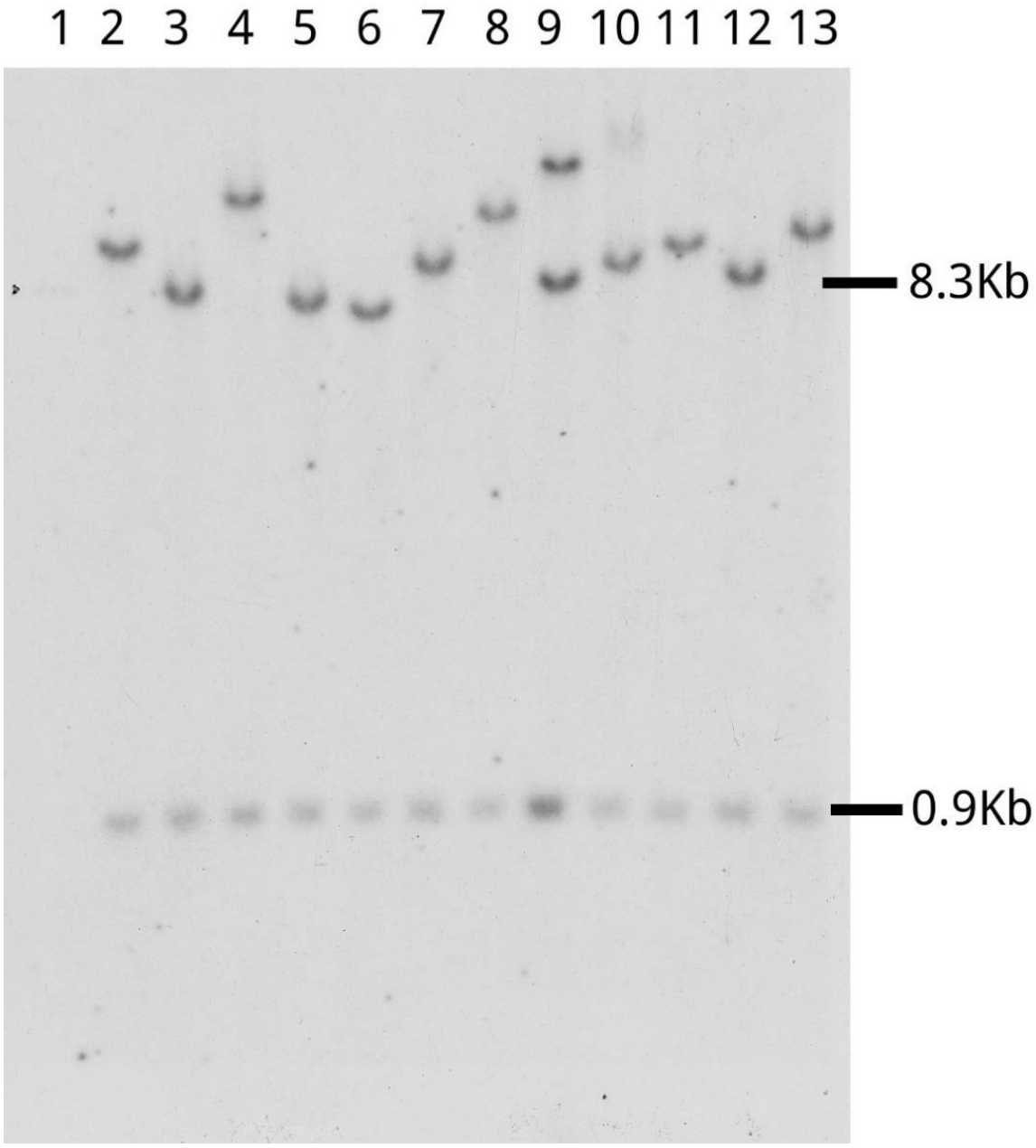

Figure 2.15 Identification of pBWW33 cell lines with correct plasmid integration through HindIII digestion.
The size and copy number of donor were examined as described above. Lane 1, HindIII digested plasmid pBWW33. Lane 2 to 13, HindIII digested genomic DNA from pBWW33 stably integrated cell lines. pBWW33-67 (lane 2) and pBWW33-48 (lane 4) were chosen for recombination experiments. The sizes of selected DNA fragments are labeled on the right side: 8.3 Kb and 0.9 Kb line up with these two fragments from HindIII digested pBWW33 plasmid. In lane 1, 8.3 Kb fragment is visible and pointed by a tiny black arrowhead, while the 0.9 Kb fragment is not visible with its location marked on the right side.

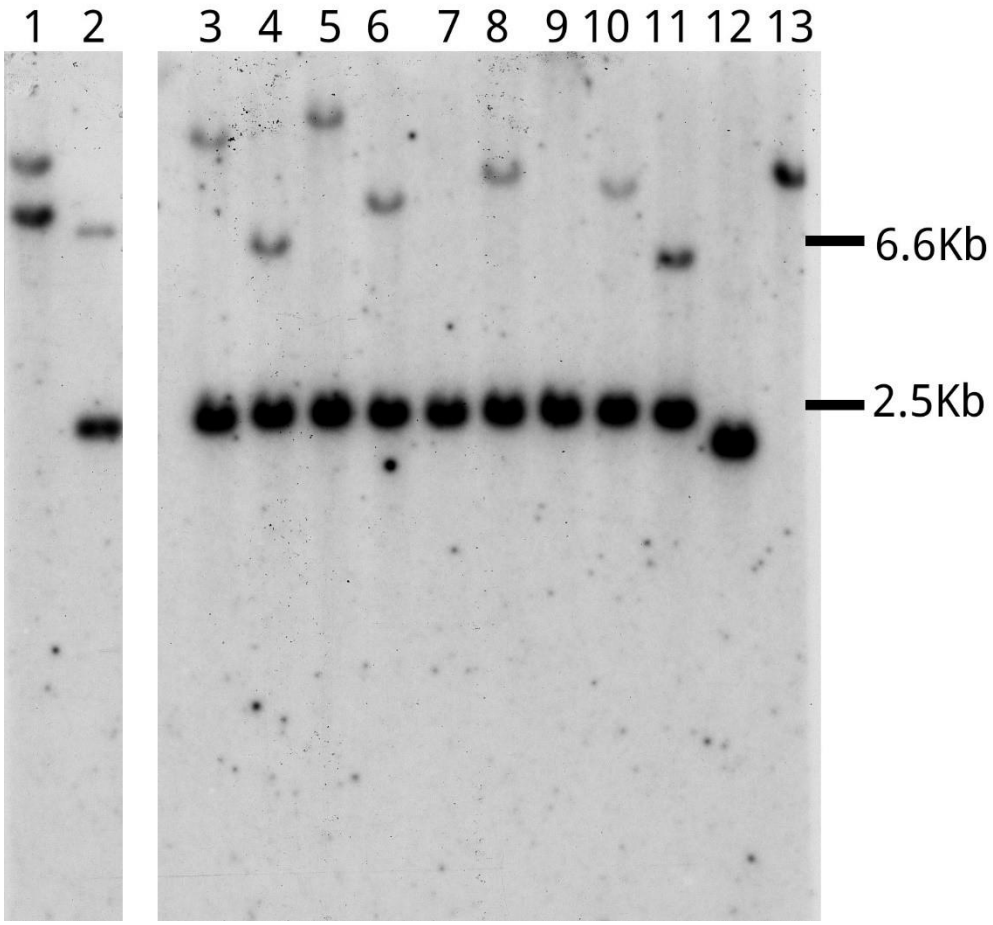

Figure 2.16 Identification of pBWW16 cell lines with correct plasmid integration through BamHI digestion.
The size and copy number of recipient were examined. Lane 2, BamHI digested plasmid pBWW16. Lane 1, 3 to 13, BamHI digested genomic DNA from pBWW16 stably integrated cell lines. pBWW16-2 (lane 3) and pBWW16-6 (lane 5) were chosen for recombination experiments. The sizes of selected DNA fragments are labeled on the right side: 6.6 Kb lines up with one fragment from HindIII digested lambda DNA, while 2.5 Kb lines up with the fragment of recipient.

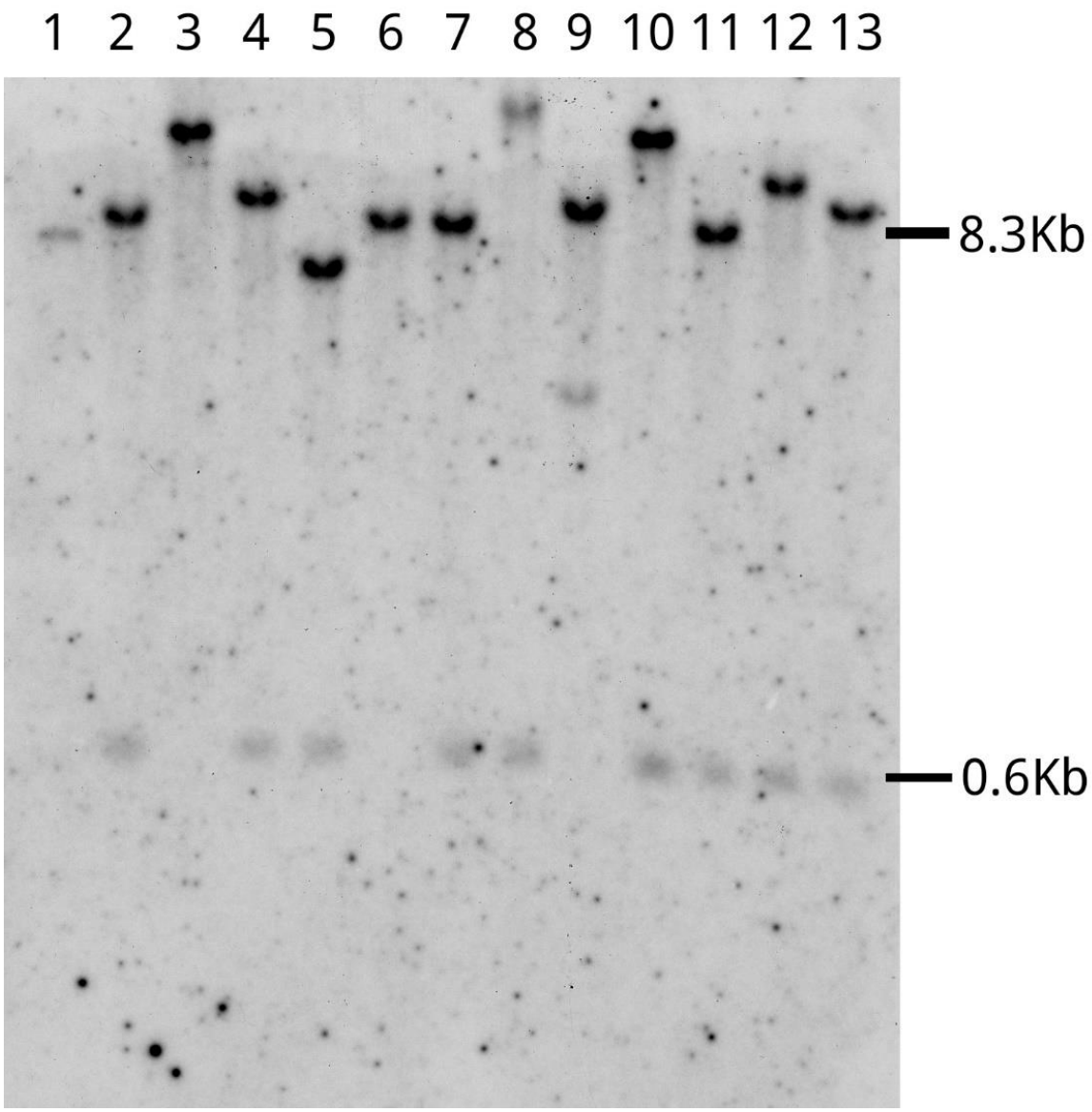

Figure 2.17 Identification of pBWW16 cell lines with correct plasmid integration through HindIII digestion.
The size and copy number of donor were examined for integration of intact substrate. Lane 1, HindIII digested plasmid pBWW16. Lane 2 to 13, HindIII digested genomic DNA from pBWW16 stably integrated cell lines. pBWW16-2 (lane 2) and pBWW16-6 (lane 4) were chosen for further experiments. The sizes of selected DNA fragments are labeled on the right side: 8.3 Kb and 0.6 Kb line up with these two fragments from HindIII digested pBWW16 plasmid

Table 2.1 Spontaneous recombination frequencies and HeR frequencies in cell line pBWW16-2 and pBWW16-6.

Cell line	Subclone	clone frequency (10^{-8})	# of colonies analyzed	# of Homeologous recombinants	Homeologous recombination frequency (10^{-8})
pBWW16-2	1	0	0	0	0
	2	0	0	0	0
	3	6.62	3	0	0
	4	3.64	2	0	0
	5	5.00	1	0	0
	6	0	0	0	0
	7	3.33	1	1	3.33
	8	0	0	0	0
	9	46.22	10	0	0
	10	18.07	6	4	12.05
	Total		23	5	
		Avg. = 8.29			Avg. = 1.54
pBWW16-6	1	0	0	0	0
	2	0	0	0	0
	3	26.67	6	4	17.78
	4	2.78	1	1	2.78
	5	6.67	2	2	6.67
	6	46.67	12	9	35.00
	7	0	0	0	0
	8	3.33	1	1	3.33
	9	0	0	0	0
	10	0	1*	0	0
	Total		23	17	
		Avg. = 8.61			Avg. = 6.56

Substrate pBWW16 has 16 mismatched nucleotides between donor and recipient.
The clone frequency was calculated for each subclone by dividing the number of HAT resistant clones by the number of cells subjected to HAT selection. Then, the clone frequencies of 10 subclones were averaged to generate the clone frequency of the cell line. Similarly, for each subclone, the HeR frequency was calculated by multiplying clone frequency of the subclone with the proportion of HeR clones in analyzed clones, and then they were averaged to generate HeR frequency of the cell line
* one HAT resistant clone of pBWW16-6 is not spontaneous recombinant as mentioned in following discussion, and excluded from calculation of recombination frequency.

Table 2.2 Spontaneous recombination frequencies and HeR frequencies in cell line pBWW25-13 and pBWW25-16.

Cell line	Subclone	clone frequency (10^{-8})	# of colonies analyzed	# of Homeologous recombinants	Homeologous recombination frequency (10-8)
pBWW25-13	1	3.75	1	0	0
	2	0	0	0	0
	3	0	0	0	0
	4	3.21	1	1	3.21
	5	3.14	1	0	0
	6	317	10	0	0
	7	24	7	7	24
	8	0	0	0	0
	9	0	0	0	0
	10	0	0	0	0
	Total	Avg. = 35.11	20	8	Avg. = 2.72
pBWW25-16	1	127.91	1	0	0
	2	62.07	3	1	20.69
	3	7.19	1	1	7.19
	4	11.49	2	0	0
	5	17.01	3	0	0
	6	117.12	1	1	117.12
	7	27.59	2*	0	0
	8	43.86	3	0	0
	9	43.35	4	1	10.84
	10	80.60	3	1	26.87
	Total	Avg. = 53.82	23	5	Avg. = 18.27

Substrate pBWW25 has 25 mismatched nucleotides between donor and recipient.
The clone frequency and HeR frequency were calculated as stated in Table 2.1.
* One HAT resistant clone of pBWW25-16 is not spontaneous recombinant as mentioned in following discussion, and excluded from calculation of recombination frequency.

Table 2.3 Spontaneous recombination frequencies and HeR frequencies in cell line pBWW33-48.

Cell line	# of Exp	Subclone	Clone frequency (10^{-8})	# of Colonies analyzed	# of Homeologous recombinants	Homeologous recombination frequency (10^{-8})
pBWW33-48	1	1	0	0	0	0
		2	8.00	0	0	0
		3	0	0	0	0
		4	0	0	0	0
		5	7.81	1	1	7.81
		6	7.52	1	1	7.52
		7	0	0	0	0
		8	0	0	0	0
		9	0	0	0	0
		10	0	0	0	0
		11	0	0	0	0
	2	1	0	0	0	0
		2	0	0	0	0
		3	0	0	0	0
		4	0	0	0	0
		5	0	0	0	0
		6	4.67	1	0	0
		7	7.69	0	0	0
		8	0	0	0	0
		9	31.25	4	4	31.25
		10	66.27	10	1	6.63
		Total		17	7	
			Avg. = 6.66			Avg. = 2.66

Substrate pBWW33 has 33 mismatched nucleotides between donor and recipient. The clone frequency and HeR frequency were calculated as stated in Table 2.1. Two fluctuation tests were carried out on pBWW33-48, and the clone frequency or HeR frequency were calculated by averaging that of every subclone.

Table 2.4 Spontaneous recombination frequencies and HeR frequencies in cell line pBWW33-67.

Cell line	# of Exp	Subclone	Clone frequency (10^{-8})	# of Colonies analyzed	# of Homeologous recombinants	Homeologous recombination frequency (10^{-8})
pBWW33-67	1	1	0	0	0	0
		2	0	0	0	0
		3	0	0	0	0
		4	0	0	0	0
		5	0	0	0	0
		6	3.66	0	0	3.66
		7	0	0	0	0
		8	0	0	0	0
		9	26.55	7	7	26.55
		10	0	0	0	0
	2	1	3.45	1	0	0
		2	0.00	0	0	0
		3	0.00	0	0	0
		4	9.62	3**	3	9.62
		5	0	0	0	0
		6	0	0	0	0
		7	4.12	5*	0	0
		8	7.25	1	0	0
		9	3.68	1	1	3.68
		10	3.40	1	1	3.40
		Total	Avg. = 3.09	19	12	Avg. = 2.35

Substrate pBWW33 has 33 mismatched nucleotides between donor and recipient. The clone frequency and HeR frequency were calculated as stated in Table 2.1. Two fluctuation tests were carried out on pBWW33-67, and the clone frequency or HeR frequency were calculated by averaging that of each subclone.

* In subclone 7 of pBWW33-67 Experiment 2, 4 out of 5 HAT^R clones are not spontaneous recombinants, so the recombination frequency is obtained through the multiplying raw clone frequency and proportion of spontaneous recombinants.

** In subclone 4 of pBWW33-67 Experiment 2, all three analyzed recombinants have original recipient as well as HeR recipient, which indicates one recombination event happened in cells with substrate duplication (confirmed later in Chapter 3). Thus, for this particular subclone, the clone frequency or homeologous recombination frequency occurring to a single copy substrate should be half of the raw clone frequency.

Table 2.5 Clone frequencies and spontaneous recombination frequencies in cell line pHR99-4.

Cell line	Subclones	Clone frequency (10^{-8})	# of Colonies analyzed	# of Homologous recombinants	Homologous recombination frequency (10^{-8})
pHR99-4	1	285.71	2*	0	0
	2	125.79	2*	1	62.89
	3	59.32	2	2	59.32
	4	196.36	0	0	0
	5	107.62	2*	1	53.81
	6	140.22	0	0	0
	7	157.89	1*	0	0
	8	204.78	0	0	0
	9	144.74	0	0	0
	10	401.22	1	1	401.22
	Total		10	5	
		Avg. = 182.37			Avg. = 57.72

Cell line pHR99-4 has homologous substrate pHR99, which does not have mismatched nucleotides between its donor and recipient.
Clone frequency was calculated for each subclone by dividing the number of HAT resistant clones by the number of cells subjected to HAT selection. Then, the clone frequencies of 10 subclones were averaged to generate the clone frequency of the cell line. Similarly, for each subclone, the HR frequency was calculated by multiplying clone frequency of the subclone with the proportion of HR clones in analyzed clones and they were averaged to generate HR frequency of the cell line.
* In cell line pHR99-4, approximately half of clones are non-recombinant, therefore, the spontaneous recombination frequency would be much different from the clone frequency and calculated separately case by case.

HeR recombination events and conversion tracts during spontaneous recombination: For each cell line, 10 to 23 clones from fluctuation tests were analyzed, and they were categorized based on their recipient sequences (Table 2.6). The vast majority of these collected clones removed the I-SceI insertion and restored the recipient to a continuous tk sequence, so they belong to spontaneous recombination events. If the recipient does not have mismatched nucleotides converted to HSV-2 TK, the recombination event is HR event, and listed in "recipient with no mismatch" in Table 2.6. A recombination event is recognized as HeR event only when its recipient does have mismatched nucleotides converted to HSV-2 TK sequence.

Among these HeR events, the restored recipients always acquired the mismatched nucleotides at position 1167, the closest mismatched nucleotide to the I-SceI site. Starting from there, the conversion tract extended continuously to the distal end regardless where it ended. These HeR events are thus further categorized based on whether the recipient acquired all the mismatched nucleotides from donor, and listed in "recipient with all mismatches" or "recipient with fewer than all mismatches" in Table 2.6. The proportion of HeR events was calculated as the percentile of HeR events within spontaneous recombination events. In examined cell lines, it is common to see the conversion of HSV-1 TK sequence to HSV-2 TK sequence, and proportion of the HeR events ranges from 22% to 75%. Cell line with the least mismatched nucleotides between donor and recipient does not have the highest proportion of HeR events. There were sporadic clones survived HAT selection via reactivation of mice thymidine kinase pseudogene rather than recombination within substrate (described later in Chapter 3), and these non-recombinants were listed in "other events" for cell lines where they arose.

Table 2.6 Recovered recombinant clones and their conversion tracts from spontaneous recombination.

Cell Lines	Analyzed clones	Recipient with all mismatches	Recipient with mismatches (fewer than all)	Recipient with no mismatch	Other events	Fraction of HeR
pBWW16-2	23	5	0	18	0	22%
pBWW16-6	23	17	0	5	1	77%
pBWW25-13	20	8	0	12	0	40%
pBWW25-16	23	4	1	17	1	23%
pBWW33-48	17	7	0	10	0	41%
pBWW33-67	19	9	0	3	7	75%
Total or Average	125	50	1	65	9	44%

If the conversion tract includes all mismatches and fewer mismatches, the event is categorized as HeR.

All recipients in HeR events acquired all mismatched nucleotides from donor except one obtained in cell line pBWW25-16: the recipient of clone pBWW25-16-4-27 only acquired the first 17 mismatched nucleotides out of total 25. The collected recombinants in current research were given a name, which is a combination of cell line, experiment and clone number, linked by dashes.

Clone frequency and HeR frequency in examined cell lines during DSB repair: Transfection was carried out on cell line pBWW16-2, pBWW16-6, pBWW25-13, pBWW25-16, pBWW33-48, pBWW33-67 and pHR99-4. pBWW16-2 and pBWW16-6 are cell lines of substrate pBWW16, which has 16 mismatched nucleotides between donor and recipient within the homeologous region. pBWW25-13 and pBWW25-16 are cell lines carrying 25 mismatched nucleotides, while pBWW33-48 and pBWW33-67 are cell lines carrying 33 mismatched nucleotides. To collect DSB-induced recombinants of these cell lines, pSce was transiently expressed in these cell lines. After I-SceI introduced a DSB into the recipient, the broken recipient sequence may be repaired and regain functional thymidine kinase. If so, colonies would form in HAT selection.

Table 2.7 lists clone frequencies and HeR frequencies of pBWW cell lines during DSB repair and Table 2.8 lists the clone frequency and HR frequency of pHR99-4 during DSB repair. The clone frequency was calculated for every cell lines as described in Materials and Methods, while the HeR frequency in pBWW cell lines is the product of clone frequency and the proportion of HeR events.

In DSB repair, the clone frequency of pBWW cell lines ranges from $1.01X10^{-5}$ to $8.53X10^{-5}$. Since almost all the HAT resistant clones in these cell lines are recombinants precisely removing the 30 bp insertion, the clone frequency is equivalent to the

Table 2.7 Clone frequencies and HeR frequencies of pBWW cell lines in DSB repair.

Cell line	Expt. Number	Clone frequency (10^{-5})	# of Colonies analyzed	Homeologous recombinants	Homeologous recombination frequency (10^{-6})
pBWW16-2	1	0.56	12	1	0.47
	2	0.63	9	1	0.70
	3	0.7	9	0	0.00
	4	1.91	5	2	7.64
	5	1.81	6	0	0.00
		Avg.=1.12			Avg.=1.76
pBWW16-6	1	2.57	14	5	9.18
	2	1.92	14	3	4.11
	3	1.3	16	4	3.25
		Avg.=1.93			Avg.=5.51
pBWW25-13	1	0.23	6	2	0.77
	2	0.25	6	3	1.25
	3	0.35	7	2	1.00
		Avg.=0.28			Avg.=1.01
pBWW25-16	1	6.49	16	1	4.06
	2	4.45	14	1	3.18
	3	4.97	15	1	3.31
		Avg.=5.30			Avg.=3.52
pBWW33-48	1	5.78	15	1	3.85
	2	8.15	15	4	21.73
	3	12.2	16	0	0.00
		Avg.=8.71			Avg.=8.53
pBWW33-67	1	0.92	16	4	2.30
	2	2.51	11	1	2.28
	3	2.35	12	5	9.79
		Avg.=1.93			Avg.=4.79

Electroporation of pSce was carried out at least 3 times on these pBWW cell lines. The table shows clone frequency, analyzed clones, HeR recombinants and HeR frequency of each experiment. For each experiment, the clone frequency was calculated by dividing the number of HAT resistant clones by the product of 5 million cells and plating efficiency. Then, all the clone frequencies were averaged to generate the clone frequency of the cell line. Similarly, for each experiment, the HeR frequency was calculated by multiplying clone frequency with the proportion of HeR clones in analyzed clones, and then they were averaged to generate HeR frequency of the cell line.

Table 2.8 Clone frequencies and HR frequencies of pHR99 during DSB repair.

Cell line	Expt. Number	Clone frequency (10^{-5})	Colonies analyzed	Homologous recombinants	Homologous recombination frequency (10^{-6})
pHR99	1	3.31	5*	3	19.86
	2	3.31	0	NA	NA
	3	3.43	5*	0	0.00
		Avg.=3.53			Avg.=9.93

Electroporation of pSce was repeated 3 times on pHR99-4 cell lines. The table shows clone frequency, analyzed clones, HR recombinants and recombination frequency of each experiment.
The clone frequency of pHR99 was calculated as in Table 2.7. Moreover, the HR frequency was calculated by multiplying clone frequency of the subclone with the proportion of HR clone, and then they were averaged for the cell line. Since there are non-recombinants collected during DSB repair, the HR frequency would be much different from the clone frequency.
* pHR99-4 has some HAT resistant clones carrying unchanged recipient, which indicate events other than DSB repair (further investigated in Chapter 3). The recombination frequency of pHR99-4 was corrected to the product of raw clone frequency and the proportion of DSB repair recombinants.
NA, identity of colonies in experiment 2 are not clear since they were lost during cell culture. Therefore, the homologous recombination frequency is not available.

recombination frequencies and used synonymous with recombination frequency. As shown in Table 2.7, longer homeologous sequence between donor and recipient does not reduce the clone frequency of the host cell lines. For each cell line, clone frequency of DSB repair is 9 times or more than that of spontaneous recombination (see Table 2.1 to 2.5), which agrees well with the fact that DSB boosts the DNA recombination between tandem repeats. Therefore, the vast majority of the collected recombinants are product of DSB repair rather than spontaneous recombination. The clone frequencies of pBWW cell lines with hybrid donors are equivalent or higher than that of pHR99-4 with pure homologous donor, however, their spontaneous recombination frequencies could be ten times lower than that of pHR99-4.

Categorization of DSB repair recombinants: Recovered clones in DSB repair experiments were categorized based on their recipient sequences. These recombination events, which restored I-SceI disrupted region of recipient to original HSV-1 TK sequence, were further divided according to the number of converted mismatched nucleotides: recipient with all mismatches, with mismatches (fewer than all) and with no mismatch. Those events that failed in restoring a continuous TK sequence were listed as other events, and they are most likely NHEJ events. In Table 2.9, these clones in column "recipient with all mismatches" and column "recipient with mismatches (fewer than all)" are HeR recombinants.

Conversion tract in homeologous recombinants: In HeR events, all or part of the mismatched nucleotides between homeologous sequences were transferred from donor to recipient, and Table 2.10 lists those mismatched nucleotides that were transferred in each HeR clone.

Table 2.9 Recovered recombination clones and conversion tracts from DSB repair.

Cell Lines	Analyzed clones	Recipient with all mismatches	Recipient with mismatches (fewer than all)	Recipient with no mismatch	Other events	Fraction of HeR
pBWW16-2	40	4	0	34	2	10%
pBWW16-6	44	11	1	32	0	27%
pBWW25-13	19	6	1	11	1	37%
pBWW25-16	42	0	3	36	3	7%
pBWW33-48	47	4	1	42	0	11%
pBWW33-67	42	5	5	31	1	24%
Total or Average	234	30	11	186	7	19%

Table 2.10 Conversion tracts in HeR recombinants from DSB repair.

Cell line	Total Mismatched Nucleotides	Converted Mismatched Nucleotides	Clone
pBWW16-2	16	M1 to M16	pBWW16-2-1-7, pBWW16-2-2-2, pBWW16-2-5-3, pBWW16-2-5-9
pBWW16-6	16	M1 to M16	pBWW16-6-1-5, pBWW16-6-1-7, pBWW16-6-1-8, pBWW16-6-1-14, pBWW16-6-1-16, pBWW16-6-2-1, pBWW16-6-2-12, pBWW16-6-2-15, pBWW16-6-3-1, pBWW16-6-3-11, pBWW16-6-3-13
		M1 to M12	pBWW16-6-3-16
pBWW25-13	25	M1 to M14	pBWW25-16-1-B2
		M1 to M25	pBWW25-16-1-C1, pBWW25-16-2-A1, pBWW25-16-2-C2, pBWW25-16-2-C3, pBWW25-16-3-B1, pBWW25-16-3-C2
pBWW25-16	25	M3 to M14	pBWW25-16-1-11
		M1 to M7	pBWW25-16-2-1
		M1 to M2	pBWW25-16-3-15
pBWW33-48	33	M1 to M33	pBWW33-48-2-16, pBWW33-48-1-5, pBWW33-48-2-2, pBWW33-48-2-6
		M1 to M2	pBWW33-48-2-14
pBWW33-67	33	M1 to M33	pBWW33-67-1-1, pBWW33-67-2-10, pBWW33-67-1-18, pBWW33-67-3-3, pBWW33-67-3-12
		M1 to M5	pBWW33-67-1-5
		M1 to M16	pBWW33-67-1-19, pBWW33-67-3-13
		M1 to M10	pBWW33-67-3-4
		M1 to M2	pBWW33-67-3-11

For convenience, these mismatched nucleotides were numbered in an order, moving away from the I-SceI site, as follows: the mismatched nucleotide closest to I-SceI recognition site is named as mismatched nucleotide 1 (M1), while the farthest one is named as mismatched nucleotide 33 (M33). The name of each recombinant in this table is a combination of cell line, experiment and clone number, linked by dashes.

These HeR recombinants restored the broken recipient back to continuous TK sequence, and also had HSV-1 TK sequence converted to HSV-2 TK sequence. In substrate pBWW16, pBWW25 and pBWW33, the homeologous region between donor and recipient has 16, 25 and 33 mismatched nucleotides respectively, and the details of these mismatched nucleotides are drawn in Figure 2.5.

HeR recombinants of six pBWW cell lines are listed in Table 2.10, and their recipient had mismatched nucleotides changed from HSV-1 TK sequence to HSV-2 TK sequences. Among recovered HeR events, the conversion tract almost always starts from the M1, and extends continuously toward M33. This rule applies to 40 HeR events except for pBWW25-16-1-11, which failed to convert the two mismatched nucleotides closest to the I-SceI site. HeR events in five pBWW cell lines have conversion tract containing all mismatched nucleotides as well as fewer mismatched nucleotides, while HeR events in cell line pBWW16-2 only have conversion tract including all mismatched nucleotides.

