THE UNIVERSITY *of* EDINBURGH

A novel CRISPR/Cas9-based assay for studying DNA repair associated mutations

Siri Aastedatter Sakya

MSc by Research
Institute of Genetics and Molecular Medicine
University of Edinburgh
2018

Declaration

I hereby declare that this thesis was composed by me, and the research presented is my own, except where otherwise stated. This work has not been submitted for any other degree or professional qualification.

Siri Aastedatter Sakya

August 2018

Acknowledgements

First, I would like to thank my main supervisor Nizar Batada for giving me the opportunity to carry out my research project in his lab under his supervision, and for his feedback during the writing process.

My sincere gratitude goes to my co-supervisor Victor Gonzalez Huici for his encouragement and excellent supervision of my experimental and written work throughout my project. His willingness to answer questions and to give his time so generously has been very much appreciated. I would also like to thank him for preparing the Illumina sequencing library and performing the Bioanalyzer analyses.

My grateful thanks are also extended to Vildan Bozok Cetintas for assisting in the preparation of the Illumina sequencing library and the conduction of the Bioanalyzer analyses.

Further, I would like to thank Noor Gannoh for kindly providing me with polyethylenimine, lentiviral transfer and packaging plasmids, and HEK293T cells; and Dr Martin Reijns for providing RPE-1 Cas9 $p53^{-/-}$ and HeLa Cas9 cells.

I am particularly grateful to Mira for her effort in designing my illustrative figures; Aaste for her insightful suggestions on my writing; and Jade for putting up with all my grammatical questions and for being my greatest support. The completion of this thesis would not have been possible without you.

Finally, I wish to thank my friends and family for their support and encouragement throughout my study.

Abstract

DNA double-strand breaks (DSBs) are a lethal type of DNA damage and are primarily repaired by two main pathways: homologous recombination (HR) and classical non-homologous end joining (C-NHEJ). Impairment of HR or C-NHEJ leads to DSB repair through a less characterised pathway termed alternative non-homologous end joining (Alt-EJ). Recent studies suggest that Alt-EJ contributes to the formation of insertions/deletions and chromosomal translocations, which may cause cancer onset and progression. However, there is currently no existing Alt-EJ assay in the native chromatin context, and thus the underlying mechanism of the Alt-EJ pathway remains poorly understood.

In this Master project, I aimed to develop and implement a novel Alt-EJ assay termed Quantitative Multiplex Analysis of Translocations (QMAT-seq) that can quantify the frequency of translocation events and identify mutational signatures at the repair junctions of translocations. The assay exploits the RNA-guided CRISPR/Cas9 system to generate multiple DSBs in the genome, and the repair of these breaks by Alt-EJ leads to translocations. In this study, five guide-RNA (gRNA) sequences were selected, allowing for assessment of outcomes at 20 distinct translocation junctions. CRISPR/Cas9-mediated DSB formation was verified by the Surveyor nuclease assay and the translocation junctions were captured by nested PCR. Sanger sequencing of the translocation amplicons revealed that translocations form as expected, thus strongly supporting the feasibility of QMAT-seq. However, QMAT-seq is still in its preliminary stage, and future perspectives include quantification of translocation frequencies, high-throughput sequencing, and an extensive bioinformatical analysis of mutational patterns at the translocation junctions. Moreover, the sensitivity of the assay needs to be quantified by perturbing known Alt-EJ factors and assessing if the resulting changes in translocation frequency and mutational patterns are consistent with what is known in the literature.

Upon completion, QMAT-seq will provide a means of identifying novel Alt-EJ factors, subtypes of Alt-EJ, and their underlying mechanism. Furthermore, characterisation of the Alt-EJ mutational signature can serve as a biomarker for deficiency in HR and C-NHEJ, which can aid in developing personalised cancer therapies. QMAT-seq thus promises to make a substantial contribution to the field of DNA repair and cancer biology.

Lay summary

DNA in our cells is constantly damaged, and survival of cells rest on their ability to repair these damages. For this purpose, they have evolved a number of repair mechanisms that deal with damages locally within a single DNA molecule. However, DNA repair can be faulty, leading to merging of DNA pieces from different molecules into a novel DNA segment known as a translocation. Translocations are commonly seen in cancer cells, and mechanisms that favour faulty repair can promote cancer development. If we can identify causes that lead to faulty repair, we can find a way to prevent it from happening. However, the set of molecules that contribute to faulty repair is poorly understood.

In this Master project, I aim to shed light on faulty DNA repair and the characteristics of translocations. When a translocation forms, small alterations termed mutations arise in the DNA. Mechanisms that lead to faulty repair leave a non-random pattern of mutations when they merge two pieces of DNA together into a translocation. In this project, I have optimised and implemented a novel laboratory-based method that permits the study of such mutational patterns, which can be useful in cancer research and clinical practice. Our assay can detect novel molecules that contribute to faulty repair by inactivating a candidate molecule and assessing how this affects the mutational pattern of translocations. A change in the mutational pattern indicates that the molecule participates in faulty repair, and such molecules may serve as targets for cancer drugs that block their action. In this way, we can prevent faulty DNA repair from happening and impede cancer development. Altogether, our assay will provide insight into how faulty DNA repair leads to translocations and how to prevent its occurrence and thus promises to make a substantial contribution to the field of DNA repair and cancer biology.

Abbreviations

Alt-EJ	Alternative non-homologous end joining
ATP	Adenosine triphosphate
bp	Base pair
ChIP	Chromatin immunoprecipitation
C-NHEJ	Classical non-homologous end joining
dH_2O	Distilled water
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
dNTP	Deoxynucleotide triphosphate
DSB	Double-strand break
E. coli	*Escherichia coli*
FBS	Fetal bovine serum
HEK293T	Human embryonic kidney cell line 293T
HR	Homologous recombination
kb	Kilobase
kDa	Kilodalton
LB	Lysogeny Broth
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEI	Polyethylenimine
PVDF	Polyvinylidene difluoride
QMAT-seq	Quantitative Multiplex Amplification of Translocations
RIPA	Radioimmunoprecipitation assay
RPE-1	Retinal pigment epithelium cell line 1
Rpm	Revolutions per minute
RT-PCR	Real-time PCR
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
ST	Barcoded synthetic translocation
v/v	Volume per unit volume
w/v	Weight per unit volume
~	Approximately

Figures

Tables

Contents

1 Introduction

Alternative non-homologous end joining (Alt-EJ) is a pathway for repair of DNA double-strand breaks (DSBs). While protecting cells from cytotoxic DNA lesions, Alt-EJ is error-prone and a significant source of chromosomal translocations, which are commonly associated with cancer onset and progression. Yet the underlying mechanisms of Alt-EJ-induced translocations and the factors involved remain poorly understood. This project aims to establish a novel assay for analysis of Alt-EJ-associated translocations that will aid in the elucidation of the Alt-EJ pathway and its role in cancer formation.

1.1 DNA double-strand breaks

The human genome is constantly subjected to damage, and one of the most detrimental types of DNA damage are DNA double-strand breaks in which both strands of the double helix are severed. These lesions frequently arise due to exogenous factors, such as ionising radiation and chemotherapeutic drugs, but can also be induced by endogenous sources, including replication stress and metabolic by-products (Mehta & Haber 2014). Although being cytotoxic, DSBs are a necessary intermediate in a number of biological processes. For instance, both crossover during meiotic recombination and V(D)J recombination, which is responsible for the diversity in antigen receptors of the vertebrate immune system, rely on programmed, site-specific DSBs (Lieber 2010). Occasionally, however, these mechanisms go awry and cause aberrant processing and an off-target introduction of DSBs (Casellas et al. 2016; Bassing & Alt 2004). If left unrepaired or misrepaired, cells may suffer genomic instability, mutations and abnormal chromosomal rearrangements, which are commonly linked to carcinogenesis onset and progression (Venkitaraman 2014; Weinberg 2014).

1.2 DNA double-strand break repair

1.2.1 Main repair pathways

To preserve genomic integrity, cells have evolved two main repair pathways for resolution of DSBs: homologous recombination (HR) and classical non-homologous end joining (C-NHEJ). HR is considered the most faithful repair mechanism and predominates in the mid-S and mid-G2 phases of the cell cycle (Jasin & Rothstein 2013). It depends on RAD51-mediated strand invasion and uses a homologous sister chromatid as a template for restoration (Figure 1,

right). Unlike HR, C-NHEJ is active throughout the cell cycle but dominates in G1 and G2 (Karanam et al. 2012). In this mechanism, DSBs are repaired through direct, homology-independent ligation of the DNA ends and requires factors such as LIG4, Artemis, DNA-PKs, XRCC4 and Ku70/80 (Figure 1, left).

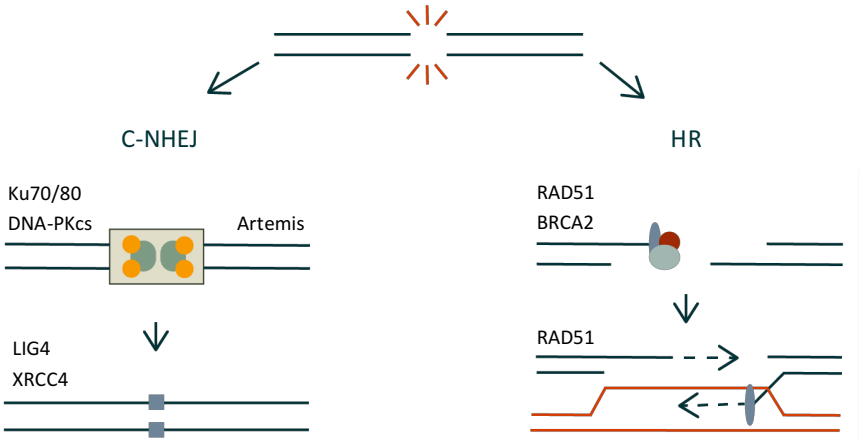

Figure 1. Main pathways for repair of DNA double-strand breaks. The classical non-homologous end joining (C-NHEJ) pathway (left) implements repair in a template-independent manner through the action of several factors, including Ku70/80, DNA-PKcs, Artemis, LIG4, and XRCC4. The homologous recombination (HR) repair pathway (right) uses a homologous sister chromatid as a template for restoration and relies on factors such as RAD51 and BRCA2. Figure adapted from Ceccaldi et al. 2016.

1.2.2 Alternative non-homologous end joining

Research in recent years demonstrates the existence of a third, less characterised form of DSB repair termed alternative non-homologous end joining (Ma et al. 2003; Wang et al. 2003; Boulton & Jackson 1996). Alt-EJ is a slow, error-prone mechanism, mainly active in the S phase of the cell cycle (Ceccaldi et al. 2016; Boulton & Jackson 1996; Mansour et al. 2010). Although a well-defined mechanistic depiction is still lacking, several lines of evidence implicate the involvement of POLQ, LIG3, LIG1, MRE11, XRCC1, CtIP, NBS1, XPF-ERCC1 and PARP-1 (Ahmad et al. 2008; Deriano et al. 2009; Lee-Theilen et al. 2011; Mateos-Gomez et al. 2015; Paul et al. 2013; Wang et al. 2005; Sharma et al. 2015; Mansour et al. 2010). It remains unclear whether Alt-EJ represents a single pathway or a set of multiple pathways with distinct biochemical mechanisms (Deriano & Roth 2013; Frit et al. 2014), but to date, two forms of Alt-EJ have been proposed (Figure 2). One form of Alt-EJ relies on the presence of short stretches of homology near or at DSB-ends and is commonly known as microhomology-mediated end joining (MMEJ). In the proposed model, DSB-ends are processed by CtIP and the MRN complex (MRE11-Rad50-NBS1) to expose microhomologies

in a mechanism known as end resection. After PARP1-mediated annealing of complementary microhomologies, 3'-end overhangs are removed by XPF-ERCC1, followed by fill-in synthesis by POLQ and sealing of DNA ends by LIG3 in a final ligation step (Decottignies 2013; Sfeir & Symington 2015). The other proposed form of Alt-EJ is not dependent on microhomologies and may rather rely on LIG1 for ligation (Boboila et al. 2010; Decottignies 2013). However, microhomologies could possibly be introduced by a, to date, unknown polymerase during gap-filling between broken DNA-ends (Simsek et al. 2011).

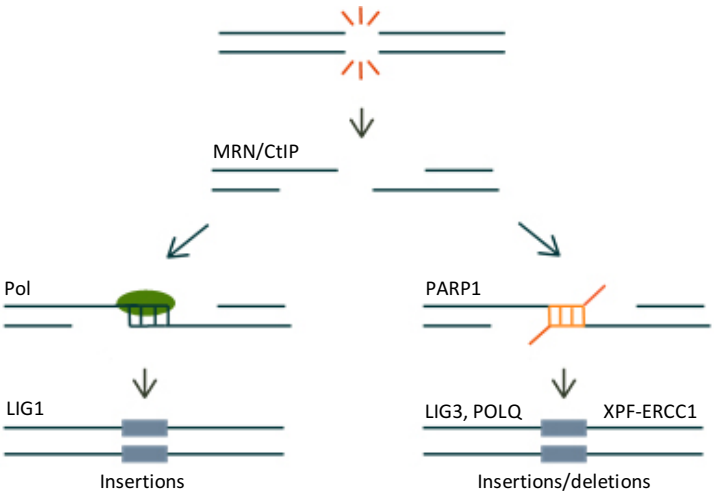

Figure 2. The alternative non-homologous end joining pathway for repair of DSBs. One proposed form of Alt-EJ relies on LIG1 and may generate insertions by a polymerase (Pol) activity (left). The other form requires the presence of microhomologies around the break and depends on factors such as PARP1, XPF-ERCC1, LIG3 and POLQ for repair, which can cause deletions and/or insertions (right). Both forms require end resection of the broken ends by MRN/CtIP prior to repair. Abbreviations: Alt-EJ, alternative non-homologous end joining; DSBs, double-strand breaks. Figure adapted from Decottignies 2013.

Although CtIP, PARP1 and MRE11 have been implicated in Alt-EJ, some studies show that inhibition of these factors does not reduce Alt-EJ activity, hence the genetic requirement for A-EJ remains a matter of debate (Bindra et al. 2013; Bothmer et al. 2013; Bunting & Nussenzweig 2013). There is also some controversy surrounding which ligase is required for the joining of broken DNA-ends and the relevance of the LIG3 cofactor XRCC1 in Alt-EJ. Substantial evidence point in the direction of LIG3 being the main executor of ligation in Alt-EJ and a role of LIG1 as a backup ligase when LIG3 is depleted (Sfeir & Symington 2015; Simsek et al. 2011). However, B lymphocytes depleted of LIG3 or XRCC1 do not show a reduction in Alt-EJ activity, indicating that XRCC1 may not be necessary for Alt-EJ repair and that LIG1 may play a more dominant role in Alt-EJ than previously anticipated (Boboila et al. 2012). This

notion is corroborated by studies demonstrating that XRCC1-deficient cells do not show a reduction in microhomology-mediated joining of DSB-ends (Verkaik et al. 2002) and that LIG1 and LIG3 contribute to Alt-EJ with similar efficiency (Paul et al. 2013). These contrasting findings substantiate the need for further elucidation of the Alt-EJ mechanism and its factors to fully understand how this pathway functions.

1.2.3 Repair pathway choice

Pathway choice in DSB repair is primarily dictated by end resection – a process in which the broken DNA ends are degraded to yield 3' single-stranded DNA tails (Figure 3). End resection shunts repair through HR or Alt-EJ, whereas repair by C-NHEJ is favoured if resection is blocked (Huertas 2010; Symington & Gautier 2011). Resection is initiated by CtIP in concert with the MRN complex and is largely dependent on the cell cycle phase (Bunting & Nussenzweig 2013; Ceccaldi et al. 2016). The resection machinery is regulated through the action of cyclin-dependent kinases (CDKs) and the ATM kinase, ensuring that its activity is restricted mainly to the S and G2 phases of the cell cycle (Bunting et al. 2010; Yun & Hiom 2009; Tomimatsu et al. 2014). Additionally, other post-translational modifications, such as acetylation and sumoylation, modulate end resection (Robert et al. 2011; Cremona et al. 2012), and the DNA damage response factor 53BP1 also serves as an important regulator of end resection and represses HR, and possibly Alt-EJ, by blocking resection (Bunting et al. 2010; Bothmer et al. 2011).

Once end resection has occurred, HR and Alt-EJ compete to repair DSBs (Ceccaldi et al. 2015; Mateos-Gomez et al. 2015). The HR factor RAD51 is a central point of regulation in the decision between these pathways. While BRCA2 facilitates HR by promoting RAD51 nucleofilament formation (Carreira & Kowalczykowski 2011), POLQ suppresses HR by impeding the accumulation of RAD51 (Mateos-Gomez et al. 2015). Furthermore, it has been suggested that POLQ shunts repair through Alt-EJ by stabilising annealed microhomologies, while the replication protein A (RPA) complex counteracts this effect (Deng et al. 2014). Interestingly, some repair factors appear to participate in several repair pathways, thus adding to the complexity of DSB repair pathway choice. For instance, a recent study demonstrates that a number of HR factors, including FA proteins, facilitate Alt-EJ (Howard et al. 2015). However, the exact role of these proteins in Alt-EJ regulation remains unclear (Ceccaldi et al. 2016). Taken together, these findings highlight the complexity of the interplay

between HR, C-NHEJ, and Alt-EJ and the need for further studies to elucidate what dictates the decision between error-prone and more faithful repair pathways.

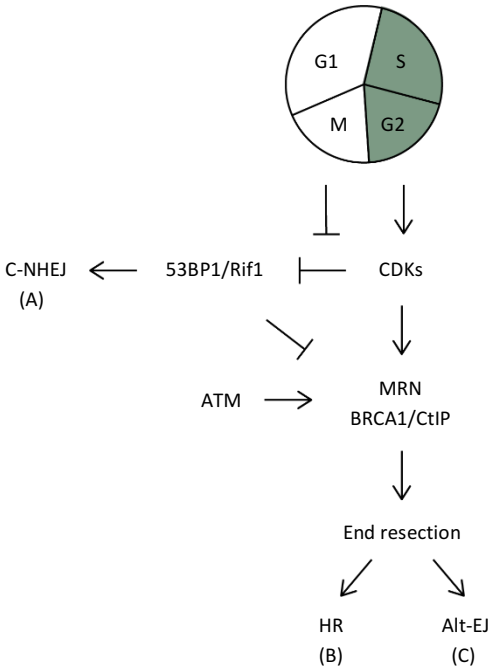

Figure 3. Factors regulating DSB repair pathway choice. The choice between C-NHEJ (A) and the resection-dependent repair pathways HR (B) and Alt-EJ (C) is largely dependent on the cell cycle phase. End resection is mainly promoted in the S/G2 phases through the action of CDKs, which phosphorylate the MRN complex and CtIP. In the G1 phase, 53BP1 and Rif1 localise to the DSB where they inhibit BRCA1 recruitment, thereby blocking DNA end resection and shunting repair through C-NHEJ. In the S/G2 phases, the ATM kinase favours HR and Alt-EJ by phosphorylating BRCA1, CtIP and the MRN complex. Abbreviations: Alt-EJ, alternative non-homologous end joining; ATM, Ataxia Telangiectasia Mutated; CDK, cyclin-dependent kinase; C-NHEJ, classical non-homologous end joining; DSB, DNA double-strand break; HR, homologous recombination. Figure adapted from Ceccaldi et al. 2016.

1.3 DNA repair and genomic instability

While protecting the cells from cytotoxic DNA damages, a common feature of NHEJ mechanisms is their error-proneness (Yousefzadeh & Wood 2013; Ghezraoui et al. 2014). If DSB repair is shunted through these pathways, cells can suffer genomic instability, particularly due to the formation of chromosomal translocations.

1.3.1 Translocations

Chromosomal translocations are abnormal genetic rearrangements that result from the fusion of genetic material, normally between two non-homologous chromosomes, due to misjoining of DSB-ends by the cells' intrinsic repair pathways (Figure 4). Translocations may cause cancer through the formation of oncogenes, inactivation of tumour suppressor genes, or by positioning proto-oncogenes under the control of other more active promoters, leading to their enhanced expression (Weinberg 2014). Among the most well-known translocations, we find the translocation between chromosomes 9 and 22, t(9;22), in which the *ABL1* and *BCR* genes are combined to form a novel fusion gene with oncogenic traits (Mitelman et al.

2007). Recurrent translocations are associated with both haematological malignancies and solid tumours, including leukaemia, lymphomas and carcinomas (Padilla-Nash et al. 2001; Küppers & Dalla-Favera 2001). Although not exclusively pathological, translocations are common in cancer cells and drive approximately 20% of all cancer cases (Bunting & Nussenzweig 2013; Mitelman et al. 2007).

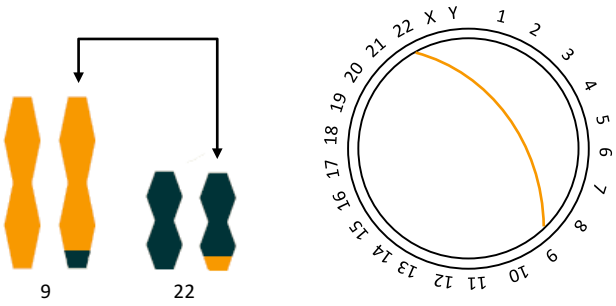

Figure 4. Chromosomal translocations. Translocations normally result from the fusion of sequences from two non-homologous chromosomes. Here, the t(9;22) translocation from chronic lymphocytic leukaemia is illustrated, in which sequences from chromosomes 9 and 22 are exchanged. In a Circos plot (right), this translocation can be displayed as a line between the breakpoints on chromosomes 9 and 22. Figure adapted from Bunting & Nussenzweig 2013.

1.3.2 Factors favouring translocation formation

Several factors affect the formation of chromosomal translocations and at what frequency they occur. Translocation formation requires the presence of multiple DSBs, and it has been previously reported that the occurrence of DSBs is non-random (Klein et al. 2011; Chiarle et al. 2011; Roukos et al. 2013). Indeed, several studies suggest that DNA features such as CpG islands, common fragile sites (CFS) and non-B DNA structures are particularly prone to breakage and translocation formation (Bacolla et al. 2016; Arlt et al. 2006; Ozeri-Galai et al. 2012; Burrow et al. 2009). Conversely, large-scale sequencing of translocation junctions shows that translocation-prone regions display substantial sequence diversity, thus confounding the relationship between specific DNA features and translocation susceptibility (Burman et al. 2015).

