2015

The role of MAR elements in DNA recombination

Kaja Kostyrko

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Institut de Biotechnologie

The role of MAR elements in DNA recombination

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Kaja KOSTYRKO

Titulaire d'un Master of Science in Biotechnology de l'Université de Poznań, Pologne

Jury

President: Prof. Tadeusz Kawecki
Thesis director: Prof. Nicolas Mermod
Expert: Dr Andrzej Stasiak
Expert: Dr Pavel Janscak

Lausanne, 2015

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président	Monsieur	Prof. Tadeusz **Kawecki**
Directeur de thèse	Monsieur	Prof. Nicolas **Mermod**
Experts	Monsieur	Dr Andrzej **Stasiak**
	Monsieur	Dr Pavel **Janscak**

le Conseil de Faculté autorise l'impression de la thèse de

Madame Kaja Kostyrko

Master of Science in Biotechnology of Adam Mickiewicz University, Poznan, Pologne

intitulée

The role of MAR elements in DNA recombination

Lausanne, le 17 août 2015

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Tadeusz Kawecki

Acknowledgements

I would like to thank my thesis advisor Professor Nicolas Mermod for his guidance and enthusiasm in uncovering the complicated nature of DNA recombination.

I would also like to thank Dr Andrzej Stasiak and Dr Pavel Janscak for agreeing to be the experts in my thesis committee. Many thanks also to Professor Tadeusz Kawecki for accepting to be the president of the committee.

I also wish to thank the members of the VitalIT group for all the in silico work throughout my project. In particular, I would like to thank Ioannis Xenarios, Samuel Neuenschwander, Emanuel Schmidt and Thomas Junier.

My special thanks go to all the members of the Laboratory of Molecular Biotechnology, especially Deborah, Simone, Ruthger, Yarick, Sandra, Yves and Nico. Thank you for all the fun times we spent in and out of the laboratory.

Finally, I would like to thank my parents, my grandparents, my brother and M. for all their love and support.

Summary

The untargeted integration of foreign DNA into the mammalian cell genome, extensively used in gene therapy and biotechnology, remains an incompletely understood process. It is believed to be based on cellular DNA double strand break (DSB) repair machinery and to involve two major steps: i) the formation of long gene arrays (concatemers), and ii) recombination of the resulting concatemer with the genome. The main DSB repair pathways in eukaryotes include non-homologous end-joining (NHEJ), homologous recombination (HR), and microhomology-mediated end-joining (MMEJ). However, it is still not clear, which of these pathways are responsible for transgene integration.

Here, we show that NHEJ is not the primary pathway used by mammalian cells in the transgene integration process, while the components of the HR pathway seem to be important for genomic integration but not concatemerization. Instead, concatemer formation appears to be mediated by a subset of the MMEJ pathway, termed synthesis-dependent MMEJ (SD-MMEJ). This mechanism also seems to be preferentially used for plasmid integration into the genome, as confirmed by the analysis of plasmid-to-genome junction sequences, which were found to display an SD-MMEJ pattern. Therefore, we propose the existence of two distinct SD-MMEJ subpathways, relying on different subsets of enzymes. One of these mechanisms appears to be responsible for concatemerization, while the other mechanism, partially dependent in HR enzymes, seems to mediate recombination with the genome.

Previous studies performed by our group suggested that matrix attachment regions (MARs), which are epigenetic regulatory DNA elements that participate in the formation of chromatin boundaries and augment transcription, may mediate increased plasmid integration into the genome of CHO cells by stimulating DNA recombination. In the present work, we demonstrate that MAR-mediated plasmid integration results from the enhanced SD-MMEJ pathway. Analysis of transgene integration loci and junction DNA sequences validated the prevalent use of this pathway by the MAR elements to target plasmid DNA into gene-rich areas of the CHO genome. We propose that this finding should in the future help to engineer cells for improved recombinant protein production.

In addition to investigating the process of transgene integration, we designed recombination assays to better characterize the components of the MMEJ and SD-MMEJ pathways. We also used CHO cells expressing cycle-sensitive reporter genes to demonstrate a potential role of HR proteins in the cell cycle regulation.

Résumé

L'intégration non ciblée d'ADN étranger dans le génome de cellules mammifères, largement utilisée en thérapie génique et en biotechnologie, reste un processus peu compris. On pense que ce type d'intégration est basé sur des mécanismes cellulaires de réparation des cassures double brin (CDB) d'ADN et implique deux étapes principales: i) la formation de longues molécules d'ADN constituées de multiples copies du transgène, et ii) la recombinaison du concatémère avec le génome. Les principales voies de réparation des CDB chez les eucaryotes sont la jonction d'extrémités non homologues (NHEJ), la recombinaison homologue (HR) et la jonction d'extrémités médiée par microhomologie (MMEJ). Cependant, il reste toujours des interrogations quant à savoir lesquelles de ces voies sont responsables de l'intégration du transgène.

Ici, nous montrons que la NHEJ n'est pas la voie principalement utilisée par les cellules de mammifères dans le processus d'intégration du transgène, tandis que des éléments de la HR semblent être importants pour l'intégration génomique mais pas pour la concatémérisation. Par contre, la formation de concatémères semble se faire par l'intermédiaire d'une réparation par MMEJ dépendant de la synthèse d'ADN appelée SD-MMEJ. Ce mécanisme apparaît également être préférentiellement utilisé pour l'intégration du plasmide dans le génome, comme le confirme l'analyse des séquences des jonctions plasmide-génome, qui révèlent un motif caractéristique de la SD-MMEJ. Par conséquent, nous proposons l'existence de deux sous-voies de la SD-MMEJ distinctes, qui dépendent de différents sous-ensembles d'enzymes. L'une d'entre elles serait responsable de concatémérisation, tandis que l'autre, en partie dépendante des enzymes de la HR, jouerait un rôle dans la recombinaison avec le génome.

Les études précédentes réalisées dans notre laboratoire ont suggéré que les MARs (matrix attachment regions), qui sont des éléments épigénétiques de contrôle de l'ADN qui participent à la formation des frontières de la chromatine et augmentent la transcription, peuvent promouvoir l'intégration de plasmides dans le génome des cellules CHO en stimulant la recombinaison d'ADN. Dans cette présente étude, nous avons pu montrer que l'intégration du plasmide par l'intermédiaire des éléments MAR résulte de la voie SD-MMEJ. L'analyse des loci d'intégration du transgène et des séquences des jonctions ont validé l'utilisation de cette voie par les éléments MAR pour cibler l'ADN plasmidique dans les régions du génome de CHO riches en gènes. Ces résultats pourraient aider dans l'avenir à modifier des cellules pour améliorer la production de protéines recombinantes.

En plus d'avoir étudié le processus d'intégration du transgène, nous avons conçu des essais de recombinaison permettant de mieux caractériser les composants des voies MMEJ et SD-MMEJ. Nous avons également démontré un rôle potentiel des protéines de HR dans la régulation du cycle cellulaire à l'aide de cellules CHO exprimant des gènes rapporteurs dépendant du cycle cellulaire.

Abbreviations

53BP1	p53-binding protein 1
aa	amino acid
ATM	ataxia telangiectasia mutated
ATR	ataxia telangiectasia and Rad3-related
BER	base excision repair
bp	base pair
BRCA	breast cancer susceptibility protein
CHO	Chinese hamster ovary
CMV	cytomegalovirus
CtIP	CtBP-interacting protein
DDR	DNA damage response
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
DSB	double-stranded break
DSBR	double strand break repair
dsDNA	double-stranded DNA
dsRed	Discosoma sp. red fluorescent protein
FA	Fanconi anemia
FANC	Fanconi anemia complementation group
FACS	Fluorescence-activated cell sorting

Fucci	ubiquitination-based cell cycle indicator
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GFP	green fluorescent protein
HR	homologous recombination
ICL	DNA interstrand crosslink
MAR	matrix attachment region
mAG	monomeric version of Azami green (fluorophore)
MDC1	mediator of DNA-damage checkpoint 1
mKO2	monomeric version of Kusabira Orange 2 (fluorophore)
MMEJ	microhomology-mediated end joining
MMR	mismatch repair
MRN	complex composed of Mre11, Rad50 and Nbs1
mRNA	messenger RNA
NER	nucleotide excision repair
NHEJ	non-homologous end joining
nt	nucleotide
OH	hydroxyl group
ORF	open reading frame
PARP1	poly (ADP-ribose) polymerase 1
PCNA	Proliferating cell nuclear antigen
Pold3	DNA polymerase delta subunit 3
qPCR	quantitative real-time polymerase chain reaction
Rad	radiation-repair gene
ROS	reactive oxygen species
RPA	replication protein A

RPKM	reads per kilobase per million
SD-MMEJ	synthesis-dependent microhomology-mediated end joining
siRNA	short interfering RNA
SSB	single-stranded break
SSBR	single strand break repair
ssDNA	single-stranded DNA
SV40	Simian vacuolating virus 40
TLS	translesion synthesis
Xrcc	X-ray repair cross-complementing

List of publications

Kostyrko, K., Neuenschwander, S., Junier, T., Regamey, A., Iseli, C., Schmid-Siegert, E., Bosshard, S., Majocchi, S., Girod, P.A., Xenarios, I., Mermod, N. MAR-Mediated Transgene Integration into Permissive Chromatin and efficient expression involve an SD-MMEJ-like DNA repair Pathway. In preparation.

Kostyrko, K., Mermod, N. Polymerase θ stimulates DSB repair by a MMEJ mechanism involving long DNA end resection. In revision.

Kostyrko, K., Bosshard, S., Urban, Z., Mermod, N. A role for homologous recombination proteins in cell cycle regulation. Cell Cycle. 2015 Sep 2;14(17):2853-61.

Grandjean, M., Girod, P.A., Calabrese, D., **Kostyrko, K.**, Wicht, M., Yerly, F., Mazza, C., Beckmann, J.S., Martinet, D., Mermod, N. High-level transgene expression by homologous recombination-mediated gene transfer. Nucleic Acids Res. 2011 Aug;39(15):e104.

Contents

Chapter 1

Introduction

Integration of foreign DNA into the genome of eukaryotic cells is one of the most commonly used methods in molecular biology. It permits to change the cell's genetic material in order to overexpress an exogenous protein. To achieve this, the DNA has to be delivered inside the cell and then transported into the nucleus, where it can be incorporated into the genome, for instance by using the cellular DNA repair machinery. Various techniques used to deliver genes into the cells as well as the different DNA vectors will be described in the first section of this chapter. The second section will review the different DNA repair mechanisms existing in eukaryotic cells.

1.1 Gene delivery methods

To arrive in the cell nucleus and integrate, the transgene first needs to cross the cell and nuclear membranes, and in case of plants or fungi also the cell wall. This process can be facilitated by the use of many physical or chemical techniques, collectively termed transfection, or by employing the natural ability of bacteria or viruses to infect eukaryotic cells, referred to as transformation or transduction, respectively. Different methods of transgene delivery can be used for different cell types and purposes. The main methods used for gene transfer in plants are Agrobacterium-mediated transformation, electroporation, particle bombardment, or direct microinjection into plant protoplasts (reviewed in [6,

268]). The methods of choice for gene delivery into yeast cells include lithium acetate transformation and electroporation (reviewed in [157]).

Animal cells, and more specifically mammalian cells which are the main focus of this work, can be efficiently transfected by calcium phosphate precipitation, lipofection, electroporation, microinjection, or viral transduction (reviewed in [161, 164]). Calcium phosphate precipitation is one of the oldest and cost effective methods of mammalian cell transfection [115]. It involves co-precipitation of DNA with calcium phosphate salts, which enter the cells by endocytosis. Lipofection is currently one of the most popular, but costly, methods of gene delivery into mammalian cells. It is based on the formation of vesicles from positively charged lipids which encapsulate the negatively charged nucleic acids and release it into the cell by fusing with its membrane [96].

The two most common physical methods of DNA transfer into mammalian cells are microinjection and electroporation. In the first method a glass micropipette or needle is used to introduce the DNA directly inside the cell [77]. This technique is used primarily to inject DNA into oocytes to generate transgenic animals. In contrast to microinjection, electroporation enables to introduce the transgene into a large numbers of cells. This later method uses short electrical pulses, which disturb the cell membrane generating small holes in the lipid bilayer enabling the DNA to enter into the cell [243].

Transduction, which uses the natural ability of viruses to infect animal cells, is probably the most efficient method of gene delivery. This method is based on modified viral particles, in which the viral DNA is partially replaced by the gene of interest. The most common vectors, used in this, and other transfection methods, will be described in the following sections.

1.2 Methods of Transgene integration into the genome

1.2.1 Vectors

1.2.1.1 Viral vectors

Gene transfer using viral vectors relies on the ability of viruses to introduce their genome into the nucleus of their host. Depending on the type and strain of the virus, the genetic material either replicates to rapidly produce new particles, or integrates into the cell genome as a provirus. This last feature is especially desirable from the point of view of molecular biology. Viral vectors for gene delivery are constructed by removing the genes responsible for the production of infectious viral particles. Instead they only contain the minimal set of genes necessary for encapsidation of the genetic material and sequences necessary for genomic integration (reviewed in [344]). These replication defective viral particles carrying the gene of interest are produced in special packaging cell lines, which express the missing viral genes necessary for the assembly of fully functional virions [214]. This enables the transgene-carrying viruses to enter the target cells and integrate their cargo into the genome, but prevents them from producing more infectious particles. However, this procedure also significantly extends the vector preparation time.

The most common types of viruses used for the production of viral vectors include adenoviruses, adeno-associated viruses (AAV), Herpes simplex virus (HSV-1), lentiviruses and retroviruses [11, 338]). The first experiments with viral gene transfer were performed using adenoviruses [322]. These viruses have a double-stranded DNA (dsDNA) genome, which does not integrate into the genome of the host, but can remain in form of episomes in non-dividing cells. However, cells that divide, gradually loose the transgene, which constitutes a major drawback for biotechnology. Adenoviruses have also proven to be highly immunogenic, and therefore are also not the preferred vector choice for gene ther-

apy. These defects can be overcome by the use of AAVs, which not only integrate their genetic material, but also do it in a very specific locus on human chromosome 19 [283]. However, they can only carry a cargo of up to 4 kb, which makes them impractical for large transgenes. Large cargo (up to 50 kb) can be delivered by vectors based on HSV-1 [213]. However, this virus does not integrate its genetic material into the genome.

RNA viruses – retro- and lentiviruses mediate very efficient gene integration in mammalian cells. However, to do so, they first need to convert their genetic material into DNA using a viral reverse-transcriptase. Retroviruses are only able to infect dividing cells and were shown to preferentially integrate their genome near transcription start sites, increasing the risk of gene-inactivation or oncogenic transformation [37]. Lentiviruses very efficiently infect dividing and non-dividing cells, and tend to integrate away from cellular promoters, making them a better choice than retroviruses for most applications [52].

Although viral vectors are very effective in delivering genes into mammalian cells, they have many limitations. Apart from aforementioned limited cargo size, risk of insertional mutagenesis and immunogenicity, they also entail high production costs and lengthy preparation procedures. Moreover, the use of viral vectors in recombinant protein production carries the risk of contaminating the therapeutic protein with viral particles. These disadvantages make viral vectors less attractive from the point of view of biotechnology and gene therapy. Therefore, many efforts are directed towards the development and improvement of non-viral vectors.

1.2.1.2 Transposon vectors

Transposons, also called "jumping genes", belong to naturally occurring mobile genetic elements that constitute around 45% of human genome, some of which can change their genomic positions (reviewed in [114, 140]). They generally only encode a minimal set

of genes necessary for their own transposition. Depending on the mechanism they use, transposons can be divided into two classes: retrotransposons and DNA transposons [352]. The first class resembles retroviruses in their mode of action, except that they lack the ability to produce infectious particles. They use a copy-and-paste mechanism of transposition, which involves an RNA-intermediate generated with the transposon-encoded reverse transcriptase.

Class II DNA transposons use a cut-and-paste mechanism, and generally encode a transposase protein flanked by inverted terminal repeats (ITRs), which serve as transposase binding sites and are necessary for transposition. This class of transposable elements was used to create gene transfer vectors by replacing the transposase gene with a gene of interest and providing the transposase in trans (on a separate plasmid or RNA, or as protein). This two-component system limits the transposon from being excised from a primary insertion site, as the expression of transposase is gradually lost with time. The two most commonly used transposon-based systems in mammalian cells, modified from salmon and moth transposons, were gracefully named Sleeping Beauty and Piggy Back, respectively [124, 139]. Although transposons were reported to integrate non-randomly, with a preference for TA and TTAA nucleotides [140, 366], the ability to deliver large transgene sizes, low production costs and lack of infectious potential, give them an advantage over viral vectors.

1.2.1.3 Plasmid vectors

Certainly the most common type of vectors used for transgene delivery into eukaryotic cells are plasmids - circular DNA molecules of bacterial origin. Ease and low costs of manipulation, large insert sizes, as well as the existence of constantly growing molecular cloning techniques makes them very powerful gene delivery tools for molecular biology.

Plasmids can be easily delivered into the cells using one of the methods described previously (section 1.1). After reaching the nucleus, plasmids can recombine with the genome in a random place, likely taking an advantage of a break in a chromosome. This process is thought to be mediated by the cellular DNA repair machinery, which joins the plasmid with the repaired fragment of the chromosomal DNA. Unlike viruses and transposons, which typically integrate only one copy of the gene per genomic locus, plasmid vectors often integrate in many copies [116, 167, 249]. This is thought to result from a two-step integration process [167]. In this model the plasmid molecules first join together forming long gene arrays, termed concatemers, which subsequently recombine with the cellular genome in a single genomic locus (Fig.1.1).

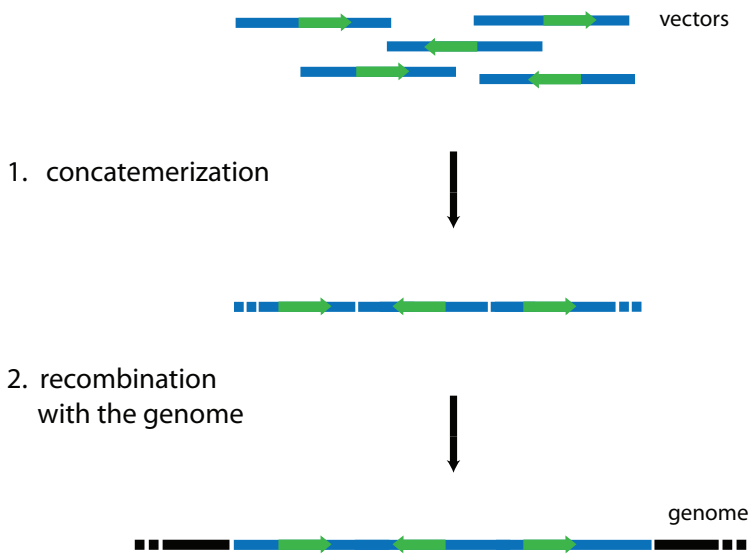


Figure 1.1: The model of the two-step plasmid integration process

While random transgene integration is relatively efficient, the major disadvantage of this method is the lack of control over the transgene integration site, which may lead to insertional mutagenesis or silencing of the gene of interest. Therefore, many alternative gene delivery systems have been developed to tackle this problem.

1.2.2 Site specific integration techniques

All the previously described vectors integrate into the genome in a more or less random fashion. As mentioned previously this is a disadvantage, as it may lead to inactivation of cellular genes or activation of proto-oncogenes. It may also result in lack of transgene expression, if the integration takes place in non-permissive heterochromatin. The following paragraphs describe the different approaches aimed at gaining control over the integration site choice, to potentially improve safety and efficiency of the transgene integration process.

1.2.2.1 Site-specific recombinases (Cre/LoxP, Flp/FRT, ϕC31/att)

Site-specific recombinases are enzymes which catalyze the recombination of specific DNA sequences, through strand cleavage, exchange and ligation (reviewed in [33, 99]). These recombinases recognize only very specific target sites composed of two inverted palindromic repeats separated by a spacer sequence, where the DNA cleavage occurs. Depending on the presence of one or two of these sites, as well as their relative orientation, the recombinases can mediate integration, excision, inversion or exchange of a DNA fragment.

The most common site-specific recombination systems are based on recombinases termed Cre (causes recombination), Flp ("flip") and ϕC31. Cre, isolated from from bacteriophage P1, recognizes 34bp target sites called loxP ("locus of crossover in phage P1") [286]. The Flp recombinase, derived from a 2μ *S. cerevisiae* plasmid, specifically recognizes 48 bp FRT (Flp recombinase recognition target) sites [113]. The target sites of both enzymes are not altered by the recombination process, which means that recombination can occur again in the same locus, potentially leading to the excision of freshly inserted transgene. This can be circumvented by using heterotypic and incompatible recognition target sites. The ϕC31 integrase, isolated from a Streptomyces phage ϕC31, mediates recombination between att (attachment) sites - attB (donor) and attP (acceptor). This process is irreversible, as

recombination results in the formation of fusion sequences termed attL and attR, which are no longer recognized by the recombinase [134, 302].

Site-specific recombinases are typically used to introduce the gene of interest into the mammalian genome by way of recombinase-mediated cassette exchange (RMCE) (reviewed in [102, 331]). In this technique a cassette with the recombinase target sites is first randomly integrated into the genome in a single copy. This acceptor cassette can then be used by the recombinases to insert a transgene flanked by the same set of recognition sites. This method is most often used to generate transgenic mice. Although it enables to precisely control the transgene integration process, it requires the pre-introduction of recombinase target sites using other techniques, as these sequences are not present in mammalian genomes. This is often requires very laborious, time consuming screening procedures, which is the greatest limitation of this technique.

1.2.2.2 Recombinant nucleases (ZFN, TALENs, CRISPR/Cas)

A novel approach to targeted integration came with the development of recombinant nucleases: zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and clustered regulatory interspaced short palindromic repeat (CRISPR)/Cas-based RNA-guided DNA endonucleases (reviewed in [98, 271]). All these systems are based on a simple idea to introduce a single- or double-stranded break in a chosen place in the genome, which is subsequently repaired by the cellular DNA repair mechanisms. Since error-prone, potentially mutagenic mechanisms are thought to repair the majority of DNA breaks in higher eukaryotes, this approach was successfully used for gene knock-out. Alternatively, these systems can be used for transgene integration. This can be achieved by providing a donor plasmid containing flanking sequences homologous to the targeted locus, which stimulates DNA recombination by homology-directed repair mechanisms.

Both ZFNs and TALENs are composed of sequence-specific DNA-binding domain (DBD) and a non-specific nuclease domain, typically the FokI restriction endonuclease. The ZFN DBDs are derived from zinc-finger proteins, while in TALENs they are based on transcription activator-like effector (TALE) proteins [98, 314, 334]. In both systems the DNA-binding region is made up of several zinc-finger or TALE modules linked together. Each module recognizes 3 bp (in ZFNs) or 1 bp (in TALENs) from the target locus. By using different combinations of modules the DBDs can be customized to recognize a wide variety of sequences. The active enzyme is a dimer of two monomers targeting the sequences flanking the FokI cleavage site.

A recently developed CRISPR/Cas system is based on a bacterial defense mechanism against foreign DNA [210]. It consists of a guide RNA, which recognizes a specific DNA target sequence, and a Cas nuclease which cleaves it. By designing a specific guide RNA the system can be easily engineered to cut most target sequences. This makes it easier and more affordable than the ZFNs and TALENs. However, its specificity and the potential for off-target effects still need to be fully assessed.

1.3 Limitations of stable transgenesis

None of the methods of gene delivery and integration fully ensure efficient and long term transgene expression in all cells having received the transgene. Regardless of the delivery type, the gene of interest becomes a part of the cellular genome upon integration, where it can be influenced by the surrounding genes, regulatory elements and changes of the chromatin structure. This phenomenon, collectively termed the position effect, can result in transgene silencing, due to heterochromatin spreading, or decreased expression caused by the neighboring repressor elements. To protect the transgene against negative position effects and stabilize its expression, epigenetic regulatory elements are often

incorporated into the vectors. These elements may act as insulators, i.e. protect the transgene from the effects of cellular enhancers or repressors, or as chromatin boundaries, i.e. prevent the spreading of heterochromatin into the transgene integration site. Among such DNA elements are Insulators, Locus Control Regions (LCRs), STabilizing Anti-Repressor (STAR) elements, Ubiquitous chromatin opening elements (UCOEs) and Matrix Attachment Regions (MARs). The first four groups have been characterized and reviewed by others [173, 209]. The properties of MARs, which are the main focus of this work, will be described in detail in the following section.

1.4 Matrix attachment regions

MARs were first identified in 1984 as chromatin fragments which co-purified with nuclear proteins after endonuclease digestion of the genomic DNA [57, 228]. They were subsequently termed matrix- or scaffold-attachment regions (MARs or SARs), for their ability to anchor the chromatin to the nuclear matrix, a putative network of protein fibers thought to constitute the inner skeleton of the cell nucleus.

MARs are highly AT-rich (above 65%), non-coding sequences of variable length, and often contain topoisomerase II cleavage sites [105]. While they have no obvious consensus sequence, MAR-containing DNA fragments were shown to possess common physical properties, like curvature, deeper DNA major groove and wider minor groove, and a high potential for unwinding, unpairing and denaturing the double helix [29, 90, 110, 131, 262]. However, all these features likely result from the high content of AT base pairs in these elements. Although there seem to be no strong sequence similarities between different MARs described so far, they appear to be evolutionarily conserved [58]. They have been identified in the flanking regions of many genes, e.g. chicken lysozyme and human interferon β [30, 258], as well as in intergenic regions and introns.

In biotechnology, MARs are best known for their beneficial role in sustaining transgene expression [5]. One of their most attractive properties from a biotechnological point of view is their proposed function of insulators, i.e. borders between active and inactive chromatin domains, which may enable them to prevent gene silencing [106, 125]. This MAR feature increased the interest in these elements in the field of genetic engineering, as many studies have shown that an addition of a MAR element in a vector containing a transgene results in increased and more stable transgene expression [111, 209, 377].

Upon their discovery, MARs were proposed to play a role in the organization of chromatin into higher-order structures by dividing chromosomes into functionally independent 50–200 kb loop domains [105, 215]. Later studies showed that MARs may directly bind to many proteins involved in chromatin organization, e.g. special (A+T)-rich binding protein 1 (SATB1), CTCF, and SWI/SNF (mating type switching and sucrose non-fermenting) complex [38, 273, 376]. Many of these proteins also function as transcription factors, which suggested a role for MAR elements in the regulation of gene expression [10, 101]. MARs were also proposed to contribute to the initiation of DNA replication [71, 259]. Recently, the data from our group also implicated these elements in the process of DNA recombination [110, 111, 116]. All this evidence suggests that MAR elements are not only important for chromatin organization within the nucleus, but may also participate in the regulation of essential cellular processes like transcription, replication and DNA repair.

1.5 DNA repair

1.5.1 DNA damage response and cell cycle regulation

All living organisms are constantly exposed to numerous factors causing DNA damage, which poses a serious threat to the integrity of the cell genome. DNA damage can lead

to cellular senescence or cell death, if left unrepaired, or to various types of mutations if misrepaired. To counteract DNA damage, cells activate a network of DNA repair and signaling pathways, collectively termed the DNA damage response (DDR) [65, 129, 132, 145, 200]. The role of the DDR is to sense DNA damage, initiate its repair through one of many specialized pathways, and if necessary, arrest cell cycle progression to allow time for the repair to be completed. Cells possess many DNA repair pathways, responsible for repairing different types of DNA lesions. The most important of these pathways operating in eukaryotic cells will be described later in this chapter.

Some types of DNA damage, e.g. chemically modified or mismatched nucleotides, can be repaired rapidly, without activating the DDR. However, more complicated lesions, like DNA interstrand crosslinks or single- and double-stranded breaks (SSBs, DSBs), which require more time to be processed, stimulate DNA damage signaling leading to cell cycle delay or arrest.

Upon SSB formation, the exposed single-stranded DNA (ssDNA) is rapidly recognized and bound by replication protein A (RPA), which in turn leads to the activation of the primary transducer kinase - ataxia telangiectasia and Rad3-related (ATR) [238]. The first sensor of DSBs, the Mre11/Rad50/Nbs1 (MRN) complex, activates ataxia telangiectasia mutated (ATM), another key protein kinase in the DDR network [294, 296]. ATR and ATM are responsible for the phosphorylation of cell cycle checkpoint kinases 1 and 2 (CHK1 and CHK2), respectively. These kinases spread the DDR signal throughout the nucleus by phosphorylating many other effector proteins. This ultimately results in the inactivation of the cell-division cycle 25 (CDC25) phosphatase and in the activation of the tumor suppressor p53 [208, 332]. Inactivation of CDC25, the phosphatase controlling entry into mitosis, results in a rapid cell cycle arrest. Phosphorylated p53 activates the transcription of its target genes, most notably a cyclin-dependent kinase inhibitor p21,

which induces a persistent G1 phase arrest [74] (Fig.1.2).

ATM also phosphorylates many downstream DNA repair and cell cycle factors, including histone H2AX, mediator of DNA-damage checkpoint 1 (MDC1), p53-binding protein 1 (53BP1) and breast cancer susceptibility protein 1 (BRCA1) [183, 296]. Activated MDC1 participates in the amplification of the DDR signal by recruiting more ATM molecules to the DNA damage site [202]. Phosphorylated 53BP1 and Brca1 activate the non-homologous end-joining (NHEJ) and homologous recombination (HR) DSB repair pathways, respectively [132]. The phosphorylation of histone H2AX (γH2AX) induces global changes in the chromatin structure leading to further recruitment of DNA repair proteins to the sites of damage [93]. This accumulation of repair factors results in the formation of microscopically-visible nuclear foci. When DSBs are repaired, dephosphorylation of γH2AX and loss of nuclear foci suppresses the DDR machinery, leading to the progression of the normal cell cycle. A failure to repair the damage results in persistent DDR signaling and triggers apoptosis [145].

1.5.2 DNA damage repair pathways

Cells evolved many different repair mechanisms against the DNA damage to which the genome is exposed every day. In case of simple chemical modifications of nucleotides, e.g. spontaneous addition of a methyl group (alkylation), the damage can usually be reversed by a single enzyme [289]. However, to repair more complex types of DNA damage cells possess specialized enzymatic pathways often involving many components.

1.5.2.1 Base excision repair (BER)

The base excision repair (BER) pathway is used to correct relatively minor damage to DNA bases, resulting from oxidation, deamination, or alkylation (reviewed in [78, 166, 170]).

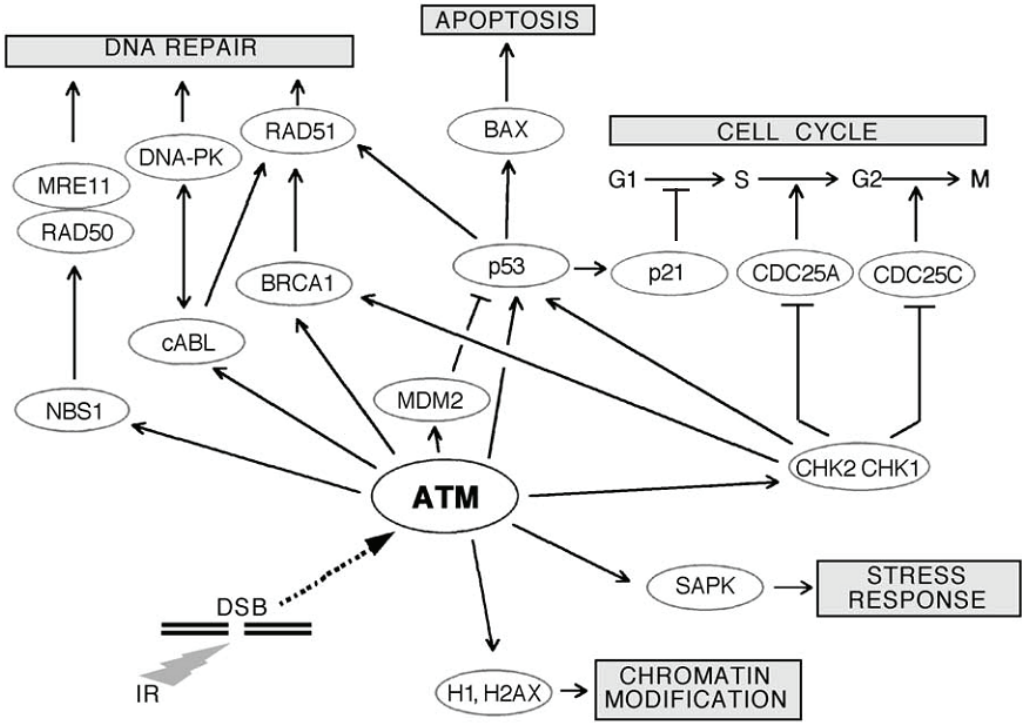


Figure 1.2: The ATM pathway in response to DSBs (from [64]).

The process is initiated by a DNA glycosylase, an enzyme which recognizes the damaged base and removes it by cleaving the N-glycosylic bond between the base and the sugar backbone [191]. Up to date, 11 different glycosylases were identified in mammals, each recognizing a specific base modification. The resulting apurinic/apyrimidinic (AP) site is processed by an AP endonuclease, which cuts the phosphodiester bond 5' to the AP site, and a phosphodiesterase, which cleaves the sugar-phosphate backbone, leading to the removal of the suger backbone (reviewed in [166]). Alternatively, an enzyme with the AP lyase activity, e.g. polymerase β, directly removes the sugar backbone overriding the need for the AP endonuclease [303]. Finally the remaining gap is filled by a DNA polymerase and sealed by a ligase.

Depending on the extent of damage, BER can proceed through one of the two sub-pathways, called the short- or long-patch BER. In short-patch repair only the damaged nucleotide is replaced by DNA polymerase β, a BER-specific member of the Pol X family,

which typically inserts single nucleotides (reviewed in [364]). In contrast, long-patch BER is mediated by DNA replication enzymes – proliferating cell nuclear antigen (PCNA) and the highly processive Pol B family DNA polymerases - ϵ and δ [310].

In the final step of BER, the remaining nick is sealed by DNA Ligase I or by the Ligase III/Xrcc1 complex [324]. Ligase I, which also plays a role in DNA replication, is thought to participate primarily in long-patch repair. Ligase III seems to be more important for short-patch BER. Yeast, which lack the homolog of mammalian DNA Ligase III as well as equivalents of polymerase β and Xrcc1, seem to rely mainly on long-patch repair (reviewed in [159]).

1.5.2.2 Nucleotide excision repair (NER)

In contrast to BER, which removes modified bases, nucleotide excision repair (NER) is responsible for the repair of large helix-distorting DNA lesions, e.g. pyrimidine dimers, intrastrand crosslinks or bulky chemical adducts (reviewed in [153, 216, 287]). This type of damage is primarily induced by UV light and some chemotherapeutic agents. NER possesses a set of multifunctional enzymes that recognize a wide variety of damage. Common features of the NER substrates are bulkiness and propensity to destabilize the DNA helix.

Depending on how the damaged nucleotides are recognized, NER can be divided into two sub-pathways: global genome NER (GG-NER) or transcription-coupled NER (TC-NER) [109, 123]. In GG-NER two damage-recognition factors, the xeroderma pigmentosum complementation group C (XPC)-RAD23B dimer and the UV-damaged DNA-binding protein (UV-DDB), scan the entire genome for helix deformations [312]. TC-NER occurs only during transcription when RNA polymerase II is stalled by a damaged nucleotide in the template DNA. This process does not require the XPC-RAD23B and UV-DDB complexes, instead it relies on many TC-NER specific factors, including Cockayne syndrome proteins

A and B (CSA and CSB) [95].

The remaining steps of the repair process are the same for both sub-pathways. After the initial recognition of the damage, the core NER complex assembles from transcription factor II H (TFIIH) and the xeroderma pigmentosum B, D, A, and G (XPB, XPD, XPA, and XPG) proteins [89]. RPA also participates in the complex by binding to the undamaged ssDNA strand. Finally the XPF/Ercc1 complex is recruited to the damage site, which leads to the cleavage 5' and 3' to the lesion by XPF and XPG respectively, resulting in the removal of a 22-30 nucleotide DNA fragment [17, 298]. The new strand is synthesized by DNA polymerases ϵ or δ, together with PCNA, or DNA polymerase κ, one of the translesion synthesis (TLS) polymerases [247]. In the last step, the nick is sealed by Ligase I or the Ligase III/Xrcc1 complex [232, 247].

The diseases caused by the deficiency in NER enzymes are Xeroderma pigmentosum, characterized by high sensitivity to sunlight and development of skin cancers, and Cockayne's syndrome, which results in premature ageing [54].

1.5.2.3 Mismatch repair (MMR)

The two repair mechanisms described in the previous sections are responsible for removing nucleotides that have an incorrect chemical structure. However, they are unable to recognize normal nucleotides incorrectly inserted during replication. A mechanism capable of doing this is termed mismatch repair (MMR). MMR detects non-Watson-Crick base pairs and insertion/deletion loops (IDLs), typically resulting from DNA polymerase slippage during the replication of a microsatellite repeats. To prevent mutations in the daughter DNA, this type of damage needs to be repaired already during replication.

MMR is a highly conserved mechanism found in most living organisms (reviewed in [59, 15, 149, 255]). The first MMR enzyme, Mutator S (MutS) was discovered in *E. coli*

[311]. Later MutS homologs (MSHs) were also identified in eukaryotes.

In mammals, the MMR process starts with the recognition of the mismatched bases by the MutSα heterodimer, composed of MSH2 and MSH6 [135] (Fig. 1.3). IDL recognition is thought to be mediated by a MutSβ complex, comprising MSH2 and MSH3 [250]. Following damage recognition MutS recruits the next MMR-specific complex: MutLα, a heterodimer of bacterial MutL homolog 1 (MLH1) and mismatch repair endonuclease PMS2 [265]. The MutS-MutL complex then moves along the DNA, until the nearest nick in the dsDNA, e.g. a gap between two Okazaki fragments bound by PCNA. This leads to the recruitment of exonuclease 1 (Exo1), which removes part of the nicked strand surrounding the mismatch [327]. The resulting ssDNA is coated by RPA, and the new strand is synthesized by DNA polymerase δ [199]. Finally the nick is sealed by DNA Ligase I [382].

The defects in MMR enzymes were linked to cancer predisposition. One of the diseases caused by mutations in MMR genes is Lynch syndrome (hereditary nonpolyposis colorectal cancer; HNPCC), a condition characterized by early-onset cancer of the colon and other organs [251].

1.5.2.4 Interstrand crosslink (ICL) repair

In addition to base damage, many mutagenic chemicals also cause the formation of DNA interstrand crosslinks (ICLs), lesions where two nucleotides from complementary DNA strands are covalently bound. This type of damage prevents DNA strand separation causing problems during replication, leading to replication fork stalling, and during transcription. This is especially detrimental for rapidly dividing cells. For this reason many cross-linking agents, like cisplatin, mitomycin C, psoralen and nitrogen mustards, are used in cancer chemotherapy [70]. ICLs may also be caused by some endogenous metabolic products, e.g.

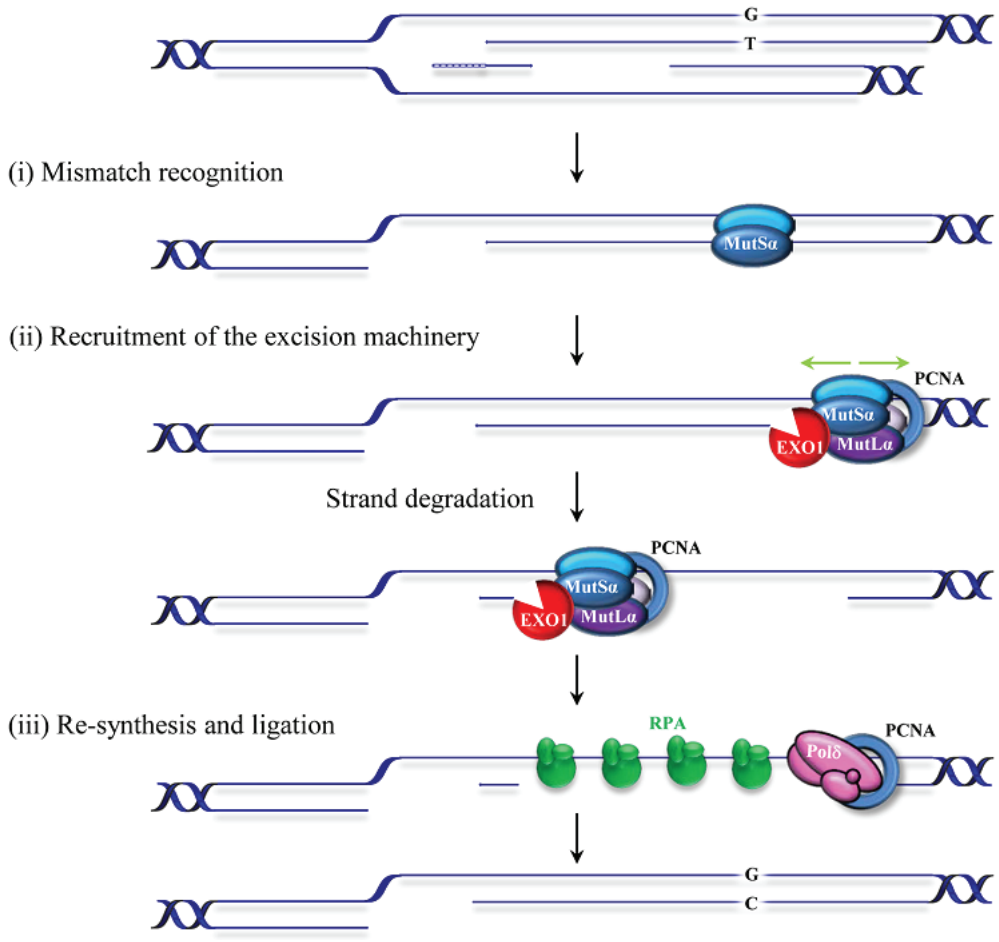


Figure 1.3: The mismatch repair pathway (from [15]).

aldehydes or nitric oxide.

In the S phase, replication fork stalling at ICLs leads to the activation of the Fanconi anemia (FA) pathway. This process was named after a disease first described in 1927 by a Swiss pediatrician Guido Fanconi [188]. FA is characterized by genomic instability, cancer predisposition, and hypersensitivity to DNA cross-linking agents (reviewed in [53, 162, 342]). Genes important for ICL repair were first identified in patients suffering from this condition.

The repair process starts when replication forks converge at the ICL, which covalently links the two DNA strands (Fig.1.4). This leads to the recruitment of Fanconi anemia complementation group M (FANCM) protein, which forms a complex with FA-associated protein 24 kDa (FAAP24) and histone fold protein 1 (MHF1) [51, 223, 365]. Upon ICL

recognition, the FANCM/FAAP24/MHF1 complex activates ATR signaling as well as the downstream FA repair factors, which together form the FA core complex. The FA core complex, composed of FANCA, -B, -C, -E, -F, -G, -L, and -M, has a ubiquitin E3 Ligase activity [104, 222]. Its role is the monoubiquitination of FANCD2 and FANCI, which is a key event in the ICL repair. The activated FANCD2-FANCI heterodimer recruits the nucleases, which cut one of the DNA strands on both sides of the ICL. Among the candidate enzymes generating the incision are Fanconi anemia-associated nuclease 1 (FAN1) and FANCP (SLX4) [165, 207]. As a result of the incisions, termed ICL unhooking, one of the DNA strands is left with the cross-linked nucleotides. TLS polymerases, REV1 or Pol ζ, synthesize a complementary strand bypassing the lesion [130]. Finally the unhooked ICL is removed by the NER pathway, described previously. The other DNA strand, containing a double stranded break (DSB), is repaired by homologous recombination (HR) using the TLS- and NER-repaired strand as a template [142, 69]. The HR pathway will be described in more detail in a later section.

1.5.2.5 Translesion synthesis (TLS)

During replication, when DNA is unwound at replication forks, repairing DNA lesions through excision may lead to DNA breaks and loss of chromosome fragments. To prevent this, the damaged strand is first replicated to restore dsDNA. This is achieved by translesion synthesis (TLS), mediated by specialized DNA polymerases, which have the ability to bypass damaged DNA bases. The polymerases implicated in this process are the Y family enzymes: polymerase η, κ, ι and Rev1, as well as the B family polymerase ζ (reviewed in [130, 281]).

TLS polymerases have a special structure, which enables them to insert nucleotides opposite chemically modified or distorted bases. This however renders them more error-

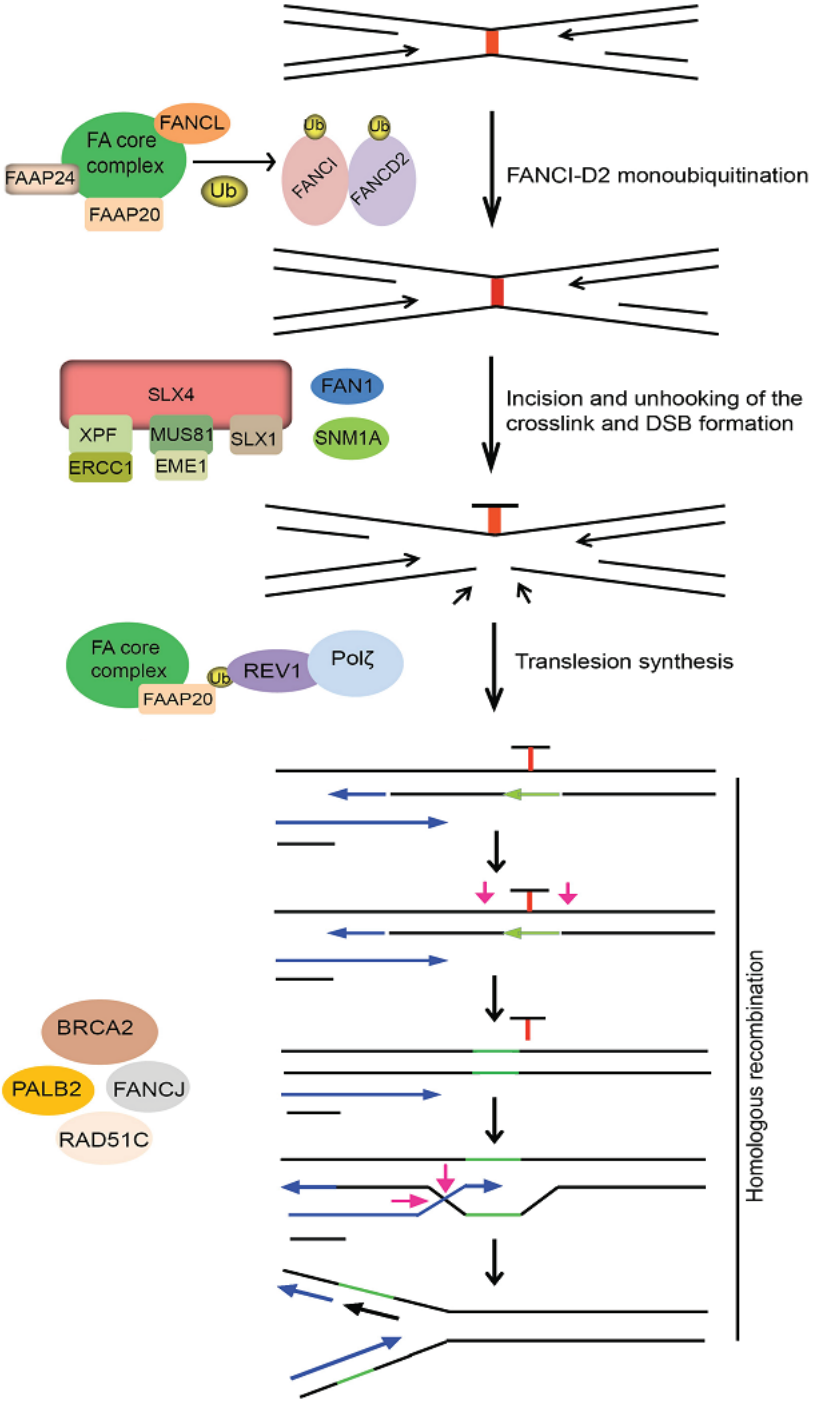


Figure 1.4: The ICL repair pathway (from [107]).

prone compared to the high fidelity replication polymerases (i.e. ϵ and δ). Therefore, their use is confined to the damage site to prevent mutagenesis. However, TLS polymerases are crucial for bypassing UV-induced lesions like DNA intrastrand crosslinks, oxidized bases or guanine N2 adducts, which block the progression of replication enzymes [130]. They also participate in the repair of ICLs, together with the FA pathway, as well as in filling-in abasic sites during BER and NER, all of which were described in the previous sections.

In addition to the main TLS enzymes, eukaryotes also possess other error-prone polymerases which are capable of translesion synthesis. These enzymes, belonging to the A and X families, are involved in a variety of DNA repair processes. The X family polymerases - μ and λ, both capable of TLS, are used mainly during non-homologous end-joining (NHEJ) repair of double stranded breaks (DSBs) [49, 91]. The A-family polymerase θ was shown to participate in BER and recently also in the microhomology-mediated end-joining (MMEJ) of DSBs [45, 47, 160, 218, 369, 372].

1.5.2.6 Single strand break repair (SSBR)

Single stranded breaks (SSBs) are a type of damage where only one of the two DNA strands is severed. These types of lesions usually result from the loss of a single nucleotide in one of the two DNA strands and are accompanied by modifications at the ends of the broken strand. SSBs can occur directly, e.g. due to reactive oxygen species (ROS) attack on the sugar backbone or as a result of abortive topoisomerase I activity (reviewed in [41, 79]). However, they can also arise indirectly, as intermediates during NER, BER or ICL repair. The most common source of SSBs are ROS, which leave 3'-phosphate and 3'-phosphoglycolate termini. Before these breaks can be repaired, DNA ends need to be restored to contain 3'-OH and 5'-phosphate groups, which are suitable targets for DNA polymerase and ligase.

The initial step of SSB repair involves break recognition by poly (ADP-ribose) polymerase-1 (PARP1), a key SSBR enzyme [87, 348]. PARP1 recruitment to SSBs leads to its activation and to that of its downstream factors by their poly(ADP)ribosylation, i.e. by the transfer of several ADP-ribose groups from NAD+ to these target proteins [42]. The main target of PARP1 is X-ray repair cross-complementing protein 1 (Xrcc1) [87, 231]. Xrcc1 serves as a docking molecule for the SSBR factors responsible for the restoration of 5' phosphate and 3' OH termini. Depending on the type of modification, different enzymes are used. Among them are polynucleotide kinase/phosphatase (PNKP), which processes ROS-induced SSBs, while Aprataxin (APTX) and tyrosyl DNA phosphodiesterase 1 (TDP1) process breaks resulting from abortive ligase and topoisomerase 1 activity (reviewed in [40]). Xrcc1 also recruits DNA polymerase β, which removes the 5' sugar remaining after the activity of the AP endonuclease 1 [303, 364]. After the modifications are removed from DNA ends, the gap is filled and ligated by BER enzymes. This most often involves DNA polymerase β, which inserts single nucleotides. However, when longer fragments need to be replaced long patch BER polymerases δ or ϵ take over. The final step of SSBR is carried out by Ligase III (in complex with Xrcc1) or Ligase I (stimulated by PCNA), the first one being used in the short- and the latter in the long-patch repair [39, 324].

1.5.2.7 Double strand break repair (DSBR)

Probably the most deleterious type of DNA damage are double stranded breaks (DSBs). DSBs are types of DNA lesions where both strands of the DNA helix are severed due to chemical agents, reactive oxygen species or ionizing radiation (reviewed in [121, 129, 144, 155]). They can also occur spontaneusly during cellular processes like replication, meiosis, V(D)J recombination or class-switch recombination. DSBs can be particularly harmful for the cells if unrepaired, leading to cell death, or to various types of mutations

if misrepaired. The two major pathways responsible for DSB repair in mammalian cells are non-homologous end-joining (NHEJ) and homologous recombination (HR) (Fig.1.5). A third group of DSB repair pathways, thought to function when the two main repair mechanisms are impaired, was collectively termed microhomology-mediated end joining (MMEJ). The choice of the repair pathway is made depending on the type of DSB and the phase of the cell cycle.

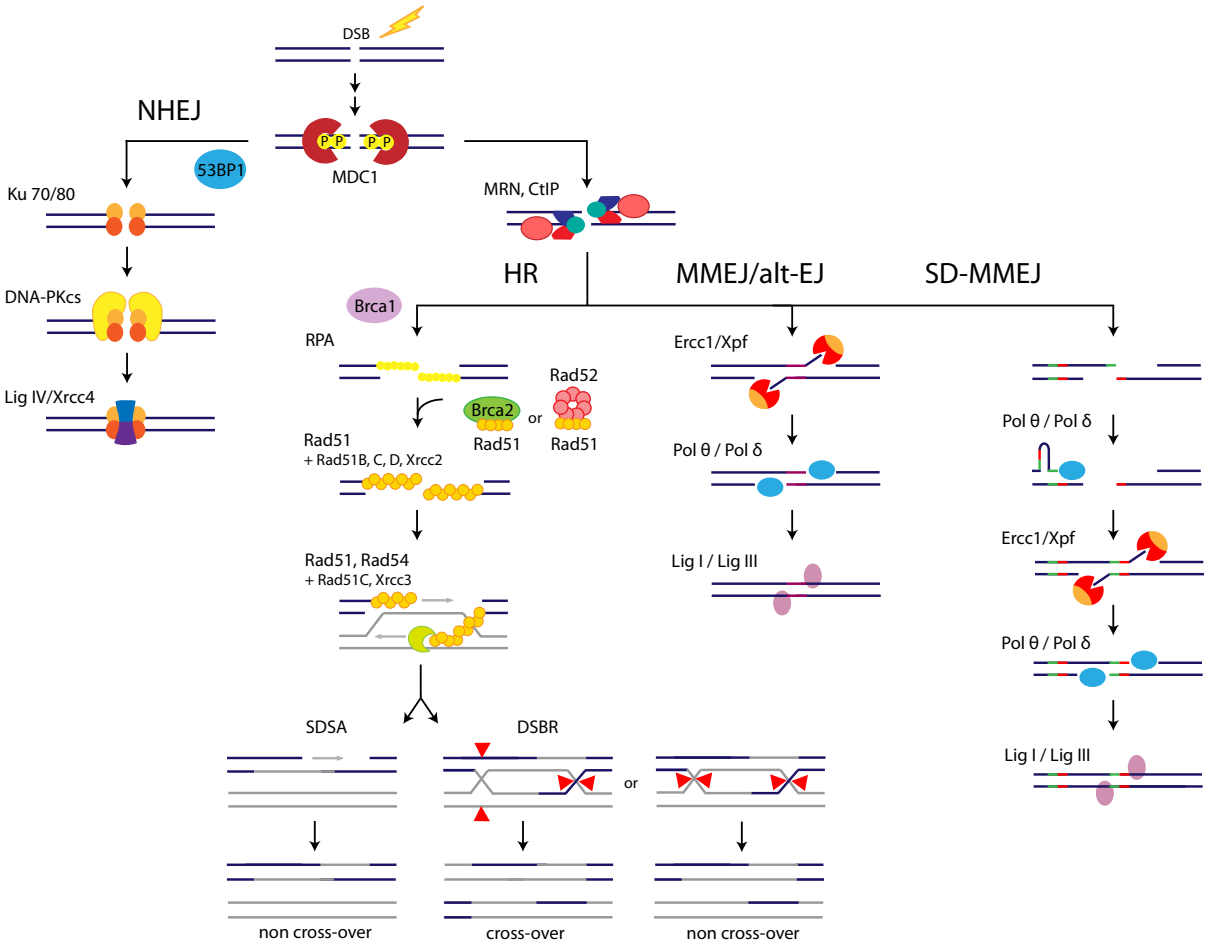


Figure 1.5: The pathways of DSB repair in higher eukaryotes. Description of the pathways, abbreviations and references are given in the text.

1.5.2.7.1 Non-homologous end joining (NHEJ)

Non-homologous end joining (NHEJ) appears to be the major mechanism responsible for DSB repair in higher eukaryotes. NHEJ operates throughout the entire cell cycle, although it becomes less important in the late S phase and G2 phase, when homologous

recombination is also used [230]. The components of this pathway were first characterized in Chinese hamster ovary (CHO) cells sensitive to ionizing radiation (reviewed in [147]). Apart from its role in the repair of radiation-induced breaks, NHEJ is also responsible for the repair of naturally arising DSBs. Among the cellular processes requiring NHEJ are V(D)J recombination, a process of antigen receptor gene rearrangement in primary immune cells, and class-switch recombination (CSR), a process of immunoglobulin heavy-chain (IgH) constant region gene rearrangement in mature B cells (reviewed in [306]).

In contrast to the other DSB repair pathways, NHEJ repairs broken ends without any need for sequence homology. It is able to very quickly and efficiently repair ends which require a simple ligation step, like blunt or cohesive ends [301]. However, when the ends are non-cohesive, additional processing is needed, rendering the process slower and more error-prone. In addition, more complex breaks, which require extensive processing, may be better substrates for other DSB repair pathways.

The initial step of NHEJ involves the Ku heterodimer (Ku70 and Ku80), which recognizes and binds to the broken DNA ends. Ku proteins were first identified by using autoantibodies from a patient with an autoimmune disease (polymyositis-scleroderma) and termed Ku after the first two letters of the patient's name [227]. Eukaryotic Ku is a dimer of two peptides: Ku70 (XRCC6) and Ku80 (XRCC5), which form a ring structure around the two broken extremities of the DNA helix [343].

DNA-bound Ku dimers serve as docking sites for the DNA-dependent protein kinase catalytic subunit (DNA-PKcs), which together with Ku forms the DNA-PK holoenzyme [147]. The DNA-PKcs molecules from the two sides of the break interact forming a bridge between the two ends [172]. This interaction is thought to trigger their catalytic activity leading to their autophosphorylation as well as to the phosphorylation of other targets, e.g. the Artemis nuclease [205, 305]. Although the exact role of the DNA-PKcs subunit is

still unclear, its kinase activity seems to be required for effective NHEJ [242].

Following DNA-PKcs activation other NHEJ factors are also recruited to the break site. These include the Artemis nuclease, DNA polymerases μ and λ and finally the Xlf and Xrcc4-Ligase IV complex [241]. Artemis and the polymerases μ and λ, participate in the processing of non-ligatable DSBs, by removing secondary structures (e.g. hairpins generated during V(D)J recombination [205]) and filling-in the gaps between non-complementary ends [253].

In the final step of NHEJ the two ends are joined by the Xrcc4-Ligase IV complex [63]. Ligase IV contains a catalytic domain at the N terminus and two BRCT domains in the C-terminal part [349]. Xrcc4 binds to Ligase IV between the BRCT domains stabilizing the complex [117]. Although Xrcc4 does not seem to have an enzymatic function, its presence is necessary for the activity of the complex. An additional factor, termed Xlf (Cernunnos) also participates in the ligation step [3, 35]. While its exact function is not entirely clear, it has been reported to stimulate the activity of the Xrcc4-Ligase IV dimer [329].

Additional factors, including 53BP1 and the MRN complex participating in the DNA damage response signaling, were proposed to modulate the NHEJ pathway [230, 345, 361]. 53BP1 has been identified as one of the proteins binding to the tumor suppressor protein p53, hence the name [141]. It is one of the first factors recruited to the site of the DNA break, forming so called nuclear foci, which colocalize with phosphorylated H2AX (γH2AX) [269]. In addition to its role in mediating the DNA damage response and checkpoint signaling, 53BP1 was shown to promote repair through NHEJ by inhibiting DNA end resection, a process generating long ssDNA tails, which are poor NHEJ substrates [383]. It was also proposed to participate in keeping the two sides of the break in close proximity, thus facilitating the ligation [80]. The MRN complex (MRX in yeast), composed of Mre11,

Rad50 and Nbs1 (homolog of yeast Xrs2), participates in the earliest steps of DSB sensing as well as in the transmission of the signal to the checkpoint signaling and DNA repair pathways (reviewed in [67, 174]). MRN was shown to participate in the choice of the DSB repair pathway by regulating the DNA end resection step [237, 285]. It has been hypothesized that in the late S and G2 phase MRN, together with other factors, stimulates end resection, reducing the efficiency of NHEJ, while in G1 it prevents extensive resection, promoting NHEJ [82].

1.5.2.7.2 Homologous recombination (HR)

Homologous recombination (HR) is regarded as the most accurate mechanism used by eukaryotic cells to repair DSBs. It is the main DSB repair pathway in yeast [304]. In higher eukaryotes it occurs almost exclusively at the end of the S-phase and in G2-phase of the cell cycle, when the genetic material is already replicated and homologous template DNA is abundant [278]. This form of recombination is also important during meiosis, e.g. in the process of crossing-over between homologous chromosomes, in the final steps of interstrand crosslink (ICL) repair, and at collapsed replication forks [282, 69, 319].

The main initial steps of HR upon DSB detection include: 1) resection of 5' DNA ends to create 3' ssDNA tails, 2) coating of the ssDNA end with the Rad51 recombinase, 3) invasion of the coated strand into a homologous DNA molecule and formation of a D-loop, a joint molecule between the broken DNA and the repair template, and 4) extension of the 3' end of the invading strand by a DNA polymerase. Once these first steps are completed the process proceeds through one of the two possible sub-pathways: DSBR or synthesis-dependent strand annealing (SDSA) [240, 319]. The DSBR mechanism involves the formation of two Holliday junctions (HJs), which after resolution give rise to either cross-over or non-crossover products. In the SDSA pathway repair occurs without the formation of HJs. Instead the D-loop dissociates and the newly synthesized strand anneals

to the second broken end. This is followed by DNA synthesis and ligation, likely mediated by the DNA replication enzymes - polymerase δ and Ligase I [112, 211]. This process, resulting in non-crossover products and gene conversion, appears to be the most common type of HR repair in mammalian cells [150].

Unlike NHEJ, HR requires extensive 5' to 3' DNA end resection. As mentioned in the previous section, this step is mediated by the MRN complex. However, since MRN does not possess the 5'-3' exonuclease activity, additional factors are required to catalyze this reaction. One of the key factors cooperating with MRN is the carboxy-terminal binding protein (CtBP)-interacting protein (CtIP) [285, 370]. CtIP is a homologue of *S. cerevisiae* Sae2, which together with the MRX complex mediates DNA end processing in yeast [55]. After the initial resection by MRN and CtIP, more extensive DNA resection is carried out by exonuclease 1 (Exo1) aided by the Bloom syndrome (BLM) helicase and the Dna2 nuclease [226, 246, 245]. In a recently proposed bidirectional model of end resection, the MRN/CtIP complex first creates a nick upstream of the DSB and catalyzes 3'-5' resection towards the break [43, 103]. Next, Exo1/Dna2 extend the resection in the 5' to 3' direction.

The long 3' ssDNA tails generated in the resection process are rapidly bound by the ssDNA-binding RPA [313]. In subsequent steps, RPA is displaced by the Rad51 recombinase [19]. Rad51 is a homologue of *E.coli* RecA protein, the first identified recombinase and an essential component of bacterial DNA repair. In mice, the disruption of the Rad51 gene leads to early embryonic lethality, emphasizing its cardinal role in mammalian HR [330]. Like RecA, Rad51 binds ssDNA ends forming long helical nucleoprotein filaments [19, 22]. The DNA-bound Rad51 searches for sequence homology along the dsDNA and finally mediates pairing between the two complementary strands.

The search for sequence homology and pairing is aided by a collection of other proteins, including Rad54/Rad54B and the five Rad51 paralogs: Rad51B, Rad51C, Rad51D, and

X-ray repair cross-complementing 2 and 3 (Xrcc2 and Xrcc3) [217, 317]. Among the factors involved in this process are also Rad52 and the breast cancer susceptibility proteins 1 and 2 (Brca1 and Brca2) [146, 233, 272]. Mutations in Brca1 and Brca2, undoubtedly among the best known HR genes, are widely associated with predisposition to breast and ovarian cancers [97, 175, 224, 239, 357].

Rad54 is a dsDNA-specific ATPase which binds directly to Rad51 and stimulates the ATP-dependent translocation of the Rad51 nucleofilament along the DNA duplex in search of homology [126]. Rad54 was also shown to stimulate the Rad51-mediated formation of the D loop [256]. A paralog of Rad54, named Rad54B, was also identified [219]. Both proteins have similar biological functions, suggesting that Rad54B may serve as backup in case of Rad54 absence [350].

The Rad51 paralogs have limited homology to the Rad51 and to each other (approx. 20-30%) and likely originated from a series of duplication events of the Rad51 gene [190]. They form two types of complexes, which act at different stages of the HR pathway [197, 217]. One of the complexes, termed BCDX2, comprises Rad51B, Rad51C, Rad51D and Xrcc2. It was proposed to take part in the formation and/or stabilization of the Rad51-DNA filament. The second complex, composed of Rad51C and Xrcc3 (CX3), seems to play a role in Holliday junction branch migration and resolution.

Rad52 in yeast plays the role of a mediator between the DSB and the recombination machinery by delivering Rad51 to ssDNA ends [252]. Until recently, mammalian Rad52 homolog did not seem to have a similar mediator function. Instead, another protein – Brca2, was believed to play this role [194]. Brca2 is a large protein with several BRC domains, which serve to bind Rad51. Its main function is Rad51 delivery and loading onto the ssDNA tails [68]. However, it was also proposed to prevent Rad51 binding to dsDNA and to help in RPA displacement (reviewed in [148]). While in mammals Brca2

seems to be the key factor responsible for Rad51 nucleofilament assembly, Rad52 was also recently reported to play a role in this process [92]. This suggested the existence of two separate pathways, one utilizing Rad52 as a mediator like in yeast, the other relying on Brca2. In the absence of a functional Brca2, Rad52 may take over its function, explaining why Rad52 inactivation is synthetically lethal in Brca2 null background. The other breast cancer susceptibility gene, Brca1, is also required for functional HR. Brca1 was reported to function as a heterodimer with Bard1, a protein necessary for Brca1 stability [151, 358]. The Brca1/Bard1 dimer is thought to play a role in the regulation of the choice between HR and NHEJ, similarly to 53BP1 (reviewed in [48]). Although the molecular mechanism of its action is unclear, Brca1 was proposed to stimulate HR by counteracting the anti-resection activity of 53BP1 [36].

Apart from repairing DSBs arising during the mitotic cell cycle, HR also plays a very important role in meiotic recombination, a process that permits to correctly align and segregate homologous chromosomes during meiosis. In the first stage of this process Spo11, a topoisomerase-related protein, introduces DSBs along the chromosomes [158]. Subsequent DNA end processing is identical to that of regular HR. However, the resulting ssDNA tails are bound by two recombinases: Rad51 and Dmc1. Dmc1 is a meiosis-specific paralog of Rad51 [26]. Both proteins share over 50% sequence identity, suggesting that the Dmc1 gene was generated by duplication of the Rad51 gene early in eukaryotic evolution [190]. Unlike Rad51 mutants, Dmc1 knock-out mice are viable, but adult animals are sterile, underlining the essential role of Dmc1 in meiosis [260]. Both Rad51 and Dmc1 play a role in formation of nucleoprotein filament and homology search during meiotic recombination. However, Rad51 was shown to play only an accessory function, while Dmc1 is absolutely crucial for this process [56]. The remaining steps of meiotic recombination resemble those of mitotic HR, although there is a higher frequency of cross-over products

than that observed in mitosis, due to the more frequent use of the DSBR pathway [8].

1.5.2.7.3 Microhomology mediated end-joining (MMEJ)

It is becoming increasingly clear that NHEJ and HR are not the only mechanisms of DSB repair in eukaryotic cells. Over the past 20 years, many groups reported the existence of alternative repair pathways, which come into play when the two main repair processes are absent or dysfunctional. This third group of still poorly characterized DSB repair pathways was collectively termed microhomology-mediated end joining (MMEJ), also referred to as alternative or backup NHEJ (alt-NHEJ or B-NHEJ) [12, 28, 73, 83, 108, 248, 254]. Here the term MMEJ will be used to emphasize the characteristic feature of this pathway, i.e. the use of 5-25 bp microhomologies during the alignment of broken DNA strands before joining [221].