Comparison of spontaneous recombinants and DSB repair recombinants: For each examined cell line, 19 to 46 DSB induced recombinants were analyzed and gene conversion of mismatched nucleotides are common across these pBWW cell lines. The proportion of the HeR events in these cell lines ranges from 7% to 37%, however, cell lines with the least mismatched nucleotides in substrate do not show the highest proportion of HeR events. Comparing spontaneous recombination and DSB induced recombination, there are two major trends regarding the proportion of HeR events. First, in the same cell line, the proportion of HeR recombinants in spontaneous recombination is always higher than that in DSB induced recombination. Take cell line pBWW33-67 for example: the proportion of HeR recombinants in spontaneous recombination is 80%,

which is much more than the 24% in DSB induced recombination. Second, a cell line with higher HeR proportion in spontaneous recombination tends to have higher HeR proportion in DSB induced recombination, too. For example, the proportion of HeR recombinants in pBWW33-67 is 80% in spontaneous recombination and 24% in DSB induced recombination, and they are higher than those 41% and 11% HeR events in pBWW33-48.

In DSB induced recombination, the restored recipient almost always acquires the mismatched nucleotide closest to I-SceI site (M1), and from there, the conversion tract extends continuously to the distal end no matter where it ends. Totally 40 recovered HeR events have continuous conversion tract beginning at M1, while the conversion tract in one event started after M1. The HeR events are categorized into "recipient with all mismatches" if all the mismatched nucleotides were transferred to recipient, or "recipient with mismatches fewer than all" if not. As summarized earlier in the chapter, 53 out of 54 spontaneous recombinants have "all mismatches" transferred to recipient, while in DSB induced recombination, 30 out of 41 recombinants have "all mismatches" transferred into recipient. After reviewing the data, apparently more HeR events with "fewer than all" mismatches appeared in DSB induced recombination compared with spontaneous recombination. The Fisher's exact test comparing HeR clones in these two sub-categories between spontaneous recombination and DSBR generates a p-value of $3X10^{-4}$, which shows a significant difference in HeR subtypes between these two events.

Duplication of substrate in cell line pBWW33-67: Sequencing results of 9 recombinants in cell line pBWW33-67 were confusing because almost all nucleotides after position 1215 have two readings: double peaks suddenly show up in the sequencing

chromatogram around the I-SceI site. A closer look of the chromatogram suggested these recombinants appears to have one original recipient and one altered recipient by DSB repair. Thus, these recombinants may arise from cells with duplication of the integrated substrate, and DSB repair only happened to one of the two copies. To learn the origin of the duplication, subcloning of cell line pBWW33-67 was carried out to obtain discrete subclones that only contain descendants from a single cell. All subclones are named by their parental cell line, initial of "subclone" and clone number. For example, subclone number 1 is called pBWW33-67S1. Copy number of substrate in these subclones was checked by Southern blot.

Southern blot was carried out to check substrate's copy number in 24 subclones of pBWW33-67. Figure 2.18 examined the substrate's copy number in the first 12 subclones of pBWW33-67. As exemplified in this figure, pBWW33-67S11 shows darker bands of donor and recipient, which indicates duplication of the substrate in this subclone. Among the remaining 12 subclones, another subclone, pBWW33-67S20, shows the same duplication of the substrate (data not shown here). To learn more about the duplication, HindIII digestion was used to cut the integrated plasmid into a 0.87 Kb donor fragment and a junction fragment composed of plasmid sequence, 2.5Kb recipient and nearby genomic sequence at integration site. Figure 2.19 demonstrates that after HindIII digestion, the bands in normal subclones (lane 2 and lane 4) are the same as those in subclones with duplicated substrate. Since junction fragments of these subclones are the same as other cell lines, two copies of substrate in pBWW33-67S11 and pBWW33-67S20 are most likely due to a DNA duplication event involving the integrated substrate, rather than two separate substrate integration events. The duplication of substrate exists

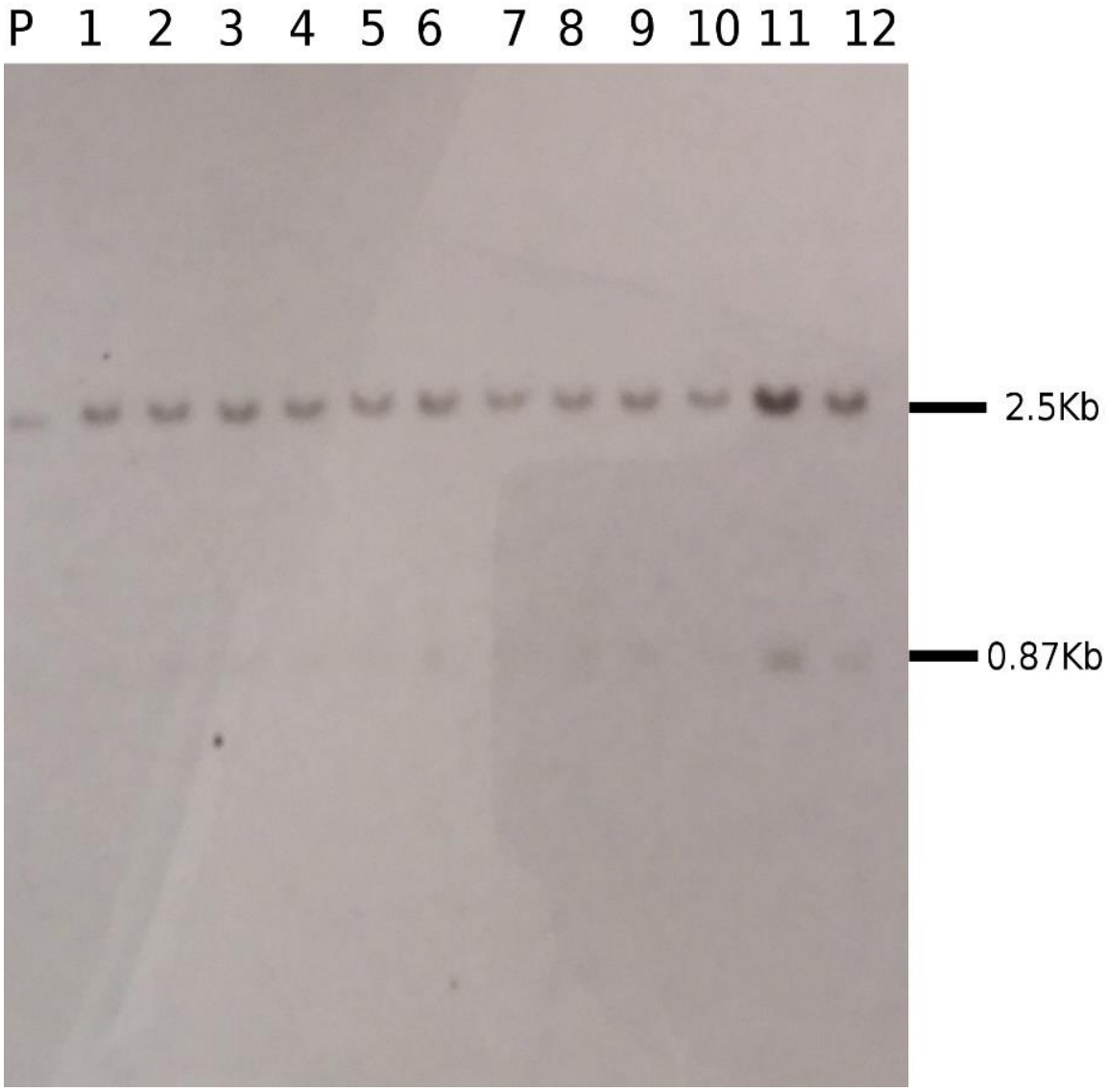

Figure 2.18 Duplication of the substrate happened in some subclones from cell line pBWW33-67.
Lane P, HindIII and BamHI digested plasmid pBWW33; Lanes 1 to 12, HindIII and BamHI digested genomic DNA from cell line pBWW33-67S1 to pBWW33-67S12.

Figure 2.19 Junction fragments in cell line pBWW33-67S11 and pBWW33-67S20 remain the same after duplication of the substrate.
Lane 1, HindIII digested pBWW33-67S11; Lane 2, HindIII digested pBWW33-67S12; Lane 3, HindIII digested pBWW33-67S20; Lane 4, HindIII digested pBWW33-67S21.

in a small portion of the cell population, independent of electroporation, so it is probably a product of a duplication event several generations after the integration of substrate. Other examined cell lines did not show the duplication of substrate.

In the current analysis, these 9 unexpected recombinants from pBWW33-67 were not listed in tables because they are a relatively small portion of the DSB induced recombinants, and the duplication confounds analysis.

Discussion:

Type of spontaneous recombinants: In current work, the majority of spontaneous recombination events are categorized into two groups: HR event and HeR event. In addition to recombination events, other type of events did happen in fluctuation tests: recipient in clone pBWW16-6-4-15 shows no sequence change, and the same thing happened in cell line pBWW33-67. These events are most likely reactivation of mouse thymidine kinase, and they were further investigated in Chapter 3.

Type of DSB-induced recombinants: DNA double-strand breaks in mammalian cells are usually repaired through non-homologous end joining (NHEJ), homologous recombination (HR) and single strand annealing (SSA) (30, 34, 78). Since the donor has truncation at both 5' and 3' end, the SSA or crossover events between donor and recipient cannot produce functional thymidine kinase. In the current experiments, recombinants only came from gene conversion events between donor and recipient, as well as minor NHEJ events (53, 55).

All collected recombinants are divided into 3 groups for clarity. HR and HeR events are separated from NHEJ events based on whether repair of DSB restored the broken DNA sequence back to a continuous tk sequence. HR events only restore the I-

SceI disrupted region of recipient using HSV-1 TK sequence from the donor, therefore the restored recipient loses the 30bp oligonucleotide insertion previously engineered into its HSV-1 TK sequence. If the restored recipient sequence has mismatched nucleotides converted to HSV-2 TK sequence besides losing the 30 bp insertion, the event is recognized as HeR. In comparison, recipient sequences in NHEJ events lose base pairs at DSB site, however, it is not precise excision of the 30 bp oligonucleotide insertion that happened in HR events. Though NHEJ events may have chance to remove the 30 bp insertion precisely like HR events, the probability and impact on current analysis are minor: previous experiments show the pure NHEJ events happened at dramatically low clone frequency, without providing a homologous donor for DSB repair.

High frequencies of HeR events were observed in pBWW substrates during spontaneous recombination: Previously, we demonstrated that HR barely happens between highly diverged sequences. Spontaneous recombination between HSV-1 TK recipient and HSV-2 TK donor sequences was never recovered from previous experiments using pHOME (55, 73). Use of a HSV-1 TK and HSV-2 TK hybrid sequence as donor allowed recombination to happen between a tandem repeat, however, only 1 of 81 recombinants from pHYB21A and pHYB12-8 had conversion of HSV-1 TK recipient to HSV2 TK sequence, far less than NHEJ or homologous recombination events that occurred within HSV-1 TK sequence (53). Using a hybrid donor composed of HSV-1 TK, HSV2 TK and HSV-1 TK sequences, the pHYB121 has recombination frequency equivalent to previous substrates. However, in 11 out of 39 recombinants converted the HSV-1 TK sequence of recipient to HSV-2 TK sequence, and the mismatched nucleotides were transferred all at once (53). These events clearly show HR restored the HSV-1 TK

recipient using the hybrid donor, and incorporated donor's HSV-2 TK sequence into recipient frequently if HSV1-TK sequences are on both sides of HSV2-TK sequence. Recombination events from these previous substrates allow us to postulate that as long as perfect homology is available on both sides, the homeologous region could be processed by HR, and leave conversion tract on the recipient sequence.

In current work, the potential surveillance against exchange between homeologous sequences was scrutinized extensively. Three new substrates have hybrid donor composed of HSV-1 TK, HSV-2 TK and HSV-1 TK sequences, and their donors differ in the length of HSV-2 TK sequences. Ranging from 60 bp to 199 bp, these HSV-2 TK sequences have 16, 25 or 33 mismatched nucleotides compared with HSV-1 TK sequence. Two cell lines of each substrate were examined in fluctuation test (summarized in Table 2.11): recombination frequencies of these substrates are around 10^{-8} and 10^{-7}, and they have HeR frequencies at 10^{-8} except pBWW25-16 is at $1.83X\ 10^{-7}$. Compared with the HeR events between HSV-1 TK recipient and hybrid donor composed of HSV-1 TK and HSV-2 TK sequence (pHYB21A or pHYB12-8), HeR events happened within these new substrates are substantial, and their clone frequency and proportion of HeR events are similar to previous substrate pHYB121. The results confirmed that homology on both sides of the homeology ensures the removal of the I-SceI insertion through HR as well as facilitates HeR events.

Recombination and HeR events are not affected by length of homeologous region in both types of recombination: Cell lines of substrate pBWW25 have average recombination frequency at $4.38X10^{-7}$, close to $5.72X10^{-7}$ from cell line of pHR99, at least 5 times more than other cell lines; while their average HeR frequency is

Table 2.11 Comparison of clone frequencies, HeR frequencies and fractions between spontaneous recombination and DSB-induced recombination.

Cell line	Spontaneous Recombination			DSB-induced Recombination		
	Recombination frequency (10^{-8})	HeR frequency (10^{-8})	HeR fraction	Recombination frequency (10^{-5})	HeR frequency (10^{-6})	HeR Fraction
pBWW33-48	6.66	2.66	41%	8.71	8.53	11%
pBWW33-67	3.09	2.35	80%	1.93	4.79	24%
pBWW25-13	35.11	2.72	40%	0.28	1.01	37%
pBWW25-16	52.44	18.3	23%	5.3	3.52	7%
pBWW16-2	8.29	1.54	22%	1.12	1.76	10%
pBWW16-6	8.61	6.56	77%	1.93	5.51	27%
pHR99-4	57.72	NA	NA	0.99	NA	NA

Cell line pHR99-4 has homologous donor and recipient, therefore, HeR frequency and HeR fraction are not applicable to this cell line. Moreover, recombination frequencies of pHR99-4 are dramatically different from its clone frequency since many HAT resistant clones from it are not recombinants.

$1.05X10^{-7}$, at least 2 times more than that of other substrates. Cell lines of pBWW16 only have slightly higher average clone frequency and HeR frequency than those in cell lines of pBWW33. Spontaneous recombination frequencies of these 6 cell lines suggest homeologous region does not necessarily bring down the clone frequency and HeR frequency. Apparently, the scrutiny and abortion against the exchange between homeologous sequences in examined cell lines are minor, compared with tens of folds fastidious rejection against evenly distributed homeology (79). If the rejection is effective within current substrates during spontaneous recombination, longer homeologous region would incur stronger rejection and should be harder to penetrate than a shorter one.

To sum up, the data of spontaneous recombination does not support substantial rejection against exchange between homeologous sequences, since no correlation was found between HeR clone frequencies and the length of homeologous region neighboring the I-SceI site. In contrast, mismatched hDNA in initiation stage faces vigorous scrutiny and rejection shown by both *in vitro* and *in vivo* assays (56, 79).

In examined substrates, the homeology does not impede the spontaneous recombination process upon successful initiation on either side. Spontaneous recombination events in substrate with even longer homeologous region could be explored in future, however, for examined substrates, no measurable impediment was observed in genetic exchange between homeologous region at least up to 200 bp.

During DSB repair, the clone frequency and HeR frequency of the two cell lines with the same substrate are not dramatically different from that of other cell lines. One-way ANOVA comparing HeR clone frequencies in 3 substrates during DSB repair give a p-value equals 0.2954, while the same test comparing clone frequencies in 3 substrates

give a p-value equals 0.0624. The performed ANOVA tests failed to find significant difference in frequencies between 3 examined substrates. Interestingly, the two cell lines carrying the 33 mismatched nucleotides have the highest average clone frequency and HeR frequency. In summary, during DSB repair, longer mismatched sequence does not decrease HeR frequency or fraction of HeR events across examined cell lines. Data in the current chapter failed to show fastidious rejection against genetic exchange between homeologous sequences during DSB repair.

Higher spontaneous clone frequency and HeR frequency in pBWW25 cell lines are likely due to its longer homology than the other two substrates. The donor of pBWW25 has the same 300 bp downstream HSV-1 TK sequence as the other two substrates, while it has 400 bp upstream HSV-1 TK sequence, about 100 bp longer than the other two substrates. As demonstrated by Huang and other researchers (26, 71), efficient recombination depends on availability of continuous uninterrupted homology, and the recombination frequency is in linear relation with the length of homology once the homology exceeds the MEPS. In these pBWW substrates, the homeologous sequence inside donor reduced the continuous homology for HR, thus they have lower clone frequency compared with substrate pHR99. The substantially longer homology of substrate pBWW25 likely brought clone frequency higher than other pBWW substrates, though the position of substrate integration may contribute to these higher frequencies as well.

HeR events are common in both spontaneous recombination and DSB repair: In Table 2.11, the proportion of spontaneous HeR recombinants ranges from 22% to 80%, showing the mismatched nucleotides are easily transferred from donor to recipient.

Similar to the observation in spontaneous recombination, analysis of DSB repair recombinants supports that DSB provoked recombination frequently proceeds from the homologous region into a homeologous region. When a DSB was introduced into HSV-1 TK recipient, 41 out of 227 recombination events have homeologous conversion tracts, indicating frequent use of homeologous sequence during DSB repair. Among the examined cell lines, 7% to 37% HeR events were found after DSB repair, so the genetic exchange between homeologous sequences is a common phenomenon during DSB repair.

HeR events are not affected by length of homeologous region in both types of recombination: In spontaneous recombination, recombination between sequences with 33 mismatched nucleotides has even higher proportion of HeR events, which contradicts with the idea that cluster of mismatches triggers rejection and destruction. Otherwise, longer homeologous sequence in pBWW33 would incur more rejection, and cause lower proportion of HeR events compared with shorter ones.

Similar to the finding in spontaneous recombination, there is no obvious rejection against homeologous exchange during DSB repair. For example, cell lines pBWW16-2 and pBWW16-6 have a substrate harboring the shortest homeologous region, but their fractions of HeR events are not the highest among examined cell lines.

Trends in proportion of HeR events: Two trends in the proportion of HeR events were described in results. First, the proportion of HeR events during DSB repair has positive correlation with that in spontaneous recombination. A cell line with more HeR events in spontaneous recombination is likely to have more HeR events during DSB repair. So, the proportions of HeR events are probably determined by a combination of multiple factors, including origin of recombination, genetic background, substrate's

integrated site and other unknown factors. Second, within the same cell line, the proportion of HeR events in spontaneous recombination is always higher than that in DSB repair. With the same genetic background, higher proportion of HeR in spontaneous recombination does suggest different underlying mechanism from DSB induced recombination. Examination of the substrates' structure reveals that DSB provoked recombination is most likely initiated in downstream HSV-1 TK sequence of donor, while spontaneous recombination has another continuous homology, the upstream HSV-1 TK sequence of donor, to start recombination events efficiently. The recombination events initiated in upstream HSV-1 TK sequence and resolved in downstream HSV-1 TK sequence only produce HeR recombinants, therefore, with these HeR events added to the pool of HR and HeR events initiated in downstream HSV-1 TK sequence, the fraction of HeR events in spontaneous recombination would surely be higher than in DSB repair.

Different homology preference of DSB repair: As postulated before, in spontaneous recombination events, conversion tract rarely ends in homeologous region, converting some of the mismatched nucleotides: out of 53 clones, only 1 clone has conversion tract terminated in middle of homeologous region (Table 2.12). In the study of DSB repair events, out of 41 identified HeR clones, 11 clones have conversion tract ended in homeology, only converting part of these mismatched nucleotides between donor and recipient. Comparing the location where conversion ends in two situations, the Fisher's exact test generates a p-value of $3X10^{-4}$, showing a dramatic difference between spontaneous recombination and DSB induced recombination. Current data clearly demonstrates that spontaneous recombination tends to resolve in the homologous region,

Table 2.12 Transfer of mismatched nucleotides between donor and recipient differs in spontaneous recombination and DSB-induced recombination.

		Spontaneous Recombination		DSB-induced Recombination	
Number of Mismatches	Cell line	Fewer than all mismatches	All mismatches	Fewer than all mismatches	All mismatches
33	pBWW33-48	0	7	1	4
	pBWW33-67	0	12	5	5
25	pBWW25-13	0	8	1	6
	pBWW25-16	1	4	3	0
16	pBWW16-2	0	5	0	4
	pBWW16-6	0	17	1	11
Total		1	53	11	30

while DSB provoked recombination frequently resolves in middle of homeologous region.

In a fluctuation test, a recombination event may generate many clones with the same conversion tract, so to get an estimation of discrete events, these identical clones arising from the same subclone are counted as one event. And the discrete events counted in this way are named as minimal spontaneous recombination events, a conservative estimation of real recombination events. If one counts the minimal spontaneous recombination events, the number of these spontaneous events converting all mismatched nucleotides is 21 while the number of spontaneous events converting part of mismatched nucleotides is only 1. If minimal spontaneous recombination events are used instead of actual recombinants number, a Fisher's exact test comparing HeR events changing partial or all mismatched nucleotides between spontaneous recombination and DSB repair generates p-value of 0.0433. The new test is more stringent than previous one using recombinants number. However, it still supports a significant difference between spontaneous recombination and DSB induced recombination.

Probable mechanistic differences behind the two types of recombination and their homology preference: Variances in mechanisms may cause the difference between spontaneous recombination and DSB provoked HR in choosing the resolution site. DSBR and SDSA models (Figure 1.1) provide satisfactory explanation of how cells repair DSB in mitotic cells (7). SDSA is considered as a major and preferable pathway to generate gene conversion events during mitosis (19). Different from DSBR, after strand invasion and DNA synthesis, the newly synthesized strand anneals back to the other resected broken end for ligation, which abrogates the second strand capture and double strand HJs

steps. SDSA tends to generate recombinants with unidirectional conversion tracts although occasionally it does allow two strand invasion and bidirectional conversion tracts (69). Moreover, without forming HJs, SDSA won not mediate crossover events between donor and recipient, which are only achieved by HJs resolution in DSBR model (7). Crossovers are not collected or discussed in the current experiments since crossover within pBWW substrates only produces recombinants carrying a truncated tk gene, and thus is unable to survive the HAT selection.

Spontaneous recombination may be due to repair of spontaneous occurring DSBs, through either SDSA or DSBR pathway, as described above. Spontaneous recombination may also start to restore the collapsed or stalled replication forks when the replication machinery meet DNA lesions on parental strands (Figure 1.2). Moreover, template switching or replication slippage may explain these obtained recombinants as well.

What remains to be explored is the underlying mechanism of relaxed homology requirement during DSB repair. Probably, DSB repair experiments collected recombinants repairing the homeologous sequence near a DSB, while the genetic exchange in collected spontaneous recombinants likely started from an unknown site some distance away. Or, the difference lies in how HR resolves the exchange between participating molecules: HJs need to be resolved during DNA recombination restoring replication forks while as a predominant pathway in DSB repair, SDSA does not employ HJ intermediate. More discussion and efforts to investigate the difference in homology requirement between spontaneous recombination and DSB-induced recombination are continued in Chapter 4.

CHAPTER 3 DNA MISMATCH REPAIR'S ROLE IN SPONTANEOUS AND DSB-INDUCED RECOMBINATION

DNA mismatch repair (MMR) is a critical pathway to maintain the genome's fidelity during cell proliferation (58). When cells produce descendants through continuous divisions, their genomes need to be replicated timely and correctly. The major DNA polymerase employed in DNA replication is faithful due to a proofreading mechanism, and the errors occur at 1 in every 100,000,000 nucleotides (80). Though the risk is low, it cannot be neglected considering size of the genome and repetitive cell divisions. Moreover, when cells are threatened by endogenous or exogenous stresses, normal replication could be stalled by altered DNA structures. In these situations, other error-prone translesion polymerases may replace the high-fidelity polymerase, and carry on DNA synthesis with a dramatically high error rate. Sometimes, the translesion polymerases may even synthesize random DNA sequence to overcome temporary lesions (81, 82). If not fixed properly, these mutations which arise from DNA replication will accumulate and cause gradual deterioration of the genome. MMR is the major strategy cells adopt to combat replicative errors. MMR deficiency in bacteria, as demonstrated by many mutants, will cause high mutation rates in these strains, and deficiency of MMR elevates spontaneous mutations in human cells as well (83). MMR deficiency associates with vast majority of the human cancers and actually leads to cancer directly as shown in certain hereditary nonpolyposis colorectal cancer (HNPCC) (60).

MMR recognizes the DNA mismatches introduced by replicative errors, makes incision on one of the mismatched DNA strands, degrades and re-synthesizes the excised strand under the guidance of the complementary strand (84, 85). During MMR, the newly synthesized strand is always chosen for degradation, therefore, cells restore the mismatched DNA sequence back to the original parental sequence (86). In gram-negative bacteria, transient hemimethylation allows MMR to recognize and destroy newly synthesized strand and keep the methylated strand as template. In other bacteria and eukaryotes, further investigation is needed to reveal how MMR targets the newly synthesized DNA strand for degradation. Some reports suggest that temporary nicks occurring on newly synthesized strands will signal the MMR for destruction while others show the asymmetric orientation of replicative machinery may assist MMR to find the newly synthesized target strand.

In bacteria, MMR has three major components working cooperatively to recognize and repair mismatched nucleotides. MutS family members perform their function as a dimer: they detect the distortion in DNA double helix caused by mismatched nucleotides, and signal for downstream effectors (57). MutLs and MutHs will form complex with the MutS dimer and recruit endonuclease, DNA helicase, and exonuclease for strand degradation. In eukaryotes, there are two types of MutS complex: Msh2/Msh6 detects base substitution and small loops while Msh2/Msh3 screens for small loops or larger loops.

MMR has been proposed to play more than one roles in HR (54). MMR may target and repair nucleotide mismatches in heteroduplex DNA (hDNA) left by HR events. At several steps of HR, the hDNA intermediates are formed through annealing DNA

strands from two homologous sequences. After HR resolved these two DNA sequences, a stretch of foreign strand might be left in one DNA molecule as hDNA. These mismatched nucleotides within hDNA would be the obvious target of MMR. Repair of the hDNA by MMR either restores the original DNA sequence or converts it to the template sequence. Second, MMR scrutinizes the strand paring between HR substrates and prevents recombination between diverged sequences as discussed in Chapter 3. In yeast, without functional MMR, the frequency of HeR events between diverged sequences will raise 100 times to that between identical sequences (66). Although the mechanism and protein components involved in rejection are still in debate, the rejection occurs as early as strand invasion and D-loop formation (57, 87).

As shown in the previous chapter, in either spontaneous recombination or DSB induced recombination, it is common to see nucleotide mismatches transferred from hybrid donor sequence to HSV1-TK recipient sequences. However, the generation of these HeR events is not clear. First, the homeologous sequence may be lost during the earlier steps in breakage, end resection and degradation before the strand invasion. So the restored recipients with HSV2-TK sequence are solely products of DNA gap repair, which re-synthesized the lost DNA according to its homologous template. Second, as described in Chapter 1, the hDNA annealing homeologous sequences could form during initial strand invasion or HJs migration, so after HJ resolution, MMR may repair nucleotide mismatches either toward HSV2-TK sequence (HeR events) or HSV1-TK sequence (HR). Third, in SDSA pathway, resected HSV-1 TK DNA strand and newly synthesized HSV2-TK strand may anneal and ligate with each other with help from nearby homologous sequences, while MMR fixes the mismatched nucleotides afterward.

Abrogation of MMR was achieved in selected cell line pBWW33-67 by stably transfection of MSH2 shRNA. Through analysis of the recombinants in absence of MMR, it is possible to reveal the molecular intermediate of HeR events and help to piece together the process for previously observed HeR events. At the same time, comparing the HeR frequency before and after MSH2 knockdown, it is possible to see whether MMR prevents HeR events from happening within pBWW substrates. Moreover, it is also possible to see whether MMR affects the resolution site of DNA recombination.

Materials and Methods:

Stable knockdown of MSH2 in cell lines pBWW33-67: To make experimental and control cell lines for the MSH2 knockdown experiment, cell line pBWW33-67 was transfected with MSH2 shRNA#3 (target sequence is 5' CCG GCC CGG CAA TCT TTC TCA GTT TCT CGA GAA ACT GAG AAA GAT TGC CGG GTT TTT G 3' for mouse Msh2 gene) or Control shRNA#1 (non human or mouse shRNA, target sequence is 5' CCG GCA ACA AGA TGA AGA GCA CCA ACT CGA GTT GGT GCT CTT CAT CTT GTT GTT TTT 3'). MSH2 shRNA #3, TRCN0000042493, was obtained from the MISSION TRC shRNA bacterial glycerol stock library (Sigma), while Control shRNA #1, SHC002, is one of the MISSION shRNA Control Vectors from Sigma. pBWW33-67 cells were plated in 35 mm dishes, and one day later, 0.8 ug shRNA plasmid was prepared and transfected into cells using Qiagen Effectene as instructed by the supplier. Two days after the transfection, the cells were trypsinized and about 2 million cells were plated in one T75 flask for puromycin selection at 5 ug/mL. Two weeks later, puromycin resistant colonies from the shRNA transfections were picked, cultured and stored separately.

6-Thioguanine (6-TG) resistance test: To check 6-TG's effect on obtained cell lines, 100 cells were plated in each T25 flask. The flasks were refilled with media containing different concentrations of 6-TG on the second day and then changed back to regular media on the third day. After 14 days, viable colonies were stained with Methylene blue and counted as described before. The colonies number in 6-TG treated flask was divided by that in control flask to give survival rate for each cell line. To screen the obtained cell lines, 4 uM 6-TG was used in the first round of tests. Selected cell lines were subjected to a second round of tests to estimate of their resistance, and the 6-TG concentrations include 0.5 uM, 2 uM, 4 uM and 8 uM.

Preparation of protein sample: To exact soluble protein samples, cells were trypsinized and suspended in 5 mL culture media. The cell suspension in a 15 mL conical tube was centrifuged at 300 x g for 3 minutes, and then the pelleted cells were washed with 1.5 mL ice cold PBS and transferred into a 1.5 mL Eppendorf tube. After the second centrifugation at 1,500 x g for 3 minutes, the newly pelleted cells were mixed with 30 uL lysis buffer (RIPA buffer: 50 nM Tris-HCl, 150 nM NaCl, 1 mM EDTA, 0.1% SDS, 1% deoxycholate and 1% Triton x-100, PH 7.4, mixed 1/10 v/v proteinase inhibitor cocktail, Sigma Cat. P8340). The cell clumps were broken up through repetitive pipetting, and the cell lysis was incubated on ice for 30 minutes. All obtained protein samples were stored at -80˚C until use. Before western blotting, the thawed cell lysis was centrifuged at 10,000 x g for 10 minutes at 4˚C, and the supernatant was transferred into a new tube.

Protein electrophoresis and transferring: Protein samples were freshly prepared before electrophoresis: 30 ug protein sample was mixed with 6 X SDS sample buffer (300 mM Tris-HCl, 600 mM DTT, 12% SDS, 30% Glycerol and 0.06%

Bromophenol blue, pH 6.8) and additional lysis buffer to reach the same final volume across samples. The sample mixture was incubated at 95˚C for 5 minutes, and then briefly spun down.

All protein samples were separated on 8% polyacrylamide gels in 1 X SDS running buffer (25 mM Tris base, 192 mM Glycine and 0.1% SDS). For one gel, the current was set at 15 mA while samples moved in the 4% stacking gel and then switched to 19 mA after samples entered the 8% separating gel. The electrophoresis was stopped once the 37 KD protein marker approached the end of the separating gel.

The 1 x transfer buffer (25 mM Tris base, 192 mM Glycine, 0.1% SDS and 20% methanol) was prepared in advance and cooled at 4˚C before use. After the electrophoresis ended, the gel was disassembled from the cassette and the stacking gel was cut off from the separating gel. The separating gel was soaked in 1 x transfer buffer and shaken slowly for 15 minutes. During the time, the nitrocellulose membrane Amersham Hybond ECL from GE was wet with deionized water as instructed by the supplier, and then soaked in 1 x transfer buffer together with filter paper, and the foam filter pad.