Several studies have investigated the impact of chromatin structure and histone modifications on translocation formation. Modifications such as mono- and tri-methylated H3K4 are proven to be enriched in translocation-prone genomic regions (Burman et al. 2015; Shimazaki et al. 2009), and whole-genome sequencing of human prostate cancer has

revealed elevated levels of the active chromatin marks H3K36me3, H3K4me3, and acetylated H3 in the *TMPRSS2-ERG* translocation breakpoint (Berger et al. 2011). However, some prostate cancer translocation regions rather displayed elevated levels of the repressive mark H3K27me3, suggesting that distinct translocations can exhibit different histone modification patterns. Thus, more research needs to be undertaken to fully understand the intricate relationship between histone modification patterns and breakage susceptibility (Roukos et al. 2013).

The spatial arrangement of chromosomes and genes has also been suggested to affect translocation formation (Roukos et al. 2013; Klein et al. 2011; Soutoglou & Misteli 2008). Studies show that chromosome regions located in close proximity with one another translocate more frequently than distant loci. For instance, the *MYC* and *BCL6* loci are found to be localised in vicinity of their prominent translocation partner *IGH* in B-cells (Roix et al. 2003). However, another study claims that even though 85% of Burkitt's lymphoma have the *IGH-MYC* translocation, more than 2,000 other genes fuse with *IGH* more frequently than *MYC*, suggesting that spatial proximity may not be a key factor in translocation formation (Bunting & Nussenzweig 2013; Hakim et al. 2012). Accordingly, more effort needs to be directed toward understanding the contextual basis for translocation formation.

1.4 Alt-EJ and cancer

1.4.1 Alt-EJ as a source of translocations and cancer

Several studies have suggested a central role of Alt-EJ in translocation formation and carcinogenesis in mammalian cells (Ferguson et al. 2000; Zhu et al. 2002; Difilippantonio et al. 2002; Zhang & Jasin 2011; Simsek et al. 2011). Yet, the exact mechanism by which Alt-EJ induces translocations remains poorly understood. For instance, a study in C-NHEJ-deficient cells shows that deletion of *LIG3*, but not *LIG1*, significantly decreases translocation frequency (Lu et al. 2016), indicating that not all Alt-EJ factors are prerequisites for translocation formation. Moreover, recent research provides conflicting evidence regarding the relative contribution of C-NHEJ and Alt-EJ to translocation formation in human cells; one study claims that C-NHEJ appears to account for most translocation rearrangements in human cells (Ghezraoui et al. 2014), whereas another study regards Alt-EJ as the main contributor to translocation formation (Soni et al. 2015). Despite this ambiguity, Alt-EJ is still of substantial clinical relevance, as evidenced by the presence of microhomology signatures

at translocation junctions in human cancer cells and the upregulation of several Alt-EJ factors in cancer patients (Tsai et al. 2008; McVey & Lee 2008). Notably, upregulated levels of LIG3 and PARP1 have been detected in breast cancer and chronic myeloid leukaemia, and POLQ is overexpressed in several types of cancer, including lung, breast, and colon cancer, and is associated with poor clinical outcomes (Wood & Doublié 2016; Higgins, Harris, et al. 2010; Tobin et al. 2012; Tobin et al. 2013; Muvarak et al. 2015; Sallmyr et al. 2008).

1.4.2 Alt-EJ as a cancer therapy target

The observed upregulation of several Alt-EJ factors in cancer patients and the contribution of Alt-EJ to translocation formation makes it an appealing target for cancer therapy (Higgins, Harris, et al. 2010; Tobin et al. 2012). Moreover, as DNA repair status modulates sensitivity to radio- and chemotherapies (Mateo et al. 2015; Fong et al. 2009; Farmer et al. 2005; Bryant et al. 2005), it is plausible to believe that therapy directed at Alt-EJ factors can make cancer cells sensitive to damage, thereby reducing the required dose of genotoxic agent and its associated side effects. Consistent with this notion, a recent study proves that targeting POLQ renders cancer cells sensitive to radiotherapy (Higgins, Prevo, et al. 2010).

Alt-EJ has also been linked to acquired therapy resistance in cancer cells and has been suggested to introduce secondary mutations in the *BRCA2* gene, thus restoring its activity in BRCA2-deficient cells and making them resistant to PARP inhibition (Edwards et al. 2008; Sakai et al. 2008). Furthermore, an increased expression of LIG3 and PARP1 has been observed in breast cancer and BCR-ABL1-positive chronic myeloid leukaemia (CML) cells resistant to anti-estrogen therapy and imatinib, respectively (Tobin et al. 2012; Tobin et al. 2013). Interestingly, these cells exhibit sensitivity to inhibition of Alt-EJ, suggesting an increased dependence on Alt-EJ for DSB repair. Thus, this opens up for the possibility of applying Alt-EJ targeted therapies when primary cancer therapies fail.

What makes Alt-EJ an even more promising therapy candidate is its ability to compensate for the loss of HR and C-NHEJ (Lee-Theilen et al. 2011; Ceccaldi et al. 2015; Kabotyanski et al. 1998; Xie et al. 2009; Zhang & Jasin 2011; Simsek & Jasin 2010). Many cancers are repair-deficient and rely on alternative repair mechanisms for survival (Lord & Ashworth 2016; Konstantinopoulos et al. 2015). Indeed, a recent study shows that HR-deficient tumours are hyper-dependent on POLQ-mediated Alt-EJ, as demonstrated by cell death following *POLQ* knockdown in HR-deficient epithelial ovarian cancer cells (Ceccaldi et al. 2015). The study also shows that POLQ is highly expressed in epithelial ovarian cancers,

consistent with the fact that approximately 50% of ovarian carcinomas are HR-defective (Network 2011). A separate study supports these findings by demonstrating that BRCA1-deficient breast cancer cells rely on POLQ for proliferation (Mateos-Gomez et al. 2015), and it has also been reported that HR-deficient cells depleted of POLQ are sensitive to PARP1 inhibitors (Ceccaldi et al. 2015). Collectively, the above data support the feasibility of synthetic lethality-based drugs that target Alt-EJ and selectively destroy cancer cells that depend on this pathway, with little effect on healthy cells with an intact repair machinery (Nickoloff et al. 2017; Ceccaldi et al. 2016; van Gent & Kanaar 2016). The PARP inhibitor olaparib was the first cancer drug based on this concept; it showed good response rates in HR-deficient ovarian cancer patients with *BRCA1/2* mutations and was granted approval in December 2014 (Gavande et al. 2016). DSB repair factors are increasingly recognised as potential cancer therapy targets, and today several repair factor inhibitors are under clinical development (O'Connor 2015).

Several strategies have been developed in order to determine HR-deficiency, including targeted sequencing to detect HR-associated mutations, protein expression assays, and functional assays to measure HR capacity (Stover et al. 2016; Davies et al. 2017; Vanderstichele et al. 2017). However, these methods have not proved very successful, substantiating the need for alternative approaches to detect HR-deficient patients. Mutational signatures representing historical and ongoing genomic alterations in cancer cells have been proposed as predictive biomarkers for HR-deficiency (Lord & Ashworth 2016; Vanderstichele et al. 2017; Ceccaldi et al. 2016), and to date, several non-random mutational patterns have been identified (Alexandrov et al. 2013). However, a more comprehensive characterisation of the Alt-EJ signature may help to better define HR-deficiency in the future.

1.5 Methods for studying Alt-EJ

Early evidence of the Alt-EJ pathway emerged from *in vitro* end joining assays in which damaged plasmid DNA was incubated with purified proteins or cell extracts containing DNA repair factors, followed by Southern blotting to detect the repair products (Pastwa et al. 2009). More recently, various fluorescence-based assays have been designed to study NHEJ mechanisms (Frit et al. 2014; Bindra et al. 2013). The majority of these assays use linearised reporter plasmids or the rare-cutting endonuclease I-SceI to introduce DSBs in intrachromosomal reporter genes. In these assays, successful end joining by Alt-EJ leads to restoration of the fluorescent reporter gene, which can be monitored by flow cytometry and

used as a measure of Alt-EJ efficiency (Truong et al. 2013; Oh et al. 2014; Wang et al. 2012). While these assays have provided important insights into the Alt-EJ mechanism, they are difficult to implement, as they require double I-SceI cuts at a single locus. It has also been reported that mutagenic repair of nuclease-induced DSBs can disrupt the open reading frame of the fluorescent reporter gene, leading to underestimation of Alt-EJ efficiency (Aubert et al. 2011; Kim et al. 2011). Moreover, none of these assays allows for analysis of Alt-EJ repair in a native chromatin context, thus highlighting the need for additional assays to further elucidate how Alt-EJ operates *in vivo*.

1.6 Summary

In the above literature review, I identify several controversies and gaps in our current knowledge on the Alt-EJ pathway and translocations. First, the genetic makeup of Alt-EJ is unclear, and there exists conflicting evidence on the relevance and the exact role of several of its factors. Nor is it clear how many distinct forms of Alt-EJ exist or how each of them is regulated. Although Alt-EJ is widely associated with translocation formation, its relative contribution to this process in human cells and what factors are involved remains unclear. There is also a need to study how DNA contextual features, such as transcription state and histone modifications, influence repair by Alt-EJ and explain why certain sites in the genome are prone to forming translocations. Yet, to date, no assay exists for analysis of Alt-EJ activity in the native chromatin context.

Being implicated in a range of cancer types, Alt-EJ is a promising cancer therapy target, but more knowledge on its regulatory mechanisms and its contribution to therapy resistance is needed in order to develop effective personalised treatments. Furthermore, we need better ways of identifying patients who will benefit from such treatments, hence the need for a more extensive characterisation of the Alt-EJ mutational signature, which can serve as a predictive biomarker of repair deficiency. These are but a few of the aspects that need to be addressed to fully understand the complex mechanisms underpinning Alt-EJ and translocation formation, its role in cancer, and its clinical relevance.

1.7 Aim

This project aims to develop and implement a novel CRISPR/Cas9-based assay that can capture, quantify, and analyse Alt-EJ-mediated translocations. The assay will: 1) enable the identification of novel Alt-EJ factors, 2) help to uncover the mechanisms of Alt-EJ and 3) aid in the discovery of mutational signatures associated with defective DNA repair for applications in personalised cancer therapy.

The assay will be implemented through the following sub-aims:

1. Generate LIG4$^{-/-}$ cell lines.
2. Construct a CRISPR/Cas9 vector and apply CRISPR/Cas9 technology in LIG4$^{-/-}$ cells to induce DSBs and translocations.
3. Optimise the Surveyor nuclease assay for validation of CRISPR/Cas9-induced DSBs.
4. Capture translocation junctions by nested PCR.
5. Analyse translocation junction sequences.

2 Experimental background and strategy

The main objective of this project is to develop and implement a novel assay for the Alt-EJ pathway. In this chapter, I first introduce the laboratory techniques on which our assay is based. Bioinformatical methods will also be employed, but as these are performed by collaborators, a detailed description is not included. In the second part of the chapter, I present our assay and its underlying experimental strategy.

2.1 Laboratory techniques

2.1.1 CRISPR/Cas9 technology

The CRISPR (clustered regulatory interspaced short palindromic repeats)/Cas (CRISPR-associated) system is a bacterial and archaeal adaptive immune system that protects against invading DNA through RNA-guided cleavage of the foreign genetic material (Hsu et al. 2014; Jinek et al. 2012). Of the three CRISPR/Cas systems that have been identified, the Type II system is the most studied and involves the integration of fragmented foreign DNA into a CRISPR locus. After the locus is transcribed, the transcript is processed into a short CRISPR RNA (crRNA) sequence, which hybridises with *trans*-activating crRNA (tracrRNA) important for further crRNA-maturation. Together, they direct the cleavage of invading DNA by the sequence-specific Cas9 endonucleases. In recent years, the Type II CRISPR/Cas system has been adapted and used as a genomic engineering tool. The crRNA and tracrRNA are often fused into a single guide-RNA (gRNA) sequence that interacts with its genomic target sequence through Watson-Crick base pairing (Figure 5). A protospacer-associated motif (PAM) on the 3'-end of the target sequence is required for strand-specific cleavage by Cas9, which in turn facilitates genome editing via DSB repair mechanisms (Hsu et al. 2014).

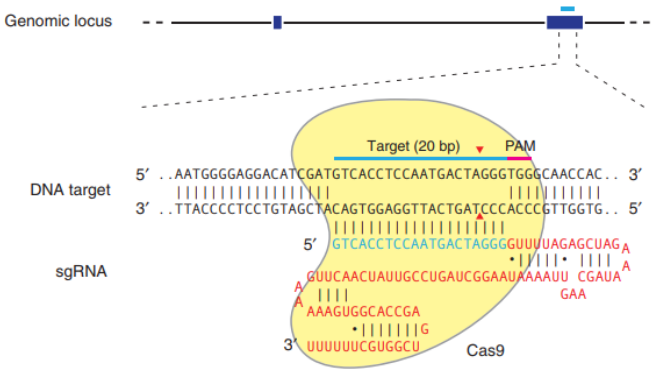

Figure 5. Illustration of the RNA-guided Cas9 nuclease (Ran et al. 2013). The Cas9 nuclease (yellow) is guided to a genomic target site by a guide-RNA consisting of a targeting sequence (blue) and a scaffold (red). The targeting sequence interacts with its genomic target sequence directly upstream of a protospacer-associated motif (PAM; pink). Cas9 introduces a double-strand break ~3 bp upstream of the PAM (red triangles).

2.1.2 Surveyor nuclease assay

The Surveyor nuclease assay is a mismatch cleavage-based gene editing validation tool that exploits mismatch-specific endonucleases to detect CRISPR/Cas9-induced mutations (Figure 6). First, wild-type (wt) and mutant DNA is PCR amplified using the same set of primers, and the resulting amplicons are denatured and rehybridised to form heteroduplexes consisting of a mutant and a wt DNA strand. Next, duplex DNA is treated with the Surveyor nuclease, which detects any present mismatches and introduces a double-strand break 3' of the mismatch. Finally, the cleavage products are resolved on an agarose gel, and gene editing is confirmed by the presence of two or several cleavage products. The intensity of the cleavage bands gives an indication of gene editing efficiency (Qiu et al. 2004).

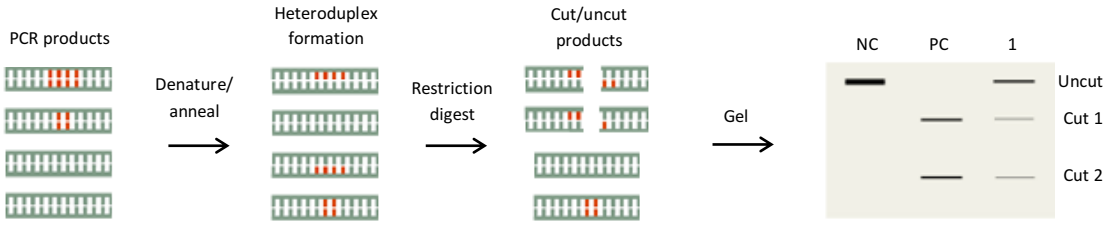

Figure 6. The Surveyor nuclease assay for validation of CRISPR/Cas9-induced mutations. First, wt (blue) and mutant (blue and red) DNA is PCR amplified using the same set of primers, and the resulting amplicons are denatured and reannealed to form heteroduplexes. Next, duplex DNA is treated with Surveyor nuclease, which detects and cleaves any present mismatches. Cleavage products are visualised on an agarose gel (lane 1), and gene editing is confirmed by the presence of two or several cleavage products (Cut 1 and 2). Abbreviations: NC, negative control; PC, positive control. Figure adapted from Johns Hopkins University 2018.

2.1.3 Golden Gate cloning

Golden Gate cloning is a cloning technique that utilises Type IIS restriction endonucleases, such as BsaI or BbsI (Gearing 2015; Kirchmaier et al. 2013). These enzymes cleave outside of their recognition sequence, and the resulting four-base overhangs can, therefore, be customised such that DNA fragments can be directionally assembled into a destination vector. Elimination of the recognition site during digestion prevents recutting and thus allows for cleaving and ligation in a single reaction.

2.1.4 Lentiviral delivery of genetic material

Lentiviral vectors are powerful tools for stable delivery of genetic material and allow for high-efficiency transfer into both dividing and non-dividing cells (Wang & McManus 2009; Giry-Laterrière et al. 2011). To increase the safety of the lentiviral system, viral elements essential

for lentiviral production are split between several packaging plasmids. Co-transfection of transfer and packaging plasmids into a packaging cell line, such as HEK293T, leads to lentivirus production, which is used to infect the target cells for stable expression of genetic material.

2.1.5 Nested PCR

Nested polymerase chain reaction (PCR) is a variant of regular PCR that ensures high specificity amplification of a target region and involves two successive rounds of PCR (Wilczynski 2009). In the initial PCR reaction, a set of external primers is used to generate an amplicon that serves as a template for the second PCR reaction, in which a set of internal primers are used to amplify a secondary target within the primary amplicon.

2.1.6 Illumina sequencing

Illumina is a high-throughput platform for DNA sequencing that involves several steps (Ambardar et al. 2016; Illumina Inc. 2017). First, a sequencing library is prepared by random fragmentation of the sample DNA, followed by ligation of adapter sequences to the 5'- and 3'-ends. Next, the DNA library is loaded into a flow cell where the fragments are captured by surface-bound oligos complementary to the adapter sequences and clonally amplified. The resulting amplicons are sequenced base-by-base using a DNA polymerase to incorporate fluorescently tagged nucleotides into the growing DNA strand. The four dNTPs emits a unique fluorescence signal that is detected by the Illumina platform. Finally, the sequence reads are aligned and compared with a reference sequence to identify sequence variations.

2.2 Experimental strategy

In this project, I will contribute to the development of a novel assay that allows for the quantification and analysis of Alt-EJ-induced translocations. The assay is termed Quantitative and Multiplex Analysis of Translocations (QMAT-seq) and exploits the CRISPR/Cas9 system to introduce multiple site-specific DSBs in the genome (Figure 7). Repair of these DSBs leads to the generation of chromosomal translocations, which are amplified, sequenced, and analysed. To ensure that the repair outcome is a direct measure of Alt-EJ activity, the C-NHEJ pathway will be inactivated so that repair is shunted through Alt-EJ. Moreover, by assessing translocations instead of single intrachromosomal breaks, one can make certain that Alt-EJ,

and not HR, conducts the majority of repair and that the repair readout will reflect true Alt-EJ damage repair events.

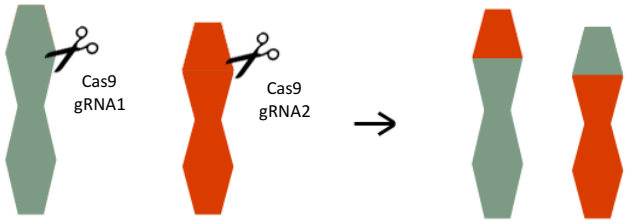

Figure 7. CRISPR/Cas9-induced translocations. Illustration of the use of CRISPR/Cas9 technology to generate desired translocations using a pair of guide-RNAs.

QMAT-seq allows for accurate quantification of translocation frequency, which is achieved by accounting for two key factors: First, as the translocation frequencies are proportional to how efficiently the CRISPR/Cas9 system generates DSBs at the different target sites, gRNA targeting efficiencies will be used to normalise the translocation read counts. Secondly, differences in priming efficiency affect the amplification of different translocation junctions, even if they exist in equal numbers. QMAT-seq addresses this by adding an equimolar pool of barcoded synthetic translocations (STs) with the exact same sequence as the endogenous translocations prior to performing PCR amplification and using the counts of STs to normalise for PCR bias. Translocation frequencies are also affected by the number of distinct gRNAs present in each cell. To ensure that the target cells receive equal quantities of each gRNA, the gRNA sequences will be expressed from a single U6 promoter as a polycistronic gene consisting of gBlocks (gRNA targeting sequence + gRNA scaffold) separated by tRNA sequences (Xie et al. 2015; Ding et al. 2018). Once inside the cell, the tRNAs are cleaved by the endogenous tRNA processing system, releasing fully functional gRNAs.

The experimental strategy for QMAT-seq is illustrated in Figure 8 and the assay will be developed as follows:

1. **Generation of C-NHEJ-deficient cell lines**

   C-NHEJ-deficient cell lines will be generated by knocking down LIG4 using lentivirus to deliver guide-RNAs (gRNAs) targeting LIG4. Subsequently, monoclonal cell lines will be isolated by serial dilution, and knockdown will be confirmed by the Surveyor nuclease assay and Western blotting.

2. **Generation of a multiplex CRISPR/Cas9 vector**

   A single lentiviral vector encoding eight gRNAs (gRNA1-8) will be generated by Golden Gate cloning, allowing for the assessment of 56 distinct translocation junctions.

3. **Expression of multiplex CRISPR/Cas9 vectors in C-NHEJ-deficient cell lines**

   A lentiviral approach will be employed to express the multiplex gRNA vector in the monoclonal LIG4$^{-/-}$ cell lines generated in step 1. The cells are then allowed to recover for 48 hours, which is sufficient time for repair to complete and translocations to form.

4. **Validation of CRISPR/Cas9-mediated DSBs and assessment of gRNA efficiencies**

   DNA from virus-infected cells (step 3) will be extracted and subjected to the Surveyor nuclease assay to validate CRISPR/Cas9-induced DSBs and assess the gRNA targeting efficiencies. Optimal assay conditions will be identified for each gRNA.

5. **Generation of barcoded synthetic translocations**

   56 barcoded synthetic translocations (STs) with the exact same sequence as the endogenous translocations will be generated. The STs consist of two short (~200 bp) DNA fragments flanking a 20 bp non-genomic barcode sequence. First, side fragments will be generated by PCR, and a restriction enzyme (RE) site is introduced on one side of the fragment using tailed primers. Treatment with appropriate REs creates overhangs compatible with those of the barcode oligo and allows for directional ligation of the three fragments. The barcode contains a HindIII site, and successful ligation will be verified by restriction enzyme analysis as well as Sanger sequencing.

6. **Capturing of translocation junctions**

   Synthetic (step 5) and endogenous translocation junctions from virus-infected cells (step 3) will be amplified by nested PCR. Initially, nested PCR will be optimised for each of the 56 translocations separately. Eventually, this step will be simplified by performing PCR amplifications in groups (‘multiplex PCR’) according to optimal PCR conditions.

7. **Sequencing of translocation junctions**

   Endogenous and synthetic translocation junctions captured by nested PCR in step 6 will be pooled and subjected to Illumina high-throughput sequencing.