MMEJ occurs throughout the cell cycle and is independent of core NHEJ and HR factors [28, 184, 204, 372]. However, it was reported to share resection steps with HR, indicating that it may partially rely on HR enzymes [72, 83, 184, 204, 328]. Indeed, MMEJ initiation was shown to require MRN and CtIP [204, 374]. Other factors that have been proposed to participate in MMEJ include PARP1, the Ligase III/ Xrcc1 complex, Ligase I, DNA polymerase θ, DNA polymerase δ, and the ERCC1/XPF complex [12, 184, 204, 372]. However, this is unlikely to be an exhaustive list of implicated proteins.

It has been suggested that in the absence of other DNA-end binding proteins (like Ku or Rad51) the DSBs are recognized by PARP1 which then initiates their repair through MMEJ [221]. The repair process, similarly to HR, starts with 5' to 3' end resection, which exposes short regions of homology on each side of the break. This processing step is mediated by the MRN/CtIP complex [229]. The complementary regions, present in the 3' ssDNA fragments, pair together and the non-complementary segments ('flaps') are removed [372]. The ERCC1/XPF complex was proposed to be responsible for mediating

flap removal [2, 204]. Gaps (if any) are then filled in by a polymerase (e.g. DNA polymerase θ or δ) and breaks joined by the Ligase I or Ligase III/Xrcc1 complex [184, 372].

In the absence of immediate microhomology regions at the DNA ends, which is often the case, a more distant fragment of the repaired molecule can be copied using a low fidelity DNA polymerase (e.g. polymerase θ). This duplicated region then participates in the alignment of DNA ends, which results in an insertion in the created junction, termed a templated insert. This more complex variant of microhomology-mediated repair has been termed synthesis-dependent MMEJ (SD-MMEJ) [372].

MMEJ is an error-prone and highly mutagenic mechanism, due to large deletions of genomic sequence characteristic of this pathway. While it seems that the use of this mechanism is limited in normal cells, it was shown to be very active in many types of cancer [323]. Indeed, cancer cells were shown to preferentially use this pathway of DSB repair, likely due to deficiency in the main repair mechanisms [23]. This may potentially make MMEJ a therapeutic target for treatment of cancers resistant to current therapies.

Although MMEJ is generally regarded as an alternative DSB repair pathway, it was recently shown to play a role in class switch recombination and V(D)J joining, which are processes previously thought to rely exclusively on NHEJ [28, 108]. This suggests that this mechanism may be more than a backup pathway. It is also possible that some DSBs, e.g. incompatible overhangs, which are poor NHEJ and/or HR targets, might be more efficiently repaired by MMEJ [380]. This opens the possibility that this repair pathway may play a more important role than initially suspected.

Chapter 2

Aim of the thesis

Untargeted integration of plasmid DNA into the mammalian cell genome is a process widely used in molecular biology and biotechnology, for instance to produce cells expressing recombinant proteins. However, it is relatively inefficient and still remains incompletely understood. While it was proposed to be mediated by DSB repair mechanisms, such as NHEJ, HR or MMEJ, which of these pathways may be primarily responsible for this process has remained unclear. A better knowledge of the molecular basis of spontaneous plasmid integration may help to make this process more efficient and less time consuming, facilitating its use in gene therapy and biotechnology. It may also allow to decrease the unwanted non-specific integration of exogenous DNA, facilitating genome editing. Therefore, one of the main goals of this study was to identify and characterize the mechanisms primarily responsible for foreign DNA integration into the mammalian cell genome.

Previous studies performed in our group demonstrated that MAR elements can stimulate integration of plasmid DNA into the CHO genome [111, 110]. In addition, we observed that successive transfections with MAR-containing vectors resulted in a very high increase of transgene expression, possibly due to enhanced recombination between individual plasmid copies leading to the formation of long concatemers [116]. Together these observations led to the hypothesis that MAR elements could play a role in transgene integration by

stimulating DNA recombination. Here, we further explored this connection between MARs and the DNA recombination machinery to determine which DNA repair pathway may be activated by these elements.

In addition to investigating the role of DSB repair pathways in transgene integration, we also sought to better understand these DNA recombination mechanisms and their link to the cell cycle. Therefore, we designed recombination assays to study the function of different variants of the MMEJ pathway, and we constructed cell cycle-sensitive reporter CHO cells to investigate the potential role of DNA repair proteins in the cell cycle regulation.

Chapter 3

The analysis of two CHO mutants deficient in the activity of Rad51D and DNA-PKcs reveals additional defects in DNA repair gene expression

This chapter includes data generated by Mélanie Grandjean and Samuel Neuenschwander.

Mélanie Grandjean performed experiments represented on Fig.3.2 and published in: Grandjean, M., Girod, P.-A., Calabrese, D., **Kostyrko, K.**, Wicht, M., Yerly, F., Mazza, C., Beckmann, J.S., Martinet, D., Mermod, N. High-level transgene expression by homologous recombination-mediated gene transfer. Nucleic Acids Res. 2011 Aug; 39(15):e104.

Samuel Neuenschwander performed the RNA-seq data analysis (Fig.3.5,3.6; Table 3S1,3S2).

3.1 Abstract

Chinese hamster cell lines sensitive to ionizing radiation and DNA damaging agents have been used for many years to characterize the components of the DNA double strand break (DSB) repair machinery in mammalian cells. Here, we used two such mutant cell lines, deficient in the activity of DNA-PKcs, and important non-homologous end-joining (NHEJ) enzyme, or Rad51D, one of the homologous recombination (HR) factors, to study the connection between DSB repair and matrix attachment regions (MARs). MARs are DNA elements proposed to play a role in chromatin organization, transcription and replication. Previous work from our group also suggested that MARs stimulate foreign DNA integration into the genome of Chinese hamster ovary (CHO) cells, opening the possibility that they may play a role in DNA recombination. Here, we showed that MAR-mediated transgene integration is enhanced in the DNA-PKcs mutant, but impaired in Rad51D-deficient cells. This suggested that MAR elements require a functional HR pathway for efficient transgene integration. However, the use of another Chinese hamster cell line, deficient in the activity of an important HR factor - Brca2, failed to support this hypothesis. Therefore, we performed whole transcriptome sequencing of the two CHO cell lines, which revealed differences in the expression of other DSB repair genes. We conclude that these mutant cell lines are not the best system to study DNA recombination and repair, and that other, more controllable tools should be used instead.

3.2 Introduction

Random genomic transgene integration is a problem in gene therapy and biotechnology, as it can lead to gene silencing, mutagenesis or inefficient expression of the target gene. Possible solutions to these problems include the use of strong or inducible promoters, gene targeting, or flanking the target gene with matrix attachment regions (MARs). MARs are polymorphic sequences in eukaryotic chromosomes that were first described in 1984, and termed scaffold-attachment regions (SARs), for their ability to anchor the chromatin to the nuclear matrix or scaffold. Upon their discovery, MARs were proposed to play a role in the organization of chromatin into higher-order structures by dividing chromosomes into functionally independent domains [215]. Later studies also showed that they may act as insulators, i.e. borders between active and inactive chromatin, and can be used to shield genes from silencing [106, 125]. Due to this property they have been successfully used for many years in biotechnology to sustain and stabilize recombinant gene expression [5, 106, 163, 377]. Apart from their role in chromatin organization, MARs were also proposed to participate in DNA replication and regulation of gene expression [5, 71, 259]. Studies from our group also suggested that MARs stimulate transgene integration, opening the possibility that they may play a role in DNA recombination [111, 110].

In eukaryotic cells, transgene integration requires cellular factors and it is believed to be triggered by DNA double stranded breaks (DSBs). The two major pathways responsible for DSB-repair in eukaryotic cells are non-homologous end-joining (NHEJ) and homologous recombination (HR). These pathways were characterized largely due to the use of cell lines sensitive to ionizing radiation or DNA-damaging chemical agents. Most of these cells were derived from Chinese hamster and were found to be deficient in HR or NHEJ components.

Here, we used some of these Chinese hamster cell lines to study the possible link between the MAR elements and DSB repair mechanisms. We first employed two Chinese

hamster ovary (CHO) mutants defective in the activity of DNA-PKcs and Rad51D, factors important for NHEJ and HR, respectively, and analyzed transgene integration and expression in these cells in the presence or absence of the human MAR 1-68, previously identified by our group [110]. These experiments demonstrated that MAR-mediated transgene integration was enhanced in the absence of DNA-PKcs, suggesting that MARs may promote integration by stimulating a mechanism alternative to NHEJ. In contrast, plasmid integration was largely impaired in the absence of Rad51D, indicating that a homology-dependent mechanism may be implicated in this process.

In parallel, we also used a different Chinese hamster mutant deficient in Brca2, another important HR factor. In contrast to Rad51D-deficient cells, transgene integration and expression was not abolished in these cells and instead was significantly increased compared to the control cells. This led us to investigate further the Rad51D and DNA-PKcs mutant cell lines. Plasmid recombination assays performed in these HR- and NHEJ-deficient cells failed to demonstrate a defect in HR and NHEJ activities, respectively. In addition, whole transcriptome sequencing of these cells revealed additional defects in the expression of several DNA repair genes. Therefore, we conclude that these two mutant cell lines are not a reliable system to study DNA recombination and that observations made in these, and likely many other, mutant cell lines should be treated with reserve. We propose that a different, more reliable system should be used to study the role of MAR elements in DNA recombination.

3.3 Materials and Methods

Vectors

The control GFP vector (pGEGFP) contains the SV40 early promoter, enhancer and vector backbone from pGL3 (Promega) driving the expression of the eGFP gene from

pEGFP-N1 (Clontech). The vector containing the human MAR 1-68 (p1-68-SV40EGFP) has been created by inserting the MAR into the pGEGFP upstream of the SV40 early promoter. The pGL3-CMV-DsRed vector was created by inserting into the pGL3-basic (Promega) the DsRed gene, together with the CMV promoter and enhancer, coming from pCMV-DsRed (Clontech). The HR and NHEJ reporter plasmids were kindly provided by V. Gorbunova (University of Rochester, New York, USA). Their construction has been described in detail elsewhere [212, 291].

Cell culture

Wild type CHO AA8 cells, mutant V3.3 [27] and 51D1 cells [128] were cultivated in DMEM/F-12+GlutaMAX^{TM} with 10% fetal bovine serum (Gibco, Invitrogen) and antibiotic-antimycotic solution (Sigma-Aldrich, #A5955). The same medium was used for Chinese hamster Brca2-deficient VC8 cells and Brca2-complemented VC8-Brca2 cells [169].

Transfections

Transfections were performed using Lipofectamine 2000 (Invitrogen) or Fugene 6 (Promega), according to the manufacturer's instructions. Transfection efficiency was monitored by fluorescence microscopy (Carl Zeiss Microscope Axio Observer.A1). Stably transfected cells expressing GFP were obtained by cotransfection of the pGEGFP or p1-68-SV40EGFP vectors with the resistance plasmid pSVpuro (Clontech) using Lipofectamine 2000 (Invitrogen). After two weeks of selection with 5 μg/ml puromycin cells were analyzed by FACS.

FACS analysis

For fluorescence-activated cell sorting (FACS) cells were harvested 24h following trans-

fection or after 2 weeks of selection, and resuspended in PBS with 2% FBS (Gibco, Invitrogen). Fluorescence acquisition was performed on the FACScalibur flow cytometer (Becton Dickinson) or the CyAn ADP Analyzer (Beckman Coulter) and data was analyzed using the FlowJo software (Tree Star).

Recombination assays

For HR and NHEJ recombination transient assays, cells were transfected with HR or NHEJ reporter plasmids digested with I-SceI, and with the pGL3-CMV-dsRed plasmid to normalize for transfection efficiency, using Fugene 6 (Promega). The pGEGFP vector was transfected in parallel as a positive control of GFP expression. Expression of GFP and DsRed was monitored by fluorescence microscopy (Carl Zeiss Microscope Axio Observer.A1) and analyzed by FACS. GFP repair efciency was calculated as a ratio of GFP-positive cells over the number of dsRed-positive cells.

PCR and quantitative PCR

To determine transgene copy number total genomic DNA was isolated from cells using the DNeasy purification kit (Qiagen). For quantitative PCR (qPCR), 6 ng of genomic DNA were analyzed using the SYBR Green I Master kit for the Light Cycler 480 machine (Roche) using primers EGFP-2F (AGCAAAGACCCCAACGAGAA) and EGFP-2R (GGCGGCGGTCACGAA). The beta-2-microglobulin (B2M) CHO gene was used as a normalization control using primers B2M_F (ACCACTCTGAAGGAGCCCA) and B2M_R (GGAAGCTCTATCTGTGTCAA). The number of integrated GFP copies was calculated using the B2M gene as a reference, as described previously [257].

To determine the presence of the Rad51D and DNA-PKcs transcripts in 51D1 and V3-3 cells, respectively, total RNA was isolated from cells using the NucleoSpin® RNA II kit

(Macherey-Nagel) and cDNA was generated using First-Strand cDNA Synthesis Kit (GE Healthcare). The full CDS of hamster Rad51D was amplified using primers CgR51D_F (AAACCATGGAAACATGGGCGTGCTCAGGG) and CgR51D_R (AAATCTAGAG-CATCATGTCTGTTTGGCAG). The final 1kb of hamster DNA-PKcs CDS was amplified using primers DNA-PKcs_F (GTGTCATGCCCATGACCT) and DNA-PKcs_R (TTA-CATCCAAGGCTCCCA). The same primers were used for Sanger sequencing of this fragment of DNA-PKcs mRNA.

Transcriptome sequencing (RNAseq)

For RNA sequencing total RNA was isolated from AA8, V3-3 and 51D1 cells using the NucleoSpin® RNA II kit (Macherey-Nagel). Libraries were prepared and sequenced by paired-end sequencing using the Illumina technology by the Genomic Technologies Facility of the University of Lausanne. Transcriptome analysis was performed by the Vital-IT computer facility at the Swiss institute of Bioinformatics in Lausanne. The reads where mapped to the annotation derived from CHO-K1 (CriGri_1.0, GCF_000223135.1) allowing for max. 2 mismatches using Bowtie [177]. The number of reads mapping on each transcript were counted and then compared among the conditions using DESeq [7], the calculated p-values were adjusted for multiple testing with the Benjamini-Hochberg procedure [20].

3.4 Results

3.4.1 Transgene integration is impaired in Rad51D-deficient cells, and increased in DNA-PKcs mutant in the presence of the MAR

To study the role of MAR elements in DNA recombination we used two CHO mutants, the V3-3 and 51D1 cells [27, 128, 351]. The 51D1 cell line was generated by knock-out of exon 4 in the gene coding for Rad51D, one of the Rad51 paralogs important for HR. We confirmed the absence of the full Rad51D transcript in 51D1 cells by PCR (Fig.3.1A). The second cell line, V3-3, was reported to be deficient in the activity of DNA-PKcs, an enzyme indispensable for NHEJ. These cells were obtained by EMS (Ethyl methanesulfonate) mutagenesis and still express the DNA-PKcs mRNA (Fig.3.1B). However, they were reported to contain a nonsense mutation in one of the DNA-PKcs alleles which gives rise to a truncated protein [264]. Sequencing results confirmed that half of the DNA-PKcs transcripts present in V3-3 cells contain a CAA to TAA mutation, potentially resulting in a protein lacking 104 amino acids at the C-terminus (Fig.3.1C).

We transfected these cells with a GFP expression vector containing a human MAR 1-68 or a MAR-devoid control, and analyzed the integration and expression of these plasmids. Stable GFP expression from the control plasmid was similar in the wild-type parental cells (AA8) and the DNA-PKcs V3-3 mutant (Fig.3.2A). However, the Rad51D-deficient 51D1 cells showed a decrease in survival upon antibiotic selection as well as diminished GFP fluorescence. The number of integrated transgene copies correlated well with GFP fluorescence, showing a similar decrease in cells lacking Rad51D (Fig.3.2B). Taken together, these results suggested that NHEJ may not be responsible for transgene integration, while the lack of a functional HR, may negatively influence this process.

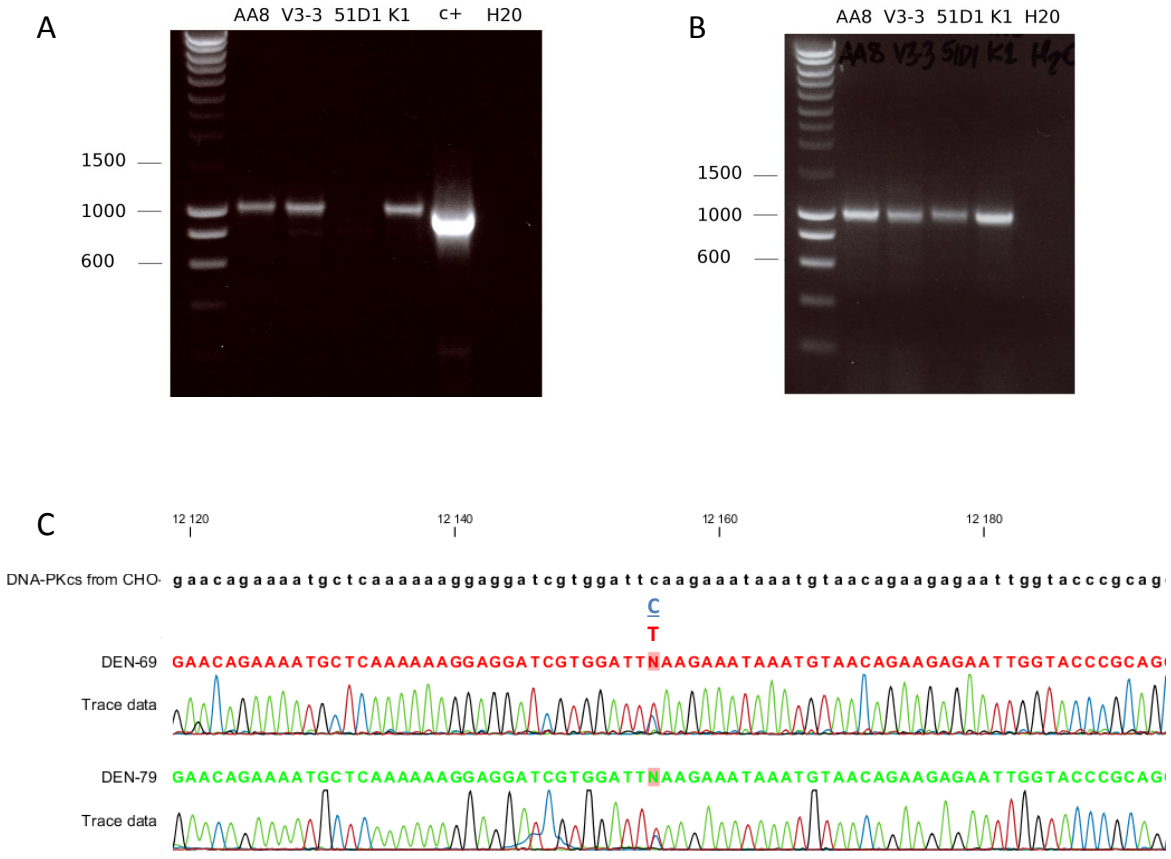


Figure 3.1: Analysis of Rad51D and DNA-PKcs transcripts in 51D1 and V3-3 cells. A) PCR of the Rad51D CDS (1015 bp) in RNA extracts from AA8, V3-3, 51D1 and CHO-K1 (K1) cells. CgRad51D from CHO-K1 cloned into a pSV40 vector served as positive control (c+). B) PCR of the final 1013 bp from the DNA-PKcs CDS in CHO cell extracts. C) Fragment of the DNA-PKcs CDS sequencing result from the V3-3 cells. Forward sequencing reaction in green, reverse in red. Results were aligned to CgDNA-PKcs (ref: XM_003509456.1).

The presence of the MAR resulted in a 3-fold increase in GFP integration and expression in wild type cells, confirming previous observations. Interestingly, in the NHEJ mutant this effect was enhanced, resulting in a 5- to 6-fold increase in GFP integration and expression when compared to AA8 cells transfected with the vector without the MAR. We therefore hypothesized that in the absence functional NHEJ, an alternative repair pathway may cooperate with the MAR to mediate very high transgene integration. In contrast, the presence of the MAR did not improve GFP expression or integration in 51D1 cells. Also in this case the number of antibiotic-resistant colonies remained very low. We therefore concluded that the MAR element may mediate transgene integration, which becomes

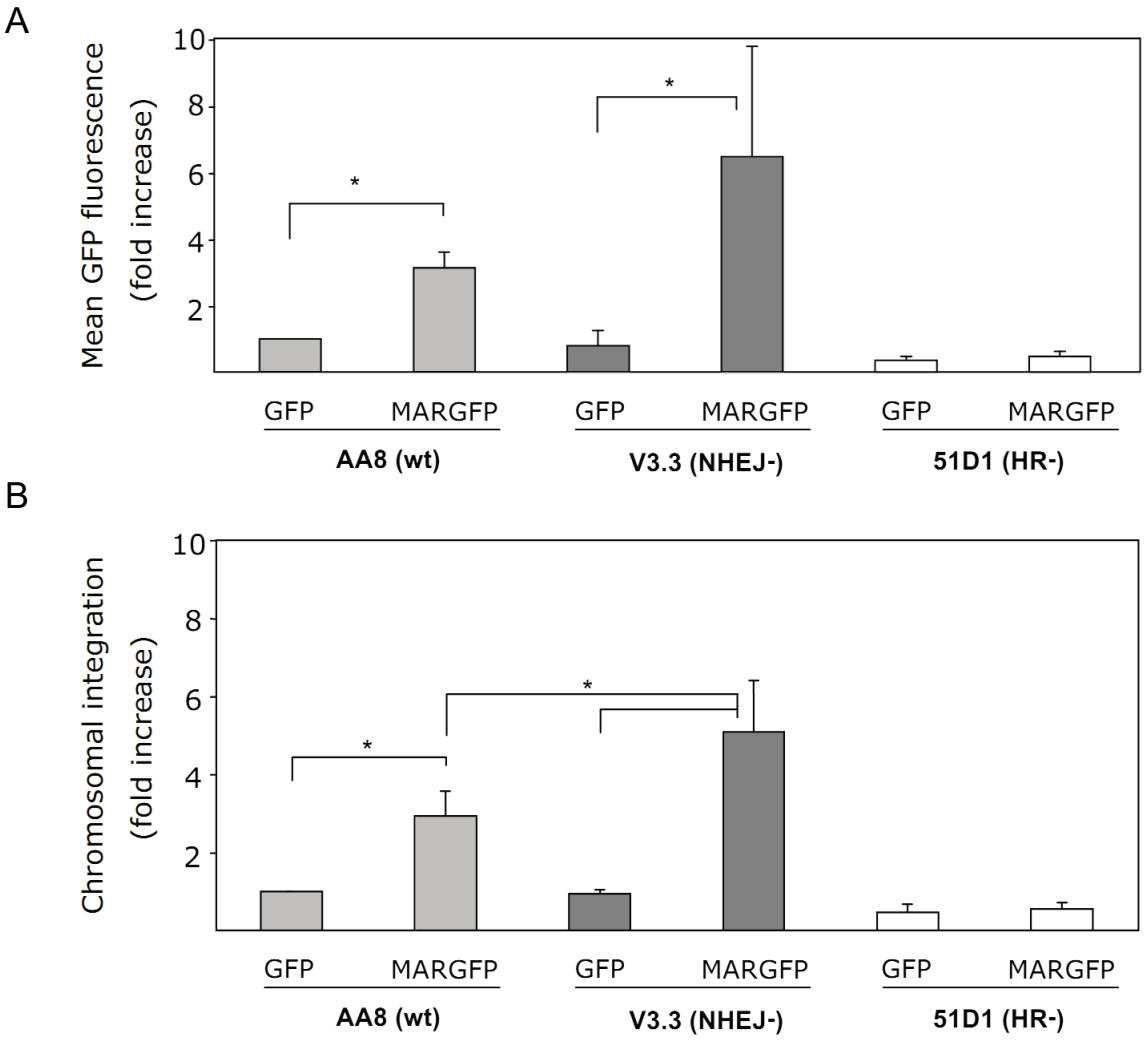


Figure 3.2: MAR-mediated transgene integration and expression is enhanced in V3-3 cells and abolished in 51D1 cells. Parental CHO cells (AA8) and mutants deficient in DNA-PKcs (V3.3) or Rad51D (51D1) were transfected with a pSV40-GFP (GFP) or the same plasmid with the MAR 1-68 (MARGFP). A) Relative mean GFP fluorescence, B) relative GFP transgene copy number. Bars represent mean fold change over the AA8 cells transfected with pSV40-GFP, s.e.m. error bars, n≥5. Asterisks indicate significant differences; (*) $p<0.05$, (**) $p<0.01$, unpaired Student's t-test. Data are from Mélanie Grandjean [116].

more efficient in the absence of NHEJ. This effect was almost completely abolished in HR deficient cells, suggesting that this pathway may be important for transgene integration in CHO cells.

3.4.2 Lack of Brca2 affects expression from genome integrated plasmids

We subsequently tested the involvement of the MAR in DNA recombination and its impact on transgene integration in another HR mutant – the Brca2-deficient VC8 cell line [169, 340]. The VC8 cells were derived from Chinese hamster lung fibroblast cell line V79 by ENU (N-ethyl-N-nitrosourea) mutagenesis. They were later found to contain premature termination mutations in both alleles of the Brca2 gene [353].

Surprisingly, the lack of functional Brca2 did not affect GFP expression or integration (Fig.3.3A,B). In addition, in the presence of the MAR 1-68, Brca2 deficiency very significantly increased expression, indicating that this defect in HR may not impair transgene genomic integration and expression but rather enhance it. Furthermore, the presence of the MAR in the plasmid resulted in higher transgene expression both in cells lacking Brca2 and Brca2-complemented (approx. 17-20 fold). This increase was much higher than ever observed with MAR 1-68 in any of the other CHO cell lines. Interestingly, GFP fluorescence in Brca2-deficient cells transfected with the MAR-containing plasmid was 2-fold higher than in cells expressing Brca2, despite the lack of significant difference in the number of integrated GFP copies between the two cell lines. Instead, GFP expression per gene copy was significantly increased in this condition (Fig.3.3C). This may suggest that the absence of Brca2 does not affect the MAR-mediated increase in transgene integration, but that it may possibly direct the transgenes to more favorable genomic loci, improving their expression. Taken together it seems that the absence of Brca2, an important HR

component, does not impair transgene integration. Moreover, lack of this protein seems to have a beneficial effect on the integration and expression of the transgene. These results were in contrast with the experiments performed in Rad51D-deficient cells. Therefore, we next tested the efficiency of HR and NHEJ in the 51D1 and V3-3 CHO mutants and we performed their whole transcriptome analysis to assess if the expression of DSB repair genes is altered in these cells.

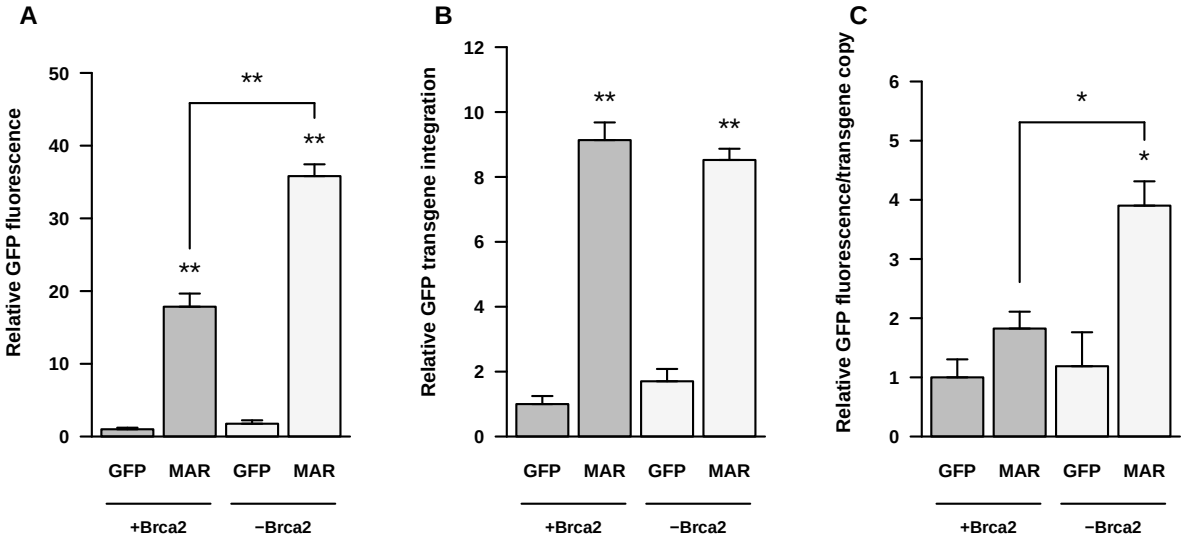


Figure 3.3: Effect of Brca2-deficiency on plasmid genomic integration and expression. Brca2-deficient (-Brca2) and complemented (+Brca2) cells were transfected with a pSV40-GFP (GFP) or the same plasmid with the MAR 1-68 (MAR). Results are shown as fold increase over the data obtained from Brca2-complemented cells transfected with the GFP plasmid without the MAR. Mean of 4 experiments. Asterisks indicate significant differences between the MAR sample and the corresponding GFP control or between the two MAR samples; (*) $p<0.05$, (**) $p<0.01$, unpaired Student's t-test. A) mean GFP fluorescence, B) GFP copy number, and C) fluorescence per GFP copy.

3.4.3 CHO 51D1 and V3-3 mutant cells are not fully deficient in HR and NHEJ activities

To verify the activity of DSB repair pathways in CHO cells, we used two plasmid assays based on the reconstitution of a functional GFP gene from two non-functional fragments by either HR or NHEJ (Fig.3S1A,B) [212, 291]. Digestion of the two vectors with the

I-SceI enzyme creates breaks imitating naturally occurring DSBs, which in case of the NHEJ reporter can be repaired by end-joining, yielding GFP expression. The HR cassette can produce a functional GFP only through inter- or intramolecular gene conversion, i.e. non-crossover and non-reciprocal gene exchange, which is the most frequent form of HR repair in mammals [150]. Without DNA cleavage and repair, the original cassettes do not express GFP.

In wild type AA8 cells, both reporters yielded GFP positive cells, although the frequency of GFP reconstitution through NHEJ was 3-fold higher than through HR, reflecting the predominance of NHEJ over HR in mammals (Fig.3S2). The frequency of HR was increased in the V3-3 mutant, albeit not significantly (Fig.3.4A). Surprisingly, we observed no change in HR in 51D1 cells. The efficiency of NHEJ was only slightly decreased in DNA-PKcs-deficient cells, and remained unaffected in the Rad51D mutant (Fig.3.4B). These results suggested that the 51D1 and V3-3 cells may not be fully deficient in HR and NHEJ activities, respectively. We thus hypothesized that the cells may possess secondary or compensatory mutations, which alter their recombination propensity, and/or that they may be mutated in other DNA repair pathways. Therefore, we proceeded to analyze the full transcriptome of these cell lines by whole RNA sequencing (RNAseq).

3.4.4 Genes mutated in 51D1 and V3-3 cells show decreased expression

We first compared the expression profiles of the two genes mutated in 51D1 and V3-3 cells using the parental AA8 cells as a reference. In 51D1 cells we observed a significant, over 70% reduction of the Rad51D mRNA compared to the wild type cells (Fig.3.5A, Table 3S1). A small number of detected hits matching Rad51D mRNA likely represents the sequencing reads mapping to the truncated transcript. Indeed, a marked reduction in the

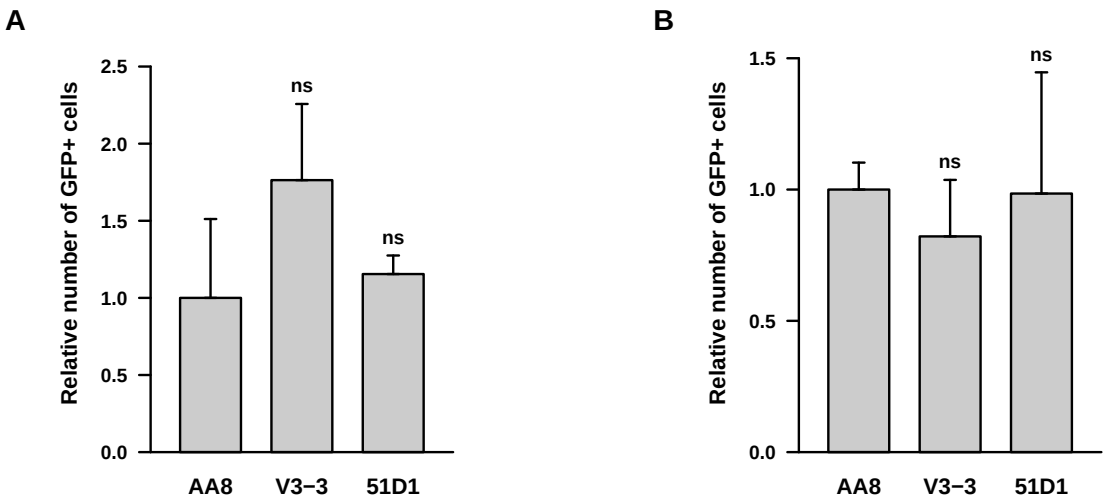


Figure 3.4: HR and NHEJ-mediated DSB repair in AA8, V3-3 and 51D1 cells. Bars represent mean fold change over AA8 cells transfected with A) the HR reporter, B) the NHEJ reporter. Mean of 3 experiments, s.e.m. error bars. Statistical signicance between the V3-3 or 51D1 cells and the AA8 cells was determined by unpaired Student's t-test, significance level $p<0.05$, ns – not significant.

transcript coverage was observed in the region corresponding to the exon 4.

In the V3-3 cell line we also observed a 50% reduction in the DNA-PKcs transcript level (Fig.3.5B, Table 3S1). Since an intact DNA-PKcs allele is still present in these cells, this decrease may result from partial degradation of the transcripts containing the premature stop codon due to nonsense-mediated mRNA decay [34]. The detailed analysis of the sequences mapping to the V3-3 DNA-PKcs mRNA revealed two nonsense mutations at positions 10 135 and 12 082 present in half of the transcripts (Fig.3.5C,D). The second mutation confirmed our previous sequencing analysis and reports published by others. The earlier mutation, giving rise to a protein lacking 748 amino acids at the C-terminus, has not been reported previously to our knowledge. However, we were unable to determine whether these two mutations, potentially yielding truncated proteins, are on the same or two different DNA-PKcs alleles.

These results further confirm that the the 51D1 cells are deficient in Rad51D. They also demonstrate that DNA-PKcs expression is decreased in V3-3 cells. However, since these mutations did not seem to significantly affect the DSB repair properties of these

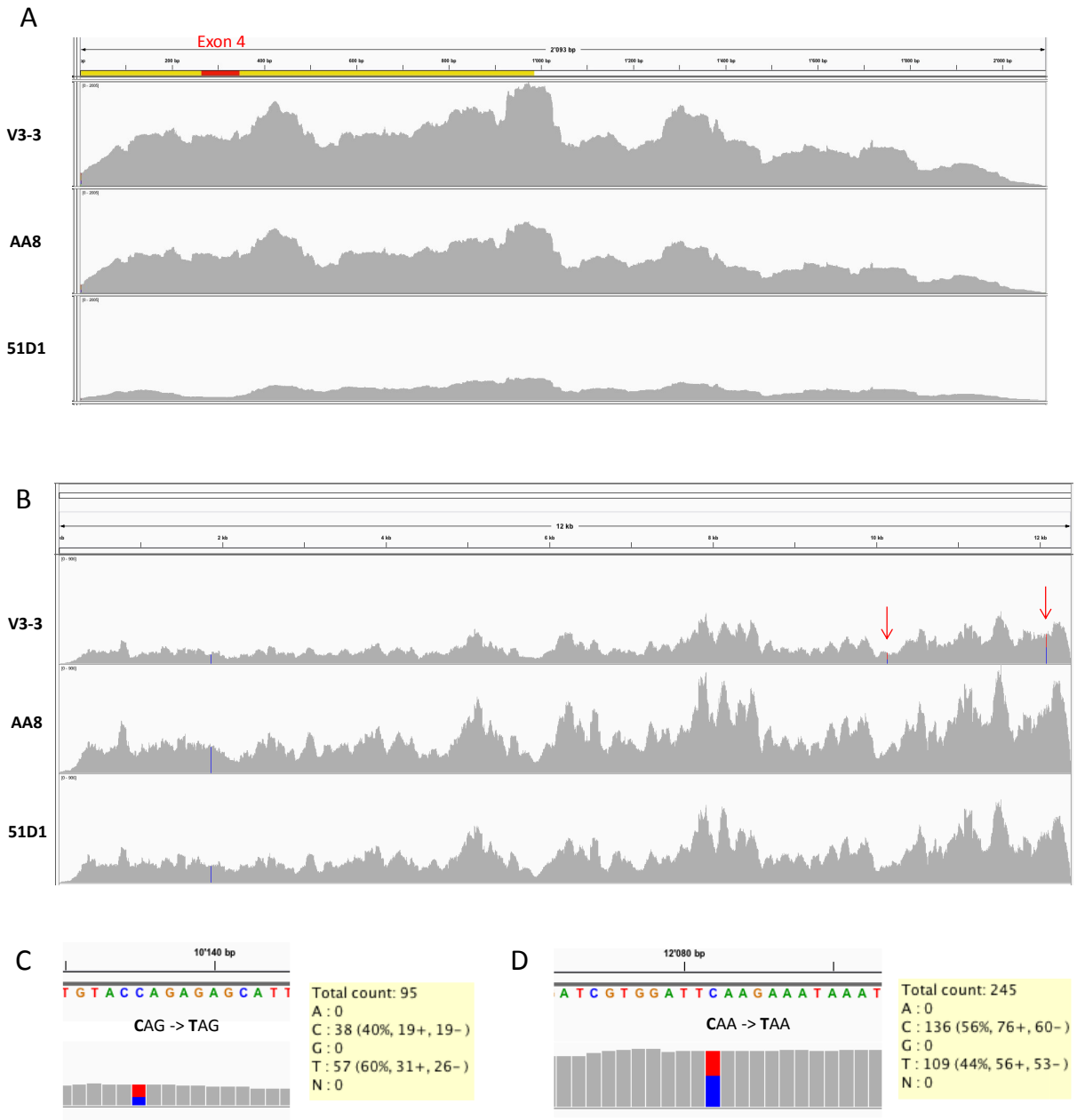


Figure 3.5: Rad51D and DNA-PKcs transcripts mapping coverage. A) Mapping of the RNAseq reads from AA8, V3-3 and 51D1 cells to the Rad51D transcript (XM_003495801.1). Yellow bar – Rad51D CDS, red bar - exon 4 knocked-out in the 51D1 cells. B) Mapping of the RNAseq reads from AA8, V3-3 and 51D1 cells to the DNA-PKcs transcript (XM_003509456.1). Red arrows point to nonsense mutations at positions 10 135 and 12 082 shown in detail in C) and D). Analysis from Samuel Neuenschwander (SIB).

cells, we decided to look at the expression of other repair genes to identify candidates that could potentially compensate for the genetic defects in these cells.

3.4.5 51D1 cells are potentially deficient in interstrand crosslink repair

We subsequently analyzed the expression of other DSB repair genes in 51D1 cells. Only the level of Rad52 mRNA was decreased compared to the control AA8 cells, albeit not significantly (Fig.3.6A, Table 3S1). However, we also observed a 30-60% increase in expression of several HR genes, most notably MDC1, Rad51, Nbs1, Xrcc2 and Xrcc3, as well as some genes attributed to other DNA repair pathways, e.g. Xrcc1, PARP1. Increased transcription of these genes, notably Rad51 and the other Rad51 paralogs - Xrcc2 and Xrcc3, could indicate that some of these factors may compensate for the lack of Rad51D in the HR pathway.

The 51D1 cells were shown to be very sensitive to γ-irradiation and UV light [128]. We also observed that they divided much slower than the AA8 or V3-3 cells, and were sensitive to transfection and antibiotic selection procedures (data not shown). Since the mutation in Rad51D does not seem to affect HR-mediated repair in these cells, we reasoned that altered expression of another gene may be responsible for their poor viability. We therefore examined all the genes that were significantly downregulated in the 51D1 cells compared to the other two cell lines (Table 3S2). Apart from Rad51D, we found significant decrease in expression of many genes important for embryogenesis, neuronal development or keratinocyte differentiation. However, all these factors are likely not essential for the survival of cultured cells. Interestingly, we also observed an over 10-fold decrease in the expression of Fanconi anemia complementation group L (Fancl), the E3 ubiquitin ligase crucial for the Fanconi anemia (FA) pathway. The FA pathway is responsible for the

removal of DNA interstrand crosslinks (ICLs), lesions where two nucleotides from opposing DNA strands are covalently bound together. This type of damage prevents DNA strand separation and therefore is especially harmful during replication and transcription. It has been estimated that mammalian cells encounter on average 10 ICLs per day [118]. While this may not appear to be a large number, it was shown that as little as 17-40 ICLs can be lethal for repair-deficient cells [179]. Therefore, a lack of an essential component of FA pathway may make these cells very susceptible to damage. Since the HR pathway is important for the downstream steps of ICL repair, the combined Fancl and Rad51D-deficiency may have a synergistic effect, potentially explaining the genetic instability and low viability of these cells.

3.4.6 Genes involved in alternative DSB repair pathways show higher expression in V3-3 cells

The lack of one functional DNA-PKcs allele was shown to strongly influence NHEJ-mediated DSB repair and V(D)J recombination in V3-3 cells [27]. However, we only observed a small decrease in NHEJ events as assessed by the GFP-reconstitution plasmid assay. Since this assay was designed to tolerate a wide range of insertions and deletions [212], we hypothesized that it likely detects repair events resulting from classical NHEJ as well as alternative end-joining, also referred to as alternative NHEJ or microhomology-mediated end-joining (MMEJ). Since this pathway was reported to be more active in NHEJ-deficient cells [28, 72, 75, 152, 186], we hypothesized that the efficient GFP repair in V3-3 cells could be due to MMEJ. Indeed, we observed an up to 2-fold increase in the expression of several genes reported to play a role in MMEJ, e.g. Ercc1, Xrcc1, Ligase III, PARP1, Pold3, Nbs1 (Fig.3.6B, Table 3S1). While these differences were not statistically significant, they seem to be in line with previous studies reporting elevated

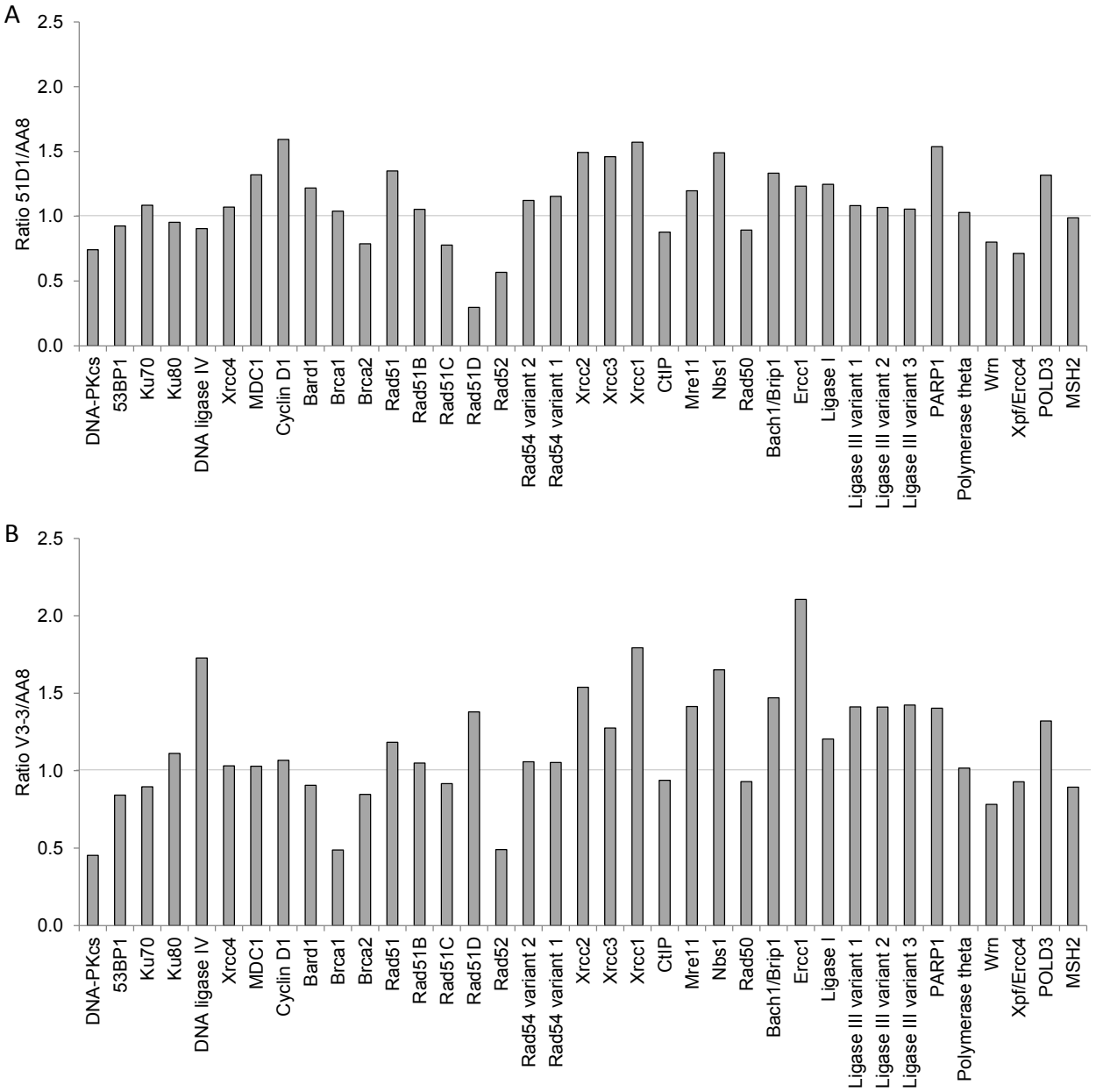


Figure 3.6: Relative expression of DSB repair genes in 51D1 and V3-3 cells. Bars represent the number of RNAseq reads mapping to the DSB repair gene transcripts relative to the number of corresponding reads from AA8 cells. A) 51D1 cells, B) V3-3 cells. Data from Samuel Neuenschwander (SIB).

use of microhomology in these cells [335, 50], a hallmark of the MMEJ pathway [221]. In conclusion, it seemed that while the V3-3 cells lack the activity of DNA-PKcs, they may be proficient in DSB repair via the MMEJ pathway.

3.5 Discussion

Matrix attachment regions are often employed in biotechnology to prevent silencing of the gene of interest and to stabilize its expression in mammalian cells. Our previous study suggested that a human MAR 1-68 may also stimulate plasmid integration into the cellular genome, potentially by cooperating with the DNA repair machinery [110]. Here, we used different Chinese hamster cell mutants deficient in the components of two DSB repair pathways - HR and NHEJ, to explore this possibility.

The experiments using Rad51D- and DNA-PKcs-deficient CHO cells suggested that MAR-mediated plasmid integration may require a pathway alternative to NHEJ, as the inactivation of this mechanism increased the integration and expression of MAR-containing plasmids. Conversely, a defect in Rad51D, one of the factors implicated in HR, nearly abolished transgene integration and expression as well as decreased the survival of stably transfected cells. This suggested that the process of transgene integration in CHO cells may be homology-dependent. Surprisingly, transgene expression in a different, Brca2-deficient cell line was higher than in control cells, both in the presence and absence of the MAR. In cells transfected with GFP only, this increase resulted from an elevated number of integrated gene copies, whereas in the presence of the MAR the increased expression could not be explained by higher copy numbers. This suggested that in the absence of Brca2, the MAR rather than further increasing the number of transgenes that integrate, may direct them to more favorable genomic loci. These results are in contrast with those obtained in with Rad51D-deficient cells, which suggested that some of these cells may

contain additional genetic defects.

CHO cell lines deficient in various recombination factors have been used for many years to study the DNA repair processes in mammalian cells. However, most of these cells were generated before high throughput sequencing technologies were developed, thus the underlying mutations remain largely unknown. Furthermore, secondary or compensatory mutations that improve culture growth have been proposed to occur, possibly affecting the recombination propensity of these cell lines [340]. We hypothesized that the CHO cell lines used in our study may have acquired mutations that restored their ability to efficiently repair DSBs. To test this hypothesis we employed two widely used GFP-reconstitution assays based on the reconstruction of a functional GFP gene through either NHEJ or HR. Surprisingly, little differences in recombination frequencies were observed between the two mutant cell lines and the parental cells. This suggested that these cells may not be fully deficient in HR or NHEJ activities, contain complementing mutations, and possess very active alternative DSB repair pathways.

To address this question we performed whole transcriptome sequencing of these cells and analyzed their DSB repair gene expression. While the two previously described mutations were apparent in these cells, we also observed additional differences in gene expression between the two mutants and the parental cell line. The 51D1 cells, in addition to the Rad51D mutation, showed a significant decrease in the expression of Fancl, a crucial factor in the FA pathway. Both functional FA and HR pathways are needed for efficient ICL repair. Therefore, ICL repair in 51D1 cells may be severely compromised due to the deficiency in the components of both pathways. In line with this, 51D1 cells complemented with wild type Rad51D remain more sensitive to cross-linking agents than the parental AA8 cells [195]. This double defect in ICL and DSB repair may account for the extreme sensitivity of these cells.

The V3-3 cells proved to contain two nonsense mutations in at least one allele of the the DNA-PKcs gene, likely rendering them defective in the NHEJ-mediated DSB repair. However, we also observed an increase in the expression of genes implicated in the MMEJ pathway. This, taken together with the lack of apparent end-joining defect, suggested that this pathway compensates for NHEJ deficiency in these cells. This conclusion is supported by other studies that reported very efficient DNA end-joining in cells with dysfunctional NHEJ [28, 152, 186, 244].

In conclusion, we uncovered previously unknown defects in two widely used recombination-deficient CHO cell lines. These results indicated that the mutant cell lines may not be the most reliable model to study DNA repair, and they highlighted the need for a more controllable system to study these processes.

3.6 Supplementary materials

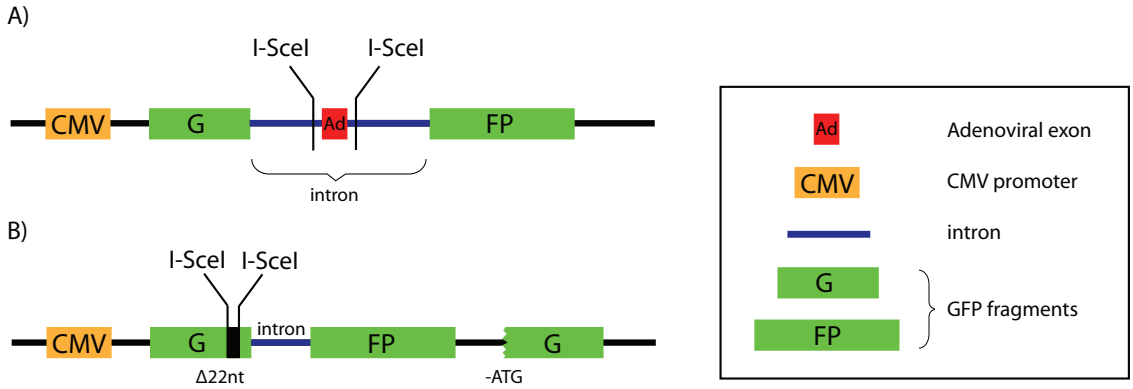


Figure 3S1: GFP-reconstitution reporter cassettes for detection of NHEJ and HR. A) The NHEJ reporter contains a GFP sequence disrupted by the intron from the rat Pem1 gene containing an adenoviral exon, which becomes inserted into the GFP mRNA after the intron is spliced out. Digestion with I-SceI removes the adenoviral exon and generates a DSB with incompatible ends, which can be repaired by NHEJ leading to the restoration of a functional GFP gene. B) The HR vector is construced similarly to the NHEJ reporter, but the first GFP fragment contains a 22nt deletion and two I-SceI restriction sites used to generate a DSB with incompatible ends. The GFP gene is followed by a copy of the first fragment lacking an ATG, which can be used as a templete to repair the GFP gene through intramolecular or intermolecular gene conversion. Components of each cassette are depicted on the legend at the right. Modified from [212]

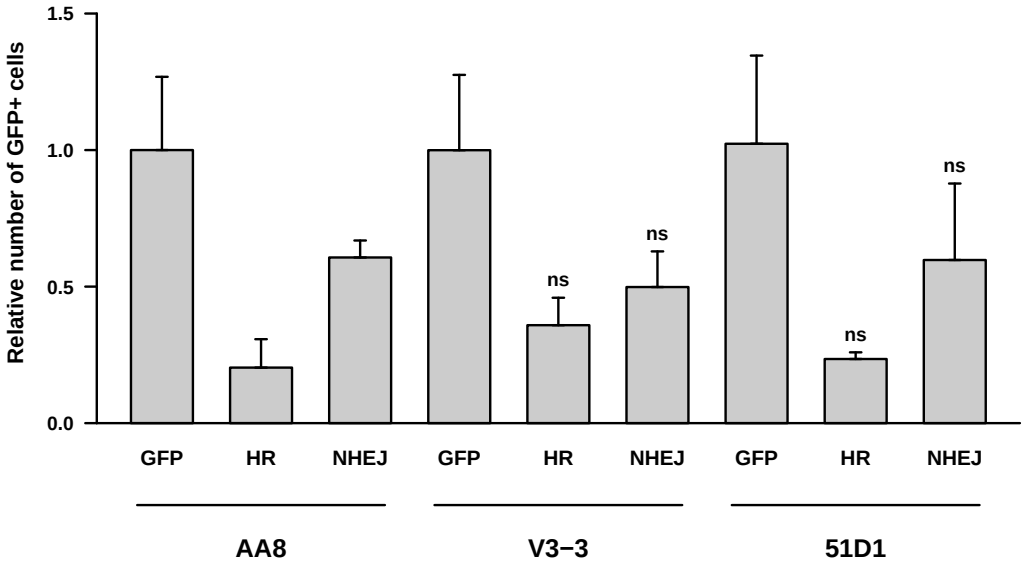


Figure 3S2: The efficiency of HR and NHEJ-mediated DSB repair in AA8, V3-3 and 51D1 cells. Bars represent mean fold change over the cells transfected with the pGEGFP plasmid (GFP) in each cell line. Mean of 3 experiments, s.e.m. error bars. The V3-3 and 51D1 cells transfected with the HR assay (HR) and the NHEJ assay (NHEJ) were compared to the corresponding samples from AA8 cells. Statistical significance was determined by unpaired Student's t-test, significance level $p<0.05$, ns – not significant.

Table 3S1: DSB repair gene expression in AA8, V3-3 and 51D1 cells. Analysis from Samuel Neuenschwander.

Gene name	Count 51D1	Count AA8	Count V33	Padj 51D1/AA8	Padj 51D1/V33	Padj V33/AA8
DNA-PKcs	24090	32496	14710	1	1	0.274314865
53BP1	26324	28488	23994	1	1	1
Ku70	25072	23120	20716	1	1	1
Ku80	12218	12816	14248	1	1	1
DNA ligase IV	1850	2044	3530	1	0.850517258	1
Xrcc4	1052	982	1012	1	1	1
MDC1	16504	12512	12866	1	1	1
Cyclin D1	22060	13848	14772	1	1	1
Bard1	4662	3828	3466	1	1	1
Brca1	96778	93154	45498	1	0.312618836	0.397143498
Brca2	15778	20052	16982	1	1	1
Rad51	10782	7990	9452	1	1	1
Rad51B	1888	1792	1880	1	1	1
Rad51C	888	1142	1046	1	1	1
Rad51D	3188	10722	14784	0.012142075	0.000213137	1
Rad52	6506	11484	5628	0.768426467	1	0.48511229
Rad54 variant 2	5542	4938	5220	1	1	1
Rad54 variant 1	5718	4958	5218	1	1	1
Xrcc2	854	572	880	1	1	1
Xrcc3	286	196	250	1	1	1
Xrcc1	14880	9472	16990	1	1	1
CtIP	17732	20196	18932	1	1	1
Mre11	23626	19742	27912	1	1	1
Nbs1	6594	4428	7312	1	1	1
Rad50	17516	19606	18232	1	1	1
Bach1/Brip1	12226	9178	13492	1	1	1
Ercc1	19240	15620	32884	1	0.950873953	0.752269087
Ligase I	25502	20452	24636	1	1	1
Ligase III variant 1	5550	5126	7234	1	1	1
Ligase III variant 2	7048	6604	9314	1	1	1
Ligase III variant 3	5998	5684	8090	1	1	1
PARP1	57968	37706	52906	1	1	1
Polymerase theta	3304	3212	3264	1	1	1
Wrn	14534	18162	14200	1	1	1
Xpf/Ercc4	2444	3428	3180	1	1	1
POLD3	15714	11930	15752	1	1	1
MSH2	17442	17668	15780	1	1	1

Table 3S2: Genes significantly downregulated in 51D1 cells compared to AA8 and V3-3 cell lines. Analysis from Samuel Neuenschwander.

Gene name	Count 51D1	Count AA8	Count V3-3	Padj 51D1/AA8	Padj 51D1/V3-3	Padj V3-3/AA8
Hox-C8	90	4308	1678	0.000000000	0.000000368	0.23
Sprr2h	2068	15694	6890	0.000000630	0.016909143	0.25
Serpinf1	778	10464	4918	0.000000000	0.000040094	0.41
Ugt1a1	946	10618	5702	0.000000005	0.000045381	0.72
PTPRZ1	2446	14114	7812	0.000028140	0.022222446	0.77
aquaporin-1	3466	14964	30480	0.000792297	0.000000007	0.86
Hox-C10	466	11032	7020	0.000000000	0.000000000	1
Lama3	3796	40600	26816	0.000000002	0.000000299	1
Tcf7l1	896	4252	6432	0.001102082	0.000003690	1
Fancl	406	4838	3190	0.000000020	0.000019269	1
Myo1d	1298	4544	7850	0.016114493	0.000019369	1
Hox-C6	104	1012	1334	0.000417922	0.000045896	1
Rad51D	3188	10722	14784	0.012142075	0.000213137	1
Slc1a4	2226	8142	9972	0.006535173	0.000515255	1
U2AF1	60	538	790	0.014007224	0.001381771	1
Tgfbr3	11586	33764	36982	0.032033409	0.010243412	1
Medag	22338	73420	70024	0.009552403	0.011007031	1
Itm2b	9040	34144	26446	0.002423083	0.026543681	1

Chapter 4

MAR-mediated transgene integration involves the SD-MMEJ DNA repair pathway

This chapter is based on a manuscript in preparation entitled "MAR-mediated transgene integration into permissive chromatin and efficient expression involve an SD-MMEJ-like DNA repair pathway" by **Kostyrko, K.**, Neuenschwander, S., Junier, T., Regamey, A., Iseli, C., Schmid-Siegert, E., Bosshard, S., Majocchi, S., Girod, P.A., Xenarios, I., Mermod, N.

4.1 Abstract

Untargeted integration of foreign DNA vectors into mammalian cell genomes is extensively used in gene therapy and biotechnology, but it remains a poorly understood and relatively inefficient process. The formation of vector concatemeric arrays and their genomic integration are commonly believed to involve DNA double strand break (DSB) repair pathways such as non-homologous end-joining (NHEJ) or homologous recombination (HR), but whether these pathways might mediate plasmid integration has remained unclear. Here, we silenced essential DSB repair genes in Chinese hamster ovary (CHO) cells and found that the lack of NHEJ activities did not affect integration or expression, while the silencing of HR factors enhanced plasmid concatemer formation and stable expression. Genomic integration of plasmids was inhibited by the silencing of specific HR proteins mediating homology search and of DNA synthesis-dependent microhomology-mediated end-joining (SD-MMEJ) activities. Analysis of transgene integration loci and junction DNA sequences validated the prevalent use of an SD-MMEJ pathway for transgene integration within gene-rich areas of the CHO genome, an effect shared with DNA elements that activate transgene integration and expression. These findings should lead to an approach mediating significantly improved gene transfer efficacy.

4.2 Introduction

Spontaneous integration of non-viral DNA vectors such as plasmids into the eukaryotic cell genome is a widely exploited process in gene therapy and biotechnology. Its molecular basis however, remains incompletely understood. It is believed to rely on cellular DNA-repair mechanisms, as it is favored and likely triggered by the presence of free DNA ends in the vector resembling double stranded breaks (DSBs). The two major pathways responsible for DSB repair in eukaryotic cells are non-homologous end-joining (NHEJ) and homologous recombination (HR) [121, 144, 155]. NHEJ, active throughout the entire cell cycle, is a fast mechanism, which efficiently joins DNA ends with little DNA end processing. In contrast, HR is a slow, multi-step process requiring resection of one of the two DNA strands and a homologous template for repair. Because of this requirement for homology, HR is traditionally regarded as an error-free pathway [25, 120]. A third group of DSB repair pathways, thought to function when the main repair mechanisms are impaired, are collectively termed microhomology-mediated end joining (MMEJ). MMEJ is a still poorly characterized family of pathways, also referred to as alternative or backup end-joining [12, 28, 73, 83, 108, 248, 254]. The MMEJ mechanisms require short (2-25 nt) homologies to align broken DNA strands before joining. Another hallmark of this process is the occurrence of large deletions and, less frequently, insertions of sequences copied from other parts of the genome, termed templated inserts [204, 225]. MMEJ was also reported to share DNA strand resection with HR, implying that it may partially rely on HR enzymes [72, 83, 184, 204, 328].

Plasmid integration into the genome of eukaryotic cells is an overall inefficient process, occurring in a minor proportion of cells that take up the exogenous DNA. It was shown to involve two major steps: i) recombination between vector molecules to form multiple transgene arrays termed concatemers, and ii) the recombination of the resulting concatemers

into the genome, usually at a single or at few chromosomal loci [94, 116, 167]. The type of DSB repair pathway that is responsible for transgene concatemerization remains currently unclear. In mammalian cells, this process was attributed to an HR mechanism [94, 356], while NHEJ appeared to be involved in zebrafish embryos and rice [66, 167]. In addition, some studies suggest that alternative pathways may also play a role in the joining of extrachromosomal DNA ends [203]. Similarly, the mechanism mediating the final recombination of the transgene with the genome remains to be fully identified. NHEJ is considered to mediate the majority of integration events in eukaryotic cells, while HR may be responsible for a smaller proportion (reviewed in [339, 359]). However, there is evidence that defects in HR strongly impair integration [201, 309], whereas other reports implicated distinct pathways in this process [136, 225].

In previous studies, we reported that plasmid integration is enhanced by the presence of matrix attachment regions (MARs), which are epigenetic regulatory DNA elements that participate in the formation of chromatin boundaries and augment transcription [101, 110, 116, 209]. MARs are thus widely used in biotechnology to sustain elevated transgene expression, as well as to prevent epigenetic silencing effects by blocking the propagation of heterochromatin [5, 125, 163, 377]. Their action to increase plasmid genomic integration and copy number suggested that a positive effect of recombination mechanisms may be an additional mechanism by which MARs improve transgene expression [110, 116]. We speculated that it involved an HR-related pathway, as transgene integration was impaired in CHO cells mutated in Rad51D, one of its components, while it was increased in cells deficient in an essential NHEJ enzyme. However, these observations based on incompletely characterized mutant cell lines did not allow the characterization of the mechanism by which plasmids integrate into the genome and how MAR elements may act to increase this process. Thus, in this study, we first sought to identify the pathway(s)

responsible for the integration of MAR containing or devoid plasmids into the genome of CHO cells.

Using a knock-down approach, we show that a subset of alternative repair mechanisms, termed synthesis-dependent MMEJ (SD-MMEJ), may be preferentially used by mammalian cell lines for the spontaneous integration of foreign DNA into their genome. This finding was confirmed by the characterization of plasmid-to-genome junction sequences, which were found to display an SD-MMEJ pattern. Finally, we show that by stimulating this mechanism, MAR elements mediate very efficient transgene integration into potentially permissive chromatin loci, and that the inhibition of competing recombination pathways can be used to significantly improve transgene expression.

4.3 Materials and Methods

Plasmids and siRNA

The MAR-devoid pGEGFP, MAR 1-68-containing p1-68-GFP, and pGL3-CMV-DsRed expression vectors were described previously [116]. The HR and NHEJ reporter plasmids were kindly provided by V. Gorbunova (University of Rochester, New York, USA) [212]. Small interfering RNA duplexes, specifically designed to target the CHO cell homologs of the recombination proteins listed in Tables 4S1 and 4S2, were designed and provided by Microsynth AG (Balgach, Switzerland). Three RNA duplexes were designed per mRNA to increase the probability of successful knock-down. Three negative (non-targeting) siRNAs were also designed as controls.

Cell culture

Adherent CHO DG44 cells [333] were cultivated in DMEM/F-12+GlutaMAX^{TM} supplemented with 1x HT and 10% fetal bovine serum (Gibco, Invitrogen, Carlsbad, CA), and

with the antibiotic-antimycotic solution (Sigma-Aldrich, St Louis, MO, #A5955). Suspension adapted CHOK1 derived cells were cultured in SFM4CHO (HyClone^{TM}) medium supplemented with 8mM L-Glutamine (PAA Laboratories GmbH, Austria) and HT (Gibco, Invitrogen, Carlsbad, CA).

Recombination assays

For HR and NHEJ recombination transient assays, adherent CHO cells were transfected with HR or NHEJ reporter plasmids digested with I-SceI, and with the pGL3-CMV-dsRed plasmid to normalize for transfection efficiency, using Fugene 6 according to manufacturer's instructions (Promega, Madison, WI). The pSV40-GFP vector (pGEGFP) was transfected in parallel as a positive control of GFP expression.

For siRNA-mediated knock-downs of recombination proteins, adherent cells were transfected with equimolar mixes of three mRNA-specific or control siRNA duplexes at a final concentration of 50nM using Lipofectamine RNAiMAX, according to the manufacturer's instructions (Invitrogen, Carlsbad, CA), using siRNAs against the genes described in Tables 4S1 and 4S2. After two days, the siRNA-treated cells were re-transfected with pGEGFP or p1-68-GFP and a puromycin resistance plasmid (pSVpuro) using Lipofectamine 2000 (Invitrogen, Carlsbad, CA). Puromycin (5 μg/ml) was added to the culture medium 24h after transfection, and stably transfected cells were selected for 2 weeks. Puromycin-resistant colonies were stained with 0.2% methylene blue and quantified using ImageJ (U.S. National Institutes of Health, Bethesda, Maryland, USA). GFP expression was analyzed by cytofluorometry using a fluorescence-activated cell sorter (FACS) (CyAn flow cytometer, Beckman Coulter), whereas aliquots of each sample were used for genomic DNA extraction.

Transgene copy number determination and quantitative PCR

To analyze the transgene genomic integration sites and copy number, total genomic DNA was isolated from cells using the DNeasy purification kit (Qiagen, Hilden, Germany). For quantitative PCR (qPCR), 6 ng of genomic DNA were analyzed using the SYBR Green I Master kit for the Light Cycler 480 machine (Roche, CA) using AGCAAAGACCCCAACGAGAA and GGCGGCGGTCACGAA as GFP-specific primers. The beta-2-microglobulin (B2M) CHO gene was amplified as a normalization control using ACCACTCTGAAGGAGCCCA and GGAAGCTCTATCTGTGTCAA as primers. The number of integrated transgene (GFP) copies was calculated using the B2M gene as a reference, as previously described [257].