The transfer "sandwich" was assembled in following order: black plate, filter pad, filter paper, gel, membrane, filter paper, filter pad and white plate. Bubbles between layers were gently driven out by a rolling test tube across the "sandwich". The transfer "sandwich" was put into the electrode assembly and then inserted into the tank with 1 x transfer buffer. The protein transferring was carried out at 100 volts for 90 minutes unless specified separately. During the transferring, the tank was incubated in an ice water bath, and an ice cooling pad was kept in the tank as well.

Western blot: After transferring, the membrane was washed briefly in 1 X Tris-buffered saline (TBS), and incubated in blocking buffer (5% non-fat milk and 0.1% Tween-20 in TBS) for 1 hour at room temperature. Primary antibody was diluted empirically in 10 mL blocking buffer, and the overnight incubation of the membrane was carried out in a suitable petri dish with slow agitation at 4˚C. On the second day, the membrane was rinsed briefly in wash buffer before three consecutive washing steps. The membrane was washed with slow agitation for 5 minutes each time in 100 mL wash buffer and the washing buffer is TBS with 0.1% Tween-20 (0.1% TBST) unless otherwise specified.

Secondary antibody was diluted empirically in 10 mL blocking buffer, and the incubation was set at room temperature for 1 hour with slow agitation. After incubation with the antibody, three washing steps were carried out as described before except the last wash took 10 minutes.

The detection was performed using Amersham's ECL Select as instructed by the supplier. Equal volume of solution A and solution B were mixed, and then the membranes were immersed in the mixture for 5 minutes in the dark. The membrane was drained of excessive liquid and wrapped in Saran wrap for chemiluminescent detection.

Western blot detecting MSH2 expression: expression of MSH2 in selected cell lines was examined through western blot using MSH2 antibody from NOVUS (Cat# NB 100-621). The western blot process was described above and the differences in loading and blotting were listed below. For electrophoresis, 30 ug Protein sample of selected cell lines was separated on 8% polyacrylamide gel, and then transferred onto Amersham Hybond ECL nitrocellulose at 400 mA for 90 minutes. The membrane was cut through

75KD to separate MSH2 and β-tubulin for subsequent immune-blotting. MSH2 primary antibody was applied at 1:15,000 in Tris buffered saline (TBS) with 3% non fat milk (NFM), and the secondary antibody (Goat polyclonal antibody to Rabbit IgG, HRP conjugated from NOVUS, Cat# NB 730-H) was added at 1: 100,000 in TBS with 3% NFM. For β-tubulin, the primary antibody (mouse monoclonal antibody from SANTA CRUZ, Cat# sc-5274) was applied at 1: 10,000 in 0.1% TBST with 5% NFM and the secondary antibody (goat anti-mouse IgG-HRP, Cat# sc-2005) was applied at 1: 10,000 in 0.1% TBST with 5% NFM as well. For both proteins, the primary antibody was incubated overnight at 4°C, and secondary antibody was incubated at room temperature for 45 minutes. The washing buffer for MSH2 was TBS with 0.05% Tween-20 (0.05% TBST), while 0.1% TBST was used for tubulin. Three washes were performed after each incubation for 5 minutes. Finally, the membranes were developed using Amersham ECL Select Western Blotting Detection Reagent (GE).

Collection and analysis of spontaneous recombination through fluctuation test: To collect spontaneous recombinants in pBWW33-67 derivative cell lines, 10 subclones were cultured from 100 cells up to 40 million cells before HAT selection, as described in Chapter 2.

Collection and analysis of spontaneous recombination through modified fluctuation test: A modified fluctuation test was attempted on pBWW33-67 to collect the recombination clones containing all descendants of a spontaneous recombination event. Each subclone was cultured from 100 cells up to 4 million cells in the modified test, and then these 4 million cells were separated into 4 T175 flasks. The cells in each T175 flask were grown for another 3 days before subjected to HAT selection without

counting and re-plating. Recombinant clones were collected and analyzed as described before.

Collect and analyze DSB repair recombinants: To collect DSB provoked recombinants of pBWW33-67 derivative cell lines, the electroporation and HAT selection were done as described in Chapter 3.

PCR amplification and sequencing: The PCR and sequencing procedures were described in Chapter 2. Moreover, nucleotide callings at every mismatched postilion (M1 to M33) were checked manually on the chromatograms when conversion occurred to recipient, in order to see whether nucleotides of donor and recipient both appear at each mismatched position.

Subcloning: To separate different cell populations within a DSB repair recombinant, subclones of existing cell lines were obtained as described in Chapter 2. The recipient sequences in newly established subclones were amplified, sequenced and compared with their parental recombinant.

Drug resistance tests on tk function: To check the function of tk in selected cell lines, the cells were tested with Trifluorothymidine (TFT) and Gancyclovir (GCV). 100,000 cells were plated into each T25 flask, and 4 hours later, these flasks were refed with media containing 5 ug/uL TFT or 10 uM/uL GCV. After 3 days, the cells in these flasks were examined under the microscope at day 3 and day 6.

RNA extraction: RNA from selected cell lines was extracted using Qiagen Rneasy Mini Kit as instructed by the supplier. About 3 million cells in a T25 flask were trypsinized, collected and lysed with supplied buffer. The cell lysate was homogenized by passing through a blunt 20-gauge needle at least 5 times.

RT-PCR and sequencing: qScripts XLT One-Step RT-PCR Kit was used with primer pair AW109 (5' TAA TAC GAC TCA CTA TAG GGA AGT AGC ACA GGC GGC ACA C 3', downstream primer matches the seventh exon of mouse TK sequence) and AW100 (5' GAT TTA GGT GAC ACT ATA GGC TGC CAT GAG CTA CAT CAA TC 3', upstream primer matches the first exon of mouse TK sequence) to amplify the mRNA transcripts of mouse thymidine kinase. The RT-PCR was carried out in a final volume of 25 uL with 500 ng sample RNA and 400 nM each primer. The RT-PCR begins with cDNA synthesis at 48ºC for 20 minutes, and then the initial denaturation at 94ºC takes 3 minutes. The subsequent 32 PCR cycles include 20 seconds denaturation at 94ºC, 1 minute annealing at 55ºC and 90 seconds elongation at 72ºC. The final elongation step is done at 72ºC for 10 minutes.

These RT-PCR products were separated on 0.8% agarose gel and the 612 bp target band in these samples was cut out for PCR amplification. The excised gel piece was frozen overnight, and then centrifuged at 16,100 x g for 3 minutes to pellet the agarose gel. The second PCR with AW109 and AW110 was carried out as described in Chapter 2 using 1 uL of the resulting supernatant.

PCR amplification of mouse thymidine kinase pseudogene: The mouse thymidine kinase pseudogene in selected cell lines was amplified using AW109 and AW100 as described in Chapter 2. Their sequences were aligned to the reference mouse tk sequence to check for any nucleotide change.

Results:

Establishing cell lines used in MSH2 knockdown experiments: Twelve stably transfected cell lines were obtained from control shRNA transfections performed on

pBWW33-67, while another 12 stably transfected cell lines were obtained from MSH2 shRNA transfections. These cell lines were named for their parent cell line, initial of experiment group and clone numbers. The cell lines with integration of control shRNA were named from pBWW33-67C1 to pBWW33-67C12, while the cell lines carrying MSH2 shRNA were named from pBWW33-67M2 to pBWW33-67M13.

Selecting cell lines for recombination assays: To initially screen for MMR-deficient cell lines and control cell lines, all established cell lines were subjected to survival test in 4uM 6-TG, whose genotoxicity is mediated by MMR (88). As listed in Table 3.1, half of the stable cell lines from MSH2 shRNA transfections have 21% or higher survival rate, showing substantial resistance to 6-TG. In contrast, almost all cell lines from control shRNA transfections were sensitive to 6-TG just like the parental cell line pBWW33-67. Only one clone pBWW33-67C2 has a 14% survival rate, but this is still lower than aforementioned 6-TG resistant clones from MSH2 shRNA transfection. pBWW33-67M13 and pBWW33-67M11 have the highest survival rates, so they were selected for western blot analysis together with control cell lines pBWW33-67C6 and pBWW33-67C8, which were sensitive to 6-TG just like the parental cell line. These four cell lines were subjected to a second round of 6-TG testing at different concentrations to better estimate their MMR functionality, and the results are listed in Table 3.2.

Western blot in Figure 3.1 shows expression of MSH2 in selected cell lines, and about 6 to 10-fold reductions were observed in MSH2 knockdown cell lines, compared with parental cell line or control cell lines. The MSH2 expression in pBWW33-67M5, pBWW33-67M11 and pBWW33-67M13 reduced dramatically while its expression in pBWW33-67C6 and pBWW33-67C8 did not. Interestingly, the MSH2 expression in

Table 3.1 Preliminary 6-TG resistance test on 22 stable cell lines carrying control shRNA or MSH2 shRNA .

Experimental groups	Cell Lines	# of colonies in 0uM 6-TG	# of colonies in 4uM 6-TG	Survival Rate
Parental	pBWW33-67	89	1	1%
Control shRNA	pBWW33-67C1	NA	NA	NA
	pBWW33-67C2	81	11	14%
	pBWW33-67C3	NA	NA	NA
	pBWW33-67C4	80	1	1%
	pBWW33-67C5	82	0	0%
	pBWW33-67C6	74	1	1%
	pBWW33-67C7	63	0	0%
	pBWW33-67C8	32	0	0%
	pBWW33-67C9	55	1	2%
	pBWW33-67C10	41	1	2%
	pBWW33-67C11	41	0	0%
	pBWW33-67C12	71	0	0%
MSH2 shRNA	pBWW33-67M2	63	0	0%
	pBWW33-67M3	67	0	0%
	pBWW33-67M4	55	2	4%
	pBWW33-67M5	52	30	58%
	pBWW33-67M6	61	0	0%
	pBWW33-67M7	103	22	21%
	pBWW33-67M8	49	1	2%
	pBWW33-67M9	65	5	8%
	pBWW33-67M10	70	28	40%
	pBWW33-67M11	75	96	128%
	pBWW33-67M12	66	31	47%
	pBWW33-67M13	75	61	81%

Number of colonies in 4 uM 6-TG was divided by number of colonies in 0 uM 6-TG to generate the survival rate for each cell line in last column.
NA: Two cell lines were lost during cell culture, not available for 6-TG test.

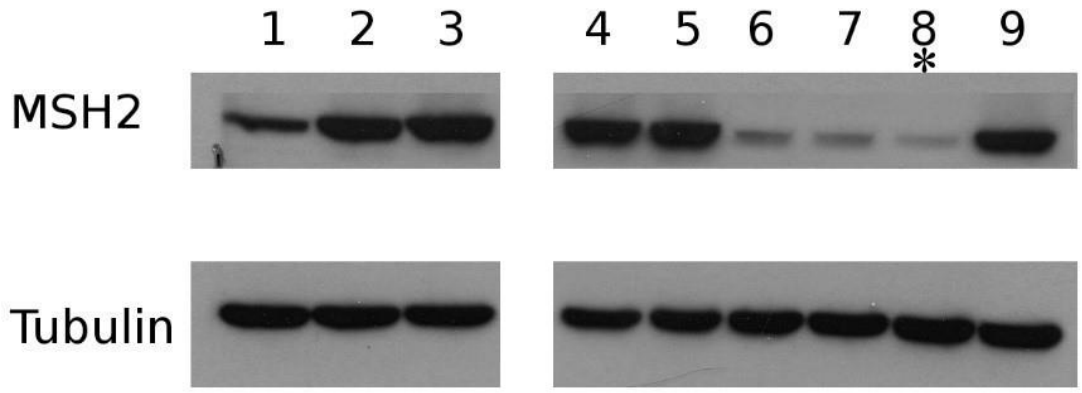

Figure 3.1 Expression of MSH2 in cell line pBWW33-67 and its derivative cell lines from control or experimental group.
Lane 1, pBWW33-67C2; Lane 2, pBWW33-67C6; Lane 3, pBWW33-67C8; Lane 4, pBWW33-67M2; Lane 5, pBWW33-67M3; Lane 6, pBWW33-67M5; Lane 7, pBWW33-67M11; Lane 8, pBWW33-67M13 and Lane 9, pBWW33-67C2.

Table 3.2 Detailed 6-TG resistance test on 5 selected cell lines.

Experiment	Cell Lines	0.5uM 6-TG	2uM 6-TG	4uM 6-TG	8uM 6-TG
1	pBWW33-67	84%	2%	3%	10%
	pBWW33-67C6	61%	3%	4%	15%
	pBWW33-67C8	54%	4%	2%	7%
	pBWW33-67M11	101%	99%	55%	25%
	pBWW33-67M13	141%	138%	117%	79%
2	pBWW33-67	54%	2%	4%	2%
	pBWW33-67M13	99%	81%	96%	58%

Two replicates were carried out for each 6-TG concentration in the second test, and their average colony number was divided by that in media only to generate survival rate. Cell line pBWW33-67M13 had survival rates over 100% in the first trial, so it was tested again together with parental cell line.

these cell lines has negative correlation with their resistance to 6-TG: the higher MSH2 expression in a cell line, the more sensitive it is to 6-TG. pBWW33-67M13 has lowest expression of MSH2 on the western blot but has highest survival rate in 6-TG media. Combining the 6-TG test and Western blot results, the cell lines pBWW33-67M11, pBWW33-67M13, pBWW33-67C6 and pbww33-67C8 are good candidates for recombination experiments. Finally, pBWW33-67M13 and two other two control cell lines, pBWW33-67C6 and pBWW33-67C8 were chosen for further recombination experiments.

Collection and analysis of spontaneous recombination and DSB induced recombination: Fluctuation tests were carried out on selected cell lines pBWW33-67M13, pBWW33-67C6 and pBWW33-67C8, and their clone frequency, analyzed clones and HeR frequency are listed in Table 3.3. DSB repair experiments were carried out on selected cell lines 3 to 5 times to collect recombinants of DSB repair, and their clone frequency, analyzed clones and HeR frequency are listed in Table 3.4.

Equivalent recombination frequency in MSH2 knockdown cells: For each pBWW33-67 derivative cell line, 12 to 22 clones from spontaneous recombination experiment were examined, and 20 to 63 clones from DSB repair were examined. pBWW33-67 and its derivative cell lines have similar spontaneous recombination frequencies at 10^{-8} and HeR events around 10^{-8}. pBWW33-67M13 has a spontaneous clone frequency of $5.71X10^{-8}$ and HeR frequency of $3.37X10^{-8}$, pBWW33-67C6 has spontaneous clone frequency of $5.95X10^{-8}$ and HeR frequency of $0.82X10^{-8}$, and pBWW33-67C8 has a spontaneous clone frequency of $3.22X10^{-8}$ and HeR frequency of $3.22X10^{-8}$.

Table 3.3 Spontaneous recombination frequencies of cell lines in MSH2 knockdown experiment.

Cell line	Subclone	Clone frequency (10^{-8})	# of Colonies analyzed	# of Homeologous recombinants	Homeologous recombination frequency (10^{-8})
pBWW33-67M13	1	4.17	2*	0	0
	2	0	0	0	0
	3	3.64	0	0	0
	4	0	0	0	0
	5	12.30	3	1	4.10
	6	19.84	9*	8	19.84
	7	0	0	0	0
	8	2.50	3*	0	0
	9	0	0	0	0
	10	14.63	8*	4	9.76
	Total		25	13	
		Avg.=5.71			Avg.=3.37
pBWW33-67C6	1	0	0	0	0
	2	13.97	4	0	0
	3	0	0	0	0
	4	0	0	0	0
	5	5.09	2	1	2.54
	6	25.82	11	0	0
	7	6.60	2	0	0
	8	0	0	0	0
	9	4.67	2	1	2.34
	10	3.33	1	1	3.33
	Total		22	3	
		Avg.=5.95			Avg.=0.82
pBWW33-67C8	1	5.26	2	2	5.26
	2	0	0	0	0
	3	0	0	0	0
	4	0	0	0	0
	5	0	0	0	0
	6	26.94	1	1	26.94
	7	0	0	0	0
	8	0	0	0	0
	9	0	0	0	0
	10	0	0	0	0
	Total		3	3	
		Avg.=3.22			Avg.=3.22

*Some HAT resistant clones of pBWW33-67M13 survived selection because of reactivated mouse tk pseudogene rather than recombination between HSV TK sequences, so the recombination frequency was adjusted as the product of raw clone frequency and proportion of recombination events.

Table 3.4 DSB-induced recombination frequencies of cell lines in MSH2 knockdown experiment.

Cell line	Expt #	Expt. Number	Clone frequency (10^{-5})	# of colonies analyzed	# of homeologous recombinants	Homeologous recombination frequency (10^{-5})
pBWW33-67M13	1	1	4.77	18	9	2.39
		2	6.14	5	2	2.46
		3	5.18	15	4	1.38
	2	1	2.7	14	4	0.77
		2	4.24	14	6	1.82
			Avg.=4.61			Avg.=1.76
pBWW33-67C6	1	1	0	0	0	0
		2	0.91	4	1	0.23
		3	0.53	6	2	0.18
	2	1	0.29	5	0	0
		2	0.29	6	1	0.05
			Avg.=0.40			Avg.=0.10
pBWW33-67C8	1	1	9.26	6	0	0
		2	6.6	6	2	2.20
		3	5.38	8	3	2.02
			Avg.=7.08			Avg.=1.41

Clone frequencies of DSB repair in these cell lines are around 10^{-5} except for cell line pBWW33-67C6 in the control group, which has an especially low clone frequency of $4X10^{-6}$. pBWW33-67M13 has a clone frequency of $4.61X10^{-5}$ and HeR frequency of $1.73X10^{-5}$, pBWW33-67C6 has a clone frequency of $0.4X10^{-5}$ and HeR frequency of $0.1X10^{-5}$.

Similar conversion tract of spontaneous recombinants in MSH2 Knockdown cells: The spontaneous HeR events in parental, control and MSH2 knockdown cell lines have these 33 mismatched nucleotides in HSV-1 TK recipient converted to HSV2-TK sequence all at once. In contrast, many DSB repair events only converted fewer than the 33 mismatched nucleotides to HSV2-TK sequence. The detail of conversion tract in DSB repair is listed in Table 3.6.

Elevated proportion of HeR events in MSH2 knockdown cells during DSB repair: The proportion of HeR events increased in MSH2 knockdown cell lines: the 39% HeR events in pBWW33-67M13 cell lines is close to two times of that in parental cell line and control cell lines (Table 3.5). However, clone frequency and HeR frequency of pBWW33-67M13 are not different from that in parental or control cell lines (Table 3.4).

Unchanged proportion of HeR events in spontaneous recombination: For spontaneous recombination, the proportion of HeR events in pBWW33-67M13 is 68%, close to pBWW33-67 and pBWW33-67C8, but different from pBWW33-67C6, which has the lowest proportion of HeR events in the 4 examined cell lines. The proportion of HeR events in pBWW33-67C6 is about 4 times lower than other cell lines in spontaneous recombination. The details of HeR events in experimental cell lines are listed in Table 3.5.

Table 3.5 HeR events of spontaneous recombination and DSB repair in MSH2 knockdown experiment.

Recombination events	Cell Line	Analyzed clones	Recipient with all mismatches	Recipient with mismatches (fewer than all)	Recipient with no mismatch	Other events	HeR Fraction
Spontaneous recombination	P	15	12	0	3	0	80%
	C6	22	3	0	19	0	14%
	C8	3	3	0	0	0	100%
	M13	19	13	0	6	0	68%
DSB repair	P	42	5	5	31	1	24%
	C6	21	2	2	16	1	19%
	C8	20	3	2	14	1	25%
	M13	64	12	13	39	0	39%

Cell lines' names are abbreviated in Table 3.5. P = parental cell line pBWW33-67, C6 = control cell line pBWW33-67C6, C8 = control cell line pBWW33-67C8 and M13 = MSH2 deficient cell line pBWW33-67M13.

Table 3.6 Converted mismatched nucleotides in HeR recombinants from DSB repair.

Cell line	Clone	Converted Mismatched Nucleotides	Coexistence of mismatched nucleotides
pBWW33-67M13	pBWW33-67M13-1-1	M1 to M17	M1 to M2
	pBWW33-67M13-1-2	M1 to M33	
	pBWW33-67M13-1-6	M1 to M7	M1 to M7
	pBWW33-67M13-1-7	M1 to M12	M1 to M12
	pBWW33-67M13-1-8	M1 to M17	M1 to M17
	pBWW33-67M13-1-9	M1 to M33	M1 to M33
	pBWW33-67M13-1-11	M1 to M33	M17 to M33
	pBWW33-67M13-1-13	M1 to M16	M1 to M16
	pBWW33-67M13-1-16	M1 to M16	M6 to M16
	pBWW33-67M13-2-7	M1 to M33	M1 to M33
	pBWW33-67M13-2-9	M1 to M33	
	pBWW33-67M13-3-3	M6 to M14 & M17 &M20	M6 to M14 & M20
	pBWW33-67M13-3-6	M1 to M33	M1 to M16, M21 to M33
	pBWW33-67M13-3-9	M1 to M16	M1 to M16
	pBWW33-67M13-3-13	M1 to M33	M1 to M33
	pBWW33-67M13-7-6	M6 to M17	M6 to M12
	pBWW33-67M13-7-7	M1 to M33	
	pBWW33-67M13-7-10	M1 to M33	M29 to M33
	pBWW33-67M13-7-16	M1 to M33	M1 to M33
	pBWW33-67M13-8-4	M1 to M30	
	pBWW33-67M13-8-7	M1 to M12	
	pBWW33-67M13-8-9	M1 to M5	M1 to M5
	pBWW33-67M13-8-11	M1 to M33	M1 to M5, M21 to M33
	pBWW33-67M13-8-14	M1 to M33	M1 to M33
	pBWW33-67M13-8-15	M1 to M12	M1 to M12
pBWW33-67C6	pBWW33-67C6-2-12	M1 to M33	
	pBWW33-67C6-3-5	M1 to M12	
	pBWW33-67C6-3-7	M1 to M33	
	pBWW33-67C6-5-10	M1 to M5 & M13 to M17	
pBWW33-67C8	pBWW33-67C6-2-3	M1 to M33	
	pBWW33-67C6-2-5	M1 to M5	
	pBWW33-67C6-3-1	M1 to M33	
	pBWW33-67C6-3-3	M1 to M22	
	pBWW33-67C6-3-6	M6 to M33	

Evidence of hDNA intermediates appeared after MSH2 knockdown: When analyze DSB repair recombinants of pBWW33-67M13 cell line, the alignments show discontinuous conversion tracts in recipient sequences where the identity of mismatched nucleotides changed between HSV-1 TK and HSV-2 TK sequences repetitively. In the chromatograms, almost all HeR events show double peaks at the positions of mismatched nucleotides, and each peak represents a nucleotide of either donor or recipient. As a result, coexistence of mismatched nucleotides causes the shifting of nucleotide calling between donor and recipient in sequencing data. The coexistence of mismatched nucleotides at each position demonstrates that the cell lines carry two set of sequence information, conceivably from both donor strand and recipient strand of a hDNA intermediate.

The chromatograms of previous recombinants were reviewed at this time and no DSB repair recombinant shows double peaks like these from pBWW33-67M13. Only one DSB repair recombinant, pBWW25-16-3-10, has mismatched nucleotides at M16 with unknown reason. Another exception is one spontaneous HeR recombinant, pBWW16-2-5-9, which has double peaks for all mismatched nucleotides.

Table 3.6 lists the conversion of mismatched nucleotide in HeR recombinants from cell line pBWW33-67M13, pBWW33-67C6 and pBWW33-67C8 during DSB repair. Spontaneous HeR events are not listed in a table because they have all mismatched nucleotides converted to HSV-2 TK sequence and they have no double peaks at mismatched position in the sequencing chromatogram.

In addition to gene conversion details, coexistence of mismatched nucleotides is labeled in a separate column. Conversion of mismatched nucleotide describes the

nucleotide positions where the donor's nucleotide is visible, while coexistence of mismatched nucleotide demands the donor's nucleotide as well as recipient's nucleotide. If the mismatched nucleotides are all from donor, they would be recorded in "converted mismatched nucleotide" column, but not marked in "coexistence of mismatched nucleotide" column.

Coexistence of mismatched nucleotides in DSB repair recombinant are separated by subcloning: To learn more about the coexistence of mismatched nucleotides in collected recombinants, 5 recombinants from pBWW33-67M13 were subcloned to separate cell populations that may carry different set of nucleotides at the mismatched positions. Sequencing results of their subclones are listed in Table 3.7.

pBWW33-67M13-1-13 and pBWW33-67M13-3-13 have continuous conversion tract starting from M1, and the nucleotides of donor and recipient coexist at each converted mismatched nucleotide position. They represent 12 out of the 20 HeR events showing evidence of hDNA. Sequencing data of subclones from these two cell lines shows that these recombinants are actually composed of two cell populations, which carry either the donor's HSV-2 TK sequence or recipient's HSV-1 TK sequence in their recipients.

Two cell lines have continuous conversion tracts starting from M1. However, mismatched nucleotides of donor and recipient only coexist at some of the converted mismatched nucleotide position. All 12 subclones of pBWW33-67M13-3-6 show donor's HSV-2 TK sequence, while all 6 subclones of pBWW33-67M13-1-16 show only HSV-1 TK sequence. Recovering one of the two cell populations is not satisfactory, however, it does not contradict with current conclusion.

Table 3.7 Separation of coexisted mismatched nucleotides was observed after subcloning recombinants from pBWW33-67M13.

Clone	Converted Mismatched Nucleotides	Coexistence of mismatched nucleotides	Mismatched nucleotides in	
			Subclone type A	Subclone type B
pBWW33-67M13-1-13	M1 to M16	M1 to M16	D1 to D16	R1 to R33
pBWW33-67M13-3-13	M1 to M33	M1 to M33	D1 to D33	R1 to R33
pBWW33-67M13-1-16	M1 to M16	M6 to M16	NA	R1 to R33
pBWW33-67M13-3-3	M6 to M14 & M17 &M20	M6 to M14 & M20	D17 to D20	NA
pBWW33-67M13-3-6	M1 to M33	M1 to M16, M21 to M33	D1 to D33	NA

The details of gene conversion tracts are listed after the recombinants, and the identities of mismatched nucleotides in their subclones followed afterward. The mismatched nucleotides of donor are labeled as D1 to D33, and the mismatched nucleotides of recipient are labeled as R1 to R33.

The last cell line pBWW33-67M13-3-3 has discontinuous conversion tract, and nucleotides of donor and recipient only coexist at some of the converted nucleotide positions. Surprisingly, all 7 recovered subclones carry the same partial HSV-2 TK sequence and the conversion of nucleotides is different from DNA sequence of their parental cell lines. It is not clear how this event with confusing sequence data happened during DSB repair.

Collection of spontaneous recombinants containing all cells from one recombination event: No spontaneous recombinant of pBWW33-67M13 was found with double peaks in chromatograms. To find evidence for hDNA in spontaneous recombination, further attempts were made to harvest a spontaneous recombination clone representing a single recombination event.

A modified fluctuation test was carried out on pBWW33-67M13 twice, and all recovered clones are listed in Table 3.8. Different from the previous fluctuation test, HAT selection was carried out on approximate 40 million cells of each subclone without re-plating, and conceivably, all descendant cells from a recombination event form a discrete clone. However, in modified fluctuation test, there was still no evidence of hDNA. Recombinants recovered in a subclone tend to have the same type of events: 3 subclones have all HR recombinants and 7 subclones have all HeR recombinants. Only subclone 9 in experiment 1 have both HR recombinant and HeR recombinant.

Reactivated mouse thymidine kinase pseudogene in non-recombinant clones: HAT resistant clones with unaltered recipient had been found sporadically in Chapter 2, but it was common for pBWW33-67M13 in spontaneous recombination. Non-recombinant clones with a reactivated mouse thymidine kinase pseudogene were

Table 3.8 Spontaneous recombinants collected from modified fluctuation test.

Experiment	Subclone	Recombinants analyzed	Homeologous Recombinants
1	1 to 6	0	0
	7	2	0
	8	2	2
	9	3	1
	10	0	0
2	1	0	0
	2	2	2
	3	0	0
	4	1	0
	5	0	0
	6	2	2
	7	0	0
	8	1	0
	9	1	1
	10	0	0
	11	1	1
	12 to 17	0	0
	18	1	1
	19	1	1
	20 to 24	0	0
	Total	17	11

All recombinants from any subclone were harvested and examined. All HeR recombinants in this chart have all 33 mismatched nucleotides converted to HSV-2 TK sequence.

observed in our laboratory before (first characterized by Jessica Burr Lea in her Master Thesis titled "Influence of the DNA Long Inverted Repeat on Recombination in Mammalian Cells"). Therefore, two drug resistant tests and RT-PCR were carried out to confirm the identities of current non-recombinant clones (89, 90).

Trifluorothymidine (TFT) is toxic to all cells with functional thymidine kinase while Ganciclovir (GCV) specifically targets cells with thymidine kinase of herpes simplex virus. The non-recombinant clone, pBWW33-67M13-4-11, reached confluence in GCV containing media but died out in TFT containing media, while control cell line grew well in either TFT or GCV containing media. Figure 3.2 shows that the mouse TK transcript was efficiently amplified in non-recombinant clones, but not from parental cell lines.

RT-PCR targeting mouse TK gene as well as pseudogene was carried out on non-recombinant clone pBWW33-67M13-4-11, parental cell line pBWW33-67M13 and previous cell lines classified in Jessica Burr Lea's master thesis. Lane 1 to 6 show RT-PCR products of non-recombinant clones in presence or absence of reverse transcriptase, while lane 7 to 9 show RT-PCR products of parental cell lines. Mouse TK sequence or pseudogene sequence was successfully amplified from pBWW33-67M13-4-11 in presence of reverse transcriptase while minor amplification happened in parental cell line pBWW33-67M13. Similarly, previously classified non-recombinant clones, pJAVA3 5c-SS-1a and pJAVA1 26E-O have the same 612 bp TK sequence amplified through RT-PCR, while the parental cell lines JAVA1 26E and pJAVA3 5c do not.

Finally, the RT-PCR products of recombinant clone pBWW33-67M13-4-5, non-recombinant clone pBWW33-67M13-4-11 and parental cell line pBWW33-67M13 in

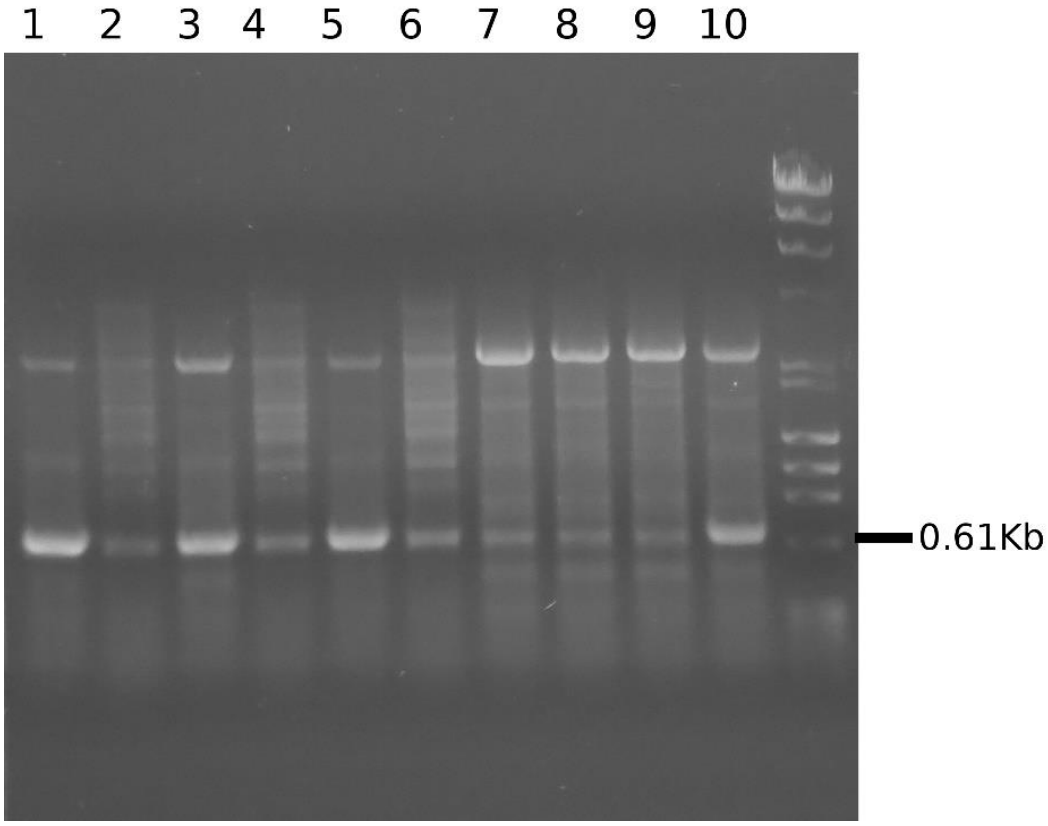

Figure 3.2 Mouse TK gene or pseudogene were successfully amplified through RT-PCR from selected non-recombinant cell lines and parental cell lines.
Lane 1, pJAVA3 5c-SS-1a; Lane 2, pJAVA3 5c-SS-1a without reverse transcriptase; lane 3, pBWW33-67M13-4-11; lane 4, pBWW33-67M13-4-11 without reverse transcriptase; lane 5, pJAVA1 26E-O; lane 6, pJAVA1 26E-O without reverse transcriptase; lane 7, pJAVA1 26E; lane 8, pJAVA3 5c; lane 9, pBWW33-67M13; lane 10, pJAVA1 5c-VV-2a.

pBWW33-67M13 experiment were excised, re-amplified and sequenced. These transcript sequences were aligned to the mouse TK gene and its pseudogene to examine their identity.