8. Development of bioinformatics pipeline for analysis of translocation junctions

Collaborators will develop a bioinformatics pipeline for analysis of translocation junctions sequenced in step 7. Differences in gRNA targeting efficiency and nested PCR priming efficiency will be accounted for when computing translocation frequencies and frequencies of the various classes of mutations.

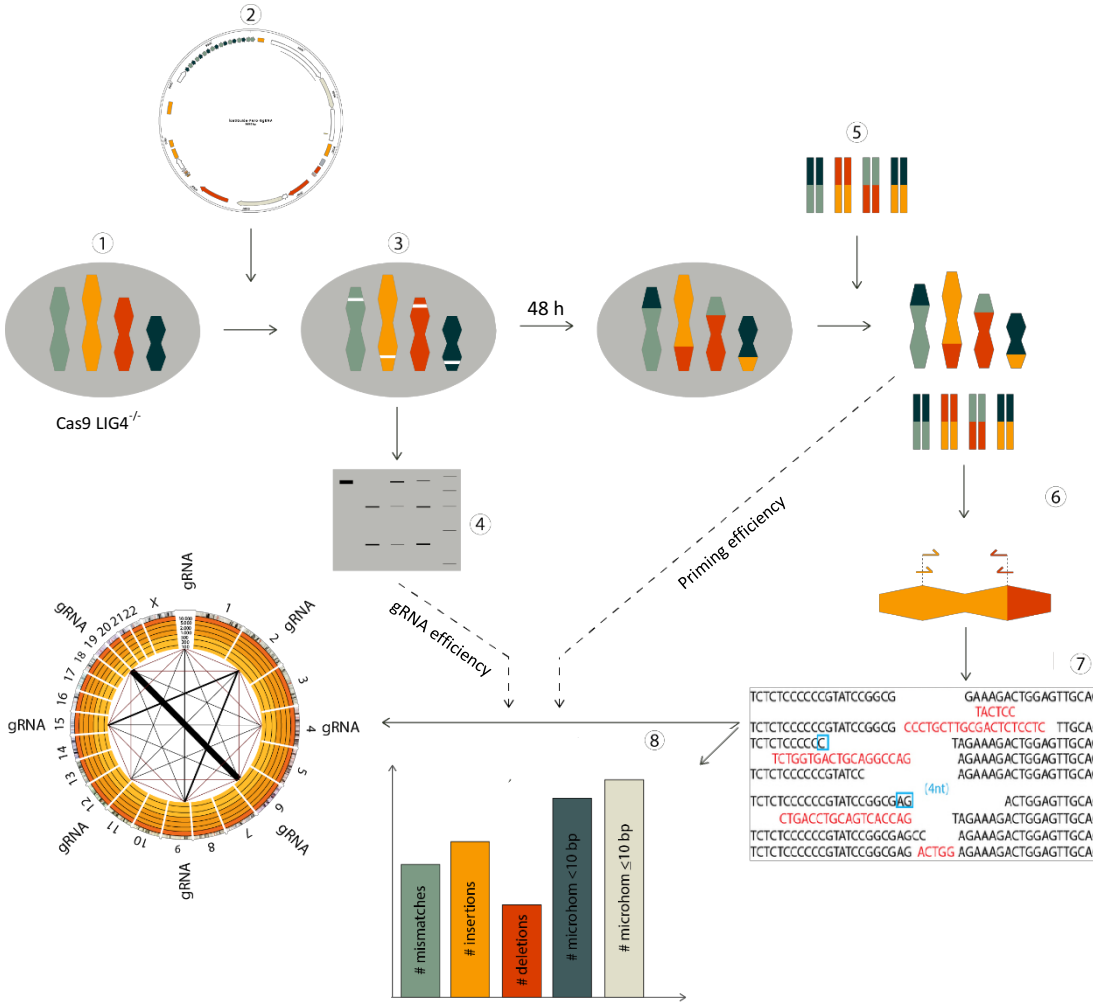

Figure 8. Illustration of steps in QMAT-seq. 1) To shunt repair through Alt-EJ, C-NHEJ-deficient cell lines will be generated by knocking down LIG4 using CRISPR/Cas9 technology. 2) CRISPR/Cas9 technology will be applied to introduce multiple DSBs in the genome of the LIG4$^{-/-}$ cell lines, repair of which will lead to translocation formation. A multiplex CRISPR/Cas9 construct encoding eight gRNAs will be generated by Golden Gate cloning, allowing for the assessment of 56 distinct translocations. 3) The multiplex vector will be expressed in the LIG4$^{-/-}$ cell lines using a lentiviral delivery approach. Cells are allowed to recover for 48 hours, which is sufficient time for repair to complete and translocations to form. 4) The Surveyor nuclease assay is performed to validate CRISPR/Cas9-induced DSBs and to assess gRNA targeting efficiencies. 5) 56 barcoded synthetic translocations with the exact same sequence as the endogenous translocations will be generated by PCR, restriction enzyme digestion and ligation. 6) Equimolar amounts of the 56 synthetic translocations are pooled with DNA extracted from virus-infected cells, and synthetic and endogenous translocation junctions are amplified by nested PCR in groups ('multiplex PCR') according to optimal PCR conditions. 7) Translocation amplicons will be pooled and subjected to Illumina sequencing. 8) Bioinformatics is performed to analyse the translocation sequences. Differences in gRNA targeting efficiency and nested PCR priming efficiency, which is estimated from the counts of reads that emerge from synthetic translocations, are accounted for when computing translocation frequencies and frequencies of the various classes of mutations. Abbreviations: Alt-EJ, alternative non-homologous end joining, DSB, double-strand break; gRNA, guide-RNA.

QMAT-seq provides a means of quantifying and analysing Alt-EJ associated translocations. To induce translocations, we have decided to use the CRISPR/Cas9 system. Other gene editing tools have also been considered, including transcription activator-like effector nucleases (TALENs) and zinc finger nucleases (ZNFs). However, the CRISPR/Cas9 system provides several advantages over these technologies: it is easy to engineer, has a high targeting efficiency, and is easily retargeted to other genomic sites. Most importantly, it has the ability to facilitate multiplex genome editing by the simultaneous delivery of multiple gRNAs, which enables translocation formation (Gaj et al. 2013; Nemudryi et al. 2014). Yet, CRISPR/Cas9 technology has some limitations. The requirement for a PAM sequence directly 3' of the target DNA sequence limits the targeting range to every ~8 bp within the human genome (Cong et al. 2013; Hsu et al. 2013). However, as QMAT-seq is a quantitative assay that does not require highly specific gene editing, this will not critically affect our study. Another limitation of the CRISPR/Cas9 system is off-target activity and the generation of undesired chromosomal translocations (Zhang et al. 2015; Cho et al. 2014). To minimise this effect, an online gRNA design tool (Ran et al. 2013) will be used to identify gRNA sequences with low off-target possibility. Moreover, since translocations are rare events in a typical CRISPR/Cas9 experiment (~1 in 10,000 cells;(Ghezraoui et al. 2014; Mateos-Gomez et al. 2015), QMAT-seq uses N gRNAs to increase the probability of a cell containing a translocation by $N\times(N-1)$. For initial development, eight gRNAs (gRNA1-8) that do not target essential or DNA repair genes will be selected, allowing for the assessment of 56 distinct translocation junctions.

The multiplex CRISPR/Cas9 vector will be constructed by Golden Gate cloning, as the unique 4-base overhangs generated by the Type IIS endonuclease permits efficient assembly of multiple DNA fragments in a single-tube reaction. Despite a large selection of unique overhangs, overhangs differing by only one base may lead to undesired ligation products (Gearing 2015). This issue will be addressed by assembling the gRNA sequences stepwise in compartmentalised reactions. Our lab has previously attempted to deliver multiplex CRISPR/Cas9 vectors into target cells by transient transfection with polyethylenimine. However, the transfection efficiency only reached 20-30%; thus, in this study, we will use a lentiviral approach.

To validate CRISPR/Cas9-induced DSBs, we have decided to employ the Surveyor nuclease assay since it is rapid and cost-efficient as compared to alternative approaches (Zischewski et al. 2017). Moreover, it allows for the assessment of gRNA targeting efficiency, which is important in terms of normalising the frequency of CRISPR/Cas9-induced

translocations. However, the Surveyor nuclease assay may underestimate the mutation frequency due to both the nuclease's cleavage properties and the formation of mutant homoduplexes that cannot be detected by the nuclease (Vouillot et al. 2015). Moreover, being a PCR-based assay, novel mutations could be introduced during the initial amplification step. These problems will be addressed by using a formula that accounts for the binomial probability distribution of duplex formation when calculating the gRNA targeting efficiencies and by using a high-fidelity polymerase, respectively.

In this study, a nested PCR approach is important to capture low-frequency translocation junctions with minimal background. Since the translocation frequency in a typical CRISPR/Cas9 experiment is low, the first round of PCR will lead to high background due to linear amplification of the native (i.e. intrachromosomal) genomic region with either the forward or the reverse external primer. A second PCR reaction will ensure many-fold enrichment of interchromosomal repair junctions (i.e. translocation junctions). The primers are designed such that translocation events with up to 200 bp deletions and 100 bp insertions can be identified. Based on the known features of Alt-EJ, these parameters should capture a large proportion of events (Chan et al. 2010; Ceccaldi et al. 2016; McVey & Lee 2008).

Translocation junctions will be sequenced by Illumina high-throughput sequencing, which permits simultaneous processing of multiple translocation junctions and highly accurate capturing of mutational patterns at the repair junctions with a low error rate, as compared to alternative approaches (Quail et al. 2012; Illumina Inc. 2017).

Overall, we believe that the composition of laboratory techniques discussed above will provide a good foundation for developing a robust and simplified Alt-EJ assay, particularly because it allows for analysis of a large number of translocations from diverse genomic contexts in relatively few steps while at the same time accounting for different types of bias.

3 Materials and methods

3.1 Materials

Table 1. Reagents and chemicals.

Name	Concentration	Supplier	Cat. no.
Blasticidin S hydrochloride	5 mg/ml	Sigma-Aldrich	15205
Buffer for T4 DNA Ligase with 10 mM ATP	10x	NEB	B0202S
CutSmart Buffer	10x	NEB	B72045
Deoxynucleotide (dNTP) Solution Mix	10 mM	NEB	N0447S
DMEM, high glucose, pyruvate	1x	TFS, Gibco	11995065
Gel Loading Dye Purple, no SDS	6x	NEB	B7025S
Golden Gate Assembly Mix	20x	NEB	M2615A
Hexadimethrine bromide (polybrene)	1000x	Sigma-Aldrich	107689
KAPA HiFi HotStart ReadyMix	2x	Roche	07958919001
LightCycler 480 SYBR Green I Master	2x	Roche	04707516001
NEB Golden Gate Buffer	10x	NEB	B2615A
NEBuffer 3.1	10x	NEB	B7203S
NuPAGE LDS Sample Buffer	4x	TFS, Invitrogen	NP0007
NuPAGE MOPS SDS Running Buffer	20x	TFS, Invitrogen	NP0001
NuPAGE Sample Reducing Agent	10x	TFS, Invitrogen	NP0004
NuPAGE Transfer Buffer	20x	TFS, Invitrogen	NP0006
Original Dried Skimmed Milk	–	MARVEL	–
Polyethylenimine	1 mg/ml	Gift from Noor Gannoh	–
Phusion HF Reaction Buffer	5x	NEB	B0518S
Pierce ECL 2 Western Blotting Substrate	–	TFS, Thermo Scientific	PI80196
Protease Inhibition Cocktail Set I	100x	Merck	539131-1VL
Puromycin Dihydrochloride	10 mg/ml	Life Technologies	A11138-03
Quick CIP	10x	NEB	M0510S
RIPA Buffer	–	TFS, Thermo Scientific	89900
SeeBlue Plus2 Prestained Standard	–	TFS, Invitrogen	LC5925
SYBR Safe DNA stain	10,000x	TFS, Invitrogen	S33102
TrypLE Express Enzyme	1x	TFS, Gibco	12605010
Tween 20, Nonionic Detergent	100%	Bio-Rad Laboratories	1706531
UltraPure Low Melting Point Agarose	–	TFS, Invitrogen	16520050
1 kb DNA Ladder	500 µg/ml	NEB	N3232S
100 bp DNA Ladder	500 µg/ml	NEB	N3231S

NEB = New England Biolabs, TFS = Thermo Fisher Scientific

Table 2. Enzymes.

Name	Concentration	Supplier	Cat. no.
BamHI-HF	20,000 U/ml	NEB	R3136S
HindIII-HF	20,000 U/ml	NEB	R31045
NdeI	20,000 U/ml	NEB	R0111S
Phusion HF DNA Polymerase	2,000 U/ml	NEB	M0530S
T4 DNA Ligase	400,000 U/ml	NEB	M0202S
T4 Polynucleotide Kinase	10 U/ml	Promega	M410A

NEB = New England Biolabs

Table 3. Bacterial strains.

Name	Supplier	Cat. no.
NEB Stable Competent *E. coli* cells	New England Biolabs	C3040I

Table 4. Cell lines.

Name	Gift from
RPE-1 Cas9 $p53^{-/-}$	Dr Martin Reijns (IGMM, Edinburgh)
HeLa Cas9	Dr Martin Reijns (IGMM, Edinburgh)
HEK293T	Noor Gammoh (IGMM, Edinburgh)

Table 5. Antibodies.

Name	Dilution	Type	Manufacturer	Cat. no.
Anti-DNA Ligase IV antibody	1:2,000	Primary	Abcam	ab193353
Rabbit Anti-GAPDH Antibody	1:500	Primary	Raybiotech	168-10591
Anti-rabbit IgG, HRP-linked Antibody	1:1,000	Secondary	Cell Signaling Technology	7074S

Table 6. Vectors.

Name	Supplier	Cat. no.	Gift from
lentiGuide-Puro	Addgene	52963	Noor Gammoh (IGMM, Edinburgh)
pAX2	Addgene	35002	Noor Gammoh (IGMM, Edinburgh)
pVSVG	Addgene	8454	Noor Gammoh (IGMM, Edinburgh)

Table 7. Commercial kits and consumables.

Name	Supplier	Cat. no.
DNeasy Blood & Tissue Kit	QIAGEN	69504
High contrast Blue Sensitive X-ray Film, 24x30 cm	SLS	–
LightCycler 480 Multiwell Plate 96, white	Roche	04729692001
NuPAGE 4-12% Bis-Tris Gel, 1.0 mm, 10-well	TFS, Invitrogen	NP0321BOX
PVDF Membrane Filter Paper Sandwich, 0.2 uM Pore Size	TFS, Invitrogen	LC2002
QIAGEN Plasmid Midi Kit	QIAGEN	12143
QIAprep Spin Miniprep Kit	QIAGEN	27104
QIAquick Gel Extraction Kit	QIAGEN	28704
QIAquick PCR Purification Kit	QIAGEN	28104
Qubit dsDNA HS Assay Kit	TFS, Invitrogen	Q32854
SD100 Cellometer Cell Counting Chambers	Nexcelom Bioscience	–
Surveyor Mutation Detection Kit for Standard Gel Electrophoresis	IDT	706025

IDT = Integrated DNA Technologies, SLS = Scientific Laboratory Supplies, TFS = Thermo Fisher Scientific

Table 8. Instruments.

Name	Supplier
Cellometer Auto T4 Bright Field Cell Counter	Nexcelom Bioscience
Curix60 Film Processor	Agfa
LightCycler 480 Instrument II	Roche
NanoDrop 1000	Thermo Fisher Scientific, Thermo Scientific
Qubit 1.0 fluorometer	Thermo Fisher Scientific, Invitrogen
U:Genius3 Gel Documentation System	Syngene International
Veriti 96-Well Thermal Cycler	Thermo Fisher Scientific, Applied Biosystems

Table 9. gRNA sequences.

Name	Sequence (5'-3')	Target gene
gRNA1-top	CACCGTAGTGACAGGTAGATCCAGG	ARHGEF1
gRNA1-bottom	AAACCCTGGATCTACCTGTCACTAC	ARHGEF1
gRNA2-top	CACCGGAGACACCTTAAGAACACCA	CD74
gRNA2-bottom	AAACTGGTGTTCTTAAGGTGTCTCC	CD74
gRNA3-top	CACCGGTTTGGAAAGAGATTCCAGT	SLC34A2
gRNA3-bottom	AAACACTGGAATCTCTTTCCAAACC	SLC34A2
gRNA4-top	CACCGCATAAAGCAGGAAACCCCCG	PAX8
gRNA4-bottom	AAACCGGGGGTTTCCTGCTTTATGC	PAX8
gRNA5-top	CACCGGTTACAGACGAACCACCCCC	CREB3L2
gRNA5-bottom	AAACGGGGGTGGTTCGTCTGTAACC	CREB3L2
gRNA6-top	CACCGCCTGGCCTCGCTGCAGCAAG	CCDC6
gRNA6-bottom	AAACCTTGCAGCGAGGCCAGGC	CCDC6
gRNA7-top	CACCGATAGATGCCACTGGGCCACG	PIGA
gRNA7-bottom	AAACCGTGGCCCAGTGGCATCTATC	PIGA
gRNA8-top	CACCGGTTGGCGTCATCATCCACCA	CD81
gRNA8-bottom	AAACTGGTGGATGATGACGCCAACC	CD81
LIG4gRNA1-top	CACCGAAGGTCGTTTACTTGCTGTA	LIG4
LIG4gRNA1-bottom	AAACTACAGCAAGTAAACGACCTTC	LIG4
LIG4gRNA2-top	CACCGGCATCTCCATGAGTTCCAGT	LIG4
LIG4gRNA2-bottom	AAACACTGGAACTCATGGAGATGCC	LIG4

Oligonucleotides were selected from the GenScript CRISPR sgRNA Database (https://www.genscript.com/gRNA-database.html) or designed using the online CRISPOR tool (http://crispor.tefor.net/). All oligonucleotides were manufactured by Integrated DNA Technologies, and 100 µM stock solutions were prepared by resuspending oligonucleotides in dH_2O.

Table 10. DNA primers.

Primer	Sequence (5'-3')	Purpose
gRNA1-Fwd Ext	GGTCAGTTAGGGCCCATTTGGA	Surveyor assay, nested PCR, ST generation
gRNA1-Rev Ext	TTGAGATACTGGTTGGAGAACTGAA	Surveyor assay, nested PCR, ST generation
gRNA2-Fwd Ext	GCTTCCTGAGAATGGCTTCCCC	Surveyor assay, nested PCR, ST generation
gRNA2-Rev Ext	GCCATCCACAGGGAGGACACT	Surveyor assay, nested PCR, ST generation
gRNA3-Fwd Ext	AAAGCATTGTGGCTAAAGTCTAACG	Surveyor assay, nested PCR, ST generation
gRNA3-Rev Ext	GAAGAATTGGGATGGCTGAGGGT	Surveyor assay, nested PCR, ST generation
gRNA4-Fwd Ext	GTGGATGAGACTGAGGCCAGAGAG	Surveyor assay, nested PCR, ST generation
gRNA4-Rev Ext	CCCTCCACATCCCACGGCTTG	Surveyor assay, nested PCR, ST generation
gRNA5-Fwd Ext	GAAGTTGCATGTCCTTAAATGTACAC	Surveyor assay, nested PCR, ST generation
gRNA5-Rev Ext	ATACTCACATGCAAGCGAATACAA	Surveyor assay, nested PCR, ST generation
gRNA6-Fwd Ext	CCACAGTAGCCTTCCCCCGCA	Surveyor assay, nested PCR, ST generation
gRNA6-Rev Ext	GCCCACTCATTCGCAGCCCG	Surveyor assay, nested PCR, ST generation
gRNA7-Fwd Ext	CCTTCAGCACCCAAAGCTCTCA	Surveyor assay, nested PCR, ST generation
gRNA7-Rev Ext	TTGGCTTCTCTCCCTCCAGGT	Surveyor assay, nested PCR, ST generation
gRNA8-Fwd Ext	ACTTGCCAGGAGGATCTCC	Surveyor assay, nested PCR, ST generation
gRNA8-Rev Ext	GGACACAAATTGTTCTTGAGCACTG	Surveyor assay, nested PCR, ST generation
gRNA1-Fwd Int	CAGGCTGCACTCTACCTCTG	Nested PCR
gRNA1-Rev Int	CTACTCAGTCCCTGACCTGCT	Nested PCR
gRNA2-Fwd Int	TAGACACCAGAGGAAAGAGAGAGTT	Nested PCR
gRNA2-Rev Int	GTGGGCTACACATTTCACACTG	Nested PCR
gRNA3-Fwd Int	GCAAGTTATCCAGGCTCGTAGCC	Nested PCR

gRNA3-Rev Int	TCTAGGATGAGAAGGAACACACCC	Nested PCR
gRNA4-Fwd Int	TGAACTGCCCGTACACGG	Nested PCR
gRNA4-Rev Int	ACCCCTCTCTCCTTGTTGTCT	Nested PCR
gRNA5-Fwd Int	TGAGGATCTATCTGTATCTTTTAATCCAACCT	Nested PCR
gRNA5-Rev Int	CAAGTTCACATGTGTGCTTGTCTTT	Nested PCR
gRNA6-Fwd Int	GGGCACTCACGATGGTCAC	Nested PCR
gRNA6-Rev Int	CATGCAGTCGTCCTGCTCGT	Nested PCR
gRNA7-Fwd Int	CCCCCAAAAGCAAGGTTATTTTCC	Nested PCR
gRNA7-Rev Int	TTGGCAGTTTTCAACTTCCTCTTC	Nested PCR
gRNA8-Fwd Int	AGTTCTATGACCAGGCCCTACAG	Nested PCR
gRNA8-Rev Int	GGATTCATCCAGGACGAGACAC	Nested PCR
gRNA1-NdeI	TCCCATATGTCAGTGCCTCGGCTGGGC	ST generation
gRNA1-BamHI	TGAGGATCCGGATCTACCTGTCACTACTCGC	ST generation
gRNA2-NdeI	TGTCATATGTTCTCCGGGAAGCTCCCCTTC	ST generation
gRNA2-BamHI	GAAGGATCCACACCTTAAGAACACCATGGAG	ST generation
gRNA3-NdeI	CCACATATGTTTCCATTTTCCCACAGATGGGC	ST generation
gRNA3-BamHI	AAAGGATCCTGGAAAGAGATTCCAGTGGGTGG	ST generation
gRNA4-NdeI	TTTCATATGAAGGGTGAGTGAGGATCTGCCG	ST generation
gRNA4-BamHI	CTTGGATCCAAAGCAGGAAACCCCCGAGGTG	ST generation
gRNA5-NdeI	TGTCATATGCTCTGTCTTGATGGATGTTGAAGG	ST generation
gRNA5-BamHI	GAGGGATCCACAGACGAACCACCCCCAGGAC	ST generation
gRNA6-NdeI	GCCCATATGTTGGTGAGCTCCTCCAGGCGG	ST generation
gRNA6-BamHI	CAAGGATCCGGCCTCGCTGCAGCAAGAGAAC	ST generation
gRNA7-NdeI	ATCCATATGAACATCAATGATAGAATCTGGAG	ST generation
gRNA7-BamHI	GTTGGATCCGATGCCACTGGGCCACGGGG	ST generation
gRNA8-NdeI	GCCCATATGGCCAAGGCTGTGGTGAAGACC	ST generation
gRNA8-BamHI	GGCGGATCCGGCGTCATCATCCACCACGGCC	ST generation
Barcode-top	GATCGAGTATCATAGAGTATCATAAAGCTT	ST generation
Barcode-bottom	TAAAGCTTTATGATACTCTATGATACTC	ST generation
Albumin-F	TTTGCAGATGTCAGTGAAAGAGA	Virus titration
Albumin-R	TGGGGAGGCTATAGAAAATAAGG	Virus titration
WPRE-F	GTCCTTTCCATGGCTGCTC	Virus titration
WPRE-R	CCGAAGGGACGTAGCAGA	Virus titration

DNA primers were manufactured by Integrated DNA Technologies. All primers were generated using the online Thermo Fisher Scientific Primer Designer Tool (https://www.thermofisher.com/uk/en/home/life-science/sequencing/sanger-sequencing/pre-designed-primers-pcr-sanger-sequencing.html) or the GenScript CRISPR sgRNA Design Tool (https://www.genscript.com/gRNA-design-tool.html). 100 µM stock solutions were prepared by resuspending primers in dH_2O. Abbreviations: Ext, external; Int, internal; PCR, polymerase chain reaction; ST, barcoded synthetic translocations.