Characterization of transgene integration sites

The genome and transcriptome of the suspension-adapted parental CHO K1 cells was determined by a combination of genomic shotgun and mate-pair sequencing using the Illumina technology, performed by Fasteris SA (Plan-Les-Ouates, Switzerland), and by the Pacific Biosystem technology at the Next Generation Sequencing Facility of the University of Lausanne. Genome assembly was performed in the Vital-IT computer facility at the Swiss institute of Bioinformatics Lausanne branch. The expressed coding sequences were annotated using the publicly available Annotation Release 101 of the Chinese hamster genome assembly (CriGri_1.0, GCF_000223135.1) [362].

For identification of plasmid integration sites in polyclonal populations, CHO K1 cells were electroporated with the pGEGFP or p1-68-GFP plasmids and the pSVpuro plasmid using the Neon transfection system (Invitrogen, Carlsbad, CA). After 3 weeks of puromycin selection total genomic DNA was isolated from polyclonal cells using the Genomic-tip

G/20 kit (Qiagen, Hilden, Germany). The DNA was sequenced using the Single Molecule Real-Time (SMRT) technology (Pacific Biosciences, Menlo Park, CA) at the Next Generation Sequencing Facility of the University of Lausanne. CHO cells transfected with p1-68-GFP were sequenced using 20 SMRT cells, and those transfected with pGEGFP required the use of 60 SMRT cells to obtain the same number of integration site sequences. Transgene integration sites were identified by a custom identification pipeline. PacBio filtered subreads were obtained using the tool DEXTRACTOR (Myers, unpublished) using the standard settings. Plasmid sequences were identified in PacBio filtered subreads with the help of the alignment tool BLASR [46]. A raw score of at least -500 was chosen as cut-off based on results using PacBio reads from untransfected CHO cells. Flanking regions of matching plasmid sequences were extracted and mapped onto the CHO-K1 genome using BLASR. 14 CHO genomic integration sites were identified per sample. 2 sets of 14 different, randomly-picked genomic scaffolds of the same length (+/-10%) as the sample scaffolds were selected as controls. The Annotation Release 101 of the Chinese hamster genome assembly (CriGri_1.0, GCF_000223135.1) was used to identify the CHO genes in the vicinity of the integration sites. The presence of genes near the plasmid integration position in each of the identified scaffolds was compared with an analogous position on a corresponding control scaffold. An exact binomial test was used to calculate statistical significance between these datasets. Based on this analysis integration within 5kb from an open reading frame (ORF) was considered intragenic, whereas integration within 35 kb from an ORF was defined as gene-proximal.

Suspension-adapted CHO K1 cells were stably transfected with plasmid vectors containing the human MAR X-29 and encoding the light and heavy chains of the trastuzumab and infliximab therapeutic antibodies, as previously described [181]. Clones expressing the highest amount of the recombinant proteins were selected for whole genome sequencing

(Illumina). Integration sites were first predicted by the in silico identification of paired reads displaying linked plasmid and genomic sequences, and the predicted junctions were subsequently validated by PCR amplification and Sanger sequencing. Identification of CHO genes near the plasmid integration sites was performed as described for the polyclonal populations.

Fluorescence in situ hybridization

Preparation of metaphase chromosomes and fluorescence in situ hybridization (FISH) were described previously [76]. FISH probes were prepared by nick translation of plasmids used to generate the clones in the presence of Orange 552 dUTP according to the manufacturer's protocol (Enzo Life Sciences, Farming dale, NY). The chromosomes were counterstained with DAPI (VECTASHIELD® Mounting Medium, Vector) and observed using a 63x oil immersion objective on an Axio Observer.A1 microscope (Zeiss).

4.4 Results

4.4.1 Plasmid integration does not rely on NHEJ or the canonical HR pathway

To assess the possible implication of NHEJ and HR in plasmid concatemer formation and spontaneous integration into the cell genome, we silenced the components of these major DSB repair pathways in CHO cells using short interfering RNA (siRNA) (Fig.1.5, Table 4S1). Efficient reduction of the target mRNA and/or protein levels by siRNA transfection was validated experimentally, to insure decreased mRNA levels of at least 2-fold (Fig.4S1). When assessed by western blotting, the Ku70, 53BP1, Rad51, and Rad51D proteins were essentially undetectable in cells treated with their cognate siRNAs, whereas they were not

significantly altered by the control siRNAs (Fig.4S2). We therefore concluded that the studied recombination proteins were efficiently down-regulated in siRNA-treated cells.

To evaluate if the knock-down of these activities may affect DSB repair and recombination mechanisms, we used previously described HR and NHEJ fluorescent reporter assays, based on the repair of transiently transfected plasmids having a I-SceI-induced DSB in the GFP coding sequence [212, 291]. These assays enable to evaluate the efficiency of extrachromosomal break repair by monitoring the restoration of GFP expression, and thereby may provide an estimation of the involvement of the HR and NHEJ pathways in plasmid concatemer formation. We observed that DSB repair of the HR reporter plasmid was significantly impaired by the knock-down of the Rad51 HR protein, whereas it was rather increased in cells treated with siRNAs targeting NHEJ factors (Fig.4S3A). This indicated that this HR protein may contribute to the repair of DSBs in episomal plasmids. In contrast, the occurrence of restored GFP expression by an NHEJ mechanism was not altered in any of the siRNA-treated cells (Fig.4S3B). Given the near-complete knock-down of critical components of the NHEJ repair mechanism such as Ku70, these results implied that NHEJ is not prominently used to rejoin plasmid DSBs in CHO cells.

The recombination mechanisms involved in plasmid concatemer formation and genomic integration were further assessed by co-transfecting the siRNA-treated cells with plasmids carrying the GFP reporter and a puromycin resistance gene. Cells having stably integrated the plasmids into their genome were selected by culture in the presence of the antibiotic. Analysis of the average number of integrated GFP copies in antibiotic-resistant polyclonal populations was used to assess the efficiency of the concatemerization process. Quantification of puromycin-resistant colonies was performed to estimate the percentage of cells which had successfully integrated transgenes into their genome, as a measure of the overall efficiency of the integration process. The level of GFP expression was measured

to assess plasmid expression per transgene copy, as an estimation of the integration of plasmids within transcription permissive or non-permissive areas of the cell genome.

The genomic integration and expression of the GFP plasmid were not significantly inhibited by the down regulation of NHEJ activities such as DNA-PKcs, Ligase IV or Xrcc4, nor did it alter the number of antibiotic resistant colonies or expression per transgene copy (Fig.4.1A-D). This indicated that NHEJ activities are not limiting for plasmid concatemerization and integration within the CHO cell genome. Rather, the plasmid copy number was increased by the down regulation of 53BP1, suggesting that NHEJ may compete with a recombination pathway mediating plasmid concatemerization (Fig.4.1B).

Stable GFP expression and/or transgene copy numbers were significantly increased by the knock-down of HR proteins, notably MDC1, Rad51, Rad52, Rad54, and Brca1 (Fig.4.1A,B). The knockdown of these proteins had overall little effect on gene expression when normalized to transgene copy, indicating that increased expression upon HR activity knockdown resulted mostly from an increased copy number rather than from the preferential integration of plasmids into transcription-permissive chromatin (Fig.4.1C). These observations indicated that HR activities may oppose a mechanism that mediates plasmid concatemerization prior to genomic integration. However, the knock-down of MDC1, Rad51 and Rad51C strongly decreased the number of puromycin resistant colonies (Fig.4.1D), indicating that these components of the HR pathway may be required for transgene integration. Interestingly, the frequency of genomic integration was not affected by the knock-down of other components of HR, such as Rad52, Rad54 or Brca1, despite their effect on transgene integration and expression. These findings implied that proteins mediating DNA homology search of the HR DNA repair pathways are required for genomic integration, but that later-acting HR proteins are not involved, suggesting the occurrence of non-canonical HR-related integration mechanisms.

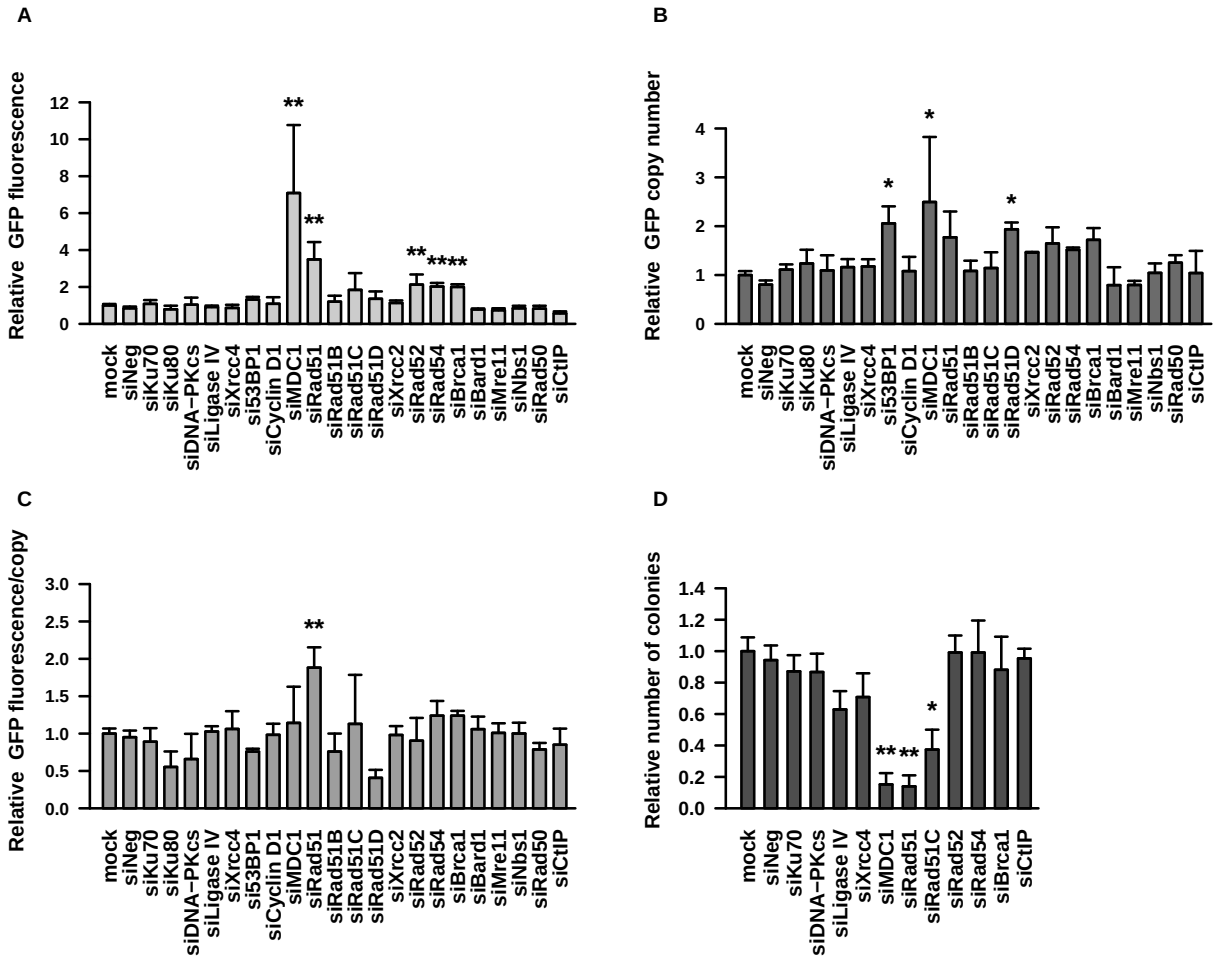


Figure 4.1: Effect of HR and NHEJ components knock-down on plasmid genomic integration and expression. CHO cells were treated with siRNAs against the indicated HR and NHEJ factors, as referenced in Table 4S1, and the cells were re-transfected with a GFP expression plasmid and with a puromycin resistance vector. Puromycin-resistant polyclonal cell populations were selected and used to assess the average GFP fluorescence (A), GFP transgene copy number (B) and GFP expression per transgene copy (C), whereas the frequency of genomic integration events was assessed by quantifying the occurrence of puromycin-resistant colonies (D). Values represent mean fold change over control cells not treated with siRNAs (mock), and error bars represent s.e.m, n≥3. Asterisks indicate significant differences between the siRNA-treated sample and mock control, (*) $p<0.05$, (**) $p<0.01$; unpaired Student's t-test with Benjamini-Hochberg correction.

4.4.2 MMEJ-type mechanism seems to mediate plasmid concatemerization and genomic integration

Given that neither the NHEJ nor the canonical HR pathway may be involved in plasmid concatemerization prior to genomic integration, we thus hypothesized that plasmid concatemer formation may be mediated by a mechanism which is suppressed by the main DSB repair pathways in untreated cells. We speculated that this could involve MMEJ-related mechanisms that were recently reported to function in eukaryotic cells with impaired NHEJ and/or HR, but that may share some early events with the HR pathway, such as the 5' strand resection (Fig.1.5) [72, 75, 204]. Thus, we next proceeded to knock-down proteins proposed to participate in MMEJ pathways (Table 4S2).

In most cases, depletion of MMEJ proteins had little effect on plasmid integration or expression, possibly because this pathway may be masked by the main repair mechanisms, especially in the absence of induced DNA damage, as was the case here. Nevertheless, we did observe some decrease of GFP expression and copy number upon the knock down of the SD-MMEJ protein DNA polymerase θ (Pol theta) (Fig.4.2A,B). This suggested that the plasmid concatemerization process may be DNA synthesis-dependent. Interestingly, we observed an increase in transgene expression and copy number upon the knock-down of DNA Ligase I (Fig.4.2B), whereas genomic integration of the plasmids was inhibited (Fig.4.2C). A recent study suggested the existence of two branches of the MMEJ-related alternative end-joining pathways, one of which may depend on Ligase I whereas the other would require Ligase III (Fig.1.5) [248, 254]. Upon Ligase I knock-down, the Ligase III-dependent branch should prevail, which may favor plasmid concatemer formation and thereby mediate the observed increase of the transgene copy number. In contrast, the Ligase I knock-down may suppress the other MMEJ-related pathway, thereby preventing genomic integration of the plasmids. Taken together, the results pointed to a Ligase III-

and DNA polymerase θ-dependent mechanism playing a role in plasmid concatemerization, whereas MDC1, Rad51 and Rad51C may be involved in a Ligase I-dependent pathway mediating genomic integration.

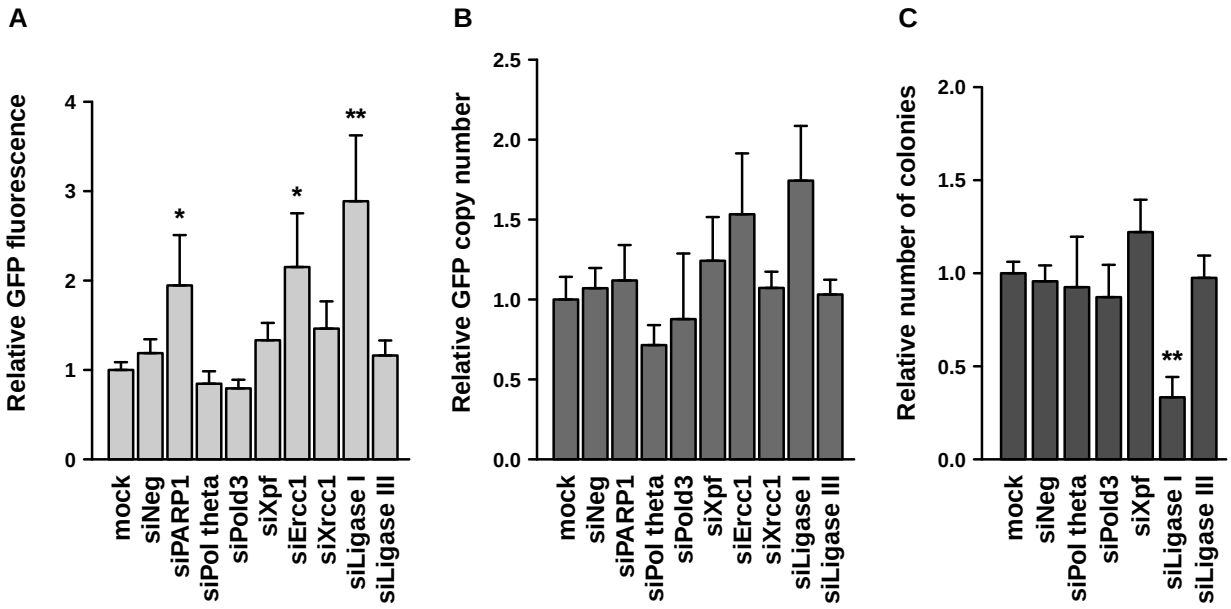


Figure 4.2: Effect of MMEJ components knock-down on plasmid genomic integration and expression. CHO cells were treated with siRNAs against the indicated MMEJ genes and processed as described in the legend to Fig. 4.1 and in Table 4S2. The average GFP fluorescence (A), GFP transgene copy number (B), and frequency of genomic integration events (D) were assessed and represented as in Fig. 4.1 (n$\geq$3).

4.4.3 Plasmid concatemerization relies on an SD-MMEJ pathway involving DNA polymerase θ and Ligase III

To directly assess whether an MMEJ-related mechanism may mediate plasmid end-joining, we constructed a MMEJ-specific GFP reporter assay based on analogous principles as for the HR and NHEJ reporter plasmids used previously. The vector, based on a previously described reporter sequence [347], contains two 9 nt microhomology sequences bracketing an I-SceI site which may be used to restore a functional GFP via a MMEJ pathway (Table 4S3). Upon transfection, the I-SceI-digested vector yielded approximately 50% of fluorescent CHO cells having successfully reconstituted a functional GFP coding sequence,

from which circularized plasmids were rescued and sequenced. Interestingly, we observed that the 9 nt homology was used to restore the functional GFP through MMEJ in only 1 out of the 12 analyzed junctions (Table 4S3, junction 1). Instead, the majority of analyzed sequences resembled the recently proposed mechanism termed synthesis dependent (SD)-MMEJ [372]. This pathway is thought to use a non-processive DNA polymerase such DNA polymerase θ to copy short homologous sequences (2-9 bp) from a different part of the repaired molecule, which can then be used to pair with the other protruding single strand [371, 372]. As a result, the junction sequence consists of a short duplication (direct or inverted) of a sequence found nearby on the repaired DNA fragment. In the remaining 11 cases there was no significant microhomology at the fused sequences, as required by the MMEJ mechanism, but we identified the direct or inverted repeat sequence which may have served as template for the DNA synthesis of SD-MMEJ, either upstream or downstream of the junction (Table 4S3 junctions 2-12). We also concluded that the repaired junctions did not result from an NHEJ mechanism, as we observed large deletions indicating extensive DNA end processing, which is not commonly observed in this latter end-joining mechanism. We therefore concluded that plasmid concatemerization relies mostly on an SD-MMEJ pathway involving the activities of DNA polymerase θ and Ligase III, and that the simple MMEJ mechanism is seldom used.

4.4.4 MAR elements promote plasmid integration by stimulating the SD-MMEJ pathways

We have previously shown that transgene integration in CHO cells is enhanced 3-4 fold in the presence of matrix attachment regions (MARs), which are DNA elements thought to be responsible for the formation of chromatin domain boundaries [110, 209]. The action of MARs to increase both the number of transgene copies as well as the frequency of

genomic integration events has been previously ascribed to HR-related mechanisms [116]. However, which HR-related recombination mechanism may be activated by MAR elements has remained elusive, as subsequent whole transcriptome sequencing analysis of the mutant CHO cell lines used in this prior study revealed alterations in the expression of a number of DNA repair genes (data not shown).

Thus, to unambiguously identify the recombination mechanism activated by such elements, we combined the addition of the human MAR 1-68 in the GFP vector with the siRNA knock down approach used earlier. As shown in previous studies, inclusion of the human MAR 1-68 in the GFP plasmid enhanced GFP expression and copy number by approximately 6-fold and 3-fold, respectively, when compared to the MAR-devoid control plasmid (Fig.4.3A,B). This indicated that the MAR acted to activate the plasmid concatemerization mechanism, whereas it concomitantly increased expression per gene copy (Fig.4.3C). The presence of the MAR also increased around 2-fold the proportion of cells having recombined the transgenes into their genome (Fig.4.3D), indicating that it also activates the transgene integration process.

In the presence of the MAR, the silencing of NHEJ factors had no effect on transgene expression or copy number, as before (Fig.4.3A,B). In contrast, the knock-down of many HR and cell cycle control factors yielded very high transgene expression, but without increasing further the transgene copy number. Consistently, we observed a significant enhancement of expression per gene copy upon the knock-down of most HR factors (Fig.4.3C). A strong inhibition of the frequency of plasmid genomic integration was again noted upon the knock-down of MDC1 and especially Rad51 in the presence of the MAR (Fig.4.3D). Taken together, these results indicated that the presence of the MAR and Rad51 inhibition acted synergistically to promote transgene integration into expression-permissive portions of the genome by a less prominent DNA repair pathway. We therefore speculated that

the MAR may act to promote a MMEJ-related pathway that directs transgenes into expression-favoring chromatin structures.

In the presence of the MAR, the knock-down of MMEJ factors had mostly similar effects on GFP expression and copy number as observed earlier for the MAR devoid plasmid, with a decrease upon the knock-down of DNA polymerase θ, and an increase in the absence of Ligase I (Fig.4.3A,B). However, the presence of the MAR element seemed to compensate for the effect of Ligase I down-regulation on transgene genomic integration, by dampening the significant inhibition resulting from the lack of this ligase on the integration of the MAR-devoid plasmid (Fig.4.3C,D). Taken together, these findings suggested that the MAR may act to increase the frequency of genomic integration events by promoting the use of the Ligase I-dependent integration pathway. Overall, we concluded that the MAR element may activate both concatemerization and genomic integration processes by stimulating SD-MMEJ related recombination pathways, and that the MAR and SD-MMEJ pathways may concur to favor integration into expression-permissive genomic loci.

4.4.5 MAR elements and the SD-MMEJ pathway direct transgenes into gene-rich chromatin regions

To further assess which of the alternative recombination pathways may mediate favorable genomic integration events, we analyzed the genomic integration loci and the DNA sequence of the genome-plasmid junctions. This was performed on three CHO clones transfected with immunoglobulin expression vectors containing a MAR element, and selected for stable expression of these therapeutic proteins at high levels. To do so, we sequenced the genome of the parental CHO-K1 cell line, as well as those of the three clones, and we devised a software to identify linked DNA sequence reads pertaining either to the plasmid or to the CHO genome. 6 integration sites in one clone (BS01) and 2 in each of the remaining

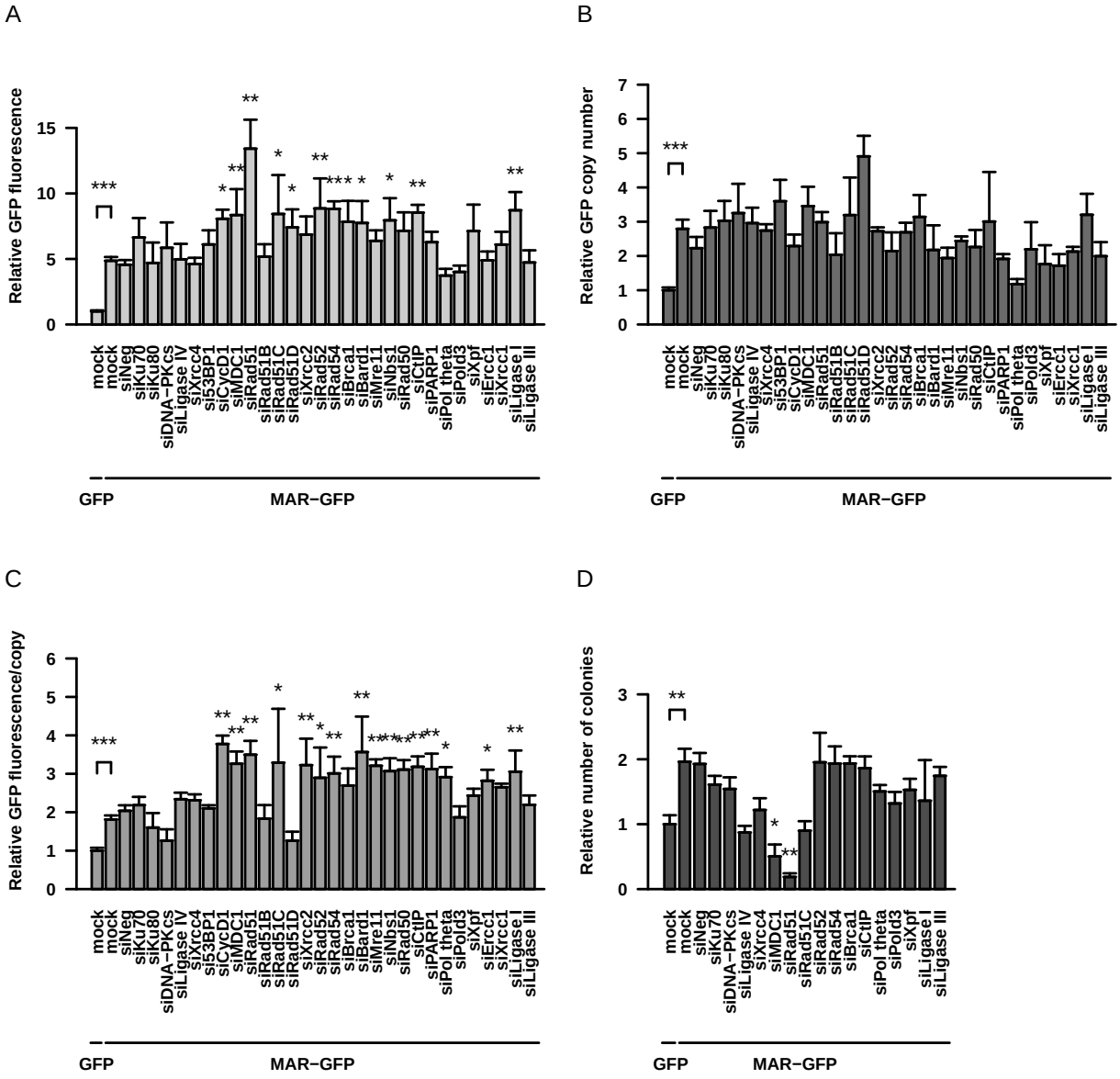


Figure 4.3: Effect of a MAR element and recombination gene knock-down on plasmid genomic integration and expression. The effect of the inclusion of a MAR element on A) stable transgene expression, B) GFP transgene copy number, C) GFP expression per transgene copy, and D) the frequency of genomic integration events, were assessed as described for Fig.4.1 and 4.2, except that siRNA-treated cells were re-transfected with the MAR-GFP, or with the MAR-devoid GFP expression construct, as indicated, and with the puromycin resistance vector (n≥3).

two clones (BS03 and Cp33/64), as predicted by the bioinformatics software, were PCR amplified from the cell genomes, which allowed the experimental validation of their junction DNA sequences. The occurrence of the predicted number of plasmid integration loci was also validated by FISH analysis for two of the analyzed clones (Fig.4S4).

From the 5 integration sites where both sides of the genomic junction sequence were validated experimentally, 2 had large deletions (913bp in BS01 and 320bp in Cp33/64), as expected from MMEJ-related mechanisms (Table 4S4, and Fig.4S5). In 5 of the 15 validated junctions, we noted the presence of short (1-3 bp) or long (60-100 bp) templated inserts, as may be explained by the involvement of a DNA polymerase such as DNA Polymerase θ in the repair process, pointing to the synthesis-dependent mechanism (Fig.4S6). All analyzed junction sequences fitted well to the SD-MMEJ model, whereas 3 junctions covered pre-existing microhomologies extending over 2 nucleotides and thus could be explained equally well by a simple MMEJ mechanism. Interestingly, we found no integration sites that could be easily explained by HR. Although NHEJ cannot be fully excluded, as it does not require homology, the SD-MMEJ mechanism seems more likely due to the presence of extended deletions and templated inserts in the junctions, and because no junction lacking any type of microhomology was observed. Overall, these results confirmed that the genomic integration of MAR-containing plasmids predominantly involves the SD-MMEJ pathway.

The majority of integration sites (8/10) at annotated loci were found within or nearby expressed genes (7 out of 8; Table 4S4), whereas only 2 were intergenic. We concluded that most integration events had occurred in expressed chromatin or in the vicinity a transcriptionally active genomic region. These results further suggested that the MAR-containing plasmids may preferably integrate within, or in the close proximity of, expressed CHO genes, providing an explanation for the proposed preferred integration of the MAR-containing vectors into permissive chromatin structures. To further assess whether this

indeed resulted from the presence of a MAR element in the immunoglobulin expression vectors or from the selection of clones having well-expressed transgenes, we determined and compared the integration loci of the MAR-containing or -devoid GFP expression vectors.

Analysis of the integration sites identified from the whole genome sequencing of polyclonal CHO populations revealed that, in the presence of the MAR, plasmids often integrated into the gene-rich areas of the genome (10/14 loci) (Table 4S5, Fig.4S7). This result was significantly different from random (p=0.05), suggesting that the MAR may stimulate genomic integration into chromatin regions potentially beneficial for transgene expression. This effect was further increased when analyzing monoclonal populations of cells selected to express the transgenes at very high levels, which displayed integration sites at even closer proximity to CHO genes (Fig.4S7B). This finding implied that the MAR and proximity to a CHO gene both acted to mediate higher transgene expression. In the cells transfected with the MAR-devoid plasmid, integration in the vicinity of genes was less frequent (8/14). Furthermore, it should be noted that these cells required 3-fold more sequencing reads to identify the same number of integration loci as obtained from the cells transfected with the MAR, suggesting that genomic integration events were less frequent without the MAR element. Interestingly, we observed that the integration sites of the MAR-devoid plasmids were more often located in or within 5kb of cellular genes (7/14 sites), which was significantly more frequent than expected by chance (p=0.02). Moreover, all of these genes were found to be expressed in the parental CHO cells (Table 4S5, Fig.4S8), suggesting that in order to survive antibiotic selection in the absence of the MAR, the cells need to integrate the transgenes into transcriptionally active chromatin regions. This may readily explain the strong decrease in cell surviving selection upon the knock-down of Rad51, as this protein was recently reported to be primarily responsible for repairing DSBs occurring in transcriptionally active chromatin [13]. In contrast, addition of the MAR

into the plasmid seemed to alleviate this need to integrate into transcribed genes, as only half of the genes in the vicinity of the integration sites were found to be expressed. This suggested that the MAR itself may ensure high expression of the transgene, likely due to its previously reported transcription-enhancing properties [1, 101]. Taken together, these results suggested that MAR elements may promote transgene integration into gene-rich chromatin regions by stimulating a Ligase I-dependent SD-MMEJ mechanism.

4.5 Discussion

Eukaryotic cells developed many defense mechanisms to detect and repair DNA double stranded breaks, one of the most dangerous types of DNA damage. The two canonical pathways responsible for DSB repair are HR and NHEJ. However, recent evidence has indicated that these two mechanisms may not be sufficient to repair all the DSBs that arise in cells, and that several alternative pathways, collectively termed MMEJ or alt-NHEJ, also exist in eukaryotic cells [108, 328]. These processes are often obscured by the main repair mechanisms, which may predominate in normal cells. Furthermore, their components are still poorly characterized and there is no simple assay to specifically detect these alternative mechanisms, rendering their study difficult. However, they are now attracting increasing attention, notably in oncology, since these 'illegitimate' recombination pathways may be involved in tumor progression and resistance to therapy, as they were shown to be more prevalent in tumor cells or cell lines defective in the main DSB repair pathways [23, 323]. They were also identified as a major cause of chromosomal rearrangements leading to cancer formation [138, 299, 381].

Here, we found that NHEJ and HR are not the main pathways responsible for non-specific recombination, which may be required for the integration of exogenous DNA in the genome of CHO cells. Rather, we found that the absence of many HR factors significantly

augmented plasmid concatemerization, as indicated by the increased number of integrated transgene copies, implying that HR proteins may compete with one or several other DSB repair pathways that more efficiently mediate plasmid concatemer formation. In contrast, specific HR proteins such as the homology-searching Rad51 were required for efficient transgene recombination with the genome, whereas the silencing of other HR proteins did not affect genomic integration. This suggested that another mechanism, which is distinct from the canonical NHEJ or HR pathways but may nevertheless require DNA homology, was mediating plasmid genomic integration.

Consistently, the knock-down of several genes involved in the SD-MMEJ pathway was found to alter plasmid genomic integration. This conclusion was also supported by the finding that the majority of rejoined plasmid extremities displayed microhomology patterns characteristic of the SD-MMEJ mechanism, as proposed by Yu and McVey [372]. Indeed, both plasmid-to-plasmid and plasmid-to-genome fusion sequences were also present as direct or inverted repeats near the junction, and templated inserts occurred occasionally at the junctions. However, the knock down of specific SD-MMEJ activities had distinct effects on plasmid concatemer formation and on genomic integration, implying the occurrence of multiple SD-MMEJ pathways. Taken together, our results imply that concatemer formation and integration of MAR-devoid plasmids are mediated by different sets of proteins that may belong to distinct branches of the SD-MMEJ pathways, as proposed in Figure 4.4. One of the SD-MMEJ pathways, which may rely on polymerase θ and Ligase III, appears to mediate plasmid concatemerization. This conclusion is supported by the finding that the knock-down of the components of the HR and of the alternative SD-MMEJ repair pathways, such as Rad genes and Ligase I, increased the number of integrated transgene copies, whereas the silencing of DNA polymerase θ decreased it. The other SD-MMEJ pathway, which may involve the activity of Ligase I, appears to mediate

the recombination of plasmid concatemer with the genome, as indicated by the finding that the lack of Ligase I nearly abolished genomic integration.

Interestingly, we observed that the knock down of Rad51 had very similar effects as the silencing of Ligase I, implying that they contribute to the same pathway mediating genomic integration of exogenous DNA. The mechanism mediating microhomology search of the SD-MMEJ pathway remains mostly uncharacterized, but it may involve DSB repair components that are common to other mechanisms, such as Rad51. In this model, the Ligase I-dependent SD-MMEJ pathway may lie downstream of the search for a homologous DNA strand by Rad51, as in the canonical HR pathway (Fig.4.4). However, the lack of extended sequence similarities may preclude the productive cooperation of Rad51 with its associated paralog proteins, preventing extended strand invasion and the successful completion of the HR pathway. End-joining would then rather be performed by the Ligase I-dependent SD-MMEJ salvage repair pathway, as it only requires short homology regions, which may be shared by the plasmid and the cell genome. When such microhomologous sequences are not available, they may be provided by an adaptor DNA stretch copied from nearby plasmid or genome sequences, leading to the insertion of a templated insert separating the joined sequences (Fig.4S6). We hypothesize that the enzyme involved in the synthesis of the templated insert may be DNA polymerase δ, which together with Ligase I participates in DNA replication and long patch base excision repair (BER) [85, 310]. This polymerase has been recently implicated in an alternative mechanism of one-ended DSB repair involving the use of microhomologies and characterized by the similar presence of templated inserts at repaired junctions [60]. Interestingly, this latter mechanism, while distinct from HR, also seemed to depend on Rad51.

The proposed existence of two parallel SD-MMEJ repair pathways, with only one requiring Rad51, points to mechanistic differences. The use of the DNA polymerase

θ in the Ligase III-mediated pathway may readily explain the lack of requirement for Rad51, as this low-fidelity translesion DNA polymerase was previously associated with the annealing of long microhomologous sequences by a Rad51-independent mechanism [47]. Consistent with this, polymerase θ was recently shown to play a role in an alternative end-joining pathway competing with the Rad51-mediated DSB repair [45]. Conversely, DNA polymerase δ is a high fidelity polymerase that does not easily switch base pairing with the replicated DNA template, as it is held at the replication fork by the proliferating cell nuclear antigen (PCNA) protein. Thus, it may need prior pairing of the DNA strand to be extended with a nearby repeated sequence by homology searching and strand invasion proteins such as Rad51, explaining the need for this HR protein in the Ligase I-dependent SD-MMEJ branch.

Inclusion of a MAR element also increased plasmid concatemerization, suggesting that it may act to activate the preferential processing of the plasmid ends by the Ligase III-dependent SD-MMEJ mechanism. Consistently, the knock-down of competing HR activities did not increase further the copy number of integrated MAR-containing plasmids. In addition, the presence of the MAR increased genomic integration and it dampened the inhibitory effect of Ligase I down regulation, whereas it did not abolish the requirement for Rad51. This finding is consistent with a preferential use of the Ligase I-dependent SD-MMEJ mechanism by the MAR-containing plasmid for genomic integration.

The molecular mechanism by which MAR elements can promote recombination by SD-MMEJ mechanisms still remains unknown. MARs contain AT-rich cores which possess a high potential to denature the double helix [29, 31, 90, 262], which may be prone to DNA strand invasion. In addition, they were shown to contain topoisomerase II cleavage sites, suggesting that they may be hot spots of DNA breakage and repair [307]. Finally, MARs were associated with so-called fragile sites, which are regions of chromosomes with

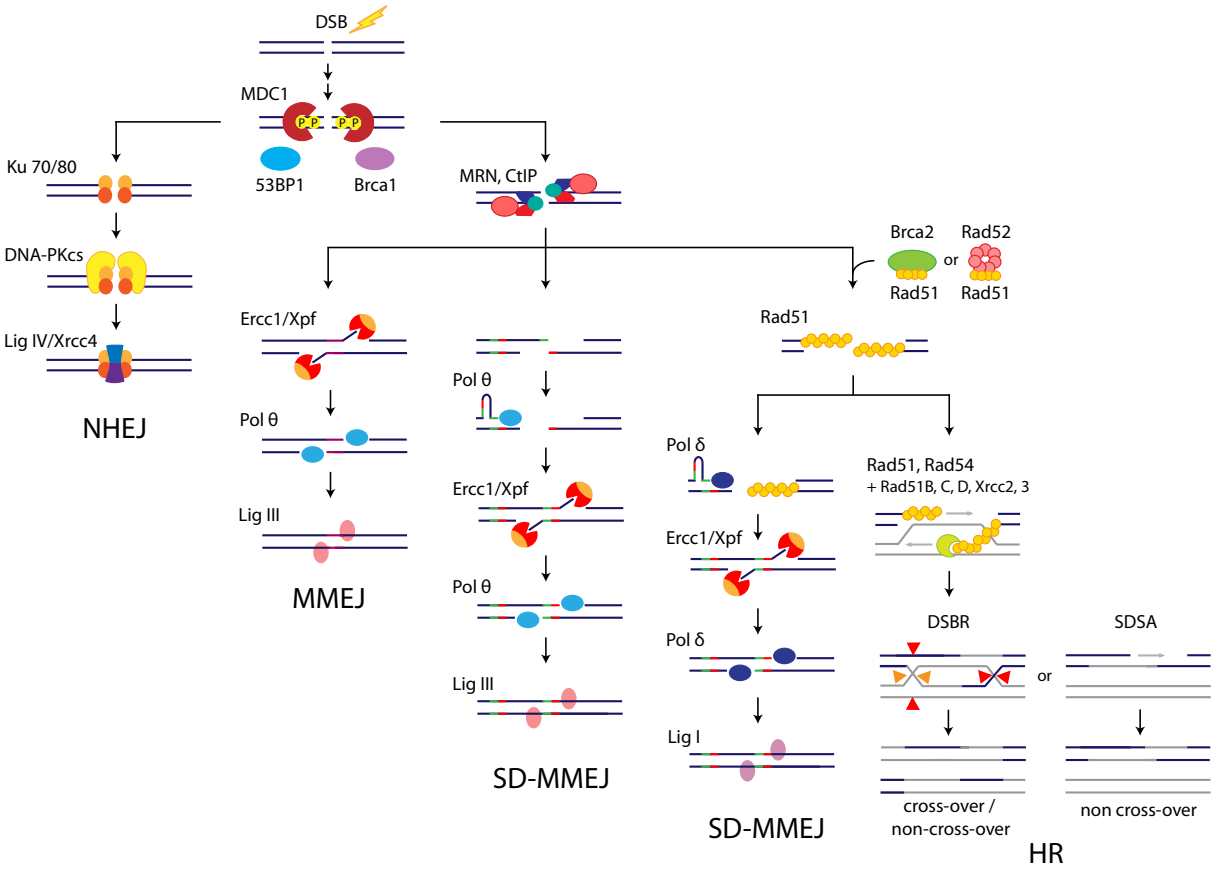


Figure 4.4: Revised model of the major eukaryotic DSB repair pathways. A novel model describing the possible interplay of the NHEJ, HR, MMEJ and two distinct SD-MMEJ pathways involved in the repair of DSBs in CHO cells is depicted, as modified from Fig.1.5. Whereas NHEJ proceeds through a ligation step without the need for sequence homology or DNA end-processing, HR involves extensive DNA 5' end exonuclease processing (resection), and it requires a homologous DNA template to complete DSB repair (DSBR) or synthesis-dependent strand annealing (SDSA). Alternative DSB repair pathways such as MMEJ or the DNA synthesis-dependent SD-MMEJ, may share the DNA end-resection machinery with HR. These alternative pathways use short regions of homology (microhomology) to align DNA ends before repair by flap removal, DNA synthesis and ligation. Whereas the MMEJ pathway utilises pre-exisiting microhomologies, SD-MMEJ requires a DNA polymerase to amplify the microhomology from a more distant region of the DNA. Although the junction sequences resulting from both SD-MMEJ branches are similar, the Ligase I-dependent SD-MMEJ branch requires the homology-searching Rad51 protein, and it may provide a fallback mechanism when an extensive region of homology is not available. Activities that may initiate MMEJ and Ligase III-dependent SD-MMEJ remain to be identified, but the presence or not of a microhomology sequence allowing the pairing of the 3' protruding strand may dictate the choice between these two pathways.

high tendency to breakage [143, 267, 316]. Consistently, these sites have been previously reported to be preferred targets of plasmid integration [270].

MAR elements were also proposed to contribute to the initiation of replication in mammalian somatic cells [71], and they were also associated with episomal DNA replication [259], suggesting that the DNA replication machinery might preferentially associate at or close to MARs. Among the factors participating in replication, several were shown to be also involved in MMEJ-related mechanisms (e.g. Pold3, Ligase I). Furthermore, these pathways were reported to be most active during the S phase, suggesting that they may be one of the the primary mechanisms used for the repair of DSBs arising during DNA replication [328]. Thus, MAR elements may be preferential sequences of SD-MMEJ initiation, which would explain their action to increase end-joining by this repair mechanism. MAR elements are also known to associate with nuclear matrix components such as SMARCAL1, a SWI/SNF family protein involved in gene expression regulation, DNA repair and stalled replication fork stabilization [273, 375]. However, whether the ATP-dependent strand-annealing helicase activity of SMARCAL1 might contribute to the SD-MMEJ mechanism has remained unexplored, as yet. Thus, additional experiments will be required to elucidate the possible interaction of MAR elements with components of the SD-MMEJ repair machinery.

Here, we identify the SD-MMEJ pathway as one of the primary mechanisms driving exogenous DNA integration in the genome of mammalian CHO cells. We also propose the existence of two distinct SD-MMEJ subpathways, relying on different subsets of enzymes. One of these mechanisms, dependent on Ligase III and polymerase θ, seems to be responsible for plasmid concatemerization. The other pathway, relying on Ligase I, polymerase δ and HR strand-invasion enzymes, seems to mediate plasmid recombination with the genome. Moreover, we propose that MAR elements may be able to stimulate

this latter mechanism to preferentially target plasmid DNA into potentially advantageous, gene-rich regions of the genome. This knowledge should in the future help to engineer cells for highly efficient recombinant protein production.

4.6 Supplementary materials

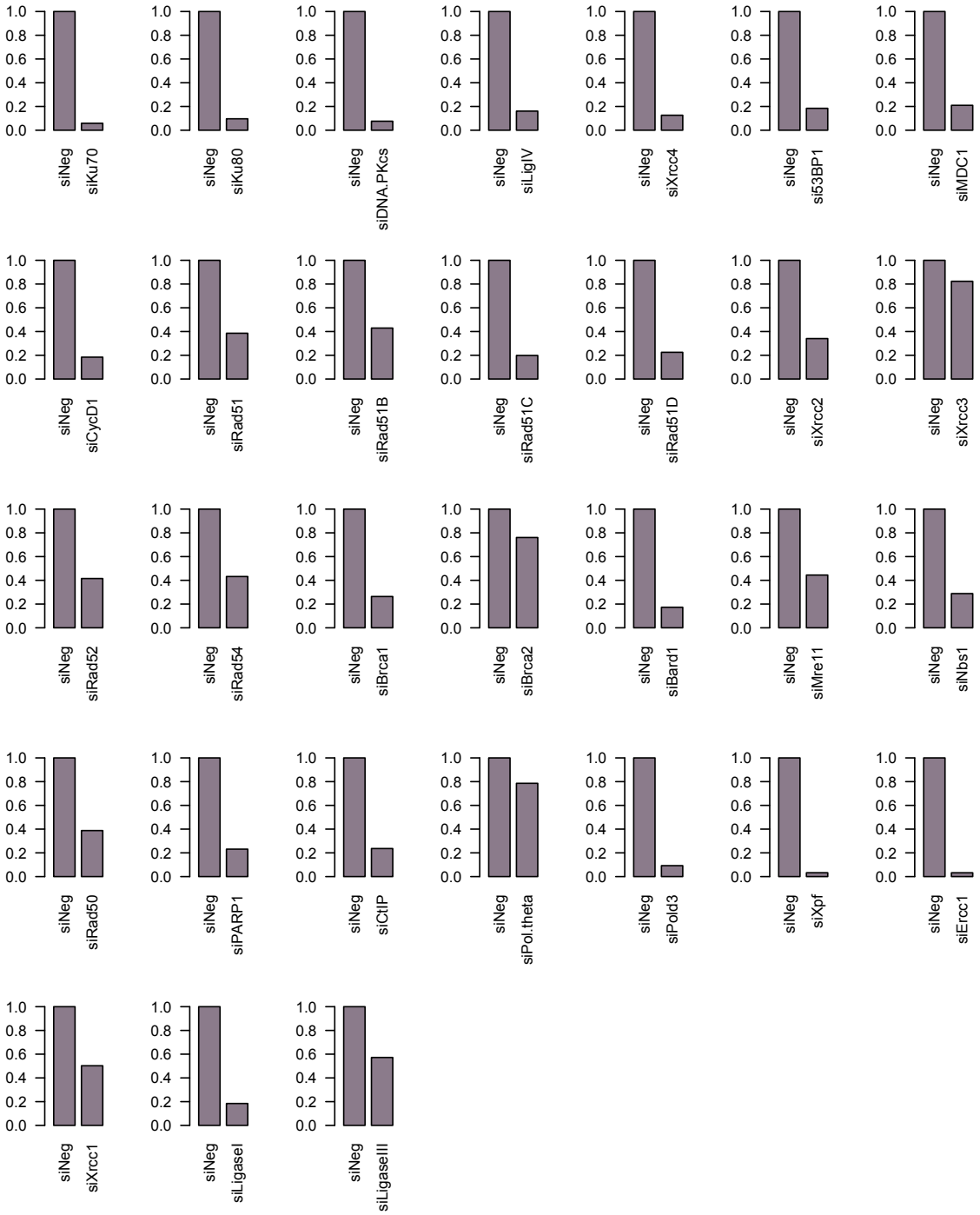


Figure 4S1: The effect of the siRNA on recombination protein mRNA level. Total mRNA was extracted from CHO DG44 cells transfected with three negative control siRNAs (siNeg) or with specific siRNAs targeting the indicated gene. The mRNA level fo the target was quantified by qPCR (following mRNA conversion to cDNA). Values were normalized to the target mRNA levels determined from cells transfected with the control siRNAs.

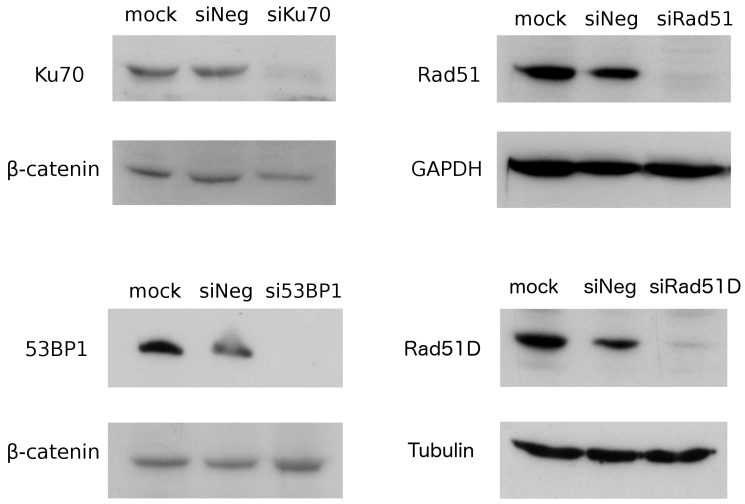


Figure 4S2: The effect of the siRNA knock-down of recombination proteins on the protein level. CHO cells were transfected with control (siNeg) or recombination protein-targeting siRNAs as for Fig.4S1, or left untreated (mock). Total cell extracts were resolved on SDS-PAGE gel, followed by immunoblotting using apropriate antibodies. Antibodies against house-keeping genes (GAPDH, β-catenin, Tubulin) were used as loading controls. Immunoblotting for A) Ku70, B) Rad51, C) 53BP1, and D) Rad51D.

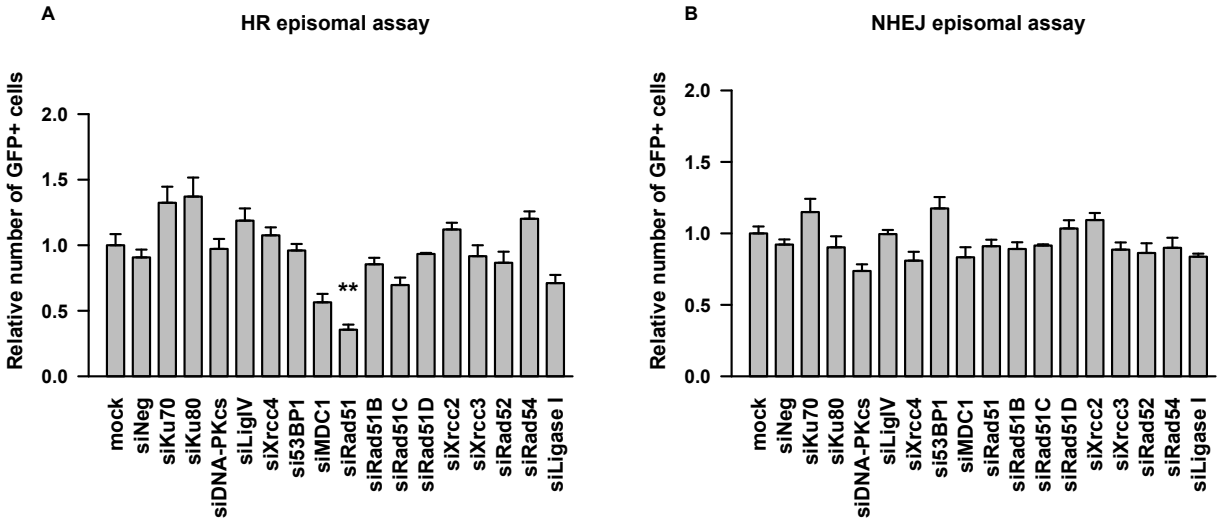
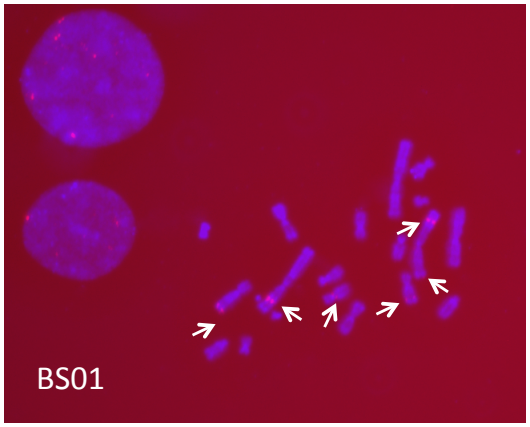
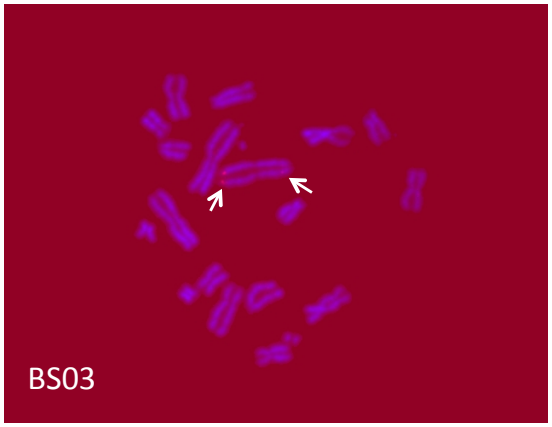


Figure 4S3: The effect of the siRNA knock-down of HR and NHEJ proteins on HR and NHEJ-mediated plasmid DSB repair. Cells left untreated (mock), transfected with control (siNeg) or recombination protein-targeting siRNAs as for Fig.4S1, and re-transfected with A) the HR reporter plasmid, or B) the NHEJ reporter plasmid. Percent of GFP-positive cells was normalized to the percent of dsRed-positive cells. Bars represent mean fold change over mock control. Mean of 3 experiments, error bars show s.e.m. Asterisks show significant differences between siRNA-treated samples and untreated (mock) control. Statistical signicance determined by unpaired Student's t-test with Benjamini-Hochberg correction; signicance level $p<0.05$ (*) and $p<0.01$ (**).

A



B



C

Clone	BS01	BS03
# integration loci (# of cells)	4 (3) 5 (4) 6 (7) 7 (3)	1 (3) 2 (5) 3 (1)

D

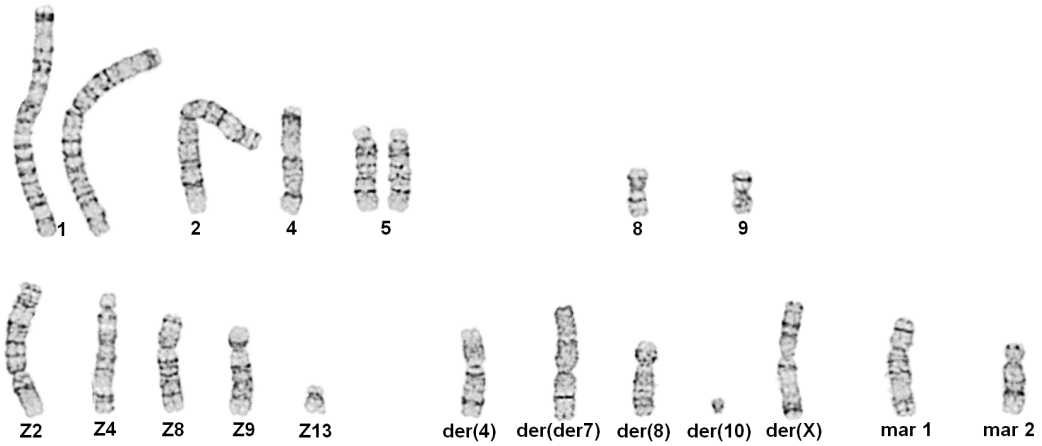


Figure 4S4: FISH analysis and karyotype of CHO-K1 cells. FISH on two CHO-K1 clones: A) BS01, and B) BS03. FISH was performed using probes targetting the vectors used to generate each clone. Chromosomes were counterstained with DAPI. Arrows point to the location of the transgene integration sites. A commercial analysis of the BS01 clone was also performed (200 nuclei analyzed; data not shown) giving a similar result. C) Quantification of integration site loci in analyzed chromosomal spreads. D) The karyotype of the CHO-K1 cells with total of 20 chromosomes (D. Martinet , Cytogenetics Lab , University Hospital, Lausanne). Metaphase chromosome spreads were stained with Giemsa. The normal hamster chromosomes are on top, followed by the Z group, derivative (der), and marker chromosomes (mar).

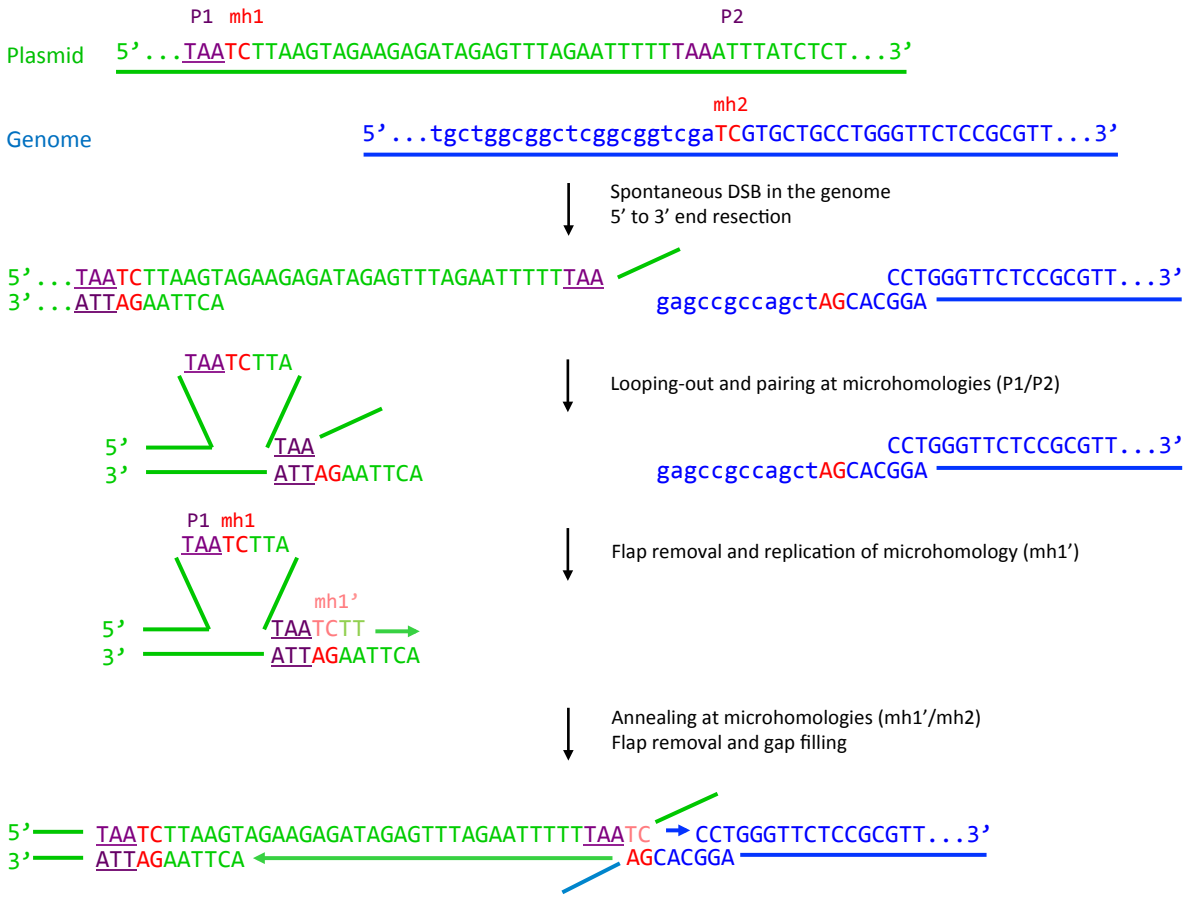


Figure 4S5: Example of a plasmid-to-genome junction and underlying SD-MMEJ mechanism. P1/P2– primer repeats, mh1/mh2 – microhomology repeats. Adapted from [372].

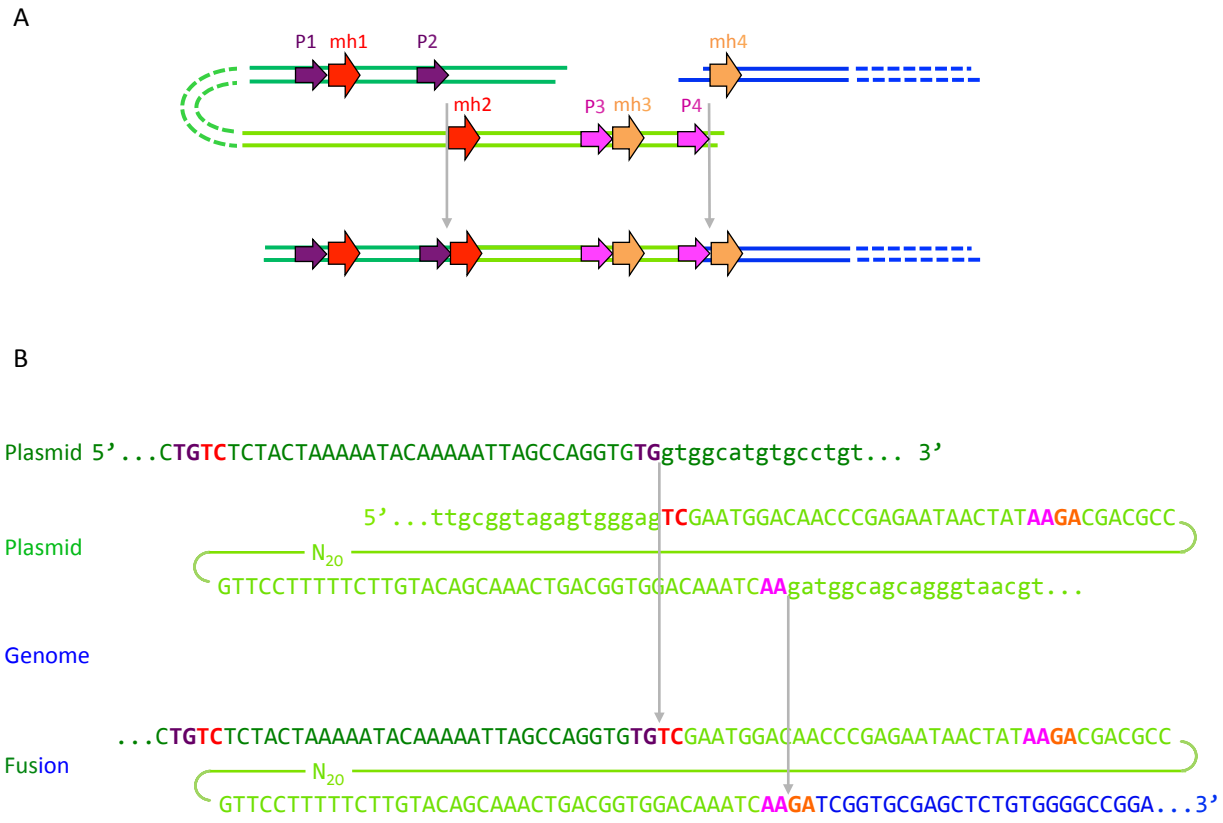


Figure 4S6: Example of a plasmid-to-genome junction and SD-MMEJ mechanism requiring a templated insertion. A) A scheme showing the mechanism of plasmid (dark green) joining with the genome (blue). Another fragment of the plasmid (light green) serves as an adaptor providing microhomlogies required for joining and becomes incorporated into the junction as a templated insert. B) Sequences of plasmid and genome fragments shown in A. P1/P2 and P3/P4 – primer repeats, mh1/mh2 and mh3/mh4 – microhomology repeats.

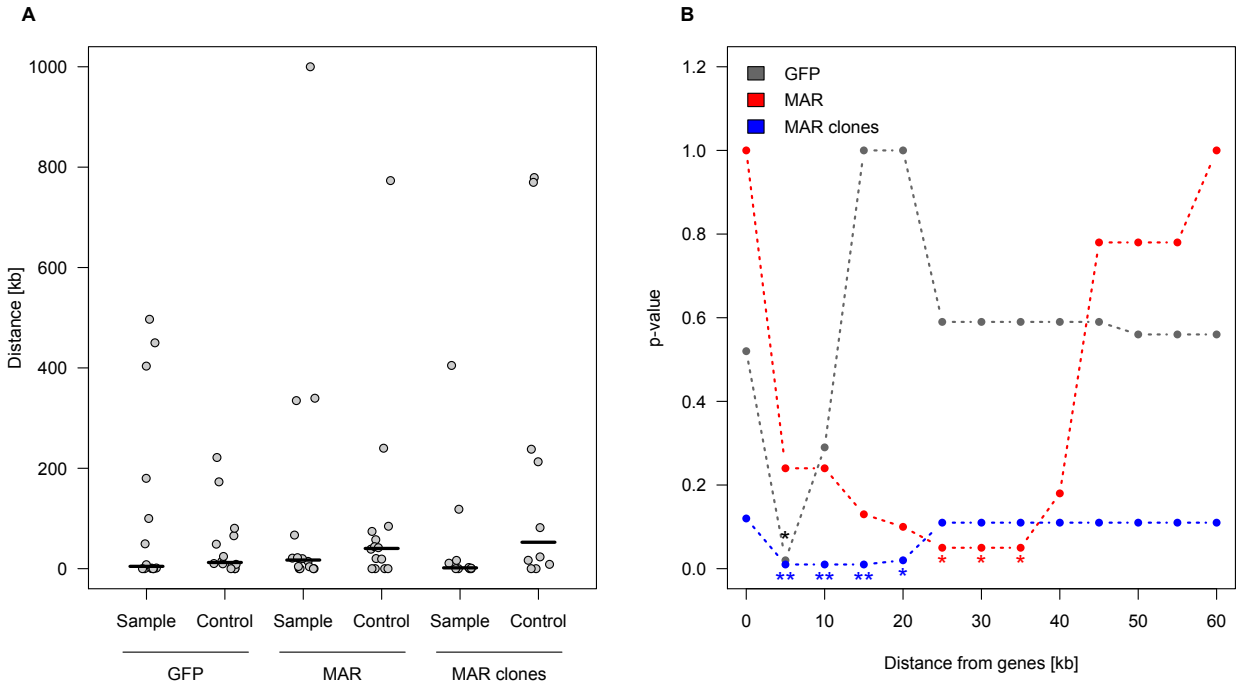


Figure 4S7: Analysis of plasmid integration loci in cells transfected with vectors containing or not a MAR. A) Plot of distances between the integration loci and the nearest CHO gene. Black lines indicate the median values. 'Sample' indicates sequenced genomic intergation sites in the samples, and 'Control' the control dataset of integration sites inserted at similar positions into randomly chosen scaffold of similar length. B) The relationship between the distance from the nearest gene and p-value was determined by an exact binomial test between each sample set and the corresponding control set. Samples represent integration loci from a polyclonal population of CHO cells transfected with the GFP plasmid without the MAR (GFP) or with the MAR (MAR), and stable high expressing CHO clones transfected with plasmids with the MAR (MAR clones). Asterisks indicate significant differences; $p \leq 0.05$ (*), $p \leq 0.01$ (**).

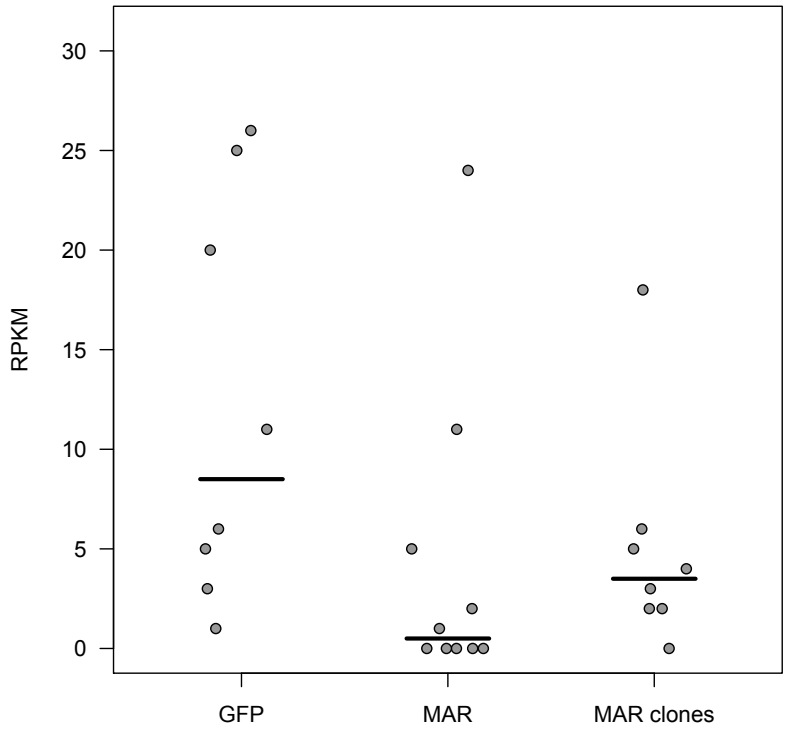


Figure 4S8: The expression of CHO genes in the proximity of transgene integration loci. The level of CHO gene expression, represented as reads per kilobase per million of mapped reads (RPKM), was assessed based on transcriptome sequencing of the parental CHO cells. Black lines indicate the median values. GFP: polyclonal cells transfected with the GFP plasmid without the MAR, MAR: polyclonal cells transfected with the MAR-containing plasmid, MAR clones: highly expressing selected CHO clones transfected with a MAR-containing plasmid.