As shown in Figure 3.3, the 612 bp transcript sequences amplified from these cell lines are mouse TK pseudogene because they carry all the 8 unique nucleotides from mouse TK pseudogene. Meanwhile, these three transcripts have 14 additional mismatches across the sequence compared with either reference mouse TK sequence. PCR and sequencing the pseudogene's genomic sequence in examined cell lines shows that most new found mutations on transcript sequence are not found on the genome. Reviewing the chromatogram of the transcripts' sequences reveals that mixed reading occurs at these 14 mutation sites. These "mutations" in transcripts' sequence is probably an artifact from prolonged RT-PCR process.

Discussion:

Generation of homeologous conversion tract: To address the mechanism of observed HeR events, experiments in the current chapter were designed to learn how these mismatched nucleotides were incorporated into the recipient sequence during recombination. In SDSA, newly synthesized recipient strand may take donor's mismatched nucleotides, anneal to the other resected recipient strand and form hDNA. In DSBR model, hDNA may form during strand invasion, second strand capture, or HJs migration that mediates strand exchange between donor and recipient. In both situations, if MMR repair the hDNA toward the foreign sequence, the recipient would receive homeologous conversion tract. Without MMR, the hDNA would be separated after DNA replication, and one of the daughter cells gains homeologous conversion tract afterward.

Figure 3.3 Amplified transcript sequences in selected cell lines were aligned to mouse TK gene and pseudogene.
M13, M13-4-11 and M13-4-5 are parental cell line, non-recombinant and recombinant respectively. Totally 8 mismatches exist between mouse TK gene and its pseudogene within the sequencing area, and these mismatched nucleotides are marked with "V". Perfectly matching sequences were highlighted in the figure for clarity.

Therefore, mismatched hDNA formed during HR could be repaired or separated to generate homeologous conversion tract.

Alternatively, excessive end resection and degradation may chew out the homeologous sequence on the recipient, and the broken recipient gets converted through DNA synthesis along the template strand. Thus, no mismatched hDNA is formed through annealing diverged sequences during the process.

Experiments in the current chapter specifically ask whether these HeR events possibly went through hDNA intermediate annealing homeologous sequences. If the hDNA intermediates do exist, whether they always include the whole homeologous region? Is it possible that the shorter conversion tracts are repair product of hDNA intermediates annealing the whole homeologous region?

Establishment of MMR impaired cell line: To answer these questions above, MSH2 in cell line pBWW33-67 was stably knockdown using shRNA. Then, spontaneous recombination and DSB induced recombination in MMR deficient background were examined and compared. Cell line pBWW33-67 was chosen for MSH2 knockdown experiment since it has higher fraction of HeR events during DSB repair, and half of the HeR events have few than all mismatched nucleotides converted. Alterations in conversion tract would be easy to see if they occur.

Obtained cell lines were screened by Western blot and 6-TG experiments. The chosen cell line of MSH2 knockdown group, pBWW33-67M13, has the lowest MSH2 expression, and highest survival rate in 6-TG experiments, indicating functional deficiency in MMR. In contrast, the control cell lines pBWW33-67C6 and pBWW33-67C8, with control shRNA integrated, have regular MSH2 expression, and they are

sensitive to 6-TG like the parental cell line. Data from these two experiments also shows the MSH2 expression in examined cell lines has negative correlation with their resistance to 6-TG, which agreed with the core role of MSH2 in MMR.

HeR recombinants of DSB repair in MSH2 deficient cells: Out of 25 HeR recombinants of pBWW33-67M13 obtained from DSB repair experiments, 20 recombinants show double peaks at mismatched nucleotide position in sequencing chromatogram. Coexistence of mismatched nucleotides in these HeR events implies that hDNA intermediates formed between homeologous region segregates into two daughter cells in MMR deficient cell, so these clones have mixed genetic information from both donor and recipient.

To learn whether these 20 recombinants are mixed cell populations from hDNA segregation events, subcloning was attempted on 5 selected cell lines. Two cell lines, pBWW33-67M13-1-13 and pBWW33-67M13-3-13, have continuous conversion tract and mixed nucleotides at every mismatched position. Subcloning of these two cell lines successfully separates two cell populations: one have all the donor's nucleotides at each converted position while the other have all the recipient's nucleotides. The result here suggested these recombinants are likely products of hDNA intermediates, which are separated into daughter cells instead of repaired by MMR. These two cell lines represent 12 out 20 HeR carrying mixed nucleotides.

pBWW33-67M13-3-6 and pBWW33-67M13-1-16 have continuous conversion tracts and mixed nucleotides at some mismatched positions, representing 7 out of 20 HeR carrying mixed nucleotides. Every examined subclones from these two cell lines have sequence from one side, donor or recipient. The results agree with segregation of two

strands with different identities, however, information of the other strand is not visualized. It is not clear why the information of one strand was lost during subcloning.

Only for clone pBWW33-67M13-3-3, the result of subcloning is not compatible with the sequencing data of its parental cell line. The recovered subclones have converted nucleotides different from parental cell, new converted nucleotides appear while some converted nucleotides from parental cell line disappears. The mystery of this particular recombinant could not be explained by simple HR models, and remains unsolved.

Again, 20 out of 25 recombinants show double peaks in sequencing chromatograms, suggesting segregation of hDNA intermediates encompassing homeologous sequences. Therefore, the vast majority of HeR events involves formation of hDNA intermediate. And the doubled percentage of HeR events after MSH2 knockdown implies that while hDNA intermediates in normal cells are repaired by MMR toward either strand, the foreign strands in hDNA intermediates are always preserved and visible in MMR deficient cells.

No change in clone frequencies after MSH2 knockdown while increase in HeR frequency: After MSH2 knockdown, clone frequencies are not dramatically different from that in parental and control cell lines. One-way ANOVA comparing frequencies in parental, control and knockdown cell lines give a p-value equals 0.3942. HeR frequencies look higher after MSH2 knockdown, while the same test give a p-value equals 0.0508. Neither frequencies are statistically significant between tested groups.

No change in conversion tracts after MSH2 knockdown: 12 out of 25 HeR events in pBWW33-67M13 have 33 mismatched nucleotides converted to HSV-2 TK sequence, close to the 50% seen in parental cell line pBWW33-67. Unchanged ratio

between partial and whole conversion tract after MSH2 knockdown suggested that MMR does affect the length of conversion tract. Therefore, partial conversion tract is repair product of hDNA intermediates annealing short homeologous sequences, while whole conversion tract is product of hDNA intermediate annealing the complete homeologous sequences.

Gene conversion tracts in MMR deficient cells support previous assumption that HeR events with fewer converted mismatched nucleotides do resolve in middle of the homeologous region. The alternative explanation that HeR events with fewer converted mismatched nucleotides came from incomplete repair of 33 mismatched nucleotides, is also barely feasible due to the mechanism of MMR: MMR always degrades hundreds or thousands of nucleotides on one strand and re-synthesize it based on the other intact DNA strand. It is unlikely that MMR switched the choice of strand within this 200 bp short sequence. Also, if hDNA intermediates always include the entire homeologous sequence but they are repaired differently, depletion of MSH2 will abrogate MMR and keep this intermediate with all 33 mismatched nucleotides for later segregation. Thus, it was predicted that in pBWW33-67M13, fraction of recombinants converting all mismatched nucleotides would rise due to deficiency of MMR. However, we observed a slightly decrease of these recombinants. To sum up, available evidences confirmed the existence of hDNA intermediates that anneal the partial homeologous region proximal to I-SceI site as well as that anneal the whole homeologous sequence.

The results in current chapter and previous two chapters clearly show the role MMR plays in collected HeR events. For examined three substrates, HeR events are common phenomenon in both spontaneous recombination and DSB-induced

recombination, and increase of homeologous sequence does not dramatically decrease either HeR frequency or proportion of HeR events. Moreover, MSH2 knockdown failed to bring dramatic boost in HeR frequency, though it increased fraction of HeR events up to 2-fold during DSB repair. Apparently, there is no vigorous rejection or destruction toward the genetic exchange between homeologous sequences by MMR in examined cell lines. It is possible that as long as nearby homology is available to initiate recombination, homeologous sequence could participate in genetic exchange without noticeable discrimination. Taken together, MMR does not target and eliminate genetic exchange between homeologous sequences in these substrates, while it rather repairs the hDNA to generate homeologous conversion tract, most likely half of the time. These results allow us to say MMR does not impede recombination between homeologous sequences but actually facilitates it between diverged sequences by converting hDNA leftover to template sequence. Though against the intuition and current scheme for MMR, it is reasonable for cells to allocate energy to reject improper recombination at the initiation stage, rather than destroy any ongoing genetic exchange between diverged sequences.

No detection of hDNA during spontaneous recombination: In pBWW33-67M13, 13 out of 19 spontaneous recombinants were products of HeR event. However, none of them show double peaks at the position of mismatched nucleotides in chromatography, which provided the evidence of hDNA intermediate during DSB repair. In fluctuation test, cells were trypsinized and plated right before the HAT selection, so if one recombination event generates two daughter cells carrying either strand of the hDNA intermediate, they are probably separated by trypsinization and form discrete colonies in

these flasks. Therefore, the fluctuation test was modified to collect clone that contains all descendant cells of a recombination event.

Fluctuation test of pBWW33-67M13 has no more than 8 clones in one subclone. If these clones were from a recombinant cell containing hDNA remnant, the recombinant cell went through divisions up to three times before HAT selection. Since most subclones grew up to about 40 million cells before HAT selection, the recombinant DNA product of a recombination event most likely still resided in a single cell when the subclone had 4 million cells, if the recombination event already happened. Therefore, in modified fluctuation test, each subclone was cultured up to 4 million cells. Then these 4 million cell were plated sparsely, cultured for another 3 days before fed with HAT media for selection. Ideally, offspring of the same recombinant cell will form a single recombinant clone with mixed cell populations. In new experiments, 8 subclones generated HeR events, but none shows direct evidence of hDNA segregation: double peaks in sequencing chromatography. Only one subclone has both HeR and HR recombinants, a sign of possible hDNA segregation, while the other 7 subclones only have HeR recombinants. The percentile of hDNA segregation events is dramatically lower than that in DSB repair. Fisher exact test comparing the ratio of hDNA segregation event versus non-segregation events in spontaneous recombination and DSBR gives p-value equal 0.0012. Therefore, hDNA segregation dominates the HeR events of DSBR while it is minor in spontaneous recombination. Currently, no direct evidence shows that spontaneous recombinants go through hDNA intermediates annealing homeologous sequences.

Two possibilities remain for further exploration. First, as described in Chapter 3, the hDNA intermediates actually formed between homeologous sequences, and they included I-SceI site as well. Therefore, segregation of hDNA in subsequent cell division produced one daughter cell carrying donor's continuous TK strand and the other daughter cell carrying I-SceI disrupted recipient strand. Since the latter daughter cell would perish in HAT selection, only the first one was able to form pure HeR recombinant in fluctuation test. Though in agreement with the mechanism of MMR and HR, the speculation still need further investigation to see whether hDNA intermediates annealing all the mismatches and I-SceI site are actually the precursors of collected HeR recombinants. It is still possible that the spontaneous recombinants in current experiments simply went through successive template switching during replication rather than hDNA intermediate formed by strand exchange.

Reactivation of mouse thymidine kinase pseudogene in collected clones: Many non-recombinant clones, mostly in fluctuation test of pBWW33-67M13, survived HAT selection without nucleotide sequence change in HSV-1 TK recipient. It was initially described in Jessica Burr Lea's Master thesis that once a processed pseudogene of mouse TK gene (91) is transcribed in high quantity, its expression can supply these Ltk- cells with functional thymidine kinase. Data from two drug tests shows these newly discovered non-recombinant clones did not grow in media with TFT, but reached confluence in presence of GCV. These drug tests indicate active thymidine kinase in these cells, but not of Herpes simplex virus origin (89, 90). The RT-PCR amplifying the transcript of mouse TK gene as well as pseudogene shows successful amplification in non-recombinant clones, but not in recombinant or parental cell lines. The transcript's

sequence in aforementioned clones have 8 nucleotides identical to mouse TK pseudogene but not mouse TK gene. Meanwhile, all these three transcripts possess many mutations different from either mouse TK gene or pseudogene. However, genomic sequence of the mouse pseudogene in aforementioned three clones are almost the same as the reference pseudogene sequence. The large amount of mutations in RT-PCR products are probably artifact after prolonged amplification steps, or less likely, vigorous RNA editing of the transcripts.

All these results suggested that the expression of mouse tk pseudogene arises dramatically in non-recombinant clones, which supplied tk function in these cells rather than the tk of herpes simplex virus in recombination substrates.

CHAPTER 4 INVESTIGATION OF THE RELAXED HOMOLOGY REQUIREMENT FOR RESOLUTION DURING DSB REPAIR

Holliday junction (HJ) refers to the cross-shaped, four-stranded DNA structure after two DNA molecules swap one strand with each other during recombination (Figure 1.1). HJ was initially introduced to describe the exchange between homologous chromosome: the HJ formed between two DNA molecules migrate to conduct strand exchange between them until its resolution (20). This hypothesis captures the nature of genetic exchange between homologous partners, and it is later on adopted in multiple theoretical models to explain DNA recombination in various circumstances, such as Meselson-Radding model, DSBR model and restoration of replication fork (7, 92). To date, the existence and dynamic of HJ during recombination have been illustrated by many genetic, cytological and molecular evidences in meiosis as well as mitosis (20, 44, 93, 94). The gene conversion events are most time unidirectional, either through gap repair or by repair of asymmetric hDNA. Sometimes in meiosis, when sequence divergence exists between the homologous partners, the HJ mediated strand exchange may leave hDNA intermediate on both molecules with mismatched nucleotides, which are repaired by MMR to produce reciprocal conversion tract (7). Without the MMR responsible for the latter process, information from both strands will be visible after post meiosis segregation of the hDNA (54, 95). Though HJs were found in mitosis at much low frequency, it is conceivable that it can mediate similar strand exchange in mitotic cells (93).

While HJs are adopted in various models explaining usually rare mutual exchange between homologous DNA sequences, SDSA serves as a prominent and simple way in which the broken DNA utilize homologous partner for repair without altering the HR template (Figure 1.1) (17, 18). Considering the recombination frequency and homology requirement in these two types of recombination, strict homology requirement appears to be required in resolution of HJ: it is likely that the resolution of HJs need significant amount of homology to properly anneal the four strands for recognition and procession. In contrast, the SDSA in mitosis may anneal strands with mismatched nucleotides, showing sloppy choice in resolution site.

Alternatively, chromatin structure around DSB site is modified due to DNA damage response (DDR), for example, the local phosphorylation of H2AX on chromatin or other potential structural changes brought by DDR factors (1, 96). Therefore, the homology requirement could be dramatically different from that proximal to a DSB lesion. In another word, the homology requirement for resolution may be relaxed by nearby DSB.

To validate aforementioned possibilities, it is critical to collect and analyze the recombinant clones solely from HJs resolution. Analyzing existing recombination events in prior chapters does not lead to an unequivocal answer. The cause of spontaneous recombination is still debatable and it might be achieved through rescuing replication fork or a two ends DSB; while DSB repair could take either SDSA or DSBR pathway. After scrutinizing all possible pathways of HR, it is found that when dHJs resolution occurs, the sequence of homologous template (donor) could be altered and converted to recipient sequence. As demonstrated in Dr. Szostak's DSBR model and his supportive

data studying HR between gapped plasmid and chromosomal gene, this particular type of events has been detected at low frequency in yeast mitotic cells (7, 97).

As shown in Chapter 2 and 3, when HR event started near a homeologous sequence, the broken DNA received homeologous conversion tract through either gap repair or repair of hDNA. Conversely, the HR template may gain mismatched hDNA by taking recipient strand during HJs migration. If aforementioned HJ branch migration and resolution in DSBR model do generate desired recombinant clones with altered donor, the conversion tract on donor would present the footprint of branch migration between initiation site and resolution site of HJ. If the HJs need adequate homology to resolve, the conversion tract should always contain the whole homeologous region, the same as observed in spontaneous recombination, while if DSB reduce the homology requirement in resolution, the conversion tract on donor will have conversion tract contain partial homeologous region as well as whole homeologous region.

In order to obtain recombinant clones from HJs resolution, and compare them with previous results from pBWW substrates, new substrate was designed to recreate the same condition for potential HR events (Figure 4.1). In previous chapters, the full length HSV1-TK sequence was broken at the I-SceI site (nucleotide 1215), and the invasion was primed by its broken ends toward a 0.8 Kb hybrid donor, which has HSV2-TK homeologous sequence about 50 bp away from the invasion site (nucleotide 1215). In the new design, though a 2.5 Kb full length tk gene serves as hybrid donor, and the I-SceI site resides in the 0.8 Kb HSV1-TK recipient, after breakage, it is the same HSV1-TK DNA ends invading a similar template with homeologous sequence about 50 bp from the

invasion site. And most importantly, the HR events are driven by the same amount of homology: nucleotide 623 to 1459 of HSV1-TK.

To demonstrate the similarity in HR events happening in pBWW33 and new pLS4 substrate, their HSV-1 TK recipients are aligned to their hybrid donors respectively (Figure 4.1): the I-SceI site in recipient is always aligned against the downstream HSV-1 TK sequence in hybrid donor, about 50 bp away from the border of HSV2-TK sequence. In both substrates, HR events rely on the 0.8 Kb homology shared by donor and recipient.

The original 2.5 Kb HSV1-TK fragment was modified extensively to make the hybrid donor for substrate pLS4. The donor would have mismatched nucleotides M33-M8 of pBWW33 substrate, and then 5 point mutations afterward introducing a BamHI recognition site and two stop codons. As a result, the hybrid donor has a cluster of 31 mismatched nucleotides about 50 bp upstream of the I-SceI insertion site (nucleotide 1215), compared with HSV1-TK recipient. In this arrangement, the invasion and repair events would be almost the same as in substrate pBWW33, where the broken HSV1-TK strands invade a 0.84 kb hybrid donor at a site 50 bp away from a cluster of 33 mismatched nucleotides.

For the new substrate pLS4, HR or NHEJ events ligating the broken recipient would not necessarily bring HAT resistance to the host cells. Recombinant clone from DSB repair will grow up only when the HR restore the broken recipient through dHJs resolution, and the HSV2-TK sequences together with stop codons in 2.5 Kb hybrid donor are converted to HSV1-TK recipient after repair of hDNA intermediate. Sequencing the HR donors in this type of recombination events will reveal the conversion tracts on participating HR donor, which in turn tells the range of HJs migration.

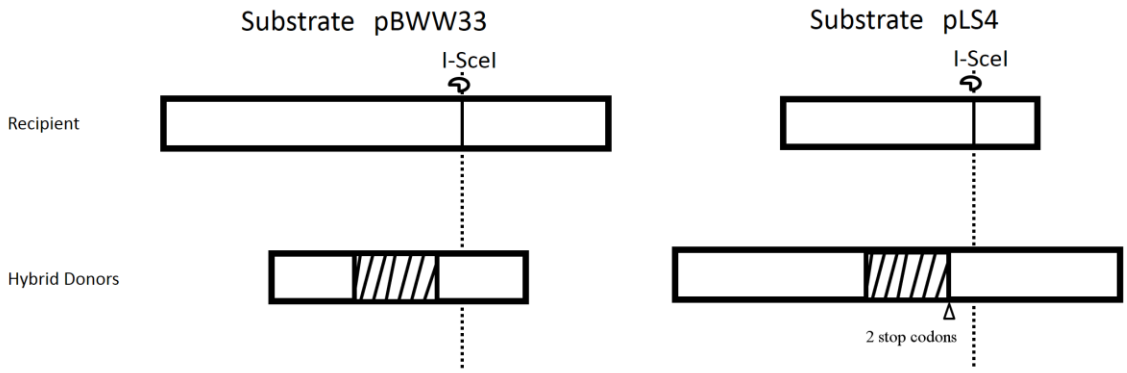

Figure 4.1 Alignment of donor and recipient in substrate pBWW33 or pLS4.
Donor and recipient in pBWW33 and pLS4 are placed side by side to show the similarity in recombination events happened between these two substrates.
The white rectangles represent HSV-1 TK sequences and the striped rectangles represent HSV-2 TK sequences. The recombination events happened within pBWW33 and pLS4 are similar because they have hybrid donor and HSV1-TK recipient sharing the same length of homology. The HSV2-TK sequence in donor of pLS4 starts from the same position as in hybrid donor of pBWW33, and it possesses the same mismatched nucleotides M33 to M8 in donor of pBWW33. The 5 additional mismatched nucleotides introduced a BamHI site and 2 stop codons right after HSV2-TK sequence.

Materials and Methods:

Structure of substrates: The new substrate pLS4 has a 2.5 Kb hybrid tk donor in tandem with a truncated HSV1-TK recipient. The hybrid donor is full length HSV1-TK sequence carrying a patch of HSV-2 TK sequence, an engineered BamHI recognition site and two engineered stop codons in middle of its coding sequence (CDS); and expression of donor's encoded thymidine kinase is aborted by these two stop codons. The new truncated recipient is made of 837 bp HSV-1 TK sequence (nucleotide 623-1459, the same length and location as the hybrid donor of pBWW33), and it has 30 bp I-SceI recognition site at the position 1215 (the same I-SceI insertion site as in recipient of pBWW33). The positions of these nucleotides are all numbered the same as in the Wagner's paper, as described in Chapter 2 (74).

Substrate pLS4 was linearized by endonuclease ClaI before transfection, and the restriction map of the linearized substrate is show in Figure 4.3. To check the completeness of these tandem repeats after genome integration, endonuclease BamHI or HindIII are used to digest genomic DNA for Southern blot.

Construction of substrate pLS4: The backbone of pLS4 is from pBWW16, and HSV1-TK recipient and donor were amplified either from pTK1 (carry complete HSV-1 TK gene sequence), pBWW16, or previously obtained recombinants pBWW33-67-1-1 (nucleotide sequence 968-1167 in its HSV-1 TK recipient was substituted by HSV-2 TK sequence). The cloning protocols were the same as described in Chapter 2 and the work flow was described in following paragraphs.

At the beginning, pBWW16 was cut by HindIII, and its 8.2 Kb backbone was purified and ligated to generate pLS2. To make pLS1, 0.87 Kb HSV1-TK recipient

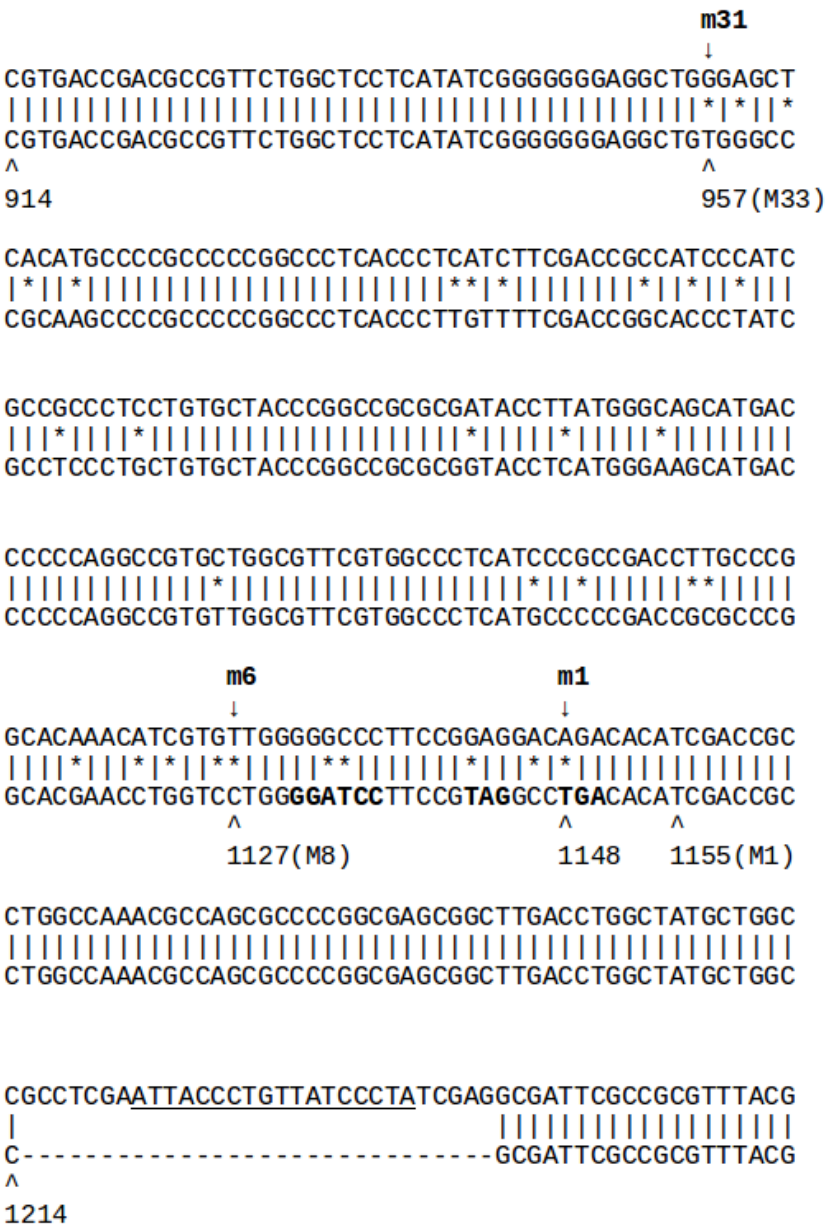

Figure 4.2 Mismatched nucleotides between recipient and donor of substrate pLS4. Compared with donor, the recipient (top sequences) has 30 bp oligonucleotide insertion including 18 bp I-SceI recognition site (underlined). Totally 31 mismatched nucleotides exist between donor and recipient (marked with * in the alignment), and they are numbered from m1 to m31 depending on their closeness to the I-SceI insertion site, the same as they were labeled in pBWW33. The mismatched nucleotides m31 to m6 in pLS4 are correspondingly M33 to M8 in pBWW33. And 5 more different nucleotides, m1 to m5, introduce BamHI site and two stop codons (all in bold, between nucleotide 1127 and nucleotide 1155). Nucleotide positions are marked below the sequence with "^", and annotated in parentheses if they are corresponding to a mismatched nucleotide in pBWW33 or pLS4.

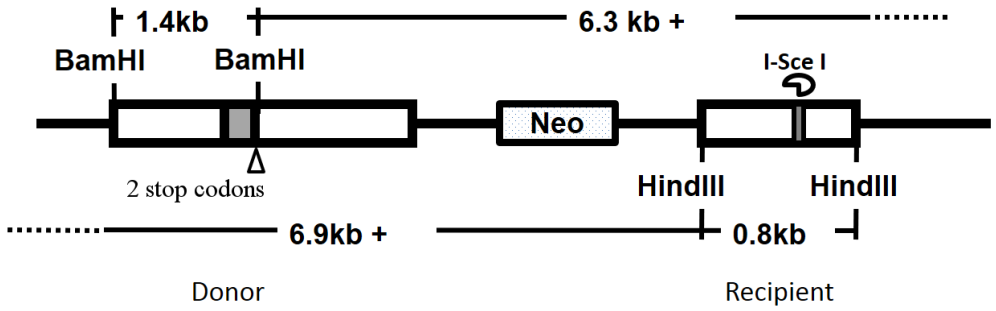

Figure 4.3 Structure and restriction map of substrate pLS4.
The linearized substrate is 9.1 Kb, and it has a 0.9 Kb HSV1-TK recipient flanked by HindIII recognition site as well as a 2.5 Kb hybrid donor carrying two BamHI site. The Neo cassette between the two repeats is for selection purposes in both prokaryotic cells and eukaryotic cells.

(623-1459) containing I-SceI insertion at position 1215 was amplified from pBWW16 using AW130 and AW93, and then the PCR product was cut by HindIII and purified for ligation. Substrate pLS2 was cut by HindIII, treated by bacterial alkaline phosphatase (BAP), gel purified and ligated with aforementioned 0.87 Kb recipient fragment to generate pLS1. In subsequent cloning steps, the plasmid backbones were always treated by SAP and gel purified for ligation with insertion fragment.

The 2.5kb full length hybrid donor was cloned into pLS plasmids through two steps of PCR and ligation, during which HSV2-TK sequence, BamHI recognition site and 2 stop codons were introduced by PCR primers and selected templates.

First, 1.1Kb HSV1-TK sequence was amplified from pTK1 using AW146 (5' CAG CCA GGA TCC TTC CGT AGG CCT GAC ACA TCG ACC G 3', matches HSV1-TK nucleotides 1132-1162) and AW147 (5' AAG TAG GGA TCC AGA TCT CGA GCT TGG CCG TGG TG 3', matches the 3' end of the 2.5Kb, BamHI flanked HSV1-TK fragment). This fragment contains partial coding sequence of HSV1-TK and downstream non-coding sequence. BamHI site and 2 stop codons were introduced at the 5' end of the fragment by primer AW146, while BglII was introduced at the 3' end by primer 147. The amplified PCR product was cut sequentially by BglII and BamHI, and then purified for ligation with 6.6 Kb plasmid backbone from BamHI digested pLS1. The ligated pLS3 has a 0.87 Kb I-SceI disrupted recipient and a 1.1 Kb HSV1-TK donor tail with 2 stop codons in coding sequence. After ligation, the intact BamHI site at the 5' end of the insertion is followed by 2 stop codons while the BamHI site at the 3' end was destroyed through the re- ligation with BglII digested sticky end.