3.2 Methods

3.2.1 Cell culture

Cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% (v/v) fetal bovine serum (DMEM-FBS) containing 2 µg/ml (HeLa) Blasticidin S hydrochlorine at 37°C and 5% CO_2 in a humidified incubator. For virus infection, cells were treated with 4 µg/ml polybrene, and virus-infected cells were selected with 2 µg/ml Puromycin Dihydrochloride.

Cells were split at 70-90% confluence by removing culture medium with an aspirator, washing once with Phosphate-buffered saline (PBS), and treating with the appropriate quantity of pre-warmed (37°C) 1x TrypLE Express Enzyme. After 5-10 min incubation at 37°C, trypsin was inactivated by adding pre-warmed DMEM-FBS, and the cell suspension was diluted to the appropriate seeding density and seeded in a new culture flask. All cells were submitted for mycoplasma testing once every month.

3.2.2 Preservation and recovery of cells from liquid nitrogen

Cells were grown to 90% confluence in a T25 flask, trypsinised, and centrifuged at 1,000 rpm for 5 minutes. The cell pellet was resuspended in 5 ml of DMEM-FBS with 10% (v/v) dimethyl sulfoxide (DMSO) and aliquoted into cryovials. Cryovials were stored in a polystyrene box overnight at -80°C for slow freezing and then transferred to a liquid nitrogen tank. For recovery of cells from liquid nitrogen, cryovials were transferred to a tray with dry ice and thawed in a 37°C water bath for ~1 min. Cells were transferred to a T25 vessel containing pre-warmed (37°C) DMEM-FBS, and the medium was replaced with fresh pre-warmed DMEM-FBS the following day.

3.2.3 Extraction of genomic DNA

Cells were trypsinised, transferred to a 15 ml Falcon tube, and centrifuged at 1,200 rpm for 5 min. 8 ml of PBS was added to the pellet without resuspending, followed by centrifugation at 1,200 rpm for 1 min. Subsequently, the supernatant was discarded, and the tube was placed on ice. Genomic DNA from harvested cells was purified using a DNeasy Blood & Tissue Kit according to the manufacturer's instructions.

3.2.4 Gel extraction

DNA separated by agarose gel electrophoresis was purified with a QIAquick Gel Extraction Kit according to the manufacturer's instructions.

3.2.5 Purification of PCR products

PCR products were purified using a QIAquick PCR Purification Kit according to the manufacturer's instructions.

3.2.6 Transformation of E. coli cells

CRISPR/Cas9 constructs and the packaging plasmids pAX2 and pVSVG were transformed into NEB Stable Competent E. coli cells according to the manufacturer's instructions, but instead of 50-100 µl, 200 µl of transformed cells was spread onto ampicillin selection plates. Cells transformed with linear lentiGuide-Puro served as negative control.

3.2.7 Inoculation and storage of bacterial stocks at -80°C

A single colony of transformed NEB Stable Competent *E. coli* cells was used to inoculate 100 ml of Mg^{2+} LB medium with 0.2 µg/ml ampicillin. Bacterial cultures were incubated overnight at 30°C and 200 rpm. Subsequently, 900 µl of inoculated bacteria was mixed with 900 µl of 50% (v/v) glycerol in a cryovial, vortexed, and stored at -80°C.

3.2.8 Plasmid preparation

CRISPR/Cas9 plasmids were isolated from overnight bacterial inoculation cultures using a QIAprep Spin Miniprep Kit, and the packaging plasmids pAX2 and pVSVG were isolated using the QIAGEN Plasmid Midi Kit according to the manufacturer's instructions.

3.2.9 DNA quantification

Concentrations of plasmid/genomic DNA were measured with a Qubit 1.0 fluorometer using a Qubit dsDNA HS Assay Kit (Invitrogen) according to the manufacturer's instructions.

3.2.10 Sanger sequencing

Tubes with 5.0 µM primer and 100-150 ng of plasmid DNA/PCR product were prepared and submitted for sequencing. All DNA sequence analyses were performed by the MRC Institute of Genetics and Molecular Medicine, Edinburgh.

3.2.11 Protein extraction

Cells were washed twice with cold PBS and lysed with cold RIPA Buffer containing Protease Inhibition Cocktail Set I for 5 min on ice. The lysate was collected in a 1.5 ml microcentrifuge tube using a cell scraper and centrifuged at 13,300 x g for 15 min. The supernatant was transferred to a new microcentrifuge tube for further analysis.

3.2.12 Protein quantification

Protein concentrations were determined by measuring absorbance at 280 nm with a NanoDrop 1000 Spectrophotometer according to the manufacturer's instructions.

3.2.13 Oligonucleotide phosphorylation and annealing

gRNA oligonucleotides (Table 9) were phosphorylated and annealed in 10 µl reactions with 10 µM top oligo, 10 µM bottom oligo, Buffer for T4 DNA Ligase with 10 mM ATP and 0.005 units of T4 Polynucleotide Kinase in dH_2O. Reactions were performed in a Veriti 96-Well Thermal Cycler with heated lid (105°C) using the following conditions: 37°C for 30 min, 95°C for 5 min, ramp-down to 25°C (-0.1°C/s). The barcode oligonucleotide was phosphorylated and annealed as described above, but with 0.01 units of T4 Polynucleotide Kinase.

3.2.14 Agarose gel electrophoresis

DNA samples were resolved on 0.6-2% (w/v) agarose gels, which were prepared by dissolving 0.6-2% (w/v) UltraPure Low Melting Point Agarose in 0.5x TBE buffer in a microwave for 1-3 min. After having cooled down to ~50°C in cold water, SYBR Safe DNA stain was added, and the solution was poured into a casting tray with a gel comb. When solidified, DNA samples and a 100 bp or 1 kb DNA Ladder mixed with Gel Loading Dye Purple were loaded into the gel. The electrophoresis was performed at 90-110 V in an electrophoresis unit filled with 0.5x TBE buffer. When the dye had run approximately ¾ the length of the gel, DNA bands were visualised under ultraviolet light using a U:Genius3 Gel Documentation System.

3.2.15 Isolation of monoclonal LIG4$^{-/-}$ cell populations

Cas9-expressing HeLa LIG4$^{-/-}$ and RPE-1 p53$^{-/-}$ LIG4$^{-/-}$ cells were grown to ~90% confluence, trypsinised, and transferred to a Falcon tube. Cells were centrifuged at 1,200 rpm for 5 min, washed with 5 ml of PBS, and centrifuged for another 5 min at 1,200 rpm. The PBS was removed, and the cell pellet was resuspended thoroughly in 2 ml of pre-warmed (37°C) DMEM-FBS. Cells were counted using a Cellometer Auto T4 Bright Field Cell Counter according to the manufacturer's instructions, and a ten-fold serial dilution was prepared to obtain a suspension with ~50 cells. 100 µl of the cell suspension was transferred into each well of a 96-well plate for an average density of 0.5 cells/well. After 7-14 days of incubation at 37°C, plates were scanned for growth, and wells with more than a single colony were

discarded. Once reaching confluence, monoclonal cell populations were trypsinised and transferred to a 24-well plate for further expansion.

3.2.16 SDS-PAGE and Western blot analysis

Samples were prepared in 20 µl volumes with ~270 µg of protein, NuPAGE LDS Sample Buffer, NuPAGE Sample Reducing Agent, and dH_2O. Samples were incubated at 70°C for 10 min, placed on ice, and resolved on a NuPAGE 4-12% Bis-Tris Gel in an electrophoresis chamber filled with NuPAGE MOPS SDS Running Buffer. 20 µl of SeeBlue Plus2 Prestained Standard was loaded into a separate well, and the electrophoresis was carried out at 4°C and 200 V for 1 h and 25 min. Following electrophoresis, a polyvinylidene fluoride (PVDF) membrane was soaked in methanol, drained, and equilibrated along with filter paper in NuPAGE Transfer Buffer with 20% (v/v) methanol. The transfer sandwich was assembled (pad-filter paper-gel-membrane-filter paper-pad) in a transfer cassette and placed in a tank filled with transfer buffer. Proteins were transferred to the PVDF membrane at 4°C and 30 V for 3.5 h. Next, the membrane was briefly soaked in PBS containing 0.1% (v/v) Tween 20 (PBS-T) and blocked with 5% (w/v) Original Dried Skimmed Milk in PBS-T at 4°C with shaking for 1 h. The membrane was washed twice with PBS-T, cut in half at 49 kDa, and probed with appropriate primary antibodies overnight with shaking at 4°C. Membranes were washed 3 x 10 min in PBS-T and incubated with secondary antibody at 4°C with shaking for 1 h. All antibodies were diluted in PBS-T 5% (v/v) milk buffer. Membranes were washed 3 x 5 min in PBS-T, treated with Pierce ECL 2 Western Blotting Substrate by mixing substrates A and B in a 40:1 ratio, and incubated in the dark for 5 min at room temperature. Blots were placed in a developing cassette with cling film, exposed for 1-10 min, and developed on High contrast Blue Sensitive X-ray Film using a Curix60 Film Processor (Agfa) in a darkroom.

3.2.17 CRISPR/Cas9 vector construction

lentiGuide-Puro vector was linearised by preparing a 20 µl reaction with 5.0 µg of lentiGuide-Puro, 15 units of BsmbI, and NEBuffer 3.1 in dH_2O. Immediately after incubation at 55°C for 1 h, the linearised vector was resolved on a 0.8% (w/v) agarose gel and gel purified (~8.0 kb band). A multiplex insert with eight gRNAs was assembled stepwise by Golden Gate cloning. All digestion/ligation reactions were performed in a Veriti 96-Well Thermal Cycler with heated lid (105°C) for 30 cycles with the following conditions: digestion at 37°C for 5 min, and ligation at 16°C for 5 min. First, DNA fragments consisting of two gRNAs each were generated

by digesting and ligating the following pairs of phosphorylated and annealed gRNA oligonucleotides in four separate reactions: gRNA1 + gRNA2, gRNA3 + gRNA4, gRNA5 + gRNA6, and gRNA7 + gRNA8. The digestion reactions were prepared in 20 µl volumes by mixing 230 ng each of the two gRNAs, NEB Golden Gate Buffer, and Golden Gate Assembly Mix in dH_2O. The resulting ligation products were pooled and ligated as follows: gRNA1/2 + gRNA3/4 and gRNA5/6 + gRNA7/8. Finally, all DNA fragments were combined in a final ligation step to give a single fragment containing all eight gRNAs. After further incubation at 16°C for 8 h, the reaction was run on a 0.8% (w/v) agarose gel. The multiplex insert was gel purified (~1.4 kb band) and ligated into the linear lentiGuide-Puro vector at 2:1, 5:1, and 25:1 molar ratios in 40 µl reactions with Buffer for T4 DNA Ligase with 10 mM ATP and 800 units of T4 DNA Ligase in dH_2O. After incubation at room temperature for 1 h, the ligase was inactivated at 65°C for 20 min.

Vectors containing single gRNAs were generated by preparing a 20 µl reaction with 50 ng of linearised lentiGuide-Puro vector, 0.05 µM annealed and phosphorylated gRNA oligonucleotide, Buffer for T4 DNA Ligase with 10 mM ATP, and 400 units of T4 DNA Ligase in dH_2O. The reaction was incubated at room temperature for 1 h, and vectors were cloned, isolated, and subjected to Sanger sequencing.

3.2.18 Lentivirus production

HEK293T cells were grown to ~70% confluence in eight 10-cm dishes and transfected with packaging plasmids and CRISPR/Cas9 constructs with a single gRNA insert in eight separate transfections. The transfection mixture contained 5 µg of the CRISPR/Cas9 construct, 1.75 µg of pVSVG, 2.75 µg of pAX2, 54 µl of 1 mg/ml polyethylenimine (PEI), and 1 ml of DMEM. The transfection mixture was incubated at room temperature for 15 min and added to the HEK293T cells. After incubation at 37°C for 48 h, the virus-containing supernatant was collected in Falcon tubes and centrifuged at 190 rpm for 3 min. The supernatant was either used for lentiviral transduction immediately or aliquoted into cryovials and stored at -80°C for later use.

3.2.19 Lentiviral transduction

HeLa Cas9, RPE-1 Cas9 $p53^{-/-}$, and RPE-1 Cas9 $p53^{-/-}$ $LIG4^{-/-}$ cells (clones 1 and 2) were seeded in a 6-well plate and grown to 20-30% confluence. Cells were transduced either with the eight viruses produced in 3.2.18 separately (HeLa Cas9 and RPE-1 Cas9 $p53^{-/-}$) or with a

combination of five viruses for delivery of gRNA1/2/3/5/7 (RPE-1 Cas9 $p53^{-/-}$ and RPE-1 Cas9 $p53^{-/-}$ $LIG4^{-/-}$). For delivery of a single gRNA, 2 ml of the appropriate virus was added to the well. For delivery of the five gRNAs combined, 400 µl of each lentivirus was added. Uninfected cells served as negative control. Plates were sealed with parafilm, centrifuged at 15°C and 1,000 rpm for 10 min, and stored at 37°C in a humidified incubator. 24 h post transduction, the medium was replaced with pre-warmed (37°C) DMEM-FBS. Some of the RPE-1 Cas9 $p53^{-/-}$ and RPE-1 Cas9 $p53^{-/-}$ $LIG4^{-/-}$ cells (clones 1 and 2) were infected with a combination of the five viruses twice. The second transduction was performed 18 h after the first transduction as described above, and the cells were incubated at room temperature for 6 h before replacing the medium with pre-warmed DMEM-FBS and storing the cells at 37°C.

3.2.20 Surveyor nuclease assay

The Surveyor nuclease assay was performed as described by Integrated DNA Technologies (http://sfvideo.blob.core.windows.net/sitefinity/docs/default-source/user-guide-manual/surveyor-kit-for-gel-electrophoresis-user-guide.pdf?sfvrsn=a9123407_6) with a few modifications using a Surveyor Mutation Detection Kit for Standard Gel Electrophoresis.

Step 1 - PCR amplification of mutant and wild-type samples

Mutant and wt DNA was amplified in separate 50 µl reactions by mixing Phusion HF Reaction Buffer, 0.5 µM forward primer, 0.5 µM reverse primer, 1.0 units of Phusion HF DNA Polymerase, 0.2 mM Deoxynucleotide (dNTP) Solution Mix, and ~150 ng DNA template in dH_2O. C and G controls were prepared in separate 50 µl reactions using 2 µl of Control C/G. All reactions were carried out in a Veriti 96-Well Thermal Cycler with heated lid (105°C). First, an initial denaturation step was performed at 98°C for 30 s followed by 30 amplification cycles with the following conditions: denaturation at 98°C for 10 s, annealing at 65°C for 30 s, and extension at 72°C for 30 s. A final extension was performed at 72°C for 5 min. PCR amplicons were analysed on a 1.8% (w/v) agarose gel, and if a single band of the correct size was visible in each sample, samples were processed further in step 2.

Step 2 - DNA duplex formation

Wt and mutant amplicons were hybridised in a 1:1 ratio to form heteroduplexes. Homoduplexes of wt or mutant DNA alone served as negative controls and were prepared in separate tubes. 20 µl reactions were prepared by mixing 150 ng each of wt and mutant DNA in dH_2O. The positive control contained 300 ng each of control C and G, and negative controls

contained 300 ng of mutant/wt DNA alone. Reactions were performed in a Veriti 96-Well Thermal Cycler with heated lid (105°C) using the following program: 10 min at 95°C, ramp-down to 85°C (-2.0°C/s), 1 min at 85°C, ramp-down to 75°C (-0.3°C/s), 1 min at 75°C, ramp-down to 65°C (-0.3 °C/s), 1 min at 65°C, ramp-down to 55°C (-0.3°C/s), 1 min at 55°C, ramp-down to 45°C (-0.3°C/s), 1 min at 45°C, ramp-down to 35°C (-0.3°C/s), 1 min at 35°C, ramp-down to 25°C (-0.3°C/s), 1 min at 25°C.

Step 3 - Treatment with Surveyor nuclease

DNA duplexes and control samples were digested in separate tubes by mixing 20 µl DNA from step 2 with 1 µl Surveyor Enhancer S and 1 (gRNA2/7) or 2 (gRNA1/3/5) µl Surveyor Nuclease S. The reactions were gently vortexed, incubated at 42°C for 1 h, and either analysed on an agarose gel immediately after digestion or treated with 2.5 µl Stop Solution and stored at -20°C.

Step 4 – Analysis of DNA fragments

DNA products from step 3 were resolved on a 2% (w/v) agarose gel, and CRISPR/Cas9-induced DSBs were verified by the presence of new cleavage products in samples treated with Surveyor nuclease.

3.2.21 Calculation of gRNA targeting efficiency

Calculations of gRNA targeting efficiency were based on relative band intensities of the cleaved and undigested bands in the gel image from step 4 of the Surveyor nuclease assay. Band intensities were quantified in ImageJ with image properties set to '8-bit' and 'Invert LUT'. First, undigested and cleaved bands were captured in a single rectangle along the length of the lane by using the 'Rectangle' tool, and plots were produced by pressing 'Ctrl+1' followed by 'Ctrl+3'. The 'Straight' and 'Wand' tools were used to remove background noise and measure the relative area under the peaks for each band. The fraction of cleaved duplex DNA (f_{cut}) was calculated for each lane by using the following formula:

$$f_{cut} = (b + c)/(a + b + c)$$

Here, *a* is the intensity of undigested duplex DNA and *b* and *c* are the intensities of the two cleavage products. Indel mutation occurrence was used as a measure for gRNA targeting efficiency and was estimated with the following formula:

$$Indel(\%) = 100\times\left(1 - \sqrt{1 - f_{cut}}\right)$$

3.2.22 Generation of barcoded synthetic translocations

The eight PCR amplicons generated from wt DNA in step 1 of the Surveyor nuclease assay (3.2.20) served as templates for the generation of the 5'-side and 3'-side fragments of the barcoded synthetic translocations (STs) by PCR. 5'-side fragments were generated by using forward external primers (gRNAX-Fwd Ext) and tailed reverse primers with a BamHI restriction site (gRNAX-BamHI). 3'-side fragments were generated by using reverse external primers (gRNAX-Rev Ext) and tailed forward primers with a NdeI restriction site (gRNAX-NdeI). Separate 50 µl reactions were prepared for each fragment by mixing Phusion HF Reaction Buffer, 0.5 µM forward primer, 0.5 µM reverse primer, 1.0 units Phusion HF DNA Polymerase, 0.2 mM Deoxynucleotide (dNTP) Solution Mix, and 1.0 µl of the appropriate DNA template in dH_2O. The reactions were carried out in a Veriti 96-Well Thermal Cycler with heated lid (105°C). PCR amplification was performed exactly as described in 3.2.20 (step 1), but with 35 cycles instead of 30. PCR products were purified and resolved on a 1.9% (w/v) agarose gel.

Side fragments were digested with appropriate restriction enzymes by mixing 48.0 µL of DNA with CutSmart Buffer and 40.0 units of either BamHI-HF or NdeI. Digestion reactions were incubated overnight at 37°C. Restriction enzymes were inactivated by placing 5'-side fragments on ice and incubating the 3'-side fragments at 65°C for 20 min. Digested side fragments were dephosphorylated by adding 1 µL of Quick CIP and incubating samples at 37°C for 30 min followed by inactivation at 85°C for 2 min.

The 56 STs were assembled by mixing equimolar amounts (1.0 pmol) of duplex barcode oligonucleotide and appropriate dephosphorylated side fragments with 400 units of T4 DNA Ligase and Buffer for T4 DNA Ligase with 10 mM ATP. dH_2O was added to a final volume of 20 µl, and samples were incubated at 16°C overnight. Reactions were inactivated at 65°C for 10 min and STs were amplified exactly as described in 3.2.20 (step 1), but with 35 cycles instead of 30. PCR reactions were prepared in 25 µL reactions with 2 µL of ST, 0.5 µM forward external primer, 0.5 µM reverse external primer, 0.2 mM Deoxynucleotide (dNTP) Solution Mix, 5.0 units of Phusion HF DNA Polymerase, and Phusion HF Reaction Buffer. Amplified STs were resolved on a 1.9% (w/v) agarose gel and gel purified.

3.2.23 Restriction enzyme analysis

Purified barcoded synthetic translocations were treated with HindIII in 25 µL reactions with 200-400 ng of DNA, 10.0 units of HindIII-HF, and CutSmart Buffer in dH_2O. The digestion was

performed at 37°C for 1 h, and digestion products were analysed on a 1.9% (w/v) agarose gel. Undigested STs served as positive controls.