Table 4S1: List of HR and NHEJ targets for siRNA knock-down.

Pathway	Target	References
NHEJ	Ku70/Ku80	[343]
	DNA-PKcs	[4, 84, 300]
	LigIV	[63]
	Xrcc4	[63, 185]
	53BP1	[361]
HR	MDC1	[202, 309, 361, 379]
	Rad51	[9, 19, 22, 341]
	Rad51B, Rad51C, Rad51D, Xrcc2, Xrcc3	[44, 192, 196, 217, 261, 315, 320]
	Rad52	[336, 337, 92]
	Rad54	[88, 126]
	Brca1	[61, 368]
	Bard1	[358]
	Brca2	[68, 92, 194, 233, 368]
	MRN (MRX in yeast)	[308]
	CtIP	[285, 370, 373]

Table 4S2: List of MMEJ targets for siRNA knock-down.

Target	References
MRN (MRX in yeast)	[73, 83, 184, 204, 380]
CtIP	[347, 374]
PARP1	[12]
Ercc1/Xpf (Rad1/Rad10 in yeast)	[184, 204]
Ligase I	[62, 187, 248, 254]
Ligase III (absent in yeast)	[12, 73, 187, 248, 254]
Xrcc1 (absent in yeast)	[73]
POLD3 (POL32 in yeast)	[60, 184]
Polymerase theta (POLQ) (absent in yeast)	[168, 372, 160]

Table 4S3: Sequences of plasmid-to-plasmid junctions.

original vector	CGGCAACTACAAGACCCGCGCCGAGTAGGGATAACAGGGTAATCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCA
Junction #1	CGGCAACTACAAGACCCGCGCCGAG----------GTGAAGTTCGAGGGCGACACCCTGGTGAACCGCA

original vector	**TCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGG**AGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAG TAGGGATAACAGGGTAATCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACC**GCATCGAGCTGAAGGGCATCGACTTCA**
Junction #2	**TCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAG**---------- ----------**GCATCGAGCTGAAGGGCATCGACTTCA**

original vector	CGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACT **TCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGG**AGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAG TAGGGATAACAGGGTAATCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCA
Junction #3	**TCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAG**---------- ----------**CCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCA**

original vector	TCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTAC CCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGA **CGGCAACTACAAGACCCGCGCCGAG**TAGGGATAACAGGGTAATCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGG**TGAACCGCA**
Junction #4	**CGGCAACTACAAGACCCGCGCCG**----------**TGAACCGCA**

original vector	**CAACTACAAGACCCGCG**CCGAGTAGGGATAACAGGGTAA**TCGCGCCGAG**GTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCG AGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATTCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATC ATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGGGGACGGCAGCGTGCAGCTCGCCGACCA CTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTACCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAG ACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAG
Junction #5	**CAACTACAAGACCCGCG**----------**ATCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCG**

original vector	CAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGC AGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACC **ATCTTCTTCAAGGACG**ACGGCAACTACAAGACCCGCGCCGAGTAGGGATAACAGGGTAATCGCGCCGAG**GTGAAGTTCGAGGGCGA**
Junction #6	**ATCTTCTTCAAGGACG**----------**TGAAGTTCGAGGGCGA**

original vector	CGGCAACTACAAGACCCGCGCCGAGTAGGGATAACAGGGTAATCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCA
Junction #7	CGGCAACTACAAGACCCGCGCCGAGT----------CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCA

original vector	**TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAA**GGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTAC AAGACCCGCGCCGAGTAGGGATAACAGGGTAAT**CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAA** GGGCATCGACTTCAAGGAGGACGGCAACATTCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAG
Junction #8	**TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAA**---------- ----------**CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAA**

original vector	CCCGCGCCGAGTAGGGATAACAGGGTAATCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGC
Junction #9	CCCGC----------CCTGGTGAACCGCATCGAGCTGAAGGGC

original vector	**GACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCG** **AAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA**CGACGGCAACTACAAGACCCGCGCCGAGTAGGGATAACAGGGTAATCGC GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGA**CTTCAAGGAGGACGGCAACATCCTGGG**
Junction #10	**AAGGCTACGTCCAGGAGCGCACCATCTT**---------- ----------**CTTCAAGGAGGACGGCAACATCCTGGG**

original vector	**CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCA** ACTACAAGACCCGCGCCGAGTAGGGATAACAGGGTAATCGC**GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAG** **CTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCAT**
Junction #11	**CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCAT**---------- ----------**GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAG**

original vector	**CCAGGAGCGC**ACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGTAGGGATAACAGGGTAAT**CGCGCCGAGGTGA** AGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTG GAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAA CATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACC ACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCC
Junction #12	**CCAGGAGCG**----------**CGCGCCGAGGTGA**

Note: The I-Sce1 cleavage site and microhomology (mh) sequences of the original vector are indicated by red and blue letters, respectively. The microhomologies of the SD-MMEJ model are underlined in the experimental junctions and original vector. The "–" signs indicate deleted bases at the junctions.

Table 4S4: Sequences of the DNA junctions of Immunoglobulin expression vector genomic integration sites.

Clone	Integration site #	Junction side	Sequence[1]	Mechanism[2]	Integration in/near a gene[3] expressed (yes/no)	Deletion in the genome (size)	Templated insert[4] (size)
BS01	1	Left	junction GCGAGCAGAACGGAGACTGAAGGGGTGGGGCCCGCGGCCGACAATGGGCGGGGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAA CGCGGAATGGCGAATGGCAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATA genome ...CTATTGTCCTTGCTC(206bp)TGGGGCCGCGGCCGATCGG... vector frgm 1 ...TTATTGACGTCAATGGGCGGGGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATG TAACGCGGAACTCCATATATGGGCTATGA... vector frgm 2 ...CCTGAATGGCGAATGGCAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAA...	SD-MMEJ (IR and DR) MMEJ (3 nt)	Ssh3 (exon), yes (RPKM=3)	no	yes (66bp)
		Right	junction AAAATACAAAAATTAGCCAGGTGTGTCGAATGGACAACCCGAGAATAACTATAAGACGACGCCACCCGTCCTTGATAGCGACGGTTCCTTTTTCTTGTA CAGCAAACTGACGGTGGACAAATCAAGATCGGTGCGAGCTCTGTGGGGCCGGACTGGGCACCAGTC vector frgm 1 ...ACCCTGTCTC(24bp)GGTGTGGTGG... vector frgm 2 ...AGACATTGCGGTAGAGTGGGAGTCGAATGGACAACCCGAGAATAACTATAAGACGACGCCACCCGTCCTTGATAGCGACGGTTCCTTTTTCTTGTA CAGCAAACTGACGGTGGACAAATCAAGATGGCAGC... genome ...GGCCGATCGGTGCGAGCTCTGTGGGGCCGGACTGGGCACCAGTCGGGCCCTTTAAGGGCCGCTCCCGA...	SD-MMEJ (DR) MMEJ (3 nt)			yes (100bp)
	2	Left	junction AAAATCCCCGCCCTTTGCTGGCGGCTCGGCGGTCGATGATGCCTCTTCGCTATTACGCCAGCCCAAGCTACCATGATAAG genome ...ATCCCCGCCCTTTGCTGGCGGCTCGGCGGTCGATCG... vector ...ATCGGTGCGGGCCTCTTCGCTATTA(1103bp)CTTGGTTGACGGCAATTTCGATGATGCAGC...	SD-MMEJ	Dph1 (exon), yes (RPKM=18)	no	yes (3bp)
		Right	junction TTAAGGTAATCTTAAGTAGAAGAGATAGAGTTTAGAATTTTTTAATCGTGCTGCCTGGGTTCTCCGCGTTCCTCCAGCGC vector ...AGGTAATCTTAAGTAGAAGAGATAGAGTTTAGAATTTTTTAAATTTATCTCT... genome ...CGATCGTGCTGCCTGGGTTCTCCGCGTTCCTCCAGCGCTGCCTTTTTGGTC...	SD-MMEJ (DR)			no
	3	Left	junction AAATTATACTGAGTAAGGTAACTCTGGACCAGAAATGAGAATCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTC genome ...TTATACTGAGTAAGGTAACTCTGGACCAGAAATGAGAATGACCTGTGTT... vector ...CAGCAGCATAGGGATCCGTCGATCGACCGATGCCCTTGAG(25bp)CGGTGGGCGCGGGGCATGACTATCGTCG...	MMEJ (2nt) SD-MMEJ (IR or DR)	intergenic (404.9kb from nearest gene)	n.a.[5]	no
	4	Left	junction GTCCCTACCCGCCTCCTGGCGCTGGTCTCCGACTTGGAGCGGCCCTGCATCTCAATTAGTCAGCAACCATAGTCCCGCCC genome ...CCTACCCGCCTCCTGGCGCTGGTCTCCGATCGGTCCTCCAGAGCGCGCG... vector ...ACGGGAGGTACTTGGAGCGGCCCTGCATCTCAATTAGTCAGCAAC(68bp)ATCGCTGACTAATTTTTTTT...	SD-MMEJ (IR or DR)	10.9kb from lin-54 homolog, yes (RPKM=4)	n.a.	no
	5	Right	junction ATGGTAGCTTGGGCTGGCGTAATAGCGAAGAGGCCCGCACCGGCGGGCCAATTTCTTTATGCATATCGCTGTGTCCTAGG vector ...TGGGCGGAGTTAGGGGCGGGACTATGGT(103bp)ACCGATCGCCCTT... genome ...TTTTCAAAAATGCCCAGGGCCAATTTCTTTATGCATATCGCTGTGTCCTAGGCCCTCGACAGGCA...	SD-MMEJ (DR)	Ankzf1 (exon) , yes (RPKM=2)	n.a.	yes (2bp)
	6	Right	junction ACATTCGACGATTATGCGATGCATTGGGTTCGTCAGATGAGGCTCTACAAGGTGAAATGTCTTCTTGCAGAGCCTCTCTA vector ...TTCGACGATTATGCGATGCATTGGGTTCGTCAGGCGCCTGGGAAGGGTC... genome ...TACTCTTTCGAGGGCTCAGTCCTATCAGATGAGGCTCTACAAGGTGAAATGTC(76bp)CTCTCTGACACGACTACCTCTTTCTC...	MMEJ (4nt) SD-MMEJ (IR)	16.5kb from Amigo1, no (RPKM=0)	n.a.	no

Table 4S4: Sequences of the DNA junctions of Immunoglobulin expression vector genomic integration sites (part2).

Clone	Integration site #	Junction side	Sequence[1]	Mechanism[2]	Integration in/near a gene[3] expressed (yes/no)	Deletion in the genome (size)	Templated insert[4] (size)
BS03	1	Left	junction TAATGCGTCCTTGCTTTGCTCTGAGCGCTTCTTGTCTCGAACAGGGGGGGTGGTCTTGTAGTTGTTCTCGGGCTGTCCGT genome ...TGCGTCCTTGCTTTGCTCTGAGCGCTTCTTGTCTCGATCGACACCCT... vector ...GGAGTCCAGCACAGGGGGGGTGGTCTTGTAGTTGTTCTCGGGCTGTCCGTTGGACTCCCATTC...	SD-MMEJ (IR)	**17.9kb from Bahcc1,** yes (RPKM=5)	yes (~913 bp)	no
		Right	junction GGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGGGGTGGGGAGCCAAGCCGCGGCGAGAGCGCAAAGCCCGCG vector ...GTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCCCACCCCAC... genome ...CGATCGGCCCGCAGGGGGTGGGGAGCCAAGCCGCGGCGAGAGCGCAAAGCCCGCG...	MMEJ (5nt) SD-MMEJ (DR)			no
	2	Right	junction GGCGAAGGGGCTCCTTAAGCGCAAGGCCTCGAACTCTCCCTCGCTTGAGTGCGGGGTCAGCTCAGGAAACCCTGTCTCTT vector ...GAAGGGGCTCCTTAAGCGCAAGGCCTCGAACTCTCCACCCACTTCC... genome ...TGGGGCACGATCGCTTGAGTGCGGGGTCAGCTCAGGAAACCCTGTCTCTTAAAAACCTTAAGC...	SD-MMEJ (DR)	**1kb from C17orf70,** yes (RPKM=6)	n.a.[5]	n.a.
Cp33/64	1	Left	junction GCTTTTCTAACTTAAATTATCTGGTTTCTCTTTAACTACAATGCTGCTGGTTTACAGACCACATGTAGAGTGGCAATGTG genome ...TTTCTAACTTAAATTATCTGGTTTCTCTTTAACTACAATTTGCCTCT... vector ...GGGCATCGGTCGATCGACGGATCCCTATGCTGCTGGTTTACAG(147bp)TTTAGCAAGACAGTGATAATGCTAATATG...	MMEJ (2nt) SD-MMEJ (DR)	**Cblb (intron),** yes (RPKM=2)	yes (320 bp)	no
		Right	junction TGCCATCCAGCACATTGCCACTCTACATGTGGTCTGTAAACCAGCAGCATACAGCATAAACAAATGTAACACAACTTAAA vector ...TGAAATTACAATACAAATGATGAT(100bp)TGTAAACCAGCAGCATAGGGATCCGTCGAT... genome ...CAAATGTTCCACAGCATAAACAAATGTAACACAACTTAAACTAATATTTCACA...	SD-MMEJ (DR)			no
	2	Left	junction GCCCTGCCACTTTGAACATACTTTCTTGTTTACTTATTACTATTTTCCTTTTATATGTTTAAATCATCATTTGTATTGTAATTTCATCTTCCA genome ...CTGCCACTTTGAACATACTTTCTTGTTTACTTATTATGTTTATCCC... vector ...GAACAATTTTTATTTTCCTTTTATATGTTTAAATCATCATTTGTATTGTAATTTCATCTTCCA...	SD-MMEJ (DR)	**intergenic** (118.4kb from nearest gene)	no	yes (1nt)
		Right	junction ACCTCCCGTACCTTAATATTACTTACTTATCATGGTAGCTTGTTTATCCCTTAGTCTTCCCACTCCTGACTTGAATGCTT vector ...TCCCGTACCTTAATATTACTTACTTATCATGGTAGCTTGGGCTGGCG... genome ...TTACTTATTATGTTTATCCCTTAGTCTTCCCACTCCTGACTTGAATGCTTACTCTTTGAGGAG...	MMEJ (2nt) SD-MMEJ (DR)			no

[1]Sequenced plasmid integration junctions are represented by blue (CHO genome) and green (vector sequence) letters, as predicted from whole genome sequencing of Illumina genomic and mate-pair libraries, and as validated experimentally by PCR amplification and direct sequencing of the junctions.
[2] This lists the most probable mechanisms accounting for the junction, consisting of the Microhomology-mediated end-joining (MMEJ) and/or synthesis-dependent MMEJ (SD-MMEJ). The length of the microhomology is indicated in parenthesis. IR and DR indicate the use of inverted or direct repeat as a template for SD-MMEJ, respectively.
[3]Genomic integration site locus. The gene nearest to the integration site is listed in bold. Gene expression was assessed by total RNA sequencing of the parental CHO cells. Reads per kilobase of transcript per million reads mapped (RPKM) are used as a measure of expression level.
[4]Insertion of nucleotides templated from another part of the vector or genome.
[5] n.a., not annotated

Table 4S5: Analysis of plasmid integration sites in cells transfected with vectors with or without the MAR element.

Sample[1]	Integration within genes[2]	Integration near genes[3]	Expressed genes[4]
Polyclonal population without MAR	7/14 (*)[5]	8/14	8/8
Polyclonal population with MAR	6/14	10/14 (*)	5/10
High expressing clones with MAR	6/10 (**)	8/10	7/8

[1] Polyclonal populations of CHO cells transfected with GFP or MAR-GFP plasmids were sequenced by high-throughput sequencing (Pacific Biosciences) and plasmid-to-genome junctions were predicted using bioinformatics tools. Integration sites in high expressing CHO clones transfected with MAR-containing plasmids were PCR-amplified and sequenced using Sanger sequencing.
[2] integration locus inside or within 5kb from a gene
[3] integration locus within 35kb from a gene
[4] number of expressed genes in the neighborhood (within 35kb) of the integration locus
[5] Statistical significance calculated between each sample set and the corresponding control set using an exact binomial test. Significance levels $p \leq 0.05$ (*), $p \leq 0.01$ (**).

Chapter 5

Polymerase θ stimulates DSB repair by a MMEJ mechanism involving long DNA end resection

This chapter is based on a manuscript in preparation entitled "Polymerase θ stimulates DSB repair by a MMEJ mechanism involving long DNA end resection" by **Kostyrko, K.**, Mermod, N.

5.1 Abstract

DNA double stranded breaks (DSBs) are one of the most deleterious type of DNA lesions. The main pathways responsible for repairing these breaks in eukaryotic cells are homologous recombination (HR) and non-homologous end-joining (NHEJ). However, it is becoming increasingly clear that a third group of still poorly characterized DSB repair pathways also exists in cells. These mechanisms are collectively termed microhomology-mediated end joining (MMEJ), highlighting their main feature, i.e. the use of short homologies in the end-joining process. Here, I constructed GFP reporter assays to characterize two variants of these alternative pathways – simple MMEJ and synthesis-dependent (SD)-MMEJ. The use of these assays in Chinese hamster ovary (CHO) cells demonstrated that while MMEJ is able to mediate relatively efficient DSB repair if longer (9 bp) microhomologies are present, the majority of DSBs are repaired using the highly error-prone SD-MMEJ pathway. I also performed siRNA knock-down of different genes proposed to play a role in MMEJ. The depletion of most of these factors did not influence the relative efficiencies of the different end joining pathways. However, the knock-down of polymerase θ inhibited DNA end resection and repair through simple MMEJ in favor of other end-joining pathways. This suggested that this enzyme may be important for MMEJ, while DNA synthesis in SD-MMEJ may be mediated by a different low fidelity polymerase.

5.2 Introduction

During their lifetime cells constantly face DNA damage caused by byproducts of normal metabolic processes or exogenous factors, such as chemical agents or ionizing radiation (IR). One of the most deleterious types of DNA damage are double-stranded breaks (DSBs), which can cause problems during replication and transcription, or lead to the loss of chromosome fragments. Eukaryotic cells possess many mechanisms to sense and repair these types of breaks. The two major pathways responsible for DSB repair are non-homologous end-joining (NHEJ) and homologous recombination (HR) [212]. These two mechanisms compete for broken DNA ends in the cell, and the choice between them is made depending on the type of the DSB and the phase of the cell cycle. NHEJ, the main pathway used in higher eukaryotes, is active throughout the cell cycle. It is a fast process, which very efficiently repairs easily ligatable DSBs. In contrast, HR is a much more complex mechanism, active mainly in the late S and G2 phase of the cell cycle. HR is considered an error-free pathway as it can repair very complicated DSBs with high fidelity. However, it requires extensive DNA end processing and a homologous DNA molecule as a template.

In recent years it became apparent that a third mechanism of DSB repair also exists in eukaryotic cells [21, 346]. This pathway, in normal cells masked by the main repair processes, has many names: alternative end-joining (alt-EJ), alternative or backup NHEJ (alt-NHEJ, a-NHEJ, B-NHEJ), or microhomology mediated end-joining (MMEJ) [12, 28, 73, 108, 254, 346]. It is still unclear whether it comprises one or more different mechanisms. There seems to be no unanimous view on the subject, as reflected by the ambiguous nomenclature. Here I will refer to these pathways collectively as MMEJ, to underline their common feature, i.e. the use of very short 2-25 nt homologies in the alignment of the broken ends before joining.

Many fundamental findings on the functioning of the DSB repair pathways have been made using in vivo plasmid end-joining assays. These assays are most commonly based on the reconstitution of a functional reporter gene by one of the DSB repair mechanisms after the induction of a break in a nonfunctional substrate. Many such assays were constructed to investigate NHEJ and HR in various types of cells [18, 122, 171, 212, 266, 279, 290, 291, 380], and recently also MMEJ [248, 328, 347]. However, with a growing number of new studies reporting mechanistically distinct MMEJ pathways, the need arises for new, more specific assays to distinguish between these pathways.

Here, I sought to design a GFP-reconstitution assay to measure synthesis-dependent (SD)-MMEJ, one of the MMEJ variants, recently proposed by Yu and McVey [372]. Both MMEJ and SD-MMEJ start with the 5' to 3' end resection [328, 372], similarly to HR, but diverge in later steps. Repair is carried out by MMEJ if the resulting ssDNA overhangs contain short regions of homology, which can pair together to mediate alignment of the two sides of the DSB. However, if the extensive resection fails to expose any microhomologies, the MMEJ machinery may be unable to re-join the two ends. Moreover, the presence of long ssDNA overhangs precludes the use of the NHEJ pathway, which cannot process these types of substrates [234, 383]. The SD-MMEJ model offers a solution in such situations. In this mechanism, a non-processive DNA polymerase copies a sequence from up- or downstream of the break, which subsequently serves to align the two sides of the break enabling the continuation of the MMEJ pathway.

Here, I attempted to construct an SD-MMEJ assay and compared it with a previously published reporter designed to measure simple MMEJ, where the rejoining of a functional GFP sequence by either mechanism can be followed by flow cytometry [347]. While MMEJ seemed to be more efficient in reconstituting the GFP sequence, sequencing analysis revealed that the repaired DNA junctions extracted from GFP-positive cells resulted from

both types of mechanisms, although SD-MMEJ was more frequently used than simple MMEJ. This suggests that SD-MMEJ, while error-prone, is a very robust mechanism able to repair difficult, incompatible DSBs without any need for pre-existing homology. When combining the MMEJ assay with siRNA knock-down of genes involved in alternative end-joining pathways, I observed that the depletion of polymerase θ decreased the efficiency of MMEJ in favor of NHEJ and SD-MMEJ. These results are in contrast with previous reports implicating this polymerase in the SD-MMEJ pathway, indicating that the SD-MMEJ mechanism may rely on the activity of more than one DNA polymerase.

5.3 Materials and Methods

Cell culture

Adherent CHO DG44 cells [333] were cultivated in DMEM/F12+GlutaMAX^{TM} supplemented with 1x HT and 10% fetal bovine serum (Gibco, Invitrogen), and with the antibiotic-antimycotic solution (Sigma-Aldrich, #A5955).

Construction of the recombination assays

The MMEJ and SD-MMEJ reporter cassettes were constructed by interrupting the GFP coding sequence present in the pSV40-GFP plasmid (described previously [116]) with restriction sites. The MMEJ vector was based on a previously described reporter [347]. Briefly, a naturally occurring 9b sequence (CGCGCCGAG) was duplicated and an 18bp I-SceI recognition site was inserted in between the two copies of the sequence. Two in frame stop codons present in the inserted sequence prevent the expression of a functional GFP from the intact vector. Digestion with I-SceI linearizes the vector and creates a DSB with 3' overhangs.

Two SD-MMEJ reporter cassettes were designed using microhomologies already present

in the GFP sequence. In the first assay (SD-MMEJ-1) the two 5bp microhomologies (CGAGG) are 7bp apart. In the second assay (SD-MMEJ-2) the two 7bp microhomologies (CCACCCT) are 5bp apart. To enable the formation of DSBs with 5' incompatible overhangs and prevent re-ligation upon digestion, two restriction sites (separated by 3 bp) were introduced into each vector inside one of the microhomologies. SpeI and AflII recognition sites were used in SD-MMEJ-1, and AflII and EcoRI in SD-MMEJ-2. In both cases, in-frame stop codons are present inside the restriction sites to prevent GFP expression from the intact vectors.

Transfection and FACS analysis

MMEJ and SD-MMEJ plasmids were digested with the appropriate restriction enzymes (New England Biolabs) for 5h and purified by ethanol precipitation. Aliquots were analyzed by gel electrophoresis to confirm complete digestion. CHO cells were transfected with the linearized plasmids, and with the pGL3-CMV-dsRed plasmid, to normalize for transfection efficiency, using Fugene 6 according to manufacturer's instructions (Promega). The pSV40-GFP vector (pGFP) was transfected in parallel as a positive control of GFP expression. Expression of GFP and DsRed was monitored by fluorescence microscopy (Carl Zeiss Microscope Axio Observer.A1) and flow cytometry. For flow cytometry cells were harvested 24h following transfection and resuspended in 0.5 ml of PBS with 2% FBS (Gibco, Invitrogen). Data was acquired using the CyAn analyzer (Beckman Coulter) and analyzed using the FlowJo software (Tree Star). GFP repair efficiency was calculated as a ratio of GFP-positive cells over the number of dsRed-positive cells.

Junction sequence analysis

For the analysis of junction sequences, CHO cells were transfected with the MMEJ and

SD-MMEJ vectors but without the pGL3-CMV-dsRed plasmid, using Fugene 6 according to manufacturer's instructions (Promega). After 24h cells were harvested and GFP-positive cells were sorted by FACS (MoFlo Astrios Cell Sorter, Beckman Coulter). Total DNA was isolated using the DNeasy Blood & Tissue Kit (Qiagen). GFP sequences were amplified by PCR using primers GFP-NcoI-F1 (ATTCCGGTACTGTTGGTAAAGCCACCA) and GFPp2-Rev (TGTATCTTATCATGTCTGCT). PCR products were cloned into the NcoI and XbaI-cleaved pSV40 vector. Ligation mixtures were transformed into recombination deficient E. coli cells (XL10-Gold Ultracompetent Cells, Stratagene) and plated on LB and ampicillin plates. Colonies were picked and analyzed by colony PCR for the presence of the insert using primers GFP-NcoI-F1 and SV40_lateR (TCCAAACTCATCAATGTATC). Plasmids isolated from positive clones were sequenced by Sanger sequencing.

siRNA and transfections

Small interfering RNA duplexes were specifically designed to target the Chinese hamster homologs of DNA-PKcs, Ku70, Ligase I, Ligase III, polymerase θ and Pold3. The siRNAs were designed and provided by Microsynth AG (Balgach, Switzerland). Three RNA duplexes were designed per mRNA to increase the probability of successful knock-down. Three negative (non-targeting) siRNAs were also designed as controls. For siRNA-mediated knock-down, CHO-DG44 cells were transfected with equimolar amounts of three targeting or non-targeting siRNA duplexes at a final concentration of 50nM using Lipofectamine RNAiMAX, according to the manufacturer's instructions (Invitrogen). After two days, the siRNA-treated cells were re-transfected with the pre-digested MMEJ vector using Fugene 6 (Promega). GFP-positive cells were sorted and junctions amplified and sequenced as described above.

5.4 Results

5.4.1 CHO cells restore GFP expression more efficiently from the MMEJ than the SD-MMEJ reporters

To measure the efficiency of extrachromosomal MMEJ and SD-MMEJ in mammalian cells, I constructed GFP reporter assays aimed at specifically detecting these pathways. The MMEJ repair substrate contains two 9bp microhomologies flanking the I-SceI restriction site (Fig.5.1A). After I-SceI digestion and end-resection, annealing at these microhomologies should enable the restoration of a functional GFP coding sequence (Fig.5S1). Two versions of the SD-MMEJ reporter were constructed, differing in size of microhomologies and distance between them. They contain a GFP sequence interrupted by two tandem restriction sites, which serve to create a DSB with non-complementary 5' overhangs (Fig.5.1B, 5S2, 5S3). Since there are no extended microhomologies in the sequence surrounding the break, they should be amplified from another fragment of the vector by a DNA polymerase, as illustrated in the SD-MMEJ mechanism shown on Fig.4S5. In the two substrates the potential 'primer' sequences (P1 and P2), as well as the microhomology (μ1), which needs to be amplified to properly align the two broken ends, are located on one side of the DSB. The use of these direct repeats should enable the restoration of a functional GFP sequence.

The linearized vectors were transiently transfected into CHO cells and the appearance of GFP-positive cells was analyzed by flow cytometry. Over 50% of cells transfected with the MMEJ reporter were able to reconstitute a functional GFP coding sequence (Fig.5.2). In contrast, only approx. 9% of the cells transfected with one of the SD-MMEJ assays (SD-MMEJ-1) successfully repaired the GFP gene. The second SD-MMEJ reporter (SD-MMEJ-2) yielded only a few GFP-positive cells (approx. 0.5%). The difference in efficiency between the two SD-MMEJ constructs could result from the choice of the

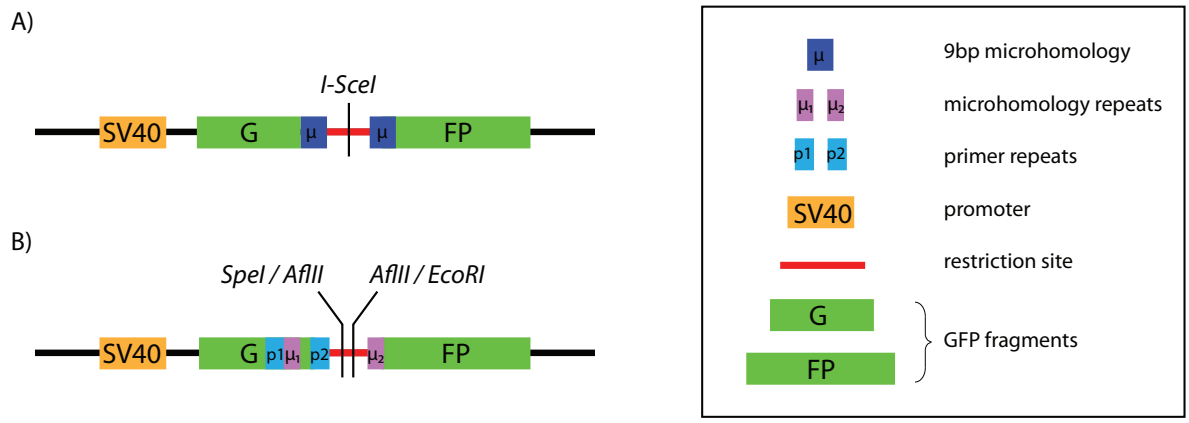


Figure 5.1: GFP-reconstitution reporter cassettes for detection of simple MMEJ and SD-MMEJ. A) MMEJ reporter, B) SD-MMEJ reporter. Description and abbreviations in text.

restriction enzymes used to generate the break or from the two 'primer' sequences not being spaced far enough in the second assay (Fig.5S2, 5S3).

These results suggested that the MMEJ mechanism is more efficient than SD-MMEJ pathway, possibly because of the presence of pre-existing, relatively long (9bp) microhomologies, which may provide a greater chance of successful GFP reconstitution. In contrast, the lack of homology in the SD-MMEJ assays may force the repair machinery to copy the microhomology from a more distant region of the plasmid, which may not necessarily lead to the reconstitution of a functional GFP. To ultimately verify if all the reporter cassettes were repaired as expected in CHO cells, and to unambiguously identify the mechanisms used, I sequenced the reconstituted plasmids isolated from the GFP-positive cells.

5.4.2 MMEJ and SD-MMEJ reporters are more frequently repaired by the SD-MMEJ pathway

The sequences of 12 junctions were obtained from cells transfected with the MMEJ vector and 5 and 4 junctions from cells transfected with either the SD-MMEJ-1 or the SD-MMEJ-2 vector, respectively. The sequences were then analyzed for the presence of microhomologies, deletions and potential primer/microhomology pairs in the vicinity of the junction. The

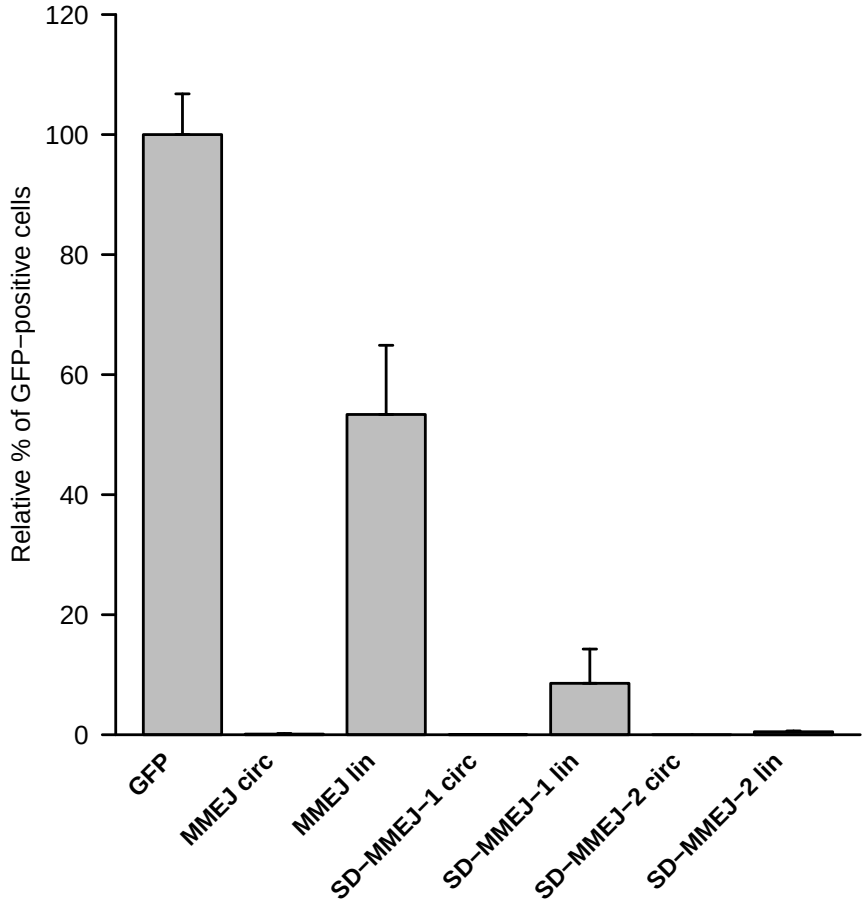


Figure 5.2: Frequency of GFP reconstitution from the MMEJ and SD-MMEJ assays in CHO cells. Cells were transfected with a control GFP plasmid (GFP), non-digested (circ) or linearized MMEJ or SD-MMEJ reporters (lin). Percentages of GFP-positive cells are shown with respect to cells transfected with the control plasmid (shown as 100%). Mean of 3 experiments, s.e.m. error bars.

presence of limited sequence loss (0-6bp) and a lack of microhomology was interpreted as the result of the NHEJ mechanism. Junctions showing evidence of long end-resection (>6 bp), with at least 2 bp of pre-existing microhomology, were classified as simple MMEJ products. If these short homologous sequences were absent, I searched for $\geq$ 2bp + $\geq$ 2bp direct or inverted repeats up- or downstream of the junction, as well as for the presence of templated inserts, which are hallmarks of the SD-MMEJ pathway.

Analysis of all the sequenced repair products revealed that the majority (60-75%) of the junctions from cells transfected with both types of assays must have resulted from the SD-MMEJ mechanism (Table 5.1, 5S1). The MMEJ mechanism seemed to account for only

25-40% of all the junctions analyzed, despite the presence of the 9 bp microhomology region in the vicinity of the cleavage site. Surprisingly, I did not observe any repair products that could be unambiguously attributed to the NHEJ pathway. These results indicate that not only the SD-MMEJ pathway is more efficient than simple MMEJ, but also that the appearance of GFP fluorescence in these kind of reporter assays cannot be used as a sole determinant of repair efficiency. Therefore, sequencing of the repaired junctions should always be used to reliably estimate the true efficiency of end-joining mechanisms in the cells.

Table 5.1: Analysis of the sequenced repair products.

Reporter	Repair mechanism number of junctions (%)		Deletion size[1] [bp]	Pre-existing microhomology size [bp] (number of junctions)	Primer repeat [bp] + microhomology [bp] (number of junctions)
	MMEJ	SD-MMEJ			
MMEJ	3 (25%)	9 (75%)	17, 21, 27, 53, 54, 54, 64, 73, 81, 91, 109, 117	9 (2), 3 (1)	2+2 (2), 3+2 (2), 2+3 (3), 3+3 (1), 2+4 (1)
SD-MMEJ-1	2 (40%)	3 (60%)	10, 20, 30, 71, 103	9 (1), 4 (1)	2+3 (2), 4+2 (1)
SD-MMEJ-2	1 (25%)	3 (75%)	24, 48, 52, 76	3 (1)	3+2 (1), 2+4 (1), 3+4 (1)

[1] Size of the sequence missing from the reporter construct after repair. Deletions are ordered from smallest to largest. A deletion of 27bp would be expected from the MMEJ mechanism using the 9bp microhomology on its cognate assay plasmid.

Out of the 6 MMEJ-type junctions obtained from all assays, half occurred at 9 bp microhomologies (Table 5S1). In the 3 remaining cases, 3 bp and 4 bp of homology were used for alignment. This indicated that the MMEJ mechanism shows a preference for longer microhomologies. Interestingly, out of all the sequences obtained from cells transfected with the MMEJ reporter, only one contained a GFP sequence successfully reconstituted using the preset 9 bp microhomology (Table 5S1, junction #4). This may suggest that precise, error-free repair events of the most proximal microhomology domains

correspond to a very low percentage of all events.

In contrast, none of the SD-MMEJ constructs was repaired using the designed microhomologies, even though the SD-MMEJ mechanism often relied on 5bp (2 + 3 bp) microhomology sequences (8/15 of SD-MMEJ-like junctions), as in the SD-MMEJ-1 assay (Table 5.1, Tables 5S1 and 5S2). In most cases (9/15) the 'primer' sequence was 2 bp long and the amplified microhomology used for bridging the breaks was 2 bp (6/15) or 3 bp (6/15) long. This suggested that even very short sequences can serve as starters for the SD-MMEJ DNA polymerase, making this process very robust in repairing breaks completely devoid of extended microhomology. This, however, likely entails poor repair fidelity, consistent with the low frequency of successful GFP reconstitution events from the SD-MMEJ reporters.

Interestingly, the repair of the MMEJ cassette was accompanied by larger deletions than the SD-MMEJ assays (Table 5.1, Fig.5.3). The deletions in the MMEJ reporter plasmid were on average 16 and 13 bp longer than in the SD-MMEJ-1 and SD-MMEJ-2 plasmids, respectively. This difference almost exactly reflects the difference between the 5' protruding ends in the SD-MMEJ vectors and recessive 5' ends in the MMEJ vector (Δ=16 bp), suggesting that the extent of the 5' to 3' end resection was similar for all three constructs. I concluded that all the reporter constructs can be used to study both types of repair mechanisms.

Taken together, these results indicate that, while the simple MMEJ pathway is potentially more precise if larger homologies are present, the SD-MMEJ pathway seems to be much more frequently used. I hypothesize that the SD-MMEJ mechanism plays a role of a salvage repair pathway when other mechanisms have failed.

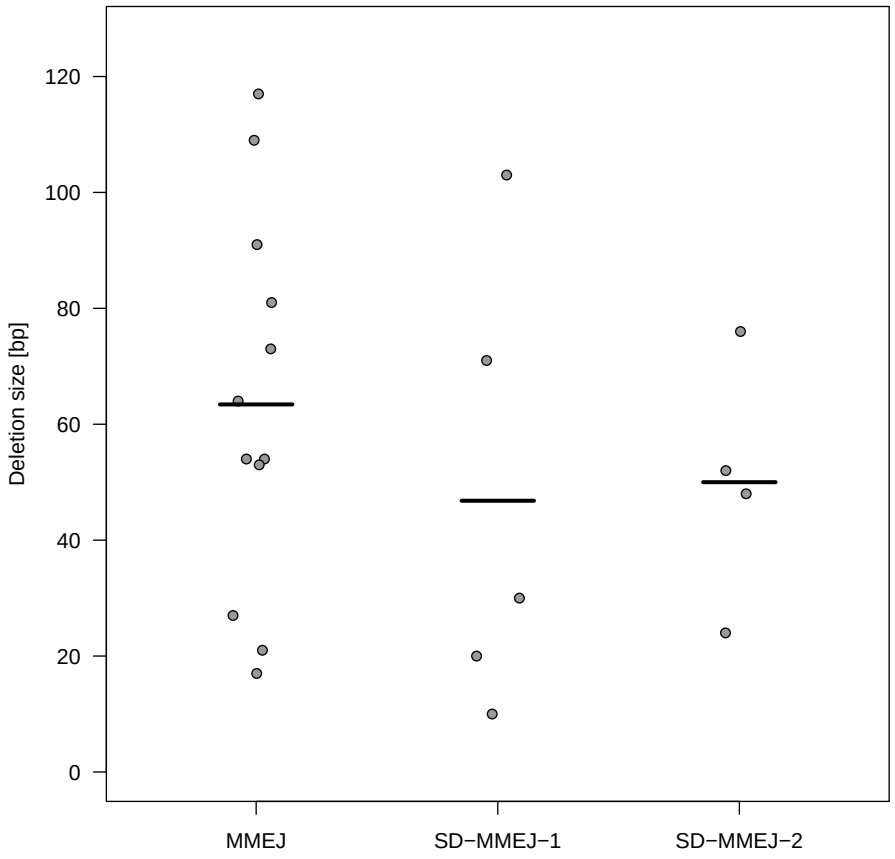


Figure 5.3: The distribution of deletion sizes in junctions from cells transfected with the MMEJ, SD-MMEJ-1 and SD-MMEJ-2 reporters. Black lines represent mean deletion sizes.

5.4.3 Depletion of polymerase θ inhibits DNA end resection and repair through MMEJ in favor of other end-joining pathways

I next used the above described strategy to assess the relative efficiency of MMEJ and SD-MMEJ pathways in cells depleted of factors thought to be implicated in these processes. I therefore used short interfering RNA (siRNA) to knock-down Ligase I, Ligase III, DNA polymerase θ and DNA polymerase δ subunit 3 (Pold3), all of which were previously reported to play a role in alternative DSB repair pathways [47, 60, 184, 187, 248, 254, 371]. I also used siRNAs against two NHEJ genes – Ku70 and DNA-PKcs, and a non-targeting control siRNA. The decrease in the target mRNA levels upon siRNA treatment was verified

by real-time PCR (Fig.5S4). Since the reporter vectors did not seem to significantly differ in the relative frequency of the MMEJ and SD-MMEJ mechanisms, I applied only the MMEJ reporter, as it allows for higher recovery of GFP-positive cells.

Surprisingly, in cells transfected with the control siRNA, the frequency of MMEJ-attributed repair (60%) was higher than that of SD-MMEJ (30%) (Fig.5.4A, Table 5S2), which contrasted with the results obtained in untreated cells. This suggested that the siRNA transfection alone could alter the frequency with which the two pathways are used. I also observed some junctions that could only be attributed to the NHEJ mechanism (10%). Similar repair patterns were observed in cells depleted in Ku70, DNA-PKcs, Ligase I and Pold3, indicating that these factors may not be essential for the repair of extrachromosomal DSBs. However, it should be noted that the knock-down of DNA-PKcs and Ku70 increased the number of GFP-positive cells about 2-fold (data not shown), in line with the view that alternative end-joining pathways are more active when NHEJ is disabled [28, 72, 152, 186]. Moreover, DNA-PKcs-depletion slightly increased the number of junctions repaired with the MMEJ mechanism. Interestingly, the knock-down of polymerase θ significantly decreased the frequency of simple MMEJ (to 20%), in favor of NHEJ (to 30%) and SD-MMEJ (to 50%), suggesting that it plays an important role in the first pathway. This may also indicate that another low fidelity DNA polymerase is used in SD-MMEJ when polymerase θ is absent. An increase in SD-MMEJ repair was also noted in cells depleted of Ligase III, although in this case the frequency of MMEJ did not change.

The analysis of deletion sizes in the recovered junctions demonstrated that loss of polymerase θ led to a decrease in medium-sized deletions in favor of very short ones (Fig.5.4B, Fig.5S5), consistent with the increased usage of the NHEJ pathway. This potentially indicates that the lack of this polymerase may limit DNA end resection. The

number of large deletions was decreased in favor of medium-sized deletions in the absence of Ligase I and III, suggesting that lack of these ligases may also lead to the repair processes that limit resection, but to a lesser extent.

In summary, it seems that polymerase θ mediates DSB repair by a MMEJ mechanism involving long DNA end resection process, as the depletion of this enzyme seems to result in an increased usage of the other end-joining pathways, i.e. NHEJ and SD-MMEJ. I hypothesize that polymerase θ-independent SD-MMEJ may rely on a different low fidelity enzyme.

5.5 Discussion

DSBs are potentially genotoxic DNA lesions, which need to be efficiently repaired by the cells to prevent chromosomal aberrations or serious DNA damage, leading to carcinogenesis or even cell death. The two major pathways responsible for DSB repair in eukaryotes are HR and NHEJ. However, in recent years it became apparent that these mechanisms are assisted by a family of alternative DSB repair pathways, collectively termed MMEJ. These mechanisms are thought to come into play when the main repair pathways are insufficient to repair all the breaks that arise in cells. This is often the case in cancer cells, which suffer from high levels of oxidative stress [193, 318]. Indeed, many reports show that such alternative MMEJ-like DSB repair pathways are more active in tumor cells [23, 323].

Studies of these alternative repair mechanisms advanced largely due to the development of plasmid recombination assays, which allow to measure the efficiency of end-joining in the cells. These assays are often based on transiently transfected or genome-integrated reporter substrates, in which DSBs are induced by restriction enzymes. Restoration of reporter gene expression serves to assess the efficiency of DSB repair. In addition the structure of the repaired products is often verified by sequencing.

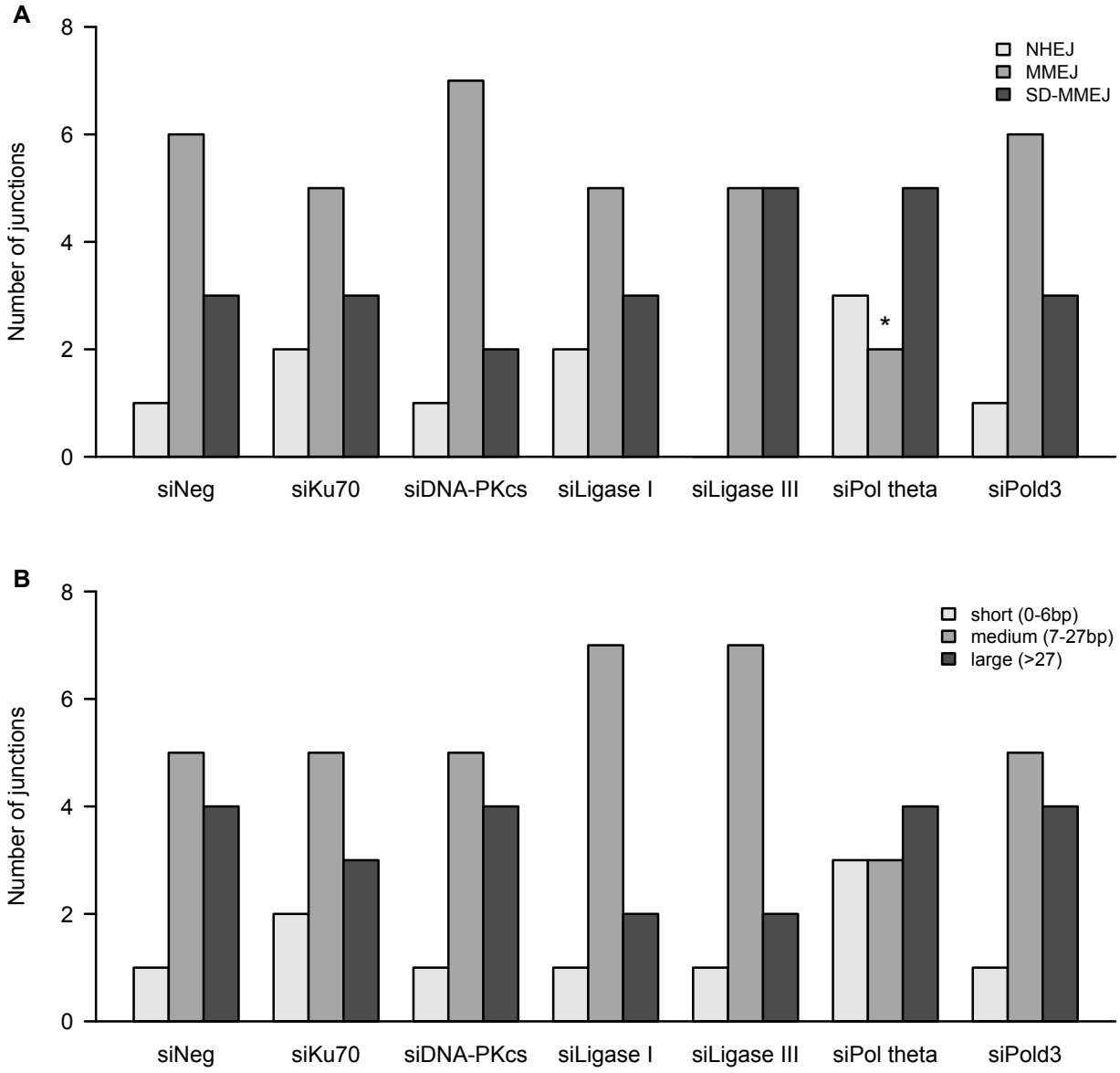


Figure 5.4: The frequency of DSB repair mechanisms and deletion sizes in cells transfected with MMEJ and SD-MMEJ reporters. A) The frequency of DSB repair with NHEJ, MMEJ and SD-MMEJ mechanisms in cells depleted in NHEJ or MMEJ genes. Statistical significance was calculated using the exact binomial test. Asterisks indicate significant differences between the siRNA treated sample and control treated with a non-targeting siRNA (siNeg), significance level $p<0.05$. B) The sizes of junctional deletions in siRNA-transfected cells. Deletions in the sequenced MMEJ reporter junctions were classified as short (0-6bp), medium (7-27bp) or large (>27bp).

Here, I constructed two GFP reporter substrates to study a recently proposed subpathway of MMEJ, termed SD-MMEJ. I also compared my two constructs with a previously described MMEJ reporter cassette. These experiments demonstrated that the MMEJ reporter, transiently transfected in CHO cells, yielded higher numbers of GFP-positive cells than the two SD-MMEJ reporters, likely due to the presence of longer microhomologies. However, sequencing of the repaired junctions revealed that while evidently less accurate and less likely to yield a functional GFP coding sequence, the SD-MMEJ mechanism is more frequently used by the cells. Therefore, I hypothesize that the availability of homology at the break site determines the pathway used for repair. If the DSB contains blunt or compatible cohesive ends, an apparently rare situation, they can be easily ligated by the NHEJ machinery (Fig.5.5). However, if the breaks are not compatible, extended 5'-3' end resection creates long ssDNA overhangs, which can anneal at pre-existing microhomologies, enabling repair by MMEJ. In the absence of homology at the break site, the SD-MMEJ mechanism can amplify the microhomology needed for alignment from another part of the repaired molecule. In conclusion, the SD-MMEJ pathway seems to be a very robust mechanism able to repair incompatible DSBs when other end-joining pathways fail.

I also analyzed the frequency of different DNA end-joining pathways in CHO cells depleted in important NHEJ and MMEJ factors. Out of the siRNAs tested, only the knock-down of polymerase θ significantly affected DNA end-joining in this transient assay. Interestingly, depletion of this enzyme resulted in decrease of MMEJ repair in favor of NHEJ and SD-MMEJ, indicating that polymerase θ may be especially important for this former pathway. These results are in line with the proposed role of this polymerase in MMEJ [45, 160, 168, 218]. While the increase in SD-MMEJ is inconsistent with previous reports assigning polymerase θ to this pathway [47, 372], it may be explained by the participation of other DNA polymerases in this process, e.g. translesion synthesis (TLS)

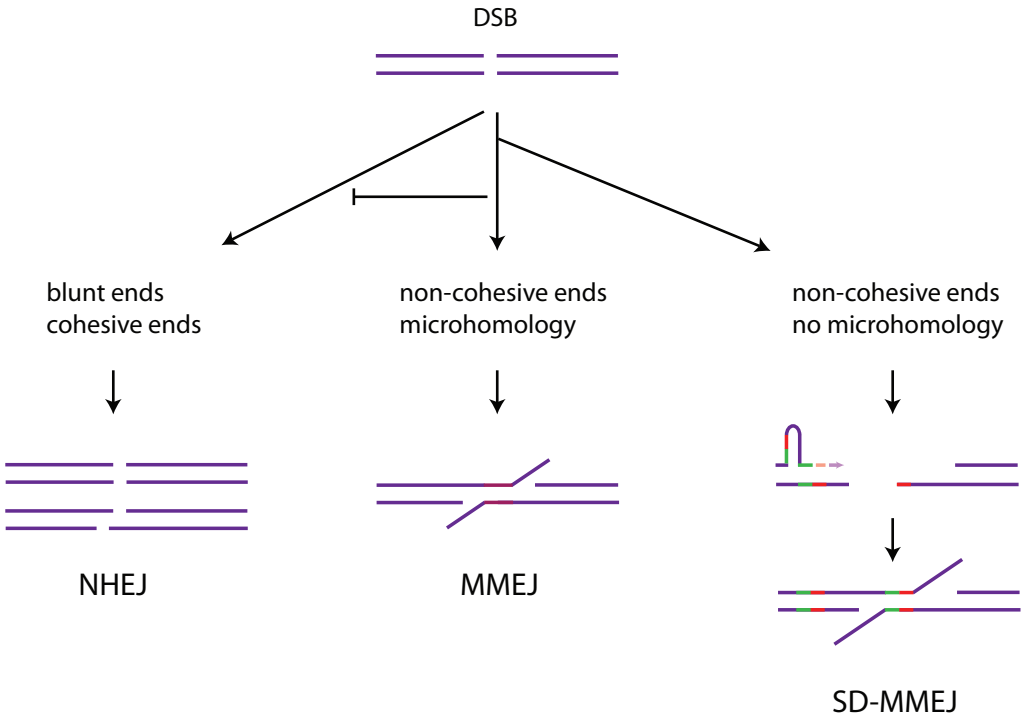


Figure 5.5: The model of DSB end-joining repair. The proposed model of DNA end-joining pathways in eukaryotic cells. The structure of the break and the availability of homology determine the repair mechanism.

or replication enzymes. Indeed, yeast homologs of TLS polymerases η and ζ (Rad30 and Rev3) as well as Pold3 (Pol32) were also reported to play a role in alternative DSB repair pathways [184] .

In the present work, I describe the construction of reporter assays designed to specifically detect repair events mediated by a sup-pathway of MMEJ – SD-MMEJ. The construction of such assay proved to be more challenging than initially anticipated. However, I was able to demonstrate that a combination of GFP fluorescence analysis and sequencing can be used to successfully measure the contribution of the different end-joining pathways to the repair of extrachromosomal DSBs. Here, due to the lack of time, I did not attempt to use the assays in the integrated form, nor to analyze more junctions, which may have enabled to obtain more statistically significant results. While this strategy would enable to study the repair of chromosomal DSBs, the transient assays are significantly less time consuming.

Moreover, they are a much better model to study the recombination of free DNA in a cellular environment, such as plasmids delivered to the cells during transfection. Finally, the efficiency of the intrachromosomal assays is often negatively influenced by the low efficiency of in vivo digestion at a single locus by the restriction enzyme. Nevertheless, the use of only transient assays may not be sufficient to draw universal conclusions concerning DSB repair in eukaryotic cells, as it was demonstrated that extrachromosomal DNA breaks are treated differently than genomic breaks [288, 354]. Therefore, to reliably assess the contribution of the different pathways to DSB repair, both methods should be used in parallel.

5.6 Supplementary materials

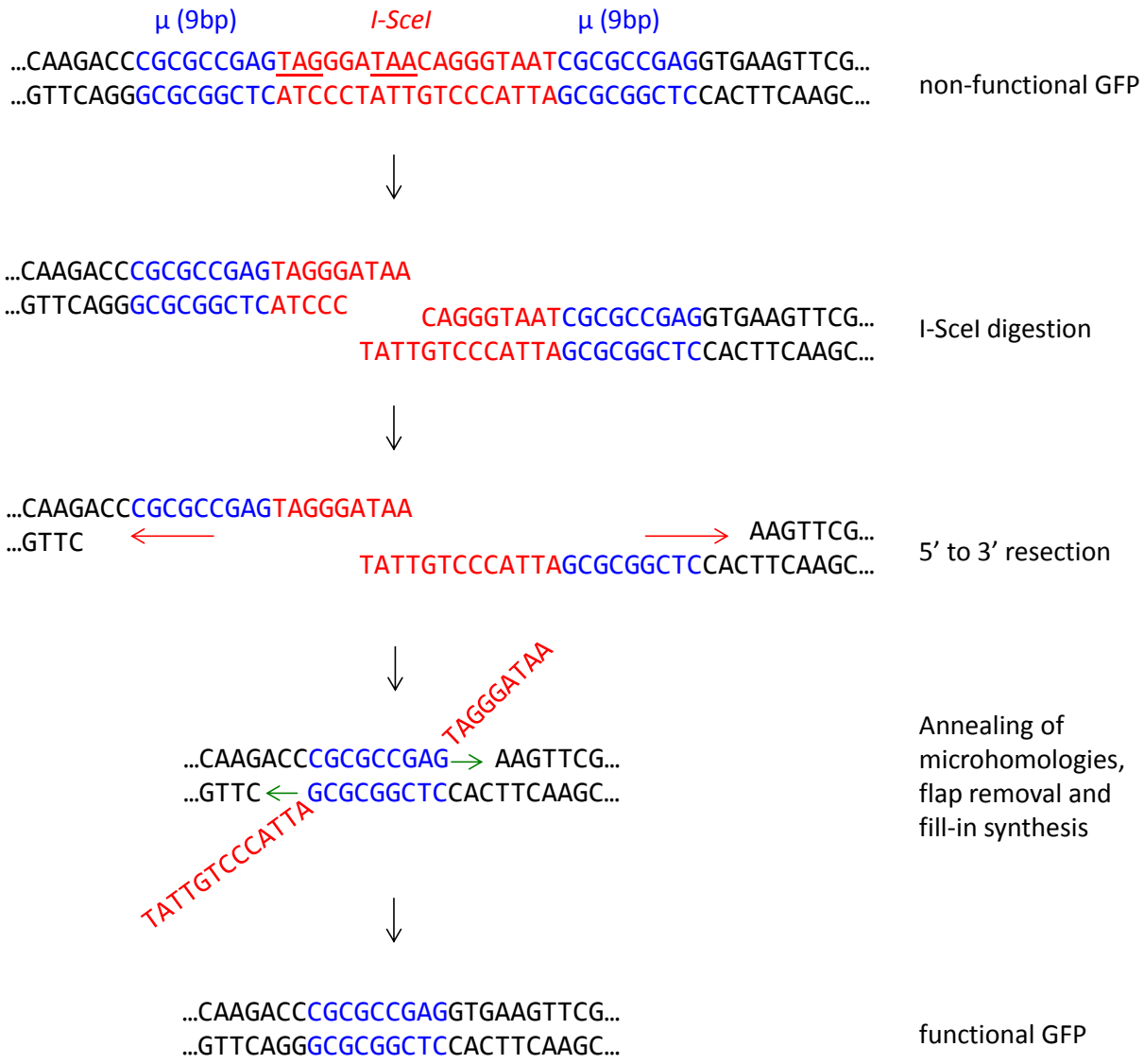


Figure 5S1: The mechanism of action of the simple MMEJ reporter assay. Stop codons are underlined. Red arrows indicate the direction of the 5’-3’ resection. Green arrows indiate the direction of DNA synthesis. Adapted from [347].

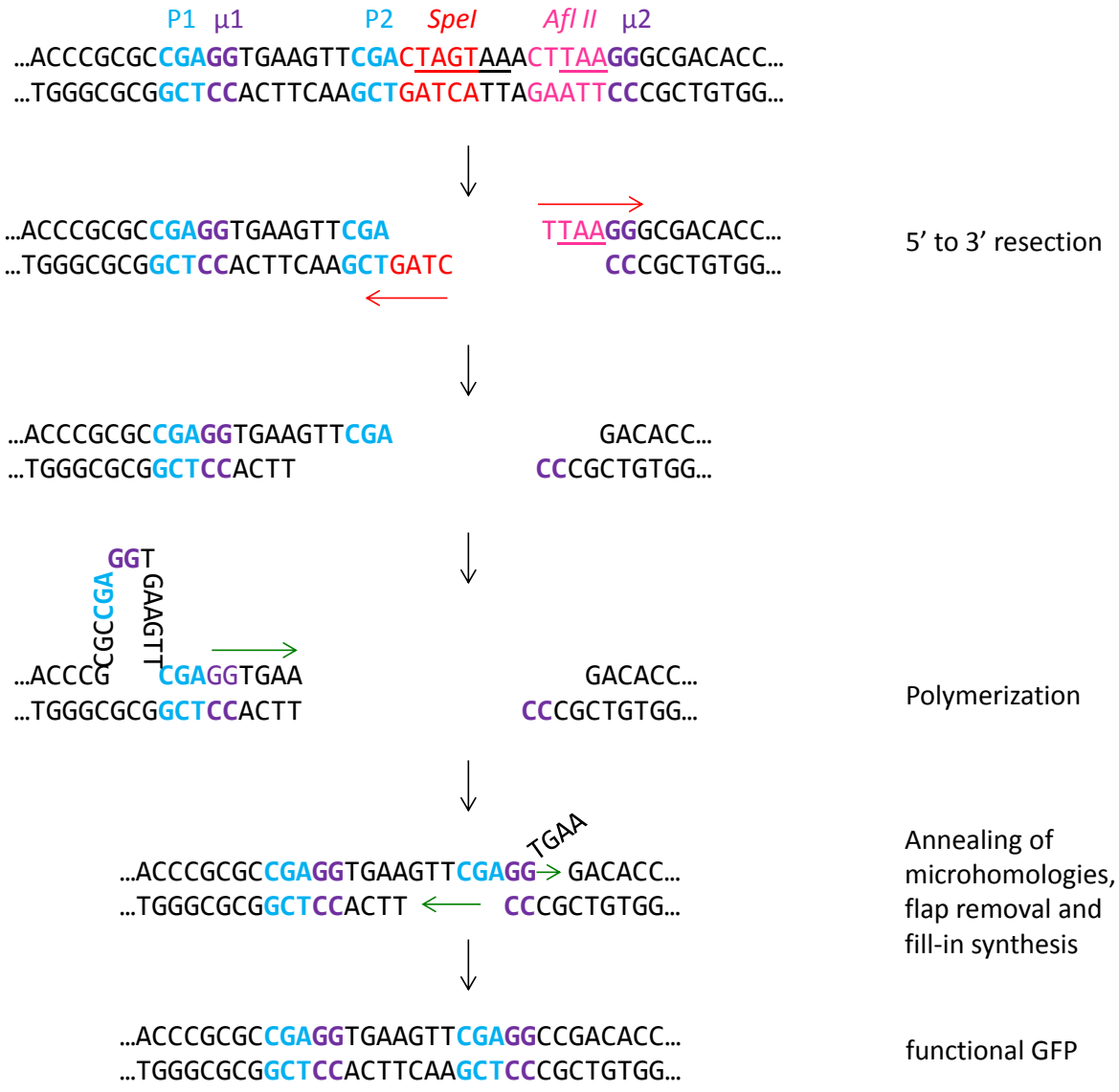


Figure 5S2: Proposed mechanism of repair for the SD-MMEJ-1 reporter. Stop codons are underlined. Red arrows indicate the direction of the 5’-3’ resection. Green arrows indiate the direction of DNA synthesis. P1/P2– primer repeats, $\mu 1/\mu 2$ – microhomology repeats. Adapted from [372].

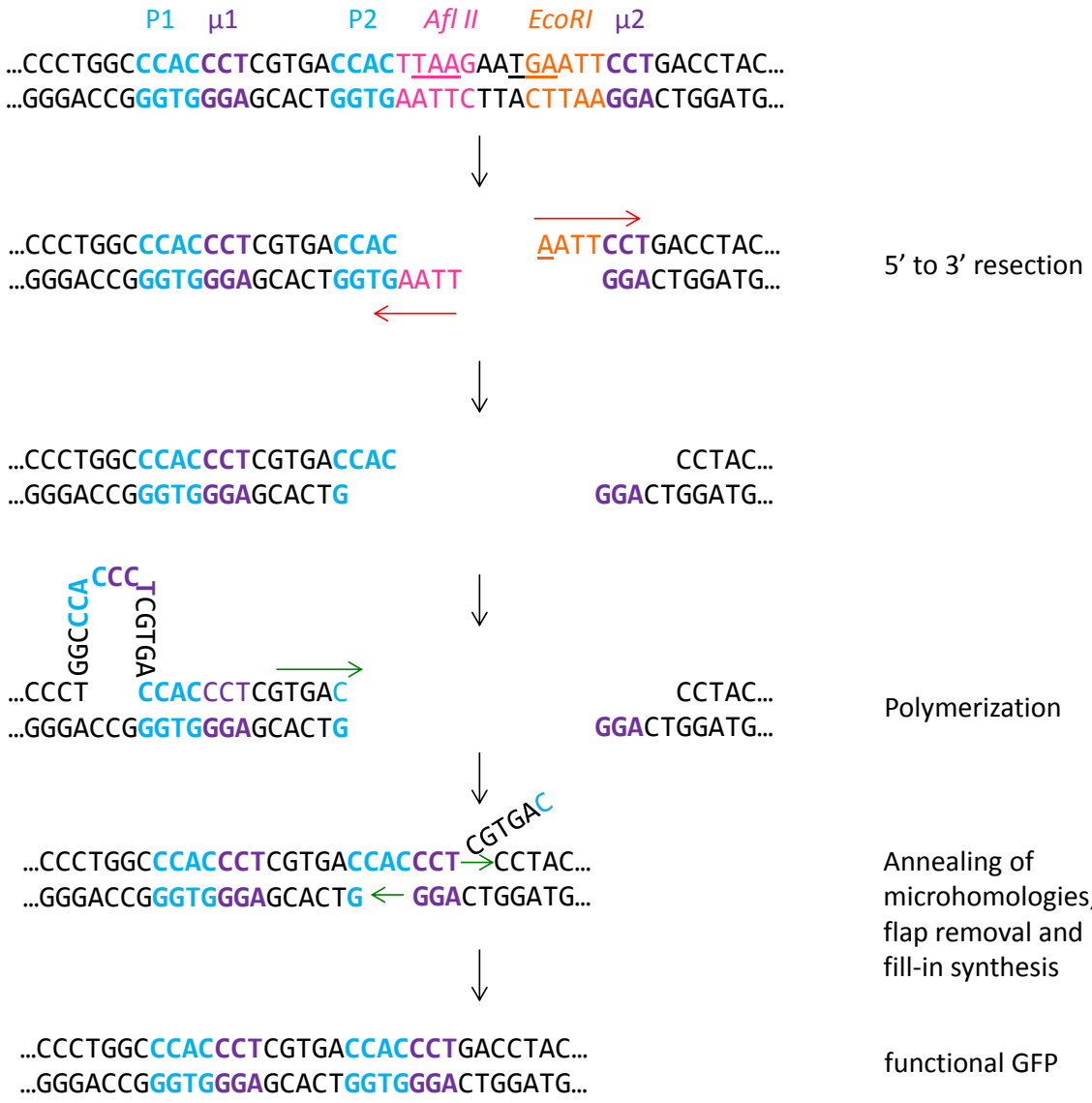


Figure 5S3: Proposed mechanism of repair for the SD-MMEJ-2 reporter. Stop codons are underlined. Red arrows indicate the direction of the 5’-3’ resection. Green arrows indiate the direction of DNA synthesis. P1/P2– primer repeats, $\mu 1/\mu 2$ – microhomology repeats. Adapted from [372].