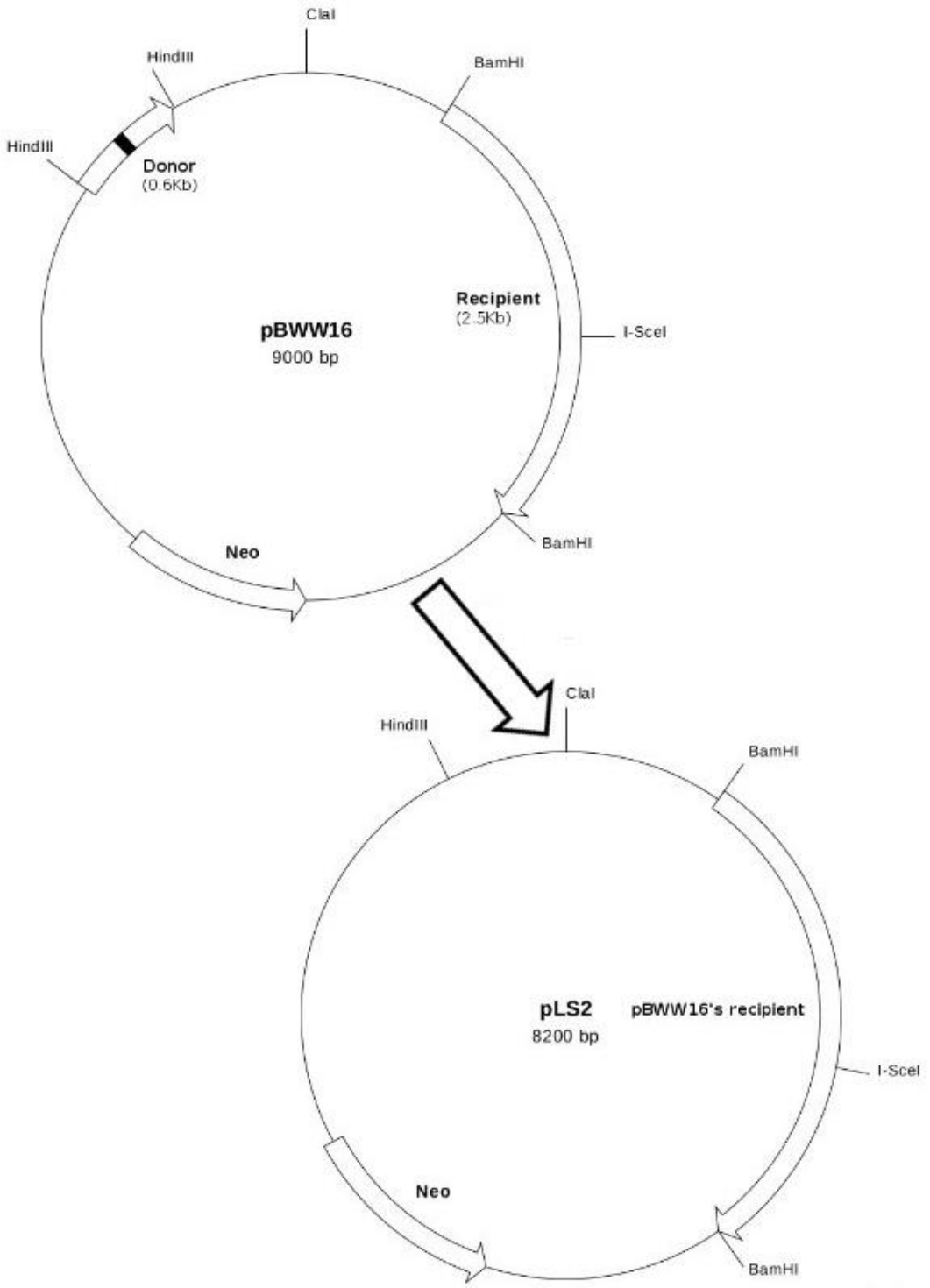

Figure 4.4 Work flow to construct substrates pLS2.
Plasmids, primers and restriction sites in use are drawn and labeled accordingly in the figure: the white bars in donors and recipient are HSV-1 TK sequences, while the solid bars in these donors are HSV-2 TK sequences. Donor fragment in pBWW16 was excised and the remaining plasmid was ligated to produce pLS2. pLS2 was then used to build plasmid pLS1.

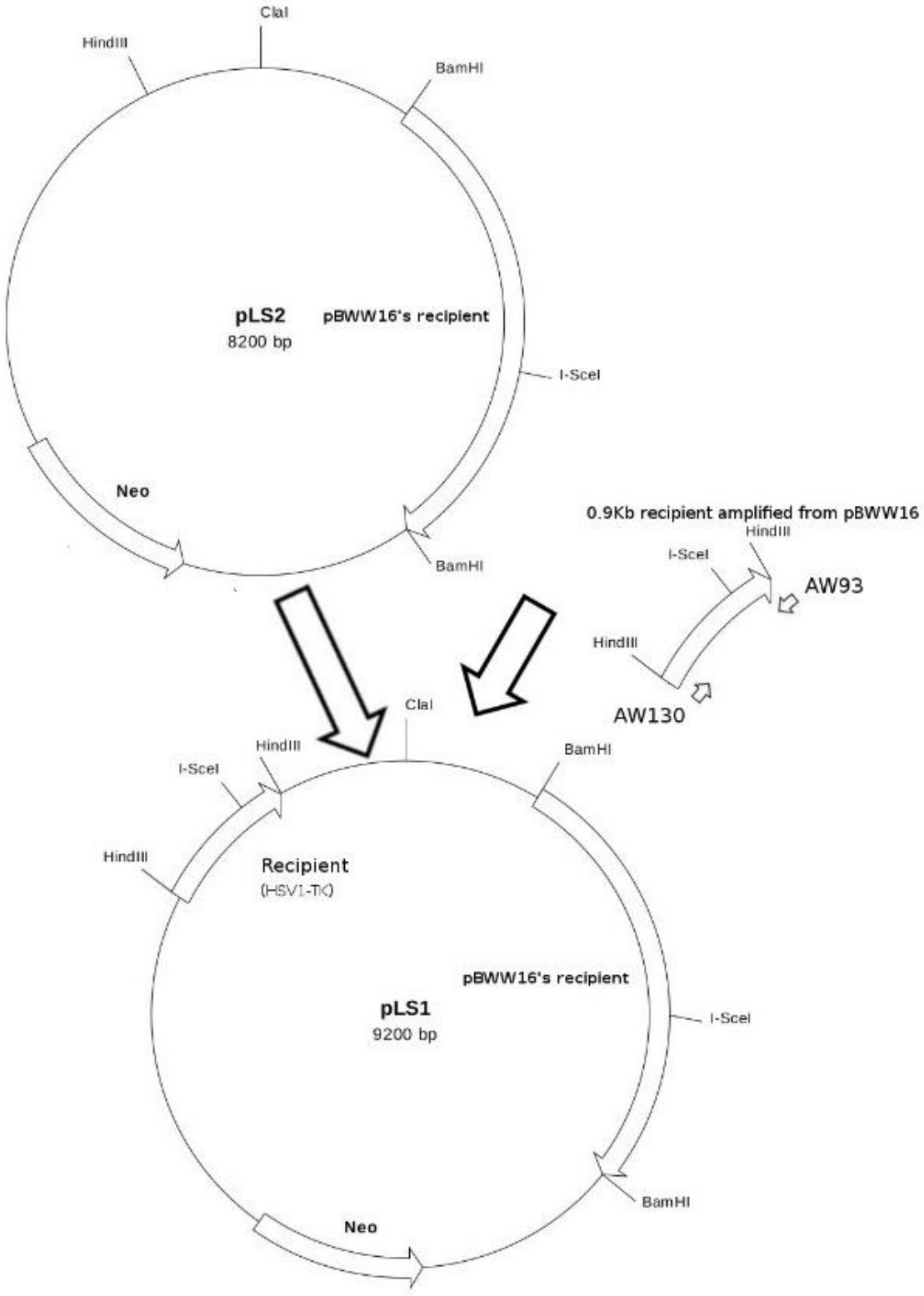

Figure 4.5 Work flow to construct substrates pLS1.
A 0.9 Kb, PCR amplified recipient was inserted into the unique HindIII site of pLS2 to produce plasmid pLS1, which was used to build plasmid pLS3.

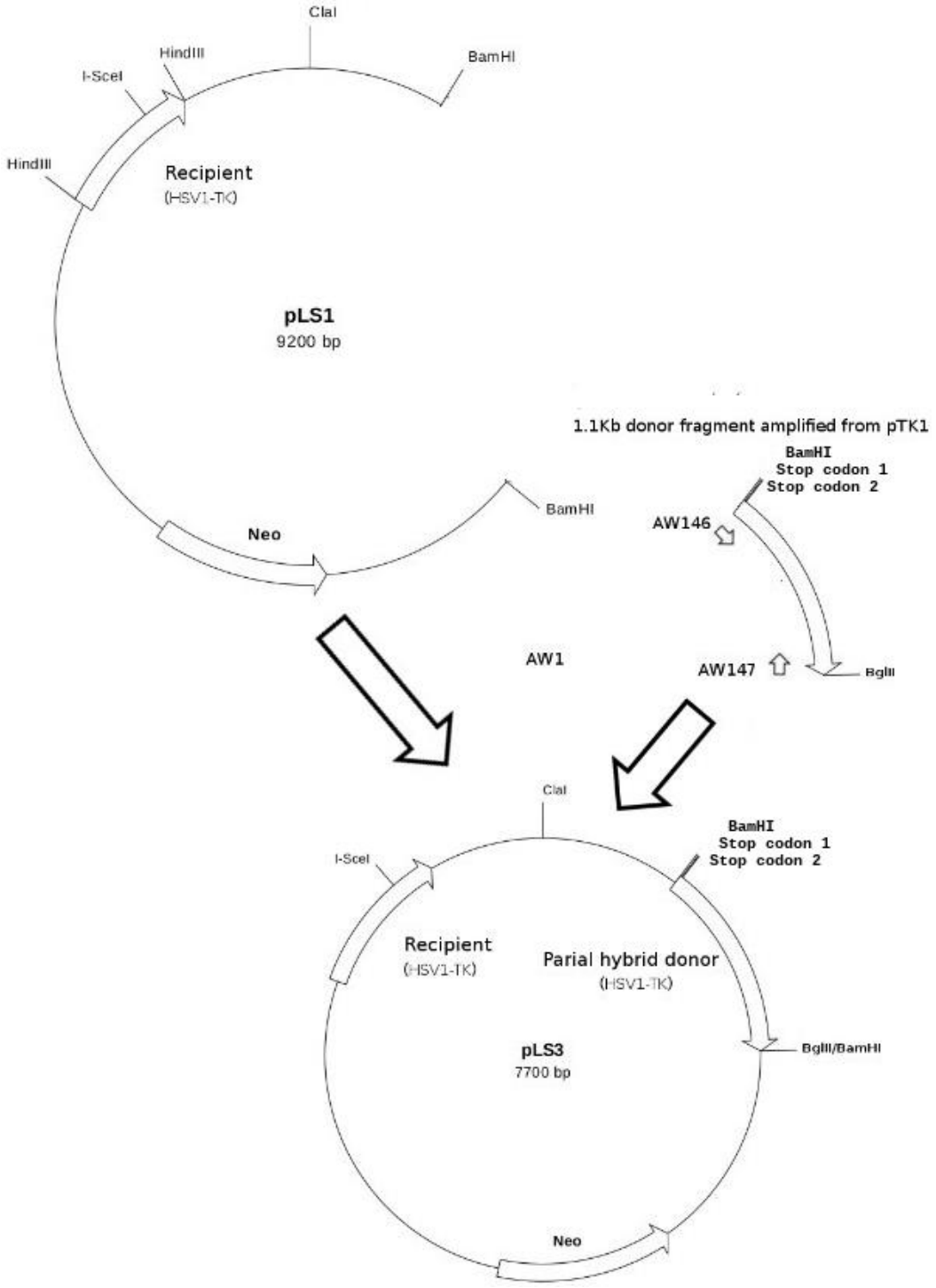

Figure 4.6 Work flow to construct substrates pLS3.
The plasmids were labeled the same way as in Figure 4.5. The 2.5 Kb TK fragment was excised from pLS1 by BamHI, and a 1.1 Kb, PCR amplified donor's downstream fragment was inserted into the open BamHI site to produce plasmid pLS3, which was used to build plasmid pLS4.

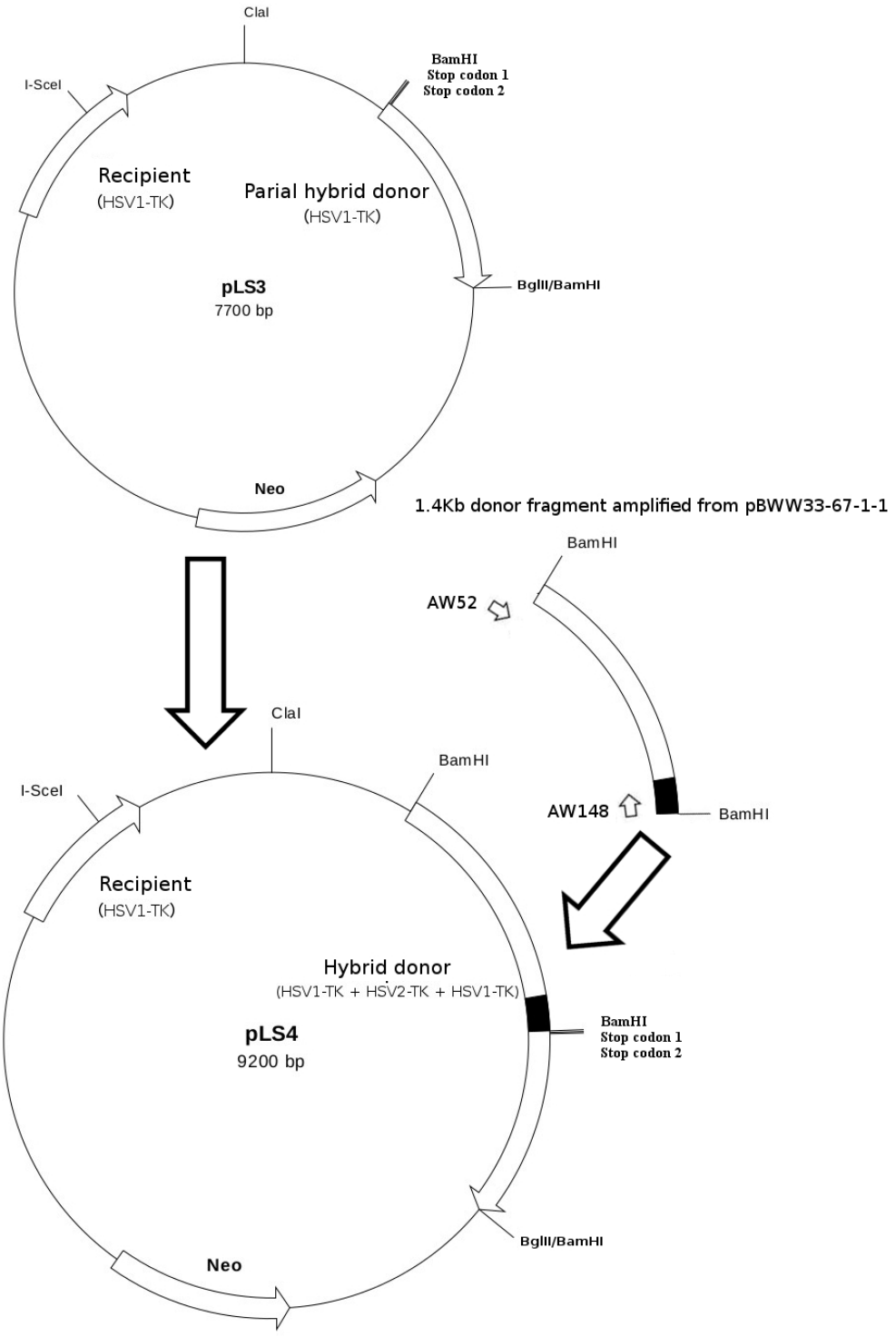

Figure 4.7 Work flow to construct substrates pLS4.
The plasmids were labeled the same way as in Figure 4.5. A 1.4 Kb, PCR amplified donor's upstream fragment was inserted into the unique BamHI site of pLS3 to produce plasmid pLS4. pLS4 is the recombination substrate used in subsequent DSB repair experiments.

Second, 1.4Kb HSV1-TK sequence containing promoter and partial coding sequence was amplified from clone pBWW33-67-1-1 using AW52 (matches plasmid backbone upstream of BamHI site) and AW148 (5' CCG GAA GGA TCC CCA GGA CCA GGT TCG TGC CGG 3', matches HSV1-TK nucleotides 1131-1111). The PCR product was cut by BamHI and purified for ligation with 7.7 Kb backbone from BamHI digested pLS3. Therefore, the pLS4 has a 2.5 Kb hybrid donor with a stretch of HSV2-TK sequence followed by 2 stop codons.

Cell culture: Mouse Ltk- cell line and its derivative cell lines are cultured as described in Chapter 2.

Cell line establishment: Mouse Ltk- cells were transfected with pLS4 through electroporation and the cells with stable substrate integration were selected using G418. The electroporation, selection and screening process were the same as described in Chapter 2. Cell lines were chosen for further experiments if they have intact tandem repeats and the signal indicates a single copy integration of the substrate.

Collection of DSB repair recombinant clones: To collect DSB repair recombinants, electroporation of selected cell lines with pSce plasmid and subsequent HAT selection were carried out as described in Chapter 3. HAT resistant clones were ready to pick after 2 -3 weeks in selection, and the clone frequencies or HeR clone frequencies were calculated the same as in Chapter 3. These clones were picked and cultured for sequence analysis as described in Chapter 2.

PCR amplification and sequencing: The PCR and sequencing procedures were described in Chapter 2.

Analysis of DSB repair recombinants: All alignments were carried out in SciEdCentral to compare recombinant sequence with reference sequences (engineered hybrid donor sequence and HSV-1 TK recipient sequence). Through alignment, conversion of mismatched nucleotide from HSV2-TK to HSV1-TK sequence on donor would be revealed. Moreover, nucleotide callings at every mismatched postilion (m1 to m30) were checked manually in chromatography.

Results:

Construction of substrates: the HindIII flanked donor was removed from pBWW16 to generate pLS2. Compared with pBWW16, new pLS2 is linearized by HindIII without releasing 0.7 kb fragment. pLS2 was confirmed through HindIII digestion in Figure 4.8.

After 0.9 kb I-SceI disrupted HSV1-TK recipient was amplified from pBWW16's recipient, it was then ligated into the unique HindIII site of pLS2. Figure 4.9 confirmed successful insertions of 0.9 Kb HSV1-TK recipient in 4 pLS1 plasmid: successful insertion was observed in pLS1-7 to pLS1-10 (Lane 3 to Lane 6). Recipient in pLS1-8 has the correct sequence and direction according to DNA sequencing data, therefore, the plasmid was renamed as pLS1 and used in following experiments.

The downstream 1.1 kb HSV1-TK sequence was amplified from pTK1 and inserted into the BamHI site of pLS1. If the PCR product was corrected inserted, new plasmid will show a longer vector fragment, and a 1 kb plasmid fragment after ClaI and BamHI double digestion, as shown in Lane 3 of Figure 4.10. Double digestion by ClaI and BamHI revealed pLS3-16 clone has 1.1 Kb HSV1-TK insertion with right direction

(see Figure 4.10). plasmid pLS3-16 was confirmed by sequencing and renamed as pLS3 for subsequent experiments.

At the last step, 1.4 kb upstream HSV1-TK sequence was amplified from a recombinant clone carrying 33 mismatched nucleotides, and then inserted into the unique BamHI site of pLS3. BamHI digestion was used to revealed pLS4 clones with correct 1.4 Kb HSV1-TK insertion (see Figure 5.11). Successful insertions were found in pLS4-5B, pLS4-6B and pLS4-8B. After sequencing, plasmid pLS4-6B has the insertion with correct sequence and direction and finally renamed as pLS4.

Establishing cell lines pLS4: pLS4 was linearized at the unique ClaI site before transfection and the substrates was transfected into mouse Ltk- cells through electroporation as described before. Totally 29 G418 resistant clone (pLS4-1 to pLS4-29) were picked, cultured and screened on Southern blots for single copy integration of the substrate.

Screening cell lines with single copy integration of substrate: Genomic DNA from each clone was digested using HindIII or BamHI for Southern blot. In HindIII digestion, the cell line with intact substrate should show a DNA fragment lining up with the 0.9 Kb recipient, and a junction fragment larger than 6.9 Kb, the size of remaining non-recipient fragment. Similarly, in BamHI digestion, the cell line should have a DNA fragment lining up with a 1.4Kb donor fragment, and a junction fragment larger than 6.3 Kb. All cell lines in use were checked by both digestion to make sure they have intact donor and recipient.

Selected Southern blot images of pLS4 stably transfected cell lines are shown in Figure 4.12 and Figure 4.13. The HSV-1 TK probe recognizes the HSV-1 TK recipient as

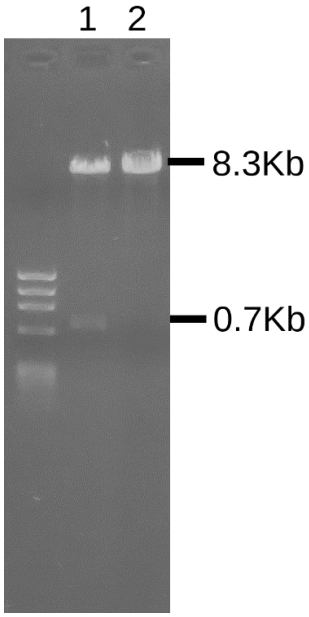

Figure 4.8 Identification of plasmid pLS2 through HindIII digestion.
Lane 1, HindIII digested pBWW16. Lane 2, HindIII digested pLS2. The 8.3 Kb and 0.7 Kb labels point to the backbone fragment and donor fragment of digested pBWW16.

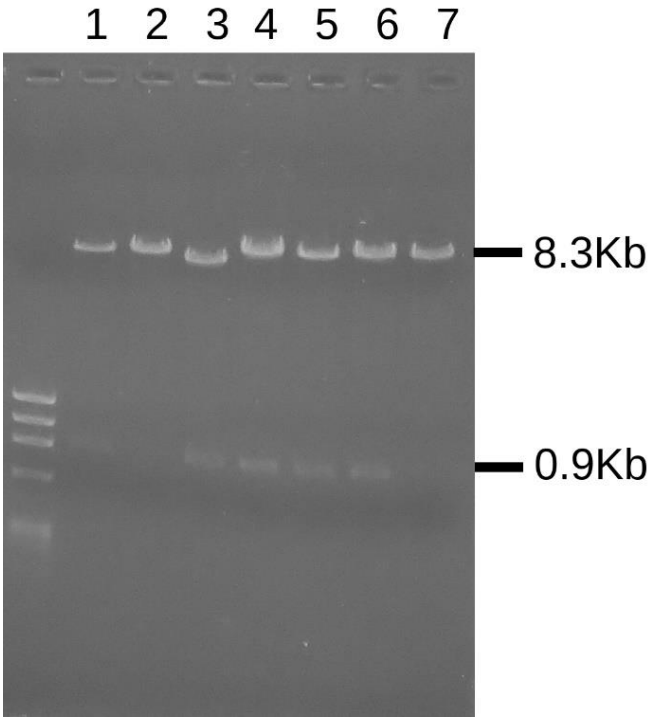

Figure 4.9 Plasmids pLS1 were confirmed with 0.9 Kb recipient insertion through HindIII digestion.
Lane 1 to 7, HindIII digestion of plasmid from clone pLS1-5 to pLS1-11.

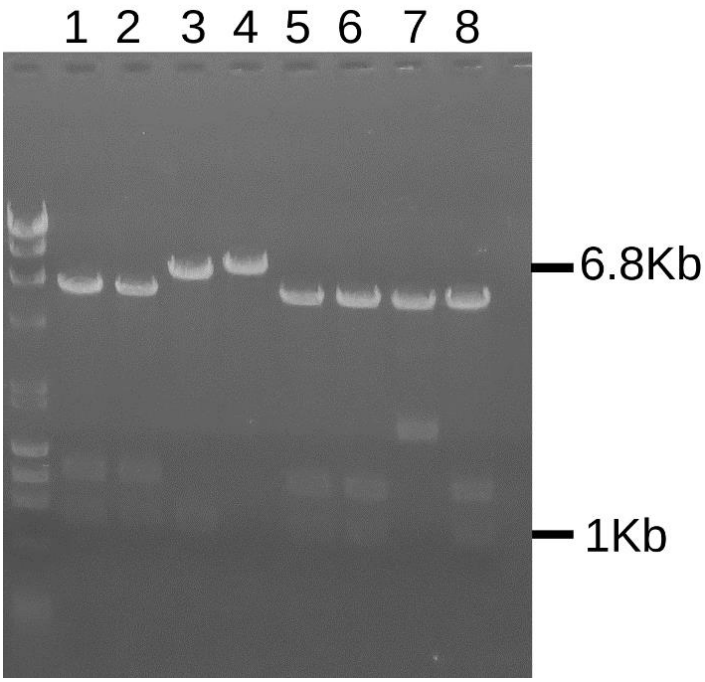

Figure 4.10 Plasmid pLS3 with 1.1 Kb donor insertion was confirmed by BamHI and ClaI double digestion.
Lane 1 to 8, double digested plasmid from clone pLS3-14 to pLS1-21.

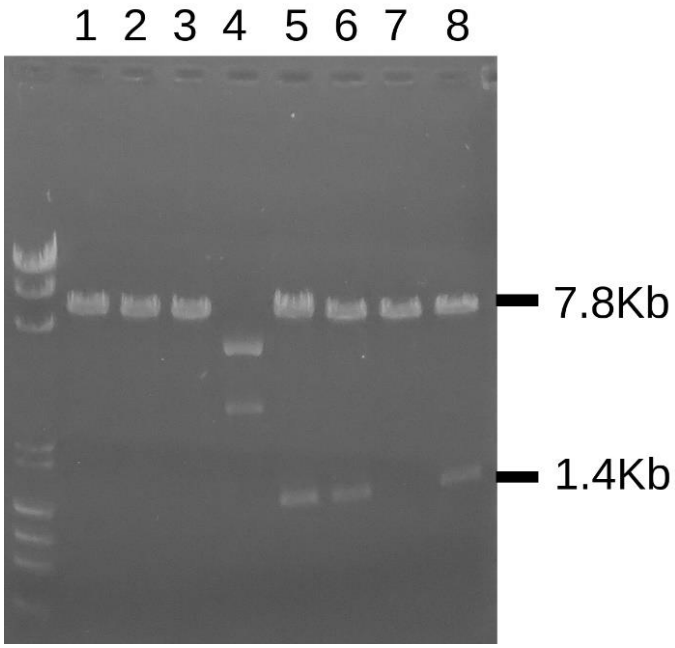

Figure 4.11 Plasmids pLS4 with 1.4 Kb donor insertion were confirmed by BamHI digestion.
Lane 1 to 8, BamHI digestion of plasmid from clone pLS4-1B to pLS4-8B.

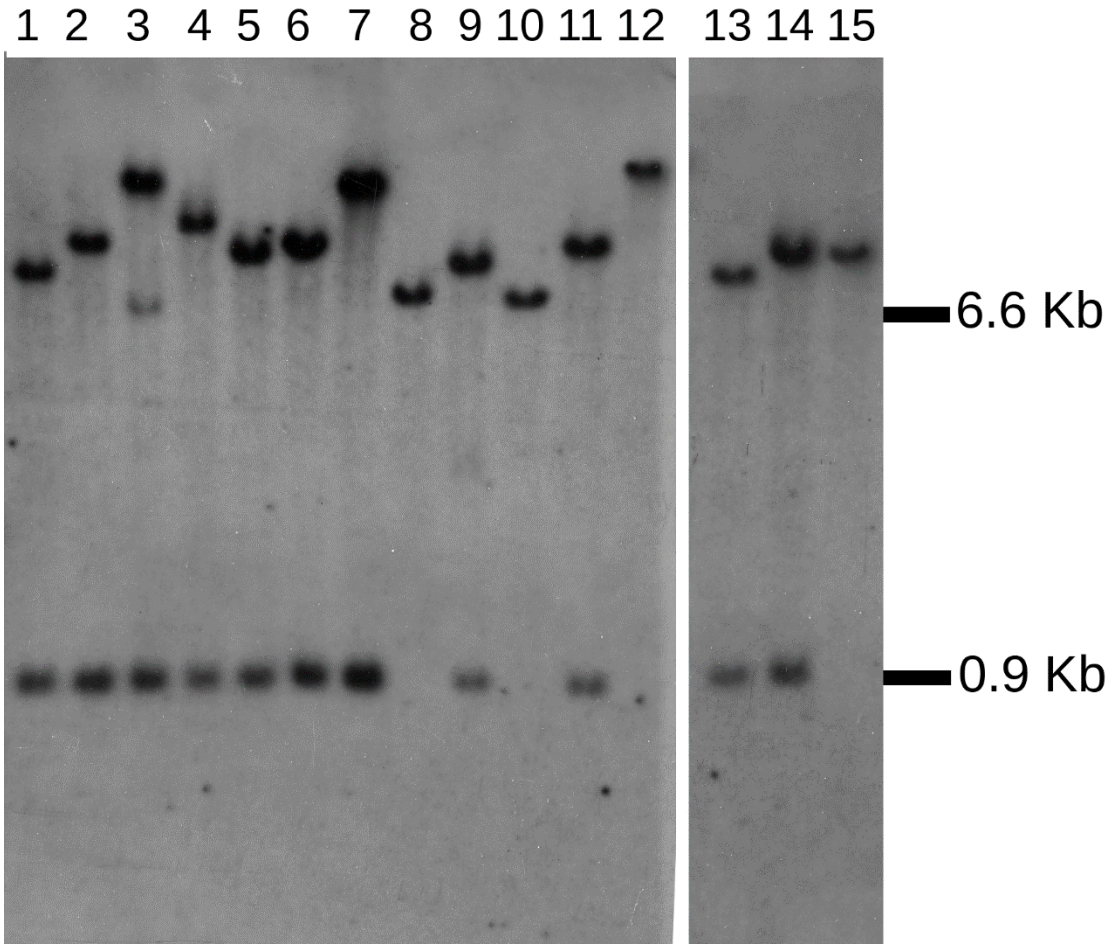

Figure 4.12 Identification of pLS4 cell lines with correct substrate integration by HindIII digestion.
The size of recipient was examined as described above. Lane 1 to 15, HindIII digested genomic DNA of pLS4 stably integrated cell lines: pLS4-2 to pLS4-11, pLS18, pLS4-20, plS4-21, pLS4-23 and pLS4-26. Cell lines pLS4-5 (lane 4), pLS4-7 (lane6) and pLS4-23 (lane 14) were chosen for following experiments.