3.2.24 Nested PCR

Negative testing of external nested PCR primers was conducted using genomic DNA from RPE-1 Cas9 $p53^{-/-}$ cells as template. All 56 combinations of forward and reverse primers were tested. PCR reactions were prepared and run exactly as described in 3.2.20 (step 1), and combinations of primers for amplification of the same gRNA target region (e.g. gRNA1-Fwd Ext and gRNA1-Rev Ext) served as positive controls. PCR products were resolved on a 1.8% (w/v) agarose gel.

External PCR reactions were performed in groups with a single external forward primer combined with the remaining external reverse primers (e.g. gRNA1-Fwd with gRNA2/3/5/7 Rev). 25 µl reactions were prepared by mixing ~100 ng DNA, Phusion HF Reaction Buffer, 0.2 mM Deoxynucleotide (dNTP) Solution Mix, 5.0 units of Phusion HF DNA Polymerase, and 1.0 µM of each primer in dH_2O. PCR amplification was performed exactly as described in 3.2.20 (step 1), but with 35 cycles instead of 30.

As for external reactions, internal reactions were performed in groups with a single internal forward primer combined with the remaining internal reverse primers. 2.5 µl of DNA from the external reaction was mixed with 1.0 µM of each primer and KAPA HiFi HotStart ReadyMix, and dH_2O was added to a final volume of 25 µl. Reactions were performed in a Veriti 96-Well Thermal Cycler with heated lid (105°C). An initial denaturation step was performed at 95°C for 3 min followed by 25 amplification cycles with the following conditions: denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 30 s. A final extension was performed at 72°C for 5 min. The resulting PCR products were resolved on a 1.8% (w/v) agarose gel.

3.2.25 Illumina library preparation

An Illumina library of translocation amplicons from RPE-1 Cas9 $p53^{-/-}$ and RPE-1 Cas9 $p53^{-/-}$ $LIG4^{-/-}$ cells was prepared by Victor Gonzalez Huici and Vildan Bozok Cetintas.

3.2.26 Virus quantification

Lentiviruses for delivery of CRISPR/Cas9 constructs with a single gRNA were quantified as described by Barczac et al. (2015) with a few modifications. RPE-1 Cas9 p53$^{-/-}$ LIG4$^{-/-}$ cells were trypsinised and centrifuged at 1,200 rpm for 5 min. The cell pellet was resuspended in pre-warmed (37°C) DMEM-FBS, and cells were counted with a Cellometer Auto T4 Bright Field Cell Counter according to the manufacturer's instructions. ~50,000 cells were seeded in each well of a 24-well plate. After 6 hours, separate wells were infected with 0, 2, and 4 µl of each lentivirus. After incubation at 37°C for 48 h, cells were harvested and subjected to genomic DNA extraction. Lentiviral titer was assessed by measuring the levels of a lentiviral-specific transgene (WPRE) and a single copy gene-specific reference gene (albumin) using real-time PCR. Levels of albumin and WPRE in cells infected with 0, 2 and 4 µl of virus were measured in separate 20 µl reactions by mixing 5 µl of DNA, 0.5 µM forward primer, 0.5 µM reverse primer, and LightCycler 480 SYBR Green I Master in dH_2O. DNA from RPE-1 Cas9 p53$^{-/-}$ cells was used to generate albumin and WPRE standard curves for calibration and reference. A ten-fold serial dilution was prepared with dH_2O, with the most concentrated and diluted samples being ~24 ng/µl and ~2.4 pg/µl, respectively. Test and standard samples were loaded into a LightCycler 480 Multiwell Plate 96, and wells with DNA from wt RPE-1 cells and dH_2O only served as negative controls. Samples were amplified in a LightCycler 480 Instrument II. After pre-incubation at 95°C for 5 min, 45 cycles with the following conditions was performed: 95°C for 10s, 60°C for 10 s, 72°C for 10 s. After amplification, assessment of single product generation was performed by melting curve analysis: 95°C for 5 s, 65°C for 1 min, continuous acquisition of fluorescence (5 acquisitions/°C) to 97°C, and cooling at 40°C for 30 s. An estimation of relative virus titer was determined by considering the volume of used lentivirus and using albumin levels to normalise WPRE levels.

3.2.27 Bioanalyzer analysis

The Illumina library of translocation amplicons from RPE-1 Cas9 p53$^{-/-}$ and RPE-1 Cas9 p53$^{-/-}$ LIG4$^{-/-}$ cells was analysed by Victor Gonzalez Huici and Vildan Bozok Cetintas using a Bioanalyzer.

4 Results

4.1 Isolation of monoclonal LIG4$^{-/-}$ cell lines

To ensure that repair outcomes and translocation frequencies are direct measures of Alt-EJ activity, QMAT-seq was carried out in cells deficient in the C-NHEJ pathway. C-NHEJ repair was inhibited in Cas9-expressing HeLa and RPE-1 p53$^{-/-}$ cells by knocking down LIG4 using CRISPR/Cas9 technology. Cleavage efficiency was validated by means of the Surveyor nuclease assay, which allows the detection of mismatches to the original sequence as a result of DSB repair. 500-600 bp amplicons containing the LIG4 gRNA target region from treated and control cells were generated by PCR and hybridised to form heteroduplexes. Duplexes were treated with 2 µl of the mismatch-specific Surveyor nuclease for 1 h at 42°C and resolved on a 2% (w/v) agarose gel (Figure 9). CRISPR/Cas9-induced mutation formation was confirmed in both RPE-1 and HeLa cells, as revealed by the presence of cleavage products in lanes marked 'T', with indel rates of 16.8% (HeLa) and 52.9% (RPE-1).

As CRISPR/Cas9-mediated gene editing results in a heterogeneous cell population, single clones were isolated to derive isogenic LIG4$^{-/-}$ cell lines. Only monoclonal RPE-1 Cas9 p53$^{-/-}$ cells managed to grow successfully, and a total of 25 clones were isolated. LIG4 knockdown was verified by subjecting protein extracts to Western blot analysis (Figure 10). Three out of 25 clones (12%; lanes 1, 2, and 3) express GAPDH (38 kDa) but not LIG4 (103 kDa), confirming successful LIG4 knockdown.

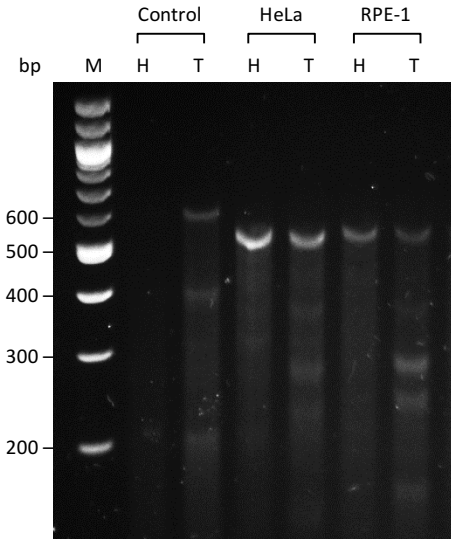

Figure 9. Assessment of CRISPR/Cas9-induced LIG4 mutations in HeLa and RPE-1 p53$^{-/-}$ cells. DNA from Cas9-expressing HeLa and RPE-1 p53$^{-/-}$ cells was subjected to the Surveyor nuclease assay. Wt and mutant 500-600 bp amplicons containing the LIG4 gRNA target region were generated by PCR and hybridised to form heteroduplexes (lanes marked 'T'). Wt or mutant amplicons alone were self-hybridised to form control homoduplexes (lanes marked 'H'). Duplexes were treated with 2 µl of the mismatch-specific Surveyor nuclease for 1 h at 42°C and resolved on a 2% (w/v) agarose gel. CRISPR/Cas9-induced mutation formation was confirmed in both cell lines, as revealed by the presence of cleavage products in lanes marked 'T'. Control samples (controls C and G) provided with the kit were amplified, hybridised and digested as described above (expected digestion products: 217 and 416 bp). Selected band sizes (given in bp) of the DNA marker (M) are indicated. Image was captured with a U:Genius3 Gel Documentation System. Abbreviations: Bp, base pair; gRNA, guide-RNA; LIG4, DNA ligase 4; wt, wild-type, w/v, weight per unit volume.

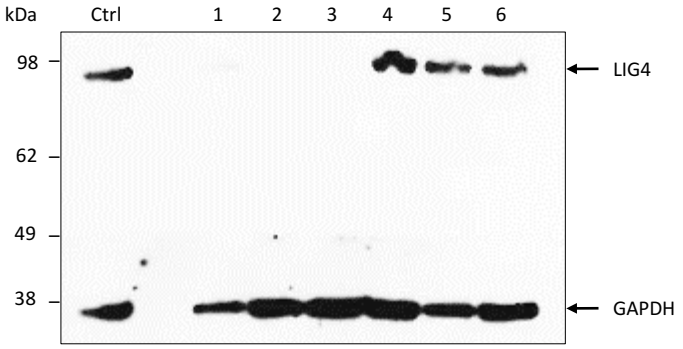

Figure 10. Assessment of LIG4 knockdown in RPE-1 Cas9 p53$^{-/-}$ cell lines. Protein lysates were resolved on a 4-12% gel by SDS-PAGE, and expression of LIG4 (103 kDa) and GAPDH (38 kDa) loading control was assessed by Western blot analysis with Anti-DNA Ligase IV (Abcam, 193353) and Rabbit Anti-GAPDH (Raybiotech, 168-10591) antibodies. Three LIG4-negative cell lines were detected (lanes 1, 2, and 3). Protein extract from HeLa Cas9 cells served as positive control (Ctrl). Abbreviations: GAPDH, Glyceraldehyde 3-phosphate dehydrogenase; LIG4, DNA ligase 4; kDa, kilodalton; SDS-PAGE, sodium dodecyl sulphate polyacrylamide gel electrophoresis.

4.2 CRISPR/Cas9 vector construction

A multiplex CRISPR/Cas9 vector for simultaneous delivery of the eight gRNAs was constructed by Golden Gate cloning. Ligation reactions with varying molar ratios of vector and insert (1:25 and 1:5) were prepared, and the construct was transformed into NEB Stable Competent *E. coli* cells, single colonies were isolated and expanded, and plasmids from multiple clones were purified and analysed on 0.6 % agarose gels. None of the resulting recombinant vectors had the anticipated molecular size (not shown). To assess whether the incorrect vector size was caused by erroneous ligation, another multiplex gRNA insert was assembled by Golden Gate cloning and analysed on a 0.8% (w/v) agarose gel (Figure 11A). A band corresponding to the expected size of the multiplex insert (~1500 bp, indicated by arrow) was detected, accompanied by seven partial ligation products (lower bands). The insert was gel purified and cloned into a BsmbI-digested lentiGuide-Puro vector in a 2:1 molar ratio, and recombinant vector (RV) purified from a colony of transformed NEB Stable Competent *E. coli* cells was resolved on a 0.6% (w/v) agarose gel (Figure 11B). The observed band (~5.0 kb) does not correspond to the expected size of the RV (~9.7 kb), thus a different approach was taken in which single gRNAs were ligated into separate lentiGuide-Puro vectors (not shown). The eight constructs were all successfully generated, as verified by Sanger sequencing.

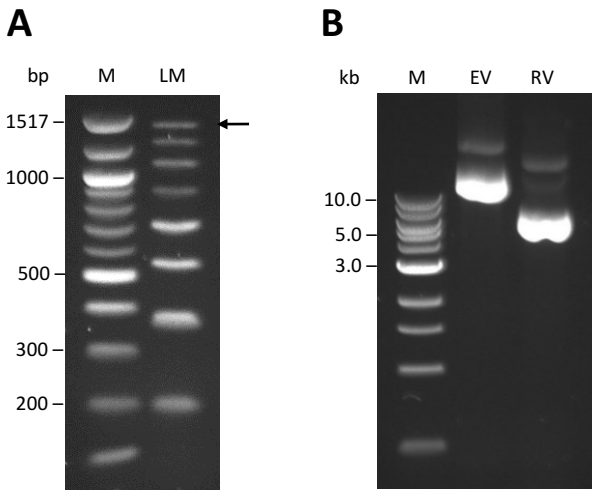

Figure 11. Construction of the multiplex CRISPR/Cas9 vector. A) A multiplex insert containing eight gRNAs (gRNAs 1-8) was generated by Golden Gate cloning. The Golden Gate ligation mix (LM) was resolved on a 0.8% (w/v) agarose gel, which showed the presence of the multiplex insert (~1500 bp, indicated by arrow) and seven partial ligation products (lower bands). B) The multiplex gRNA insert from A was gel purified and cloned into a BsmbI-digested lentiGuide-Puro vector in a 2:1 molar ratio. The recombinant vector (RV) was resolved on a 0.6% (w/v) agarose gel, and empty lentiGuide-Puro vector (EV) served as positive control. The observed band (~5.0 kb) does not correspond to the expected size of the RV (~9.7 kb), indicating unsuccessful vector construction. Selected band sizes (given in bp/kb) of the DNA markers (M) are indicated. Images were captured with a U:Genius3 Gel Documentation System. Abbreviations: bp, base pair; kb, kilobase; w/v, weight per unit volume.

4.3 CRISPR/Cas9 validation

Cas9-expressing HeLa and RPE-1 $p53^{-/-}$ cells were infected with lentiviruses for the delivery of single gRNAs in eight separate transductions. In order to verify the generation of CRISPR/Cas9-induced DSBs and assess the gRNA targeting efficiencies, genomic DNA from infected cells was subjected to the Surveyor nuclease assay. The assay was optimised for each of the gRNAs, with the four variables being PCR annealing temperature, number of PCR cycles, amount of Surveyor nuclease, and digestion incubation time (Table 11). 400-600 bp wild-type and mutant amplicons containing the gRNA target regions were generated by PCR and hybridised to form heteroduplexes. Duplexes were treated with the mismatch-specific Surveyor nuclease and resolved on 2% agarose gels. Under optimal conditions, DSB generation induced by gRNAs 1, 2, 3, 5, and 7 was validated in both HeLa (Figure 12) and RPE-1 (Figure 13) cells, as confirmed by the presence of new cleavage products in lanes marked

'T'. Optimal PCR conditions could not be identified for gRNAs 4 and 8, resulting in insufficient amounts of amplicons for duplex formation, and optimal digestion conditions were not identified for gRNA6. Accordingly, gene editing mediated by these gRNAs could not be verified, and they were therefore left out of the remainder of the study.

Table 11. Optimal conditions for the Surveyor nuclease assay.

Parameter	gRNA1	gRNA2	gRNA3	gRN4	gRNA5	gRNA6	gRNA7	gRNA8
Annealing temp. (°C)	65	65	65	–	65	65	65	–
No. of PCR cycles	30	30	30	–	30	30	30	–
Digestion time (h)	1	1	1	–	1	–	1	–
Surveyor nuclease (µl)	2	1	2	–	2	–	1	–

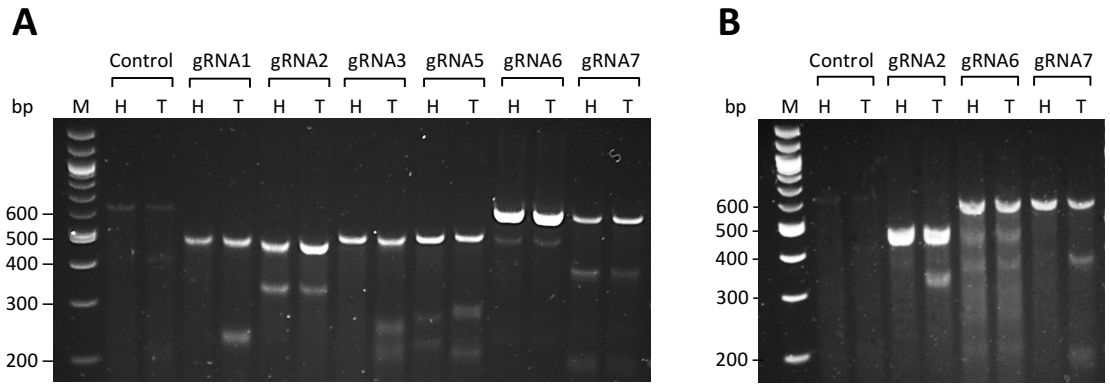

Figure 12. Assessment of CRISPR/Cas9-induced DSBs in HeLa Cas9 cells. Cells were infected separately with lentiviruses for the delivery of single gRNAs. DNA was extracted and subjected to the Surveyor nuclease assay. 400-600 bp wt and mutant amplicons containing the gRNA target regions were generated by PCR and hybridised to form heteroduplexes (lanes marked 'T'). Wt or mutant amplicons alone were self-hybridised to form control homoduplexes (lanes marked "H"). Duplexes were treated with 2 µl (panel A) or 1 µl (panel B) of the mismatch-specific Surveyor nuclease for 1 h at 42 °C, except for the gRNA6 duplexes in panel B which were digested with 2 µl of nuclease for 2 h at 42 °C. Duplexes were resolved on 2% (w/v) agarose gels, and generation of DSBs induced by gRNAs 1, 2, 3, 5 and 7 was validated, as confirmed by the presence of new cleavage products in lanes marked 'T'. Control samples (controls C and G) provided with the Surveyor Nuclease kit were amplified, hybridised and digested as described above (expected digestion products: 217 and 416 bp). Selected band sizes (given in bp) of the DNA marker (M) are indicated. Images were captured with a U:Genius3 Gel Documentation System. Abbreviations: Bp, base pair; DSB, double-strand break; gRNA, guide-RNA; wt, wild-type; w/v, weight per unit volume.

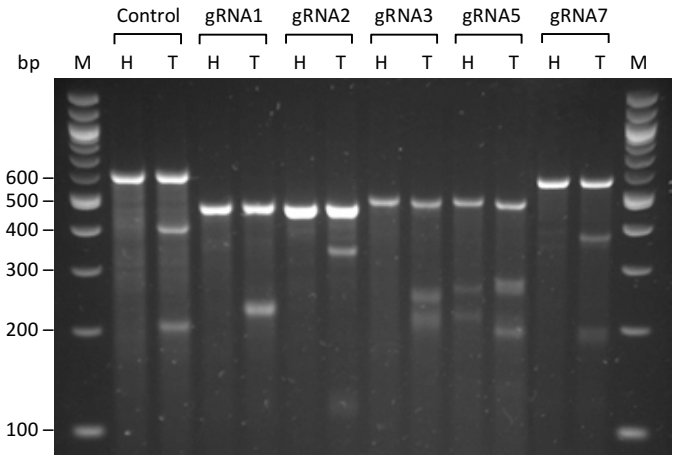

Figure 13. Assessment of CRISPR/Cas9-induced DSBs in RPE-1 Cas9 p53$^{-/-}$ cells. Cells were infected separately with lentiviruses for the delivery of single gRNAs. DNA was extracted and subjected to the Surveyor nuclease assay. 400-600 bp wt and mutant amplicons containing the gRNA target regions were generated by PCR and hybridised to form heteroduplexes (lanes marked 'T'). Wt or mutant amplicons alone were self-hybridised to form control homoduplexes (lanes marked 'H'). Duplexes were treated with 2 µl (gRNA1/3/5/6) or 1 µl (gRNA2/7) of the mismatch-specific Surveyor nuclease for 1 h at 42°C and resolved on 2% (w/v) agarose gels. Gene editing mediated by gRNAs 1, 2, 3, 5, and 7 was validated, as confirmed by the presence of new cleavage products in lanes marked 'T'. Control samples (controls C and G) provided with the kit were amplified, hybridised and digested as described above (expected digestion products: 217 and 416 bp). Selected band sizes (given in bp) of the DNA marker (M) are indicated. Image was captured with a U:Genius3 Gel Documentation System. Abbreviations: bp, base pair; DSB, double-strand break; gRNA, guide-RNA; wt, wild-type; w/v, weight per unit volume.

4.4 Assessment of gRNA targeting efficiency

The targeting efficiencies of gRNAs 1, 2, 3, 5, and 7 were calculated using ImageJ and are listed in Table 12. Targeting efficiencies appear relatively consistent across the two cell types.

Table 12. gRNA targeting efficiencies.

	Targeting efficiency (%)	
	HeLa Cas9	**RPE-1 Cas9 p53$^{-/-}$**
gRNA1	20.7	19.2
gRNA2	12.9	13.4
gRNA3	15.5	19.4
gRNA5	18.8	29.3
gRNA7	23.8	20.6

Abbreviations: gRNA, guide-RNA.

4.5 Generation of barcoded synthetic translocations

Barcoded synthetic translocations (STs) for normalisation of the nested PCR step were generated by directional ligation of 5'-side and 3'-side DNA fragments to a 20 bp non-genomic barcode oligo containing a HindIII restriction site (Figure 14). All STs are denoted in the 5' to 3' direction (e.g. 1-2 is made up of sequences from the 5'-side and 3'-side of the genes targeted by gRNAs 1 and 2, respectively). To confirm correct ligation of the three fragments, STs were treated with HindIII for 1 h at 37°C and resolved on a 1.9% agarose gel along with undigested samples (Figure 15). Digestion products of the correct molecular sizes (Table 13) were detected for all samples in lanes marked 'D'. Furthermore, the presence of bands corresponding to the expected molecular sizes (Table 14) of the STs (red boxes in lanes marked 'C') indicates successful ligation, which was also confirmed by Sanger sequencing. Figure 15 only displays selected STs; however, all STs except for 1-8, 2-6, 6-2, and 7-6 were confirmed by restriction enzyme analysis and Sanger sequencing.

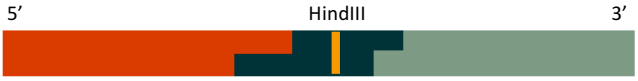

Figure 14. Illustration of a barcoded synthetic translocation. The barcoded synthetic translocations have the exact same sequences as the endogenous CRISPR/Cas9-induced translocations. They consist of a 5'-side fragment (red) and a 3'-side fragment (light blue) connected by a unique, non-genomic 20-bp barcode sequence (dark blue) to distinguish them from the endogenous translocations. The barcode sequence contains a HindIII restriction site (yellow).