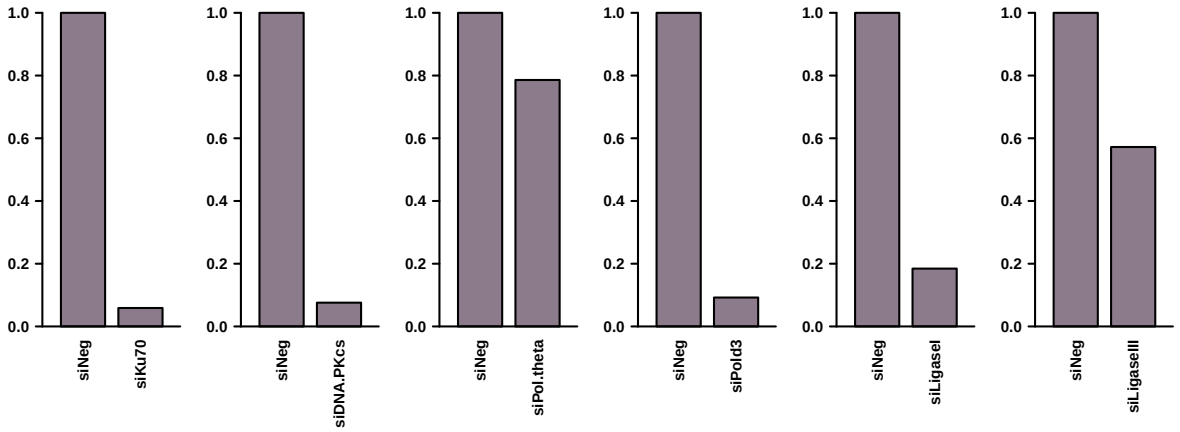


Figure 5S4: The effect of siRNA knock-down on target mRNA level. Total mRNA was extracted from CHO cells transfected with three negative control siRNAs (siNeg) or with specific siRNAs targeting the indicated gene. The mRNA level of the target was quantified by qPCR. Values were normalized to the target mRNA levels in cells transfected with the control siRNAs.

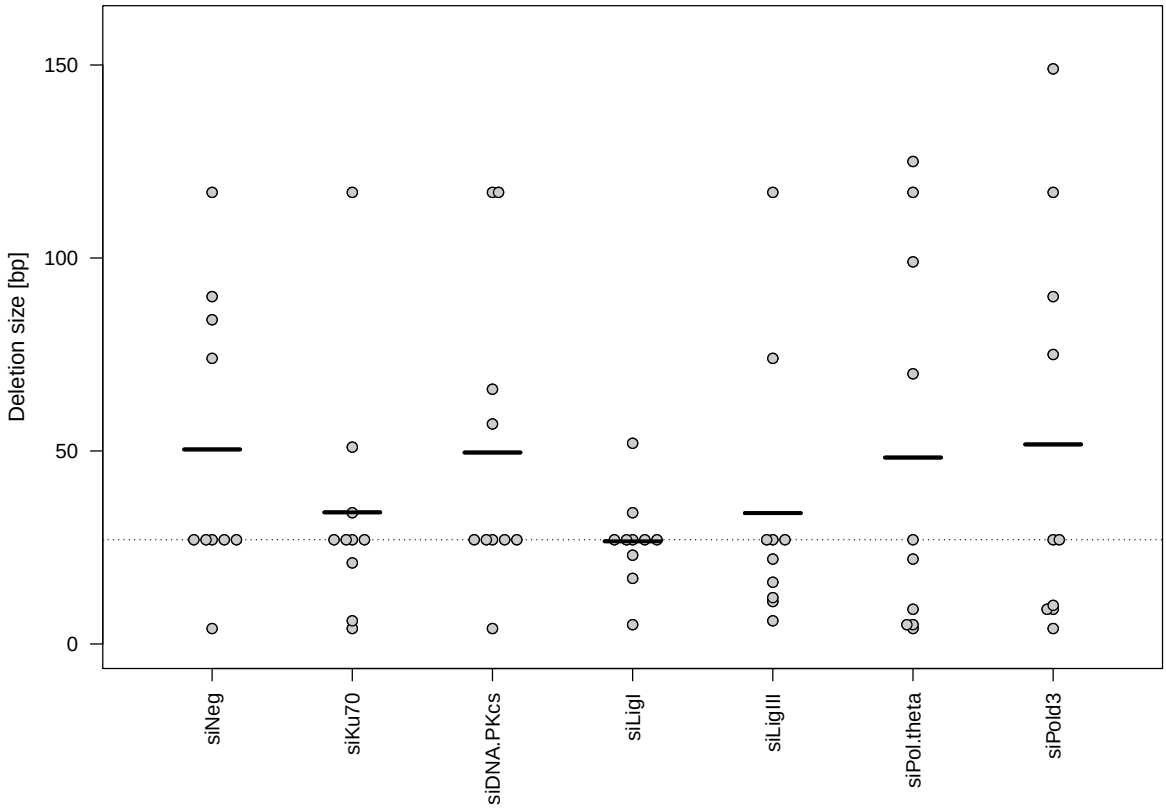


Figure 5S5: The distribution of deletion sizes in cells depleted in MMEJ and NHEJ genes. Black lines represent mean deletion sizes. The dotted line represents the expected deletion size from a MMEJ mechanism relying on the preset 9 bp microhomology.

Table 5S1: Sequenced repair products from cells transfected with the MMEJ and SD-MMEJ assays. A) MMEJ assay, B) SD-MMEJ-1 assay, C) SD-MMEJ-2 assay.

A

microhomology I-SceI microhomology

5'…CAAGACCCGCGCCGAGTAGGGATAA CAGGGTAATCGCGCCGAGGTGAAGTTCG…3'

3'…GTTCAGGGCGCGGCTCATCCC TATTGTCCCATTAGCGCGGCTCCACTTCAAGC…5'

MMEJ #1 junction

original vector	**TCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAG**GAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAG TAGGGATAACAGGGTAATCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACC**GCATCGAGCTGAAGGGCATCGACTTCA**
junction	**TCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAG**---------- ----------**GCATCGAGCTGAAGGGCATCGACTTCA**

MMEJ #2 junction

original vector	CGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACT **TCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAG**GAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAG TAGGGATAACAGGGTAATCGCG**CCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCA**
junction	**TCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAG**---------- ----------**CCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCA**

MMEJ #3 junction

original vector	TCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTAC CCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGA **CGGCAACTACAAGACCCGCGCCG**AGTAGGGATAACAGGGTAATCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGG**TGAACCGCA**
junction	**CGGCAACTACAAGACCCGCGCCG**----------**TGAACCGCA**

MMEJ #4 junction

original vector	CGGCAACTACAAGACCCGCGCCGAGTAGGGATAACAGGGTAATCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCA
junction	CGGCAACTACAAGACCCGCGCCGAG----------GTGAAGTTCGAGGGCGACACCCTGGTGAACCGCA

MMEJ #5 junction

original vector	**CAACTACAAGACCCGCG**CCGAGTAGGGATAACAGGGTA**ATCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCG** AGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATTCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATC ATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGGGGACGGCAGCGTGCAGCTCGCCGACCA CTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTACCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAG ACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAG
junction	**CAACTACAAGACCCGCG**----------**ATCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCG**

MMEJ #6 junction

original vector	CAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGC AGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACC **ATCTTCTTCAAGGACG**ACGGCAACTACAAGACCCGCGCCGAGTAGGGATAACAGGGTAATCGCGCCGAGG**TGAAGTTCGAGGGCGA**
junction	**ATCTTCTTCAAGGACG**----------**TGAAGTTCGAGGGCGA**

MMEJ #7 junction

original vector	CGGCAACTACAAGACCCGCGCCGAGTAGGGATAACAGGGTAATCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCA
junction	CGGCAACTACAAGACCCGCGCCGAGT----------CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCA

Table 5S1: Sequenced repair products from cells transfected with the MMEJ and SD-MMEJ assays (continued).

MMEJ #8 junction

original vector	**TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAA**GGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTAC AAGACCCGCGCCGAGTAGGGATAACAGGGTAAT**CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAA** GGGCATCGACTTCAAGGAGGACGGCAACATTCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAG
junction	**TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAA**---------- ----------**CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAA**

MMEJ #9 junction

original vector	CCCGCGCCGAGTAGGGATAACAGGGTAATCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGC
junction	CCCGC----------CCTGGTGAACCGCATCGAGCTGAAGGGC

MMEJ #10 junction

original vector	**GACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCG** **AAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA**CGACGGCAACTACAAGACCCGCGCCGAGTAGGGATAACAGGGTAATCGC GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGA**CTTCAAGGAGGACGGCAACATCCTGGG**
junction	**AAGGCTACGTCCAGGAGCGCACCATCTT**---------- ----------**CTTCAAGGAGGACGGCAACATCCTGGG**

MMEJ #11 junction

original vector	**CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCA** ACTACAAGACCCGCGCCGAGTAGGGATAACAGGGTAATCGC**GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAG** **CTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCAT**
junction	**CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCAT**---------- ----------**GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAG**

MMEJ #12 junction

original vector	**CCAGGAGCG**CACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGTAGGGATAACAGGGTAAT**CGCGCCGAGGTGA** AGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTG GAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAA CATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACC ACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCC
junction	**CCAGGAGCG**----------**CGCGCCGAGGTGA**

B

SpeI AflII cut

…ACCCGCGCCGAGGTGAAGTTCGA TTAAGGGCGACACC…

…TGGGCGCGGCTCCACTTCAAGCTGATC CCCGCTGTGG…

micro6 #1

original vector	TACAAGACCCGCGCCGAGGTGAAGTTCGACTAGTAAACTTAAGGGCGACACCCTGGTGAACCGCATCGAGCTGAA
junction	TACAAGACCCGCGCCGAGGTGAAGTTCGAC----------AAGGGCGACACCCTGGTGAACCGCATCGAGCTGAA

micro6 #2

original vector	TACAAGACCCGCGCCGAGGTGAAGTTCGACTAGTAAACTTAAGGGCGACACCCTGGTGAACCGCATCGAGCTGAA
junction	TACAAGACCCGCG----------GGCGACACCCTGGTGAACCGCATCGAGCTGAA

Table 5S1: Sequenced repair products from cells transfected with the MMEJ and SD-MMEJ assays (continued).

micro6 #3

original vector	**CCGAAGGCTACGTCCAGGAGCGCACCA**TCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAC TAGTAAACTTAAGGGCGA**CACCCTGGTGAACCGCATCGAGCTGAA…**
junction	**CCGAAGGCTACGTCCAGGAGCGCACCA**---------- ----------**CACCCTGGTGAACCGCATCGAGCTGAA…**

micro6 #4

original vector	**CCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAAC**TACAAGACCCGCGCCGAGGTGAAGTTCGAC TAGTAAACTTAAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAAC**ATCCTGG GGCACAAGCTGGAG…**
junction	**CCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAAC**---------- **ATCCTGGGGCACAAGCTGGAG…**

micro6 #5

original vector	TACAAGACCCGCGCCGAGGTGAAGTTCGACTAGTAAACTTAAG GGCGACACCCTGGTGAACCGCATCGAGCTGAA
junction	TACAAGACCCGCGCCGAGGTGAAGTTCGAC----------ACCCTGGTGAACCGCATCGAGCTGAA

C

AflII EcoRI

…CCCTGGCCCACCCTCGTGACCAC AATTCCTGACCTAC…

…GGGACCGGGTGGGAGCACTGGTGAATT GGACTGGATG…

micro16 #1

original vector	GTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACTTAAGAATGAATTCCTGACCTACGGCGTGCA GTGCTTCAGCCGCTACCCCGACCACATG
junction	GTTCATCTGCACTAC----------CCTGACCTACGGCGTGCA GTGCTTCAGCCGCTACCCCGACCACATG…

micro16 #2

original vector	CACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACTTAAGAATGAATTCCTGACCTACGGCGTGCAGTGCTTC
junction	CACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGT----------CCTACGGCGTGCAGTGCTTC

micro16 #3

original vector	**CACCACCGGCAAGCTG**CCCGTGCCCTGGCCCACCCTCGTGACCACTTAAGAATGAATTCCTGACCTACGGCGTGCAGTGCTTC AGCCGCTAC**CCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTT**
junction	**CACCACCGGCAAGCTG**---------- ------**CCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTT**

micro16 #4

original vector	AAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACTTAAGAATGAATTCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCC
junction	AAGCTGCCCGTGCCCTGGCCCACCCTCG----------CGCTACCCC

Note: The microhomologies of the SD-MMEJ model are underlined with a double line, the microhomologies of the MMEJ model are underlined with a single line. The "-" signs indicate deleted bases.

Table 5S2: Analysis of repair products from siRNA-treated cells transfected with the MMEJ reporter.

siRNA	junction #	Mechanism	Length of microhomology	Deletion/insertion (bp)
Neg	1	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	2	SD-MMEJ	4bp: CG/GG (IR, 3bp upstream of junction) 7bp: CGCC/GGG (DR, 297bp downstream of junction)	**-74bp**
	3	SD-MMEJ	5bp: CTT/GG (DR, 326 bp upstream of junction)	**-84bp**
	4	SD-MMEJ	5bp: CGC/GG (IR, 388bp downstream of junction)	**-90bp**
	5	MMEJ	9bp: GACGGCAAC	**-117bp**
	6	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	7	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	8	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	9	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	10	NHEJ SD-MMEJ	blunt join 6bp: GGG/CAG (IR, 173bp upstream from junction)	**-4bp**
Ku70	1	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	2	MMEJ	9bp: CTTCAAGGA	**-117bp**
	3	SD-MMEJ SD-MMEJ	1) 5bp: CC/GCT (DR, 69bp upstream from junction) 2) 4bp: CG/CA (DR, 46bp downstream from junction)	**-51bp** 38bp templated insert from GFP (215bp downstream)
	4	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	5	SD-MMEJ	5bp: CG/GCG (DR, 127bp upstream from junction)	**-21bp**
	6	NHEJ SD-MMEJ	blunt join 6bp: GGG/CAG (IR, 173bp upstream from junction)	**-4bp**
	7	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	8	NHEJ SD-MMEJ	blunt join GGG/GG (IR, 236 bp downstream of junction)	**-6bp**
	9	SD-MMEJ	5bp: GG/GGC (DR, 62bp downstream from junction)	**-34bp**
	10	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**

Table 5S2: Analysis of repair products from siRNA-treated cells transfected with the MMEJ reporter (2).

siRNA	junction #	Mechanism	Length of microhomology	Deletion/insertion (bp)
LigI	1	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	2	NHEJ	blunt join	**-23bp**
	3	SD-MMEJ	5bp: GCA/GT (DR, 102bp upstream of junction)	**-52bp**
	4	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	5	SD-MMEJ	5bp: AG/CGC (DR, 46bp upstream from junction)	**-17bp**
	6	SD-MMEJ	7bp: GC/TGAAG (DR, 32bp downstream from the junction)	**-34bp**
	7	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	8	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	9	NHEJ	1bp join	**-5bp**
	10	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
LigIII	1	MMEJ SD-MMEJ	2bp: AC 7bp: GAC/CCTG (DR, 171bp upstream from junction)	**-74bp**
	2	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	3	SD-MMEJ	5bp: AG/AAT (IR, 81bp downstream from junction)	**-12bp**
	4	SD-MMEJ	7bp: CCGA/AGG (DR, 56p upstream from junction)	**-11bp**
	5	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	6	SD-MMEJ	GTAG/CAG (IR, 261 bp downstream from junction) AG/CAG (DR, 94 bp upstream from the junction)	**-6bp**
	7	MMEJ	9bp: CTTCAAGGA	**-117bp**
	8	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	9	SD-MMEJ	5bp: AG/CCG (DR, 113bp upstream from junction)	**-22bp**
	10	SD-MMEJ	5bp: GC/GGT (IR, 40 bp downstream from junction)	**-16bp**

Table 5S2: Analysis of repair products from siRNA-treated cells transfected with the MMEJ reporter (3).

siRNA	junction #	Mechanism	Length of microhomology	Deletion/insertion (bp)
Pol theta	1	SD-MMEJ	5bp: CA/CCG (DR, 166bp upstream of junction)	**-70bp**
	2	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	3	NHEJ	blunt join	**-4bp**
		SD-MMEJ	6bp: GGG/CAG (IR, 173bp upstream from junction)	
	4	SD-MMEJ	1) 6bp: CA/CGCC (IR, 56bp upstream from the junction)	**-99bp**
		SD-MMEJ	2) 4bp: CG/TG (on the insert, DR, 2bp upstream from the junction)	31bp templated insert from GFP (168bp downstream)
	5	MMEJ	4bp: AGGG	**-9bp**
	6	NHEJ	1bp join	**-5bp**
	7	SD-MMEJ	4bp: CG/TA (IR, 49 bp upstream from junction)	**-125bp**
	8	NHEJ	1bp join	**-5bp**
	9	SD-MMEJ	6bp: CCC/TAA (DR, 497bp upstream from junction)	**-22bp**
	10	SD-MMEJ	6bp: AGC/AGG (IR, 277bp downstream from junction)	**-117bp**
Pold3	1	MMEJ	2bp: GG	**-10bp**
	2	MMEJ	4bp: AGGG	**-9bp**
	3	SD-MMEJ	7bp: CC/CGACA (DR, 233bp downstream from the junction)	**-149bp**
	4	MMEJ	3bp: AGG	**-90bp**
		SD-MMEJ	5bp: TCC/AG (IR, 75bp downstream from junction)	
	5	SD-MMEJ	5bp: AG/CAG (DR, 91bp upstream from the junction)	**-9bp**
	6	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	7	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	8	NHEJ	blunt join	**-4bp**
		SD-MMEJ	6bp: GGG/CAG (IR, 173bp upstream from junction)	
	9	SD-MMEJ	5bp: CG/TGA (DR, 124bp upstream of junction)	**-75bp**
	10	MMEJ	9bp: GACGGCAAC	**-117bp**

Table 5S2: Analysis of repair products from siRNA-treated cells transfected with the MMEJ reporter (4).

siRNA	junction #	Mechanism	Length of microhomology	Deletion/insertion (bp)
DNA-PKcs	1	MMEJ	9bp: GACGGCAAC	**-117bp**
	2	SD-MMEJ	5bp: AGA/AT (IR, 81bp downstream from junction)	**-27bp**
	3	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	4	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	5	MMEJ	3bp: CGC	**-66bp**
	6	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	7	SD-MMEJ	5bp: GCG/TG (DR, 121bp upstream from junction)	**-57bp**
	8	MMEJ	9bp (Wang): CGCGCCGAG	**-27bp**
	9	NHEJ	blunt join	**-4bp**
		SD-MMEJ	6bp: GGG/CAG (IR, 173bp upstream from junction)	
	10	MMEJ	4bp: CATC	**-117bp**

Note: '9 bp (Wang)' indicates repair with the MMEJ mechanism relying on the preset 9 bp microhomology.

Chapter 6

A role for homologous recombination proteins in cell cycle regulation

This chapter is based on a manuscript submitted for publication entitled "A role for homologous recombination proteins in cell cycle regulation" by **Kostyrko, K.**, Bosshard, S., Urban, Z., Mermod, N.

This chapter contains data obtained together with Zuzanna Urban, during her Summer Undergraduate Program internship project, and Sandra Bosshard, during her Master thesis project.

Zuzanna Urban participated in the characterization of the CHO Fucci cells (Fig.6S1, 6S2, 6S5). Sandra Bosshard performed siRNA knock-down experiments (Fig.6.2, 6.3, 6.4, 6S3A, 6S4, 6S6, 6S7). I generated and characterized the CHO Fucci cells, performed part of the siRNA knock-down experiments and supervised all the work.

6.1 Abstract

Eukaryotic cells respond to DNA breaks, especially double-stranded (DSBs), by activating the DNA damage response (DDR), which encompasses DNA repair and cell cycle checkpoint signaling. The DNA damage signal is transmitted to the checkpoint machinery by a network of specialized DNA damage-recognizing and signal-transducing molecules. However, recent evidence suggests that DNA repair proteins themselves may also directly contribute to the checkpoint control. Here, we investigated the role of homologous recombination (HR) proteins in normal cell cycle regulation in the absence of exogenous DNA damage. For this purpose, we used Chinese Hamster Ovary (CHO) cells expressing the Fluorescent ubiquitination-based cell cycle indicator (Fucci). Systematic siRNA-mediated knockdown of HR genes in these cells demonstrated that the absence of several of these factors alters cell cycle distribution, albeit differentially. The knock-down of MDC1, Rad51 and Brca1 caused the cells to arrest in the G2 phase, suggesting that they may be required for the G2/M transition. In contrast, inhibition of the other HR factors, including several Rad51 paralogs and Rad50, led to the arrest in the G1/G0 phase. Moreover, the absence of Rad51B, Rad51C, CtIP and Rad50 seemed to induce entry into the quiescent G0 phase. In conclusion, the lack of many HR factors may lead to cell cycle checkpoint deficiency, even in the absence of exogenous DNA damage, indicating that these proteins may play an essential role both in DNA repair and checkpoint signaling.

6.2 Introduction

DNA double strand breaks (DSBs), one of the most deleterious types of DNA lesions, can result from ionizing radiation or chemical agents, or from natural cellular processes such as DNA replication or maturation of the immune system genes. If left unrepaired, they constitute a major threat to genetic integrity and stability, leading to cell death or carcinogenesis [295]. In response to DSBs, cells activate a network of DNA repair and signaling pathways, collectively termed the DNA damage response (DDR) [129, 145, 200]. To allow time for DNA repair, the DDR machinery activates cell cycle checkpoints that arrest cell cycle progression until genome integrity is restored. The DDR-activated checkpoints include the G1/S, the intra-S and the G2/M transitions. The G1/S checkpoint, the most sensitive to DNA damage, is defective in most human cancers [178, 198].

The Mre11/Rad50/Nbs1 (MRN) complex is among the first sensors of DSBs, subsequently activating Ataxia telangiectasia mutated (ATM) [183]. ATM, a key protein kinase in the DDR network, is responsible for phosphorylation of many downstream DNA repair and cell cycle factors, including tumor suppressor p53, mediator of DNA-damage checkpoint 1 (MDC1), cell cycle checkpoint kinase 2 (Chk2), and breast cancer susceptibility protein 1 (Brca1) [294, 296]. The activation of these factors results in signaling cascades ultimately leading to cell cycle arrest. ATM-dependent phosphorylation of histone H2AX (γH2AX) also induces global changes in the chromatin structure leading to the recruitment of DNA repair proteins to the sites of damage.

Several specialized pathways exist in higher eukaryotic cells to repair DNA breaks. One of the main pathways responsible for DSB repair is non-homologous end-joining (NHEJ). NHEJ is a fast process, based on a simple ligation of the two broken DNA ends, active throughout the entire cell cycle [253]. In the absence of functional NHEJ, cells were shown to use a highly error-prone, backup mechanism termed microhomology mediated

end joining (MMEJ) [137, 152]. The third pathway, considered to be the most precise of all the DSB repair mechanisms, was termed homologous recombination (HR) [284]. HR requires extensive homology for repair, and thus is primarily used in late S and G2 phases of the cell cycle, when genetic material is already replicated and sister chromatids are available as repair template.

A key role in eukaryotic HR is played by the Rad51 recombinase, which coats ssDNA ends resulting from the initial processing of the DSB [19, 293]. The DNA-bound Rad51 then searches for sequence homology along the complementary DNA strand and mediates pairing between the two strands. The Rad51 protein is essential for cell proliferation, as targeted knock-out of this gene leads to embryonic lethality in mice [330]. Other proteins involved in HR include CtIP, Brca2, Rad52, Rad54 and the five Rad51 paralogs: Rad51B, Rad51C, Rad51D, Xrcc2, and Xrcc3 [92, 233, 285, 315, 317]. Rad51B, Rad51C, Rad51D and Xrcc2 together form the BCDX2 complex, which was proposed to facilitate the formation and stabilization of the Rad51 nucleofilament [217]. Rad51C also participates in the formation of a second complex with Xrcc3, termed CX3 [197]. The CX3 dimer was reported to play an essential role in the final resolution of recombination intermediates. The MRN complex, MDC1, and Brca1, which are components of the DDR response, also play a role in HR [146, 174, 379].

It has been recently proposed that HR proteins, in addition to their role in DNA repair, could also directly contribute to cell cycle control [14, 276, 367]. The knock-down of Rad51 was shown to induce G2/M arrest, suggesting that this protein is required for the progression from the G2 phase to mitosis [180, 86, 297], and Brca1 was reported to play a role in the regulation of the G2/M and intra-S checkpoints [355, 367]. Rad51C was also proposed to contribute to cell cycle regulation, although there are conflicting reports as to its exact role. Rodrigue and others observed that the knock-down of Rad51C in

human cells leads to the arrest at the G2/M checkpoint [276]. This indicated that Rad51C, similarly to Rad51, is needed for the progression through the G2 phase. In another study, Rad51C knock-down caused the cells to escape the intra-S and G2/M checkpoints, thus allowing them to enter mitosis [14]. This work suggested that Rad51C may play a role in the activation of G2/M checkpoint in response to DNA damage. Several other DSB repair proteins have also been proposed to participate in cell cycle progression [176, 363].

Here, to systematically asses the role of HR factors in cell cycle regulation in the absence of exogenous DNA damage, we silenced several important HR genes in Chinese Hamster Ovary (CHO) cells expressing the fluorescent ubiquitination-based cell cycle indicator (Fucci) [280]. We showed that the knock-down of many HR factors, including Rad51, MDC1, Brca1, several Rad51 paralogs, CtIP, and Rad50 significantly affected cell cycle progression, albeit differentially. The knock-down of MDC1, Rad51 and Brca1 caused the cells to arrest at the G2/M checkpoint, suggesting that these factors may be required for the transition through the G2 phase and entry into mitosis. In contrast, the absence of the remaining HR proteins increased the proportion of G1/G0 phase cells, indicating that their deficiency may cause the cells to escape the G2/M checkpoint, divide and subsequently become arrested in the G1 phase. We also observed that knock-down of Rad51B, Rad51C, CtIP and Rad50 increased the proportion of G0 cells, suggesting that in the absence of these factors cells may enter a quiescent state. In conclusion, it seems that many HR proteins are not only important for DSB repair, but possibly also play a role in the regulation of cell cycle progression.

6.3 Materials and Methods

CHO cells expressing Fucci probes

Adherent CHO DG44 cells [333] were cultivated in DMEM/F12+GlutaMAX^{TM} sup-

plemented with 1x HT and 10% fetal bovine serum (FBS) (Gibco, Invitrogen), and with antibiotic-antimycotic solution (Sigma-Aldrich, #A5955). CHO Fucci cells were constructed using lentiviral vectors carrying the red and green fluorescent ubiquitination-based cell cycle indicator (Fucci) cassettes [280]. The red Fucci cassette contains a monomeric version of Kusabira Orange 2 (mKO2) reporter gene fused to a truncated human Cdt1 (hCdt1, amino acids 30–120). The mKO2-hCdt1(30/120) protein is expressed in G1 phase and degraded at the onset of the S phase. The green Fucci cassette contains the monomeric version of Azami green (mAG) reporter gene fused to the 110 amino acid N-terminus of human Geminin (hGem amino acids 1-110). The mKO2-hGem(1/110) protein accumulates through S, G2 and M phases of the cell cycle and is degraded in the metaphase/anaphase transition of mitosis. The lentiviral constructs were kindly provided by M. Lutolf (EPFL, Lausanne, Switzerland). Briefly, the cells were transduced with a 1:1 ratio of mKO2 and mAG vectors at MOI 50. Three weeks after transduction double positive (mKO2+mAG+) clones were single-cell sorted by fluorescence-activated cell sorting (FACS) (FACSAria II sorter, Becton-Dickinson). A single clone (#21) with similar levels of mKO2 and mAG fluorescence intensity was selected for subsequent experiments.

Cell synchronization

For starvation synchronization CHO cells were grown for 72h in medium supplemented with 0.2% FBS (Gibco, Invitrogen). For synchronization through contact inhibition cells were grown for 3-5 days until complete confluency. Both methods synchronize the cells in G1/G0 phase. To reinitiate cell cycle progression cells were replated at lower density in complete medium. For early S phase synchronization 10 000 double positive (mAG+mKO2+) cells were sorted by FACS (FACSAria II sorter, Becton-Dickinson) into each well of a 12-well plate.

siRNA transfection

Small interfering RNA duplexes were specifically designed to target the Chinese hamster homologs of HR genes. The siRNAs were designed and provided by Microsynth AG (Balgach, Switzerland). Three RNA duplexes were designed per gene to increase the probability of successful knock-down. Three negative (non-targeting) siRNAs were also designed as controls. For siRNA-mediated knock-down, CHO-Fucci cells were transfected with equimolar amounts of three siRNA duplexes at a final concentration of 50nM using Lipofectamine RNAiMAX, according to the manufacturer's instructions (Invitrogen). After 72h cells were analyzed by microscopy (Carl Zeiss Microscope Axio Observer.A1).

Flow cytometry

For flow cytometry cells were harvested 72h following siRNA transfection, resuspended in 0.5 ml of PBS with 2% FBS (Gibco, Invitrogen), and analyzed using the CyAn analyzer (Beckman Coulter). Acquired data was analyzed using the FlowJo software (Tree Star). For the time point experiment cells were harvested every 4h at 40, 44, 48, 64, 68 and 72h post transfection, fixed in PBS with 4% PFA (1:2 v/v) and analyzed by flow cytometry.

6.4 Results

6.4.1 Characterization of CHO Fucci cells

The fluorescent ubiquitination-based cell cycle indicator (Fucci) system enables the simultaneous observation of multiple cell cycle phases in living cells [280]. The system is based on the expression and ubiquitination of cell cycle-dependent human proteins Cdt1 and Geminin, fused to fluorescent markers monomeric Kusabira Orange 2 (mKO2)

and monomeric Azami Green (mAG), respectively. Depending on the expression and accumulation of the fluorophores four main cell subpopulations can be visualized – the early G1 phase cells, G1/G0 cells, early S cells and late S, G2 and M phase cells (Fig.6.1). The freshly divided early G1 phase cells are non-fluorescent, but as the cells progress through the G1 phase they start expressing and accumulate the mKO2-hCdt1(30/120) chimeric protein. In the early S phase, while the mKO2-hCdt1(30/120) becomes ubiquitinated and degraded, the cells start producing the mAG-hGem(1/110) fusion protein, which makes them appear yellow under the microscope. Expression of mAG-hGem(1/110) persists through the S, G2 and M phases, causing the cells to emit only green fluorescence until the end of the cell cycle.

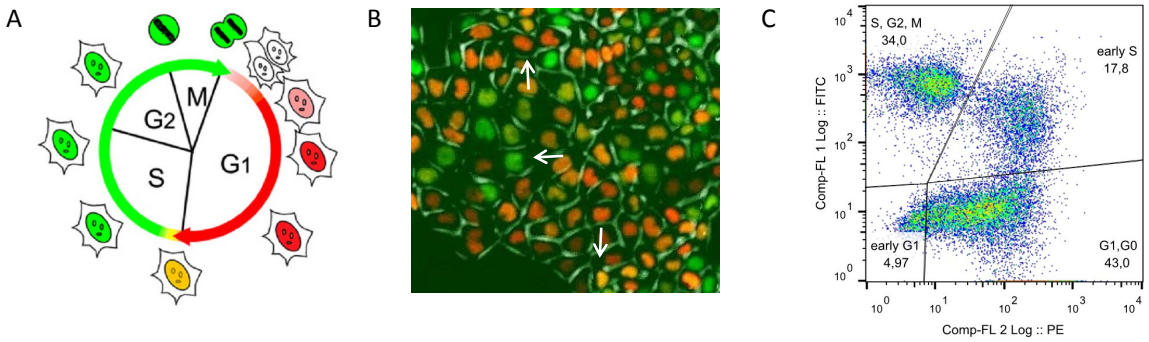


Figure 6.1: Characterization of CHO Fucci cells. A) Scheme of the Fucci cell cycle [280]. B) Typical uorescence image of CHO Fucci cells. Arrows point to cells in G1/G0 (top), late S, G2, and M (middle) and early S phase cells (bottom). C) Flow cytometry analysis of CHO Fucci cells with four subpopulations of cells: mKO2-positive G1/G0 cells (PE channel), mAG-positive late S, G2 and M cells (FITC channel), double positive (mKO2+mAG+) early S cells, and a double negative (mKO2- , mAG-) early G1 cells.

To confirm that the CHO Fucci cells express the fluorescent probes in a cell cycle dependent manner, we synchronized the culture by serum deprivation, which induces a cell cycle arrest in the G1/G0 phase, or by contact inhibition, which causes G1 arrest [119, 277]. We observed that over 90% of serum-deprived cells displayed red fluorescence indicative of the G1/G0 phase (Fig.6S1,6S2). In cells grown to confluence this number was lower (50%), indicating that this method is less efficient in synchronizing cells than

serum starvation (Fig.6S1B,6S2B). We subsequently released the cells from cycle arrest and monitored the cell cycle distribution 16-24h and 40-48h following release. Within 20h the starved cells progressed through the S phase into the G2 and M phases, as evidenced by the accumulation of green fluorescent cells. Cells grown to confluence reached this phase faster (within 16h), which is consistent with the fact that this synchronization method does not induce the cells to enter the G0 phase, which therefore need less time to resume the cycle. Most starved cells remained synchronized until 48h post release, whereas the cells grown to confluence became desynchronized. In conclusion, the CHO Fucci cells appeared to express the fluorescent probes in a cell cycle dependent manner and therefore may serve as a model to study cell cycle regulation.

6.4.2 Knock-down of HR factors differentially influences cell cycle distribution of CHO Fucci cells

To assess the role of HR factors in cell cycle regulation, we subsequently treated the CHO Fucci cells with a panel of short interfering RNAs (siRNAs) directed against these genes. The efficiency of siRNA knock-down was assessed by qPCR (Fig.6S3A). Rad51 and Rad51D knock-down was also confirmed by western blot (Fig.6S3B). A non-targeting siRNA as well as an siRNA against Cyclin D1, which is required for the progression through the G1/S checkpoint [16], were used as a negative and positive control, respectively.

The different siRNA treatments did not affect the number of freshly divided early G1 phase cells (Fig.6.2A). However, we noted a decrease in this sub-population, albeit not significant, with Rad51C siRNA. The number of G1/G0 cells was significantly increased upon treatment with the Cyclin D1 siRNA, as expected from the role of this cyclin in the cell cycle (Fig.6.2B). Interestingly, the knock-down of three Rad51 paralogs - Rad51B, -C, and -D, as well as Rad50, one of the MRN components, and to a lower extant CtIP,

had a similar effect. The accumulation of G1/G0 cells upon the knock-down of these factors suggests that the progression through the G1 phase may be perturbed in their absence, possibly involving the G1/S checkpoint. This is surprising, as these proteins are thought to operate primarily in the late S and G2 phases of the cell cycle, when HR is most active. However, it is possible that these factors may be necessary for the G2/M checkpoint activation in response to DNA damage. In this scenario, the absence of these factors would cause the cells to circumvent arrest and enter mitosis despite the presence of unrepaired breaks, which later on would activate the G1/S checkpoint.

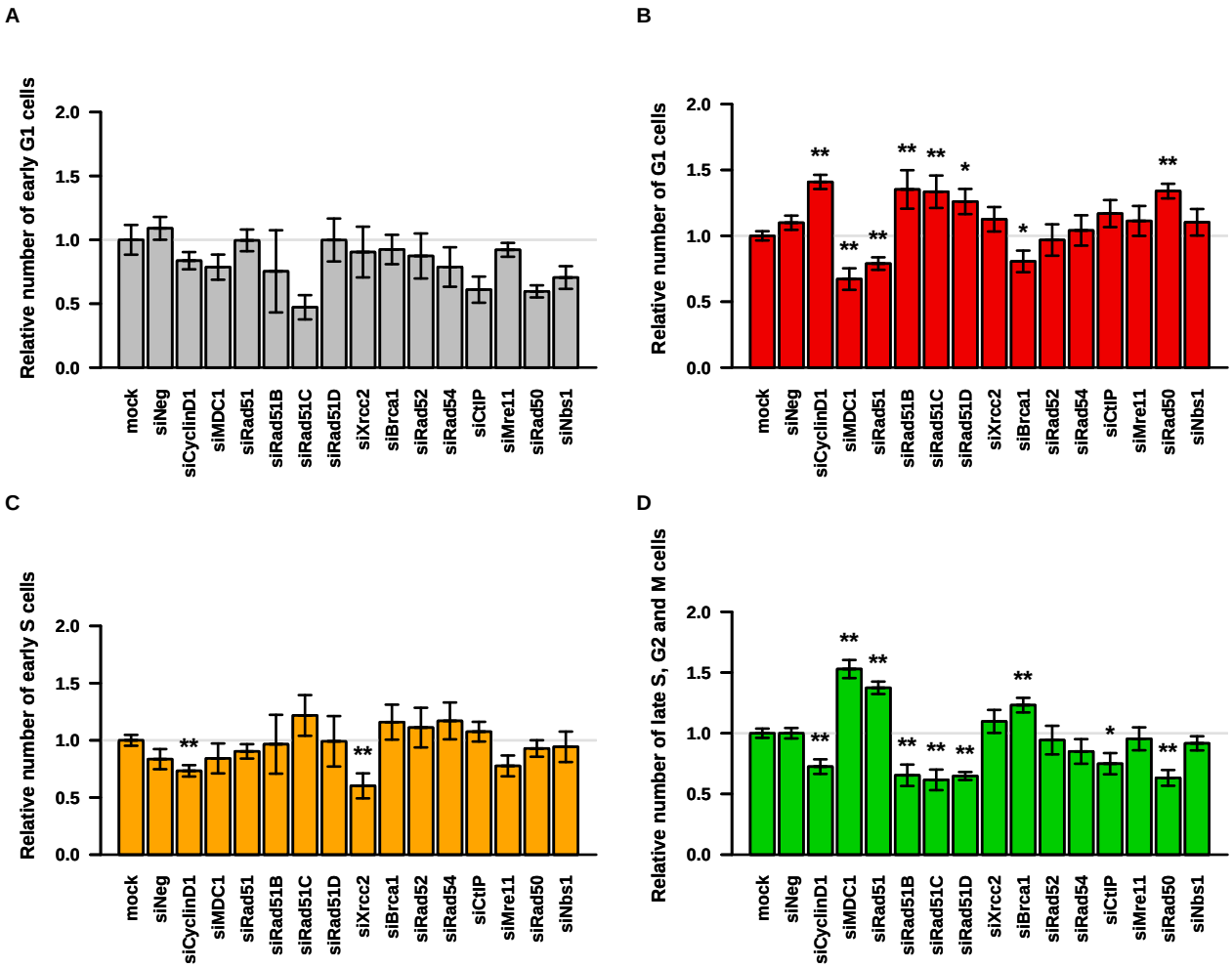


Figure 6.2: Knock-down of HR factors affects cell cycle distribution of CHO Fucci cells. Results are shown as fold change over the data obtained from mock-treated cells (mock). Mean of $\geq$3 experiments, error bars show s.e.m. Asterisks indicate significant differences between siRNA-treated samples and mock control. Statistical signicance relative to mock was determined by unpaired Student's t-test with Benjamini-Hochberg correction; signicance level $p<0.05$ (*), $p<0.01$ (**).

As expected, the proportion of early S phase cells was significantly decreased in the absence of Cyclin D1, due to the defective G1/S transition (Fig.6.2C). We also noted a decrease in this subpopulation in the presence of Xrcc2 siRNA, which, however, did not correlate with an increase in the number of G1/G0 phase cells.

The knock-down of MDC1, Rad51 and Brca1 resulted in a significant accumulation of green-fluorescent late S, G2 or M phase cells (Fig.6.2D). This further supported the view that these proteins are required for the progression through the G2/M checkpoint [86, 202, 297, 326, 367]. The percentage of green fluorescent cells was also slightly increased in the presence of Xrcc2 siRNA. In contrast, the silencing of Cyclin D1 as well as Rad51B, -C, -D, Rad50 and CtIP, resulted in a significant decrease in this subpopulation, due to the aforementioned accumulation of G1/G0 phase cells.

In conclusion, several HR proteins appear to be involved in cell cycle regulation, albeit differentially. MDC1, Rad51 and Brca1 seem to be essential for the progression from S and G2 phases into mitosis, while Rad51 paralogs, Rad51B, -C and -D, as well as the DNA end resection enzymes, Rad50 and CtIP may be required for activating the G2/M checkpoint in response to damage and/or progression through the G1/S checkpoint.

6.4.3 Knock-down of Rad51B, Rad51C, CtIP and Rad50 induces entry into the G0 phase

In addition to increasing the number of mKO2-positive cells, we also observed that the knock-down of Cyclin D1 and several HR factors increased the level of mKO2 fluorescence (Fig.6S4). A detailed analysis of these cells revealed two sub-populations with different fluorescence intensities (Fig.6.3A). A recently published report identified these low- and high mKO2-expressing cells as cycling G1 and quiescent G0 cells, respectively [325]. In line with this, we observed an increase in the number of bright red fluorescent cells upon

serum starvation (Fig.6S5), further supporting the view that this sub-population represents non-cycling G0 cells.

We therefore set out to quantify the G1 and G0 sub-populations in cells transfected with HR siRNAs. In controls, as well as in most siRNA-treated samples, the G0 phase cells constituted only approx. 5-10% of the population (Fig.6.3B). However, treatment with Cyclin D1 siRNA increased the number of quiescent cells to 30%. This is consistent with previous studies showing that Cyclin D1 deficiency causes the cells to enter the G0 phase [133]. Interestingly, we also observed a very significant increase in the number of G0 cells upon the knock-down of Rad51B, Rad51C, CtIP, and Rad50. This was especially striking in the presence of Rad51B and Rad51C siRNAs, where G0 cells comprised up to 40% of the entire population. This implied that the absence of these HR factors may constitute a signal to enter the quiescent state.

6.4.4 Rad51 and Rad51C depletion leads to cell cycle arrest

We next sought to investigate whether the aberrant cell cycle distribution upon siRNA knock-down of HR factors results from a cell cycle arrest or a delayed cycle progression. We focused our attention on Rad51 and Rad51C, the two HR proteins with pronounced, but distinct effects on the cell cycle. We synchronized siRNA-treated cells in early S phase by sorting double positive mKO2+mAG+ cells and analyzed their cell cycle distribution 1 and 2 days after sorting (Fig.6S6A). Since the normal doubling time of CHO DG44 cells is approx. 12-14h [116, 119], we estimated that in the 18h between sorting and the first measurements, the cells should have completed the cell cycle, divided and nearly completed another cycle. According to these calculations the majority of cells would be green fluorescent at the beginning of the analysis (40h post transfection). This was indeed the case for untreated cells, and for cells transfected with the non-targeting siRNA

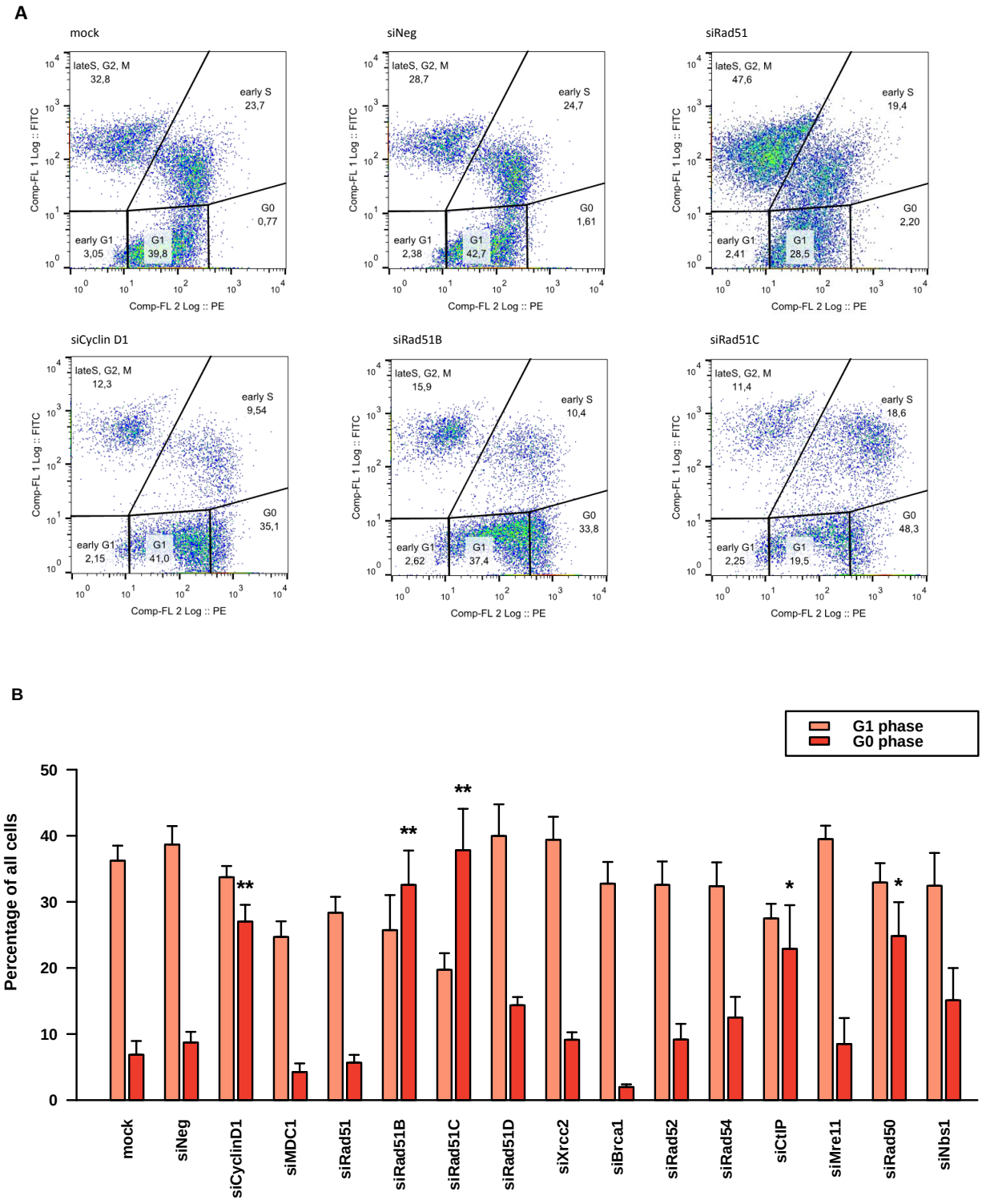


Figure 6.3: Knock-down of Cyclin D1, Rad51B, Rad51C, CtIP and Rad50 induces senescence. A) FACS plots of siRNA-treated cells. B) Percentages of cells in G1 and G0 phases. Mean of ≥3 experiments, error bars show s.e.m. Asterisks show significant differences between siRNA-treated samples and untreated (mock) control. Statistical signicance relative to mock was determined by unpaired Student's t-test with Benjamini-Hochberg correction; signicance level $p<0.05$ (*), $p<0.01$ (**).

(Fig.6S6B,C; Fig.6S7A,B). Both populations also showed a similar cell cycle distribution over time with two more divisions, one between 44h and 48h and another less than 64h post transfection. The time between these divisions was also consistent with the normal CHO cell cycle duration.

Cells treated with the Cyclin D1 siRNA initially displayed a similar cell cycle pattern as the controls, with one division 44-48h post transfection (Fig.6S6D, Fig.6S7C). However, the next division seemed delayed to 64-68h, and the accumulation of G1 and G0 phase cells started to become apparent. In Rad51-depleted cells, already the first division was delayed compared to the controls (48-64h), after which most cells appeared not to divide anymore (Fig.6S6E, Fig.6S7D). A small portion of cells (approx. 20%) underwent a second division at around 64-68h after transfection. The percentage of green fluorescent cells stayed high throughout the experiment indicating that most cells were arrested in late S, G2 or M phase. This could indicate that a strong Rad51 deficiency causes the cells to arrest in the late phases of the cycle, likely at the G2/M checkpoint. Mild Rad51 depletion, possibly due to lower knock-down efficiency, may result in delayed cell cycle progression.

Cells treated with Rad51C siRNA also divided later than the controls (between 48h and 64h) pointing to delay in the cell cycle (Fig.6S6F, Fig.6S7E). However, Rad51C-deficiency resulted in a steadily elevated proportion of cells in G1 and G0 phases, resembling the effect of Cyclin D1 knock-down. Indeed, cluster analysis demonstrated that Rad51C-depleted cells grouped together with Cyclin D1 siRNA-treated cells (data not shown). However, the majority of these red-fluorescent cells appeared to be quiescent, reaching 50% of the entire population (Fig.6.4). This was more than observed in the presence of Cyclin D1 siRNA (10-20%). After the delayed first cycle, most of the Rad51C-deficient cells underwent a second division within a normal time of 12-14h, although some cells remained arrested in the G1 or G0 phase. This indicated that out of all the cells initially arrested in a

quiescent G0 phase, some were later released from the block and re-entered the cell cycle, possibly due to the loss of the knock-down effect. At 64h post transfection most of the cells reverted to a profile resembling untreated cells.

Overall, we concluded that Rad51 knock-down causes an arrest at the late phases of the cell cycle, most probably due to the G2/M checkpoint, or delayed cell cycle progression in cells circumventing the arrest. In contrast, Rad51C depletion causes entry into the quiescent G0 phase, most likely due to the activation of the G1/S checkpoint.

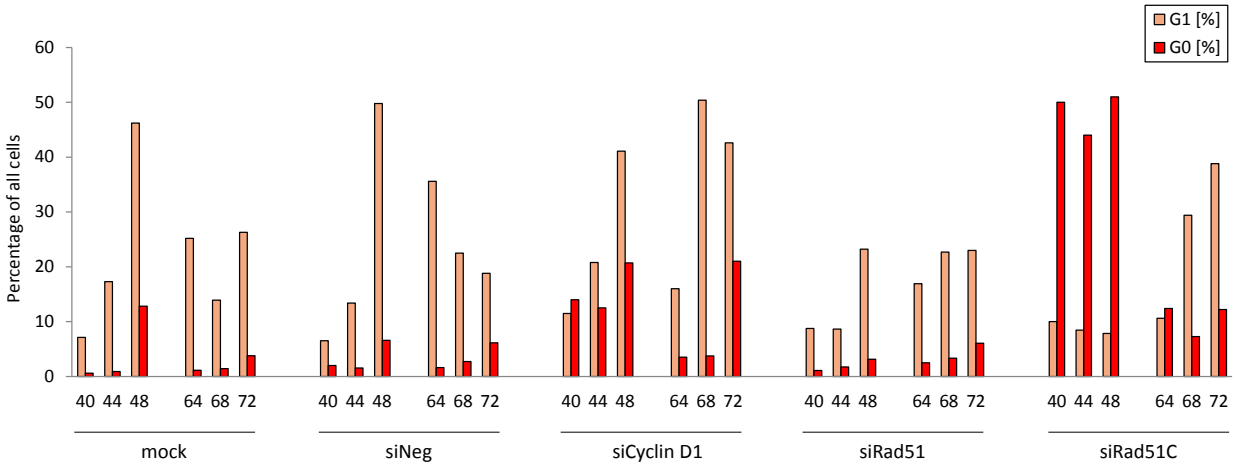


Figure 6.4: Effect of HR protein depletion on the percentages of quiescent cells in CHO Fucci cells synchronized in early S phase. Numbers below the bars represent time (in hours) post siRNA transfection. The siRNA targets are indicated below the plot. Mock – cells treated with transfection reagent only.

6.5 Discussion

The DNA damage response encompasses many functionally interconnected pathways, including cell cycle checkpoint signaling and DNA repair. Many proteins participating in cell cycle regulation are also known to control DNA repair, whereas the converse was not known to be true for DNA repair factors. However, recent evidence suggested that some DSB repair proteins may also be implicated in cell cycle regulation. To further explore this connection between DNA repair proteins and cell cycle machinery, we analyzed the effect of HR protein depletion on the CHO cell cycle in the absence of induced DNA damage. We

found that the knock-down of several HR factors altered cell cycle progression, suggesting that they may play a role in cell cycle control. However, lack of these proteins affected the cell cycle differentially, despite their being part of the same DNA repair pathway.

In the late S and G2 phases, the presence of unrepaired DNA breaks results in the activation of the G2/M cell cycle checkpoint by specialized DNA damage sensing factors, e.g. ATM, ATR, p53, and Chk1 [189, 321]. We anticipated that the absence of crucial HR factors would lead to the accumulation of unrepaired DSBs, thereby stimulating the DDR to arrest the cells before division. Surprisingly, the knock-down of the majority of HR proteins, including three Rad51 paralogs, CtIP, and Rad50, failed to arrest the cells at the G2/M checkpoint. Instead, the presence of these factors seemed to be necessary for the progression from G1 to S phase. Only the loss of MDC1, Rad51 and Brca1 led to an accumulation of late S, G2 and M cells, likely due to the activation of the intra-S and/or G2/M checkpoints. This could indicate that the knock-down of these three genes, important for the early steps of the HR pathway, leads to the accumulation of enough endogenous DNA damage to trigger the G2/M checkpoint.

The absence of the remaining HR factors may potentially inhibit the repair process incompletely, instead rendering it more error-prone. These cells would still enter mitosis, but due to the accumulation of imprecisely repaired DSBs, they may become arrested at the next G1/S checkpoint, explaining the accumulation of cells in the G1/G0 phase. However, in a recent study, Shibata and co-workers estimated that only about 15% of DSBs occurring in the G2 phase are repaired by the HR pathway, while the remaining breaks are efficiently repaired by other mechanisms [292]. This, together with our results obtained in the absence of induced DNA damage, may indicate that the effects of HR protein knock-down described here are not merely due to the accumulation of unrepaired DSBs. Instead, these HR proteins may play a more direct role in the cell cycle, for instance

by interacting with cell cycle signaling factors, cyclins or cyclin-dependent kinases.

Our results obtained with Rad51B and -C siRNAs are in contrast with recently published observations that the knock-down of these Rad51 paralogs in HeLa cells blocks progression through the G/M checkpoint [276]. However, it is possible that the absence of these proteins in human cells may have a different impact on HR and/or the cell cycle, than it does in CHO cells, which divide two times faster than HeLa cells and display different kinetics of DSB repair [154, 263].

In the present study, we observed that apart from a defect in the G1/S transition, the knock-down of Rad51B, Rad51C, CtIP and Rad50 also caused entry into the non-proliferative G0 phase, resembling Cyclin D1 knock-down. This suggests that the absence of these proteins may constitute a signal for the cells to withdraw from the cell cycle. This may partially explain the characteristic enhanced proliferation of cancer cells, in which HR proteins are often overexpressed. Taken together, this work indicates that many HR components may be necessary for the normal cell cycle progression, at least in CHO cells. It will therefore be of interest to decipher how these proteins transmit the signals to the cell cycle control machinery, and what are their putative targets, to further understand their role in the cell cycle regulation network.

6.6 Supplementary materials

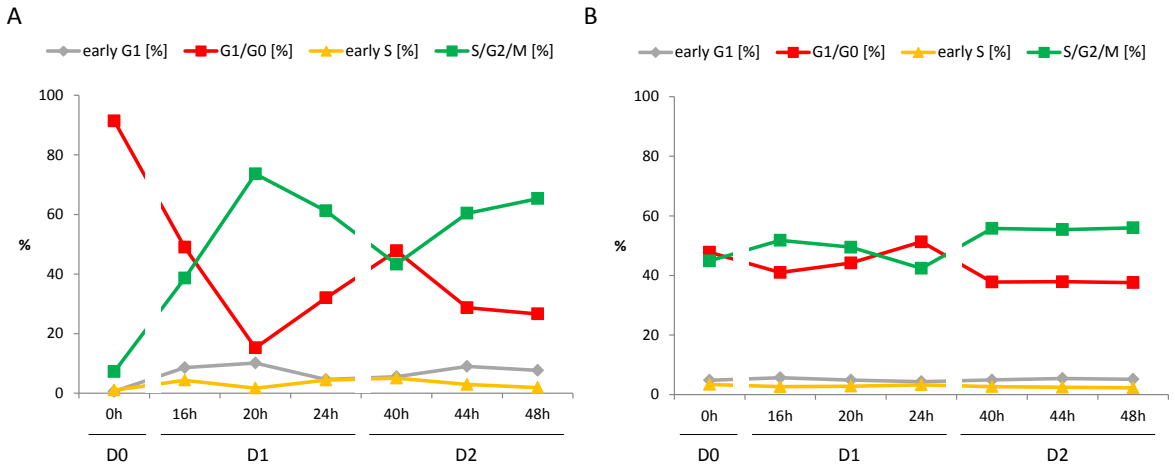


Figure 6S1: Cell cycle progression of synchronized CHO Fucci cells. Cells were synchronized by A) serum starvation, or B) contact inhibition. Cell cycle distribution data is represented as percentages of cells in a given cell cycle phase at the indicated time post release from the cell cycle arrest. Flow cytometry measurements were taken on the day of release (D0), and during the first and second mitotic cycle (D1 and D2).

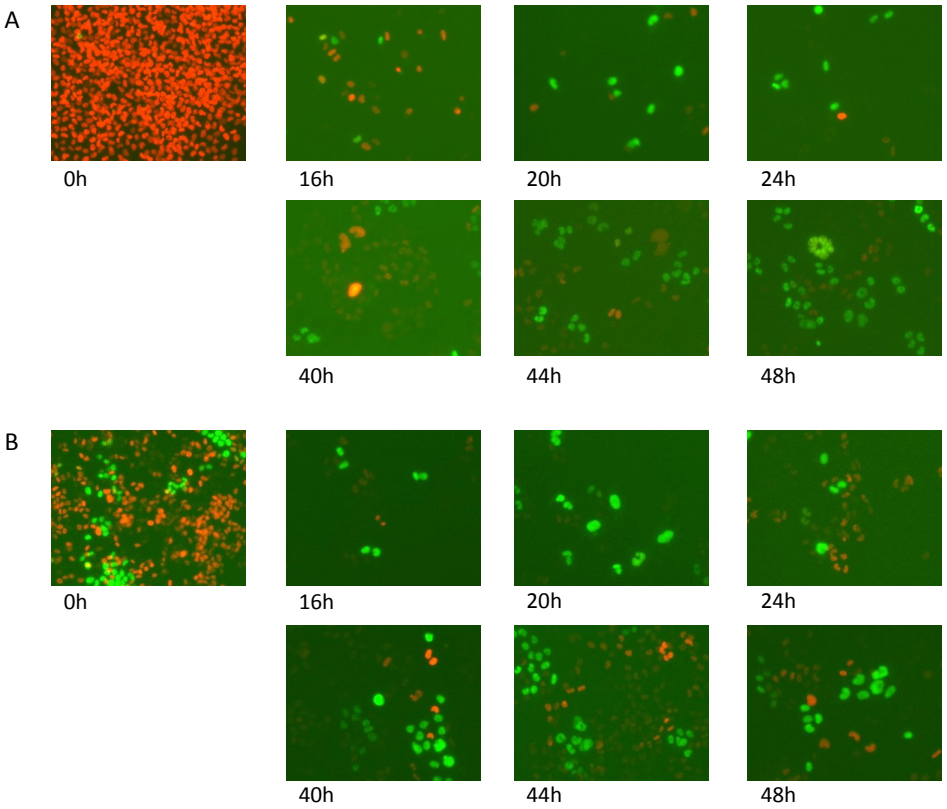


Figure 6S2: Microscope images of synchronized CHO Fucci cells. A) Serum starvation, B) contact inhibition. Time post release from cell cycle arrest is indicated below each image. All images were acquired using the same exposure time.

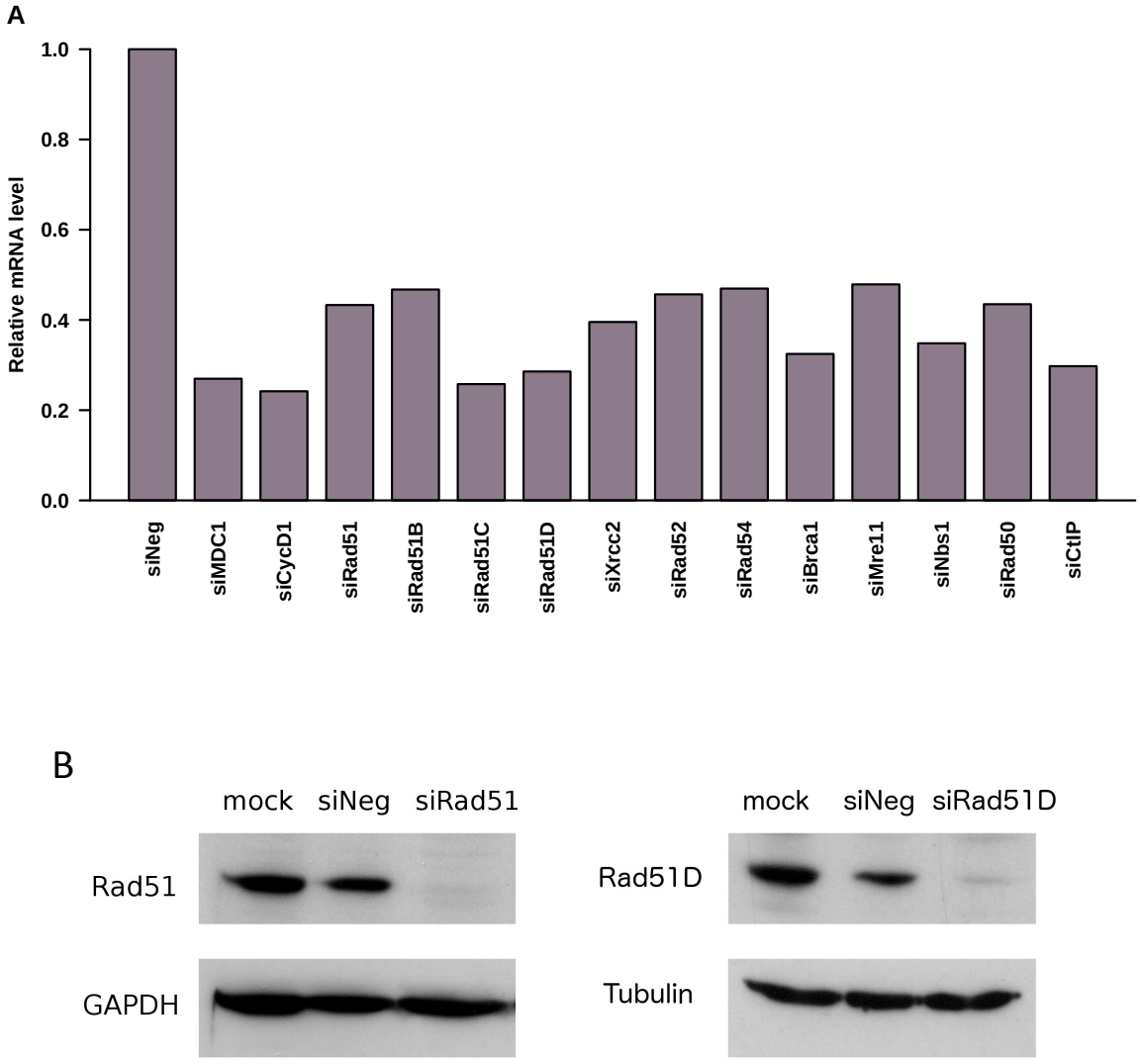


Figure 6S3: The effect of siRNA knock-down on target mRNA and protein levels. CHO cells transfected with three siRNAs targeting the indicated gene, with three non-targeting siRNAs (siNeg) or left untreated (mock). A) The mRNA level of the target was quantified by qPCR using specific primers. Values were normalized to the target mRNA levels in cells transfected with the non-targeting siRNA. B) Western blotting for Rad51 (left) and Rad51D (right) in total protein extracts isolated from CHO cells. GAPDH and Tubulin – loading controls.

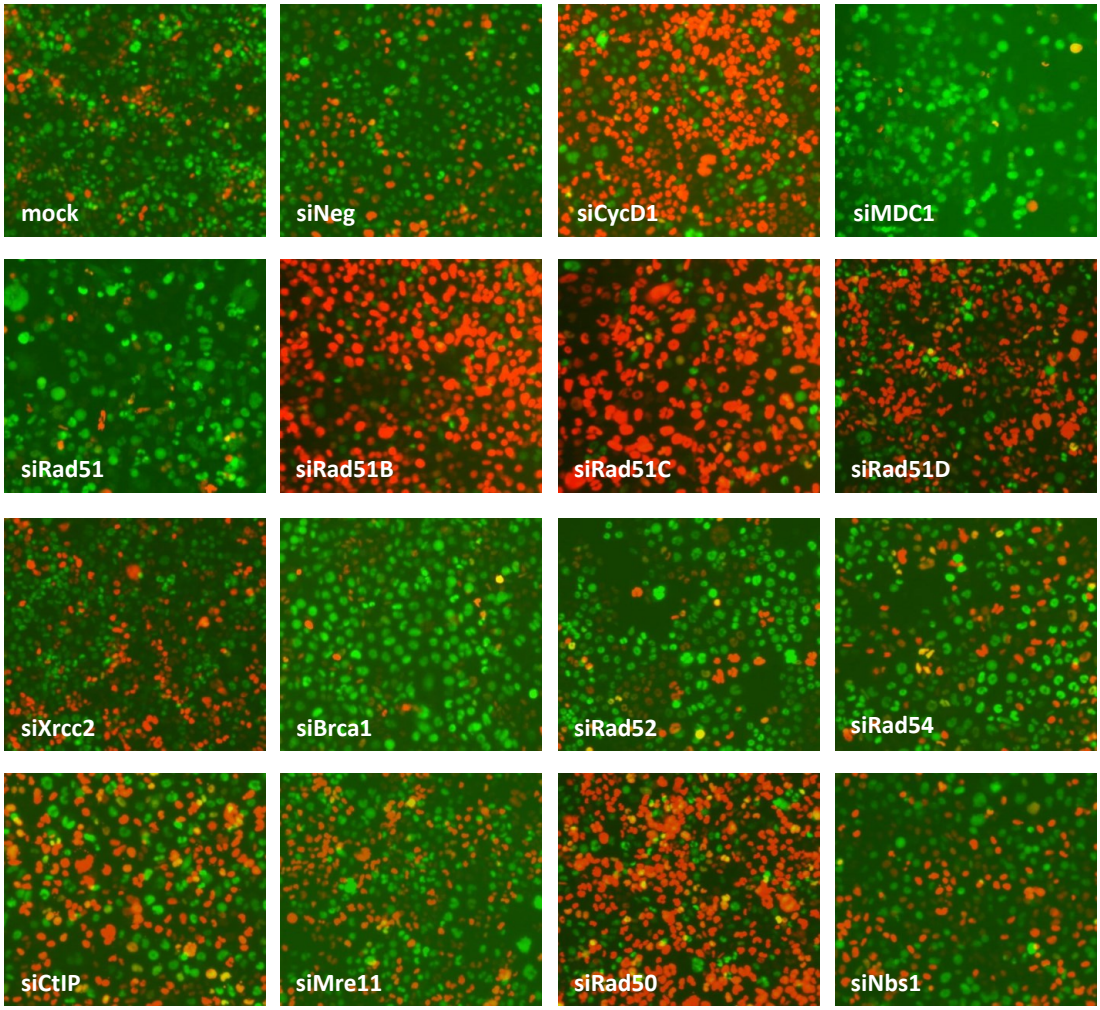


Figure 6S4: Microscope images of CHO Fucci transfected with siRNAs against HR factors. Images acquired 72h after the indicated siRNA transfection, or with the transfection reagent only (mock).