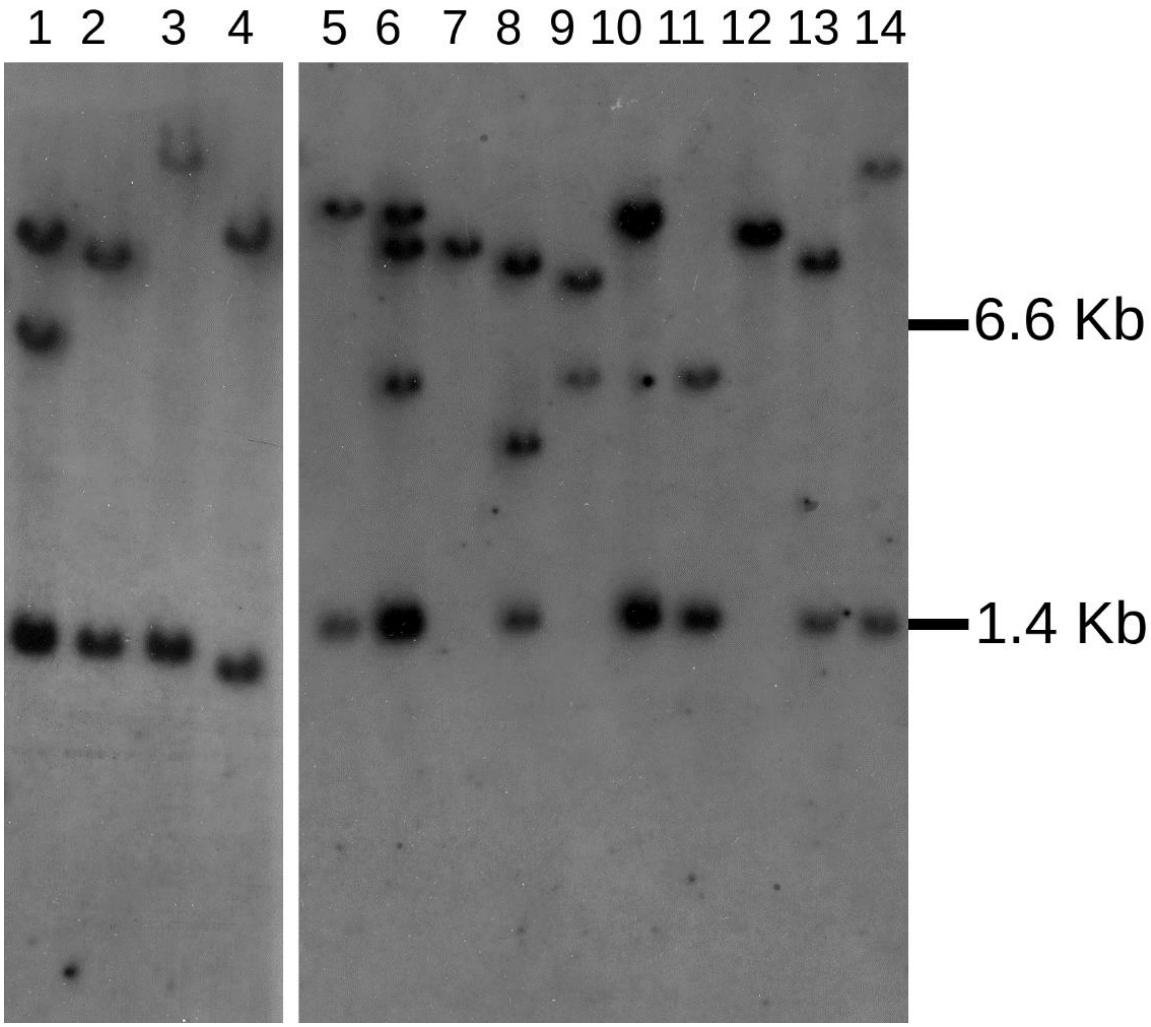

Figure 4.13 Identification of pLS4 cell lines with correct substrate integration by BamHI digestion.
The size of donor fragment was screened as described above. Lane 1 to 14, BamHI digested genomic DNA of pLS4 stably integrated cell lines: pLS4-3 to pLS4-11, pLS18, pLS4-20, plS4-21, pLS4-23 and pLS4-26. Cell lines pLS4-5 (lane 3), pLS4-7 (lane5) and pLS4-23 (lane 13) were chosen for following experiments.

well as the hybrid donors. Cell lines were chosen for future experiments only when they have intact donor and recipient fragments.

The Southern blot results of selected cell lines pLS4-5, pLS4-7, pLS4-23 were labeled in these figures (Figure 4.12 and Figure 4.13). The plasmid digestions on these blots were not as expected, and failed to indicate the correct molecular weight. Therefore, the 6.6 kb MW maker was used to estimate the size of junction fragment. Junction fragments in chosen cell lines were much larger than 6.6 kb, supposedly having intact donor and recipient.

Collection of DSB repair recombination events: for each selected cell line, 4 electroporation experiments were initially attempted to collect DSBR recombinants. All three cell lines have similar plating efficiency (around 30%) after electroporation, while only pLS4-7 generated HAT resistant clones within 4 attempts. Totally, 23 resistant clones were selected from roughly 20 million cells of pLS4-7 after I-SceI expression.

When electroporation was carried out on cell lines pLS4-5 and pLS4-23, 5 million untreated cells were plated in HAT selection to check potential existed spontaneous recombinant. No recombinant grew up in these selections either. Four mock electroporation experiments were carried out later on for pLS4-7 cell line using 1 X TE instead of pSce, while surprisingly, 83 resistant clones grew up in HAT selection from 20 million cells.

The hybrid donor sequences in pLS4-7 cell line, and 4 resistant clones from pSce transfection were amplified and sequenced as before. The results show that they all carry the original hybrid donor sequence with stop codons. So far, no induction of DSBR

recombinants was observed in examined cell lines and no desired DSBR recombinants were recovered in performed experiments.

Discussion:

Designed substrate to collect recombinants from dHJs resolution: As proposed in introduction, current recombination substrate was designed to collect only recombinant clones from dHJs resolution. Analysis of this type of recombinant could tell whether HJs resolution requires proper homology to resolve, and whether DSB affects homology requirement of HJ resolution.

Ideally, if DSB-induced recombination have HJs established in nearby homology, the HJ may anneal the homeologous sequences into mismatched hDNA. Upon establishment, as long as HJs need proper homology to resolve, they have to resolve outside the homeologous region and include the whole homeologous sequences in the conversion tract.

Failure in recovery of desired DSBR recombinants: Because induction of DSB induced recombinant clones was not observed across examined cell lines, the attempts failed to collect recombinants solely from dHJs resolution during DSB repair.

Without obtaining conversion tracts on donor through Holliday junction resolution, it remains unclear what relaxed homology requirement during DSB repair, the nearby DSB or the pathway choice of SDSA.

Explanation and future direction: In meiosis, reciprocal conversion of two participating DNA molecules were occasionally recovered after HR, while there are not many reports studying the possible change in HR template in mitosis. In Dr. Orr-weaver and Dr. Szostak's original study on DSB-induced recombination between gapped plasmid

and chromosome, 3 out of 72 recombination events possibly changed the chromosomal donor after they repaired the gapped plasmid through gene conversion (98). However, this phenomenon was barely reported in mitotic cells afterward even if periodically observed.

The desired gene conversion events changing donor as well as the recipient DNA are expected to occur at low frequency during mitosis. First, the predominant HR pathway in DSB repair is assumed to be SDSA, and only a small fraction of HR would progress into HJs intermediate (19). Second, Holliday junction dissolution could resolve dHJs without altering donor (21), and then the rest events may diverge into crossover or gene conversion events with even smaller probability.

MMR may reduce the chance to collect desired gene conversion events. Even HJ migration and resolution do leave recipient's DNA strand in donor, in the form of hDNA intermediate, it may be restored back to original donor sequence. If the MMR prefers to repair the hDNA intermediate toward the original sequence, it is even harder to recover the gene conversion events on donor without intervention in MMR. Since it is demonstrated that the deficiency of MMR does not alter the resolution site in spontaneous recombination or DSB-induced recombination (Chapter 3), depleting MSH2 could be attempted next to preserve information from both strands of the hDNA intermediate after HR. Ideally, after HJ resolution, hDNA intermediates left on donor are not corrected by MMR, and their exogenous recipient strand would be preserved and segregated to produce viable recombinants with conversion tract on donor sequence.

As discussed in introduction, the desired recombinants from HJ resolution are probably too low to be detected in current experimental condition, and multiple rounds of

electroporation may be needed to collect the expected recombinants. Second, if the MMR prefers restoring donor's hDNA toward its original sequences, the desired recombinants are not recoverable in current condition. Depleting MSH2 may be attempted next to reveal potential sequence change in donor brought by recipient through dHJs resolution.

CHAPTER 5 RECQ4 AFFECTS THE PATHWAY CHOICE IN HOMOLOGOUS RECOMBINATION DURING DSB REPAIR

RecQ4 helicase belongs to a conservative DNA helicase family which shares homology with RecQ of bacteria (39, 99). There are five human RecQ like helicase, while three of them, WRN, BLM and RecQ4 are directly linked to genetic diseases characterized by genome instability, premature aging or cancer predisposition. Deficiencies in RecQ helicases have such great impacts on genome maintenance that they have been postulated to play roles in DNA replication, DSB-induced recombination and telomeric maintenance. Current research and opinions on their functions are illustrated in following paragraphs.

Werner syndrome arises from deficiency in RecQ helicase WRN, and the patients show premature aging and cancer predisposition (39). WRN has multiple catalytic activities on DNA, including 3' to 5' DNA helicase, DNA dependent ATPase, 5' to 3' exonuclease, and strand annealing. *In vitro* experiments show that WRN interacts with DNA-PKcs and inhibits its function unless assembled with DNA and Ku70, while conversely, DNA-PKcs also phosphorylates WRN *in vitro* and *in vivo*, though the impact is not fully understood yet (65). Besides these direct interactions with NHEJ core components, WRN has been proven critical in NHEJ events: WRN's deficiency brings more repair events with large deletions, indicating delayed NHEJ and a shift to A-NHEJ (31, 100). A recent report revealed more details of WRN's function and its effect on NHEJ: study of NHEJ and A-NHEJ events in WRN deficient cells suggests that WRN

promotes NHEJ, reduces A-NHEJ, and decreases the use of microhomology in end joining. WRN's catalytic activities, exonuclease and helicase are critical to NHEJ while its solely presence inhibits A-NHEJ. Moreover, WRN significantly reduces CtIP's recruitment to DSB site, and inhibits progressive resection needed for HR and A-NHEJ (41). Without WRN, the repair shifts toward A-NHEJ pathway, which promote telomere fusion and chromosomal translocation. Consistent with existing reports, nearly half of NHEJ events in WRN deficient cells collected in our laboratory have chromosomal rearrangements, which were confirmed by Southern blot and cytological analysis (unpublished data).

Bloom syndrome arises from mutations in BLM helicase, and BLM syndrome patients have cancer predisposition in multiple tissues, while their cells have obvious cytogenetic abnormality showing high frequency of SCE during mitosis (12, 13, 70). Without proper function of BLM, crossover becomes preferred outcome of HR in BLM deficient cells (70). As described before, crossover events in meiosis promote genetic diversity and ensure correct separation of homologous chromosomes, however, when somatic cells are forced to repair lethal DSB lesions, crossover is neither necessary nor optimal outcome. In BLM deficient cells, elevated crossover between homologous chromosome is conceivable, which may facilitate LOH and contribute to the development of a wide spectrum of cancer (9, 101).

The similar roles BLM plays at two HR steps explain the bias toward crossover in BLM deficient cells. First, BLM could unwind newly synthesized DNA strand from its template, allowing it to anneal back to the other resected DNA strand for ligation (102, 103). Therefore, BLM directs HR events toward the SDSA pathway instead of a dHJs

intermediate, which has great potency to generate crossover (16, 19). Similar to aforementioned function, BLM may cooperate with MMR to prevent homeologous invasion at this earlier step of HR. Sgs1 (yeast homolog of BLM) interacts with MMR's components, recognizes and efficiently unwinds strand annealing with mismatched nucleotides during SSA (68), so at strand invasion step, similar mismatched hDNA within D-loop intermediate is likely unwound the same way. Although helicase UvrD of E.coli interacts with MutS-MutL and unwinds the hDNA containing nucleotide mismatches in vitro (87), the choice of helicase in mammalian cells could be different and still need further investigation.

Secondly, BLM propels convergent movement of dHJs to hemicatenated DNA, with two DNA duplexes conjoining through a single-strand interlock, and then the hemicatenated linkage gets resolved by decatenation instead of HJs resolution (21, 99). Hemicatenated or catenated structure frequently arises in DNA replication or DNA recombination, and it needs prompt unlocking for subsequent segregation (104). In yeast, epistatic effect was observed between sgs1, topoisomerase IIIα and RIM1 in resolving this particular replicative lesion, which confirmed sgs1's participation in DNA decatenation (105–107). In following up biochemical assays, the complex was reconstituted to drive convergent branch migration of HJs along DNA substrates and detach DNA molecules in catenation (104, 108, 109). Similarly in mammalian cells, BLM mediates the convergent branch migration of dHJs to a single-strand interlock, which is readily resolved by associated topoisomerase IIIa and RMI1-RIM2 (22), therefore, BLM manages to reduce crossover events during HR.

RecQ4 deficiency causes Rothmund-Thomson syndrome, and the patients have skeletal abnormalities and predisposition to malignant tumors (110). Cytological evidences show RecQ4 deficient cells have issues in genome stability and proper segregation of chromosome (111). RecQ4 participates in telomeric maintenance, and protects chromosomes from telomere erosion. RecQ4 is also assembled in initiation complex for replication (112). Overexpression of RecQ4 has been found in prostate cancer, and its expression level correlates with the tumor grade, while depletion in RecQ4 reduces the tumor growth and invasiveness (113).

Moreover, RecQ4 participates in DSB repair. RecQ4 moves to the repair loci earlier than other RecQ helicase members (BLM or WRN) (110) and the recruitment depends on MRE11 (15). RecQ4's unwinding activity is vital for the 5' end resection and it also helps to recruit CtIP to the repair site (15). RecQ4 also shows high affinity to branched DNA substrate, especially HJs structure, therefore, it could help to resolve DNA intermediates during replication or recombination (114). In Arabidopsis, RecQ4 is reported with anti-crossover activity in meiosis similar to Top3a or FANCM (115).

Though suggested with multiple functions, RecQ4's role during DSB repair still lack systematic studies. Limited studies on RecQ4 in DNA repair were conducted in mammalian cells, especially possible impact on DSB repair outcome. A recent published paper described the interaction and potential cooperation between BLM and RecQ4 supported by cytological evidence (116), In their paper, SCE events in control cells and patient's BLM deficient cells were briefly compared before and after RecQ4 knockdown. Due to the chronic stress BLM deficient cells went through, potential genetic changes make them not optional experimental system for further manipulation. The comparison of

recombination events immediately after the protein depletion, as conducted in current chapter, would be better to avoid potential distortion from genetic background. Moreover, the use of substrate pLB4-11 could present changes in multiple pathways for DSB repair, and also subtle details of HR. If RecQ4 plays roles in DSB repair and affects the repair outcome, different composition of DNA recombination events would be expected in RecQ4 deficient cells.

pLB4-11 cell line has a chromosomal integrated recombination substrate pLB4, which was designed to collect all types of DSB repair events. As shown in Figure 6.1, pLB4 has a 2.5 kb full length tk gene and a 3.9 kb tk-neo fusion gene in tandem, and they serve as donor and recipient in a HR events. The 2 kb DNA sequence separating tandem repeats contains hygromycin resistance gene, which was used in substrate integration. An I-SceI recognition site was engineered into the tk-neo fusion gene, and these 22 bp oligonucleotide insertion shift the reading frame, so the tk-neo gene only produces a truncated thymidine kinase protein. Initially the downstream neo gene is not expressed due to the shifted reading frame of tk gene, until a recombination event restores the correct reading frame for the fusion gene.

When I-SceI introduce a DSB into the tk-neo fusion gene, cells have chance to rejoin the two broken ends and sometimes, regain the original reading frame for the fusion gene. Therefore, DSB repair events may allow host cells to survive G418 selection, and form clones for sequence analysis. Reading frame could be restored in several ways. First, HR could direct the resected DNA overhang to the nearby homologous template, the tk gene sequence of donor, and the broken recipient sequence would get restored to a continuous HSV-1 TK sequence, allowing translation of tk-neo

fusion gene. The crossover events between donor and recipient would be recovered as well because the upstream donor has full length tk gene carrying intact promoter. Second, some NHEJ events may restore the reading frame to correctly translate neo gene. When the number of deleted nucleotides happens to be 3N+1 bp, the rejoined DNA sequence will resume the original reading frame, and the produced tk-neo fusion protein still render host cells G418 resistance even with polypeptide loss in tk region. Third, SSA may rejoin the broken DNA molecule through long stretch of homology shared by the two tandem repeats. These events cause fusion of the two tk sequences and loss of DNA sequence in-between. For substrate pLB4, SSA events cause 2.4 Kb deletion and generate a 3.9 Kb fusion fragment. Technically, the recombinant clones from SSA events are indistinguishable from crossover events, since they both fuse the donor and recipient.

Restriction map in Figure 5.1 demonstrates the expected bands from endonuclease digested pLB4, and changes brought by different recombination events. If digested by BamHI, the original substrate or gene conversion product will show a 3.9 Kb recipient fragment and a 4.5 Kb junction fragment containing donor and hygromycin resistance gene. The NHEJ product have the same 4.5 kb junction fragment as original pLB4, while the recipient fragment would be smaller than 3.9 kb due to sequence deletion (not obvious with short deletion). The crossover or SSA events would fuse aforementioned two fragments into a 3.9 Kb "recipient" fragment, the only visible band on the blot.

To unveil the details of recombination events, thymidine kinase gene sequences from two different strains of type 1 herpes simplex virus were placed in pLB4. Totally 13 nucleotide polymorphisms exist between the tk genes in donor and recipient (see Figure 5.2). These 13 nucleotides were numerated by their position along the tk sequence.

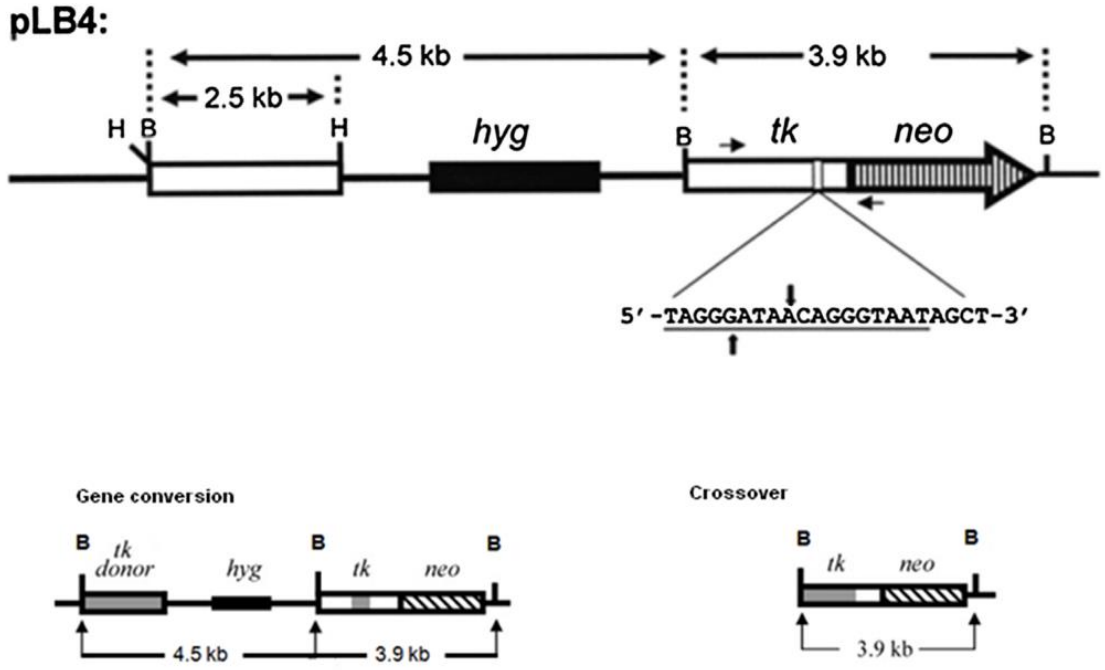

Figure 5.1 Restriction map of substrate pLB4 and possible recombinants. Recombination substrate pLB4 has two tandem repeats: the 3.9 Kb recipient of tk-neo fusion gene and the 2.5 Kb donor of tk gene. Recognition site of I-SceI was inserted into the tk sequence of the recipient. The 22 bp oligonucleotide insertion shifts the original reading frame and aborts the function of downstream neo gene. The restriction map of the gene conversion and crossover products is shown below. Adapted from Wang et al. 2016 (70).

There are 4 mismatched nucleotides upstream of the I-SceI site: the first 3 mismatches cluster together about 400 bp away from the I-SceI site while the last one located within 20 bp. The fifth nucleotide is 2 bp after the I-SceI site, and the rest 8 mismatches start 200 bp downstream of the I-SceI site, spanning around 400 bp. The nucleotide polymorphism within tk-neo recipient helps to delineate NHEJ and HR events, because HR may convert recipient's nucleotide to donor at these positions.

Materials and Methods:

Transient knockdown of RecQ4, BLM or both in human fibroblast cell line pLB4-11: Transfections of RecQ4 siRNA were carried out using Qiagen HiPerfect Transfection Reagent following the supplier's instructions. On the first day, 1.5 x10^5 cells were plated in 35 mm dish with 2.3 mL culture media (a-MEM supplied with 10% FBS). On the second day, 12 uL of 10 uM siRNA solution was mixed gently with 88 uL a-MEM in 1.5 mL Eppendorf tube, and then with 12 uL HiPerfect Transfection Reagent. The mixture was gently vortexed and kept at room temperature for 10 minutes to allow formation of transfection complex. Afterward, the siRNA mixture was added drop-wisely onto the cells, and the dish was gently swirled back and forth to ensure even distribution. On the third day, the dish was refilled with fresh media 2-3 hours before the second siRNA transfection. The transfection procedure was the same as performed a day before. The calculated siRNA concentration is always 50 nM in RecQ4 knockdown experiments.

The double knockdown experiments employed the same procedure as described for RecQ4 knockdown experiments, while the amount of siRNA was adjusted accordingly. Control siRNA, BLM siRNA and RecQ4 siRNA were added at a final concentration of 75 nM, 25 nM and 50 nM in corresponding experimental groups, while

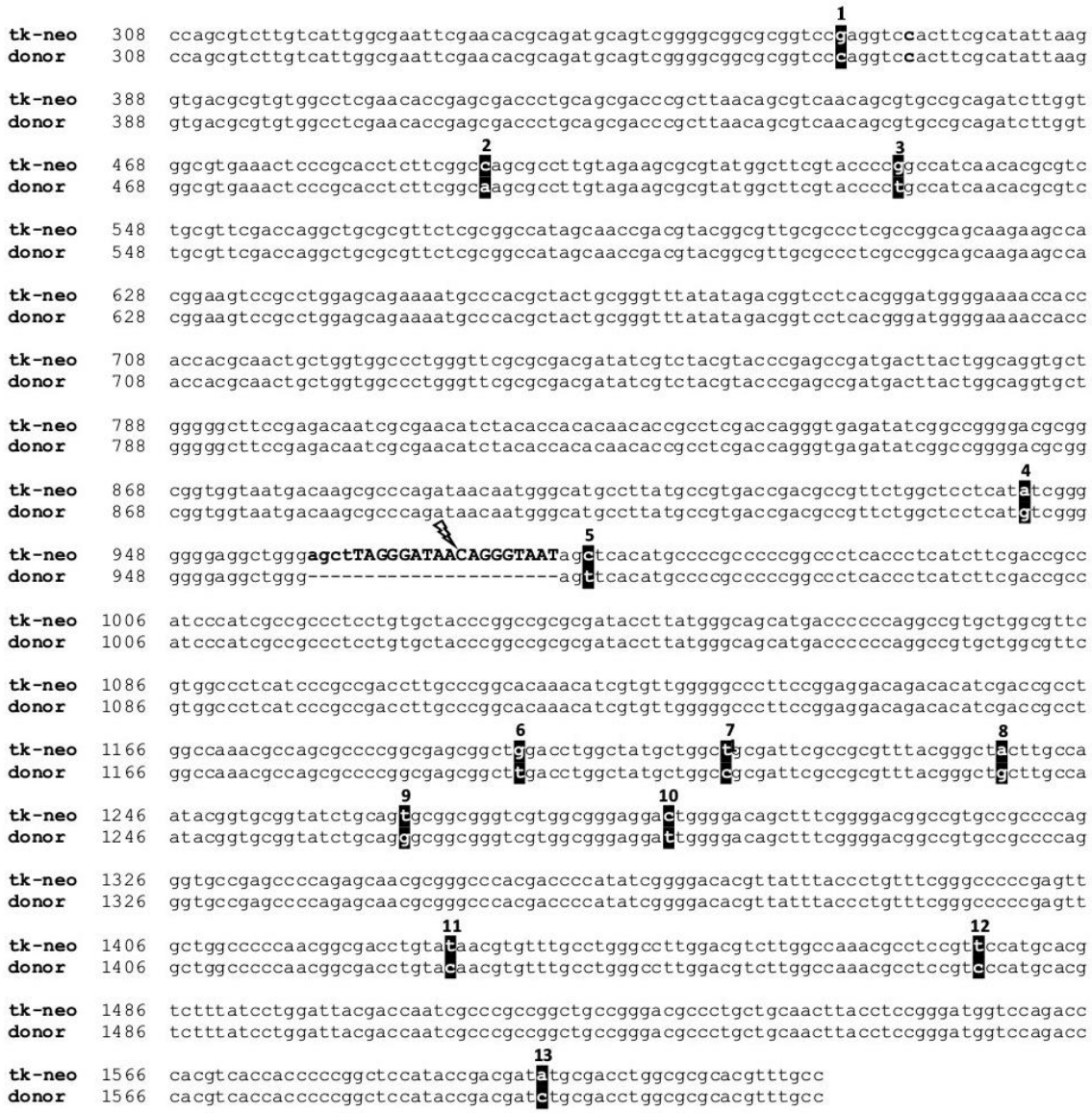

Figure 5.2 Mismatched nucleotides between donor and recipient in pLB4.
Totally 13 nucleotide mismatches were highlighted and numbered in the sequence alignment. Bold font was used to show the 22 bp insertion with I-SceI recognition site capitalized. Adapted from Wang et al. 2016 (70).

in BLM and RecQ4 double knockdown group, two siRNA were added at the same concentrations as in their own knockdown groups.

All mentioned siRNAs targeting RecQ4 or BLM with commercial name FlexiTube siRNA, were bought from Qiagen. RecQ siRNA#2 (Cat. No SI00061852) and RecQ siRNA#3 (Cat. No SI00061859) target RecQ4 transcript in human cells. BLM siRNA#3 (Cat. No SI00000952) targets BLM transcript in human cells. A negative control siRNA not targeting human transcripts was also purchased from Qiagen (Cat. No 1027310).

Recovery of DSB repair recombinant clones: To collect recombinant clones from DSB repair, pSce was electroporated into cells 24 hours after the second siRNA transfection. The cells were trypsinized from 35 mm dish, suspended in culture media and then counted on a hemocytometer. For each experimental group, 2.5 $x10^5$ cells were transferred into a 15 mL conical tube for electroporation while all the rest cells, approximately 3.5-4.5 $x10^5$ cells, were cultured in T25 for another day before lyzed to extract protein sample.

For transfection, the conical tubes containing 2.5 $x10^5$ cells were centrifuged at 300 X g for 3 minutes. After removing the supernatant, the pelleted cells were suspended in 3 mL phosphate buffered saline (PBS), and then centrifuged as described above. The newly pelleted cells were suspended in 300 uL PBS before electroporation. To start the electroporation, 2.4 uL DNA solution containing 7.5 ug pSce plasmid, and then the 300 uL cells suspension were added into a cuvette with 0.4 mm gap width. The DNA and cells were mixed together by pipetting 3 to 4 times. All electroporation experiments were carried out using Bio-Rad Gene Pulser set to 700 volts and 25 uF. Immediately after

electroporation, cells were transferred into a T25 with 5 mL culture medium, and they were incubated for two days before G418 selection.

The G418 selections were carried out at 1000 ug/mL from previous experience, and for an appropriate colony density, no more than 5 $x10^4$ cells were plated per T75 flask. Discrete colonies usually formed within 14 days and they were randomly picked and cultured for further analysis.

Preparation of protein sample: The procedure to exact soluble protein samples was the same as described in Chapter 4.

Protein electrophoresis and transferring: The procedure has been carried out the same as described in Chapter 3. The transferring condition for BLM was described in previous paragraph but the transferring was set at 450 mA for 90 minutes. The transferring buffer for RecQ4 does not contain 0.1% SDS and the transferring was set 100 volts for 90 minutes.

RecQ4 deficient lymphocyte cell lines AG18465 from Coriell Institute for Medical Research was used as negative control in western blotting, while BLM deficient human fibroblast cells GM08505 was obtained from NIGMS.

Western blot: The Western blot procedure was described in details in Chapter 4. The primary and secondary antibodies were both diluted at 1: 10, 000 to detect tubulin, while they were diluted at 1: 500 and 1: 25, 000 for RecQ4, or diluted at 1: 8, 000 and 25, 000 for BLM. As instructed by the supplier, the blocking and incubation of RecQ4 primary antibody were carried out in TBS solution with 5% BSA instead of 0.1% TBST

Calculation of clone frequency: After clone picking, all T75 flasks were stained and the remaining clones were counted. For each experimental group, the total clone number was divided by the cells number in G418 selection to obtain the clone frequency.

PCR amplification: The PCR procedures was described in Chapter 2. AW85 and AW133 were used to amplify tk-neo sequence including the 13 nucleotide mismatches between donor and recipient, so any nucleotide change within the area would be revealed by DNA sequencing.

Sequencing and alignment: The sequencing procedure was described in Chapter 3. Nucleotide sequences from these recombinant clones were aligned to recipient sequence and donor sequences for any nucleotide change using SciEdCentral.

Southern blot: The genomic DNA of these recombinant clones was digested by BamHI, and then separated on 0.8% agarose gel for Southern blot. The procedure was the same as described in Chapter 2.

Results:

Knockdown of RecQ4 in human fibroblast cell line pLB4-11: Successful knockdown of RecQ4 was achieved in pLB4-11 cell line after two consecutive siRNA transfections. Figure 5.3 shows the RecQ4 expression in the RecQ4 knockdown experiment 1. Cellular RecQ4 expression in Mock (only treated with transfection reagent but without siRNA) and control siRNA group is high while its expression is barely detectable in RecQ4 siRNA #2 group and RecQ4 siRNA #3 group.

Figure 5.4 demonstrated that RecQ4 knockdown was performed successfully in second experiment. Moreover, the lane 6 was loaded with 15 ug protein sample from RecQ4 deficient cells and 15 ug protein sample from mock to mimic 50% RecQ4

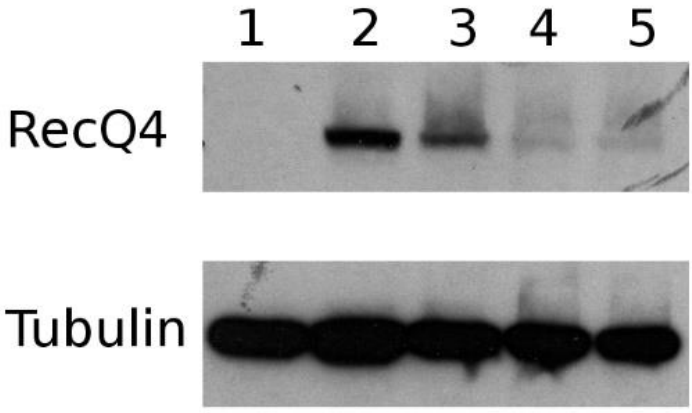

Figure 5.3 Knockdown of RecQ4 was achieved in RecQ4 knockdown experiment 1. Protein extract (30ug) from following cells are shown in western blot: Lane 1, RecQ4 deficient cells; lane 2, control siRNA group; lane 3, Mock; lane 4, RecQ4 siRNA #2 group; lane 5, RecQ4 siRNA #3 group.

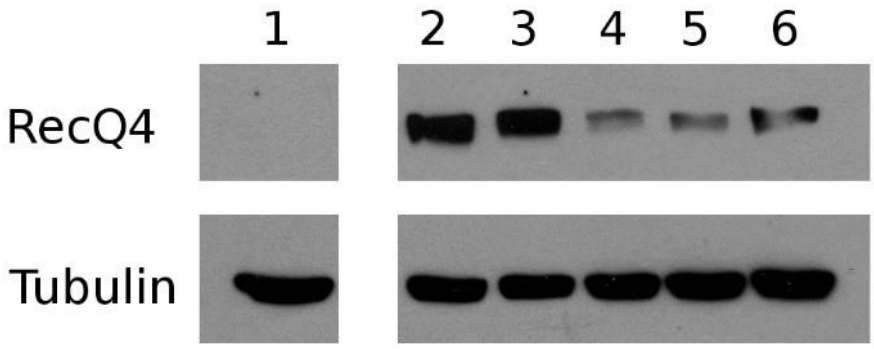

Figure 5.4 Knockdown of RecQ4 was achieved in RecQ4 knockdown experiment 2. Protein extract (30ug) from following cells are shown in western blot: Lane 1, RecQ4 deficient cells; lane 2, control siRNA group; lane 3, Mock; lane 4, RecQ4 siRNA #2 group; lane 5, RecQ4 siRNA #3 group; lane 6, mixture of 15ug protein sample from RecQ4 deficient cell and Mock.

expression. RecQ4 expression in both RecQ4 siRNA groups is obviously lower than the last lane, and it is estimated that 75% knockdown efficiency was achieved in both RecQ4 knockdown groups.