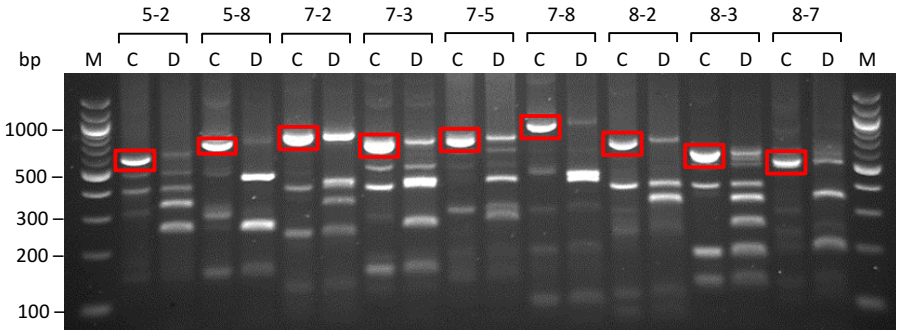

Figure 15. Restriction enzyme analysis of selected barcoded synthetic translocations. 200-400 ng of the STs were digested with 10.0 units of HindIII-HF for 1 h at 37°C. Digestion products (lanes marked 'D') were resolved on a 1.9% (w/v) agarose gel, and undigested STs (lanes marked 'C') served as controls. Here, a selection of nine STs are shown, all named in the 5' to 3' direction (e.g. 8-7 is made up of sequences from the 5'-side and 3'-side of the genes targeted by gRNAs 8 and 7, respectively). Successful generation of STs was confirmed both by the presence of bands with sizes matching the expected sizes of the STs (red boxes in lanes marked 'C') and the presence of digestion products in lanes marked 'D'. Selected band sizes (given in bp) of the DNA marker (M) are indicated. Image was captured with a U:Genius3 Gel Documentation System. Abbreviations: bp, base pair; ST, barcoded synthetic translocation; w/v, weight per unit volume.

Table 13. Expected sizes of ST side fragments.

5'-fragment	Size (bp)	3'-fragment	Size (bp)
1	256	1	246
2	147	2	332
3	255	3	259
4	359	4	237
5	235	5	278
6	359	6	264
7	409	7	194
8	318	8	453

Bp, base pair; ST, barcoded synthetic translocation.

Table 14. Expected sizes of barcoded synthetic translocations.

ST	Size (bp)	ST	Size (bp)	ST	Size (bp)	ST	Size (bp)
1-2	614	3-1	527	5-1	507	7-1	681
1-3	541	3-2	613	5-2	593	7-2	767
1-4	519	3-4	518	5-3	520	7-3	694
1-5	560	3-5	559	5-4	374	7-4	672
1-6	546	3-6	545	5-6	525	7-5	713
1-7	476	3-7	475	5-7	455	7-6	699
1-8	735	3-8	734	5-8	714	7-8	888
2-1	419	4-1	631	6-1	631	8-1	590
2-3	432	4-2	628	6-2	717	8-2	676
2-4	410	4-3	644	6-3	644	8-3	603
2-5	451	4-5	663	6-4	622	8-4	457
2-6	437	4-6	560	6-5	663	8-5	622
2-7	367	4-7	579	6-7	579	8-6	608
2-8	626	4-8	838	6-8	838	8-7	538

Bp, base pair; ST, barcoded synthetic translocation.

4.6 Capturing of translocation junctions

Prior to capturing CRISPR/Cas9-induced translocations by nested PCR, the external nested PCR primers were subjected to negative testing to confirm that no amplification occurred unless a translocation had formed. Genomic DNA from wt RPE-1 cells was used as template, and all 56 combinations of forward and reverse primers were tested. A combination of primers for amplification of the same gRNA target region (e.g. gRNA1-Fwd Ext and gRNA1-Rev Ext) served as positive controls, and PCR products were resolved on a 1.8% (w/v) agarose gel (Figure 16). Primer combinations are denoted in the 5' to 3' direction, e.g. "1-2" is a combination of gRNA1-Fwd Ext and gRNA2-Rev Ext primers. Most primer combinations had a high specificity; however, a few non-specific bands were observed for combinations with the gRNA4/6/8 forward or reverse primers (e.g. 2-4, 4-3, and 8-4). This, in addition to the fact

that optimal Surveyor nuclease assay conditions could not be identified (see above), contributed to the decision to omit gRNAs 4, 6, and 8 from the study. A non-specific band for the primer combination 7-3 was also observed, but as the band did not fall within the molecular size range of the translocations (350-900 bp), this deviancy was disregarded and PCR conditions were kept unchanged.

Once negative testing of the external nested PCR primers was conducted, they were used to capture CRISPR/Cas9-induced translocations in Cas9-expressing RPE-1 $p53^{-/-}$ cells (Figure 17). Two LIG4-negative clones (Clone 1 and 2) and LIG4-positive RPE-1 cells (wt) were infected either once (Panel A) or twice (Panel B) with equal volumes of lentiviruses for the simultaneous delivery of gRNA1/2/3/5/7. Both external and internal PCR reactions were performed in five separate reactions using a single forward primer combined with the remaining reverse primers (Table 15). For instance, in lane 1F, the gRNA1 forward primer was combined with the gRNA2/3/5/7 reverse primers to capture translocations 1-2, 1-3, 1-5, and 1-7, respectively. PCR products were resolved on 1.8% (w/v) agarose gels, and DNA from non-infected cells served as negative controls (Panel C). The expected molecular sizes of the translocation amplicons lie in the range 400-800 bp. Although some weak bands within this range can be detected, a number of strong, non-specific bands of low molecular size (150-300 bp) are present in most of the lanes. These bands are also observed in the non-infected samples (Panel C).

Prior to transduction, the lentiviruses were titrated to ensure that the target cells were treated with equal quantities of each virus. Virus titration was performed by measuring the levels of a lentiviral-specific transgene (WPRE) and a single copy gene-specific reference gene (albumin) using real-time PCR. However, due to low DNA concentrations, the results from this experiment were not conclusive and are therefore not presented here. Thus, in this pilot experiment, we decided to use equal volumes of each virus to infect the target cells.

Table 15. Primer combinations for capturing of translocation junctions by nested PCR.

F1	F2	F3	F5	F7
gRNA1-Fwd	gRNA2-Fwd	gRNA3-Fwd	gRNA5-Fwd	gRNA7-Fwd
gRNA2-Rev	gRNA1-Rev	gRNA1-Rev	gRNA1-Rev	gRNA1-Rev
gRNA3-Rev	gRNA3-Rev	gRNA2-Rev	gRNA2-Rev	gRNA2-Rev
gRNA5-Rev	gRNA5-Rev	gRNA5-Rev	gRNA3-Rev	gRNA3-Rev
gRNA7-Rev	gRNA7-Rev	gRNA7-Rev	gRNA7-Rev	gRNA5-Rev

Abbreviations: Fwd, forward; Rev, reverse.

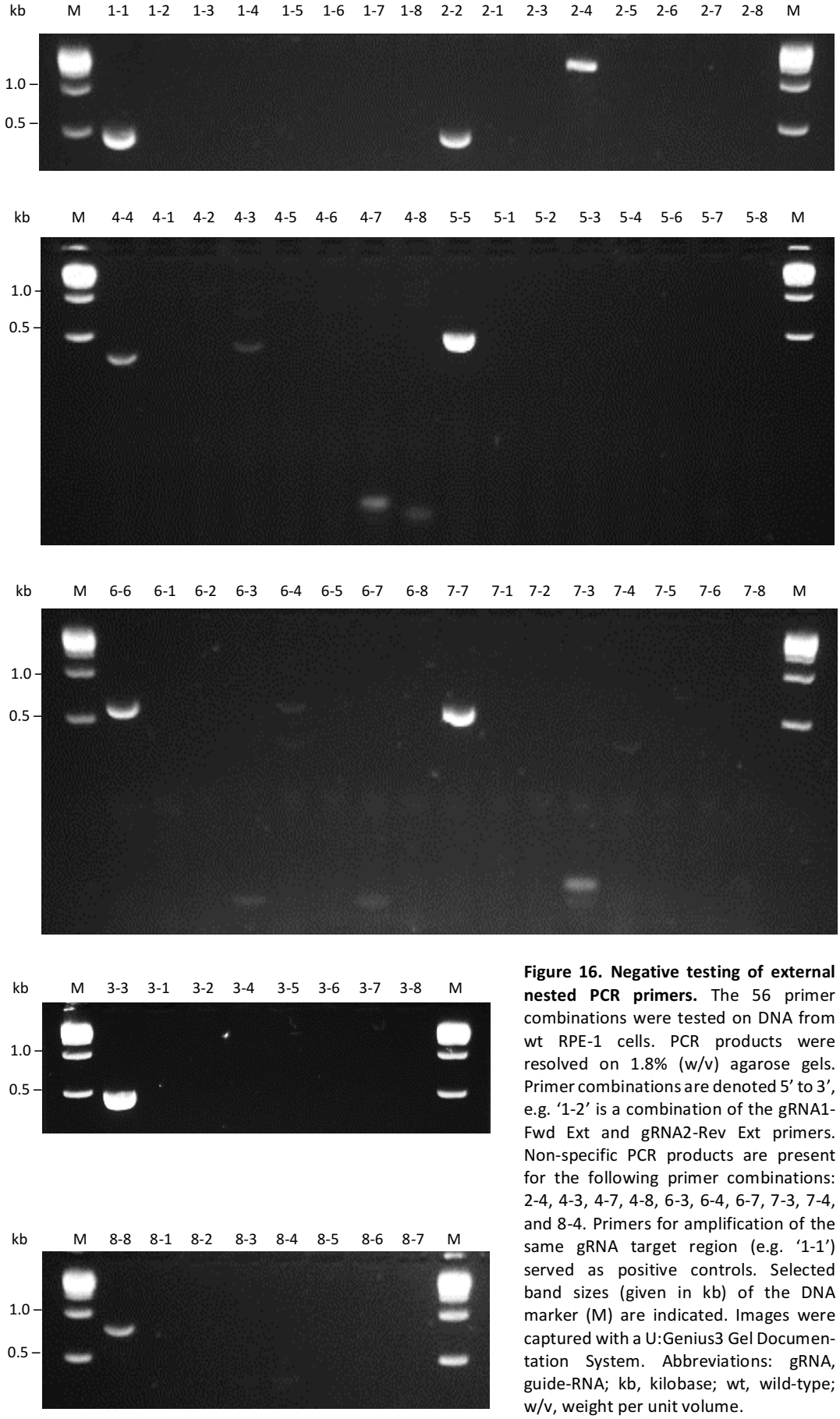

Figure 16. Negative testing of external nested PCR primers. The 56 primer combinations were tested on DNA from wt RPE-1 cells. PCR products were resolved on 1.8% (w/v) agarose gels. Primer combinations are denoted 5' to 3', e.g. '1-2' is a combination of the gRNA1-Fwd Ext and gRNA2-Rev Ext primers. Non-specific PCR products are present for the following primer combinations: 2-4, 4-3, 4-7, 4-8, 6-3, 6-4, 6-7, 7-3, 7-4, and 8-4. Primers for amplification of the same gRNA target region (e.g. '1-1') served as positive controls. Selected band sizes (given in kb) of the DNA marker (M) are indicated. Images were captured with a U:Genius3 Gel Documentation System. Abbreviations: gRNA, guide-RNA; kb, kilobase; wt, wild-type; w/v, weight per unit volume.

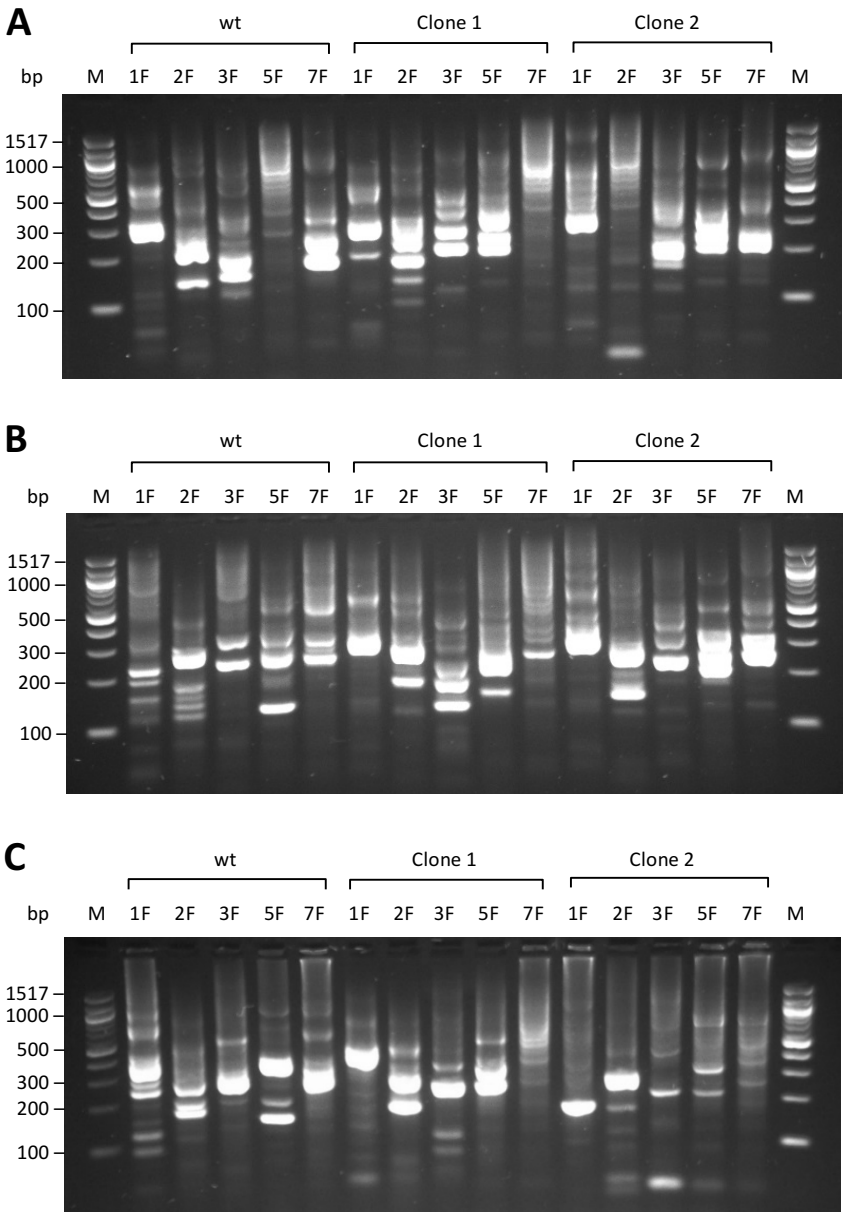

Figure 17. Capturing of CRISPR/Cas9-induced translocations formed in RPE-1 Cas9 p53-/- cells. LIG4-positive (wt) and two LIG4-negative clones (Clone 1 and Clone 2) were transduced with a combination of lentiviruses for the delivery of gRNAs 1, 2, 3, 5, and 7 either once (panel A) or twice (panel B). Non-transduced cells (panel C) served as controls. Genomic DNA was extracted, and the 20 translocations were captured by nested PCR. Both external and internal PCR reactions were performed in five separate reactions, in which each of the gRNA forward primers (1F, 2F, 3F, 5F, and 7F) were combined with the remaining gRNA reverse primers (e.g. in lane 1F, the gRNA1 forward primer is combined with the reverse primers for gRNAs 2, 3, 5, and 7). Amplified translocation junctions were resolved on 1.8% (w/v) agarose gels. Bands corresponding to the expected sizes of the translocation amplicons (400-800 bp) are observed; however, a number of non-specific products are present in all lanes, indicating too low PCR stringency. Selected band sizes (given in bp) of the DNA marker (M) are indicated. Images were captured with a U:Genius3 Gel Documentation System (Syngene International). Abbreviations: bp, base pair; ST, barcoded synthetic translocation; w/v, weight per unit volume.

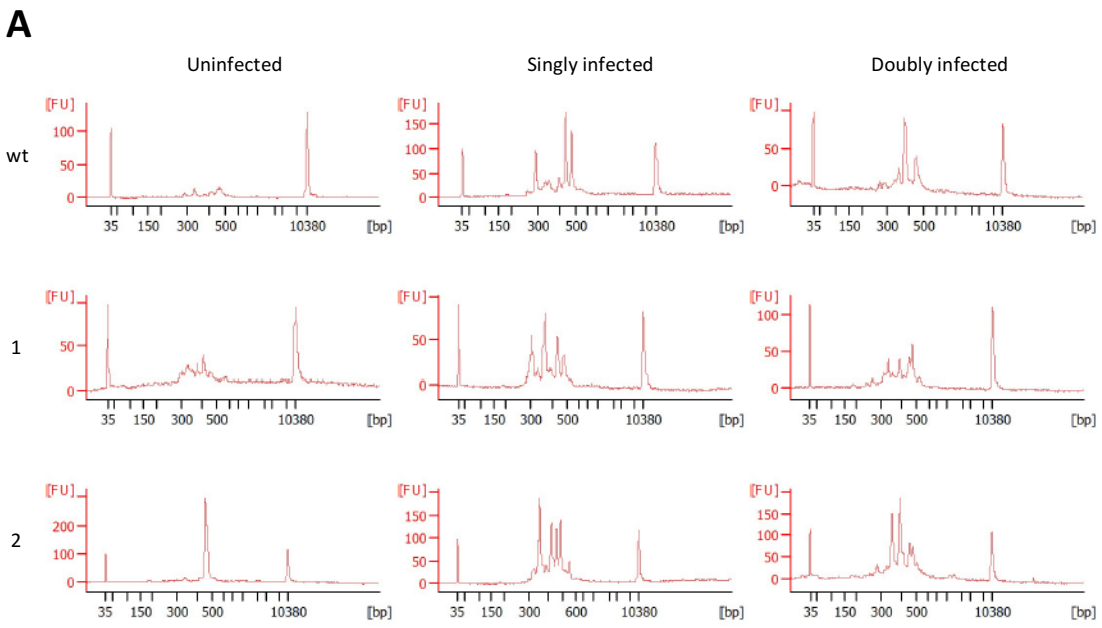

Figure 18. Analysis of CRISPR/Cas9-induced translocation junctions from RPE-1 Cas9 p53$^{-/-}$ cells. Nested PCR amplicons from uninfected (U), singly infected (S) and doubly infected (D) wt and LIG4$^{-/-}$ (clone 1/2) cells were indexed for Illumina sequencing and analysed in a Bioanalyzer (Agilent) with a lower (~35 bp) and upper (~10,380 bp) marker. DNA fragments can be visualised as peaks in an electropherogram where fluorescence intensity (FU) is plotted against size/migration time (Panel A) or as bands in a gel-like densitometry plot (Panel B). Translocation junctions are detected in the infected samples as indicated by the appearance of novel peaks/bands absent in uninfected samples. Notably, peak/band patterns differ between infected wt and LIG4$^{-/-}$ samples, demonstrating that inactivation of C-NHEJ gives a distinct set of translocation events.

Nested PCR amplicons from wt and LIG4$^{-/-}$ (Clone 1 and 2) RPE-1 Cas9 p53$^{-/-}$ cells were indexed for Illumina sequencing and analysed in a Bioanalyzer (Agilent) by Victor Gonzalez Huici and Vildan Bozok Cetintas (Figure 18). Both uninfected (U), singly infected (S) and doubly infected (D) samples were analysed. DNA fragments can be visualised as peaks in an electropherogram where fluorescence intensity (FU) is plotted against size/migration time (Panel A) or as bands in a gel-like densitometry plot (Panel B). Translocation junctions are detected in the infected samples, as indicated by the appearance of novel peaks in the electropherograms and novel, individualised bands in the densitometry plot absent in the uninfected samples. Notably, peak/band patterns differ between infected wt and LIG4$^{-/-}$

samples, demonstrating that inactivation of C-NHEJ gives a distinct set of translocation events.

4.7 Sanger sequencing of translocation junctions

Nested PCR of DNA from RPE-1 Cas9 $p53^{-/-}$ cells produced several non-specific PCR products, making it difficult to identify genuine translocation junctions. However, some of the band patterns of the infected samples differed from the uninfected control samples (Figure 17) and were therefore subjected to Sanger sequencing. Sequencing revealed that four of the translocation junctions (1-5, 1-7, 5-3 and 5-7) were successfully captured by nested PCR from cells transduced with virus once, with tidy and unambiguous sequences that corresponded exactly to those of the expected translocations. In cells transduced twice, two translocations were confirmed by sequencing (5-3 and 5-7) and some unexpected sequences were identified given the primers used.

5 Discussion

In recent years, the alternative non-homologous end joining (Alt-EJ) pathway for repair of DNA double-strand breaks has been implicated in translocation formation, which may contribute to cancer onset and progression. However, the underlying mechanism of Alt-EJ-induced translocations and the factors involved remain poorly understood due to the lack of a general Alt-EJ assay in the native chromatin complex.

The topic of my thesis was technology development rather than addressing a biological question. Thus, the main output of my thesis is the progress I have contributed to in the development of a genomic assay termed QMAT-seq for analysis of mutations associated with the Alt-EJ pathway (Figure 19). In particular, I have generated LIG4$^{-/-}$ cell lines and synthetic translocations, cloned gRNAs, and checked their efficiency in generating double-strand breaks. I have also performed some pilot experiments with a subset of gRNA sequences and captured the resulting translocation junctions by nested PCR. High-throughput sequencing libraries have been generated by Victor Gonzalez Huici and Vildan Bozok Cetintas, and the sequencing data is currently being analysed by Katerina Bouefa (PhD student in our lab).

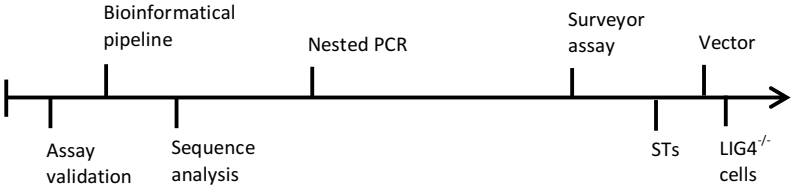

Figure 19. Approximate progress of QMAT-seq development. To date, I have successfully generated LIG4$^{-/-}$ cell lines and eight CRISPR/Cas9 vectors encoding single gRNAs. I have also produced 52/56 synthetic translocations (STs) and identified optimal Surveyor assay conditions for five of the eight gRNAs. Moreover, I have performed some pilot experiments with a subset of gRNAs and managed to capture four distinct translocation junctions by nested PCR. Future perspectives include further PCR optimisation, analysis of translocation sequences, development of a bioinformatical pipeline and assay validation.