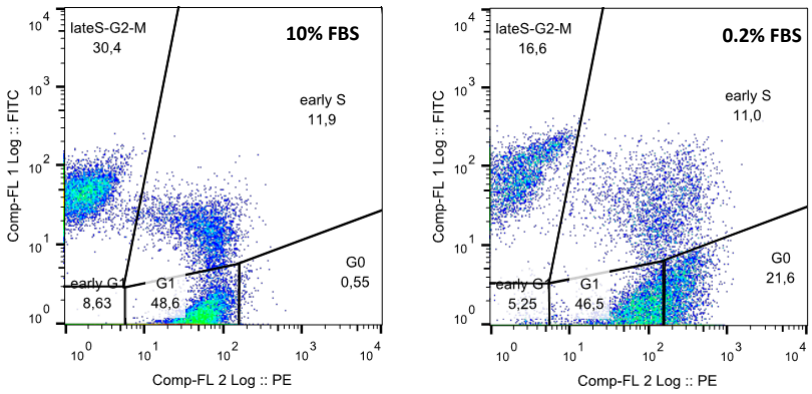


Figure 6S5: Flow cytometry profiles of CHO Fucci cells grown in full medium and with low serum content. The amount of fetal bovine serum (FBS) in the medium is indicated on each plot. Cells grown in full medium (left), cells grown low serum medium (right).

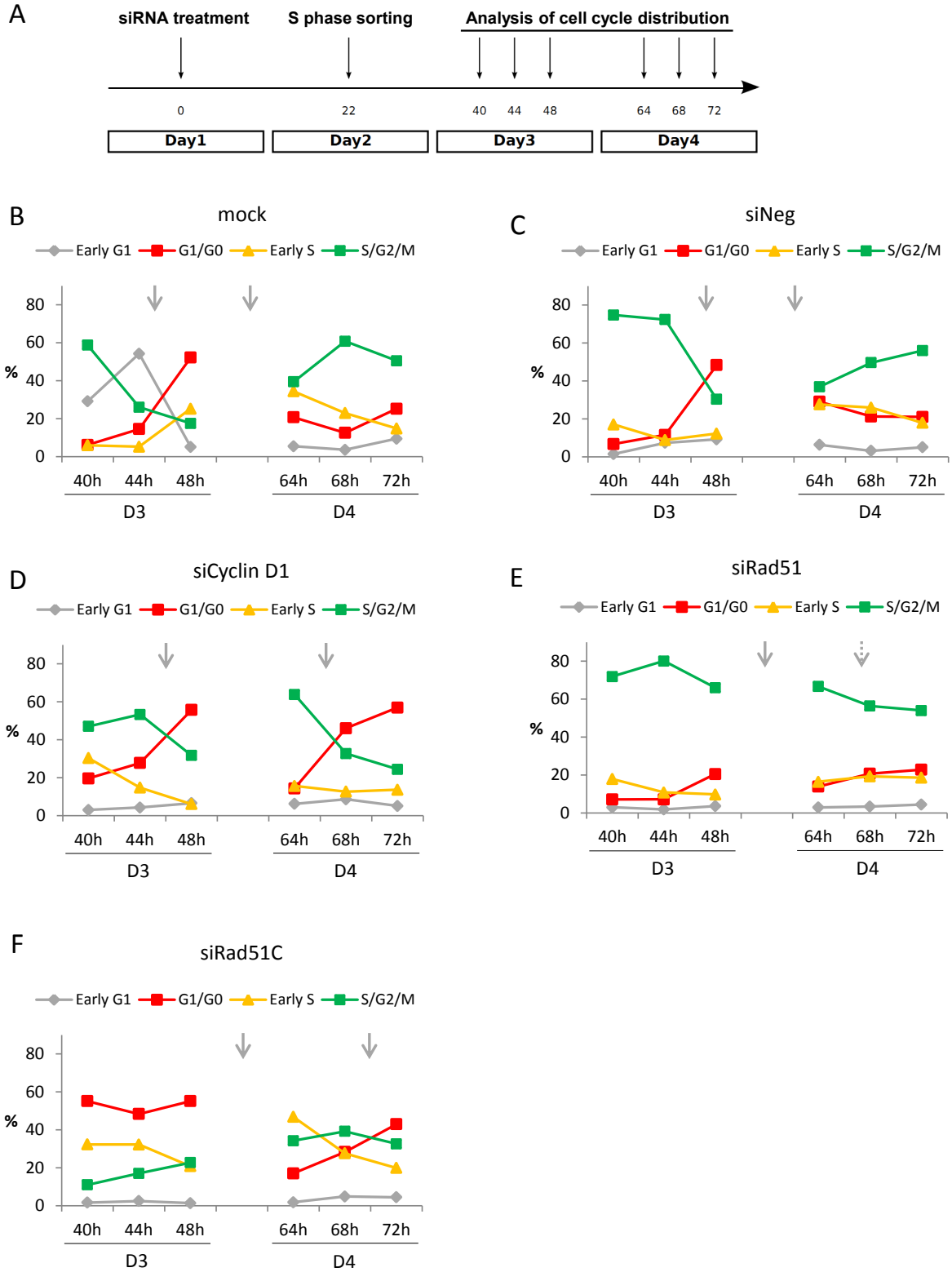


Figure 6S6: Effect of Rad51 and Rad51C knock-down on cell cycle progression of CHO Fucci cells synchronized in early S phase. A) General outline of the procedure. B-F) Percentages of cells in a given cell cycle phase at the indicated time post siRNA transfection (day 3 and 4 –D3 and D4). Arrows indicate the estimated average time of cell division, dotted arrow indicates a cell division of a subpopulation of cells.

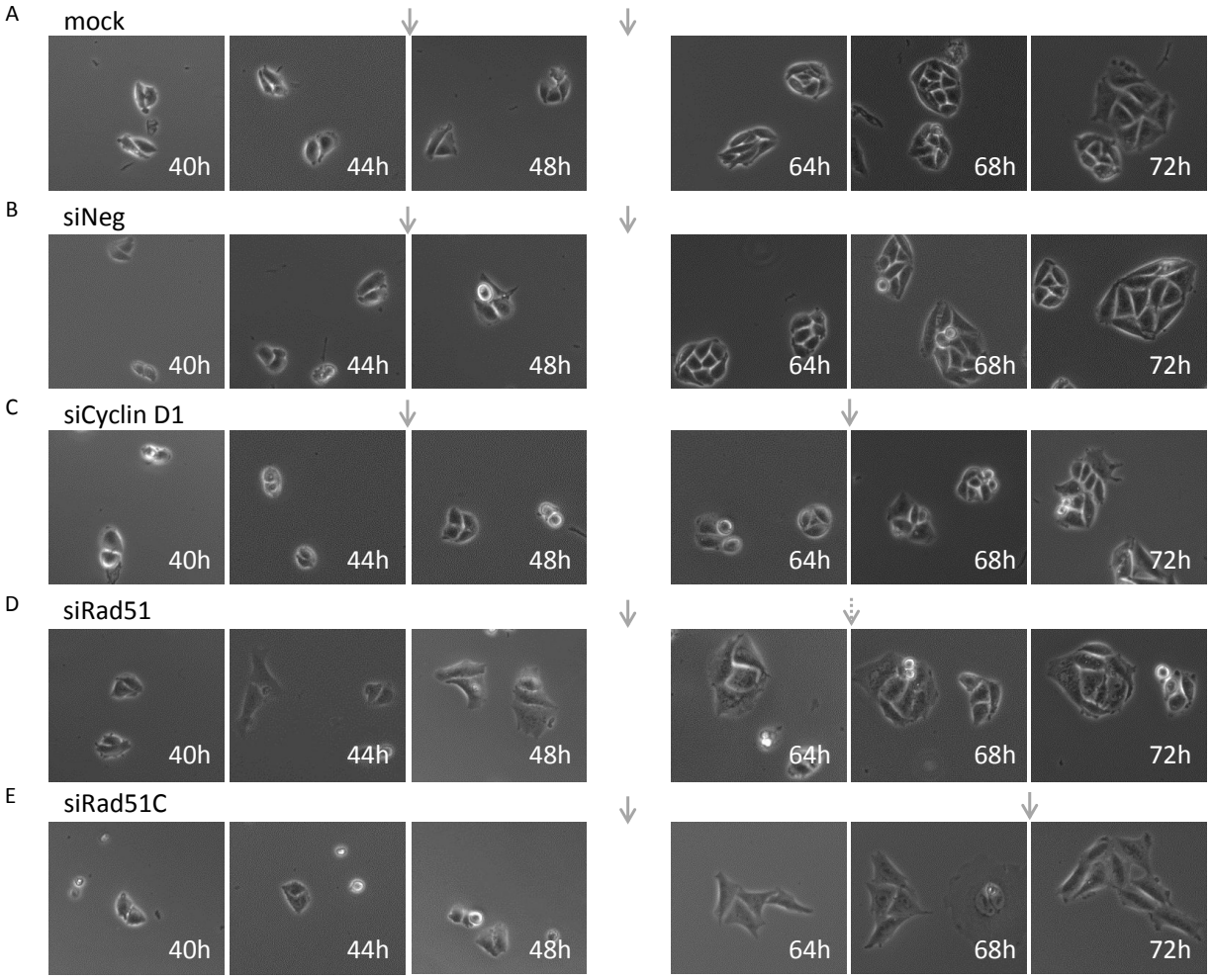


Figure 6S7: Phase contrast images of cells synchronized in early S phase. Images were acquired at the indicated time post siRNA transfection. Arrows indicate the estimated time of cell division, dotted arrow - a division of a cell subpopulation.

Chapter 7

Conclusions

7.1 Plasmid integration in mammalian cells

Introduction of foreign DNA into the eukaryotic cells can lead to its incorporation into the genome [94, 156, 275]. However, the enzymatic machinery responsible for this phenomenon remains largely undefined. In lower eukaryotes, plasmids containing genome-derived sequences were shown to preferentially recombine with the homologous region of the host cell chromosome [127, 249], pointing to a HR-dependent mechanism of integration. However, in mammalian cells, this process does not seem to rely on extensive homology between the exogenous and cellular DNA [156]. Moreover, the integration loci do not appear to be restricted to any specific chromosomes or chromosomal regions [275]. This suggested a random and homology-independent mode of integration.

According to the most widely accepted integration mechanism, the exogenous DNA fragments are joined with the cellular DNA during the repair of naturally occurring genomic lesions [167, 359]. In this model, the free plasmid ends are treated by the cell DNA repair machinery like chromosomal DSBs. In the attempt to 'repair' them, the cellular enzymes connect them to any neighboring, plasmid or genomic, DNA ends. The former case leads to the formation of long plasmid concatemers, the latter results in the incorporation of the vector into the genome. Since NHEJ was shown to very efficiently

join broken DNA ends in mammalian cells, this pathway was proposed to be responsible for the integration process.

However, the use of sequencing techniques revealed the presence of rearrangements, deletions, insertions, and microhomologies in the regions surrounding the integration sites [136, 203, 225]. These findings seem inconsistent with NHEJ, since it is considered to be a fairly conservative and precise process (reviewed in [25]). Indeed, in the past years many groups reported the existence of an alternative, highly error-prone end-joining pathway relying on a different set of enzymes than NHEJ and employing microhomologies in the repair process. This pathway, termed alternative NHEJ or MMEJ, was also associated with spontaneous genomic rearrangements, loss of chromosome fragments and duplications in cancer cells [138, 244, 299, 381]. Therefore, it seems possible that this mechanism, rather than classical NHEJ, may be primarily responsible for plasmid genomic integration.

In the present study we find that a specific subset of MMEJ, namely the SD-MMEJ pathway, allows DNA transfected into CHO cells to integrate into the genome. Furthermore, we propose the existence of two distinct SD-MMEJ subpathways, relying on different subsets of enzymes. One of these mechanisms, dependent on base- and nucleotide-excision repair factors and translesion synthesis polymerases, seems to be responsible for forming plasmid concatemers. The other pathway, partially dependent on HR and/or DNA replication enzymes, seems to mediate plasmid recombination with the genome.

7.2 The role of MAR elements in DNA recombination

Previous studies performed by our group suggested that matrix attachment regions may mediate increased plasmid integration into the genome of CHO cells by stimulating DNA recombination. In the present work, we demonstrated that MAR-mediated plasmid integration results from the enhanced SD-MMEJ pathway of DSB repair. However,

the precise molecular mechanism by which MARs promote this process remains to be elucidated.

MARs are very AT-rich and possess a high potential to unwind and denature the DNA double helix [29, 31, 90, 262]. They were also shown to contain multiple target sites of topoisomerase II, enzyme responsible for removing DNA supercoils by inducing DSBs [57]. These properties may potentially make MAR-containing chromatin regions more prone to breaking. Indeed, MARs were previously associated with hotspots of DNA cleavage and illegitimate recombination [307]. Recently Myers and co-workers also reported the identification of a consensus motif enriched in recombination hot spots in human cells [235, 236]. Interestingly, the two human MAR elements used in our study also seem to contain these short sequences. Finally, MARs were also shown to be present in so-called fragile sites, chromosomal loci with high tendency to breakage [143, 267]. These sites were associated with genomic rearrangements, including deletions, translocations and inversions, suggesting that DSBs may frequently arise in these regions [81, 274]. Moreover, plasmid DNA was shown to preferentially integrate inside these sites [270]. All this taken together suggests that MAR elements may constitute recombination hotspots, making them preferred sites for illegitimate integration of foreign DNA.

The MMEJ pathway was reported to be most active during the S phase [328], suggesting that it may be one of the primary mechanisms responsible for repairing breaks arising at replication forks. In fact, several MMEJ-like mechanisms, characterized by the occurrence of genomic rearrangements and microhomology junctions, were proposed to repair these one-ended DSBs [60, 182]. MAR elements were previously reported to participate in the initiation of DNA replication in mammalian cells, suggesting that these elements may recruit replication enzymes [71, 259]. Among the factors mediating replication several were shown to play a role in MMEJ, including Ligase I and Pold3. This points to a potential

link between DNA replication, MMEJ-mediated repair and MAR elements.

Finally, it is possible that MARs are able to directly recruit MMEJ factors. It has already been proposed that MARs associate with NHEJ enzymes, including Ku, DNA-PKcs, Xrcc4, Ligase IV, as well as PARP1, an enzyme implicated in MMEJ [100, 220]. However, this was not assessed for other DSB repair factors. Therefore, additional experiments are needed to fully elucidate the possible direct interaction between MARs and the MMEJ machinery.

7.3 The problems associated with studying DNA recombination

The studies of DNA recombination mechanisms in mammalian cells encounter many difficulties stemming from the considerable genome sizes as well as from the abundance of DNA repair mechanisms operating in these cells. Much of the research on DSB repair pathways, including this work, was done in immortalized cultured cells. These models usually originate from cancer cells and likely encounter more DNA damage than normal cells, due to potential mutations of DNA repair genes, oxidative stress and repeated replication and division cycles. Therefore, it can be expected that these cells may possess more efficient DSB repair mechanisms and that the relative participation of different pathways in the repair process is distinct from normally growing cells. Indeed, studies show that alternative DSB repair pathways are often upregulated in tumor cells [23, 323]. Many cancers were also reported to possess a more active HR pathway [206, 360]. Therefore, the conclusions derived from experiments performed in these cells should not be extrapolated to all cell types.

Many DSB repair proteins were discovered and described by using mutant cell lines sensitive to DNA damaging agents or ionizing radiation. However, most of these cell lines were generated by chemical mutagenesis and many of them were never sequenced at the

genomic level. Consequently, their underlying mutations remain largely unknown. In this study we analyzed two CHO mutants reported to be deficient in the components of the NHEJ and HR pathways. Full transcriptome sequencing of these cells revealed significant differences in the expression of other DNA repair genes, calling into question the reliability of these cell lines and their usefulness in the studies of DSB repair mechanisms.

Plasmid recombination assays are another tool frequently used to study DNA recombination in mammalian cells. These assays often rely on the reconstitution of an inactive reporter gene by repairing an enzymatically induced DSB. While these assays certainly contributed to our understanding of the DSB repair in mammalian cells, they are not always very specific and often fail to detect all types of repair events. Unless junction sequencing is performed this may falsify the true contribution of the studied pathway to the DSB repair process. This proved to be the case for the published reporter assays used in this study. Moreover, these kind of assays only allow to study the repair of DSBs generated by restriction enzyme digestion, which require little processing and can be efficiently repaired by simple re-ligation. However, in vivo DSBs are often chemically modified, possess incompatible DNA ends or lack terminal 3'-hydroxyl or 5'-phosphate groups precluding ligation. Such ends require modification or resection prior to repair, and may therefore be often repaired by other pathways.

The investigation of alternative DSB repair pathways is particularly complicated due to the reported existence of many distinct sub-pathways. Moreover, the factors thought to be implicated in MMEJ also participate in other DNA repair mechanisms and/or DNA replication. This makes it very difficult to interpret the effects of their overexpression, knock-down or knock-out, and is likely the reason why many conflicting reports have been published on the subject. Interestingly, the results included in this work demonstrate that particular HR proteins may also have pleiotropic effects, and that they may also contribute

to one type of SD-MMEJ pathways. Furthermore, in addition to their roles in DNA repair, they may also influence the cell cycle progression. In conclusion, all these factors should be taken into account when studying the function of particular DSB repair pathways.

7.4 Final remarks

The drawback of random plasmid integration is the lack of control over the chromosomal integration site. This can result in insertional mutagenesis, proto-oncogene activation or silencing of the gene of interest. Gene targeting enables transgene integration at a predefined genomic locus, allowing for safer generation of stably transfected cells. However, this procedure is extremely inefficient in mammalian cells. Until now, the efforts to enhance this process were aimed at inhibiting NHEJ. Many studies show that blocking the NHEJ pathway enables to enhance HR. However, it only allows to increase targeted integration 2-30 fold, depending on the cell type [24, 136, 378]. This can be regarded as a moderate improvement considering the fact that random integration events were estimated to occur about 100-10000 times more frequently than targeted events (reviewed in [32, 301, 339]). Moreover, while this strategy enables to increase HR-mediated integration, it results in little or no decrease of random integration, likely due to very efficient alternative end-joining pathways in the target cells [24, 136]. Here, we identify the SD-MMEJ pathway as one of the primary mechanisms driving illegitimate integration in mammalian cells. This knowledge should help to modify the DNA recombination potential of cultured mammalian cells and increase the efficiency of gene targeting.

Bibliography

[1] Abranches, R., Shultz, R. W., Thompson, W. F., and Allen, G. C. Matrix attachment regions and regulated transcription increase and stabilize transgene expression. *Plant biotechnology journal 3*, 5 (Sept. 2005), 535–43. [cited at p. 79]

[2] Ahmad, A., Robinson, A. R., Duensing, A., van Drunen, E., Beverloo, H. B., Weisberg, D. B., Hasty, P., Hoeijmakers, J. H. J., and Niedernhofer, L. J. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. *Molecular and cellular biology 28*, 16 (Aug. 2008), 5082–92. [cited at p. 31]

[3] Ahnesorg, P., Smith, P., and Jackson, S. P. XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell 124*, 2 (Jan. 2006), 301–13. [cited at p. 25]

[4] Allen, C., Kurimasa, A., Brenneman, M. A., Chen, D. J., and Nickoloff, J. A. DNA-dependent protein kinase suppresses double-strand break-induced and spontaneous homologous recombination. *Proceedings of the National Academy of Sciences of the United States of America 99*, 6 (Mar. 2002), 3758–63. [cited at p. 93]

[5] Allen, G. C., Spiker, S., and Thompson, W. F. Use of matrix attachment regions (MARs) to minimize transgene silencing. *Plant molecular biology 43*, 2-3 (June 2000), 361–76. [cited at p. 11, 37, 62]

[6] Anami, S., Njuguna, E., Coussens, G., Aesaert, S., and Van Lijsebettens, M. Higher plant transformation: principles and molecular tools. *The International journal of developmental biology 57*, 6-8 (Jan. 2013), 483–94. [cited at p. 1]

[7] Anders, S., and Huber, W. Differential expression analysis for sequence count data. *Genome biology 11*, 10 (2010), R106. [cited at p. 41]

[8] Andersen, S. L., and Sekelsky, J. Meiotic versus Mitotic Recombination: Two Different Routes for Double-Strand Break Repair. *Bioassays 32*, 12 (2010), 1058–1066. [cited at p. 30]

[9] Arnaudeau, C., Rozier, L., Cazaux, C., Defais, M., Jenssen, D., and Helleday, T. RAD51 supports spontaneous non-homologous recombination in mammalian cells, but not the corresponding process induced by topoisomerase inhibitors. *Nucleic acids research 29*, 3 (Feb. 2001), 662–7. [cited at p. 93]

[10] Arope, S., Harraghy, N., Pjanic, M., and Mermod, N. Molecular characterization of a human matrix attachment region epigenetic regulator. *PLoS ONE 8*, 11 (2013), e79262. [cited at p. 11]

[11] Atkinson, H., and Chalmers, R. Delivering the goods: viral and non-viral gene therapy systems and the inherent limits on cargo DNA and internal sequences. *Genetica 138*, 5 (May 2010), 485–98. [cited at p. 3]

[12] Audebert, M., Salles, B., and Calsou, P. Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. *The Journal of biological chemistry 279*, 53 (Dec. 2004), 55117–26. [cited at p. 30, 61, 94, 101]

[13] Aymard, F., Bugler, B., Schmidt, C. K., Guillou, E., Caron, P., Briois, S., Iacovoni, J. S., Daburon, V., Miller, K. M., Jackson, S. P., and Legube, G. Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nature structural & molecular biology 21*, 4 (2014), 366–74. [cited at p. 78]

[14] Badie, S., Liao, C., Thanasoula, M., Barber, P., Hill, M. A., and Tarsounas, M. RAD51C facilitates checkpoint signaling by promoting CHK2 phosphorylation. *Journal of Cell Biology 185*, 4 (2009), 587–600. [cited at p. 132, 133]

[15] Bak, S. T., Sakellariou, D., and Pena-Diaz, J. The dual nature of mismatch repair as antimutator and mutator: for better or for worse. *Frontiers in genetics 5*, August (Jan. 2014), 287. [cited at p. 16, 18]

[16] Baldin, V., Lukas, J., Marcote, M. J., Pagano, M., and Draetta, G. Cyclin D1 is a nuclear protein required for cell cycle progression in G1. *Genes & development 7*, 5 (1993), 812–821. [cited at p. 137]

[17] Bardwell, A. J., Bardwell, L., Tomkinson, A. E., and Friedberg, E. C. Specific Cleavage of Model Recombination and Repair Intermediates by the Yeast Rad1-Rad10 DNA Endonuclease. *Science (New York, N.Y.) 265*, 5181 (1994), 2082–2085. [cited at p. 16]

[18] Bärtsch, S., Dücker, K., Würgler, F. E., and Sengstag, C. Ectopic mitotic recombination in Drosophila probed with bacterial beta-galactosidase gene-based reporter transgenes. *Nucleic acids research 25*, 19 (Oct. 1997), 3917–24. [cited at p. 102]

[19] Baumann, P., and West, S. C. Role of the human RAD51 protein in homologous recombination and double-stranded-break repair. *Trends in biochemical sciences 23*, 7 (July 1998), 247–51. [cited at p. 27, 93, 132]

[20] Benjamini, Y., and Hochberg, Y. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B 57*, 1 (1995), 289–300. [cited at p. 41]

[21] Bennardo, N., Cheng, A., Huang, N., and Stark, J. M. Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. *PLoS genetics 4*, 6 (June 2008), e1000110. [cited at p. 101]

[22] Benson, F. E., Stasiak, A., and West, S. C. Purification and characterization of the human Rad51 protein, an analogue of E.coli RecA. *The EMBO journal 13*, 23 (1994), 5764–5771. [cited at p. 27, 93]

[23] Bentley, J., Diggle, C. P., Harnden, P., Knowles, M. A., and Kiltie, A. E. DNA double strand break repair in human bladder cancer is error prone and involves microhomology-associated end-joining. *Nucleic acids research 32*, 17 (Jan. 2004), 5249–59. [cited at p. 31, 79, 113, 154]

[24] Bertolini, L. R., Bertolini, M., Maga, E. A., Madden, K. R., and Murray, J. D. Increased gene targeting in Ku70 and Xrcc4 transiently deficient human somatic cells. *Molecular biotechnology 41*, 2 (Feb. 2009), 106–14. [cited at p. 156]

[25] Bétermier, M., Bertrand, P., and Lopez, B. S. Is non-homologous end-joining really an inherently error-prone process? *PLoS genetics 10*, 1 (Jan. 2014), e1004086. [cited at p. 61, 152]

[26] Bishop, D. K., Park, D., Xu, L., and Kleckner, N. DMCI: A Meiosis-Specific Yeast Homolog of E. coli recA Required for Recombination, Synaptonemal Complex Formation, and Cell Cycle Progression. *Cell 69*, May 1 (1992), 439–456. [cited at p. 29]

[27] Blunt, T., Finnie, N. J., Taccioli, G. E., Smith, G. C., Demengeot, J., Gottlieb, T. M., Mizuta, R., Varghese, A. J., Alt, F. W., Jeggo, P. A., and Jackson, S. P. Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA repair defects associated with the murine scid mutation. *Cell 80*, 5 (Mar. 1995), 813–23. [cited at p. 39, 42, 51]

[28] Boboila, C., Yan, C., Wesemann, D. R., Jankovic, M., Wang, J. H., Manis, J., Nussenzweig, A., Nussenzweig, M., and Alt, F. W. Alternative end-joining catalyzes class switch recombination in the absence of both Ku70 and DNA ligase 4. *The Journal of experimental medicine 207*, 2 (Feb. 2010), 417–27. [cited at p. 30, 31, 51, 55, 61, 101, 112]

[29] Bode, J., Kohwi, Y., Dickinson, L., Joh, T., Klehr, D., Mielke, C., and Kohwi-Shigematsu, T. Biological significance of unwinding capability of nuclear matrix-associating DNAs. *Science (New York, N.Y.) 255*, 5041 (Jan. 1992), 195–197. [cited at p. 10, 82, 153]

[30] Bode, J., and Maass, K. Chromatin domain surrounding the human interferon-beta gene as defined by scaffold-attached regions. *Biochemistry 27*, 13 (June 1988), 4706–11. [cited at p. 10]

[31] Bode, J., Winkelmann, S., Götze, S., Spiker, S., Tsutsui, K., Bi, C., A K, P., and Benham, C. Correlations between scaffold/matrix attachment region (S/MAR) binding activity

and DNA duplex destabilization energy. *Journal of molecular biology 358*, 2 (Apr. 2006), 597–613. [cited at p. 82, 153]

[32] Bollag, R., Waldman, A. S., and Liskay, R. M. Homologous recombination in mammalian cells. *Annual review of genetics 23* (1989), 199–225. [cited at p. 156]

[33] Branda, C. S., and Dymecki, S. M. Talking about a Revolution : The Impact of Site-Specific Recombinases on Genetic Analyses in Mice. *Developmental Cell 6*, 1 (2004), 7–28. [cited at p. 7]

[34] Brogna, S., and Wen, J. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nature structural & molecular biology 16*, 2 (2009), 107–113. [cited at p. 48]

[35] Buck, D., Malivert, L., de Chasseval, R., Barraud, A., Fondanèche, M.-C., Sanal, O., Plebani, A., Stéphan, J.-L., Hufnagel, M., le Deist, F., Fischer, A., Durandy, A., de Villartay, J.-P., and Revy, P. Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell 124*, 2 (Jan. 2006), 287–99. [cited at p. 25]

[36] Bunting, S. F., Callén, E., Wong, N., Chen, H.-T., Polato, F., Gunn, A., Bothmer, A., Feldhahn, N., Fernandez-Capetillo, O., Cao, L., Xu, X., Deng, C.-X., Finkel, T., Nussenzweig, M., Stark, J. M., and Nussenzweig, A. 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell 141*, 2 (2010), 243–254. [cited at p. 29]

[37] Bushman, F. D. Retroviral integration and human gene therapy. *The Journal of clinical investigation 117*, 8 (2007), 2083–2086. [cited at p. 4]

[38] Cai, S., Han, H.-J., and Kohwi-Shigematsu, T. Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nature genetics 34*, 1 (May 2003), 42–51. [cited at p. 11]

[39] Caldecott, K. W. Mammalian single-strand break repair: mechanisms and links with chromatin. *DNA repair 6*, 4 (Apr. 2007), 443–53. [cited at p. 22]

[40] Caldecott, K. W. Single-strand break repair and genetic disease. *Nature reviews. Genetics 9*, 8 (Aug. 2008), 619–31. [cited at p. 22]

[41] Caldecott, K. W. DNA single-strand break repair. *Experimental cell research 329*, 1 (Nov. 2014), 2–8. [cited at p. 21]

[42] Caldecott, K. W. Protein ADP-ribosylation and the cellular response to DNA strand breaks. *DNA repair 19* (July 2014), 108–13. [cited at p. 22]

[43] Cannavo, E., and Cejka, P. Sae2 promotes dsDNA endonuclease activity within Mre11–Rad50–Xrs2 to resect DNA breaks. *Nature 514*, 7520 (Sept. 2014), 122–125. [cited at p. 27]

[44] Cartwright, R., Dunn, A. M., Simpson, P. J., Tambini, C. E., and Thacker, J. Isolation of novel human and mouse genes of the recA/RAD51 recombination-repair gene family. *Nucleic acids research 26*, 7 (Apr. 1998), 1653–9. [cited at p. 93]

[45] Ceccaldi, R., Liu, J. C., Amunugama, R., Hajdu, I., Primack, B., Petalcorin, M. I. R., O'Connor, K. W., Konstantinopoulos, P. A., Elledge, S. J., Boulton, S. J., Yusufzai, T., and D'Andrea, A. D. Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. *Nature Feb 2* (2015), doi: 10.1038/nature14184. [cited at p. 21, 82, 115]

[46] Chaisson, M. J., and Tesler, G. Mapping single molecule sequencing reads using Basic Local Alignment with Successive Refinement (BLASR): Theory and Application. *BMC Bioinformatics 13*, 1 (2012), 238. [cited at p. 66]

[47] Chan, S. H., Yu, A. M., and McVey, M. Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila. *PLoS genetics 6*, 7 (July 2010), 1–16. [cited at p. 21, 82, 111, 115]

[48] Chapman, J. R., Taylor, M. R. G., and Boulton, S. J. Playing the end game: DNA double-strand break repair pathway choice. *Molecular cell 47*, 4 (Aug. 2012), 497–510. [cited at p. 29]

[49] Chayot, R., Montagne, B., and Ricchetti, M. DNA polymerase μ is a global player in the repair of non-homologous end-joining substrates. *DNA repair 11*, 1 (Jan. 2012), 22–34. [cited at p. 21]

[50] Chen, B. P. C., Chan, D. W., Kobayashi, J., Burma, S., Asaithamby, A., Morotomi-Yano, K., Botvinick, E., Qin, J., and Chen, D. J. Cell cycle dependence of DNA-dependent protein kinase phosphorylation in response to DNA double strand breaks. *Journal of Biological Chemistry 280*, 15 (2005), 14709–14715. [cited at p. 53]

[51] Ciccia, A., Ling, C., Coulthard, R., Yan, Z., Xue, Y., Meetei, A. R., Laghmani, E. H., Joenje, H., McDonald, N., de Winter, J. P., Wang, W., and West, S. C. Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. *Molecular cell 25*, 3 (Feb. 2007), 331–43. [cited at p. 18]

[52] Ciuffi, A. Mechanisms Governing Lentivirus Integration Site Selection. *Current Gene Therapy 8*, 6 (Dec. 2008), 419–429. [cited at p. 4]

[53] Clauson, C., Schärer, O. D., and Niedernhofer, L. Advances in understanding the complex mechanisms of DNA interstrand cross-link repair. *Cold Spring Harbor perspectives in biology 5*, 10 (Oct. 2013), a012732. [cited at p. 18]

[54] Cleaver, J. E., Lam, E. T., and Revet, I. Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. *Nature reviews. Genetics 10*, 11 (Nov. 2009), 756–68. [cited at p. 16]

[55] Clerici, M., Mantiero, D., Lucchini, G., and Longhese, M. P. The Saccharomyces cerevisiae Sae2 protein promotes resection and bridging of double strand break ends. *The Journal of biological chemistry 280*, 46 (Nov. 2005), 38631–8. [cited at p. 27]

[56] Cloud, V., Chan, Y.-L., Grubb, J., Budke, B., and Bishop, D. K. Rad51 is an accessory factor for Dmc1-mediated joint molecule formation during meiosis. *Science (New York, N.Y.) 337*, 6099 (Sept. 2012), 1222–5. [cited at p. 29]

[57] Cockerill, P. N., and Garrard, W. T. Chromosomal Loop Anchorage of the Kappa lmmunoglobulin a Region Containing Topoisomerase II Sites. *Cell 44*, January 31 (1986), 273–282. [cited at p. 10, 153]

[58] Cockerill, P. N., and Garrard, W. T. Chromosomal loop anchorage sites appear to be evolutionarily conserved. *FEBS letters 204*, 1 (Aug. 1986), 5–7. [cited at p. 10]

[59] Conde-Pérezprina, J. C., León-Galván, M. A., and Konigsberg, M. DNA mismatch repair system: repercussions in cellular homeostasis and relationship with aging. *Oxidative medicine and cellular longevity 2012* (Jan. 2012), 728430. [cited at p. 16]

[60] Costantino, L., Sotiriou, S. K., Rantala, J. K., Magin, S., Mladenov, E., Helleday, T., Haber, J. E., Iliakis, G., Kallioniemi, O. P., and Halazonetis, T. D. Break-induced replication repair of damaged forks induces genomic duplications in human cells. *Science (New York, N.Y.) 343*, 6166 (Jan. 2014), 88–91. [cited at p. 81, 94, 111, 153]

[61] Cousineau, I., Abaji, C., and Belmaaza, A. BRCA1 regulates RAD51 function in response to DNA damage and suppresses spontaneous sister chromatid replication slippage: implications for sister chromatid cohesion, genome stability, and carcinogenesis. *Cancer research 65*, 24 (Dec. 2005), 11384–91. [cited at p. 93]

[62] Crespan, E., Czabany, T., Maga, G., and Hübscher, U. Microhomology-mediated DNA strand annealing and elongation by human DNA polymerases λ and β on normal and repetitive DNA sequences. *Nucleic acids research 40*, 12 (July 2012), 5577–90. [cited at p. 94]

[63] Critchlow, S. E., Bowater, R. P., and Jackson, S. P. Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Current Biology 7*, 8 (Aug. 1997), 588–598. [cited at p. 25, 93]

[64] Daboussi, F., Dumay, A., Delacôte, F., and Lopez, B. S. DNA double-strand break repair signalling: the case of RAD51 post-translational regulation. *Cellular signalling 14*, 12 (Dec. 2002), 969–75. [cited at p. 14]

[65] d'Adda di Fagagna, F. Living on a break: cellular senescence as a DNA-damage response. *Nature reviews. Cancer 8*, 7 (July 2008), 512–22. [cited at p. 12]

[66] Dai, J., Cui, X., Zhu, Z., and Hu, W. Non-homologous end joining plays a key role in transgene concatemer formation in transgenic zebrafish embryos. *International journal of biological sciences 6*, 7 (Jan. 2010), 756–68. [cited at p. 62]

[67] D'Amours, D., and Jackson, S. P. The Mre11 complex: at the crossroads of DNA repair and checkpoint signalling. *Nature reviews. Molecular cell biology 3*, 5 (May 2002), 317–27. [cited at p. 26]

[68] Davies, A. A., Masson, J. Y., McIlwraith, M. J., Stasiak, A. Z., Stasiak, A., Venkitaraman, A. R., and West, S. C. Role of BRCA2 in control of the RAD51 recombination and DNA repair protein. *Molecular cell 7*, 2 (Feb. 2001), 273–82. [cited at p. 28, 93]

[69] De Silva, I. U., McHugh, P. J., Clingen, P. H., and Hartley, J. A. Defining the Roles of Nucleotide Excision Repair and Recombination in the Repair of DNA Interstrand Cross-Links in Mammalian Cells. *Molecular and cellular biology 20*, 21 (2000), 7980–7990. [cited at p. 19, 26]

[70] Deans, A. J., and West, S. C. DNA interstrand crosslink repair and cancer. *Nature reviews. Cancer 11*, 7 (July 2011), 467–80. [cited at p. 17]

[71] Debatisse, M., Toledo, F., and Anglana, M. Replication Initiation in Mammalian Cells: Changing Preferences. *Cell Cycle 3*, 1 (2004), 19–21. [cited at p. 11, 37, 84, 153]

[72] Decottignies, A. Microhomology-mediated end joining in fission yeast is repressed by pku70 and relies on genes involved in homologous recombination. *Genetics 176*, 3 (July 2007), 1403–15. [cited at p. 30, 51, 61, 71, 112]

[73] Della-Maria, J., Zhou, Y., Tsai, M.-S., Kuhnlein, J., Carney, J. P., Paull, T. T., and Tomkinson, A. E. Human Mre11/human Rad50/Nbs1 and DNA ligase IIIalpha/XRCC1 protein complexes act together in an alternative nonhomologous end joining pathway. *The Journal of biological chemistry 286*, 39 (Sept. 2011), 33845–53. [cited at p. 30, 61, 94, 101]

[74] Deng, C., Zhang, P., Harper, J. W., Elledge, S. J., and Leder, P. Mice lacking p21CIP1/WAF1 undergo normal development, but are defective in G1 checkpoint control. *Cell 82*, 4 (1995), 675–84. [cited at p. 13]

[75] Deng, S. K., Gibb, B., de Almeida, M. J., Greene, E. C., and Symington, L. S. RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. *Nature structural & molecular biology 21*, 4 (Apr. 2014), 405–12. [cited at p. 51, 71]

[76] Derouazi, M., Martinet, D., Besuchet Schmutz, N., Flaction, R., Wicht, M., Bertschinger, M., Hacker, D. L., Beckmann, J. S., and Wurm, F. M. Genetic characterization of CHO production host DG44 and derivative recombinant cell lines. *Biochemical and biophysical research communications 340*, 4 (Mar. 2006), 1069–77. [cited at p. 67]

[77] Diacumakos, E. G. Methods For Micromanipulation Of Human Somatic Cells In Culture. *Methods in Cell Biology 7* (1973), 287–311. [cited at p. 2]

[78] Dianov, G. L., and Hübscher, U. Mammalian base excision repair: the forgotten archangel. *Nucleic acids research 41*, 6 (Apr. 2013), 3483–90. [cited at p. 13]

[79] Dianov, G. L., and Parsons, J. L. Co-ordination of DNA single strand break repair. *DNA repair 6*, 4 (Apr. 2007), 454–60. [cited at p. 21]

[80] Difilippantonio, S., Gapud, E., Wong, N., Huang, C.-Y., Mahowald, G., Chen, H. T., Kruhlak, M. J., Livak, F., Nussenzweig, M., Sleckman, B. P., and Nussenzweig, A. 53BP1 facilitates long-range DNA end-joining during V(D)J recombination. *Nature 456*, 7221 (2008), 529–533. [cited at p. 25]

[81] Dillon, L. W., Burrow, A. A., and Wang, Y.-H. DNA instability at chromosomal fragile sites in cancer. *Current genomics 11*, 5 (2010), 326–337. [cited at p. 153]

[82] Dimitrova, N., and de Lange, T. Cell cycle-dependent role of MRN at dysfunctional telomeres: ATM signaling-dependent induction of nonhomologous end joining (NHEJ) in G1 and resection-mediated inhibition of NHEJ in G2. *Molecular and cellular biology 29*, 20 (Oct. 2009), 5552–63. [cited at p. 26]

[83] Dinkelmann, M., Spehalski, E., Stoneham, T., Buis, J., Wu, Y., Sekiguchi, J. M., and Ferguson, D. O. Multiple functions of MRN in end-joining pathways during isotype class switching. *Nature structural & molecular biology 16*, 8 (2009), 808–813. [cited at p. 30, 61, 94]

[84] Dobbs, T. A., Tainer, J. A., and Lees-Miller, S. P. A structural model for regulation of NHEJ by DNA-PKcs autophosphorylation. *DNA repair 9*, 12 (2010), 1307–1314. [cited at p. 93]

[85] Dresler, S. L., and Frattini, M. G. DNA replication and UV-induced DNA repair synthesis in human fibroblasts are much less sensitive than DNA polymerase alpha to inhibition by butylphenyl-deoxyguanosine triphosphate. *Nucleic acids research 14*, 17 (1986), 7093–7102. [cited at p. 81]

[86] Du, L.-Q., Wang, Y., Wang, H., Cao, J., Liu, Q., and Fan, F.-Y. Knockdown of Rad51 expression induces radiation- and chemo-sensitivity in osteosarcoma cells. *Medical oncology (Northwood, London, England) 28*, 4 (Dec. 2011), 1481–7. [cited at p. 132, 139]

[87] El-Khamisy, S. F., Masutani, M., Suzuki, H., and Caldecott, K. W. A requirement for PARP-1 for the assembly or stability of XRCC1 nuclear foci at sites of oxidative DNA damage. *Nucleic Acids Research 31*, 19 (Oct. 2003), 5526–5533. [cited at p. 22]

[88] Essers, J., Hendriks, R. W., Wesoly, J., Beerens, C. E. M. T., Smit, B., Hoeijmakers, J. H. J., Wyman, C., Dronkert, M. L. G., and Kanaar, R. Analysis of mouse Rad54 expression and its implications for homologous recombination. *DNA repair 1*, 10 (Oct. 2002), 779–93. [cited at p. 93]

[89] Evans, E., Moggs, J. G., Hwang, J. R., Egly, J.-M., and Wood, R. D. Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *The EMBO journal 16*, 21 (1997), 6559–6573. [cited at p. 16]

[90] Evans, K., Ott, S., Hansen, A., Koentges, G., and Wernisch, L. A comparative study of S/MAR prediction tools. *BMC bioinformatics 8* (Jan. 2007), 71. [cited at p. 10, 82, 153]

[91] Fan, W., and Wu, X. DNA polymerase lambda can elongate on DNA substrates mimicking non-homologous end joining and interact with XRCC4-ligase IV complex. *Biochemical and biophysical research communications 323*, 4 (Oct. 2004), 1328–33. [cited at p. 21]

[92] Feng, Z., Scott, S. P., Bussen, W., Sharma, G. G., Guo, G., Pandita, T. K., and Powell, S. N. Rad52 inactivation is synthetically lethal with BRCA2 deficiency. *Proceedings of the National Academy of Sciences of the United States of America 108*, 2 (Jan. 2011), 686–91. [cited at p. 29, 93, 132]

[93] Fernandez-Capetillo, O., Lee, A., Nussenzweig, M., and Nussenzweig, A. H2AX: the histone guardian of the genome. *DNA repair 3*, 8-9 (2004), 959–67. [cited at p. 13]

[94] Folger, K. R., Wong, E. A., Wahl, G., and Capecchi, M. R. Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. *Molecular and cellular biology 2*, 11 (Nov. 1982), 1372–87. [cited at p. 62, 151]

[95] Fousteri, M., Vermeulen, W., van Zeeland, A. A., and Mullenders, L. H. F. Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo. *Molecular cell 23*, 4 (Aug. 2006), 471–82. [cited at p. 16]

[96] Fraley, R., Subramani, S., Berg, P., and Papahadjopoulos, D. Introduction of Liposome-encapsulated SV40 DNA into Cells. *The Journal of biological chemistry 255*, 21 (1980), 10431–10435. [cited at p. 2]

[97] Friedman, L., Ostermeyer, E., Szabo, C., Dowd, P., Lynch, E., Rowell, S., and King, M. Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families. *Nature genetics 8*, 4 (1994), 399–404. [cited at p. 28]

[98] Gaj, T., Gersbach, C. A., and Barbas, C. F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology 31*, 7 (July 2013), 397–405. [cited at p. 8, 9]

[99] Gaj, T., Sirk, S. J., and Barbas, C. F. Expanding the scope of site-specific recombinases for genetic and metabolic engineering. *Biotechnology and bioengineering 111*, 1 (Jan. 2014), 1–15. [cited at p. 7]

[100] Galande, S., and Kohwi-Shigematsu, T. Caught in the Act: Binding of Ku and PARP to MARs Reveals Novel Aspects of Their Function Interaction. *Critical Reviews^{TM} in Eukaryotic Gene Expression 10*, 1 (2000), 63–72. [cited at p. 154]

[101] Galbete, J. L., Buceta, M., and Mermod, N. MAR elements regulate the probability of epigenetic switching between active and inactive gene expression. *Molecular bioSystems 5*, 2 (Feb. 2009), 143–50. [cited at p. 11, 62, 79]

[102] Gama-Norton, L., Riemer, P., Sandhu, U., Nehlsen, K., Schucht, R., Hauser, H., and Wirth, D. Defeating Randomness - Targeted Integration as a Boost for Biotechnology. *Cell Line Development 6* (2009), 53–82. [cited at p. 8]

[103] Garcia, V., Phelps, S. E. L., Gray, S., and Neale, M. J. Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. *Nature 479*, 7372 (Nov. 2011), 241–4. [cited at p. 27]

[104] Garcia-Higuera, I., Taniguchi, T., Ganesan, S., Meyn, M., Timmers, C., Hejna, J., Grompe, M., and D'Andrea, A. D. Interaction of the Fanconi Anemia Proteins and BRCA1 in a Common Pathway. *Molecular Cell 7*, 2 (Feb. 2001), 249–262. [cited at p. 19]

[105] Gasser, S. M., and Laemmli, U. K. The organisation of chromatin loops: characterization of a scaffold attachment site. *The EMBO journal 5*, 3 (1986), 511–518. [cited at p. 10, 11]

[106] Gerasimova, T. I., and Corces, V. G. Chromatin insulators and boundaries: effects on transcription and nuclear organization. *Annual review of genetics 35*, 30 (Jan. 2001), 193–208. [cited at p. 11, 37]

[107] Ghosal, G., and Chen, J. DNA damage tolerance: a double-edged sword guarding the genome. *Translational cancer research 2*, 3 (Jan. 2013), 107–129. [cited at p. 20]

[108] Gigi, V., Lewis, S., Shestova, O., Mijušković, M., Deriano, L., Meng, W., Luning Prak, E. T., and Roth, D. B. RAG2 mutants alter DSB repair pathway choice in vivo and illuminate the nature of 'alternative NHEJ'. *Nucleic acids research 42*, 10 (Jan. 2014), 6352–64. [cited at p. 30, 31, 61, 79, 101]

[109] Gillet, L. C. J., and Schärer, O. D. Molecular Mechanisms of Mammalian Global Genome Nucleotide Excision Repair. *Chemical reviews 106*, 2 (2006), 253–76. [cited at p. 15]

[110] Girod, P.-A., Nguyen, D.-Q., Calabrese, D., Puttini, S., Grandjean, M., Martinet, D., Regamey, A., Saugy, D., Beckmann, J. S., Bucher, P., and Mermod, N. Genome-wide prediction of matrix attachment regions that increase gene expression in mammalian cells. *Nature Methods 4*, 9 (Sept. 2007), 747–753. [cited at p. 10, 11, 33, 37, 38, 53, 62, 73]

[111] Girod, P.-A., Zahn-Zabal, M., and Mermod, N. Use of the chicken lysozyme 5' matrix attachment region to generate high producer CHO cell lines. *Biotechnology and bioengineering 91*, 1 (July 2005), 1–11. [cited at p. 11, 33, 37]

[112] Goetz, J. D.-M., Motycka, T. A., Han, M., Jasin, M., and Tomkinson, A. E. Reduced repair of DNA double-strand breaks by homologous recombination in a DNA ligase I-deficient human cell line. *DNA repair 4*, 6 (June 2005), 649–54. [cited at p. 27]

[113] Gorman, S. O., Fox, D. T., and Wahl, G. M. Recombinase-Mediated Gene Activation and Site-Specific Integration in Mammalian Cells. *Science (New York, N.Y.) 251*, 4999 (1991), 1351–1355. [cited at p. 7]

[114] Grabundzija, I., Irgang, M., Mátés, L., Belay, E., Matrai, J., Gogol-Döring, A., Kawakami, K., Chen, W., Ruiz, P., Chuah, M. K. L., VandenDriessche, T., Izsvák, Z., and Ivics, Z. Comparative analysis of transposable element vector systems in human cells. *Molecular therapy : the journal of the American Society of Gene Therapy 18*, 6 (June 2010), 1200–9. [cited at p. 4]

[115] Graham, F. L., and van der Eb, A. J. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology 52* (1973), 456–467. [cited at p. 2]

[116] Grandjean, M., Girod, P.-A., Calabrese, D., Kostyrko, K., Wicht, M., Yerly, F., Mazza, C., Beckmann, J. S., Martinet, D., and Mermod, N. High-level transgene expression by homologous recombination-mediated gene transfer. *Nucleic acids research 39*, 15 (Aug. 2011), e104. [cited at p. 6, 11, 33, 44, 62, 63, 74, 103, 140]

[117] Grawunder, U., Zimmer, D., and Lieber, M. R. DNA ligase IV binds to XRCC4 via a motif located between rather than within its BRCT domains. *Current Biology 8*, 15 (July 1998), 873–879. [cited at p. 25]

[118] Grillari, J., Katinger, H., and Voglauer, R. Contributions of DNA interstrand cross-links to aging of cells and organisms. *Nucleic Acids Research 35*, 22 (2007), 7566–7576. [cited at p. 51]

[119] Grosjean, F., Batard, P., Jordan, M., and Wurm, F. M. S-phase synchronized CHO cells show elevated transfection efficiency and expression using CaPi. *Cytotechnology 38*, 1-3 (2002), 57–62. [cited at p. 136, 140]

[120] Guirouilh-Barbat, J., Lambert, S., Bertrand, P., and Lopez, B. S. Is homologous recombination really an error-free process? *Frontiers in genetics 5*, June (Jan. 2014), 175. [cited at p. 61]

[121] Haber, J. E. Partners and pathways - repairing a double-strand break. *Trends in genetics : TIG 16*, 6 (2000), 259–264. [cited at p. 22, 61]

[122] Hamilton, A. A., and Thacker, J. Gene recombination in X-ray-sensitive hamster cells. *Molecular and cellular biology 7*, 4 (Apr. 1987), 1409–14. [cited at p. 102]

[123] Hanawalt, P. C., and Spivak, G. Transcription-coupled DNA repair: two decades of progress and surprises. *Nature reviews. Molecular cell biology 9*, 12 (Dec. 2008), 958–70. [cited at p. 15]

[124] Handler, A. M., McCombs, S. D., Fraser, M. J., and Saul, S. H. The lepidopteran transposon vector, piggyBac, mediates germ-line transformation in the Mediterranean fruit fly. *Proceedings of the National Academy of Sciences 95*, June (1998), 7520–7525. [cited at p. 5]

[125] Harraghy, N., Gaussin, A., and Mermod, N. Sustained Transgene Expression Using MAR Elements. *Current Gene Therapy 8*, 5 (Oct. 2008), 353–366. [cited at p. 11, 37, 62]

[126] Heyer, W.-D., Li, X., Rolfsmeier, M., and Zhang, X.-P. Rad54: the Swiss Army knife of homologous recombination? *Nucleic acids research 34*, 15 (Jan. 2006), 4115–25. [cited at p. 28, 93]

[127] Hinnen, A., Hicks, J. B., and Fink, G. R. Transformation of yeast. *Proceedings of the National Academy of Sciences of the United States of America 75*, 4 (1978), 1929–1933. [cited at p. 151]

[128] Hinz, J. M., Tebbs, R. S., Wilson, P. F., Nham, P. B., Salazar, E. P., Nagasawa, H., Urbin, S. S., Bedford, J. S., and Thompson, L. H. Repression of mutagenesis by Rad51D-mediated homologous recombination. *Nucleic acids research 34*, 5 (Jan. 2006), 1358–68. [cited at p. 39, 42, 50]

[129] Hiom, K. Coping with DNA double strand breaks. *DNA repair 9*, 12 (Dec. 2010), 1256–63. [cited at p. 12, 22, 131]

[130] Ho, T. V., and Schärer, O. D. Translesion DNA Synthesis Polymerases in DNA Interstrand Crosslink Repair. *Environmental and Molecular Mutagenesis 51*, 6 (2010), 552–566. [cited at p. 19, 21]

[131] Homberger, H. Bent DNA is a structural feature of scaffold-attached regions in Drosophila melanogaster interphase nuclei. *Chromosoma 98*, 2 (Jan. 1989), 99–104. [cited at p. 10]

[132] Hosoya, N., and Miyagawa, K. Targeting DNA damage response in cancer therapy. *Cancer science 105*, 4 (Apr. 2014), 370–88. [cited at p. 12, 13]

[133] Hsu, L.-C., Huang, X., Seasholtz, S., Potter, D. M., and Gollin, S. M. Gene amplification and overexpression of protein phosphatase 1alpha in oral squamous cell carcinoma cell lines. *Oncogene 25*, 40 (2006), 5517–5526. [cited at p. 140]

[134] Hu, Z.-P., Chen, L.-S., Jia, C.-Y., Zhu, H.-Z., Wang, W., and Zhong, J. Screening of potential pseudo att sites of Streptomyces phage ΦC31 integrase in the human genome. *Acta pharmacologica Sinica 34*, 4 (Apr. 2013), 561–9. [cited at p. 8]

[135] Iaccarino, I., Palombo, F., Drummond, J., Totty, N. F., Hsuan, J., Modrich, P., and Jiricny, J. MSH6, a Saccharomyces cerevisiae protein that binds to mismatches as a heterodimer with MSH2. *Current Biology 6*, 4 (Apr. 1996), 484–486. [cited at p. 17]

[136] Iiizumi, S., Kurosawa, A., So, S., Ishii, Y., Chikaraishi, Y., Ishii, A., Koyama, H., and Adachi, N. Impact of non-homologous end-joining deficiency on random and targeted DNA integration: implications for gene targeting. *Nucleic acids research 36*, 19 (Nov. 2008), 6333–42. [cited at p. 62, 152, 156]

[137] Iliakis, G. Backup pathways of NHEJ in cells of higher eukaryotes: cell cycle dependence. *Radiotherapy and Oncology 92*, 3 (Sept. 2009), 310–5. [cited at p. 132]

[138] Iliakis, G., Wang, H., Perrault, A. R., Boecker, W., Rosidi, B., Windhofer, F., Wu, W., Guan, J., Terzoudi, G., and Pantelias, G. Mechanisms of DNA double strand break repair and chromosome aberration formation. *Cytogenetic and genome research 104*, 1-4 (Jan. 2004), 14–20. [cited at p. 79, 152]

[139] Ivics, Z., Hackett, P. B., Plasterk, R. H., and Izsva, Z. Molecular Reconstruction of Sleeping Beauty, a Tc1-like Transposon from Fish, and Its Transposition in Human Cells. *Cell 91*, 4 (1997), 501–510. [cited at p. 5]

[140] Ivics, Z., Li, M. A., Mátés, L., Boeke, J. D., Bradley, A., and Izsvák, Z. Transposon-mediated Genome Manipulations in Vertebrates. *Nature Methods 6*, 6 (2009), 415–422. [cited at p. 4, 5]

[141] Iwabuchi, K., Bartel, P. L., Li, B., Marraccino, R., and Fields, S. Two cellular proteins that bind to wild-type but not mutant p53. *Proceedings of the National Academy of Sciences 91*, June (1994), 6098–6102. [cited at p. 25]

[142] Jachymczyk, W. J., von Borstel, R. C., Mowat, M. R. A., and Hastings, P. J. Repair of interstrand cross-links in DNA of Saccharomyces cerevisiae requires two systems for DNA repair: The RAD3 system and the RAD51 system. *Molecular and General Genetics MGG 182*, 2 (July 1981), 196–205. [cited at p. 19]

[143] Jackson, J. A., Trevino, A. V., Herzig, M. C., Herman, T. S., and Woynarowski, J. M. Matrix attachment region (MAR) properties and abnormal expansion of AT island minisatellites in FRA16B fragile sites in leukemic CEM cells. *Nucleic Acids Research 31*, 21 (Nov. 2003), 6354–6364. [cited at p. 84, 153]

[144] Jackson, S. P. Sensing and repairing DNA double-strand breaks. *Carcinogenesis 23*, 5 (May 2002), 687–96. [cited at p. 22, 61]

[145] Jackson, S. P., and Bartek, J. The DNA-damage response in human biology and disease. *Nature 461*, 7267 (2009), 1071–1078. [cited at p. 12, 13, 131]

[146] Jasin, M. Homologous repair of DNA damage and tumorigenesis: the BRCA connection. *Oncogene 21*, 58 (Dec. 2002), 8981–93. [cited at p. 28, 132]

[147] Jeggo, P. A. DNA-PK: at the cross-roads of biochemistry and genetics. *Mutation Research/DNA Repair 384*, 1 (June 1997), 1–14. [cited at p. 24]

[148] Jensen, R. B. BRCA2: One Small Step for DNA Repair, One Giant Protein Purified. *Yale Journal of Biology and Medicine 86* (2013), 479–489. [cited at p. 28]

[149] Jiricny, J. Postreplicative mismatch repair. *Cold Spring Harbor perspectives in biology 5*, 4 (Apr. 2013), a012633. [cited at p. 16]

[150] Johnson, R. D., and Jasin, M. Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *The EMBO journal 19*, 13 (July 2000), 3398–407. [cited at p. 27, 47]

[151] Joukov, V., Chen, J., Fox, E. A., Green, J. B., and Livingston, D. M. Functional communication between endogenous BRCA1 and its partner, BARD1, during Xenopus laevis development. *Proceedings of the National Academy of Sciences of the United States of America 98*, 21 (Oct. 2001), 12078–83. [cited at p. 29]

[152] Kabotyanski, E. B., Gomelsky, L., Han, J.-O., Stamato, T. D., and Roth, D. B. Double-strand break repair in Ku86- and XRCC4-deficient cells. *Nucleic acids research 26*, 23 (1998), 27–31. [cited at p. 51, 55, 112, 132]

[153] Kamileri, I., Karakasilioti, I., and Garinis, G. A. Nucleotide excision repair: new tricks with old bricks. *Trends in genetics : TIG 28*, 11 (Nov. 2012), 566–73. [cited at p. 15]

[154] Kampinga, H. H., Hiemstra, Y. S., Konings, A. W., and Dikomey, E. Correlation between slowly repairable double-strand breaks and thermal radiosensitization in the human HeLa S3 cell line. *International journal of radiation biology 72*, 3 (1997), 293–301. [cited at p. 145]

[155] Kanaar, R., Hoeijmakers, J. H., and van Gent, D. C. Molecular mechanisms of DNA double strand break repair. *Trends in cell biology 8*, 12 (Dec. 1998), 483–9. [cited at p. 22, 61]

[156] Kato, S., Anderson, R. A., and Camerini-Otero, R. D. Foreign DNA introduced by calcium phosphate is integrated into repetitive DNA elements of the mouse L cell genome. *Molecular and cellular biology 6*, 5 (1986), 1787–1795. [cited at p. 151]

[157] Kawai, S., Hashimoto, W., and Murata, K. Transformation of Saccharomyces cerevisiae and other fungi: methods and possible underlying mechanism. *Bioengineered bugs 1*, 6 (2010), 395–403. [cited at p. 2]

[158] Keeney, S., Giroux, C. N., and Kleckner, N. Meiosis-Specific DNA Double-Strand Breaks Are Catalyzed by Spo11, a Member of a Widely Conserved Protein Family. *Cell 88*, 3 (Feb. 1997), 375–384. [cited at p. 29]

[159] Kelley, M. R., Kow, Y. W., and Wilson III, D. M. Disparity between DNA Base Excision Repair in Yeast and Mammals: Translational Implications. *Cancer research 63*, 3 (2003), 549–554. [cited at p. 15]

[160] Kent, T., Chandramouly, G., Mcdevitt, S. M., Ozdemir, A. Y., and Pomerantz, R. T. Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase θ. *Nature Structural & Molecular Biology Feb 2* (2015), doi: 10.1038/nsmb.2961. [cited at p. 21, 94, 115]

[161] Khan, K. Gene Transfer Technologies and their Applications : Roles in Human Diseases. *Asian Journal of Experimental Biological Sciences 1*, 1 (2010), 208–218. [cited at p. 2]

[162] Kim, H., and D'Andrea, A. D. Regulation of DNA cross-link repair by the Fanconi anemia / BRCA pathway. *Genes & development 26*, 13 (2012), 1393–1408. [cited at p. 18]

[163] Kim, J.-S. J.-M., Park, D.-H., Kang, H. S., Yoon, J., Baek, K., and Yoon, Y. Improved recombinant gene expression in CHO cells using matrix attachment regions. *Journal of Biotechnology 107*, 2 (Jan. 2004), 95–105. [cited at p. 37, 62]

[164] Kim, T. K., and Eberwine, J. H. Mammalian cell transfection: the present and the future. *Analytical and bioanalytical chemistry 397*, 8 (Aug. 2010), 3173–8. [cited at p. 2]

[165] Kim, Y., Lach, F. P., Desetty, R., Hanenberg, H., Auerbach, A. D., and Smogorzewska, A. Mutations of the SLX4 gene in Fanconi anemia. *Nature genetics 43*, 2 (Mar. 2011), 142–6. [cited at p. 19]

[166] Kim, Y.-J., and Wilson III, D. M. Overview of Base Excision Repair Biochemistry. *Current Molecular Pharmacology 5*, 1 (2012), 3–13. [cited at p. 13, 14]

[167] Kohli, A., Leech, M., Vain, P., Laurie, D. A., Christou, P., and Brill, W. J. Transgene organization in rice engineered through direct DNA transfer supports a two-phase integration mechanism mediated by the establishment of integration hot spots. *Proceedings of the National Academy of Sciences of the United States of America 95*, June (1998), 7203–7208. [cited at p. 6, 62, 151]

[168] Koole, W., van Schendel, R., Karambelas, A. E., van Heteren, J. T., Okihara, K. L., and Tijsterman, M. A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nature communications 5* (Jan. 2014), 3216. [cited at p. 94, 115]

[169] Kraakman-van der Zwet, M., Overkamp, W. J. I., van Lange, R. E. E., Essers, J., van Duijn-Goedhart, A., Wiggers, I., Swaminathan, S., van Buul, P. P. W., Errami, A., Tan, R. T. L., Jaspers, N. G. J., Sharan, S. K., Kanaar, R., and Zdzienicka, M. Z. Brca2 (XRCC11) Deficiency Results in Radioresistant DNA Synthesis and a Higher Frequency of Spontaneous Deletions. *Molecular and cellular biology 22*, 2 (2002), 669–679. [cited at p. 39, 45]

[170] Krokan, H. E., and Bjørås, M. Base excision repair. *Cold Spring Harbor perspectives in biology 5*, 4 (Apr. 2013), a012583. [cited at p. 13]

[171] Kucherlapati, R. S., Eves, E. M., Song, K. Y., Morse, B. S., and Smithies, O. Homologous recombination between plasmids in mammalian cells can be enhanced by treatment of input DNA. *Proceedings of the National Academy of Sciences of the United States of America 81*, 10 (May 1984), 3153–7. [cited at p. 102]

[172] Kurimasa, A., Kumano, S., Boubnov, N. V., Story, M. D., Tung, C. S., Peterson, S. R., and Chen, D. J. Requirement for the kinase activity of human DNA-dependent protein kinase catalytic subunit in DNA strand break rejoining. *Molecular and cellular biology 19*, 5 (May 1999), 3877–84. [cited at p. 24]

[173] Kwaks, T. H. J., and Otte, A. P. Employing epigenetics to augment the expression of therapeutic proteins in mammalian cells. *Trends in biotechnology 24*, 3 (Mar. 2006), 137–42. [cited at p. 10]

[174] Lamarche, B. J., Orazio, N. I., and Weitzman, M. D. The MRN complex in double-strand break repair and telomere maintenance. *FEBS letters 584*, 17 (Sept. 2010), 3682–95. [cited at p. 26, 132]

[175] Lancaster, J., Wooster, R., Mangion, J., Phelan, C. M., Cochran, C., Gumbs, C., Seal, S., Barfoot, R., Collins, N., Bignell, G., Patel, S., Hamoudi, R., Larsson, C., Wiseman, R., Berchuck, A., Iglehart, J., Marks, J., Ashworth, A., Stratton, M., and Futreal, P. BRCA2 mutations in primary breast and ovarian cancers. *Nature genetics 13*, 2 (1996), 238–240. [cited at p. 28]

[176] Langerak, P., and Russell, P. Regulatory networks integrating cell cycle control with DNA damage checkpoints and double-strand break repair. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences 366*, 1584 (Dec. 2011), 3562–3571. [cited at p. 133]

[177] Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology 10*, 3 (2009), R25. [cited at p. 41]

[178] Lapenna, S., and Giordano, A. Cell cycle kinases as therapeutic targets for cancer. *Nature reviews. Drug discovery 8*, 7 (2009), 547–566. [cited at p. 131]

[179] Lawley, P. D., and Phillips, D. H. DNA adducts from chemotherapeutic agents. *Mutation research 355*, 1-2 (1996), 13–40. [cited at p. 51]

[180] Le Cigne, A., Menil-Philippot, V., Fleury, F., Takahashi, M., and Thiriet, C. Transient expression of RAD51 in the late G2-phase is required for cell cycle progression in synchronous Physarum cells. *Genes to Cells 19* (2014), 755–765. [cited at p. 132]

[181] Le Fourn, V., Girod, P.-A., Buceta, M., Regamey, A., and Mermod, N. CHO cell engineering to prevent polypeptide aggregation and improve therapeutic protein secretion. *Metabolic engineering 21*, January (Jan. 2014), 91–102. [cited at p. 66]

[182] Lee, J. A., Carvalho, C. M. B., and Lupski, J. R. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell 131*, 7 (Dec. 2007), 1235–47. [cited at p. 153]

[183] Lee, J.-H., and Paull, T. T. ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science (New York, N.Y.) 308*, 5721 (Apr. 2005), 551–4. [cited at p. 13, 131]

[184] Lee, K., and Lee, S. E. Saccharomyces cerevisiae Sae2- and Tel1-dependent single-strand DNA formation at DNA break promotes microhomology-mediated end joining. *Genetics 176*, 4 (Aug. 2007), 2003–14. [cited at p. 30, 31, 61, 94, 111, 116]

[185] Li, Z., Otevrel, T., Gao, Y., Cheng, H.-L., Seed, B., Stamato, T. D., Taccioli, G. E., and Alt, F. W. The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell 83*, 7 (Dec. 1995), 1079–1089. [cited at p. 93]

[186] Liang, F., Romanienko, P. J., Weaver, D. T., Jeggo, P. A., and Jasin, M. Chromosomal double-strand break repair in Ku8O-deficient cells. *Proceedings of the National Academy of Sciences of the United States of America 93*, 17 (1996), 8929–8933. [cited at p. 51, 55, 112]

[187] Liang, L., Deng, L., Nguyen, S. C., Zhao, X., Maulion, C. D., Shao, C., and Tischfield, J. A. Human DNA ligases I and III, but not ligase IV, are required for microhomology-mediated end joining of DNA double-strand breaks. *Nucleic acids research 36*, 10 (June 2008), 3297–310. [cited at p. 94, 111]

[188] Lichtman, M. A., Spivak, J. L., Boxer, L. A., Shattil, S. J., and Henderson, E. S. *Commentary on and reprint of Fanconi G, Familiäre infantile perniziosaartige Anämie (perniziöses Blutbild und Konstitution) [Familial infantile pernicious-like anemia (pernicious blood picture and constitution)], in Jahrbuch für Kinderheilkunde (1927) 117.* Elsevier, 2000. [cited at p. 18]

[189] Lin, J. J., and Dutta, A. ATR pathway is the primary pathway for activating G2/M checkpoint induction after re-replication. *Journal of Biological Chemistry 282*, 42 (2007), 30357–30362. [cited at p. 144]