Knockdown of RecQ4, BLM or Both in human fibroblast cell line pLB4-11: Knockdown of RecQ4, BLM or both was achieved in pLB4-11 cell line before transient expression of I-SceI. In first double knockdown experiment, as shown in Figure 5.5, knockdown of RecQ4, BLM and both in pLB4-11 were achieved in corresponding experimental groups. BLM and RecQ4 expression in parental cell line, untreated group and control siRNA group are equivalent; expression of BLM is dramatically lower in BLM siRNA group or RecQ4 and BLM double siRNA groups; and minor RecQ4 expression is seen in RecQ4 siRNA groups or RecQ4 and BLM double siRNA groups. Tubulin served as internal control across these samples, and its image in Figure 5.5 came from the blot of RecQ4. Detection of Tubulin was not stable after protein transferring optimized for BLM (400 mA, 90 minutes with 0.1% SDS), however, stained acrylamide gel or stained nitrocellulose membrane after transferring showed that protein loadings were equal across samples (data not shown). RecQ4 and BLM expression in double knockdown experiment 2 are shown in Figure 5.6. BLM and RecQ4 expression were dramatically reduced in corresponding groups before induction of DNA DSB.

Collection of DSB-induced recombinant clones: To collect DSB-induced recombinant clones from these experimental groups, I-SceI was transiently expressed in cells through electroporation of pSce. After I-SceI introduced a DSB within the recipient, the broken DNA sequence may be repaired through HR, SSA or NHEJ events. If the

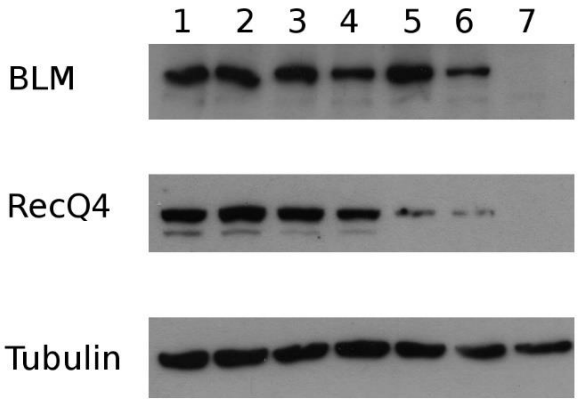

Figure 5.5 Knockdown of BLM and RecQ4 was confirmed in RecQ4 and BLM double knockdown experiment 1.
Protein extract (30ug) from the following cells are shown in Western blot: Lane 1, cell line pLB4-11; lane 2, untreated group; lane 3, control siRNA group; lane 4, BLM siRNA group; lane 5, RecQ4 siRNA group; lane 6, RecQ4 and BLM double siRNA group; lane 7, BLM deficient cells in top panel while RecQ4 deficient cells in middle and low panel The minor bands in RecQ4 panel are likely artifact due to intense signal of main bands.

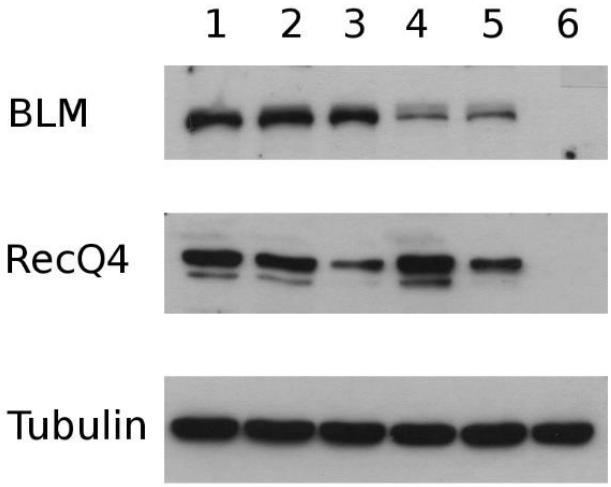

Figure 5.6 Knockdown of BLM and RecQ4 was confirmed in RecQ4 and BLM double knockdown experiment 2.
Protein extract (30ug) from the following cells are shown in western blot: Lane 1, untreated group; lane 2, control siRNA group; lane 3, RecQ4 siRNA group; lane 4, BLM siRNA group; lane 5, RecQ4 and BLM double siRNA group; lane 6, BLM deficient cells in top panel while RecQ4 deficient cells in middle and low panel. The minor bands in RecQ4 panel are likely artifact during film exposure due to intense signal.

DNA repair restores the correct reading frame for the tk-neo gene in recipient, the host cell will become G418 resistant and form recombinant clone.

DSB-induced recombinant clones from different experimental groups were recovered from two times of RecQ4 knockdown experiment and two times of RecQ 4 and BLM knockdown experiments. The clone frequencies and clones analyzed for each experimental group in each experiment are all listed in Table 5.1. Clone frequencies observed in these experimental groups range from $1.74X10^{-3}$ to $1.10X10^{-2}$. In each experiment, the clone frequencies across the experimental groups are close though cells may have different genetic background. The clone frequencies after RecQ4 knockdown are not different from those in control groups, with a p-value equals 0.413 from student's test, while BLM knockdown brings slightly higher clone frequencies than the rest groups.

Categorization of DSB repair recombinant clones: Recombinant clones from DSB repair are categorized based on their recipient sequences. If the recipients are restored to continuous HSV-1 TK sequence, losing the 22 bp oligonucleotide insertion, these events are classified as HR. HR events also convert some of the 13 nucleotide mismatches to donor, and left a conversion tract in recipient for each HR event. The HR events are further separated into gene conversion and crossover events: crossover results in the fusion of donor and recipient, in addition to aforementioned 22 bp deletion and conversion of nucleotide.

If the recipient lost a stretch of nucleotide around the break site after end rejoining, they are categorized as NHEJ events. NHEJ events do not have any nucleotide mismatches converted to donor, and almost all NHEJ events have sequence deletions other than the loss of 22 bp oligonucleotide in HR. Occasionally, there might be NHEJ

Table 5.1 Clone frequencies and analyzed clones in knockdown experiments.

Experiment	Expt #	Experimental group	Clone frequency (10^{-3})	Colonies analyzed
RecQ4 knockdown	1	Control siRNA	1.74	58
		Mock	3.73	55
		RecQ4 siRNA #2	4.38	58
		RecQ4 siRNA #3	2.76	51
	2	Control siRNA	4.90	28
		Mock	6.00	25
		RecQ4 siRNA #2	4.38	27
		RecQ4 siRNA #3	4.76	27
RecQ4 and BLM double knockdown	1	Untreated	6.19	28
		Control siRNA	4.15	26
		BLM siRNA	6.60	27
		RecQ4 siRNA	4.30	30
		RecQ4&BLM siRNA	5.11	25
	2	Untreated	8.38	16
		Control siRNA	7.89	14
		BLM siRNA	10.97	19
		RecQ4 siRNA	6.55	19
		RecQ4&BLM siRNA	9.54	18

The clone frequencies were calculated as described in Materials and Methods.

events precisely removing the 22 bp oligonucleotide insertion without conversion of mismatches 5, which should be always converted in HR events due to its extremely closeness to the DSB site.

There are also complicated events which could not be simply assigned to either HR or NHEJ. Some sporadic events were found to be combinations of HR and NHEJ repair. They converted nucleotide mismatches to donor, and they also lost oligonucleotide other than the 22 bp engineered insertion.

Changes in DSB repair recombinant clones after RecQ4 knockdown: As shown on top in Table 5.2, the majority of DSB repair recombinant clones in RecQ4 knockdown experiments are product of HR events, with nucleotide mismatches converted to donor. NHEJ events compose about one third of all DSB repair events, either in control cells or in RecQ4 knockdown cells: there are totally 118 HR events and 41 NHEJ events recovered in control cells, while 102 HR events and 45 NHEJ events in RecQ4 knockdown cells.

The proportion of gene conversion in HR events changed after RecQ4 knockdown. With reduced RecQ4 expression, more recombinant clones of gene conversion event appeared as the result. Polling data from two RecQ4 knockdown experiments, the control groups have 53 gene conversion recombinant clones and 65 crossover recombinant clones, while RecQ4 knockdown cells have 60 gene conversion recombinant clones and 43 crossover recombinant clones. After RecQ4 knockdown, the increase of gene conversion events is substantial, however, the Fisher's exact test comparing the ratio between HR subtypes in control and RecQ4 knockdown cells gives

Table 5.2 DSB repair recombinants in knockdown experiments.

Experiment	Expt #	Experimental group	Clone analyzed	GC	CO	NHEJ	COMPLEX
RecQ4 knockdown	1	Control siRNA	58	19	25	11	3
		Mock	55	14	19	19	3
		RecQ4 siRNA #2	58	19	15	16	8
		RecQ4 siRNA #3	51	20	16	11	4
	2	Control siRNA	28	10	8	9	1
		Mock	25	10	13	2	0
		RecQ4 siRNA #2	27	9	6	9	3
		RecQ4 siRNA #3	27	12	6	9	0
RecQ4 and BLM knockdown	1	Untreated	28	8	12	6	2
		Control siRNA	26	14	8	3	1
		BLM siRNA	27	7	15	4	1
		RecQ4 siRNA	30	13	10	7	0
		RecQ4&BLM siRNA	25	12	9	3	1
	2	Untreated	16	6	7	2	1
		Control siRNA	14	7	4	2	1
		BLM siRNA	19	5	14	0	0
		RecQ4 siRNA	19	5	9	4	1
		RecQ4&BLM siRNA	18	4	9	5	0

GC: gene conversion; CO: crossover; NHEJ: non-homologous end joining
Recombinant clones were grouped into NHEJ, HR or complex events based on their recipient sequence as described before. If the events are combinations of NHEJ and HR, or have other rare alterations of sequence, they are grouped as complex events.

p=0.0590, which states the difference is probably close to, however, not statistically significant.

Changes in DSB repair recombinant clones after BLM knockdown: in double knockdown experiments, DSB repair recombinant clones were collected from untreated, control, BLM knockdown, RecQ4 knockdown or BLM and RecQ4 double knockdown cells. In BLM knockdown groups, vast majority of HR events are crossover. Fisher exact test comparing HR subtypes before or after BLM knockdown gives p=0.0177, while the same test comparing HR subtypes before or after BLM and RecQ4 double knockdown gives p=0.6867.

Conversion tracts in gene conversion events: During gene conversion, all or part of the 13 nucleotide mismatches were transferred from donor to recipient. Table 5.3 and 6.4 lists the transferring of donor's nucleotides to recipient in these gene conversion events. Short conversion tracts dominate in all experimental groups, and most of them just converted mismatches 4 and 5, these two closest to I-SceI site. Some short conversion tracts only converted mismatches 5. All short conversion tracts described above are shorter than 600 bp (distance between mismatches 3 and 6). For each experimental group, recombinant clones with the same conversion tract are listed in one cell at the end of the chart.

The gene conversion tracts in RecQ4 and BLM double knockdown experiments are listed in Table 5.4 and they were labeled the same way as in Table 5.3.

Donor's mismatched nucleotides left in fusion sequence after crossover events: In crossover events, fusion sequences always possess donor's nucleotides before the I-SceI, and beyond that, they may have more donor's nucleotides until crossover

Table 5.3 Gene conversion tracts in RecQ4 knockdown experiments.

Expt. #	Experimental group	Converted nucleotides in GC events	Clone number	Clone name
1	Control siRNA	4, 5	18	RC1J, RC1K, RC1O, RC1R, RC1S, RC1T, RC1U, RC1BA, RC1BN, RC1BP, RC1BQ, RC1BR, RC1BS, RC1BW, RC1CA, RC1CC, RC1CD, RC1CI
		4 to 11	1	RC1M
	Mock	5	2	RM1BG, RM1BY
		4, 5	12	RM1D, RM1F, RM1I, RM1J, RM1O, RM1Q, RM1BC, RM1BD, RM1BL, RM1BQ, RM1BV, RM1CI
	RecQ4 siRNA #2	5	2	RR21T, RR21CD
		4, 5	15	RR21E, RR21F, RR21G, RR21Q, RR21U, RR21W, RR21BA, RR21BC, RR21BE, RR21BT, RR21BV, RR21BW, RR21BX, RR21CA, RR21CE
		1 to 5	1	RR21BI
		1 to 12	1	RR21O
	RecQ4 siRNA #3	5	2	RR31V, RR31BA
		4, 5	14	RR31E, RR31I, RR31L, RR31N, RR31R, RR31T, RR31U, RR31W, RR31BE, RR31BG, RR31BM, RR31BT, RR31CA, RR31H
		1 to 5	2	RR31Q, RR31BY
		4 to 8	1	RR31A
		1 to 12	1	RR31M
2	Control siRNA	4, 5	10	RC3W, RC3U, RC3T, RC3O, RC3J, RC3I, RC3H, RC3BD, RC3B, RC3A
	Mock	5	2	RM3M, RM3BC
		4, 5	6	RM3X, RM3Z, RM3V, RM3S, RM3J, RM3E
		4 to 11	1	RM3D
		4 to 12	1	RM3H
	RecQ4 siRNA #2	5	3	R23Z, R23J, R23BC
		4, 5	5	R23S, R23P, R23L, R23G, R23C
		1 to 5	1	R23V
	RecQ4 siRNA #3	5	2	R33Q, R33C
		4, 5	9	R33N, R33M, R33L, R33K, R33J, R33I, R33G, R33E, R33D
		1 to 5	1	R33T

Table 5.4 Gene conversion tracts of DSB repair recombinants in RecQ4 and BLM double knockdown experiments.

Expt #	Experimental group	Converted nucleotides in GC events	Clone number	Clone name
1	Untreated	4, 5	8	RBU1Y, RBU1R, RBU1Q, RBU1O, RBU1N, RBU1F, RBU1D, RBU1A
	Control siRNA	5	1	RBC1O
		4, 5	11	RBC1C, RBC1D, RBC1H, RBC1I, RBC1N, RBC1P, RBC1S, RBC1T, RBC1X, RBC1Y, RBC1BD
		4 to 10	2	RBC1BB, RBC1E
	BLM siRNA	4 5	4	RBB1K, RBB1L, RBB1P, RBB1Z,
		1 to 10	1	RBB1X
		1 to 11	1	RBB1I
		4 to 12	1	RBB1G
	RecQ4 siRNA	4, 5	11	RBR1BB, RBR1BC, RBR1BD, RBR1E, RBR1J, RBR1K, RBR1L, RBR1Q, RBR1R, RBR1S, RBR1X
		1 to 5	1	RBR1O
		4 to 12	1	RBR1W
	BLM&RecQ4 siRNA	5	3	RBRB1B, RBRB1BC, RBRB1D
		4, 5	8	RBRB1A, RBRB1BD, RBRB1H, RBRB1I, RBRB1O, RBRB1R, RBRB1S, RBRB1Y
		4 to 12	1	RBRB1C
2	Untreated	4, 5	6	RBU3B, RBU3F, RBU3H, RBU3L, RBU3P, RBU3T
	Control siRNA	4, 5	6	RBC3B, RBC3F, RBC3K, RBC3M, RBC3N, RBC3T
		4 to 7	1	RBC3C
	BLM siRNA	5	1	RBB3K,
		4, 5	4	RBB3F, RBB3G, RBB3O, RBB3Q
	RecQ4 siRNA	5	1	RBR3J,
		4, 5	3	RBR3C, RBR3I, RBR3N
		3 to 10	1	RBR3G
	BLM&RecQ4 siRNA	5	2	RBRB3I, RBRB3Q
		4, 5	2	RBRB3L, RBRB3R

occurs. The possession of donor's mismatched nucleotides in crossover events is listed in Table 5.5 and Table 5.6. The crossover events are categorized into multiple groups with the same amount of donor's mismatched nucleotide: for example, crossover events that have donor's nucleotide 1 to nucleotide 5 (right after I-SceI), and other crossover events that have specific amount of donor's nucleotides. The possession of donor's mismatched nucleotides in crossover events are not that different across experimental groups, and no obvious correlation was found between the number of donor nucleotides and the genetic background.

NHEJ events: The NHEJ events and their clone information are listed in Table 5.7 and Table 5.8. For each clone, deletion size and potential usage of microhomology are described. The sequence deletions in NHEJ events range from 1 bp to 1018 bp, and they are always 3N+1 bp in length, which restored the original reading frame for the tk-neo fusion gene. Frequently, microhomology was utilized to rejoin the broken DNA ends during NHEJ: if the broken ends of DNA go through end resection, short complementary DNA sequences could be revealed on both side and they are able to anneal for end rejoining. NHEJ events in pLB4-11 cells are not different from that in RecQ4 knockdown conditions: they are minority in DSB repair events. NHEJ events in pLB4-11 cells tends to be lower after BLM knockdown, however, it is not statistically significant compared with that in control cells.

Discussion:

Types of DSB-induced recombination: NHEJ and HR are two prominent pathways to repair DNA double-strand breaks in mammalian cells (30, 32, 34). In agreement with previous reports, these two types of events comprise the vast majority of

Table 5.5 Position of donor's mismatched nucleotides in crossover events (RecQ4 knockdown experiments).

Expt #	Group	Donor's nucleotide	Clone number	Clone name
1	Control siRNA	1 to 5	13	RC1V, RC1Y, RC1BC, RC1BG, RC1BH, RC1BL, RC1BM, RC1BT, RC1BU, RC1BZ, RC1CB, RC1CF, RC1CH
		1 to 8	3	RC1BF, RC1BJ, RC1BK
		1 to 9	1	RC1BX
		1 to 10	1	RC1P
		1 to 11	2	RC1E, RC1BD
		1 to 12	3	RC1X, RC1CE, RC1Z
		1 to 13	2	RC1N, RC1BO
	Mock	1 to 5	3	RM1V, RM1Y, RM1CE
		1 to 8	1	RM1CD
		1 to 8	1	RM1BP
		1 to 9	2	RM1CC, RM1BN
		1 to 10	5	RM1E, RM1G, RM1S, RM1BF, RM1BX
		1 to 11	1	RM1W
		1 to 12	4	RM1X, RM1Z, RM1BM, RM1BT
		1 to 13	2	RM1BI, RM1CB
	RecQ4 siRNA #2	1 to 5	5	RR21V, RR21BH, RR21BR, RR21BS, RR21BY
		1 to 8	1	RR21R
		1 to 9	1	RR21BO
		1 to 10	3	RR21J, RR21BK, RR21H
		1 to 11	1	RR21L
		1 to 12	3	RR21N, RR21BG, RR21BQ
		1 to 13	1	RR21P
	RecQ4 siRNA #3	1 to 5	5	RR31J, RR31Y, RR31BR, RR31BS, RR31CH
		1 to 7	1	RR31BZ
		1 to 9	3	RR31B, RR31G, RR31BN
		1 to 10	3	RR31P, RR31BH, RR31BP
		1 to 12	3	RR31CD, RR31CI, RR31BO
		1 to 13	1	RR31BU

Table 5.5 continued

Expt #	Group	Donor's nucleotide	Clone number	Clone name
2	Control	1 to 5	3	RC3Y, RC3N, RC3E
	siRNA	1 to 8	1	RC3C
		1 to 10	1	RC3P
		1 to 11	1	RC3D
		1 to 12	1	RC3S
		1 to 13	1	RC3BC
	Mock	1 to 5	3	RM3Q, RM3G, RM3B
		1 to 8	2	RM3P, RM3O
		1 to 9	1	RM3W
		1 to 10	1	RM3U
		1 to 11	2	RM3I, RM3A
		1 to 12	4	RM3BA, RM3Y, RM3N, RM3BB
	RecQ4	1 to 8	1	R23H
	siRNA #2	1 to 10	1	R23N
		1 to 12	2	R23Q, R23BD
		1 to 13	2	R23F, R23A
	RecQ4	1 to 5	1	R33Z
	siRNA #3	1 to 8	1	RR31O
		1 to 9	2	R33B, R33H,
		1 to 12	2	R33S, R33P

Table 5.6 Position of donor's mismatched nucleotides in crossover events (RecQ4&BLM knockdown experiments).

Expt #	Group	Donor's nucleotide	Clone number	Clone
1	Untreated	1 to 5	3	RBU1Z, RBU1V, RBU1K
		1 to 9	1	RBU1B
		1 to 10	2	RBU1J, RBU1X
		1 to 11	1	RBU1T
		1 to 12	4	RBU1BD, RBU1M, RBU1G, RBU1BC,
		1 to 13	1	RBU1W
	Control siRNA	1 to 5	1	RBC1A
		1 to 6	1	RBC1Q
		1 to 9	1	RBC1B
		1 to 10	3	RBC1F, RBC1G, RBC1K,
		1 to 12	2	RBC1V, RBC1BC
	BLM siRNA	1 to 5	3	RBB1BB, RBB1O, RBB1U
		1 to 9	2	RBB1S, RBB1BD
		1 to 10	4	RBB1E, RBB1Q, RBB1R, RBB1A
		1 to 11	1	RBB1B
		1 to 12	3	RBB1J, RBB1M, RBB1C
		1 to 13	2	RBB1BC, RBB1D
	RecQ4 siRNA	1 to 5	6	RBR1P, RBR1A, RBR1C, RBR1G, RBR1N, RBR1U
		1 to 8	1	RBR1Y
		1 to 9	2	RBR1BA, RBR1M
		1 to 12	1	RBR1T
	BLM&RecQ4 siRNA	1 to 5	1	RBRB1V
		1 to 8	1	RBRB1T
		1 to 10	2	RBRB1L, RBRB1U
		1 to 11	1	RBRB1F
		1 to 12	4	RBRB1X, RBRB1G, RBRB1J, RBRB1Q

Table 5.6 continued

Expt #	Group	Donor's nucleotide	Clone number	Clone
2	Untreated	1 to 5	3	RBU3A, RBU3K, RBU3M
		1 to 10	1	RBU3S
		1 to 12	2	RBU3G, RBU3J
		1 to 13	1	RBU3N
	Control siRNA	1 to 5	1	RBC3A,
		1 to 9	1	RBC3I
		1 to 10	1	RBC3D
		1 to 12	1	RBC3O
	BLM siRNA	1 to 5	2	RBB3E, RBB3N
		1 to 7	1	RBB3C
		1 to 8	2	RBB3D, RBB3P
		1 to 9	1	RBB3T
		1 to 10	3	RBB3I, RBB3B, RBB3H
		1 to 11	1	RBB3L
		1 to 12	4	RBB3A, RBB3M, RBB3R, RBB3S
	RecQ4 siRNA	1 to 5	2	RBR3A, RBR3Q
		1 to 8	2	RBR3F, RBR3K
		1 to 9	1	RBR3R
		1 to 10	2	RBR3B, RBR3L
		1 to 12	1	RBR3O
		1 to 13	1	RBR3S
	BLM&RecQ4 siRNA	1 to 5	6	RBRB3F, RBRB3G, RBRB3H, RBRB3M, RBRB3N, RBRB3S,
		1 to 10	1	RBRB3E
		1 to 12	2	RBRB3J, RBRB3O

Table 5.7 NHEJ events and clone information in RecQ4 knockdown experiments.

Expt #	Group	Deletion size (bp)	Microhomology	Clone name
1	Control siRNA	10	GG	RC1I
		22	AGCT	RC1A
		22	AGCT	RC1G
		22	A	RC1Q
		28	GGG	RC1W
		133	CAGGGT	RC1CG
		169	GCCGT	RC1L
		241	GC	RC1BE
		256	G	RC1H
		295	TA	RC1BI
	Mock	1	A	RM1T
		1	A	RM1CA
		4	A	RM1CF
		7	A	RM1BS
		7	G	RM1BZ
		10	GG	RM1C
		10	A	RM1BK
		22	0	RM1H
		22	AGCT	RM1R
		22	AGCT	RM1BH
		22	AGCT	RM1BJ
		43	GG	RM1BW
		118	AGC	RM1K
		145	AACA	RM1L
		148	ACA	RM1A
		307	C	RM1BU
		343	GCCC	RM1BR
		685	ACG	RM1BA
		1018	C	RM1CH

Table 5.7 continued

Expt #	Group	Deletion size (bp)	Microhomology	Clone name
1	RecQ4 siRNA #2	1	A	RR21A
		1	A	RR21B
		1	G	RR21BJ
		1	A	RR21CC
		4	0	RR21CB
		10	GG	RR21M
		10	GG	RR21S
		22	AGCT	RR21C
		22	AGCT	RR21I
		22	AGCT	RR21BD
		22	AGCT	RR21CF
		25	G	RR21BM
		133	A	RR21CG
		157	C	RR21BZ
		244	GC	RR21D
		427	G	RR21Z
	RecQ4 siRNA #3	1	A	RR31Z
		1	0	RR31BF
		1	A	RR31BQ
		1	G	RR31CG
		4	0	RR31O
		7	G	RR31F
		22	AGCT	RR31D
		22	AGCT	RR31S
		22	AGCT	RR31BD
		22	AGCT	RR31BL
		193	CAGG	RR31BW

Table 5.7 continued

Expt #	Group	Deletion size (bp)	Microhomology	Clone name
2	Control siRNA	4	0	RC3X
		4	0	RC3BB
		10	GG	RC3Q
		10	GG	RC3M
		22	AGCT	RC3Z
		76	GGG	RC3V
		145	G	RC3F
		271	GAT	RC3L
		400	ATAGC	RC3BA
	Mock	19	GG	RM3BD
		286	ATCG	RM3R
	RecQ4 siRNA #2	1	A	R23T
		4	0	R23D
		4	0	R23BB
		22	AGCT	R23E
		28	GGG	R23U
		37	GC	R23B
		64	GC	R23I
		389	GCCG	R23O
		607	GGGAT	R23Y
	RecQ4 siRNA #3	7	A	R33W
		10	GG	R33U
		16	G	R33BB
		22	AGG	R33BC
		133	CAGGGT	R33Y
		166	0	R33X
		166	GA	R33V
		337	C	R33BD
		430	G	R33F

Table 5.8 NHEJ events and their clone information in RecQ4 and BLM double knockdown experiments.

Expt #	Groups	Deletion size (bp)	Microhomology	Clone name
1	Untreated	1	A	RBU1C
		7	A	RBU1S
		19	A	RBU1U
		22	AGCT	RBU1L
		25	G	RBU1P
		25	G	RBU1H
	Control siRNA	7	A	RBC1L
		19	GG	RBC1M
		22	0	RBC1U
	BLM siRNA	1	G	RBB1W
		8	TAA	RBB1Y
		10	GG	RBB1BA
		118	0	RBB1V
	RecQ4 siRNA	1	A	RBR1F
		22	AGCT	RBR1B
		22	AGCT	RBR1I
		76	GGG	RBR1D
		133	A	RBR1V
		133	CAGGGT	RBR1Z
		310	AC	RBR1H
	BLM&RecQ4 siRNA	1	G	RBRB1M
		1	G	RBRB1N
		22	A	RBRB1BA
2	Untreated	1	A	RBU3O
		133	CAGGGT	RBU3Q
	Control siRNA	16	0	RBC3J
		268	ATC	RBC3E
	BLM siRNA	NA	NA	NA
	RecQ4 siRNA	1	G	RBR3E
		4	0	RBR3H
		10	GG	RBR3D
		445	CAT	RBR3M
	BLM&RecQ4 siRNA	22	AGCT	RBRB3A
		22	A	RBRB3B
		22	AGCT	RBRB3T
		394	TG	RBRB3D
		430	TG	RBRB3K

recovered DSB repair recombinant clones. As mentioned in introduction, SSA could produce recombinant during DSB repair, which is indistinguishable from that of crossover events.

In pLB4-11 cell line, NHEJ events survived G418 selection as long as they rejoined the broken recipient sequence with 3N+1 bp deletion, which restore the original reading frame for tk-neo fusion gene. According to the data presented here, NHEJ events are only a small proportion of the recovered recombinant clones, the actual NHEJ events are most likely 3 times higher. This estimation suggests that NHEJ events induced by the break are roughly as many as HR events. NHEJ events do not convert the aforementioned 13 nucleotide mismatches, and the homologous donor does not serve the "template" role in NHEJ events.

Different from NHEJ events, HR events restore the disrupted recipient back to a continuous HSV-1 TK sequence through removing the 22 bp engineered insertion. Moreover, donor's unique nucleotides would be transferred into recipient sequence: the mismatched nucleotide 5, 3 bp downstream the I-SceI, was always converted while the rest mismatched nucleotides were converted to donor as the conversion tract move further. The HR events were further divided into gene conversion events and crossover events. Crossover events had donor sequence spliced to recipient sequence, causing 4.5 Kb fragment deletion between these two tandem repeats.

Crossover events fusing tandem repeats: Aforementioned crossover events could be achieved by SSA: if the broken ends went through extensive unwinding and resection, the donor strand would eventually have chance to anneal with recipient strand. After flaps cleavage and ligation, the rejoined DNA molecule would have a fusion of

donor and recipient, the same as crossover events generated by HR. It is unknown whether recovered fusion events arise primarily from crossover or SSA, however, we found an increase in crossover events after BLM knockdown, which agree well with BLM's roles in Holliday junction dissolution or D-loop rejection to prevent crossover, but contradict with BLM's roles in long-range resection (117) and strand annealing to promote SSA (118).

Clone frequency not affected by RecQ4 knockdown: The DSB repair clone frequencies are equivalent in control and RecQ4 knockdown groups in each experiment, while variations in clone frequency were observed across experiments. In BLM knockdown experiments, no obvious change in clone frequency was observed after BLM knockdown or BLM and RecQ4 double knockdown.

No dramatic change in NHEJ events after RecQ4 knockdown: NHEJ events comprise a small portion of recovered events, and their proportions in RecQ4 knockdown groups are not dramatically different from that in control groups. They have sequence deletion from several nucleotides to thousands of nucleotides, and the deletion sizes are similar across experimental groups. Most NHEJ events rejoin the broken ends with potential use of microhomology, while only 13 out of 122 NHEJ events rejoined the broken ends without using short complementary ends.

Microhomology used in these events is generally short, from 1 to 6 nucleotides. About half of the NHEJ events used microhomology with 2 or more nucleotides, about one third used microhomology with 1 nucleotide, while the rest one sixth did not use microhomology. No obvious difference in utilizing microhomology was observed across experimental groups, and if any difference did exist, it does not correlate with genetic

background of examined cells. Though knockdown of RecQ4 or BLM does not affect NHEJ events, current data confirmed that NHEJ prefer to use terminal microhomology to ligate two broken ends.

Ratio of gene conversion to crossover changed after RecQ4 knockdown: In control groups, more crossover events were recovered than gene conversion events after DSB repair (65 vs 53), while this trend was reversed after RecQ4 knockdown (43 vs 60). The substantial difference between control cells and knockdown cells suggests the HR outcome shift from crossover to gene conversion after RecQ4 knockdown. However, the fisher exact test comparing these two HR subtypes between control and knockdown groups does not support significant difference between these two conditions.

RecQ4 knockdown alleviate the effect of BLM deficiency on HR: In RecQ4 and BLM double knockdown experiments, the DSB repair in BLM knockdown cells were examined as well as in RecQ4 and BLM double knockdown cells. BLM knockdown increased the proportion of crossover events, supported by fisher exact test comparing HR subtypes between BLM knockdown groups and control groups ($p=0.0499$). The same observation has been reported in our recent publication (70). The ratio of HR subtypes between other groups are not dramatically different. In RecQ4 and BLM double knockdown cells, though RecQ4 deficiency failed to increase the gene conversion events, it does abrogate potential increase of crossover event in BLM deficient cells. Since RecQ4 depletion alleviates effect from BLM deficiency, RecQ4 likely acts against BLM in HJ dissolution or D-loop rejection steps, which need further investigation.

Resolution site of crossover events is away from DSB: Most recovered gene conversion events have mismatches 4 and 5 converted to donor, with maximal 700 bp

conversion tract surrounding the break site. Some gene conversion events only converted mismatch 5, which means maximal 200 bp conversion tract mostly downstream of the break site.