5.1 Cell line choice

Cas9-expressing HeLa and RPE-1 p53$^{-/-}$ cell lines were selected for this study due to their fast proliferation rates, high plating-efficiency, and easy maintenance. However, only RPE-1 cells managed to grow successfully when knocking down LIG4 and isolating single clones to derive isogenic LIG4$^{-/-}$ cell lines. The HeLa cells were p53-proficient, thus the reduced cell viability observed in this cell line may be attributed to p53-induced apoptosis upon loss of C-NHEJ

(Menon & Povirk 2014). Consistent with this notion, both LIG4- and XRCC4-deficiency in mice leads to embryonic lethality via the p53 cell cycle checkpoint pathway, and depletion of p53 rescues embryonic lethality and impaired cellular proliferation (Frank et al. 2000; Gao et al. 2000). Hence, it might be advantageous to apply QMAT-seq in p53-deficient cells in future experiments to ensure cell viability.

Moreover, QMAT-seq should be tested across a panel of different cell lines to assess whether there are any cell-dependent differences that might affect the assay outcome. For instance, it has been reported that HeLa cells exhibit a different gene expression profile from normal human cells, including signs of chromothripsis and high expression of numerous repair factors (Landry et al. 2013). Thus, QMAT-seq outcomes should be interpreted with care and it could be advantageous to use the genome of the model cell line as a genomic reference when characterising the mutational signature of Alt-EJ.

5.2 CRISPR/Cas9 vector construction

Originally, we sought to use a multiplex CRISPR/Cas9 approach for the simultaneous delivery of the eight gRNAs, but the cloning of the multiplex vector proved problematic; all transformants harboured vectors of incorrect molecular size, suggesting that a recombination event might have occurred. Indeed, even though using a $RecA^-$ strain, which greatly reduces recombination of cloned DNA, other RecA-independent recombination mechanisms exist (Ozgenc et al. 2005; Azpiroz & Laviña 2017; Dutra et al. 2007). Moreover, the multiplex insert contains repetitive DNA sequences and may, therefore, be particularly susceptible to recombination (Al-Allaf et al. 2013; Bzymek & Lovett 2001).

In this study, I show that co-expression of CRISPR/Cas9 constructs with single gRNAs is a feasible alternative to the multiplex approach. However, despite a straightforward cloning step, this approach is less accurate and more time-consuming. First, it requires an additional virus titration step to ensure that the target cells receive equal quantities of each gRNA. Secondly, only a subset of the target cells will receive all gRNA sequences, making it hard to accurately measure and compare translocation frequencies. Therefore, I believe future research should focus on generating a multiplex vector, particularly since several research groups show this to be feasible. For instance, Xie et al. have been successful in cloning multiplex vectors with up to eight gRNAs under the control of a single promoter in DH5α *E. coli* cells (Xie et al. 2015), thus it might be useful to test other bacterial strains suitable for difficult clone constructs. Alternatively, the gRNA sequences could be expressed from

independent promoters to reduce the probability of recombination events between the repetitive sequences (Kabadi et al. 2014; Gou et al. 2007).

5.3 Lentiviral infection

In this study, RPE-1 Cas9 $p53^{-/-}$ $LIG4^{-/-}$ cells were either singly or doubly infected with viruses to assess whether it would be possible to increase the number of cells receiving all of the gRNAs, and thus increase translocation frequencies, by infecting the cells twice. However, only two translocation junctions were confirmed by sequencing in doubly infected cells as compared to four translocation junctions in singly infected cells. Moreover, unexpected sequences were identified in samples from doubly infected cells given the primers used for sequencing, which may indicate the occurrence of additional translocation events. Thus for future QMAT-seq analyses, a single virus infection seems appropriate as to prevent multiple translocation events and to preserve the initial mutational pattern left by Alt-EJ.

5.4 Optimisation of the Surveyor nuclease assay

Optimal Surveyor nuclease assay conditions could not be identified for gRNAs 4, 6, and 8; the concentrations of the DNA amplicons generated in step 1 of the assay were consistently too low, and the corresponding gel images (not shown) often showed the presence of multiple non-specific PCR products, suggesting poor primer design. Thus, for future experiments, it could be advantageous to test a set of different PCR primers for validation of these gRNAs. I would also like to point out the signs of unspecific cleavage in the gRNA5 homoduplex control sample (Figure 12A and Figure 13). To increase the cleavage specificity, I would suggest reducing the amount of Surveyor nuclease, shortening the digestion time, or a combination of the two in future experiments.

5.5 Assessment of gRNA targeting efficiency

ImageJ analyses of Surveyor nuclease cleavage band intensities is a rapid way to identify gRNA targeting efficiencies. However, this approach is somewhat inaccurate as band intensities are found by manually closing off the area under the curve and the decision on where the peak starts and ends is subjective and may cause inaccuracies of up to approximately 5%. Thus, in the future, we need a more accurate way of determining the efficiency of CRISPR/Cas9-induced double-strand breaks to normalise the translocation read

counts. DNA double-strand breaks trigger phosphorylation of the histone variant H2AX as part of the cells' repair response (Kuo & Yang 2008; Scully & Xie 2013). As the intensity of DNA damage is proportional to the levels of γH2AX, one approach could be to perform chromatin immunoprecipitation (ChIP) with anti-γH2AX antibody followed by quantification by quantitative PCR to measure gRNA targeting efficiencies.

5.6 Optimisation of nested PCR

When capturing translocation junctions from RPE-1 Cas9 $p53^{-/-}$ $LIG4^{-/-}$ cells by nested PCR, a number of strong, non-specific bands of low molecular size (150-300 bp) were produced. These bands were also observed in the uninfected samples in which no amplicons should be present, suggesting a too low PCR stringency and the need for further optimisation. To identify optimal conditions for the multiplex nested PCR, I would set up multiple reactions and systematically change variables such as template concentration, annealing temperature, and extension time. Moreover, the appropriate amounts of barcoded synthetic translocations (STs) to be added to the nested PCR reaction should be identified by setting up PCR reactions with different ST-to-template ratios; for instance no STs, equimolar amounts of STs, and tenfold as much STs.

5.7 Validation of QMAT-seq

Once QMAT-seq has been established, the assay should be validated by perturbing multiple proteins known to regulate Alt-EJ and translocation formation, such as POLQ and LIG3, and assess whether they change the translocation frequencies consistent with what is known in the literature (Figure 20). Validation will be performed in $LIG4^{-/-}$ cells to ensure that repair outcomes are a measurement of Alt-EJ activity only, and the junction sequences will be analysed to detect the presence of microhomologies, which are a characteristic of Alt-EJ repair junctions (Schimmel et al. 2017; Ottaviani et al. 2014). If sensitive enough, QMAT-seq should show a reduction in translocation frequency upon knockdown of known Alt-EJ factors and a significant change in the mutational pattern at the translocation junctions.

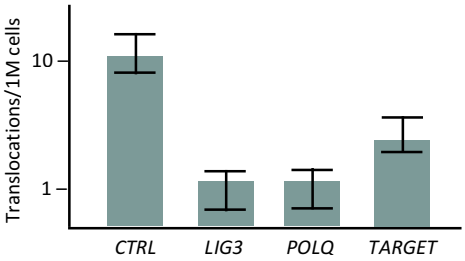

Figure 20. Validation and future applications of QMAT-seq. QMAT-seq will be validated by knocking down several genes whose protein products regulate Alt-EJ, such as LIG3 and POLQ, to assess if they alter the translocation frequencies consistent with what is known in the literature. *TARGET* represents a novel Alt-EJ gene that we hope to identify using QMAT-seq in the future.

5.8 Implications for future practice and research

QMAT-seq raises a number of opportunities for future research and practice, both in terms of fundamental research and clinical practice. When fully developed and validated, QMAT-seq will allow for the identification of novel Alt-EJ factors by assessing changes in translocation frequencies and mutational patterns at the repair junctions upon knockdown of potential Alt-EJ factors. A significant alteration of the mutational pattern, as well as a reduction in translocation frequencies, may indicate that the candidate protein is implicated in Alt-EJ repair and translocation formation (Figure 20). As Alt-EJ is associated with cancer onset and progression, this would have great clinical relevance in terms of identifying novel targets for cancer therapy.

QMAT-seq also has the potential to better characterise the mutational signature of Alt-EJ through deep sequencing and analysis of a large number of translocatsion junctions, which could have implications for cancer diagnostics and therapy. As Alt-EJ functions as a backup pathway when HR and C-NHEJ are impaired, the non-random mutational pattern of Al-EJ can serve as a biomarker in patients with C-NHEJ- or HR-deficiency (Simsek & Jasin 2010; Ceccaldi et al. 2015). Currently, the Alt-EJ signature is widely associated with the presence of microhomologies at the repair junction (Ottaviani et al. 2014; Schimmel et al. 2017; Ghezraoui et al. 2014); however, not all forms of Alt-EJ appear to rely on the presence of microhomologies (Decottignies 2013; Mansour et al. 2010). A better-defined Alt-EJ signature may aid in predicting which patients would benefit from therapies directed at targeting the

Alt-EJ pathway for a synthetic lethal effect in which cancer cells that depend on this pathway are selectively killed. Furthermore, Alt-EJ has been implicated in acquired therapy resistance (Edwards et al. 2008; Sakai et al. 2008), but mutational patterns indicative of resistance mechanisms have yet to be characterised (Vanderstichele et al. 2017). Thus, it would be interesting to apply QMAT-seq in therapy-resistant cancer cells to define a biomarker for acquired therapy resistance, which could aid in developing more efficient personalised cancer treatments with fewer side effects.

It would also be interesting to assess the role of epigenetics in translocation formation and Alt-EJ repair. Histone modifications have been implicated in regulation of the DNA damage response and in the recruitment of important repair factors to sites of DNA damage (Price & D'Andrea 2013; Dimitrova et al. 2015). Moreover, chromatin architecture is largely dictated by histone marks, and it has been suggested that the accessibility of DNA influences susceptibility to DSBs and translocation formation (Roukos et al. 2013; Bannister & Kouzarides 2011). By using epigenetic inhibitors, such as histone deacetylase (HDAC) and DNA methyltransferase (DNMT) inhibitors, and assess the effect of these changes on translocation frequencies, QMAT-seq will allow for studies on how specific histone marks and transcription state influence repair Alt-EJ and translocation formation.

Other molecular features that make certain regions of the genome particularly susceptible to breakage and translocation formation is still an active field of research. By targeting genomic regions in the native chromatin context, QMAT-seq enables studies of genomic context-dependent differences, such as sequence-specific differences (e.g. Alu repeats and CpG islands). Moreover, as QMAT-seq targets genes on different chromosomes with distinct nuclear locations, the assay can also provide insight into the effects of the spatial arrangement of chromosomes and genes on translocation formation.

5.9 Strengths and limitations of QMAT-seq

QMAT-seq is a robust assay with several strengths. Unlike most current assays, which study recurrent chromosomal translocations (Brunet et al. 2009; Torres et al. 2014), QMAT-seq allows for assessment of repair outcomes of translocations at user-defined sites in the genome. This, combined with the fact that it works at native chromosomal loci rather than on plasmids as the majority of current DNA repair assays do (Frit et al. 2014; Bindra et al. 2013), enables assessment of a large number of translocations across diverse genomic contexts and determination of genomic context-dependent differences. QMAT-seq also

permits unbiased assessment of repair outcomes rather than specific types of microhomology as several established assays do (Sharma et al. 2015; Kostyrko & Mermod 2016), which will contribute to a better characterisation of the Alt-EJ mutational signature. Furthermore, in addition to providing information on the change in translocation frequency upon perturbation of certain proteins and histone modifications, QMAT-seq also gives information on mutagenic outcome, thus giving a more comprehensive picture of how these factors influence translocation formation and Alt-EJ repair. Finally, in contrast to many previously described assays (Seluanov et al. 2010; Kuhar et al. 2014; Frit et al. 2014), QMAT-seq is not dependent on fluorescent reporter proteins. This has several benefits as these assays often show low frequencies of fluorescent cells after DSB induction and are limited by the requirement for two DSBs at a single locus. Moreover, it has been suggested that mutagenic NHEJ repair of nuclease-induced DSBs may disrupt expression of the fluorescent reporter gene, leading to underestimation of repair activity (Bindra et al. 2013; Aubert et al. 2011).

Although QMAT-seq presents several advantages over pre-existing assays, it comes with a few limitations. As the CRISPR/Cas9 system is not removed after initial DSB induction, repair junctions could possibly be re-cleaved, thus disturbing the mutational pattern observed at these junctions. Moreover, the assay is not as rapid as other existing assays; it could take up to one month to obtain assay results due to time-consuming library preparation and sequence analysis, and the high-throughput sequencing is relatively costly. Lastly, it requires more expertise in terms of bioinformatical analysis of the translocation sequences. Nonetheless, QMAT-seq will be an important asset for analysis of Alt-EJ repair and translocations that will complement pre-existing assays and strengthen the research conclusions drawn from them.

5.10 Summary

The main objective of my thesis was to develop and implement a novel CRISPR/Cas9-based assay for analysis of Alt-EJ associated translocations. In the present study, I have accomplished several sub-aims of this process: I have generated C-NHEJ-deficient cell lines, cloned and validated gRNA sequences for induction of double-strand breaks and translocation formation, generated synthetic translocations, performed pilot experiments with a subset of gRNAs, and amplified the resulting translocation junctions. Collectively, my results strongly support the feasibility of QMAT-seq. However, QMAT-seq is still in its

preliminary stage, and future perspectives include nested PCR optimisation, quantification of translocation frequencies, sequencing by Illumina technology, and an extensive bioinformatics analysis of the translocation junctions. Moreover, the assay needs to be validated by perturbing known Alt-EJ factors and assessing if the resulting changes in translocation frequency and mutational patterns are consistent with what is known in the literature. Upon completion, QMAT-seq will provide a means of identifying novel Alt-EJ factors, subtypes of Alt-EJ, and uncovering their underlying mechanisms. Furthermore, characterisation of the Alt-EJ mutational signature can serve as a biomarker for HR- and C-NHEJ-deficiency, which can aid in developing personalised cancer therapies. In conclusion, QMAT-seq is a versatile assay that will provide insight into the Alt-EJ pathway, its contribution to translocation formation, and its role in carcinogenesis and thus promises to make a substantial contribution to the field of DNA repair and cancer biology.

References

Ahmad, A. et al., 2008. ERCC1-XPF Endonuclease Facilitates DNA Double-Strand Break Repair . *Molecular and Cellular Biology*, 28(16), pp.5082–5092. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2519706/.

Al-Allaf, F.A. et al., 2013. Remarkable stability of an instability-prone lentiviral vector plasmid in Escherichia coli Stbl3. *3 Biotech*, 3(1), pp.61–70. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3563744/.

Alexandrov, L.B. et al., 2013. Signatures of mutational processes in human cancer. *Nature*, 500(7463), pp.415–421. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3776390/.

Ambardar, S. et al., 2016. High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian Journal of Microbiology*, 56(4), pp.394–404. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5061697/.

Arlt, M.F. et al., 2006. Common fragile sites as targets for chromosome rearrangements. *DNA Repair*, 5(9), pp.1126–1135. Available at: http://www.sciencedirect.com/science/article/pii/S1568786406001509.

Aubert, M. et al., 2011. Successful Targeting and Disruption of an Integrated Reporter Lentivirus Using the Engineered Homing Endonuclease Y2 I-AniI M. Semple, ed. *PLoS ONE*, 6(2), p.e16825. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3036713/.

Azpiroz, M.F. & Laviña, M., 2017. Analysis of RecA-independent recombination events between short direct repeats related to a genomic island and to a plasmid in Escherichia coli K12 V. Souza, ed. *PeerJ*, 5, p.e3293. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5426353/.

Bacolla, A. et al., 2016. Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences. *Nucleic Acids Research*, 44(12), pp.5673–5688. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4937311/.

Bannister, A.J. & Kouzarides, T., 2011. Regulation of chromatin by histone modifications. *Cell Research*, 21(3), pp.381–395. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3193420/.

Bassing, C.H. & Alt, F.W., 2004. The cellular response to general and programmed DNA double strand breaks. *DNA Repair*, 3(8), pp.781–796. Available at: http://www.sciencedirect.com/science/article/pii/S1568786404001752.

Berger, M.F. et al., 2011. The genomic complexity of primary human prostate cancer. *Nature*, 470(7333), pp.214–220. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3075885/.

Bindra, R.S. et al., 2013. Development of an assay to measure mutagenic non-homologous end-joining repair activity in mammalian cells. *Nucleic Acids Research*, 41(11), pp.e115–e115. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3675474/.

Boboila, C. et al., 2010. Alternative end-joining catalyzes class switch recombination in the absence of both Ku70 and DNA ligase 4. *The Journal of Experimental Medicine*, 207(2), pp.417–427. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2822597/.

Boboila, C. et al., 2012. Robust chromosomal DNA repair via alternative end-joining in the absence of X-ray repair cross-complementing protein 1 (XRCC1). *Proceedings of the National Academy of Sciences of the United States of America*, 109(7), pp.2473–2478. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3289296/.

Bothmer, A. et al., 2013. Mechanism of DNA resection during intrachromosomal recombination and immunoglobulin class switching. *The Journal of Experimental Medicine*, 210(1), pp.115–123. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3549709/.

Bothmer, A. et al., 2011. Regulation of DNA End Joining, Resection, and Immunoglobulin Class Switch Recombination by 53BP1. *Molecular cell*, 42(3), pp.319–329. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3142663/.

Boulton, S.J. & Jackson, S.P., 1996. Saccharomyces cerevisiae Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *The EMBO Journal*, 15(18), pp.5093–5103. Available at:

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC452249/.
Brunet, E. et al., 2009. Chromosomal translocations induced at specified loci in human stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 106(26), pp.10620–10625. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2700748/.
Bryant, H.E. et al., 2005. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature*, 434, p.913. Available at: http://dx.doi.org/10.1038/nature03443.
Bunting, S.F. et al., 2010. 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell*, 141(2), pp.243–254. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2857570/.
Bunting, S.F. & Nussenzweig, A., 2013. End-joining, translocations and cancer. *Nature Reviews Cancer*, 13, p.443. Available at: http://dx.doi.org/10.1038/nrc3537.
Burman, B. et al., 2015. Histone modifications predispose genome regions to breakage and translocation. *Genes & Development*, 29(13), pp.1393–1402. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4511214/.
Burrow, A.A. et al., 2009. Over half of breakpoints in gene pairs involved in cancer-specific recurrent translocations are mapped to human chromosomal fragile sites. *BMC Genomics*, 10, p.59. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2642838/.
Bzymek, M. & Lovett, S.T., 2001. Instability of repetitive DNA sequences: The role of replication in multiple mechanisms. *Proceedings of the National Academy of Sciences*, 98(15), p.8319 LP-8325. Available at: http://www.pnas.org/content/98/15/8319.abstract.
Carreira, A. & Kowalczykowski, S.C., 2011. Two classes of BRC repeats in BRCA2 promote RAD51 nucleoprotein filament function by distinct mechanisms. *Proceedings of the National Academy of Sciences of the United States of America*, 108(26), pp.10448–10453. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3127913/.
Casellas, R. et al., 2016. Mutations, kataegis and translocations in B cells: understanding AID promiscuous activity. *Nature Reviews Immunology*, 16, p.164. Available at: http://dx.doi.org/10.1038/nri.2016.2.
Ceccaldi, R. et al., 2015. Homologous recombination-deficient tumors are hyper-dependent on POLQ-mediated repair. *Nature*, 518(7538), pp.258–262. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4415602/.
Ceccaldi, R., Rondinelli, B. & D'Andrea, A.D., 2016. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends in cell biology*, 26(1), pp.52–64. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4862604/.
Chan, S.H., Yu, A.M. & McVey, M., 2010. Dual Roles for DNA Polymerase Theta in Alternative End-Joining Repair of Double-Strand Breaks in Drosophila. *PLOS Genetics*, 6(7), p.e1001005. Available at: https://doi.org/10.1371/journal.pgen.1001005.
Chiarle, R. et al., 2011. Genome-Wide Translocation Sequencing Reveals Mechanisms of Chromosome Breaks and Rearrangements in B Cells. *Cell*, 147(1), pp.107–119. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3186939/.
Cho, S.W. et al., 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Research*, 24(1), pp.132–141. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3875854/.
Cong, L. et al., 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (New York, N.Y.)*, 339(6121), pp.819–823. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3795411/.
Cremona, C.A. et al., 2012. Extensive DNA Damage-Induced Sumoylation Contributes to Replication and Repair and Acts in Addition to the Mec1 Checkpoint. *Molecular Cell*, 45(3), pp.422–432. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3340930/.
Davies, H. et al., 2017. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nature Medicine*, 23, p.517. Available at: http://dx.doi.org/10.1038/nm.4292.
Decottignies, A., 2013. Alternative end-joining mechanisms: a historical perspective. *Frontiers in Genetics*, 4, p.48. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3613618/.
Deng, S.K. et al., 2014. RPA Antagonizes Microhomology-Mediated Repair of DNA Double-Strand Breaks. *Nature structural & molecular biology*, 21(4), pp.405–412. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3980576/.

Deriano, L. et al., 2009. Roles for NBS1 in alternative nonhomologous end joining of V(D)J recombination intermediates. *Molecular cell*, 34(1), pp.13–25. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2704125/.

Deriano, L. & Roth, D.B., 2013. Modernizing the Nonhomologous End-Joining Repertoire: Alternative and Classical NHEJ Share the Stage. *Annual Review of Genetics*, 47(1), pp.433–455. Available at: https://doi.org/10.1146/annurev-genet-110711-155540.

Difilippantonio, M.J. et al., 2002. Evidence for Replicative Repair of DNA Double-Strand Breaks Leading to Oncogenic Translocation and Gene Amplification. *The Journal of Experimental Medicine*, 196(4), pp.469–480. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2196056/.

Dimitrova, E., Turberfield, A.H. & Klose, R.J., 2015. Histone demethylases in chromatin biology and beyond. *EMBO Reports*, 16(12), pp.1620–1639. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4687429/.

Ding, D. et al., 2018. Engineering Introns to Express RNA Guides for Cas9- and Cpf1-Mediated Multiplex Genome Editing. *Molecular Plant*, 11(4), pp.542–552. Available at: https://doi.org/10.1016/j.molp.2018.02.005.

Dutra, B.E., Sutera, V.A. & Lovett, S.T., 2007. RecA-independent recombination is efficient but limited by exonucleases. *Proceedings of the National Academy of Sciences*, 104(1), p.216 LP-221. Available at: http://www.pnas.org/content/104/1/216.abstract.