[190] Lin, Z., Kong, H., Nei, M., and Ma, H. Origins and evolution of the recA/RAD51 gene family: Evidence for ancient gene duplication and endosymbiotic gene transfer. *PNAS 103*, 27 (2006), 10328–10333. [cited at p. 28, 29]

[191] Lindahl, T. An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues. *Proceedings of the National Academy of Sciences of the United States of America 71*, 9 (1974), 3649–3653. [cited at p. 14]

[192] Lio, Y.-C., Schild, D., Brenneman, M. A., Redpath, J. L., and Chen, D. J. Human Rad51C deficiency destabilizes XRCC3, impairs recombination, and radiosensitizes S/G2-phase cells. *The Journal of biological chemistry 279*, 40 (Oct. 2004), 42313–20. [cited at p. 93]

[193] Liou, G.-Y., and Peter, S. Reactive oxygen species in cancer. *Free Radical Research 44*, 5 (2010), 479–496. [cited at p. 113]

[194] Liu, J., Doty, T., Gibson, B., and Heyer, W.-D. Human BRCA2 protein promotes RAD51 filament formation on RPA-covered ssDNA. *Nature structural & molecular biology 17*, 10 (2010), 1260–1262. [cited at p. 28, 93]

[195] Liu, X., Wang, Y., Benaissa, S., Matsuda, A., Kantarjian, H., Estrov, Z., and Plunkett, W. Homologous recombination as a resistance mechanism to replication-induced double-strand breaks caused by the antileukemia agent CNDAC. *Blood 116*, 10 (2010), 1737–1746. [cited at p. 54]

[196] Liu, Y., Masson, J.-Y., Shah, R., O'Regan, P., and West, S. C. RAD51C is required for Holliday junction processing in mammalian cells. *Science (New York, N.Y.) 303*, 5655 (Jan. 2004), 243–6. [cited at p. 93]

[197] Liu, Y., Tarsounas, M., O'Regan, P., and West, S. C. Role of RAD51C and XRCC3 in genetic recombination and DNA repair. *The Journal of biological chemistry 282*, 3 (Jan. 2007), 1973–9. [cited at p. 28, 132]

[198] Löbrich, M., and Jeggo, P. A. The impact of a negligent G2/M checkpoint on genomic instability and cancer induction. *Nature reviews. Cancer 7*, 11 (2007), 861–869. [cited at p. 131]

[199] Longley, M. J., Pierce, A. J., and Modrich, P. DNA Polymerase delta Is Required for Human Mismatch Repair in Vitro. *The Journal of biological chemistry 272*, 16 (1997), 10917–10921. [cited at p. 17]

[200] Lord, C. J., and Ashworth, A. The DNA damage response and cancer therapy. *Nature 481*, 7381 (Jan. 2012), 287–94. [cited at p. 12, 131]

[201] Lou, Z., Chen, B. P.-C., Asaithamby, A., Minter-Dykhouse, K., Chen, D. J., and Chen, J. MDC1 regulates DNA-PK autophosphorylation in response to DNA damage. *The Journal of biological chemistry 279*, 45 (Nov. 2004), 46359–62. [cited at p. 62]

[202] Lou, Z., Minter-Dykhouse, K., Franco, S., Gostissa, M., Rivera, M. A., Celeste, A., Manis, J. P., van Deursen, J., Nussenzweig, A., Paull, T. T., Alt, F. W., and Chen, J. MDC1 maintains genomic stability by participating in the amplification of ATM-dependent DNA damage signals. *Molecular cell 21*, 2 (Jan. 2006), 187–200. [cited at p. 13, 93, 139]

[203] Lundberg, R., Mavinakere, M., and Campbell, C. Deficient DNA end joining activity in extracts from fanconi anemia fibroblasts. *The Journal of biological chemistry 276*, 12 (Mar. 2001), 9543–9. [cited at p. 62, 152]

[204] Ma, J.-L., Kim, E. M., Haber, J. E., and Lee, S. E. Yeast Mre11 and Rad1 Proteins Define a Ku-Independent Mechanism To Repair Double-Strand Breaks Lacking Overlapping End Sequences. *Molecular and cellular biology 23*, 23 (2003), 8820–8828. [cited at p. 30, 31, 61, 71, 94]

[205] Ma, Y., Pannicke, U., Schwarz, K., and Lieber, M. R. Hairpin Opening and Overhang Processing by an Artemis/DNA-Dependent Protein Kinase Complex in Nonhomologous End Joining and V(D)J Recombination. *Cell 108*, 6 (Mar. 2002), 781–794. [cited at p. 24, 25]

[206] Maacke, H., Jost, K., Opitz, S., Miska, S., Yuan, Y., Hasselbach, L., Lüttges, J., Kalthoff, H., and Stürzbecher, H. W. DNA repair and recombination factor Rad51 is over-expressed in human pancreatic adenocarcinoma. *Oncogene 19*, 23 (2000), 2791–2795. [cited at p. 154]

[207] Mackay, C., Déclais, A.-C., Lundin, C., Agostinho, A., Deans, A. J., MacArtney, T. J., Hofmann, K., Gartner, A., West, S. C., Helleday, T., Lilley, D. M. J., and Rouse, J. Identification of KIAA1018 / FAN1 , a DNA Repair Nuclease Recruited to DNA Damage by Monoubiquitinated FANCD2. *Cell 142*, 1 (2010), 65–76. [cited at p. 19]

[208] Mailand, N., Falck, J., Lukas, C., Syljuåsen, R. G., Welcker, M., Bartek, J., and Lukas, J. Rapid Destruction of Human Cdc25A in Response to DNA Damage. *Science (New York, N.Y.) 288*, 5470 (2000), 1425–1429. [cited at p. 12]

[209] Majocchi, S., Aritonovska, E., and Mermod, N. Epigenetic regulatory elements associate with specific histone modifications to prevent silencing of telomeric genes. *Nucleic acids research 42*, 1 (Jan. 2014), 193–204. [cited at p. 10, 11, 62, 73]

[210] Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. RNA-guided human genome engineering via Cas9. *Science (New York, N.Y.) 339*, 6121 (Feb. 2013), 823–6. [cited at p. 9]

[211] Maloisel, L., Fabre, F., and Gangloff, S. DNA polymerase delta is preferentially recruited during homologous recombination to promote heteroduplex DNA extension. *Molecular and cellular biology 28*, 4 (Feb. 2008), 1373–82. [cited at p. 27]

[212] Mao, Z., Bozzella, M., Seluanov, A., and Gorbunova, V. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA repair 7*, 10 (Oct. 2008), 1765–71. [cited at p. 39, 46, 51, 56, 63, 68, 101, 102]

[213] Marconi, P., Argnani, R., Epstein, A. L., and Manservigi, R. HSV as a vector in vaccine development and gene therapy. *Human Vaccines 4*, 2 (Oct. 2014), 91–105. [cited at p. 4]

[214] Markowitz, D., Goff, S., and Bank, A. A Safe Packaging Line for Gene Transfer : Separating Viral Genes on Two Different Plasmids. *Journal of virology 62*, 4 (1988), 1120–1124. [cited at p. 3]

[215] Marsden, M. P. F., and Laemmli, U. K. Metaphase Chromosome Structure: Evidence for a Radial Loop Model. *Cell 17*, August (1979), 849–858. [cited at p. 11, 37]

[216] Marteijn, J. A., Lans, H., Vermeulen, W., and Hoeijmakers, J. H. J. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nature reviews. Molecular cell biology 15*, 7 (June 2014), 465–81. [cited at p. 15]

[217] Masson, J. Y., Tarsounas, M. C., Stasiak, A. Z., Stasiak, A., Shah, R., McIlwraith, M. J., Benson, F. E., and West, S. C. Identification and purification of two distinct complexes containing the five RAD51 paralogs. *Genes & development 15*, 24 (Dec. 2001), 3296–307. [cited at p. 28, 93, 132]

[218] Mateos-Gomez, P. A., Gong, F., Nair, N., Miller, K. M., Lazzerini-Denchi, E., and Sfeir, A. Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature Feb 2* (2015), doi: 10.1038/nature14157. [cited at p. 21, 115]

[219] Matsuda, M., Miyagawa, K., Takahashi, M., Fukuda, T., Kataoka, T., Asahara, T., Inui, H., Watatani, M., Yasutomi, M., Kamada, N., Dohi, K., and Kamiya, K. Mutations in the RAD54 recombination gene in primary cancers. *Oncogene 18*, 22 (1999), 3427–3430. [cited at p. 28]

[220] Mauldin, S. K., Getts, R. C., Liu, W., and Stamato, T. D. DNA-PK-dependent binding of DNA ends to plasmids containing nuclear matrix attachment region DNA sequences: evidence for assembly of a repair complex. *Nucleic acids research 30*, 18 (Sept. 2002), 4075–87. [cited at p. 154]

[221] McVey, M., and Lee, S. E. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends in genetics : TIG 24*, 11 (Nov. 2008), 529–38. [cited at p. 30, 53]

[222] Meetei, A. R., de Winter, J. P., Medhurst, A. L., Wallisch, M., Waisfisz, Q., van de Vrugt, H. J., Oostra, A. B., Yan, Z., Ling, C., Bishop, C. E., Hoatlin, M. E., Joenje, H., and Wang, W. A novel ubiquitin ligase is deficient in Fanconi anemia. *Nature genetics 35*, 2 (Oct. 2003), 165–70. [cited at p. 19]

[223] Meetei, A. R., Medhurst, A. L., Ling, C., Xue, Y., Singh, T. R., Bier, P., Steltenpool, J., Stone, S., Dokal, I., Mathew, C. G., Hoatlin, M., Joenje, H., de Winter, J. P., and Wang, W. A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nature genetics 37*, 9 (Sept. 2005), 958–63. [cited at p. 18]

[224] Merajver, S. D., Frank, T. S., Xu, J., Pham, T. M., Calzone, K. A., Bennett-Baker, P., Chamberlain, J., Boyd, J., Garber, J. E., Collins, F. S., and Weber, B. L. Germline BRCA1 mutations and loss of the wild-type allele in tumors from families with early onset breast and ovarian cancer. *Clinical Cancer Research 1*, May (1995), 539–544. [cited at p. 28]

[225] Merrihew, R. V., Marburger, K., Pennington, S. L., Roth, D. B., and Wilson, J. H. High-frequency illegitimate integration of transfected DNA at preintegrated target sites in a mammalian genome. *Molecular and cellular biology 15*, 1 (1996), 10. [cited at p. 61, 62, 152]

[226] Mimitou, E. P., and Symington, L. S. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature 455*, 7214 (Oct. 2008), 770–4. [cited at p. 27]

[227] Mimori, T., Akizuki, M., Yamagata, H., Inada, S., Yoshida, S., and Homma, M. Characterization of a High Molecular Weight Acidic Nuclear Protein Recognized by Autoantibodies in Sera from Patients with Polymyositis-Scleroderma Overlap. *The Journal of clinical investigation 68*, September (1981), 611–620. [cited at p. 24]

[228] Mirkovitch, J., Mirault, M. E., and Laemmli, U. K. Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell 39*, 1 (Nov. 1984), 223–32. [cited at p. 10]

[229] Mladenov, E., and Iliakis, G. Induction and repair of DNA double strand breaks: the increasing spectrum of non-homologous end joining pathways. *Mutation research 711*, 1-2 (June 2011), 61–72. [cited at p. 30]

[230] Moore, J. K., and Haber, J. E. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Molecular and Cellular Biology 16*, 5 (1996), 2164–2173. [cited at p. 24, 25]

[231] Mortusewicz, O., Amé, J.-C., Schreiber, V., and Leonhardt, H. Feedback-regulated poly(ADP-ribosyl)ation by PARP-1 is required for rapid response to DNA damage in living cells. *Nucleic acids research 35*, 22 (Jan. 2007), 7665–75. [cited at p. 22]

[232] Moser, J., Kool, H., Giakzidis, I., Caldecott, K., Mullenders, L. H. F., and Fousteri, M. I. Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III alpha in a cell-cycle-specific manner. *Molecular cell 27*, 2 (July 2007), 311–23. [cited at p. 16]

[233] Moynahan, M. E., Pierce, A. J., and Jasin, M. BRCA2 is required for homology-directed repair of chromosomal breaks. *Molecular cell 7*, 2 (Feb. 2001), 263–72. [cited at p. 28, 93, 132]

[234] Muñoz Galván, S., López-Saavedra, A., Jackson, S. P., Huertas, P., Cortés-Ledesma, F., and Aguilera, A. Competing roles of DNA end resection and non-homologous end joining functions in the repair of replication-born double-strand breaks by sister-chromatid recombination. *Nucleic Acids Research 41* (2013), 1669–1683. [cited at p. 102]

[235] Myers, S., Bottolo, L., Freeman, C., McVean, G., and Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science (New York, N.Y.) 310*, 5746 (Oct. 2005), 321–4. [cited at p. 153]

[236] Myers, S., Freeman, C., Auton, A., Donnelly, P., and McVean, G. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nature genetics 40*, 9 (Sept. 2008), 1124–9. [cited at p. 153]

[237] Nairz, K., and Klein, F. mre11S - a yeast mutation that blocks double-strand-break processing and permits nonhomologous synapsis in meiosis. *Genes & Development 11*, 17 (Sept. 1997), 2272–2290. [cited at p. 26]

[238] Nam, E. A., and Cortez, D. ATR signaling: more than meeting at the fork. *The Biochemical journal 436*, 3 (2011), 527–536. [cited at p. 12]

[239] Narod, S., Lynch, H., Conway, T., Watson, P., Feunteun, J., and Lenoir, G. Increasing incidence of breast cancer in family with BRCA1 mutation. *The Lancet 341*, April 24 (1993), 1101–1102. [cited at p. 28]

[240] Nassif, N., Penney, J., Pal, S., Engels, W. R., and Gloor, G. B. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Molecular and cellular biology 14*, 3 (1994), 1613–1625. [cited at p. 26]

[241] Neal, J. A., Dang, V., Douglas, P., Wold, M. S., Lees-Miller, S. P., and Meek, K. Inhibition of homologous recombination by DNA-dependent protein kinase requires kinase activity, is titratable, and is modulated by autophosphorylation. *Molecular and cellular biology 31*, 8 (Apr. 2011), 1719–33. [cited at p. 25]

[242] Neal, J. A., and Meek, K. Choosing the right path: does DNA-PK help make the decision? *Mutation research 711*, 1-2 (June 2011), 73–86. [cited at p. 25]

[243] Neumann, E., Schaefer-Ridder, M., Wang, Y., and Hofschneider, P. H. Gene transfer into mouse lyoma cells by electroporation in high electric fields. *The EMBO journal 1*, 7 (1982), 841–845. [cited at p. 2]

[244] Newman, E. A., Lu, F., Bashllari, D., Wang, L., Opipari, A. W., Castle, V., and Arbor, A. Alternative NHEJ Pathway Components are Therapeutic Targets in High-risk Neuroblastoma. *Molecular cancer research : MCR*, Jan 6 (2015), pii: molcanres.0337.2014. [cited at p. 55, 152]

[245] Nimonkar, A. V., Genschel, J., Kinoshita, E., Polaczek, P., Campbell, J. L., Wyman, C., Modrich, P., and Kowalczykowski, S. C. BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. *Genes & development 25*, 4 (Feb. 2011), 350–62. [cited at p. 27]

[246] Nimonkar, A. V., Ozsoy, A. Z., Genschel, J., Modrich, P., and Kowalczykowski, S. C. Human exonuclease 1 and BLM helicase interact to resect DNA and initiate DNA repair. *Proceedings of the National Academy of Sciences of the United States of America 105*, 44 (Nov. 2008), 16906–11. [cited at p. 27]

[247] Ogi, T., Limsirichaikul, S., Overmeer, R. M., Volker, M., Takenaka, K., Cloney, R., Nakazawa, Y., Niimi, A., Miki, Y., Jaspers, N. G., Mullenders, L. H. F., Yamashita, S., Fousteri, M. I., and Lehmann, A. R. Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. *Molecular cell 37*, 5 (Mar. 2010), 714–27. [cited at p. 16]

[248] Oh, S., Harvey, A., Zimbric, J., Wang, Y., Nguyen, T., Jackson, P. J., and Hendrickson, E. A. DNA ligase III and DNA ligase IV carry out genetically distinct forms of end joining in human somatic cells. *DNA repair 21* (Sept. 2014), 97–110. [cited at p. 30, 61, 71, 94, 102, 111]

[249] Orr-Weaver, T. L., and Szostak, J. W. Multiple, Tandem Plasmid Integration in Saccharomyces cerevisiae. *Molecular and cellular biology 3*, 4 (1983), 747–749. [cited at p. 6, 151]

[250] Palombo, F., Iaccarino, I., Nakajima, E., Ikejima, M., Shimada, T., and Jiricny, J. hMutSbeta, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. *Current Biology 6*, 9 (1996), 1181–1184. [cited at p. 17]

[251] Papadopoulos, N., and Lindblom, A. Molecular Basis of HNPCC: Mutations of MMR Genes. *Human Mutation 10*, 2 (1997), 89–99. [cited at p. 17]

[252] Pâques, F., and Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiology and molecular biology reviews : MMBR 63*, 2 (June 1999), 349–404. [cited at p. 28]

[253] Pastwa, E., and Błasiak, J. Non-homologous DNA end joining . *Acta Biochemica Polonica 50*, 4 (2003), 891–908. [cited at p. 25, 131]

[254] Paul, K., Wang, M., Mladenov, E., Bencsik-Theilen, A., Bednar, T., Wu, W., Arakawa, H., and Iliakis, G. DNA ligases I and III cooperate in alternative non-homologous end-joining in vertebrates. *PloS one 8*, 3 (Jan. 2013), e59505. [cited at p. 30, 61, 71, 94, 101, 111]

[255] Peña Diaz, J., and Jiricny, J. Mammalian mismatch repair: error-free or error-prone? *Trends in biochemical sciences 37*, 5 (May 2012), 206–14. [cited at p. 16]

[256] Petukhova, G., Stratton, S., and Sung, P. Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins. *Nature 393*, May (1998), 91–94. [cited at p. 28]

[257] Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic acids research 29*, 9 (May 2001), e45. [cited at p. 40, 65]

[258] Phi-Van, L., von Kries, J. P., Ostertag, W., and Strätling, W. H. The chicken lysozyme 5' matrix attachment region increases transcription from a heterologous promoter in heterologous cells and dampens position effects on the expression of transfected genes. *Molecular and cellular biology 10*, 5 (May 1990), 2302–7. [cited at p. 10]

[259] Piechaczek, C., Fetzer, C., Baiker, A., Bode, J., and Lipps, H. J. A vector based on the SV40 origin of replication and chromosomal S/MARs replicates episomally in CHO cells. *Nucleic Acids Research 27*, 2 (Jan. 1999), 426–428. [cited at p. 11, 37, 84, 153]

[260] Pittman, D. L., Cobb, J., Schimenti, K. J., Wilson, L. A., Cooper, D. M., Brignull, E., Handel, M. A., Schimenti, J. C., and Harbor, B. Meiotic Prophase Arrest with Failure of Chromosome Synapsis in Mice Deficient for Dmc1, a Germline-Specific RecA Homolog. *Molecular cell 1*, April (1998), 697–705. [cited at p. 29]

[261] Pittman, D. L., Weinberg, L. R., and Schimenti, J. C. Identification, characterization, and genetic mapping of Rad51d, a new mouse and human RAD51/RecA-related gene. *Genomics 49*, 1 (Apr. 1998), 103–11. [cited at p. 93]

[262] Platts, A. E., Quayle, A. K., and Krawetz, S. A. In-silico prediction and observations of nuclear matrix attachment. *Cellular & molecular biology letters 11*, 2 (Jan. 2006), 191–213. [cited at p. 10, 82, 153]

[263] Posakony, J. W., England, J. M., and Attardi, G. Mitochondrial growth and division during the cell cycle in HeLa cells. *The Journal of Cell Bology 74*, 2 (1977), 468–491. [cited at p. 145]

[264] Priestley, A., Beamish, H. J., Gell, D., Amatucci, A. G., Muhlmann-Diaz, M. C., Singleton, B. K., Smith, G. C., Blunt, T., Schalkwyk, L. C., Bedford, J. S., Jackson, S. P., Jeggo, P. A., and Taccioli, G. E. Molecular and biochemical characterisation of DNA-dependent protein kinase-defective rodent mutant irs-20. *Nucleic acids research 26*, 8 (Apr. 1998), 1965–73. [cited at p. 42]

[265] Prolla, T. A., Christie, D.-M., and Liskay, R. M. Dual Requirement in Yeast DNA Mismatch Repair for MLH1 and PMSJ, Two Homologs of the Bacterial mutL Gene. *Molecular and cellular biology 14*, 1 (1994), 407–415. [cited at p. 17]

[266] Puchta, H., Kocher, S., and Hohn, B. Extrachromosomal Homologous DNA Recombination in Plant Cells Is Fast and Is Not Affected by CpG Methylation. *Molecular and cellular biology 12*, 8 (1992), 3372–3379. [cited at p. 102]

[267] Puliti, A., Rizzato, C., Conti, V., Bedini, A., Gimelli, G., Barale, R., and Sbrana, I. Low-copy repeats on chromosome 22q11.2 show replication timing switches, DNA flexibility peaks and stress inducible asynchrony, sharing instability features with fragile sites. *Mutation research 686*, 1-2 (Apr. 2010), 74–83. [cited at p. 84, 153]

[268] Rafsanjani, M. S. O., Alvari, A., Samim, M., Hejazi, M. A., and Abdin, M. Z. Application of Novel Nanotechnology Strategies in Plant Biotransformation : A Contemporary Overview. *Recent Patents on Biotechnology 6*, 1 (2012), 69–79. [cited at p. 2]

[269] Rappold, I., Iwabuchi, K., Date, T., and Chen, J. Tumor Suppressor p53 Binding Protein 1 (53BP1) Is Involved in DNA Damage–signaling Pathways. *The Journal of Cell Bology 153*, 3 (2001), 613–620. [cited at p. 25]

[270] Rassool, F. V., McKeithan, T. W., Neilly, M. E., van Melle, E., Espinosa III, R., and Le Beau, M. M. Preferential integration of marker DNA into the chromosomal fragile site at 3p14: An approach to cloning fragile sites. *Proceedings of the National Academy of Sciences of the United States of America 88*, August (1991), 6657–6661. [cited at p. 84, 153]

[271] Reh, W. A., and Vasquez, K. M. Gene Targeting by Homologous Recombination. *eLS. John Wiley & Sons, Ltd*, April (May 2014), 1–10. [cited at p. 8]

[272] Resnick, M. A., and Martin, P. The repair of double-strand breaks in the nuclear DNA of Saccharomyces cerevisiae and its genetic control. *MGG Molecular & General Genetics 143*, 2 (1976), 119–129. [cited at p. 28]

[273] Reyes, J. C., Muchardt, C., and Yaniv, M. Components of the Human SWI/SNF Complex Are Enriched in Active Chromatin and Are Associated with the Nuclear Matrix. *The Journal of cell biology 137*, 2 (1997), 263–274. [cited at p. 11, 84]

[274] Richards, R. I. Fragile and unstable chromosomes in cancer: Causes and consequences. *Trends in Genetics 17*, 6 (2001), 339–345. [cited at p. 153]

[275] Robins, D. M., Ripley, S., Henderson, A. S., and Axel, R. Transforming DNA integrates into the host chromosome. *Cell 23*, January (1981), 29–39. [cited at p. 151]

[276] Rodrigue, A., Coulombe, Y., Jacquet, K., Gagne, J.-P., Roques, C., Gobeil, S., Poirier, G., and Masson, J.-Y. The RAD51 paralogs ensure cellular protection against mitotic defects and aneuploidy. *Journal of Cell Science 126*, Pt1 (2013), 348–359. [cited at p. 132, 133, 145]

[277] Rosner, M., Schipany, K., and Hengstschläger, M. Merging high-quality biochemical fractionation with a refined flow cytometry approach to monitor nucleocytoplasmic protein expression throughout the unperturbed mammalian cell cycle. *Nature protocols 8*, 3 (2013), 602–26. [cited at p. 136]

[278] Rothkamm, K., Krüger, I., Thompson, L. H., and Löbrich, M. Pathways of DNA Double-Strand Break Repair during the Mammalian Cell Cycle. *Molecular and cellular biology 23*, 16 (2003), 5706–5715. [cited at p. 26]

[279] Rubnitz, J., and Subramani, S. The minimum amount of homology required for homologous recombination in mammalian cells. *Molecular and cellular biology 4*, 11 (Nov. 1984), 2253–8. [cited at p. 102]

[280] Sakaue-Sawano, A., Kurokawa, H., Morimura, T., Hanyu, A., Hama, H., Osawa, H., Kashiwagi, S., Fukami, K., Miyata, T., Miyoshi, H., Imamura, T., Ogawa, M., Masai, H., and Miyawaki, A. Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell 132*, 3 (Feb. 2008), 487–98. [cited at p. 133, 134, 135, 136]

[281] Sale, J. E. Translesion DNA synthesis and mutagenesis in eukaryotes. *Cold Spring Harbor perspectives in biology 5*, 3 (Mar. 2013), a012708. [cited at p. 19]

[282] Saleh-Gohari, N., Bryant, H. E., Schultz, N., Parker, K. M., Cassel, T. N., and Helleday, T. Spontaneous homologous recombination is induced by collapsed replication forks that are caused by endogenous DNA single-strand breaks. *Molecular and cellular biology 25*, 16 (Aug. 2005), 7158–69. [cited at p. 26]

[283] Samulski, R. J., Zhu, X., Xiao, X., Brook, J. D., Housman, D. E., Epstein, N., and Hunter, L. A. Targeted integration of adeno-associated virus (AAV) into human chromosome 19. *The EMBO journal 10*, 12 (1991), 3941–3950. [cited at p. 4]

[284] San Filippo, J., Sung, P., and Klein, H. Mechanism of eukaryotic homologous recombination. *Annual review of biochemistry 77* (Jan. 2008), 229–57. [cited at p. 132]

[285] Sartori, A. A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., and Jackson, S. P. Human CtIP promotes DNA end resection. *Nature 450*, 7169 (Nov. 2007), 509–14. [cited at p. 26, 27, 93, 132]

[286] Sauer, B. Inducible Gene Targeting in Mice Using the Cre/lox System. *Methods 14*, 4 (1998), 381–392. [cited at p. 7]

[287] Schärer, O. D. Nucleotide excision repair in eukaryotes. *Cold Spring Harbor perspectives in biology 5*, 10 (Oct. 2013), a012609. [cited at p. 15]

[288] Schulte-Uentrop, L., El-Awady, R. A., Schliecker, L., Willers, H., and Dahm-Daphi, J. Distinct roles of XRCC4 and Ku80 in non-homologous end-joining of endonuclease- and ionizing radiation-induced DNA double-strand breaks. *Nucleic acids research 36*, 8 (May 2008), 2561–9. [cited at p. 117]

[289] Sedgwick, B., Bates, P. A., Paik, J., Jacobs, S. C., and Lindahl, T. Repair of alkylated DNA: recent advances. *DNA repair 6*, 4 (Apr. 2007), 429–42. [cited at p. 13]

[290] Seidman, M. M. Intermolecular homologous recombination between transfected sequences in mammalian cells is primarily nonconservative. *Molecular and cellular biology 7*, 10 (Oct. 1987), 3561–5. [cited at p. 102]

[291] Seluanov, A., Mittelman, D., Pereira-Smith, O. M., Wilson, J. H., and Gorbunova, V. DNA end joining becomes less efficient and more error-prone during cellular senescence. *Proceedings of the National Academy of Sciences of the United States of America 101*, 20 (May 2004), 7624–9. [cited at p. 39, 46, 68, 102]

[292] Shibata, A., Barton, O., Noon, A. T., Dahm, K., Deckbar, D., Goodarzi, A. A., Löbrich, M., and Jeggo, P. A. Role of ATM and the damage response mediator proteins 53BP1 and MDC1 in the maintenance of G(2)/M checkpoint arrest. *Molecular and cellular biology 30*, 13 (July 2010), 3371–83. [cited at p. 144]

[293] Shibata, A., Conrad, S., Birraux, J., Geuting, V., Barton, O., Ismail, A., Kakarougkas, A., Meek, K., Taucher-Scholz, G., Löbrich, M., and Jeggo, P. A. Factors determining DNA double-strand break repair pathway choice in G2 phase. *The EMBO journal 30*, 6 (Mar. 2011), 1079–92. [cited at p. 132]

[294] Shiloh, Y. The ATM-mediated DNA-damage response: taking shape. *Trends in biochemical sciences 31*, 7 (July 2006), 402–10. [cited at p. 12, 131]

[295] Shiloh, Y., and Lehmann, A. R. Maintaining integrity. *Nature cell biology 6*, 10 (2004), 923–928. [cited at p. 131]

[296] Shiloh, Y., and Ziv, Y. The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. *Nature reviews. Molecular cell biology 14*, 4 (Apr. 2013), 197–210. [cited at p. 12, 13, 131]

[297] Shtam, T. A., Kovalev, R. A., Varfolomeeva, E. Y., Makarov, E. M., Kil, Y. V., and Filatov, M. V. Exosomes are natural carriers of exogenous siRNA to human cells in vitro. *Cell communication and signaling : CCS 11* (2013), 88. [cited at p. 132, 139]

[298] Sijbers, A. M., de Laat, W. L., Ariza, R. R., Biggerstaff, M., Wei, Y.-F., Moggs, J. G., Carter, K. C., Shell, B. K., Evans, E., de Jong, M. C., Rademakers, S., de Rooij, J., Jaspers, N. G., Hoeijmakers, J. H., and Wood, R. D. Xeroderma Pigmentosum Group F Caused by a Defect in a Structure-Specific DNA Repair Endonuclease. *Cell 86*, 5 (Sept. 1996), 811–822. [cited at p. 16]

[299] Simsek, D., Brunet, E., Wong, S. Y.-W., Katyal, S., Gao, Y., McKinnon, P. J., Lou, J., Zhang, L., Li, J., Rebar, E. J., Gregory, P. D., Holmes, M. C., and Jasin, M. DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. *PLoS genetics 7*, 6 (June 2011), e1002080. [cited at p. 79, 152]

[300] Sipley, J. D., Menninger, J. C., Hartley, K. O., Ward, D. C., Jackson, S. P., and Anderson, C. W. Gene for the catalytic subunit of the human DNA-activated protein kinase maps to the site of the XRCC7 gene on chromosome 8. *Proceedings of the National Academy of Sciences of the United States of America 92*, 16 (Aug. 1995), 7515–9. [cited at p. 93]

[301] Smith, J., Baldeyron, C., Oliveira, I. D., Sala-Trepat, M., and Papadopoulo, D. The influence of DNA double-strand break structure on end-joining in human cells. *Nucleic acids research 29*, 23 (2001), 4783–4792. [cited at p. 24, 156]

[302] Smith, M. C. M., Brown, W. R. A., McEwan, A. R., and Rowley, P. A. Site-specific recombination by phiC31 integrase and other large serine recombinases. *Biochemical Society transactions 38*, 2 (Apr. 2010), 388–94. [cited at p. 8]

[303] Sobol, R. W., Prasad, R., Evenski, A., Baker, A., Yang, X.-P., Horton, J. K., and Wilson, S. H. The lyase activity of the DNA repair protein beta-polymerase protects from DNA-damage-induced cytotoxicity. *Nature 405*, 6788 (2000), 807–810. [cited at p. 14, 22]

[304] Sonoda, E., Hochegger, H., Saberi, A., Taniguchi, Y., and Takeda, S. Differential usage of non-homologous end-joining and homologous recombination in double strand break repair. *DNA repair 5*, 9-10 (Sept. 2006), 1021–9. [cited at p. 26]

[305] Soubeyrand, S., Pope, L., De Chasseval, R., Gosselin, D., Dong, F., de Villartay, J.-P., and Haché, R. J. G. Artemis phosphorylated by DNA-dependent protein kinase associates preferentially with discrete regions of chromatin. *Journal of molecular biology 358*, 5 (May 2006), 1200–11. [cited at p. 24]

[306] Soulas-Sprauel, P., Rivera-Munoz, P., Malivert, L., Le Guyader, G., Abramowski, V., Revy, P., and de Villartay, J.-P. V(D)J and immunoglobulin class switch recombinations: a paradigm to study the regulation of DNA end-joining. *Oncogene 26*, 56 (Dec. 2007), 7780–91. [cited at p. 24]

[307] Sperry, A. O., Blasquez, V. C., and Garrard, W. T. Dysfunction of chromosomal loop attachment sites: illegitimate recombination linked to matrix association regions and topoisomerase II. *Proceedings of the National Academy of Sciences of the United States of America 86*, 14 (July 1989), 5497–501. [cited at p. 82, 153]

[308] Stracker, T. H., and Petrini, J. H. J. The MRE11 complex: starting from the ends. *Nature reviews. Molecular cell biology 12*, 2 (Feb. 2011), 90–103. [cited at p. 93]

[309] Stucki, M., Clapperton, J. A., Mohammad, D., Yaffe, M. B., Smerdon, S. J., and Jackson, S. P. MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell 123*, 7 (Dec. 2005), 1213–26. [cited at p. 62, 93]

[310] Stucki, M., Pascucci, B., Parlanti, E., Fortini, P., Wilson, S. H., Hübscher, U., and Dogliotti, E. Mammalian base excision repair by DNA polymerases δ and ϵ. *Oncogene 17*, 7 (1998), 835–843. [cited at p. 15, 81]

[311] Su, S.-S., and Modrich, P. Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs. *Proceedings of the National Academy of Sciences 83*, July (1986), 5057–5061. [cited at p. 17]

[312] Sugasawa, K., Ng, J. M., Masutani, C., Iwai, S., van der Spek, P. J., Eker, A. P., Hanaoka, F., Bootsma, D., and Hoeijmakers, J. H. Xeroderma Pigmentosum Group C Protein Complex Is the Initiator of Global Genome Nucleotide Excision Repair. *Molecular Cell 2*, 2 (Aug. 1998), 223–232. [cited at p. 15]

[313] Sugiyama, T., Zaitseva, E. M., and Kowalczykowski, S. C. A Single-stranded DNA-binding Protein Is Needed for Efficient Presynaptic Complex Formation by the Saccharomyces cerevisiae Rad51 Protein. *Journal of Biological Chemistry 272*, 12 (Mar. 1997), 7940–7945. [cited at p. 27]

[314] Sun, N., and Zhao, H. Transcription activator-like effector nucleases (TALENs): a highly efficient and versatile tool for genome editing. *Biotechnology and bioengineering 110*, 7 (July 2013), 1811–21. [cited at p. 9]

[315] Suwaki, N., Klare, K., and Tarsounas, M. RAD51 paralogs: roles in DNA damage signalling, recombinational repair and tumorigenesis. *Seminars in cell & developmental biology 22*, 8 (Oct. 2011), 898–905. [cited at p. 93, 132]

[316] Svetlova, E. Y., Razin, S., and Debatisse, M. Mammalian recombination hot spot in a DNA loop anchorage region: A model for the study of common fragile sites. *Journal of Cellular Biochemistry 81*, S36 (2001), 170–178. [cited at p. 84]

[317] Swagemakers, S. M. A., Essers, J., de Wit, J., Hoeijmakers, J. H. J., and Kanaar, R. The Human Rad54 Recombinational DNA Repair Protein Is a Double-stranded DNA-dependent ATPase. *Journal of Biological Chemistry 273*, 43 (Oct. 1998), 28292–28297. [cited at p. 28, 132]

[318] Szatrowski, T. P., and Nathan, C. F. Production of Large Amounts of Hydrogen Peroxide by Human Tumor Cells. *Cancer research 51*, 3 (1991), 794–798. [cited at p. 113]

[319] Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., and Stahl, F. W. The double-strand-break repair model for recombination. *Cell 33*, 1 (May 1983), 25–35. [cited at p. 26]

[320] Takata, M., Sasaki, M., Tachiiri, S., Fukushima, T., Sonoda, E., Schild, D., Thompson, L. H., and Takeda, S. Chromosome instability and defective recombinational repair in knockout mutants of the five Rad51 paralogs. *Molecular and cellular biology 21*, 8 (2001), 2858–2866. [cited at p. 93]

[321] Taylor, W. R., and Stark, G. R. Regulation of the G2/M transition by p53. *Oncogene 20*, 15 (2001), 1803–1815. [cited at p. 144]

[322] Thummel, C., Tjian, R., and Grodzicker, T. Expression of SV40 T Antigen under Control of Adenovirus Promoters. *Cell 23*, March (1981), 825–836. [cited at p. 3]

[323] Tobin, L. A., Robert, C., Nagaria, P., Chumsri, S., Twaddell, W., Ioffe, O. B., Greco, G. E., Brodie, A. H., Tomkinson, A. E., and Rassool, F. V. Targeting abnormal DNA repair in therapy-resistant breast cancers. *Molecular cancer research : MCR 10*, 1 (Jan. 2012), 96–107. [cited at p. 31, 79, 113, 154]

[324] Tomkinson, A. E., Vijayakumar, S., Pascal, J. M., and Ellenberger, T. DNA ligases: structure, reaction mechanism, and function. *Chemical reviews 106*, 2 (Feb. 2006), 687–99. [cited at p. 15, 22]

[325] Tomura, M., Sakaue-Sawano, A., Mori, Y., Takase-Utsugi, M., Hata, A., Ohtawa, K., Kanagawa, O., and Miyawaki, A. Contrasting Quiescent G0 Phase with Mitotic Cell Cycling in the Mouse Immune System. *PLoS ONE 8*, 9 (2013), 1–10. [cited at p. 139]

[326] Townsend, K., Mason, H., Blackford, A. N., Miller, E. S., Chapman, J. R., Sedgwick, G. G., Barone, G., Turnell, A. S., and Stewart, G. S. Mediator of DNA damage checkpoint 1 (MDC1) regulates mitotic progression. *The Journal of biological chemistry 284*, 49 (Dec. 2009), 33939–48. [cited at p. 139]

[327] Tran, P. T., Erdeniz, N., Symington, L. S., and Liskay, R. M. EXO1-A multi-tasking eukaryotic nuclease. *DNA repair 3*, 12 (Dec. 2004), 1549–59. [cited at p. 17]

[328] Truong, L. N., Li, Y., Shi, L. Z., Hwang, P. Y.-H., He, J., Wang, H., and Razavian, N. Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *PNAS 110*, 19 (2013), 7720–7725. [cited at p. 30, 61, 79, 84, 102, 153]

[329] Tsai, C. J., Kim, S. A., and Chu, G. Cernunnos-XLF promotes the ligation of mismatched and noncohesive DNA ends. *PNAS 104*, 19 (2007), 7851–7856. [cited at p. 25]

[330] Tsuzuki, T., Fujii, Y., Sakumi, K., Tominaga, Y., Nakao, K., Sekiguchi, M., Matsushiro, A., Yoshimura, Y., and Morita, T. Targeted disruption of the Rad51 gene leads to lethality in embryonic mice. *Proceedings of the National Academy of Sciences of the United States of America 93*, 13 (June 1996), 6236–40. [cited at p. 27, 132]

[331] Turan, S., Zehe, C., Kuehle, J., Qiao, J., and Bode, J. Recombinase-mediated cassette exchange (RMCE) - a rapidly-expanding toolbox for targeted genomic modifications. *Gene 515*, 1 (Feb. 2013), 1–27. [cited at p. 8]

[332] Turenne, G. A., Paul, P., Laflair, L., and Price, B. D. Activation of p53 transcriptional activity requires ATM's kinase domain and multiple N-terminal serine residues of p53. *Oncogene 20*, 37 (2001), 5100–5110. [cited at p. 12]

[333] Urlaub, G., and Chasin, L. A. Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. *Proceedings of the National Academy of Sciences of the United States of America 77*, 7 (July 1980), 4216–20. [cited at p. 63, 103, 133]

[334] Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., and Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nature reviews. Genetics 11*, 9 (Sept. 2010), 636–46. [cited at p. 9]

[335] van der Burg, M., Ijspeert, H., Verkaik, N. S., Turul, T., Wiegant, W. W., Morotomi-Yano, K., Mari, P.-O., Tezcan, I., Chen, D. J., Zdzienicka, M. Z., van Dongen, J. J. M., and van Gent, D. C. A DNA-PKcs mutation in a radiosensitive T–B– SCID patient inhibits Artemis activation and nonhomologous end-joining. *Journal of Clinical Investigation 119*, 1 (2009), 91–98. [cited at p. 53]

[336] Van Dyck, E., Hajibagheri, N. M., Stasiak, A., and West, S. C. Visualisation of human rad52 protein and its complexes with hRad51 and DNA. *Journal of molecular biology 284*, 4 (Dec. 1998), 1027–38. [cited at p. 93]

[337] Van Dyck, E., Stasiak, A. Z., Stasiak, A., and West, S. C. Binding of double-strand breaks in DNA by human Rad52 protein. *Nature 398*, September (1999), 728–731. [cited at p. 93]

[338] Vannucci, L., Lai, M., Chiuppesi, F., Ceccherini-Nelli, L., and Pistello, M. Viral vectors: a look back and ahead on gene transfer technology. *New Microbiologica 36*, 1 (2013), 1–22. [cited at p. 3]

[339] Vasquez, K. M., Marburger, K., Intody, Z., and Wilson, J. H. Manipulating the mammalian genome by homologous recombination. *Proceedings of the National Academy of Sciences of the United States of America 98*, 15 (July 2001), 8403–10. [cited at p. 62, 156]

[340] Verhaegh, G. W., Jongmans, W., Morolli, B., Jaspers, N. G., van der Schans, G. P., Lohman, P. H., and Zdzienicka, M. Z. A novel type of X-ray-sensitive Chinese hamster cell mutant with radioresistant DNA synthesis and hampered DNA double-strand break repair. *Mutation research 337*, 2 (Sept. 1995), 119–29. [cited at p. 45, 54]

[341] Vispé, S., Cazaux, C., Lesca, C., and Defais, M. Overexpression of Rad51 protein stimulates homologous recombination and increases resistance of mammalian cells to ionizing radiation. *Nucleic acids research 26*, 12 (June 1998), 2859–64. [cited at p. 93]

[342] Walden, H., and Deans, A. J. The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. *Annual review of biophysics 43* (Jan. 2014), 257–78. [cited at p. 18]

[343] Walker, J. R., Corpina, R. A., and Goldberg, J. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature 412*, 6847 (Aug. 2001), 607–14. [cited at p. 24, 93]

[344] Walther, W., and Stein, U. Viral Vectors for Gene Transfer A Review of Their Use in the Treatment of Human Diseases. *Drugs 60*, 2 (2000), 249–271. [cited at p. 3]

[345] Wang, B., Matsuoka, S., Carpenter, P. B., and Elledge, S. J. 53BP1, a mediator of the DNA damage checkpoint. *Science (New York, N.Y.) 298*, 5597 (Nov. 2002), 1435–8. [cited at p. 25]

[346] Wang, H., Perrault, A. R., Takeda, Y., Qin, W., and Iliakis, G. Biochemical evidence for Ku-independent backup pathways of NHEJ. *Nucleic Acids Research 31*, 18 (Sept. 2003), 5377–5388. [cited at p. 101]

[347] Wang, H., Shao, Z., Shi, L. Z., Hwang, P. Y.-H., Truong, L. N., Berns, M. W., Chen, D. J., and Wu, X. CtIP protein dimerization is critical for its recruitment to chromosomal DNA double-stranded breaks. *The Journal of biological chemistry 287*, 25 (June 2012), 21471–80. [cited at p. 72, 94, 102, 103, 118]

[348] Wang, Z., Wang, F., Tang, T., and Guo, C. The role of PARP1 in the DNA damage response and its application in tumor therapy. *Frontiers of medicine 6*, 2 (June 2012), 156–64. [cited at p. 22]

[349] Wei, Y.-F., Robins, P., Carter, K., Caldecott, K., Pappin, D. J. C., Yu, G.-L., Wang, R.-P., Shell, B. K., Nash, R. A., Schär, P., Barnes, D. E., Haseltine, W. A., and Lindahl, T. Molecular cloning and expression of human cDNAs encoding a novel DNA ligase IV and DNA ligase III, an enzyme active in DNA repair and recombination. *Molecular and cellular biology 15*, 6 (1995), 3206–3216. [cited at p. 25]

[350] Wesoly, J., Agarwal, S., Sigurdsson, S., Bussen, W., Van Komen, S., Qin, J., van Steeg, H., van Benthem, J., Wassenaar, E., Baarends, W. M., Ghazvini, M., Tafel, A. A., Heath, H., Galjart, N., Essers, J., Grootegoed, J. A., Arnheim, N., Bezzubova, O., Buerstedde, J.-M., Sung, P., and Kanaar, R. Differential Contributions of Mammalian Rad54 Paralogs to Recombination, DNA Damage Repair, and Meiosis. *Molecular and cellular biology 26*, 3 (2006), 976–989. [cited at p. 28]

[351] Whitmore, G. F., Varghese, A. J., and Gulyas, S. Cell cycle responses of two X-ray sensitive mutants defective in DNA repair. *International journal of radiation biology 56*, 5 (1989), 657–665. [cited at p. 42]

[352] Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., Sanmiguel, P., and Schulman, A. H. A

unified classification system for eukaryotic transposable elements. *Nature reviews Genetics 8*, 12 (2007), 973–982. [cited at p. 5]

[353] Wiegant, W. W., Overmeer, R. M., Godthelp, B. C., van Buul, P. P. W., and Zdzienicka, M. Z. Chinese hamster cell mutant, V-C8, a model for analysis of Brca2 function. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis 600*, 1-2 (2006), 79–88. [cited at p. 45]

[354] Willers, H., McCarthy, E. E., Hubbe, P., Dahm-Daphi, J., and Powell, S. N. Homologous recombination in extrachromosomal plasmid substrates is not suppressed by p53. *Carcinogenesis 22*, 11 (Nov. 2001), 1757–63. [cited at p. 117]

[355] Wiltshire, T., Senft, J., Wang, Y., Konat, G. W., Wenger, S. L., Reed, E., and Wang, W. BRCA1 contributes to cell cycle arrest and chemoresistance in response to the anticancer agent irofulven. *Molecular pharmacology 71*, 4 (2007), 1051–1060. [cited at p. 132]

[356] Wong, E. A., and Capecchi, M. R. Homologous recombination between coinjected DNA sequences peaks in early to mid-S phase. *Molecular and cellular biology 7*, 6 (June 1987), 2294–5. [cited at p. 62]

[357] Wooster, R., Neuhausen, S. L., Mangion, J., Quirk, Y., Ford, D., Collins, N., Nguyen, K., Seal, S., Tran, T., Averill, D., Fields, P., Marshall, G., Narod, S., Lenoir, G. M., Lynch, H., Feunteun, J., Devilee, P., Cornelisse, C. J., Menko, F. H., Daly, P. A., Ormiston, W., Mcmanus, R., Pye, C., Lewis, C. M., Cannon-Albright, L. A., Peto, J., Ponder, B. A. J., Skolnick, M. H., Easton, D. F., Goldgar, D. E., and Stratton, M. R. Localization of a Breast Cancer Susceptibility Gene, BRCA2, to Chromosome 13q1 2-13. *Science 265*, September 30 (1994), 15–17. [cited at p. 28]

[358] Wu, L. C., Wang, Z. W., Tsan, J. T., Spillman, M. A., Phung, A., Xu, X. L., Yang, M.-C. W., Hwang, L.-Y., Bowcock, A. M., and Baer, R. Identification of a RING protein that can interact in vivo with the BRCA1 gene product. *Nature genetics 14*, 4 (1996), 430–40. [cited at p. 29, 93]

[359] Würtele, H., Little, K. C. E., and Chartrand, P. Illegitimate DNA integration in mammalian cells. *Gene therapy 10*, 21 (Oct. 2003), 1791–9. [cited at p. 62, 151]

[360] Xia, S. J., Shammas, M. A., and Shmookler Reis, R. J. Elevated recombination in immortal human cells is mediated by HsRAD51 recombinase. *Molecular and cellular biology 17*, 12 (1997), 7151–7158. [cited at p. 154]

[361] Xie, A., Hartlerode, A., Stucki, M., Odate, S., Puget, N., Kwok, A., Nagaraju, G., Yan, C., Alt, F. W., Chen, J., Jackson, S. P., and Scully, R. Distinct roles of chromatin-associated proteins MDC1 and 53BP1 in mammalian double-strand break repair. *Molecular cell 28*, 6 (Dec. 2007), 1045–57. [cited at p. 25, 93]

[362] Xu, X., Nagarajan, H., Lewis, N. E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., Hammond, S., Andersen, M. R., Neff, N., Passarelli, B., Koh, W., Fan, H. C., Wang, J., Gui, Y., Lee, K. H., Betenbaugh, M. J., Quake, S. R., Famili, I., Palsson, B. O., and Wang, J. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nature biotechnology 29*, 8 (Aug. 2011), 735–41. [cited at p. 65]

[363] Xu, Z.-Y., Loignon, M., Han, F.-Y., Panasci, L., and Aloyz, R. Xrcc3 induces cisplatin resistance by stimulation of Rad51-related recombinational repair, S-phase checkpoint activation, and reduced apoptosis. *The Journal of pharmacology and experimental therapeutics 314*, 2 (2005), 495–505. [cited at p. 133]

[364] Yamtich, J., and Sweasy, J. B. DNA Polymerase Family X: Function, Structure, and Cellular Roles. *Biochimica et Biophysica Acta 1804*, 5 (2010), 1136–1150. [cited at p. 15, 22]

[365] Yan, Z., Delannoy, M., Ling, C., Daee, D., Osman, F., Muniandy, P. A., Shen, X., Oostra, A. B., Du, H., Steltenpool, J., Lin, T., Schuster, B., Decaillet, C., Stasiak, A., Stasiak, A. Z., Stone, S., Hoatlin, M. E., Schindler, D., Woodcock, C., Joenje, H., Sen, R., de Winter, J. P., Li, L., Seidman, M. M., Whitby, M. C., Myung, K., Constantinou, A., and Wang, W. A histone-fold complex and FANCM form a conserved DNA-remodeling complex to maintain genome stability. *Molecular cell 37*, 6 (2010), 865–878. [cited at p. 18]

[366] Yant, S. R., Wu, X., Huang, Y., Garrison, B., Burgess, S. M., and Kay, M. A. High-resolution genome-wide mapping of transposon integration in mammals. *Molecular and cellular biology 25*, 6 (Mar. 2005), 2085–94. [cited at p. 5]

[367] Yarden, R. I., Pardo-Reoyo, S., Sgagias, M., Cowan, K. H., and Brody, L. C. BRCA1 regulates the G2/M checkpoint by activating Chk1 kinase upon DNA damage. *Nature genetics 30*, February (2002), 285–289. [cited at p. 132, 139]

[368] Yoshida, K., and Miki, Y. Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. *Cancer science 95*, 11 (Nov. 2004), 866–71. [cited at p. 93]

[369] Yoshimura, M., Kohzaki, M., Nakamura, J., Asagoshi, K., Sonoda, E., Hou, E., Prasad, R., Wilson, S. H., Tano, K., Yasui, A., Lan, L., Seki, M., Wood, R. D., Arakawa, H., Buerstedde, J.-M., Hochegger, H., Okada, T., Hiraoka, M., and Takeda, S. Vertebrate POLQ and POLbeta cooperate in base excision repair of oxidative DNA damage. *Molecular cell 24*, 1 (2006), 115–125. [cited at p. 21]

[370] You, Z., Shi, L. Z., Zhu, Q., Wu, P., Zhang, Y.-W., Basilio, A., Tonnu, N., Verma, I. M., Berns, M. W., and Hunter, T. CtIP links DNA double-strand break sensing to resection. *Molecular cell 36*, 6 (Dec. 2009), 954–69. [cited at p. 27, 93]

[371] Yousefzadeh, M. J., Wyatt, D. W., Takata, K.-I., Mu, Y., Hensley, S. C., Tomida, J., Bylund, G. O., Doublié, S., Johansson, E., Ramsden, D. a., McBride, K. M., and Wood,

R. D. Mechanism of Suppression of Chromosomal Instability by DNA Polymerase POLQ. *PLoS genetics 10*, 10 (Oct. 2014), e1004654. [cited at p. 73, 111]

[372] Yu, A. M., and McVey, M. Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. *Nucleic acids research 38*, 17 (Sept. 2010), 5706–17. [cited at p. 21, 30, 31, 73, 80, 89, 94, 102, 115, 119, 120]

[373] Yu, X., and Chen, J. DNA Damage-Induced Cell Cycle Checkpoint Control Requires CtIP , a Phosphorylation-Dependent Binding Partner of BRCA1 C-Terminal Domains. *Molecular and cellular biology 24*, 21 (2004), 9478–9486. [cited at p. 93]

[374] Yun, M. H., and Hiom, K. CtIP-BRCA1 modulates the choice of DNA double-strand break repair pathway throughout the cell cycle. *Nature 459*, 7245 (2009), 460–463. [cited at p. 30, 94]

[375] Yusufzai, T., and Kadonaga, J. T. HARP Is an ATP-Driven Annealing Helicase. *Science 322*, October (2008), 748–750. [cited at p. 84]

[376] Yusufzai, T. M., and Felsenfeld, G. The 5'-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. *PNAS 101*, 23 (2004), 8620–8624. [cited at p. 11]

[377] Zahn-Zabal, M., Kobr, M., Girod, P.-A., Imhof, M., Chatellard, P., de Jesus, M., Wurm, F., and Mermod, N. Development of stable cell lines for production or regulated expression using matrix attachment regions. *Journal of biotechnology 87*, 1 (Apr. 2001), 29–42. [cited at p. 11, 37, 62]

[378] Zaunbrecher, G. M., Mir, B., Dunne, P. W., Breen, M., and Jorge, A. Enhancement of extra chromosomal recombination in somatic cells by affecting the ratio of homologous recombination (HR) to non-homologous end joining (NHEJ). *Animal Biotechnology 19*, 1 (2008), 6–21. [cited at p. 156]

[379] Zhang, J., Ma, Z., Treszezamsky, A., and Powell, S. N. MDC1 interacts with Rad51 and facilitates homologous recombination. *Nature structural & molecular biology 12*, 10 (Oct. 2005), 902–9. [cited at p. 93, 132]

[380] Zhang, X., and Paull, T. T. The Mre11/Rad50/Xrs2 complex and non-homologous end-joining of incompatible ends in S. cerevisiae. *DNA repair 4*, 11 (Nov. 2005), 1281–94. [cited at p. 31, 94, 102]

[381] Zhang, Y., and Jasin, M. An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. *Nature structural & molecular biology 18*, 1 (Jan. 2011), 80–4. [cited at p. 79, 152]

[382] Zhang, Y., Yuan, F., Presnell, S. R., Tian, K., Gao, Y., Tomkinson, A. E., Gu, L., and Li, G.-M. Reconstitution of 5'-directed human mismatch repair in a purified system. *Cell 122*, 5 (Sept. 2005), 693–705. [cited at p. 17]

[383] Zhou, Y., Caron, P., Legube, G., and Paull, T. T. Quantitation of DNA double-strand break resection intermediates in human cells. *Nucleic acids research 42*, 3 (Feb. 2014), e19. [cited at p. 25, 102]

Appendices

Appendix A

Appendix

Published online 7 June 2011

Nucleic Acids Research, 2011, Vol. 39, No. 15 ***e104***
doi:10.1093/nar/gkr436

High-level transgene expression by homologous recombination-mediated gene transfer

Mélanie Grandjean[1], Pierre-Alain Girod[1], David Calabrese[1], Kaja Kostyrko[1], Marianne Wicht[2], Florence Yerly[3], Christian Mazza[3], Jacques S. Beckmann[2], Danielle Martinet[2] and Nicolas Mermod[1,*]

[1]Laboratory of Molecular Biotechnology, Center for Biotechnology UNIL-EPFL, University of Lausanne, [2]Service of Medical Genetics, Center Hospitalier Universitaire Vaudois (CHUV), 1015 Lausanne and [3]Department of Mathematics, University of Fribourg, 1700 Fribourg, Switzerland

Received August 24, 2010; Revised May 12, 2011; Accepted May 13, 2011

ABSTRACT

Gene transfer and expression in eukaryotes is often limited by a number of stably maintained gene copies and by epigenetic silencing effects. Silencing may be limited by the use of epigenetic regulatory sequences such as matrix attachment regions (MAR). Here, we show that successive transfections of MAR-containing vectors allow a synergistic increase of transgene expression. This finding is partly explained by an increased entry into the cell nuclei and genomic integration of the DNA, an effect that requires both the MAR element and iterative transfections. Fluorescence *in situ* hybridization analysis often showed single integration events, indicating that DNAs introduced in successive transfections could recombine. High expression was also linked to the cell division cycle, so that nuclear transport of the DNA occurs when homologous recombination is most active. Use of cells deficient in either non-homologous end-joining or homologous recombination suggested that efficient integration and expression may require homologous recombination-based genomic integration of MAR-containing plasmids and the lack of epigenetic silencing events associated with tandem gene copies. We conclude that MAR elements may promote homologous recombination, and that cells and vectors can be engineered to take advantage of this property to mediate highly efficient gene transfer and expression.

INTRODUCTION

A major impediment to efficient and stable transgene expression is the variability of expression noted in independently transformed mammalian cells and organisms, both in experimental biology and for therapeutic applications. The high degree of expression variability is thought to depend on the number of transgene copies that integrate within the host genome and on the site of transgene integration (1,2). Indeed, transgene expression may be influenced by the fortuitous presence of regulatory elements at the random integration locus in the host genome. In addition, transgene expression is thought to reflect the influence of particular chromatin structure coming from adjacent chromosomal domains (3–5). Finally, the co-integration of multiple transgene copies at the same genomic locus may lead to silencing, possibly because of the formation of small inhibitory RNAs from antisense transgene transcription (6).

To increase and stabilize transgene expression in mammalian cells, epigenetic regulators such as matrix attachment regions (MAR) are increasingly used to protect transgenes from silencing effects (7). MAR were first discovered two decades ago for their association with the nuclear matrix or scaffold (8,9), a poorly characterized structural network that may consist of various non-histone nuclear proteins such as lamins, topoisomerases and components of transcription machinery (10). Eukaryotic chromosomes are organized in independent loops of chromatin that may control DNA replication, transcriptional regulation and chromosomal packaging (11–15). MARs were proposed to be the specific DNA sequences that anchor the chromosomes to the matrix

*To whom correspondence should be addressed. Tel: +41 21 693 61 51; Fax: +41 21 693 76 10; Email: nicolas.mermod@unil.ch
Present addresses:
Pierre-Alain Girod, Selexis SA, 18 Ch. des Aulx, Geneva, Switzerland.
David Calabrese, Selexis SA, 18 Ch. des Aulx, Geneva, Switzerland.

and partition chromosomes into these 50–200 kb DNA loop structures (16–18).

MARs are polymorphic 300–3000 bp-long DNA elements composed essentially of non-coding AT-rich sequences, and they are estimated to be 50 000–100 000 in the mammalian genomes (10). Their activity is thought to relate to their structural properties rather than to their primary sequence. Although no consensus MAR sequence has been found, they often have AT-rich sequences (19) and they may adopt particular conformations and physicochemical properties, such as a natural curvature (20), a deep major groove and a narrow minor groove (21), a high DNA strand unwinding and unpairing susceptibility (12), and a high potential to double-helix denaturation (22,23).

Besides providing a topological structure to the chromatin, MARs also contribute to regulate key genomic functions (24), as they were involved in the control of activities such as DNA replication and gene transcription (25,26). For instance, several origins of replication have been mapped within MARs in various eukaryotic genomes (27). Moreover, MARs are able to recruit endogenous replication factors and may allow sustained episomal replication when placed within an active transcription unit (28,29). Similarly, the ability of MARs to influence gene expression has been associated to the binding of protein factors in addition to the intrinsic properties of their DNA sequence (8,30,31). MARs associate with specific ubiquitous and tissue-specific transcription factors such as special AT-rich binding protein1 [SATB-1; (32)], NMP4 (33) and CTCF (34), which may in turn recruit regulatory proteins such as histone acetyl transferases, topoisomerases and ATP-dependent chromatin remodeling complexes to mediate a more expression-permissive chromatin state (35,36), as well as components of the transcription machinery and splicing factors (37,38). Thus, in addition to defining chromatin loop domains and organizing chromosomal architecture, MARs may contribute to control chromatin structure and gene expression.

MAR elements were shown to increase transgene expression and to decrease the clonal variability in stable transfections of mammalian cell lines and in transgenic plants and animals (21,39–42). MARs were proposed to act as insulators that protect transgenes from the repressive effects of surrounding heterochromatic area of the chromosomes and/or to relocate transgenes in an active compartment of the nucleus (7). MAR were also incorporated into viral vectors to reduce their susceptibility to silencing (43). A transgene flanked with MAR elements may thus constitute an autonomous chromatin domain whose expression would remain independent of the adjacent chromosomal environment.

MAR elements may also constitute targets of DNA recombination or rearrangement events, as exemplified by many MAR-related deletions and translocations involved in leukaemia and breast cancer (44,45). Additionally, it was shown that retroviral integration often occurs close to or within MARs (46,47). MARs were shown to bind DNA topoisomerase II (48), an enzyme that catalyzes double-strand breaks, as may be required to initiate the recombination pathways of DNA repair (49,50). However, whether MARs might mediate or regulate DNA recombination remains to be demonstrated.

In this study, we show that successive transfections of MAR-containing expression vectors mediate an unexpectedly high increase in transgene expression when the DNA reaches the nucleus at phases of the cell division cycle when the homologous recombination (HR) pathway is most active. This results in part from an increased transport of the DNA in the nucleus and more efficient genomic integration. This effect is abrogated in cells deficient in HR as opposed to non-homologous end-joining. This study thus allows to propose a new function of MAR elements, which may be to promote homologous DNA recombination. It also identifies some of the limitations to efficient gene transfer and expression in eukaryotic cells, and it provides new avenues for more efficient and more reliable gene expression, for instance to express therapeutic proteins or for gene and cell therapies.

MATERIALS AND METHODS

Plasmids and constructs

pGEGFPcontrol contains the SV40 early promoter, enhancer and vector backbone from pGL3 (Promega) driving the expression of the eGFP gene from pEGFP-N1 (Clontech). pPAG01SV40EGFP results from the insertion of the chicken lysozyme MAR fragment upstream of the SV40 early promoter of pGEGFPcontrol (51). The human MAR 1–68 was identified by the SMARScan program using DNA structural properties. It was cloned from human bacterial artificial chromosomes in pBluescript and then inserted into pGEGFPcontrol upstream the SV40 early promoter, resulting in the p1–68(NcoI filled)SV40EGFP plasmid (21).

pGL3-CMV-DsRed was created by inserting the DsRed gene, under the control of the CMV promoter (including the enhancer), from pCMV-DsRed (Clontech) in pGL3-basic (Promega). pGL3-CMV-DsRed-kan was then created by exchanging the ampicillin gene of pGL3-CMV-DsRed for kanamycin resistance gene from pCMV-DsRed (Clontech) by digestion of both plasmids with BspHI. Then, the chicken lysozyme or the human 1–68 MAR were inserted into the pGL3-CMV-DsRed-kan plasmid. They were inserted as KpnI/BglII fragment containing the chicken lysozyme fragment, or as KpnI/BamHI human 1–68MAR fragment, upstream of the CMV promoter in pGL3-CMV-DsRed-kan, resulting in pPAG01GL3-CMV-DsRed and p1–68(NcoI) filledGL3-CMV-DsRed, respectively.

Cell culture and transfection

The CHO DG44 cell line (52) was cultivated in DMEM: F12 (Gibco) supplemented with HT (Gibco) and 10% fetal bovine serum (FBS, Gibco). Parental CHO cells AA8, NHEJ deficient cells V3.3 (53) and HR deficient cells 51D1 (54) were kindly provided by Dr Fabrizio Palitti (University of Tuscia, Italy) and were cultivated in DMEM: F12 medium with 10% FBS and antibiotics.

Transfections were performed using Lipofect-AMINE 2000, according to the manufacturer's instructions (Invitrogen). GFP or DsRed fluorescence levels were analyzed using a fluorescence-activated cell sorter (FACS), 1, 2 or 3 days post-transfection (transient transfections). Stable pools of CHO-DG44 cells expressing GFP and/or DsRed were obtained by cotransfection of the resistance plasmid pSVpuro (Clontech). After two weeks of selection with 5 μg/ml puromycin for CHO-DG44 (8 μg/ml puromycin for AA8, V3.3 and 51D1), cells were analyzed by FACS.

Cell subjected to multiple transfections were treated as follows: after the first transfection, the cells were transfected a second time according to the protocole detailed above, except that the resistance plasmid carried a different resistance gene (pSV2 neo, Clontech). The two transfections were at a 21 h interval, to be in phase with the cell cycle progression, unless otherwise indicated in the text. Twenty-four hours after the second transfection, cells were passaged and selected with 250 μg/ml G418 and/or 2.5 μg/ml puromycin (250 μg/ml G418 and 4 μg/ml puromycin for AA8, V3.3 and 5A1D1). After 3 weeks of selection, GFP and/or DsRed expression was analyzed by cytofluorometry.

Fluorescence activated cell sorting

Transient expression of eGFP and DsRed proteins was quantified at 24, 48 or 72 h after transfection using a FACScalibur flow cytometer (Becton Dickinson), whereas expression of stable cell pools was determined after at least 2 weeks of antibiotic selection. Cells were washed with PBS, harvested in trypsin-EDTA, pooled, and resuspended in serum-free synthetic ProCHO5 medium (Cambrex corporation). Fluorescence analyses were acquired on the FACScalibur flow cytometer (Becton Dickinson) with the settings of 350 V on the GFP channel (FL-1) and 450 V on the DsRed channel (FL-3) for transient expression. Settings of 240 V for FL-1 and 380 V for FL-3 were used to analyze stable expression. 100 000 events were acquired for stable transfections and 10 000 for transient transfections. Data processing was performed using the WinMDI 2.8 software.

Cell cycle analysis

At the indicated times, the cell cycle status was analyzed by flow cytometry after staining the nuclear DNA with propidium iodide (PI). Cells were first washed with a (PBS), trypsinized and harvested in 1 ml of growth media by centrifugation for 5 min at 1500 rpm (250 *g*) in a microcentrifuge. After an additional PBS wash, cells were resuspended in 1 ml of PBS before fixing with ethanol by the addition of 500 μl of cold 70% ethanol dropwise to the cell suspension under agitation in a vortex. Samples were incubated for 30 min at −20°C and cells were centrifuged as before. The resulting cell pellet was resuspended in 500 μl of cold PBS, supplemented with 50 μg/ml of RNaseA and the DNA was stained with 40 μg/ml of PI for 30 min in the dark. Cells were then washed with PBS, centrifuged and resuspended in 500 μl of ProCHO5 medium (Cambrex corporation) before analysis in a FACScalibur flow cytometer (FL-3 channel; Becton Dickinson). Ten thousand events were acquired for each sample.

Fluorescent in situ hybridization

Fluorescent *in situ* hybridization (FISH) were performed as previously described in Derouazi *et al.* (55) and Girod *et al.* (21). Briefly, metaphase chromosomal spreads were obtained from cells transfected with or without the 1–68 human MAR and treated with colchicine. FISH was performed using hybridization probes prepared by the direct nick translation of pSV40GEGFP plasmid without the MAR.

Isolation of nuclei and DNA

Nuclei were isolated 1, 2 or 3 days after transient transfection(s), from proliferating and confluent CHO DG44 cells grown in 6-well plates. Cells (1×10^6) were washed twice with cold PBS, resuspended in 2 volumes of cold buffer A [10 mM HEPES (pH 7.5), 10 mM KCl, 1.5 mM $Mg(OAc)_2$, 2 mM dithiothreitol] and allowed to swell on ice for 10 min. Cells were disrupted using a Dounce Homogenizer. The homogenate was centrifuged for 2 min at 2000 rpm (370 *g*) at 4°C. The pellet of disrupted cells was then resuspended in 150 μl of PBS and deposited on a cushion of buffer B [30% sucrose, 50 mM Tris–HCl (pH 8.3), 5 mM $MgCl_2$, 0.1 mM EDTA] and centrifuged for 9 min at 3500 rpm (1200 *g*). The pellets of nuclei were resuspended in 200 μl of buffer C [40% glycerol, 50 mM Tris–HCl (pH 8.3), 5 mM $MgCl_2$, 0.1 mM EDTA] and stored frozen at −80°C until use (56).