In contrast, crossover events have more donor's unique nucleotide downstream of the I-SceI insertion site in their recombinant DNA sequences. Polling data from control cells, 50 out 54 gene conversion events in control cells have conversion tract ended between mismatches 5 and 6, only gaining one donor's unique nucleotide after break site. However, in 43 out of 62 crossover events, their recombinant DNA sequences have two or more donor's unique nucleotides after I-SceI site. Fisher exact test comparing the resolution sites in gene conversion events and crossover events comes out statistically significant ($p=0.0001$). The results suggest that the crossover events may went through extensive strand exchange and DNA synthesis, so the area of genetic exchange tends to be dramatically longer than that in gene conversion events.

Complex recombination events encountered in current experiments:

Sporadically, there are complex recombination events which cannot be simply grouped into either HR or NHEJ events. There are totally 30 such events out of 551 examined events across experiments and they are generally listed in Table 5.9.

Many complex recombinants have additional modifications on their DNA sequences besides the repair at the DSB site. They might be a combination of NHEJ and HR event, a combination of NHEJ and a secondary deletion event, or a combination of HR and a secondary unknown sequence changing event. There are also other gross rearrangement events that could not be identified through sequencing and southern blotting. All these events demonstrated that during DSB repair, cells probably made multiple attempts to fix

the lesion and adopted more than one pathway to save them from stalled cell cycle and cell death.

Concluding Remarks:

HR protects genome stability by fixing severe DNA damage affecting both strands of a DNA molecule, while occasionally it brings detrimental mutation, deletion or duplication after recombining diverged DNA sequences. In the current dissertation, recombination between diverged sequences has been examined using a new set of recombination substrates, which have different lengths of homeologous sequences between donor and recipient surrounded by perfect homology. The frequency and the fraction of HeR events were compared among these substrates to reveal how the length of homeologous sequences affects the HeR events between donor and recipient, or conversely, how cells process these homeologous sequences during DNA recombination.

The obtained results clearly demonstrate that when aided by surrounding homology, HeR events occur at significantly higher frequencies than that between purely homeologous sequences, and they are common regardless of the length of homeologous sequences. Additionally, recombinants' DNA sequences after MSH2 knockdown suggested hDNA as the recombination intermediate in these HeR recombinants of DSB repair. Moreover, MMR depletion brought a minor boosting effect on HeR frequency in cell line pBWW33-67, in contrast with its dramatic boost on recombination between diverged sequences. Available data presented evidence supporting that cells carry out recombination ignoring the mismatched nucleotides between donor and recipient as long as enough homology is available on both side of the homeologous sequence to mediate the recombination.

Table 5.9 Complex recombination events collected in all experiments.

Clone Number	1st event	Event Details	2nd event
RBU1I	NHEJ	37 bp deletion with no microhomology	CO with conversion tract to 3
RR21BF	NHEJ	993 bp deletion with microhomology GCG	CO at location beyond 13 mismatched nucleotides
RC1D	NHEJ	193 bp deletion with microhomology CAGG	GC with conversion tract to 3
RC1B	NHEJ	22 bp deletion with microhomology AGG	GC with conversion tract to 4
RR31CC	NHEJ	10 bp deletion with no microhomology	a string of G to C mutation near NHEJ
RR21BL	NHEJ	8 bp deletion with microhomology TAA	467 bp deletion with microhomology AC
RC1C	NHEJ	8 bp deletion with microhomology TAA	13 bp insertion near NHEJ
RR31BX	NHEJ	23 bp deletion with no microhomology	DEL&INS with 209 bp deletion and 207 bp insertion near NHEJ, 0.9 kb deletion beyond 13 mismatched nucleotides in recipient.
RR31BI	CO	with donor's nucleotide from 1 to 11	Possibly 0.3 kb insertion in recipient beyond 13 mismatched nucleotides
RR31BK	CO	with donor's nucleotide from 1 to 12	unknown sequence alteration at a second location
RR21CH	CO	with donor's nucleotide from 1 to 12	Possibly 0.4 kb deletion in recipient beyond 13 mismatched nucleotides
RM1U	CO	with donor's nucleotide from 1 to 5	Possibly sequence alteration affecting the terminal enzyme sites
RM1BO	CO	with donor's nucleotide from 1 to 11	Possibly 0.1 kb insertion in recipient beyond 13 mismatched nucleotides
RBB1F	DEL&INS	24 bp deletion with 5 bp insertion	CO with conversion tract to 4 at an upstream location
RR21BU	DEL&INS	602 bp deletion with 266 bp insertion	Possibly CO at location beyond 13 mismatched nucleotides
R23W	NHEJ	18 bp deletion	1 bp second deletion between mismatched nucleotides 4 and 5
RBC3G	INS	2 bp insertion	NA
RC3R	INS	137 bp insertion	NA
RR21Y	INS	11 bp insertion	NA
RBR3T	DEL&INS	2 bp deletion and 1 bp insertion	NA

Table 5.9 continued

Clone Number	1st event	Event Details	2nd event
RBU3I	DEL&INS	21 bp deletion and 2 bp insertion	NA
RBC1Z	DEL&INS	11 bp deletion and 13 bp insertion	NA
R23R	DEL&INS	8 bp deletion and 1 bp insertion	NA
RR21K	DEL&INS	13 bp deletion and 6 bp insertion	NA
RR21CI	DEL&INS	41 bp deletion and 4 bp insertion	NA
RBRB1K	UNKNOWN	gross rearrangement	NA
RBU1E	UNKNOWN	gross rearrangement affecting recipient	NA
RR21BN	UNKNOWN	gross rearrangement affecting recipient	NA
RM1B	UNKNOWN	gross rearrangement affecting recipient	NA
R23K	DEL0	0 bp deletion in recipient	NA

Existing data indicates that MMR vigorously inhibits recombination between homeologous sequences and that MSH2-MSH6 interacts with hDNA within D-loop *in vitro*. Data in this dissertation strongly suggested that MMR helps to terminate HeR events at initiation step, while once invasion and synthesis established along the DNA template, the strand exchange and recombination between nearby homeologous sequence seem not a problem to the cells. This further sketch of HeR rejection mechanism does have its own virtue since the rejection of homeologous recombination is better to be at an earlier stage considering the long lasting, complicated procedure to achieve recombination. Timely check and abortion of recombination between diverged sequences greatly benefit for cells to combat a deadly lesion without wasting unnecessary time and energy. To sum up, results and conclusion in this dissertation provide detailed genetic evidence supporting homeologous tolerance at later stages of HR.

Our findings also have great significance in understanding HeR events documented in rare medical conditions. The recombination events within these artificial, chromosomally integrated substrates are conceivably similar to the recombination events between evolutionarily related, diverged sequences *in vivo*, such as recombination between Alu sequences or between members of a gene family. Our results highlighted the risk of recombination between these repeats since they still share a significant amount of homology. On the other side, from the evolution point of view, the results also showed the recombinational exchange between slightly diverged sequence could be very active in aforementioned favorable conditions, which probably contributes to the homogenization of a gene family, termed as concerted evolution.

Another important conclusion of the current research is that homology requirements of spontaneous and DSB-induced recombination are dramatically different, and DSB is sloppy in choosing resolution site. The results emphasize the high risk in DSB repair, where HeR events are easily obtained even without assistance of homology on both sides of the homeologous sequences. Currently, the mechanism underlying the different homology requirements, whether SDSA or nearby DSB, during DSB repair relaxed homology requirement, remains unknown; the desired HeR events resulting solely from Holliday junction resolution are not yet recovered. If the products of this sub-pathway really exist and are able to be recovered in the future, they would provide invaluable help to clarify the cause of relaxed homology requirements during DSB repair. It remains necessary to conduct more DSB experiments or MSH2 depletion experiments in order to recover the aforementioned recombination events. And meanwhile, *in vitro* studies are also informative and valuable if they can determine whether Holliday junctions could be efficiently resolved in a region of homeologous sequences.

Moreover, attempts have been carried out to see RecQ4's function in DSB repair. The recombination substrate pLB4 allows us to collect a broad spectrum of DNA repair events, and check the details of sequence changes in recovered events. Agonistic roles of BLM and RecQ4 were found in generating crossover events. Since RecQ4 joins the replisome, unwinds DNA strands for DNA synthesis and has high affinity to Holliday junctions, it is postulated that RecQ4 counteracts BLM in D-loop rejection or convergent branch migration of Holliday junctions. There is a conflicting report stating RecQ4 and BLM cooperate to reduce crossover events in studies on BLM patients' cells. Therefore,

more experiments, both *in vitro* and *in vivo*, are required to verify and pursue the current results.

The work presented in this dissertation moves forward our understanding of the mechanism and regulation of HR, and warrants further investment to answer these yet-resolved and newly emerging questions. Better understanding of HR will inevitably contribute to our knowledge about genome instability, carcinogenesis, and has potential applications to improve human health.

REFERENCE

1. Jackson SP, Bartek J. 2009. The DNA-damage response in human biology and disease. Nature 461:1071–1078.
2. Garinis GA, van der Horst GTJ, Vijg J, Hoeijmakers JHJ. 2008. DNA damage and ageing: new-age ideas for an age-old problem. Nat Cell Biol 10:1241–1247.
3. Smith JA, Bannister LA, Bhattacharjee V, Wang Y, Waldman BC, Waldman AS. 2007. Accurate homologous recombination is a prominent double-strand break repair pathway in mammalian chromosomes and is modulated by mismatch repair protein Msh2. Mol Cell Biol 27:7816–7827.
4. San Filippo J, Sung P, Klein H. 2008. Mechanism of eukaryotic homologous recombination. Annu Rev Biochem 77:229–257.
5. Krejci L, Altmannova V, Spirek M, Zhao X. 2012. Homologous recombination and its regulation. Nucleic Acids Res 40:5795–5818.
6. Li X, Heyer WD. 2008. Homologous recombination in DNA repair and DNA damage tolerance. Cell Res 18:99–113.
7. Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. 1983. The double-strand-break repair model for recombination. Cell 33:25–35.
8. Frankenberg-Schwager M, Frankenberg D. 1990. DNA double-strand breaks: their repair and relationship to cell killing in yeast. Int J Radiat Biol 58:569–575.
9. Vilenchik MM, Knudson AG. 2003. Endogenous DNA double-strand breaks:

production, fidelity of repair, and induction of cancer. Proc Natl Acad Sci U S A 100:12871–12876.

10. Nakanishi K, Cavallo F, Perrouault L, Giovannangeli C, Moynahan ME, Barchi M, Brunet E, Jasin M. 2011. Homology-directed Fanconi anemia pathway cross-link repair is dependent on DNA replication. Nat Struct Mol Biol 18:500–503.
11. Friedrich K, Lee L, Leistritz DF, Nürnberg G, Saha B, Hisama FM, Eyman DK, Lessel D, Nürnberg P, Li C, Garcia-F-Villalta MJ, Kets CM, Schmidtke J, Cruz VT, Van den Akker PC, Boak J, Peter D, Compoginis G, Cefle K, Ozturk S, López N, Wessel T, Poot M, Ippel PF, Groff-Kellermann B, Hoehn H, Martin GM, Kubisch C, Oshima J. 2010. WRN mutations in Werner syndrome patients: genomic rearrangements, unusual intronic mutations and ethnic-specific alterations. Hum Genet 128:103–111.
12. Wang Y, Smith K, Waldman BC, Waldman AS. 2011. Depletion of the bloom syndrome helicase stimulates homology-dependent repair at double-strand breaks in human chromosomes. DNA Repair (Amst) 10:416–426.
13. Latt SA, Schreck RR. 1980. Sister chromatid exchange analysis. Am J Hum Genet 32:297–313.
14. Krejci L, Altmannova V, Spirek M, Zhao X. 2012. Homologous recombination and its regulation. Nucleic Acids Res 40:5795–5818.
15. Lu H, Shamanna RA, Keijzers G, Anand R, Rasmussen LJ, Cejka P, Croteau DL, Bohr VA. 2016. RECQL4 Promotes DNA End Resection in Repair of DNA Double-Strand Breaks. Cell Rep 16:1–13.
16. Ferguson DO, Holloman WK. 1996. Recombinational repair of gaps in DNA is

asymmetric in Ustilago maydis and can be explained by a migrating D-loop model. Proc Natl Acad Sci U S A 93:5419–5424.

17. Andersen SL, Sekelsky J. 2010. Meiotic versus mitotic recombination: Two different routes for double-strand break repair. BioEssays 32:1058–1066.

18. Nassif N, Penney J, Pal S, Engels WR, Gloor GB. 1994. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. Mol Cell Biol 14:1613–25.

19. Johnson RD, Jasin M. 2000. Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. EMBO J 19:3398–407.

20. Haber JE, Ira G, Malkova A, Sugawara N. 2004. Repairing a double-strand chromosome break by homologous recombination: revisiting Robin Holliday's model. Philos Trans R Soc Lond B Biol Sci.

21. Bizard AH, Hickson ID. 2014. The dissolution of double Holliday junctions. Cold Spring Harb Perspect Biol 6:a016477.

22. Bocquet N, Bizard AH, Abdulrahman W, Larsen NB, Faty M, Cavadini S, Bunker RD, Kowalczykowski SC, Cejka P, Hickson ID, Thomä NH. 2014. Structural and mechanistic insight into Holliday-junction dissolution by topoisomerase IIIα and RMI1. Nat Struct Mol Biol 21:261–8.

23. Neelsen KJ, Lopes M. 2015. Replication fork reversal in eukaryotes: from dead end to dynamic response. Nat Rev Mol Cell Biol 16:207–220.

24. Saleh-Gohari N, Bryant HE, Schultz N, Parker KM, Cassel TN, 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.

25. Marians KJ, Pasero P, Yeeles JTP. 2016. Rescuing Stalled or Damaged Replication Forks.pdf 1–16.

26. Waldman AS, Liskay RM. 1988. Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. Mol Cell Biol 8:5350–5357.

27. Luria S, Delbrück M. 1943. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. Genetics 28:491–511.

28. Helleday T. 2003. Pathways for mitotic homologous recombination in mammalian cells. Mutat Res - Fundam Mol Mech Mutagen 532:103–115.

29. Ward JF. 1988. DNA Damage Produced by Ionizing Radiation in Mammalian Cens : Identities , Mechanisms of Formation , and Reparability. Prog nucelic acid Res Mol Biol 35:95–125.

30. Helleday T, Lo J, van Gent DC, Engelward BP. 2007. DNA double-strand break repair: From mechanistic understanding to cancer treatment. DNA Repair (Amst) 6:923–935.

31. Dueva R, Iliakis G. 2013. Alternative pathways of non-homologous end joining (NHEJ) in genomic instability and cancer. Transl Cancer Res 2:163–177.

32. Guirouilh-Barbat J, Lambert S, Bertrand P, Lopez BS. 2014. Is homologous recombination really an error-free process? Front Genet 5:175.

33. Lin Y, Waldman BC, Waldman AS. 2003. Suppression of high-fidelity double-strand break repair in mammalian chromosomes by pifithrin-a, a chemical inhibitor of p53. DNA Repair (Amst) 2:1–11.

34. Lin Y, Lukacsovich T, Waldman AS. 1999. Multiple pathways for repair of DNA

double-strand breaks in mammalian chromosomes. Mol Cell Biol 19:8353–60.

35. Lazzerini-Denchi E, Sfeir A. 2016. Stop pulling my strings - what telomeres taught us about the DNA damage response. Nat Rev Mol Cell Biol 17:364–378.

36. Schubert I, Schubert V, Fuchs J. 2011. No evidence for “break-induced replication” in a higher plant - but break-induced conversion may occur. Front Plant Sci 2:8.

37. Lin WY, Wilson JH, Lin Y. 2013. Repair of chromosomal double-strand breaks by precise ligation in human cells. DNA Repair (Amst) 12:480–487.

38. Bétermier M, Bertrand P, Lopez BS. 2014. Is Non-Homologous End-Joining Really an Inherently Error-Prone Process? PLoS Genet 10.

39. Keijzers G, Maynard S, Shamanna RA, Rasmussen LJ, Croteau DL, Bohr VA. 2014. The role of RecQ helicases in non-homologous end-joining. Crit Rev Biochem Mol Biol 49:463–472.

40. Ferguson DO, Alt FW. 2001. DNA double strand break repair and chromosomal translocation: lessons from animal models. Oncogene 20:5572–5579.

41. Shamanna RA, Lu H, de Freitas JK, Tian J, Croteau DL, Bohr VA. 2016. WRN regulates pathway choice between classical and alternative non-homologous end joining. Nat Commun 7:13785.

42. Fishman-Lobell J, Rudin N, Haber JE. 1992. Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. Mol Cell Biol 12:1292–1303.

43. Hu L, Kim TM, Son MY, Kim SA, Holland CL, Tateishi S, Kim DH, Yew PR, Montagna C, Dumitrache LC, Hasty P. 2013. Two replication fork maintenance

pathways fuse inverted repeats to rearrange chromosomes. Nature 501:569–572.

44. McEachern MJ, Haber JE. 2006. Break-induced replication and recombinational telomere elongation in yeast. Annu Rev Biochem 75:111–135.

45. Donnianni RA, Symington LS. 2013. Break-induced replication occurs by conservative DNA synthesis. Proc Natl Acad Sci U S A 110:13475–80.

46. Yamamoto A, Yagi H, Habu T, Yoshimura Y, Matsushiro A, Nishimune Y, Morita T, Taki T, Yoshida K, Yamamoto K, Matsushiro A. 1996. Cell cycle-dependent expression of the mouseRad51 gene in proliferating cells. Mol Gen Genet MGG 251:1–12.

47. Flygare J, Benson F, Hellgren D. 1996. Expression of the human RAD51 gene during the cell cycle in primary human peripheral blood lymphocytes. Biochim Biophys Acta - Mol Cell Res 1312:231–236.

48. Huertas P, Jackson SP. 2009. Human CtIP Mediates Cell Cycle Control of DNA End Resection and Double Strand Break Repair 284:9558–9565.

49. Yu X, Chen J. 2004. DNA Damage-Induced Cell Cycle Checkpoint Control Requires CtIP , a Phosphorylation-Dependent Binding Partner of BRCA1 C-Terminal Domains 24:9478–9486.

50. Ciccia A, Elledge SJ. 2010. Review The DNA Damage Response : Making It Safe to Play with Knives. Mol Cell 40:179–204.

51. Chiarle R, Zhang Y, Frock RL, Lewis SM, Molinie B, Ho YJ, Myers DR, Choi VW, Compagno M, Malkin DJ, Neuberg D, Monti S, Giallourakis CC, Gostissa M, Alt FW. 2011. Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. Cell 147:107–119.

52. Chen JM, Cooper DN, Chuzhanova N, Ferec C, Patrinos GP. 2007. Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet 8:762–775.

53. Yang D, Goldsmith EB, Lin Y, Waldman BC, Kaza V, Waldman AS. 2006. Genetic exchange between homeologous sequences in mammalian chromosomes is averted by local homology requirements for initiation and resolution of recombination. Genetics 174:135–144.

54. Spies M, Fishel R. 2015. Mismatch Repair during Homologous and Homeologous Recombination. CSH Perspect Biol 1–22.

55. Bhattacharjee V, Lin Y, Waldman BC, Waldman AS. 2014. Induction of recombination between diverged sequences in a mammalian genome by a double-strand break. Cell Mol Life Sci 117:759–769.

56. Tham KC, Hermans N, Winterwerp HHK, Cox MM, Wyman C, Kanaar R, Lebbink JHG. 2013. Mismatch Repair Inhibits Homeologous Recombination by Coordinated Directional Unwinding of Trapped DNA Structures. Mol Cell 3:326–337.

57. Honda M, Okuno Y, Hengel SR, Martín-López J V, Cook CP, Amunugama R, Soukup RJ, Subramanyam S, Fishel R, Spies M. 2014. Mismatch repair protein hMSH2-hMSH6 recognizes mismatches and forms sliding clamps within a D-loop recombination intermediate. Proc Natl Acad Sci U S A 111:E316-25.

58. Li GM. 2008. Mechanisms and functions of DNA mismatch repair. Cell Res 18:85–98.

59. Goldfarb T, Alani E. 2005. Distinct roles for the Saccharomyces cerevisiae mismatch repair proteins in heteroduplex rejection, mismatch repair and

nonhomologous tail removal. Genetics 169:563–574.

60. Li L, McVety S, Younan R, Liang P, Du Sart D, Gordon PH, Hutter P, Hogervorst FBL, Chong G, Foulkes WD. 2006. Distinct patterns of germ-line deletions in MLH1 and MSH2: the implication of Alu repetitive element in the genetic etiology of Lynch syndrome (HNPCC). Hum Mutat 27:388.

61. Pink RC, Wicks K, Caley DP, Punch EK, Jacobs L, Raul D, Carter F. 2011. Pseudogenes : Pseudo-functional or key regulators in health and disease ? 17:792–798.

62. Balakirev ES, Ayala FJ. 2003. Pseudogenes: are they "junk" or functional DNA? Annu Rev Genet 37:123–151.

63. Bengesser K, Vogt J, Mussotter T, Mautner VF, Messiaen L, Cooper DN, Kehrer-Sawatzki H. 2014. Analysis of crossover breakpoints yields new insights into the nature of the gene conversion events associated with large NF1 deletions mediated by nonallelic homologous recombination. Hum Mutat 35:215–226.

64. Waldman AS. 2008. Ensuring the fidelity of recombination in mammalian chromosomes. BioEssays 30:1163–1171.

65. Štambuk S, Radman M. 1998. Mechanism and control of interspecies recombination in Escherichia coli. I. Mismatch repair, methylation, recombination and replication functions. Genetics 150:533–542.

66. Datta A, Adjiri A, New L, Crouse GF, Jinks Robertson S. 1996. Mitotic crossovers between diverged sequences are regulated by mismatch repair proteins in Saccaromyces cerevisiae. Mol Cell Biol 16:1085–93.

67. Datta a, Hendrix M, Lipsitch M, Jinks-Robertson S. 1997. Dual roles for DNA

sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. Proc Natl Acad Sci U S A 94:9757–9762.

68. Sugawara N, Goldfarb T, Studamire B, Alani E, Haber JE. 2004. Heteroduplex rejection during single-strand annealing requires Sgs1 helicase and mismatch repair proteins Msh2 and Msh6 but not Pms1. Proc Natl Acad Sci U S A 101:9315–9320.

69. Elliott B, Jasin M. 2001. Repair of double-strand breaks by homologous recombination in mismatch repair-defective mammalian cells. Mol Cell Biol 21:2671–2682.

70. Wang Y, Li S, Smith K, Waldman BC, Waldman AS. 2016. Intrachromosomal recombination between highly diverged DNA sequences is enabled in human cells deficient in Bloom helicase. DNA Repair (Amst) 41:73–84.

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

72. Liskay RM, Letsou A, Stachelek JL. 1987. Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. Genetics 115:161–167.

73. Waldman AS, Liskay RM. 1987. Differential effects of base-pair mismatch on intrachromosomal versus extrachromosomal recombination in mouse cells. Proc Natl Acad Sci U S A 84:5340–5344.

74. Wagner MJ, Sharp JA, Summers WC. 1981. Nucleotide sequence of the thymidine kinase gene of herpes simplex virus type 1. Proc Natl Acad Sci U S A 78:1441–1445.

75. Waldman AS, Waldman BC. 1998. Stable transfection of mammalian cells by syringe-mediated mechanical loading of DNA. Anal Biochem 258:216–22.

76. Letsou A, Liskay RM. 1987. Effect of the molecular nature of mutation on the efficiency of intrachromosomal gene conversion in mouse cells. Genetics 769:759–769.

77. Lukacsovich T, Waldman AS. 1999. Suppression of intrachromosomal gene conversion in mammalian cells by small degrees of sequence divergence. Genetics 151:1559–1568.

78. Guirouilh-Barbat J, Huck S, Bertrand P, Pirzio L, Desmaze C, Sabatier L, Lopez BS. 2004. Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells. Mol Cell 14:611–623.

79. Larocque JR, Jasin M. 2010. Mechanisms of recombination between diverged sequences in wild-type and BLM-deficient mouse and human cells. Mol Cell Biol 30:1887–1897.

80. Tago Y, Imai M, Ihara M, Atofuji H, Nagata Y, Yamamoto K. 2005. Escherichia coli Mutator ΔpolA is Defective in Base Mismatch Correction: The Nature of in Vivo DNA Replication Errors. J Mol Biol 351:299–308.

81. Belousova EA, Lavrik OI. 2015. DNA polymerases β and λ and their roles in cell. DNA Repair (Amst) 29:112–126.

82. Makarova A V., Burgers PM. 2015. Eukaryotic DNA polymerase ζ. DNA Repair (Amst) 29:47–55.

83. Glaab W, Tindall K. 1997. Mutation rate at the hprt locus in human cancer cell lines with specific mismatch repair-gene defects. Carcinogenesis 18:1–8.

84. Kolodner RD, Marsischky GT. 1999. Eukaryotic DNA mismatch repair. Curr Opin Genet Dev 9:89–96.

85. Kunkel T, Erie D. 2005. DNA Mismatch Repair. Annu Rev Biochem 74:681–710.

86. Lenhart JS, Pillon MC, Guarné A, Biteen JS, Simmons LA. 2016. Mismatch repair in Gram-positive bacteria. Res Microbiol 167:4–12.

87. Tham KC, Hermans N, Winterwerp HHK, Cox MM, Wyman C, Kanaar R, Lebbink JHG. 2013. Mismatch repair inhibits homeologous recombination via coordinated directional unwinding of trapped DNA structures. Mol Cell 51:326–337.

88. Yan T, Berry SE, Desai AB, Kinsella TJ. 2003. DNA mismatch repair (MMR) mediates 6-thioguanine genotoxicity by introducing single-strand breaks to signal a G2-M arrest in MMR-proficient RKO cells. Clin Cancer Res 9:2327–2334.

89. Goring DR, DuBow MS. 1985. A cytotoxic effect associated with 9-(1,3-dihydroxy-2-propoxymethyl)-guanine is observed during the selection for drug resistant human cells containing a single herpesvirus thymidine kinase gene. Biochem Biophys Res Commun 133:195–201.

90. Suzuki N, Emura T, Fukushima M. 2011. Mode of action of trifluorothymidine (TFT) against DNA replication and repair enzymes. Int J Oncol 39:263–270.

91. Christian S, Wintersberger E. 1989. Mouse thymidine kinase : sequence of a processed pseudogene. Nucleic Acids Res 17:19135.

92. Meselson MS. 1975. General Model for Genetic Recombination 72:358–361.

93. Bzymek M, Thayer NH, Oh SD, Kleckner N, Hunter N. 2010. Double Holliday junctions are intermediates of DNA break repair. Nature 464:937–41.

94. Schwacha A, Kleckner N. 1995. Identification of double Holliday junctions as intermediates in meiotic recombination. Cell 83:783–791.

95. Hunter N, Flemming W, Boveri T. 2015. Meiotic Recombination : The Essence of Heredity.

96. Heyer WD. 2015. Regulation of Recombination and Genomic.

97. Orr-Weaver TL, Szostak JW. 1983. Yeast recombination: the association between double-strand gap repair and crossing-over. Proc Natl Acad Sci U S A 80:4417–4421.

98. Orr-Weaver TL, Szostak JW. 1983. Multiple, tandem plasmid integration in Saccharomyces cerevisiae. Mol Cell Biol 3:747–9.

99. Bochman ML. 2014. Roles of DNA helicases in the maintenance of genome integrity. Mol Cell Oncol 1:e963429.

100. Chiruvella KK, Liang Z, Wilson E, Webb CJ, Wu Y, Virginia A, Wilson TE. 2013. Repair of Double-Strand Breaks by End Joining 1–22.

101. Yusa K, Horie K, Kondoh G, Kouno M, Maeda Y, Kinoshita T, Takeda J. 2004. Genome-wide phenotype analysis in ES cells by regulated disruption of Bloom's syndrome gene. Nature 429:896–899.

102. Adams MD, McVey M, Sekelsky JJ. 2003. Drosophila BLM in double-strand break repair by synthesis-dependent strand annealing. Science 299:265–267.

103. Bugreev D V, Yu X, Egelman EH, Mazin A V. 2007. Novel pro- and anti-recombination activities of the Bloom's syndrome helicase. Genes {&} Dev 21:3085–3094.

104. Cejka P, Plank JL, Dombrowski CC, Kowalczykowski SC. 2012. Decatenation of

DNA by the S. cerevisiae Sgs1-Top3-Rmi1 and RPA Complex: A Mechanism for Disentangling Chromosomes. Mol Cell 47:886–896.

105. Chang M, Bellaoui M, Zhang C, Desai R, Morozov P, Delgado-Cruzata L, Rothstein R, Freyer GA, Boone C, Brown GW. 2005. RMI1/NCE4, a suppressor of genome instability, encodes a member of the RecQ helicase/Topo III complex. EMBO J 24:2024–33.

106. Mullen JR, Nallaseth FS, Lan YQ, Slagle E, Brill SJ, Slagle CE. 2005. Yeast Rmi1 / Nce4 Controls Genome Stability as a Subunit of the Sgs1-Top3 Complex Yeast Rmi1 / Nce4 Controls Genome Stability as a Subunit of the Sgs1-Top3 Complex 25:4476–4487.

107. Gangloff S, McDonald JP, Bendixen C, Arthur L, Rothstein R. 1994. The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase. Mol Cell Biol 14:8391–8.

108. Cejka P, Plank JL, Bachrati CZ, Hickson ID, Kowalczykowski SC. 2010. Rmi1 stimulates decatenation of double Holliday junctions during dissolution by Sgs1-Top3. Nat Struct Mol Biol 17:1377–82.

109. Cejka P, Kowalczykowski SC. 2010. The full-length Saccharomyces cerevisiae Sgs1 protein is a vigorous DNA helicase that preferentially unwinds holliday junctions. J Biol Chem 285:8290–8301.

110. Singh DK, Karmakar P, Aamann M, Schurman SH, May A, Croteau DL, Burks L, Plon SE, Bohr VA. 2010. The involvement of human RECQL4 in DNA double-strand break repair. Aging Cell 9:358–371.

111. Durand F, Castorina P, Morant C, Delobel B, Barouk E, Modiano P. 2002.

Rothmund-Thomson syndrome, trisomy 8 mosaicism and RECQ4 gene mutation. Ann Dermatol Venereol 129:892–895.

112. Im J-S, Ki S-H, Farina A, Jung D-S, Hurwitz J, Lee J-K. 2009. Assembly of the Cdc45-Mcm2-7-GINS complex in human cells requires the Ctf4/And-1, RecQL4, and Mcm10 proteins. Proc Natl Acad Sci U S A 106:15628–32.

113. Su Y, Meador JA, Calaf GM, De-Santis LP, Zhao Y, Bohr VA, Balajee AS. 2010. Human RecQL4 helicase plays critical roles in prostate carcinogenesis. Cancer Res 70:9207–9217.

114. Sedlackova H, Cechova B, Mlcouskova J, Krejci L. 2015. RECQ4 selectively recognizes Holliday junctions. DNA Repair (Amst) 30:80–89.

115. Séguéla-Arnaud M, Crismani W, Larchevêque C, Mazel J, Froger N, Choinard S, Lemhemdi A, Macaisne N, Van Leene J, Gevaert K, De Jaeger G, Chelysheva L, Mercier R. 2015. Multiple mechanisms limit meiotic crossovers: TOP3α and two BLM homologs antagonize crossovers in parallel to FANCM. Proc Natl Acad Sci U S A 112:4713–8.

116. Singh DK, Ghosh AK, Croteau DL, Bohr VA. 2012. RecQ helicases in DNA double strand break repair and telomere maintenance. Mutat Res - Fundam Mol Mech Mutagen 736:15–24.

117. Sturzenegger A, Burdova K, Kanagaraj R, Levikova M, Pinto C, Cejka P, Janscak P. 2014. DNA2 cooperates with the WRN and BLM RecQ helicases to mediate long-range DNA end resection in human cells. J Biol Chem 289:27314–27326.

118. Cheok CF, Wu L, Garcia PL, Janscak P, Hickson ID. 2005. The Bloom's syndrome helicase promotes the annealing of complementary single-stranded DNA. Nucleic Acids Res 33:3932–3941.