Edwards, S.L. et al., 2008. Resistance to therapy caused by intragenic deletion in BRCA2. *Nature*, 451, p.1111. Available at: http://dx.doi.org/10.1038/nature06548.

Farmer, H. et al., 2005. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*, 434, p.917. Available at: http://dx.doi.org/10.1038/nature03445.

Ferguson, D.O. et al., 2000. The nonhomologous end-joining pathway of DNA repair is required for genomic stability and the suppression of translocations. *Proceedings of the National Academy of Sciences of the United States of America*, 97(12), pp.6630–6633. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC18682/.

Fong, P.C. et al., 2009. Inhibition of Poly(ADP-Ribose) Polymerase in Tumors from BRCA Mutation Carriers. *New England Journal of Medicine*, 361(2), pp.123–134. Available at: https://doi.org/10.1056/NEJMoa0900212.

Frank, K.M. et al., 2000. DNA Ligase IV Deficiency in Mice Leads to Defective Neurogenesis and Embryonic Lethality via the p53 Pathway. *Molecular Cell*, 5(6), pp.993–1002. Available at: https://doi.org/10.1016/S1097-2765(00)80264-6.

Frit, P. et al., 2014. Alternative end-joining pathway(s): Bricolage at DNA breaks. *DNA Repair*, 17(Supplement C), pp.81–97. Available at: http://www.sciencedirect.com/science/article/pii/S1568786414000445.

Gaj, T., Gersbach, C.A. & Barbas, C.F., 2013. ZFN, TALEN and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology*, 31(7), pp.397–405. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3694601/.

Gao, Y. et al., 2000. Interplay of p53 and DNA-repair protein XRCC4 in tumorigenesis, genomic stability and development. *Nature*, 404, p.897. Available at: http://dx.doi.org/10.1038/35009138.

Gavande, N.S. et al., 2016. DNA repair targeted therapy: The past or future of cancer treatment? *Pharmacology & Therapeutics*, 160, pp.65–83. Available at: http://www.sciencedirect.com/science/article/pii/S0163725816000322.

Gearing, M., 2015. Plasmids 101: Golden Gate Cloning. Available at: https://blog.addgene.org/plasmids-101-golden-gate-cloning [Accessed September 9, 2018].

van Gent, D.C. & Kanaar, R., 2016. Exploiting DNA repair defects for novel cancer therapies D. Kellogg, ed. *Molecular Biology of the Cell*, 27(14), pp.2145–2148. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4945134/.

Ghezraoui, H. et al., 2014. Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. *Molecular cell*, 55(6), pp.829–842. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4398060/.

Giry-Laterrière, M., Verhoeyen, E. & Salmon, P., 2011. Lentiviral Vectors BT - Viral Vectors for Gene Therapy: Methods and Protocols. In O.-W. Merten & M. Al-Rubeai, eds. Totowa, NJ: Humana

Press, pp. 183–209. Available at: https://doi.org/10.1007/978-1-61779-095-9_8.
Gou, D. et al., 2007. A novel approach for the construction of multiple shRNA expression vectors. *The Journal of Gene Medicine*, 9(9), pp.751–763. Available at: https://doi.org/10.1002/jgm.1080.
Hakim, O. et al., 2012. DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. *Nature*, 484(7392), pp.69–74. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3459314/.
Higgins, G.S., Prevo, R., et al., 2010. A siRNA Screen of Genes Involved in DNA Repair Identifies Tumour Specific Radiosensitisation by POLQ Knockdown. *Cancer research*, 70(7), pp.2984–2993. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2848966/.
Higgins, G.S., Harris, A.L., et al., 2010. Overexpression of POLQ Confers a Poor Prognosis in Early Breast Cancer Patients. *Oncotarget*, 1(3), pp.175–184. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2917771/.
Howard, S.M., Yanez, D.A. & Stark, J.M., 2015. DNA Damage Response Factors from Diverse Pathways, Including DNA Crosslink Repair, Mediate Alternative End Joining N. Maizels, ed. *PLoS Genetics*, 11(1), p.e1004943. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4309583/.
Hsu, P.D. et al., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*, 31(9), pp.827–832. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3969858/.
Hsu, P.D., Lander, E.S. & Zhang, F., 2014. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*, 157(6), pp.1262–1278. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4343198/.
Huertas, P., 2010. DNA resection in eukaryotes: deciding how to fix the break. *Nature structural & molecular biology*, 17(1), pp.11–16. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2850169/.
Illumina Inc., 2017. An introduction to Next-Generation Sequencing Technology. Available at: https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf [Accessed September 9, 2018].
Jasin, M. & Rothstein, R., 2013. Repair of Strand Breaks by Homologous Recombination. *Cold Spring Harbor Perspectives in Biology*, 5(11), p.a012740. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3809576/.
Jinek, M. et al., 2012. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, 337(6096), p.816 LP-821. Available at: http://science.sciencemag.org/content/337/6096/816.abstract.
Johns Hopkins University, 2018. crispr.technology/resources/quantification.html. Available at: http://crispr.technology/resources/quantification.html [Accessed September 3, 2018].
Kabadi, A.M. et al., 2014. Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Research*, 42(19), pp.e147–e147. Available at: http://dx.doi.org/10.1093/nar/gku749.
Kabotyanski, E.B. et al., 1998. Double-strand break repair in Ku86- and XRCC4-deficient cells. *Nucleic Acids Research*, 26(23), pp.5333–5342. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC147996/.
Karanam, K. et al., 2012. Quantitative live cell imaging reveals a gradual shift between DNA repair mechanisms and a maximal use of HR in mid-S phase. *Molecular cell*, 47(2), pp.320–329. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3494418/.
Kim, H. et al., 2011. Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nature Methods*, 8, p.941. Available at: http://dx.doi.org/10.1038/nmeth.1733.
Kirchmaier, S., Lust, K. & Wittbrodt, J., 2013. Golden GATEway Cloning – A Combinatorial Approach to Generate Fusion and Recombination Constructs. *PLOS ONE*, 8(10), p.e76117. Available at: https://doi.org/10.1371/journal.pone.0076117.
Klein, I.A. et al., 2011. Translocation-Capture Sequencing Reveals the Extent and Nature of Chromosomal Rearrangements in B Lymphocytes. *Cell*, 147(1), pp.95–106. Available at: https://doi.org/10.1016/j.cell.2011.07.048.
Konstantinopoulos, P.A. et al., 2015. Homologous recombination deficiency: Exploiting the fundamental vulnerability of ovarian cancer. *Cancer discovery*, 5(11), pp.1137–1154. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4631624/.

Kostyrko, K. & Mermod, N., 2016. Assays for DNA double-strand break repair by microhomology-based end-joining repair mechanisms. *Nucleic Acids Research*, 44(6), pp.e56–e56. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4824085/.

Kuhar, R. et al., 2014. Novel fluorescent genome editing reporters for monitoring DNA repair pathway utilization at endonuclease-induced breaks. *Nucleic Acids Research*, 42(1), pp.e4–e4. Available at: http://dx.doi.org/10.1093/nar/gkt872.

Kuo, L.J. & Yang, L.-X., 2008. γ-H2AX - A Novel Biomarker for DNA Double-strand Breaks. *In Vivo*, 22(3), pp.305–309. Available at: http://iv.iiarjournals.org/content/22/3/305.abstract.

Küppers, R. & Dalla-Favera, R., 2001. Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene*, 20, p.5580. Available at: http://dx.doi.org/10.1038/sj.onc.1204640.

Landry, J.J.M. et al., 2013. The Genomic and Transcriptomic Landscape of a HeLa Cell Line. *G3: Genes|Genomes|Genetics*, 3(8), pp.1213–1224. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3737162/.

Lee-Theilen, M. et al., 2011. CtIP promotes microhomology-mediated alternative end-joining during class switch recombination. *Nature structural & molecular biology*, 18(1), pp.75–79. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3471154/.

Lieber, M.R., 2010. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End Joining Pathway. *Annual review of biochemistry*, 79, pp.181–211. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3079308/.

Lord, C.J. & Ashworth, A., 2016. BRCAness revisited. *Nature Reviews Cancer*, 16, p.110. Available at: http://dx.doi.org/10.1038/nrc.2015.21.

Lu, G. et al., 2016. Ligase I and ligase III mediate the DNA double-strand break ligation in alternative end-joining. *Proceedings of the National Academy of Sciences of the United States of America*, 113(5), pp.1256–1260. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4747774/.

Ma, J.-L. et al., 2003. 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), pp.8820–8828. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC262689/.

Mansour, W.Y., Rhein, T. & Dahm-Daphi, J., 2010. The alternative end-joining pathway for repair of DNA double-strand breaks requires PARP1 but is not dependent upon microhomologies. *Nucleic Acids Research*, 38(18), pp.6065–6077. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2952854/.

Mateo, J. et al., 2015. DNA-Repair Defects and Olaparib in Metastatic Prostate Cancer. *New England Journal of Medicine*, 373(18), pp.1697–1708. Available at: https://doi.org/10.1056/NEJMoa1506859.

Mateos-Gomez, P.A. et al., 2015. Mammalian Polymerase Theta Promotes Alternative-NHEJ and Suppresses Recombination. *Nature*, 518(7538), pp.254–257. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4718306/.

McVey, M. & Lee, S.E., 2008. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends in genetics : TIG*, 24(11), pp.529–538. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5303623/.

Mehta, A. & Haber, J.E., 2014. Sources of DNA Double-Strand Breaks and Models of Recombinational DNA Repair. *Cold Spring Harbor Perspectives in Biology*, 6(9), p.a016428. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4142968/.

Menon, V. & Povirk, L., 2014. Involvement of p53 in the repair of DNA double strand breaks: Multifaceted roles of p53 in homologous recombination repair (HRR) and non-homologous end joining (NHEJ). *Sub-cellular biochemistry*, 85, pp.321–336. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4235614/.

Mitelman, F., Johansson, B. & Mertens, F., 2007. The impact of translocations and gene fusions on cancer causation. *Nature Reviews Cancer*, 7, p.233. Available at: http://dx.doi.org/10.1038/nrc2091.

Muvarak, N. et al., 2015. c-MYC Generates Repair Errors via Increased Transcription of Alternative-NHEJ Factors, LIG3 and PARP1, in Tyrosine Kinase-activated Leukemias. *Molecular cancer research : MCR*, 13(4), pp.699–712. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4398615/.

Nemudryi, A.A. et al., 2014. TALEN and CRISPR/Cas Genome Editing Systems: Tools of Discovery.

Acta Naturae, 6(3), pp.19–40. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4207558/.

Network, T.C.G.A.R., 2011. Integrated Genomic Analyses of Ovarian Carcinoma. *Nature*, 474(7353), pp.609–615. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3163504/.

Nickoloff, J.A. et al., 2017. Drugging the Cancers Addicted to DNA Repair. *JNCI: Journal of the National Cancer Institute*, 109(11), pp.djx059-djx059. Available at: http://dx.doi.org/10.1093/jnci/djx059.

O'Connor, M.J., 2015. Targeting the DNA Damage Response in Cancer. *Molecular Cell*, 60(4), pp.547–560. Available at: https://doi.org/10.1016/j.molcel.2015.10.040.

Oh, S. et al., 2014. DNA ligase III and DNA ligase IV carry out genetically distinct forms of end joining in human somatic cells. *DNA repair*, 21, pp.97–110. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4125535/.

Ottaviani, D., LeCain, M. & Sheer, D., 2014. The role of microhomology in genomic structural variation. *Trends in Genetics*, 30(3), pp.85–94. Available at: http://www.sciencedirect.com/science/article/pii/S016895251400002X.

Ozeri-Galai, E., Bester, A.C. & Kerem, B., 2012. The complex basis underlying common fragile site instability in cancer. *Trends in Genetics*, 28(6), pp.295–302. Available at: http://www.sciencedirect.com/science/article/pii/S0168952512000303.

Ozgenc, A.I., Szekeres, E.S. & Lawrence, C.W., 2005. In Vivo Evidence for a recA-Independent Recombination Process in Escherichia coli That Permits Completion of Replication of DNA Containing UV Damage in Both Strands. *Journal of Bacteriology*, 187(6), pp.1974–1984. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1064058/.

Padilla-Nash, H.M. et al., 2001. Jumping translocations are common in solid tumor cell lines and result in recurrent fusions of whole chromosome arms. *Genes, Chromosomes and Cancer*, 30(4), pp.349–363. Available at: https://doi.org/10.1002/gcc.1101.

Pastwa, E. et al., 2009. In vitro non-homologous DNA end joining assays—The 20th anniversary. *The International Journal of Biochemistry & Cell Biology*, 41(6), pp.1254–1260. Available at: http://www.sciencedirect.com/science/article/pii/S1357272508004810.

Paul, K. et al., 2013. DNA Ligases I and III Cooperate in Alternative Non-Homologous End-Joining in Vertebrates A. J. Lustig, ed. *PLoS ONE*, 8(3), p.e59505. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3610672/.

Price, B.D. & D'Andrea, A.D., 2013. Chromatin Remodeling at DNA Double-Strand Breaks. *Cell*, 152(6), pp.1344–54.

Qiu, P. et al., 2004. Mutation detection using Surveyor™ nuclease. *BioTechniques*, 36(4), pp.702–707. Available at: https://doi.org/10.2144/04364PF01.

Quail, M.A. et al., 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, 13(1), p.341. Available at: https://doi.org/10.1186/1471-2164-13-341.

Ran, F.A. et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, 8(11), pp.2281–2308. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3969860/.

Robert, T. et al., 2011. HDACs link the DNA damage response, processing of double-strand breaks and autophagy. *Nature*, 471(7336), pp.74–79. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3935290/.

Roix, J.J. et al., 2003. Spatial proximity of translocation-prone gene loci in human lymphomas. *Nature Genetics*, 34, p.287. Available at: http://dx.doi.org/10.1038/ng1177.

Roukos, V., Burman, B. & Misteli, T., 2013. The cellular etiology of chromosome translocations. *Current opinion in cell biology*, 25(3), pp.357–364. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3688675/.

Sakai, W. et al., 2008. Secondary mutations as a mechanism of cisplatin resistance in BRCA2-mutated cancers. *Nature*, 451(7182), pp.1116–1120. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2577037/.

Sallmyr, A., Tomkinson, A.E. & Rassool, F. V, 2008. Up-regulation of WRN and DNA ligase IIIα in chronic myeloid leukemia: consequences for the repair of DNA double-strand breaks. *Blood*, 112(4), pp.1413–1423. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2967309/.

Schimmel, J. et al., 2017. Mutational signatures of non-homologous and polymerase theta-mediated

end-joining in embryonic stem cells. *The EMBO Journal*, 36(24), pp.3634–3649. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5730883/.

Scully, R. & Xie, A., 2013. Double strand break repair functions of histone H2AX. *Mutation research*, 750(0), p.10.1016/j.mrfmmm.2013.07.007. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3818383/.

Seluanov, A., Mao, Z. & Gorbunova, V., 2010. Analysis of DNA Double-strand Break (DSB) Repair in Mammalian Cells. *Journal of Visualized Experiments : JoVE*, (43), p.2002. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3157866/.

Sfeir, A. & Symington, L.S., 2015. Microhomology-mediated end joining: a back-up survival mechanism or dedicated pathway? *Trends in biochemical sciences*, 40(11), pp.701–714. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4638128/.

Sharma, S. et al., 2015. Homology and enzymatic requirements of microhomology-dependent alternative end joining. *Cell Death & Disease*, 6(3), p.e1697. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4385936/.

Shimazaki, N., Tsai, A.G. & Lieber, M.R., 2009. H3K4me3 Stimulates the V(D)J RAG Complex for Both Nicking and Hairpinning in Trans in Addition to Tethering in Cis: Implications for Translocations. *Molecular cell*, 34(5), pp.535–544. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2920491/.

Simsek, D. et al., 2011. DNA Ligase III Promotes Alternative Nonhomologous End-Joining during Chromosomal Translocation Formation J. E. Haber, ed. *PLoS Genetics*, 7(6), p.e1002080. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3107202/.

Simsek, D. & Jasin, M., 2010. Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4/ligase IV during chromosomal translocation formation. *Nature structural & molecular biology*, 17(4), pp.410–416. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3893185/.

Soni, A. et al., 2015. Marked contribution of alternative end-joining to chromosome-translocation-formation by stochastically induced DNA double-strand-breaks in G2-phase human cells. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 793, pp.2–8. Available at: http://www.sciencedirect.com/science/article/pii/S1383571815001801.

Soutoglou, E. & Misteli, T., 2008. On the Contribution of Spatial Genome Organization to Cancerous Chromosome Translocations. *JNCI Monographs*, 2008(39), pp.16–19. Available at: http://dx.doi.org/10.1093/jncimonographs/lgn017.

Stover, E.H. et al., 2016. Biomarkers of Response and Resistance to DNA Repair Targeted Therapies. *Clinical Cancer Research*, 22(23), p.5651 LP-5660. Available at: http://clincancerres.aacrjournals.org/content/22/23/5651.abstract.

Symington, L.S. & Gautier, J., 2011. Double-Strand Break End Resection and Repair Pathway Choice. *Annual Review of Genetics*, 45(1), pp.247–271. Available at: https://doi.org/10.1146/annurev-genet-110410-132435.

ThermoFisher Scientific, PCR Methods - Top Ten Strategies. Available at: https://www.thermofisher.com/uk/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/pcr-education/pcr-reagents-enzymes/pcr-methods.html [Accessed September 9, 2018].

Tobin, L.A. et al., 2013. Targeting abnormal DNA double strand break repair in tyrosine kinase inhibitor-resistant chronic myeloid leukemias. *Oncogene*, 32(14), pp.1784–1793. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3752989/.

Tobin, L.A. et al., 2012. Targeting abnormal DNA repair in therapy-resistant breast cancers. *Molecular Cancer Research*, 10(1), pp.96–107. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3319138/.

Tomimatsu, N. et al., 2014. Phosphorylation of EXO1 by CDKs 1 and 2 regulates DNA end resection and repair pathway choice. *Nature communications*, 5, p.3561. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4041212/.

Torres, R. et al., 2014. Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR–Cas9 system. *Nature Communications*, 5, p.3964. Available at: http://dx.doi.org/10.1038/ncomms4964.

Truong, L.N. et al., 2013. Microhomology-mediated End Joining and Homologous Recombination

share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, 110(19), pp.7720–7725. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3651503/.

Tsai, A.G. et al., 2008. Human Chromosomal Translocations at CpG Sites and a Theoretical Basis for their Lineage and Stage Specificity. *Cell*, 135(6), pp.1130–1142. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2642632/.

Vanderstichele, A. et al., 2017. Genomic signatures as predictive biomarkers of homologous recombination deficiency in ovarian cancer. *European Journal of Cancer*, 86, pp.5–14. Available at: https://doi.org/10.1016/j.ejca.2017.08.029.

Venkitaraman, A.R., 2014. Cancer Suppression by the Chromosome Custodians, BRCA1 and BRCA2. *Science*, 343, pp.1470–5.

Verkaik, N.S. et al., 2002. Different types of V(D)J recombination and end-joining defects in DNA double-strand break repair mutant mammalian cells. *European Journal of Immunology*, 32(3), pp.701–709. Available at: https://doi.org/10.1002/1521-4141(200203)32:3%3C701::AID-IMMU701%3E3.0.CO.

Vouillot, L., Thélie, A. & Pollet, N., 2015. Comparison of T7E1 and Surveyor Mismatch Cleavage Assays to Detect Mutations Triggered by Engineered Nucleases. *G3: Genes|Genomes|Genetics*, 5(3), pp.407–415. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4349094/.

Wang, H. et al., 2003. Biochemical evidence for Ku-independent backup pathways of NHEJ. *Nucleic Acids Research*, 31(18), pp.5377–5388. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC203313/.

Wang, H. et al., 2012. CtIP Protein Dimerization Is Critical for Its Recruitment to Chromosomal DNA Double-stranded Breaks. *The Journal of Biological Chemistry*, 287(25), pp.21471–21480. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3375568/.

Wang, H. et al., 2005. DNA Ligase III as a Candidate Component of Backup Pathways of Nonhomologous End Joining. *Cancer Research*, 65(10), p.4020 LP-4030. Available at: http://cancerres.aacrjournals.org/content/65/10/4020.abstract.

Wang, X. & McManus, M., 2009. Lentivirus Production. *Journal of Visualized Experiments : JoVE*, (32), p.1499. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2865973/.

Weinberg, R.A., 2014. *The Biology of Cancer* 2nd ed., New York: Garland Science, Taylor & Francis Group.

Wilczynski, S.P., 2009. Chapter 7 - Molecular Biology. In N. Weidner et al., eds. *Modern Surgical Pathology*. Philadelphia: W.B. Saunders, pp. 85–120. Available at: http://www.sciencedirect.com/science/article/pii/B9781416039662000060.

Wood, R.D. & Doublié, S., 2016. DNA polymerase θ (POLQ), double-strand break repair, and cancer. *DNA repair*, 44, pp.22–32. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5114520/.

Xie, A., Kwok, A. & Scully, R., 2009. Role of mammalian Mre11 in classical and alternative non-homologous end joining. *Nature structural & molecular biology*, 16(8), pp.814–818. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2730592/.

Xie, K., Minkenberg, B. & Yang, Y., 2015. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proceedings of the National Academy of Sciences*, 112(11), p.3570 LP-3575. Available at: http://www.pnas.org/content/112/11/3570.abstract.

Yousefzadeh, M.J. & Wood, R.D., 2013. DNA polymerase POLQ and cellular defense against DNA damage. *DNA repair*, 12(1), pp.1–9. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3534860/.

Yun, M.H. & Hiom, K., 2009. CtIP-BRCA1 modulates the choice of DNA double-strand break repair pathway throughout the cell cycle. *Nature*, 459(7245), pp.460–463. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2857324/.

Zhang, X.-H. et al., 2015. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Molecular Therapy - Nucleic Acids*, 4, p.e264. Available at: http://www.sciencedirect.com/science/article/pii/S216225311630049X.

Zhang, Y. & Jasin, M., 2011. An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. *Nature structural & molecular biology*, 18(1), pp.80–84. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3261752/.

Zhu, C. et al., 2002. Unrepaired DNA Breaks in p53-Deficient Cells Lead to Oncogenic Gene Amplification Subsequent to Translocations. *Cell*, 109(7), pp.811–821. Available at: https://doi.org/10.1016/S0092-8674(02)00770-5.

Zischewski, J., Fischer, R. & Bortesi, L., 2017. Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnology Advances*, 35(1), pp.95–104. Available at: http://www.sciencedirect.com/science/article/pii/S0734975016301586.