Total cell DNA was isolated from CHO DG44 stable cell pools or from isolated cell nuclei using the DNeasy Tissue Kit from Qiagen. For stable cell pools, 1×10^6 confluent CHO DG44 cells growing in 6-well plates were collected. DNA extraction was performed according to the manufacturer's instruction for the isolation of total DNA from cultured Animal cells. DNA isolation was performed on frozen pellets of isolated cell nuclei, which were first thawed and centrifuged at 1700 rpm (300 *g*) for 5 min to remove buffer C before beginning DNA extraction following the same protocol as for the isolation of DNA from stable cell lines.

Transgene copy number determination and quantitative PCR

To determin the copy number of transgenes integrated in the genome, ~6 ng of genomic DNA were analyzed by quantitative PCR using the SYBR Green-*Taq* polymerase kit from Eurogentec Inc and ABI Prism 7700 PCR machine. The following primers were used to quantify the GFP gene: GFP-For: ACATTATGCCGGACAAA GCC and GFP-Rev: TTGTTTGGTAATGATCAGCAA GTTG, while primers GAPDH-For: CGACCCCTTCAT-TGACCTC and GAPDH-Rev: CTCCACGACATACTC AGCACC were used to amplify the GAPDH gene. The ratios of the GFP target gene copy number were calculated relative to that of the GAPDH reference gene as described previously (57). To determine import of the transgene into nuclei following transfection, quantitative

PCR was performed on DNA extracted from purified nuclei using the same GFP and GAPDH primer pairs as described above.

The number of GAPDH gene and pseudogene copies used as reference was estimated for the mouse genome, as the CHO genome sequence is not available as yet. Alignment to the mouse genome of the DNA sequence of the 190 bp amplicon generated by the GAPDH primers was performed using NCBI BLAST software. A number of 88 hits was found per haploid genome. As the CHO DG44 are near-diploid cells (55), we estimated that 176 copies of the GAPDH genes and pseudogenes occur in the genome of CHO DG44 cells. This number was used as a normalization standard for the quantification of the GFP transgene copy number.

Confocal microscopy

pGEGFPcontrol and p1-68(NcoI filled)SV40EGFP plasmids were labeled either with rhodamine by the *Label* IT Tracker TH-Rhodamine Kit or with Cy5 by the *Label* IT Tracker Cy 5 Kit (Mirus, Mirusbio) according to the manufacturer's protocol, and purified by ethanol precipitation. For transfection, DNA transfection was carried out with the Lipofectamine 2000 reagent (Invitrogen) according to the supplier's instructions. At 3, 6 and 21 h after transfection, the medium was removed and the cells were fixed with 4% paraformaldehyde at room temperature for 15 min. When indicated, cells were treated for 30 min with LysoTracker Red DND-99 (Molecular Probes, Invitrogen) at a final concentration of 75 nM before fixation, to stain the acidic organelles (e.g. endosomes and lysosomes) according to the manufacturer's instructions. The fixed cells were then washed twice with PBS and mounted in a DAPI/Vectashied solution to stain the nuclei.

Fluorescence and bright-field images were captured using a Carl Zeiss LSM 510 Meta inverted confocal laser-scanning microscope, equipped with a 63× NA 1.4 planachromat objective. *Z*-series images were obtained from the bottom of the coverslip to the top of the cells. Each 8-bit TIFF image was transferred to the ImageJ software to quantify the total brightness and pixel area of each region of interest. For data analysis, the pixel areas of each cluster in the cytosol s_i(cyt), nucleus s_i(nuc) and lysosome s_i(lys) were separately summed in each XY plane. Theses values [$S'_{Z=j}$(cyt), $S'_{Z=j}$(nuc) and $S'_{Z=j}$(lys), respectively] were further summed through all of *Z*-series of images and denoted *S*(cyt), *S*(nuc) and *S*(lys), respectively. The total pixel area for the clusters of labeled pDNA in the cells, *S*(tot), was calculated as the sum of *S*(cyt), *S*(nuc) and *S*(lys). The fraction of pDNA in each compartment was calculated as $F(k) = S(k)/S(\text{tot})$, where represents each subcellular compartment (nucleus, cytosol or lysosome).

RESULTS

Effect of iterative transfection on transgene expression

Previous work has led to the screening of human MARs to identify one, termed MAR 1–68, that was found to potently increase and stabilize gene expression in cultured cells and in mice when inserted upstream of the promoter/enhancer sequences (21,39). Co-transfection of a GFP expression vector and an antibiotic resistance plasmid, followed by antibiotic selection of cells having stably integrated the transgenes in their genome, typically yields a bimodal distribution of the fluorescence in polyclonal cell populations when analyzed by flow cytometry (Figure 1A). A first cell subpopulation, which overlaps the *y*-axis in this experimental setting, corresponds to cells expressing GFP at undetectable levels, while another subpopulation of cells expresses significant GFP levels. Inclusion of MAR 1–68 increased the level of expression from fluorescent cells and concomitantly reduced the proportion of silent cells (15 versus 36%, Figure 1B). The increase in expression did not depend on the use of the viral SV40 promoter, as it was also obtained when expressing GFP from the the cellular glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene promoter (Supplementary Figure S1).

We next tested whether two consecutive co-transfections might further increase GFP expression from the MAR-containing plasmids (Figure 1A). When the same GFP expression vector was co-transfected again 2 weeks later with a distinct antibiotic resistance gene, a 2.4-fold increase of fluorescence was observed on average after selection for resistance to the second antibiotic, which is close to the expected 2-fold increase (Figure 1A and C). In contrast, an unexpectedly higher (4- to −5-fold) increase of GFP expression was observed from two successive transfections performed on consecutive days followed by selection with both antibiotics. When averaging over all cells of the polyclonal population, a 20-fold increase of expression was gained by successive transfections of MAR-containing plasmids when compared to a single transfection without a MAR (Figure 1C). Furthermore, some of the cells displayed very high levels of expression, and the occurrence of silent cells was almost fully abrogated from the polyclonal population (0.5%, Figure 1B). Consecutive transfections without a MAR yielded modest GFP expression, resulting in a 3.2-fold average increase of the overall fluorescence level when compared to a single transfection, and it did not abrogate the occurrence of silent cells (Figure 1C and data not shown). Thus, the presence of the MAR and the iterative transfection act synergistically to mediate elevated expression levels.

Overall, the expression levels obtained from the two consecutive transfections of MAR-containing plasmids were so high that the GFP fluorescence could be readily seen from the cell culture monolayers in the daylight, without excitation with UV light (Supplementary Figure S2A). This effect was not limited to the human MAR 1–68, as both the relatively less potent chicken lysozyme MAR (cLysMAR) and MAR X-29, a potent MAR isolated from human chromosome X (21), yielded an increase in expression when comparing double to single transfections (Supplementary Figure S3). This indicated that the elevated expression upon successive transfections may be a general property of MAR elements. However, the cLysMAR element yielded relatively lower transgene

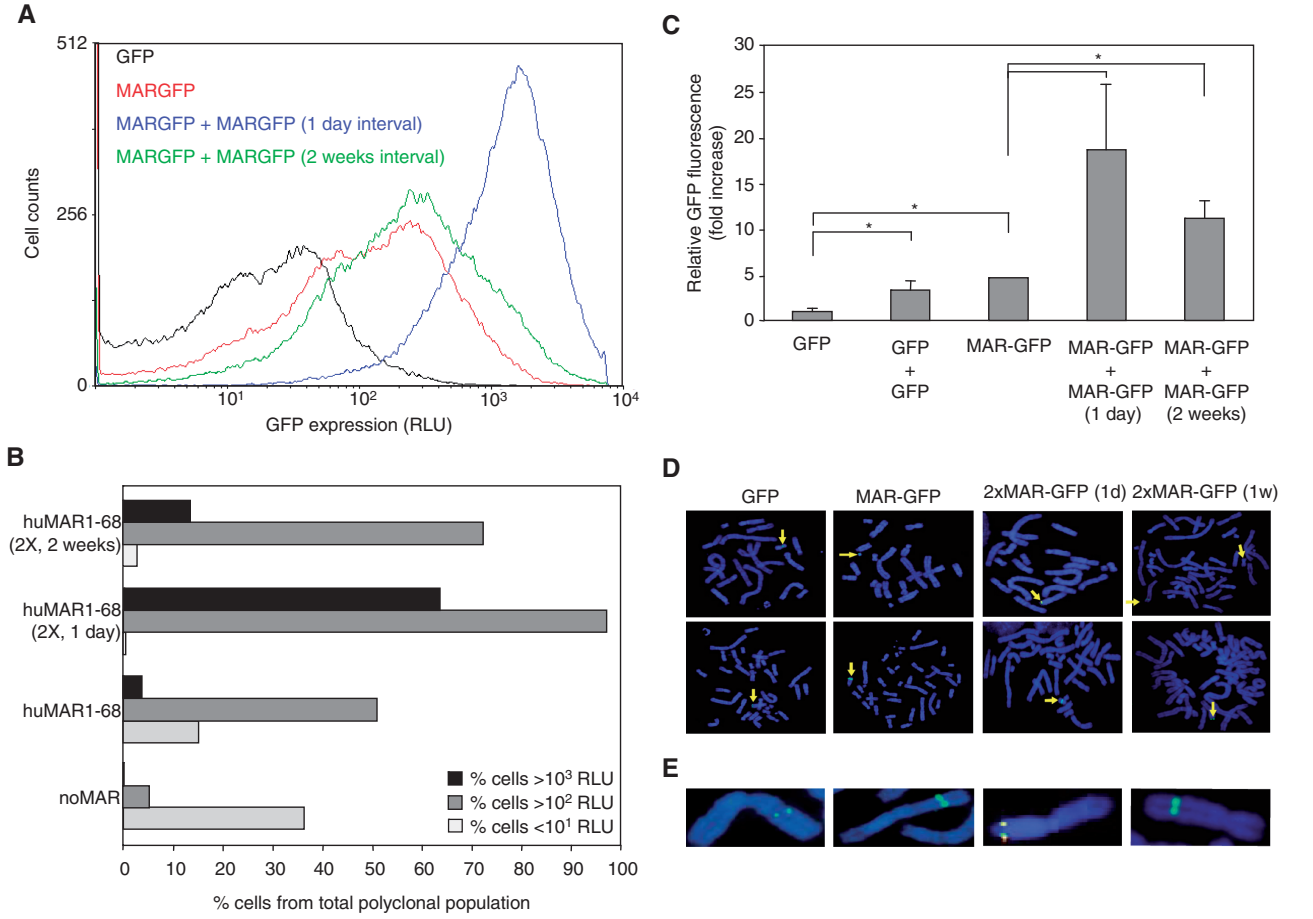


Figure 1. Analysis of the effect of MARs and successive stable transfections on gene transfer and expression. CHO DG44 cells were co-transfected with the GFP expression vector devoid of MAR element (GFP, dark line), or with the vector containing MAR 1–68 (MAR1–68GFP, red line), and with the pSVpuro plasmid mediating resistance to puromycin. Some of these cells were submitted to a second transfection with the same GFP expression vector but with a selection plasmid mediating neomycin resistance, either on the day following the first transfection (blue line) or after 2 weeks of selection for puromycin resistance (green line). After 2 weeks of selection for puromycin (single transfections), or 3 weeks of selection for both puromycin and neomycin resistance (double transfections at 1 day interval), or 2 weeks of selection for puromycin followed by the second transfection and two weeks of neomycin resistance selection (double transfections at 2 weeks interval), eGFP fluorescence was quantified by cytofluorometry. (**A**) Fluorescence distribution in polyclonal populations of GFP-expressing cells. The cell fluorescence profiles shown are representative of four independent experiments. (**B**) Histogram showing the percentage of total cells corresponding to non/low-expressors that display <10 relative light units (RLU), or cells that display medium and high (>100 RLU) or very-high (>1000 RLU) GFP fluorescence, as determined from the analysis of stable cell pools as shown in panel A. (**C**) The mean GFP fluorescence of each stable polyclonal cell pool was normalized to that obtained from the transfection of MARGFP and the average and standard deviation of four independent transfections is shown as a fold increase over the fluorescence obtained by one transfection without a MAR. Asterisks indicate significant differences in GFP expression (Student's *t*-test, $P < 0.05$). (**D**) FISH analysis of eGFP transgene chromosomal integration sites in cells singly or doubly transfected with or without the human MAR. Metaphase chromosomes spreads of stable cell pools were hybridized with the GFP plasmid without MAR, and representative illustrations of the results are shown. (**E**) Enlargements of chromosomes are shown to illustrate differences in fluorescence intensities.

activation after a single transfection, and the increase in expression observed after the successive transfections was comparably lower. The converse was true of MAR X-29, indicating that MARs ability to mediate each of these effects may vary similarly.

The effect of the double transfection of MAR-containing plasmids was not dependent either on the GFP transgene or on the SV40 promoter used to express GFP, as similar results were obtained when a CMV promoter was used to express the DsRed reporter gene or the immunoglobulin light and heavy chains (Supplementary Figure S2B and data not shown). Interestingly, the very high levels of immunoglobulins expressed by monoclonal CHO cell clones often correlated with an increased cell division time. This indicates that the cells were likely reaching their physiological limits in terms of protein synthesis. This may be expected, as cells synthesizing similar amounts of the recombinant protein as compared to their own cellular proteins (~100 pg per cell) should double the energetic input required for each cell division. Nevertheless, a large proportion of clones were found to express the heterologous protein at very high levels without interfering with their own metabolism, as they did not slow down cell division significantly (Supplementary Figure S2B).

Cointegration of transgenes upon iterative transfections

An important parameter driving high expression upon iterative transfection was found to be the time interval between the two transfections. The synergistic effect on expression was not observed when re-transfecting cells after 2 weeks. Rather, the two transfections behaved as two independent and thus additive events (Figure 1C). This suggested that the plasmid DNAs from each transfection may have to interact as episomes within the nucleus and potentially form mixed concatemers before co-integrating into the cell genome. This possibility was assessed by FISH analysis of metaphase chromosomal spreads from stable polyclonal populations. Eighty individual metaphases of cells transfected once either with or without the MAR element were hybridized with a probe consisting of the GFP plasmid without a MAR. A single integration site was observed, but higher fluorescence intensities were observed from cells transfected with the MAR (Figure 1D and E). Fluorescence intensity was further increased by the double transfection process, suggesting that a higher number of transgene copies had integrated. Unique integration sites were noted in all cases after single or two consecutive transfections. However, double integration events were observed in approximately half of the cells transfected twice at an interval of 1 week, when little episomal DNA should remain from the first transfection. This indicates that independent integration events may occur if DNA integration from the first transfection has been completed before the second transfection is performed. Double transfections did not lead to apparently increased aneuploidies nor to detectable chromosomal rearrangements, and they did not detectably lead to insertions at a preferred chromosomal locus, as none of the analyzed cells had an identical integration site. Thus, transgene integration upon two transfections does not appear to be targeted to any specific chromosomes or chromosomal sites, as reported earlier for single transfections of MAR-containing plasmids (21).

High transgene expression requires phasing of the cell cycle and transfections

As timing between transfections seemed to play a role in high transgene expression, we analyzed the effect of systematic variations of the time interval. The highest GFP expression level was observed when the second transfection was performed 21 h after the first one, yielding consistently a 5-fold increase of fluorescence as compared to a single transfection. When the second transfection was performed after 18, 24 and 27 h, a 3- to 3.5-fold increase of expression was obtained as compared to a single transfection, but these were significantly lower than those obtained after 21 h (Figure 2A). As this timing is close to the duration of the first cell division cycle after cell passaging (Supplementary Figure S4), this suggested that high transgene expression upon consecutive transfections might be linked to particular phases of the cell division cycle.

The distribution of the cells along the division cycle was determined by PI staining of the DNA. This analysis indicated an over-representation of cells at the G1 phase 18 h after cell passaging, and this was found to correspond to the timing that yields the highest expression from a single transfection (Figure 2B, Supplementary Figure S4A and B and data not shown). A similar pattern and over-representation of G1 cells was obtained 21 h after the first transfection, which again corresponds to the timing that yields the highest expression levels upon a second transfection (Figure 2B). If expression is indeed linked to cell cycle phasing, another optimum for transgene expression should be observed if a second transfection was performed at an interval corresponding to two cell divisions. After 42 h, the synergistic effect of the two transfections was lost, as expression was similar to that obtained for one transfection. However, a second, albeit lower, synergistic increase of transgene expression was observed after 48 h. The higher levels of expression observed from a first transfection at 18 h and for a second transfection performed with a 21 or 48 h interval imply that optimal DNA transfer and/or expression may occur at specific cell division stages.

Effect of MAR and consecutive transfections on cellular DNA uptake

FISH analysis suggested that elevated expression upon successive transfections may result in part from the integration of a higher number of the transgene copies in the genome (Figure 1D). Consecutive transfections at an interval of one day might lead to an increase of the concentration of plasmid episomes in the nucleus, thereby augmenting the probability of transgene integration within the cell genome. To assess the amount of transgene entering nuclei at each transfection, we performed transient single or double transfections followed by plasmid extraction from nuclei isolated 1 or 2 days after the second transfection and quantification of the transgenes by real-time quantitative PCR (qPCR). Cells doubly transfected with MAR-GFP exhibited 3.8-fold more GFP transgene copies in their nuclei than cells transfected just once with MAR-GFP (Figure 3A). When comparing cells transfected with these different plasmids expressing either GFP or DsRed, we observed that the nuclear delivery resulting from the second transfection of MAR-GFP was 4.2-fold higher than the one observed from a single transfection of this plasmid. However, the nuclear transport of the firstly transfected GFP plasmid was not increased significantly by performing a second transfection. We concluded that DNA transport to the nucleus from the second transfection is favored by performing a prior first transfection.

These conclusions were strengthened by confocal imaging of DNA transport, where plasmids used for the first transfection were labeled with rhodamine while the secondly transfected plasmids were labeled with Cy5 (red and white labels, respectively, Figure 4A). Similar numbers of rhodamine-labeled plasmid clusters were observed in cell nuclei after a first transfection with or without a MAR, which correlates well with the lack of effect of the MAR on DNA transport as assessed by qPCR (Figures 3A and 4A). Nuclear plasmid clusters

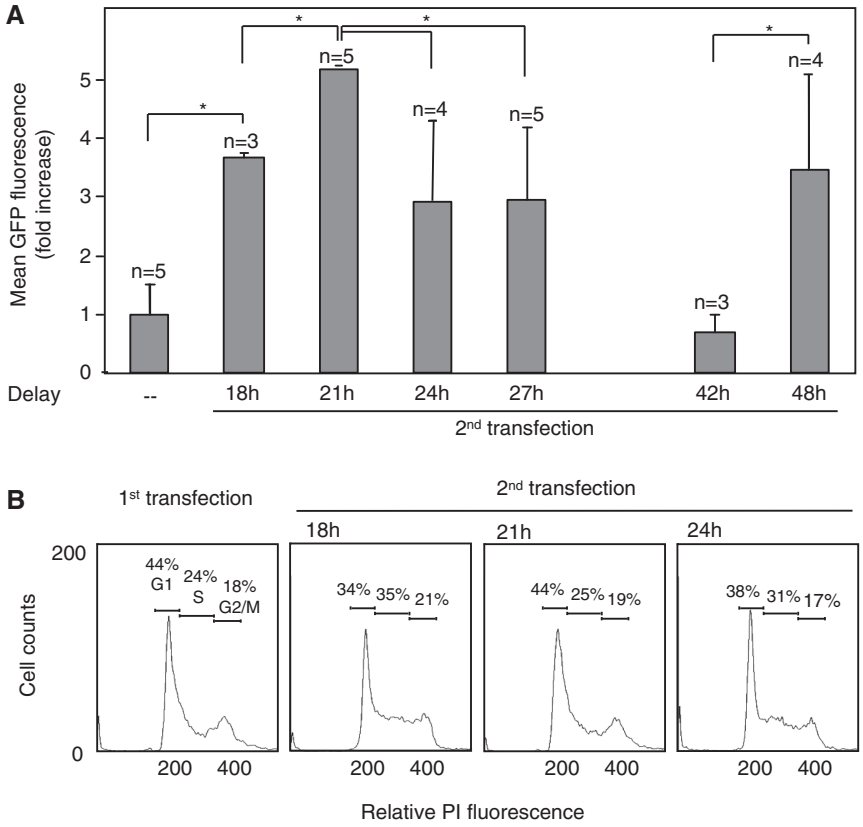


Figure 2. Determination of the optimal timing between successive transfections. (**A**) Stable polyclonal populations were generated by a single transfection (minus symbol) or by two consecutive transfections of the MAR-GFP expression plasmid with the indicated time intervals. After 2 weeks of selection, mean GFP expression of the total polyclonal populations was determined. Fluorescence levels were normalized to the maximal values obtained and they are displayed as the fold increase over the expression obtained from a single transfection where *n* corresponds to the number of independent transfections. Asterisks indicate significant differences in GFP expression (Student's *t*-test, $P < 0.05$). (**B**) Analysis of cell cycle progression. At the time of first and second transfections, CHO cells were harvested and stained with propidium iodide (PI) and fluorescence was analyzed by cytofluorometry. The distribution of relative PI fluorescence represents the amount of genomic DNA per cell. The percentage of the population associated to each cell cycle state (G1, S, G2/M) is as indicated.

were observed in essentially all the cells after two transfections. However only few cells expressed GFP, in agreement with previous observations that only a minority of cells are able to express transiently transfected genes (58).

The transport of transfected plasmid DNA in CHO cells, which is known to consist of cellular uptake, lysosomal escape and nuclear import, is limited by endosomal/lysosomal degradation (58). Thus, we next assessed the intracellular trafficking of transfected plasmid DNA by quantifying its distribution in cellular organelles and in the cytosol after each transfection, after specific staining of the endosomal/lysosomal and nuclear compartments to distinguish them from the cytosol. Results summarized in Figure 4B show a similar subcellular distribution of plasmid DNA with or without MAR 21 h after a first transfection, although nuclear transport of MAR-containing plasmids seemed somewhat faster at the earlier time points. Performing a second transfection of the MAR-devoid plasmid did not yield an improved nuclear transport. However, plasmids bearing a MAR element escaped lysosomal retention and entered nuclei much more efficiently, as 80% of the total Cy5-labeled pDNA was located in the nuclei in presence of the MAR 21 h after the second transfection, as compared to <40% of the plasmid devoid of MAR (Figure 4B). Rather, most of the MAR-devoid plasmid ended up in the lysosomal/endosomal compartment, as found also for the first transfection. The unexpected finding of a cooperative effect of the MAR and of the iterative gene transfer on lysosomal escape thus provides an explanation for the increased concentration of plasmids in isolated nuclei (Figures 3A and 4B). The reason for this phenomenon is unclear at present, but it might result in part from the saturation of the cellular degradation compartments by the DNA of first transfection, thus allowing plasmids of the second transfection to remain in the cytoplasm where the MAR may promote plasmid transport into the nucleus.

MAR elements increase the copy number of genome-integrated transgenes

Next, we examined whether the increased transport of plasmid DNA elicited by the MAR and the consecutive transfections may increase transgene integration into the genome of CHO cells. Stable polyclonal cell populations

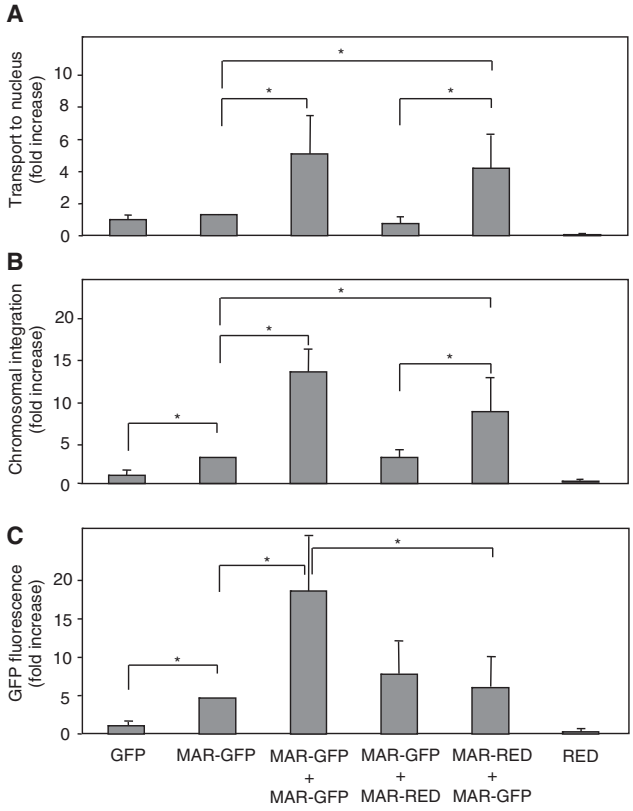


Figure 3. DNA transport, integration and expression upon successive transfections. (**A**) Amount of GFP transgenes transport into cell nuclei during single and double transient transfections with GFP or DsRed plasmids with or without a MAR. MAR-GFP + MAR-RED corresponds to a double transfection where MAR-GFP is transferred during the first transfection, whereas MAR-RED was used in the second transfection. Nuclei were isolated and total DNA was extracted one day after a single or after the second transfection, and the number of GFP transgenes transported into the nuclei was quantified by qPCR. Results were normalized to that of the reference CHO cell genomic GAPDH gene and represent the mean of four independent transfections. (**B**) Effect of the MAR and successive transfections on integrated GFP transgene copy number. Total genome-integrated transgene DNA was extracted from the previously described GFP-expressing cells after 3 weeks of selection of stable polyclonal cell pools, and DNA was quantified as for panel A. (**C**) Effect of MAR and successive transfections on GFP expression. The GFP fluorescence levels of the stable cell pools analyzed in panel B were assayed by cytofluorometry.

were selected as for Figure 1, and the average numbers of stably integrated GFP transgene copies per genome were determined on total cell DNA using qPCR. Inclusion of a MAR element in transfected plasmids significantly increased the number of transgenes integrating in the genome of stable cell pools (Figure 3B). Because the MAR does not act to increase nuclear transport after single transfections (Figure 3A), we concluded that the MAR may increase genomic integration of the plasmid *per se*. This finding supports previous indications that the use of MARs may increase the number of transgene copies that integrate in the genome of recipient cells (51,59).

Successive transfections also mediated a 4-fold increase of plasmid integration, which is commensurate to the increase in free episomes noted in transient transfections (Figure 3A and B). We estimated that 48 GFP plasmid copies had integrated on average when transfecting once without a MAR, while ~163 copies and 676 copies on average were obtained from one or two successive transfections with the MAR, respectively. Overall, the increased nuclear transport synergistically elicited by both the MAR and the successive transfections yielded a >10-fold increase in transgene copy number when combined to the MAR-driven increase of plasmid integration. Note, however, that the double transfection of MAR-containing plasmids yielded yet an even higher increase in transgene expression (~20-fold, Figure 3C). This indicated that increased expression did not result solely from increased transgene copy number, but that it must also stem from a MAR-mediated increase of expression per integrated transgene copy, as expected from the previously observed antisilencing and transcription activation effects of this MAR element (39).

When assessing GFP expression and transgene copy number in individual cell clones isolated from the polyclonal populations, a correlation was found between transgene expression and copy number, in that higher levels of integration and of expression were observed from single or multiple transfections of MAR-containing plasmids (Supplementary Figure S5A). Furthermore, no significant decrease of expression could be detected from MAR-containing clones having co-integrated very high numbers of transgene copies and MARs (Supplementary Figure S5B), and stable elevated expression was maintained upon adaptation of several of these clones to growth in suspension and further culturing (data not shown). These results indicated that the MAR was able to prevent inhibitory effects that may result from the repetitive nature of the co-integrated plasmids and/or from antisense transcription, an effect that can be attributed to the potent anti-silencing properties of this MAR element (39). However, the average levels of expression did not match perfectly with copy number, as noted when analyzing individual cell clones (Supplementary Figure S5A), or when comparing GFP expression from the firstly or secondly transfected DNA, in co-transfection experiments with the dsRED vector (Figure 3B and C). We therefore conclude that the enhanced transgene expression observed after two successive transfections of MAR-containing plasmids can be explained in part by the improved nuclear import and genomic integration of the plasmid DNA, which results in increased transgene copy number, as well as by the lack of silencing and higher transgene transcription. However, other effects may also influence transgene expression in individual clones depending on the transfection history and conditions, clonal fitness heterogeneity and/or possible effects pertaining to the genomic integration locus.

Effects of DNA homology on plasmid integration and expression

As the high GFP fluorescence observed from successive transfections of MAR-containing plasmids results in part from the increased transgene integration at a single

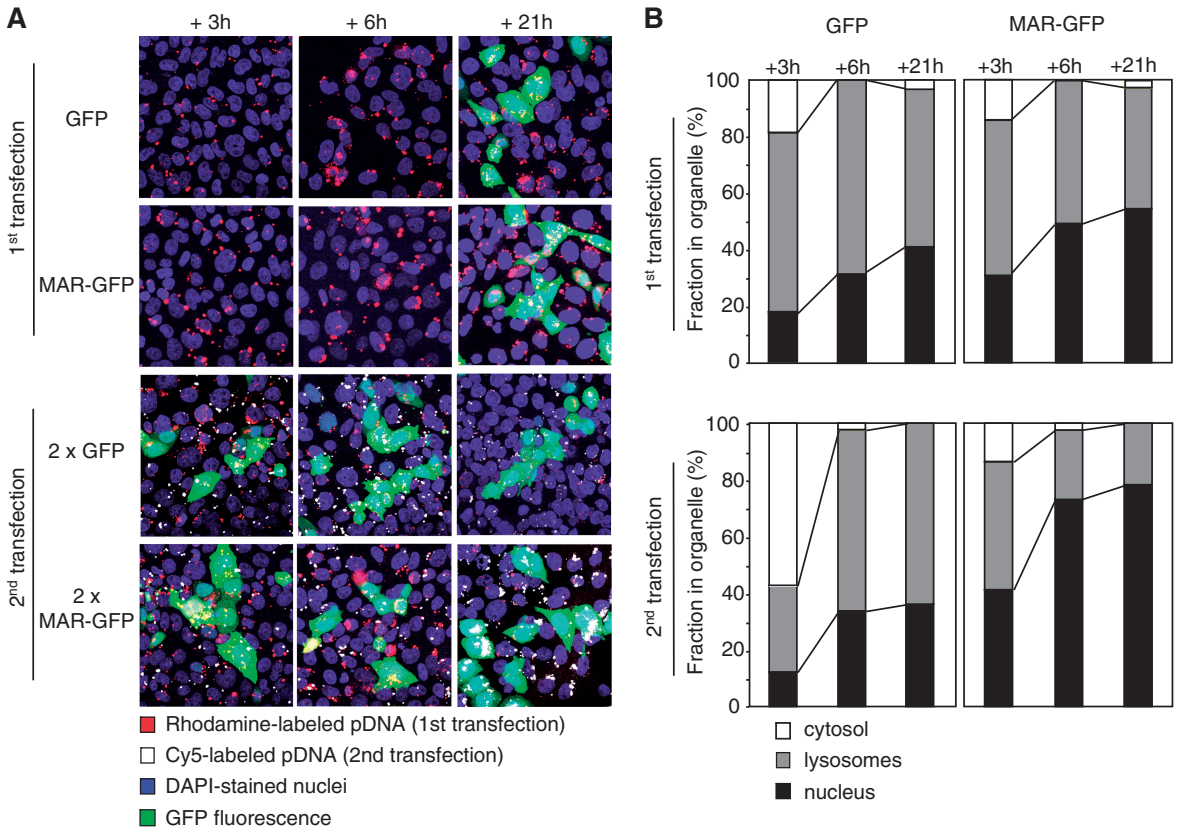


Figure 4. Subcellular distribution of transfected DNA. (**A**) Confocal microscopy analysis of DNA intracellular trafficking. Transient single or double transfections were performed in CHO cells using plasmids bearing or not a MAR labeled with Rhodamine and Cy5 fluorophores, as indicated. Transfected cells were fixed and stained with DAPI (blue) 3, 6, 21 h post-transfections. Cells expressing GFP appear in green on the pictures. (**B**) Quantification of the subcellular plasmid DNA distribution was performed on confocal laser microscopy performed for panel A, except that endosome/lysosome compartments were stained with LysoTracker Red DND-99. The pixel area of clusters derived from rhodamine or Cy5 fluorescence were used to estimate the amount of plasmid DNA in ~120 cells.

chromosomal locus, we next examined the molecular basis of this effect. A possibility is that the integration of a MAR-containing plasmid during the first transfection might promote secondary integration at the same genomic locus during the second transfection. This may be expected from the ability of the MAR to maintain chromatin in an accessible state and thus to provide proper targets for HR. Alternatively, the high number of integrated transgenes may result from a more efficient concatemerization of the plasmids introduced during both transfections, as may be mediated by the high concentration of episomes found in the nucleus. Indeed, HR was proposed to mediate the formation of large concatemers of transfected plasmids (60), which may lead to the co-integration of multiple plasmid copies upon recombination with the genomic DNA. In the latter model, HR may occur between similar DNA sequences on the plasmids used during the first and second transfections, and thus the efficacy of transgene integration and expression should critically depend on DNA sequence homologies.

This latter possibility was first assessed by analyzing the effect of plasmid homology on transgene expression by performing successive transfections with different combinations of transgenes (GFP or DsRed), plasmid backbones (ampicillin or kanamycin baterial resistance) and/or MARs (chicken lysozyme MAR or the human MAR 1–68). Transfection of distinct MARs, transgenes, or bacterial resistance all decreased the high expression normally observed with successive transfections (Figure 5A). The double transfection effect was almost fully abolished when using different MARs, transgenes and vector elements (MAR_1-GFP + MAR_2-RED constructs), suggesting that plasmid homology is required to achieve high expression from successive transfections.

HR is often elicited as a DNA repair mechanism of double-stranded breaks (DSB), in a process that was termed homologous recombination repair (61,62). Thus, we tested whether plasmid linearization prior to transfection mediates the high expression obtained from successive transfections. A more than additive increase of transgene expression was also observed with circular plasmids, however, the overall expression was lower than that obtained using linear plasmids (Supplementary Figure S6). This is consistent with the increased recombinogenic properties of linear DNA in HR processes (63).

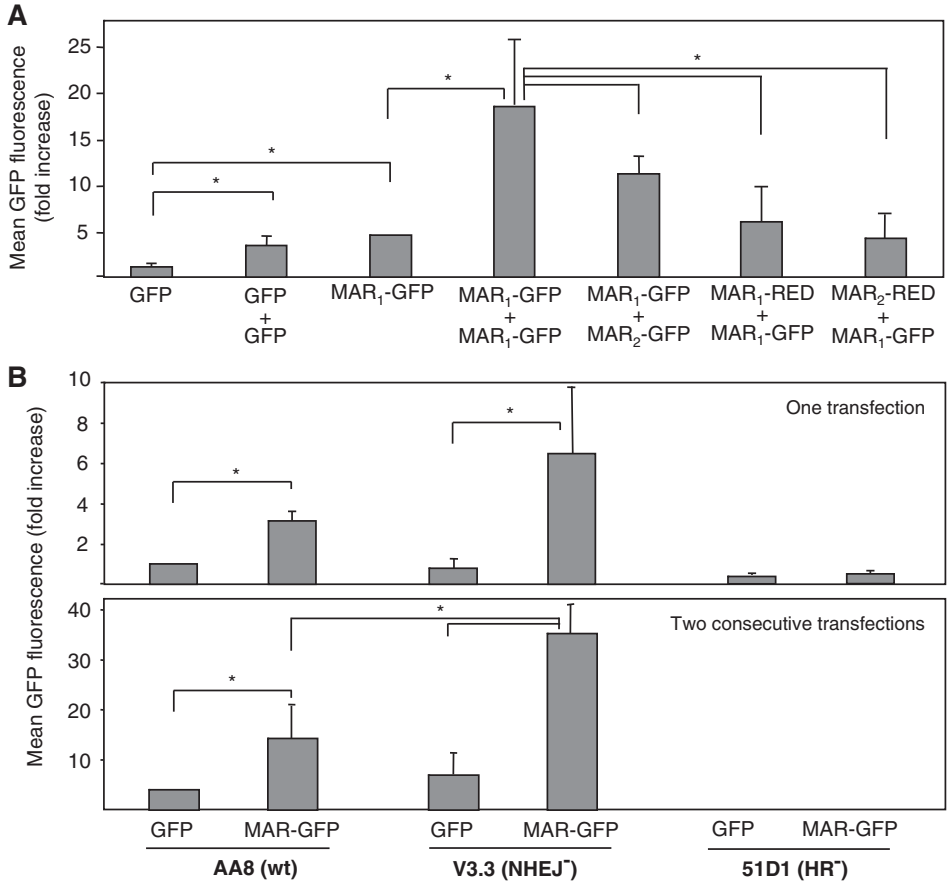


Figure 5. The MAR, plasmid homology and homologous recombination mediate high transgene expression. (**A**) Stable polyclonal cell pools were generated by the transfection of plasmids bearing different transgenes (GFP or DsRed), MAR (MAR1 for the human 1–68 or MAR2 for the chicken lysozyme MAR), and/or bacterial resistance gene (ampicillin or kanamycin), and the relative average GFP fluorescence of four independent transfections are shown as the fold increase over that obtained from one transfection without MAR. Asterisks show significant differences in GFP expression (Student's *t*-test, $P < 0.05$). (**B**) Stable transfections with GFP or MAR1-68GFP plasmids were performed in the parental CHO cell line (AA8) and in mutants deficient either in the homologous recombination (51D1) or non-homologous end-joining (V3.3) pathway. The mean GFP fluorescence of each stable polyclonal cell pool generated from single (top panel) or two consecutive (bottom panel) transfections were normalized to that obtained from AA8 cells singly transfected with the MAR-devoid plasmid. Asterisks indicate significant differences in GFP expression (Student's *t*-test, $P < 0.05$). No stably transfected cells were obtained form the double transfection of 51D1 cells.

HR mediates increased expression

The requirement of plasmid homology and double-strand breaks to achieve the higher expression levels upon the double transfection implied that HR may be involved. Transgenes were proposed to integrate into the cell genome using two families of antagonistic pathways, termed non-homologous end-joining (NHEJ) or HR. These pathways are more active during specific phases of the cell cycle, as exemplified by HR, which is used to repair DNA damages during or after DNA replication in the S and G2/M phases of the cell cycle (64). Cells lacking classical NHEJ genes show a double-stranded break repair biased in favor of HR, suggesting that these two major pathways normally compete to repair these DNA lesions (65). Thus, one way to activate HR is to suppress or genetically inactivate NHEJ, as seen in yeast and mammalian cells (65–68). A possible implication of HR-related mechanisms in the increased transgene expression that results from the MAR and/or successive transfections was thus directly assessed using CHO cell lines mutated in a key component of either pathways, and which are thus only competent for either HR or NHEJ.

The 51D1 CHO mutant derivative lacks the RAD51 strand transferase and is thus deficient in HR, while V3.3 CHO cells lack the catalytic activity of the DNA-dependent protein kinase (DNA-PK) that plays an essential role to initiate the NHEJ pathway (13,54,69). In the wild-type parental cell lines (AA8), the MAR mediated a 3-fold increase of the overall GFP fluorescence in the polyclonal population (Figure 5B). However, few stably transfected colonies survived after selection for antibiotic resistance in the 51D1 cell line and GFP expression remained very low. In contrast, an exacerbated MAR-driven activation of transgene expression was observed in NHEJ-deficient cells, resulting in a >6-fold increase of transgene expression when compared to cells transfected once with the GFP expression vector without MAR.

Similar trends were noted for successive transfections, in that GFP expression from V3.3 cells was increased in

the presence of the MAR as compared to the parental AA8 cells (Figure 5B, note the different scales of the top and bottom panels). Inactivation of the NHEJ pathway had little effect on the expression of the MAR-devoid plasmid but it further exacerbated the expression increase elicited by the MAR, indicating that presence of the MAR and high HR activity are both necessary to obtain very high transgene expression. Cells deficient in HR did not yield stable colonies from the double transfection, demonstrating the requirement of the HR pathway in the efficient integration and maintenance of transgenes in the successive gene transfer process. Analysis of the number of integrated transgene copies revealed a 25-fold higher number of integrated transgene copies upon the double transfection of MARGFP in NHEJ-deficient V3.3 cells (Supplementary Figure S7). However, a relatively larger (over 35-fold, Figure 5B) increase in expression was obtained from two consecutive transfections of the V3.3 cells with the MAR. Again, this indicated that the observed MAR-mediated increase in expression results both from an increase in transgene genomic integration and in an augmentation of the expression per transgene copy.

DISCUSSION

The variability in gene expression among independently transformed cells or organisms is well documented. Inconsistent expression levels have been associated to the variable number of genes that have stably integrate in the host cell genome, to properties of the sites of integration, and/or to the gene transfer procedures. Non-viral gene transfer remains characterized by variable expression efficiencies and by an uncertain outcome in terms of expression levels and stabilities. Gene amplification has been used to augment the copy number and hence expression of the exogenous genes (70), however this often leads to unstable expression when the selection reagent is removed (71). An alternative has been to optimize expression vectors by inserting synthetic or natural regulatory sequences that increase and/or stabilize expression, such as MAR elements (39). However, even in the most favorable conditions, transfections have lead to the occurrence of cells that integrate the transgene in their genome but express it at low levels. Hence, the identification of stable cell lines is usually associated with the tedious isolation and characterization of many clones to identify one with the desired expression properties. Here, we show that very high levels of transgene expression can be consistently obtained in nearly all cells of a polyclonal cell population after successive transfections of MAR-containing constructs. Our results further indicate that efficient gene transfer and expression of MAR-driven vectors require a functional HR or repair pathway.

In this study, we show that MAR elements can act in part to increase the number of copies of exogenous genes that integrate in the genome, substantiating previous non-quantitative observations (21,59). However, the intuition that a high copy number always supports stronger expression is often non-valid, as the presence of multiple gene copies co-integrated at one or a few loci of the host genome has been reported to favor silencing. The propensity of repeated elements to pair and assemble in heterochromatin or to generate double-stranded and/or small interfering RNAs from antisens transcription of adjacent transgenes is a frequent cause of gene silencing (6,72). Here, the copy number of integrated transgenes and cell fluorescence levels were shown to correlate well when comparing single or double transfections performed in the presence of the MAR. We find also that the MAR-mediated increase in transgene expression results in part from the integration of more transgene copies in the host cell genome. However, we also find that the relative increase of transgene expression is higher than the increase of transgene copy number upon inclusion of the MAR. This indicates that MAR 1-68 acts both to increase transgene integration and to increase expression per transgene copy. Overall, we estimate that increased integration mediates ~70% of the observed effect on the increase of expression, while the augmentation of transgene transcription per se, as shown to occur in presence of MAR 1-68 (39), accounts for the remaining 30% of the increase of expression observed upon successive transfections. Thus, both effects concur to mediate the very high expression levels observed after successive transfections of MAR-containing plasmids.

We also find that transgene genomic integration is synergistically increased by successive gene transfers and by the MAR. Successive transfections of MAR-containing plasmids result in improved plasmid transport to the nucleus, as explained by the decreased targeting of the second DNA load to the degradation compartment. While the greater amounts of episomal plasmids in the nucleus may readily explain the increased co-integration of independent DNA loads in successive transfections (Supplementary Figure S8), we find that the MAR-mediated increase in transgene integration does not result from an increased plasmid targeting or transport to the nucleus. This indicated that the MAR may directly promote the recombination of the exogenous DNA with the host cell genome.

In vertebrates, HR and NHEJ differentially contribute to repairing abnormal DNA structures such as double-stranded breaks, depending on the nature of the DNA damage and the phase of the cell cycle (64). This study shows that plasmid integration critically depend on a functional HR pathway, as inactivation of Rad51, a key initiator of HR (73), nearly abolishes transgene integration and expression. Conversely, inactivation of the antagonistic NHEJ pathway, which is associated to increased HR, concomitantly increases the integration and expression of MAR-containing plasmids, but not of the MAR-devoid control. In addition, increased integration and expression upon successive MAR transfections require the plasmids to have homologous DNA sequences. Furthermore, the cells must be in the G1 cell cycle phase at the time of transfection, and we find that DNA transport to the nucleus is nearly completed within 6 h after DNA transfer, at which time cells have progressed to exit S and enter G2 (see Supplementary Data), at which time HR is most active. Taken together, these findings provide strong

support for the involvement of HR in the increased integration, and, consequently, the higher expression observed in presence of the MAR and/or upon double transfections. They also suggest that plasmid integration by a HR repair pathway may be increased in the presence of a MAR element on the DNA. However, a direct demonstration of the role of MAR elements in activating HR will require further experimentations.

Mammalian cells contain the enzymatic machinery required to mediate recombination between newly introduced plasmid DNA molecules, and HR between co-injected plasmid molecules is an efficient process in cultured mammalian cell lines, approaching 100% of the molecules (60). Plasmid concatemers thereby formed may integrate at one or a few sites in the host chromosome, the integration site being different in independently transformed cells (74). The orientation of the copies within the concatemer is not random, but usually organized as tandem head-to-tail arrays, as generated by the HR of independent plasmid copies in the cell (75). HR between the newly introduced DNA and its homologous chromosomal sequence has been reported to occur infrequently in the absence of MAR element, at a frequency of 1:1000 cells receiving DNA (76). However, estimations based on transgene expression, such as antibiotic selection, may significantly underestimate the true frequency of integration. Whether plasmid concatemers may be able to undergo HR with previously integrated transgene copies, or whether a single integration event of a large plasmid concatemer occurs, remains difficult to assess experimentally. In any case, the contribution of MAR elements to promote HR proposed in this study readily provides an explanation to the previously observed MAR-mediated increase in the occurrence of cells that stably integrate the transfected genes, and thus in the number of antibiotic-resistant colonies (21).

The requirement of a functional HR pathway to mediate efficient integration of MAR-containing constructs implies that these DNA elements might be preferred sites of homologous pairing and DNA recombination. A number of molecular mechanisms have been proposed to explain the mode of action of MARs. For instance, they may reduce the occurrence of lowly-expressing cells by protecting transgenes from silencing effects linked to integration at heterochromatic loci (7). MARs may recruit regulatory proteins that modify chromatin to adopt a more transcriptionally permissive state by mediating histone hyperacetylation, they may change subnuclear localization of the transgene, or they may facilitate the association of transcription factors that activate gene transcription (2,8,36,39). Although not mutually exclusive, our results indicate that a significant part of the effect of MAR elements on transgene expression might in fact be attributed to increased transgene integration into the host cell genome by HR.

The proposed role of MARs as HR-promoting genomic signals might result from their action on chromatin and/or on DNA accessibility, thereby providing an access to DNA binding proteins such as topoisomerases (48). As reported in this study and by prior reports, all cells take-up the transgene during a transfection although few express it at a detectable level. This effect remains poorly understood, but it may result from the slow or inefficient unpacking of the DNA from complexes generated by the transfection reagents (77). Thus, the MAR might potentially act in part by facilitating the release of the DNA from complexes with the transfection agent and/or with repressive protein structures. However, specific DNA structures may also act to promote recombination. For instance, endonuclease-mediated double-stranded DNA breaks mediate improved rates of HR in eukaryotes, bacteria and archae (78–82). Thus, an involvement of the unwinding and DNA strand unpairing potential of MARs might directly or indirectly promote the association of recombination-initiating proteins (22,23,83,84). Finally, MAR elements have been associated with the promotion of DNA replication of episomally maintained vectors (28,29). However, this latter effect requires a promoter mediating transcription of the MAR, which is not the case in this study. Furthermore, the MAR sequence used in this study did not lead to increased episomal DNA, and transgene maintenance was clearly associated to chromosomal integration. Thus, if a direct role for DNA replication appears unlikely in the settings used in this study, the requirement of proteins such as replication promoting activity in both replication and HR would be consistent with a role of MARs in promoting both activities.

Overall, our findings imply that successive transfections and the MAR may mediate very efficient expression by promoting HR between individual plasmid molecules, thereby favoring the chromosomal integration of larger concatemers, and by maintaining a permissive chromatin structure after genomic integration. An interesting but as yet unexplored possibility might be that the MAR and HR-mediated events may lead to transgene integration at regions of homologies with the cellular genome. One tantalizing possibility might even be that genomic integration might occur at the endogenous cellular MAR elements, as may result from the fact that MARs share similar AT-rich and highly repetitive sequences across species. Thus, the transgenes would be expected to integrate at more accessible or privileged regions of the genome, explaining the efficient and stable expression. It will be interesting to evaluate whether MAR elements may contribute to regulating HR in a chromosomal context and whether this can be exploited to facilitate gene replacement strategies.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Tha authors thank Arnaud Paradis for skilful help in confocal microscopy and imaging, and Dr Fabrizio Palitti for the kind gift of materials. The authors thank Nathalie Besuchet-Schmutz for excellent technical assistance.

FUNDING

This project was funded by a grant from the Swiss Commission for Technology and Innovation and Selexis SA, and by the University of Lausanne. Funding for open access charge: University of Lausanne.

Conflict of interest statement. Some genetic elements used in this study are also being used by a biotech SME, Selexis SA, co-founded by N.M. After the completion of their work, the company hired some co-authors (P.A.G., D.C.) to work in its own laboratories. This work was financed jointly by a Swiss government agency and be the University, and for a smaller part by the company. We are confident that this has not affected our objectivity in presenting these results, as these were obtained and written-up independently of the authors who joined the company.

REFERENCES

1. Kalos,M. and Fournier,R.E. (1995) Position-independent transgene expression mediated by boundary elements from the apolipoprotein B chromatin domain. *Mol. Cell. Biol.*, **15**, 198–207.
2. Recillas-Targa,F., Pikaart,M.J., Burgess-Beusse,B., Bell,A.C., Litt,M.D., West,A.G., Gaszner,M. and Felsenfeld,G. (2002) Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci USA*, **99**, 6883–6888.
3. Robertson,G., Garrick,D., Wu,W., Kearns,M., Martin,D. and Whitelaw,E. (1995) Position-dependent variegation of globin transgene expression in mice. *Proc. Natl Acad. Sci. USA*, **92**, 5371–5375.
4. Henikoff,S. (1996) Dosage-dependent modification of position-effect variegation in Drosophila. *Bioessays*, **18**, 401–409.
5. Wakimoto,B.T. (1998) Beyond the nucleosome: epigenetic aspects of position-effect variegation in Drosophila. *Cell*, **93**, 321–324.
6. Martienssen,R.A. (2003) Maintenance of heterochromatin by RNA interference of tandem repeats. *Nat. Genet.*, **35**, 213–214.
7. Bell,A.C. and Felsenfeld,G. (1999) Stopped at the border: boundaries and insulators. *Curr. Opin. Genet. Dev.*, **9**, 191–198.
8. Hart,C.M. and Laemmli,U.K. (1998) Facilitation of chromatin dynamics by SARs. *Curr. Opin. Genet. Dev.*, **8**, 519–525.
9. Bode,J., Goetze,S., Heng,H., Krawetz,S.A. and Benham,C. (2003) From DNA structure to gene expression: mediators of nuclear compartmentalization and dynamics. *Chromosome Res.*, **11**, 435–445.
10. Bode,J., Benham,C., Knopp,A. and Mielke,C. (2000) Transcriptional augmentation: modulation of gene expression by scaffold/matrix-attached regions (S/MAR elements). *Crit. Rev. Eukaryot. Gene Expr.*, **10**, 73–90.
11. Gasser,S.M. and Laemmli,U.K. (1987) Improved methods for the isolation of individual and clustered mitotic chromosomes. *Exp. Cell Res.*, **173**, 85–98.
12. Bode,J., Schlake,T., Rios-Ramirez,M., Mielke,C., Stengert,M., Kay,V. and Klehr-Wirth,D. (1995) Scaffold/matrix-attached regions: structural properties creating transcriptionally active loci. *Int. Rev. Cytol.*, **162A**, 389–454.
13. Jackson,D.A. (1997) Chromatin domains and nuclear compartments: establishing sites of gene expression in eukaryotic nuclei. *Mol. Biol. Rep.*, **24**, 209–220.
14. Stein,G.S., van Wijnen,A.J., Stein,J.L. and Lian,J.B. (1999) Interrelationships of transcriptional machinery with nuclear architecture. *Crit. Rev. Eukaryot. Gene Expr.*, **9**, 183–190.
15. Heng,H.H., Krawetz,S.A., Lu,W., Bremer,S., Liu,G. and Ye,C.J. (2001) Re-defining the chromatin loop domain. *Cytogenet. Cell Genet.*, **93**, 155–161.
16. Heng,H.H., Goetze,S., Ye,C.J., Liu,G., Stevens,J.B., Bremer,S.W., Wykes,S.M., Bode,J. and Krawetz,S.A. (2004) Chromatin loops are selectively anchored using scaffold/matrix-attachment regions. *J. Cell Sci.*, **117**, 999–1008.
17. Mirkovitch,J., Mirault,M.E. and Laemmli,U.K. (1984) Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell*, **39**, 223–232.
18. Capelson,M. and Corces,V.G. (2004) Boundary elements and nuclear organization. *Biol. Cell*, **96**, 617–629.
19. Boulikas,T. (1993) Nature of DNA sequences at the attachment regions of genes to the nuclear matrix. *J. Cell Biochem.*, **52**, 14–22.
20. Homberger,H.P. (1989) Bent DNA is a structural feature of scaffold-attached regions in Drosophila melanogaster interphase nuclei. *Chromosoma*, **98**, 99–104.
21. Girod,P.A., Nguyen,D.Q., Calabrese,D., Puttini,S., Grandjean,M., Martinet,D., Regamey,A., Saugy,D., Beckmann,J.S., Bucher,P. *et al.* (2007) Genome-wide prediction of matrix attachment regions that increase gene expression in mammalian cells. *Nat. Methods*, **4**, 747–753.
22. Platts,A.E., Quayle,A.K. and Krawetz,S.A. (2006) In-silico prediction and observations of nuclear matrix attachment. *Cell Mol. Biol. Lett.*, **11**, 191–213.
23. Evans,K., Ott,S., Hansen,A., Koentges,G. and Wernisch,L. (2007) A comparative study of S/MAR prediction tools. *BMC Bioinformatics*, **8**, 71.
24. Ottaviani,D., Lever,E., Takousis,P. and Sheer,D. (2008) Anchoring the genome. *Genome Biol.*, **9**, 201.
25. Koina,E. and Piper,A. (2005) An inactive X specific replication origin associated with a matrix attachment region in the human X linked HPRT gene. *J. Cell Biochem.*, **95**, 391–402.
26. Mearini,G., Chichiarelli,S., Zampieri,M., Masciarelli,S., D'Erme,M., Ferraro,A. and Mattia,E. (2003) Interaction of EBV latent origin of replication with the nuclear matrix: identification of S/MAR sequences and protein components. *FEBS Lett.*, **547**, 119–124.
27. Amati,B.B. and Gasser,S.M. (1988) Chromosomal ARS and CEN elements bind specifically to the yeast nuclear scaffold. *Cell*, **54**, 967–978.
28. Piechaczek,C., Fetzer,C., Baiker,A., Bode,J. and Lipps,H.J. (1999) A vector based on the SV40 origin of replication and chromosomal S/MARs replicates episomally in CHO cells. *Nucleic Acids Res.*, **27**, 426–428.
29. Jenke,A.C., Stehle,I.M., Herrmann,F., Eisenberger,T., Baiker,A., Bode,J., Fackelmayer,F.O. and Lipps,H.J. (2004) Nuclear scaffold/matrix attached region modules linked to a transcription unit are sufficient for replication and maintenance of a mammalian episome. *Proc. Natl Acad. Sci. USA*, **101**, 11322–11327.
30. Liebich,I., Bode,J., Reuter,I. and Wingender,E. (2002) Evaluation of sequence motifs found in scaffold/matrix-attached regions (S/MARs). *Nucleic Acids Res.*, **30**, 3433–3442.
31. Pederson,T. (1998) Thinking about a nuclear matrix. *J. Mol. Biol.*, **277**, 147–159.
32. Cai,S., Han,H.J. and Kohwi-Shigematsu,T. (2003) Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nat. Genet.*, **34**, 42–51.
33. Feister,H.A., Torrungruang,K., Thunyakitpisal,P., Parker,G.E., Rhodes,S.J. and Bidwell,J.P. (2000) NP/NMP4 transcription factors have distinct osteoblast nuclear matrix subdomains. *J. Cell Biochem.*, **79**, 506–517.
34. Dunn,K.L., Zhao,H. and Davie,J.R. (2003) The insulator binding protein CTCF associates with the nuclear matrix. *Exp. Cell Res.*, **288**, 218–223.
35. Fernandez,L.A., Winkler,M. and Grosschedl,R. (2001) Matrix attachment region-dependent function of the immunoglobulin mu enhancer involves histone acetylation at a distance without changes in enhancer occupancy. *Mol. Cell. Biol.*, **21**, 196–208.
36. Yasui,D., Miyano,M., Cai,S., Varga-Weisz,P. and Kohwi-Shigematsu,T. (2002) SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature*, **419**, 641–645.
37. Donev,R., Horton,R., Beck,S., Doneva,T., Vatcheva,R., Bowen,W.R. and Sheer,D. (2003) Recruitment of heterogeneous nuclear ribonucleoprotein A1 in vivo to the LMP/TAP region of the major histocompatibility complex. *J. Biol. Chem.*, **278**, 5214–5226.

38. Rajaiya,J., Nixon,J.C., Ayers,N., Desgranges,Z.P., Roy,A.L. and Webb,C.F. (2006) Induction of immunoglobulin heavy-chain transcription through the transcription factor Bright requires TFII-I. *Mol. Cell. Biol.*, **26**, 4758–4768.
39. Galbete,J.L., Buceta,M. and Mermod,N. (2009) MAR elements regulate the probability of epigenetic switching between active and inactive gene expression. *Mol. Biosyst.*, **5**, 143–150.
40. Kwaks,T.H. and Otte,A.P. (2006) Employing epigenetics to augment the expression of therapeutic proteins in mammalian cells. *Trends Biotechnol.*, **24**, 137–142.
41. Allen,G.C., Spiker,S. and Thompson,W.F. (2000) Use of matrix attachment regions (MARs) to minimize transgene silencing. *Plant Mol. Biol.*, **43**, 361–376.
42. Namciu,S.J. and Fournier,R.E. (2004) Human matrix attachment regions are necessary for the establishment but not the maintenance of transgene insulation in Drosophila melanogaster. *Mol. Cell. Biol.*, **24**, 10236–10245.
43. Dang,Q., Auten,J. and Plavec,I. (2000) Human beta interferon scaffold attachment region inhibits de novo methylation and confers long-term, copy number-dependent expression to a retroviral vector. *J. Virol.*, **74**, 2671–2678.
44. Welcsh,P.L. and King,M.C. (2001) BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. *Hum. Mol. Genet.*, **10**, 705–713.
45. Iarovaia,O.V., Shkumatov,P. and Razin,S.V. (2004) Breakpoint cluster regions of the AML-1 and ETO genes contain MAR elements and are preferentially associated with the nuclear matrix in proliferating HEL cells. *J. Cell Sci.*, **117**, 4583–4590.
46. Kulkarni,A., Pavithra,L., Rampalli,S., Mogare,D., Babu,K., Shiekh,G., Ghosh,S. and Chattopadhyay,S. (2004) HIV-1 integration sites are flanked by potential MARs that alone can act as promoters. *Biochem. Biophys. Res. Commun.*, **322**, 672–677.
47. Johnson,C.N. and Levy,L.S. (2005) Matrix attachment regions as targets for retroviral integration. *Virol. J.*, **2**, 68.
48. Blasquez,V.C., Sperry,A.O., Cockerill,P.N. and Garrard,W.T. (1989) Protein: DNA interactions at chromosomal loop attachment sites. *Genome*, **31**, 503–509.
49. Bae,Y.S., Kawasaki,I., Ikeda,H. and Liu,L.F. (1988) Illegitimate recombination mediated by calf thymus DNA topoisomerase II in vitro. *Proc. Natl Acad. Sci. USA*, **85**, 2076–2080.
50. Keeney,S., Giroux,C.N. and Kleckner,N. (1997) Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell*, **88**, 375–384.
51. Girod,P.A., Zahn-Zabal,M. and Mermod,N. (2005) Use of the chicken lysozyme 5′ matrix attachment region to generate high producer CHO cell lines. *Biotechnol. Bioeng.*, **91**, 1–11.
52. Urlaub,G., Kas,E., Carothers,A.M. and Chasin,L.A. (1983) Deletion of the diploid dihydrofolate reductase locus from cultured mammalian cells. *Cell*, **33**, 405–412.
53. Blunt,T., Finnie,N.J., Taccioli,G.E., Smith,G.C., Demengeot,J., Gottlieb,T.M., Mizuta,R., Varghese,A.J., Alt,F.W., Jeggo,P.A. *et al.* (1995) Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA repair defects associated with the murine scid mutation. *Cell*, **80**, 813–823.
54. Hinz,J.M., Tebbs,R.S., Wilson,P.F., Nham,P.B., Salazar,E.P., Nagasawa,H., Urbin,S.S., Bedford,J.S. and Thompson,L.H. (2006) Repression of mutagenesis by Rad51D-mediated homologous recombination. *Nucleic Acids Res.*, **34**, 1358–1368.
55. Derouazi,M., Martinet,D., Besuchet Schmutz,N., Flaction,R., Wicht,M., Bertschinger,M., Hacker,D.L., Beckmann,J.S. and Wurm,F.M. (2006) Genetic characterization of CHO production host DG44 and derivative recombinant cell lines. *Biochem. Biophys. Res. Commun.*, **340**, 1069–1077.
56. Milligan,L., Antoine,E., Bisbal,C., Weber,M., Brunel,C., Forne,T. and Cathala,G. (2000) H19 gene expression is up-regulated exclusively by stabilization of the RNA during muscle cell differentiation. *Oncogene*, **19**, 5810–5816.
57. Karlen,Y., McNair,A., Perseguers,S., Mazza,C. and Mermod,N. (2007) Statistical significance of quantitative PCR. *BMC Bioinformatics*, **8**, 131.
58. Akita,H., Ito,R., Kamiya,H., Kogure,K. and Harashima,H. (2007) Cell cycle dependent transcription, a determinant factor of heterogeneity in cationic lipid-mediated transgene expression. *J. Gene Med.*, **9**, 197–207.
59. Kim,J.M., Kim,J.S., Park,D.H., Kang,H.S., Yoon,J., Baek,K. and Yoon,Y. (2004) Improved recombinant gene expression in CHO cells using matrix attachment regions. *J. Biotechnol.*, **107**, 95–105.
60. Folger,K.R., Thomas,K. and Capecchi,M.R. (1985) Nonreciprocal exchanges of information between DNA duplexes coinjected into mammalian cell nuclei. *Mol. Cell. Biol.*, **5**, 59–69.
61. Liang,F., Han,M., Romanienko,P.J. and Jasin,M. (1998) Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc. Natl Acad. Sci. USA*, **95**, 5172–5177.
62. Wyman,C., Ristic,D. and Kanaar,K. (2004) Homologous recombination-mediated double-strand break repair. *DNA Repair*, **3**, 827–833.
63. Wong,E.A. and Capecchi,M.R. (1986) Analysis of homologous recombination in cultured mammalian cells in transient expression and stable transformation assays. *Somat. Cell Mol. Genet.*, **12**, 63–72.
64. Takata,M., Sasaki,M.S., Sonoda,E., Morrison,C., Hashimoto,M., Utsumi,H., Yamaguchi-Iwai,Y., Shinohara,A. and Takeda,S. (1998) Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J.*, **17**, 5497–5508.
65. Delacote,F., Han,M., Stamato,T.D., Jasin,M. and Lopez,B.S. (2002) An xrcc4 defect or Wortmannin stimulates homologous recombination specifically induced by double-strand breaks in mammalian cells. *Nucleic Acids Res.*, **30**, 3454–3463.
66. Clikeman,J.A., Khalsa,G.J., Barton,S.L. and Nickoloff,J.A. (2001) Homologous recombinational repair of double-strand breaks in yeast is enhanced by MAT heterozygosity through yKU-dependent and -independent mechanisms. *Genetics*, **157**, 579–589.
67. Allen,C., Kurimasa,A., Brenneman,M.A., Chen,D.J. and Nickoloff,J.A. (2002) DNA-dependent protein kinase suppresses double-strand break-induced and spontaneous homologous recombination. *Proc. Natl Acad. Sci. USA*, **99**, 3758–3763.
68. Pierce,A.J., Hu,P., Han,M., Ellis,N. and Jasin,M. (2001) Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mammalian cells. *Genes Dev.*, **15**, 3237–3242.
69. Jeggo,P.A. (1997) DNA-PK: at the cross-roads of biochemistry and genetics. *Mutat Res.*, **384**, 1–14.
70. Kaufman,R.J. (2000) Overview of vector design for mammalian gene expression. *Mol. Biotechnol.*, **16**, 151–160.
71. Chusainow,J., Yang,Y.S., Yeo,J.H., Toh,P.C., Asvadi,P., Wong,N.S. and Yap,M.G. (2008) A study of monoclonal antibody-producing CHO cell lines: what makes a stable high producer? *Biotechnology and Bioengineering*, **102**, 1182–1196.
72. Henikoff,S. (1998) Conspiracy of silence among repeated transgenes. *Bioessays*, **20**, 532–535.
73. West,S.C. (2003) Molecular views of recombination proteins and their control. *Nat. Rev. Mol. Cell Biol.*, **4**, 435–445.
74. Robins,D.M., Ripley,S., Henderson,A.S. and Axel,R. (1981) Transforming DNA integrates into the host chromosome. *Cell*, **23**, 29–39.
75. Folger,K.R., Wong,E.A., Wahl,G. and Capecchi,M.R. (1982) Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. *Mol. Cell Biol.*, **2**, 1372–1387.
76. Thomas,K.R., Folger,K.R. and Capecchi,M.R. (1986) High frequency targeting of genes to specific sites in the mammalian genome. *Cell*, **44**, 419–428.
77. Chen,H.H., Ho,Y.P., Jiang,X., Mao,H.Q., Wang,T.H. and Leong,K.W. (2008) Quantitative comparison of intracellular unpacking kinetics of polyplexes by a model constructed from quantum dot-FRET. *Mol. Ther.*, **16**, 324–332.
78. Jacquier,A. and Dujon,B. (1985) An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell*, **41**, 383–394.

79. Kostriken,R., Strathern,J.N., Klar,A.J., Hicks,J.B. and Heffron,F. (1983) A site-specific endonuclease essential for mating-type switching in Saccharomyces cerevisiae. *Cell*, **35**, 167–174.
80. Szostak,J.W., Orr-Weaver,T.L., Rothstein,R.J. and Stahl,F.W. (1983) The double-strand-break repair model for recombination. *Cell*, **33**, 25–35.
81. Smih,F., Rouet,P., Romanienko,P.J. and Jasin,M. (1995) Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Res.*, **23**, 5012–5019.
82. Chevalier,B.S. and Stoddard,B.L. (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.*, **29**, 3757–3774.
83. Bode,J., Kohwi,Y., Dickinson,L., Joh,T., Klehr,D., Mielke,C. and Kohwi-Shigematsu,T. (1992) Biological significance of unwinding capability of nuclear matrix-associating DNAs. *Science*, **255**, 195–197.
84. Bode,J., Winkelmann,S., Gotze,S., Spiker,S., Tsutsui,K., Bi,C., Prashanth,A.K. and Benham,C. (2006) Correlations between scaffold/matrix attachment region (S/MAR) binding activity and DNA duplex destabilization energy. *J. Mol. Biol.*, **358**, 597–613.