



MEDICAL UNIVERSITY
OF VIENNA

Fundamental Insights of DNA Repair for CRISPR-Cas9-Mediated Genome Editing

Doctoral thesis at the Medical University of Vienna
for obtaining the academic degree

Doctor of Philosophy

Submitted by

MSc, Joana Ferreira da Silva

Supervisor:
Joanna Loizou, PhD

CeMM Research Center for Molecular Medicine of the Austrian Academy of
Sciences
Lazarettgasse 14, 1090 Vienna, Austria

Institute of Cancer Research, Department of Medicine I, Comprehensive Cancer
Centre, Medical University of Vienna
Borschkegasse 8a, 1090 Vienna, Austria

Vienna, November 2021

Acknowledgements

This thesis marks the end of a chapter of my life. A challenging yet rewarding journey that I will forever remember fondly. This work would not have been possible without the continuous support of multiple people who, through bigger or smaller gestures and often without even realising, kept me going.

First and foremost, I would like to sincerely thank my Mentor, Joanna Loizou. Your continuous support and mentorship allowed me to be here today. Thank you for always believing in me and for showing me what a strong, passionate and resilient scientist looks like. You have truly been a role model and I am proud to carry your name with me in future career endeavors. Thank you for everything.

The Loizou Lab has been instrumental for the completion of this thesis. It has always amazed me how many creative, kind and smart people can fit into the same room. Thank you to all past and present members of the lab, with whom I have shared benches, lab meetings, retreats, discussions, Halloween parties and cake. You are all amazing and I cannot wait to see what the future holds for all of you.

I would like to thank the CeMM Faculty for useful input and guidance throughout my entire PhD. A very special thank you to Georg Winter and Joerg Menche, who not only supported me through thesis committee meetings, but much further than that. Thank you to everyone that makes CeMM special. Not only scientific staff, but also admin team, back office support, cleaning staff. The work you do in the background and the willingness that you put on always taking the extra mile with us is what truly makes CeMM different. This is something I have learned not to take from granted in the 'real' world.

Going through this PhD would have been very difficult without the support of multiple friends. Thank you to my home friends and the friends that made Vienna feel like home. Thank you all for being present in the best and the worst moments. A very special thank you to my extended Zozo, who very rapidly became my Viennese family. Thank you, Tala, for your wisdom and warmth. Thank you, Matthias, for always being my person and my rock.

Chris, thank you for your love and support throughout these 4 years. Your ability to make me laugh, even when I'm the grumpiest, still amazes me to this day. I couldn't imagine a better adventure partner and I cannot wait to see what the future holds for us.

Finally, a very special thanks to my Mum and Dad, Grandparents and Brother. You have always allowed me to see the world full of infinite possibilities, but you never let me forget where I come from. And that unconditional love I carry with me, everyday. This thesis is for you.

Declaration

The work of this thesis was developed at several academic institutions and with assistance from different collaborators. In the following, all contributions are listed in detail.

Chapter 3.1 has been published by Ferreira da Silva et al., 2019. The author of this thesis has designed and performed most of the experiments, analysed, interpreted the data, made the figures and wrote the manuscript. Detailed information on the contribution of other authors is listed on the last page of the manuscript. All the work has been carried out in the laboratory of Dr Joanna Loizou, at the CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences.

The data presented in chapter 3.2 has been published in a pre-print server (<https://www.biorxiv.org/content/10.1101/2021.09.30.462548v1.full.pdf>) and is intended for publication in 2021-2022. All experimental work, analysis and interpretation were conducted by the author, together with Dr Gonçalo Oliveira, and with assistance of Dr Amandine Moretton and MSc Emili Arasa. The author wrote the manuscript with input from Dr Joanna Loizou, who conceived the project together with Prof. Josef Jiricny (Institute of Biochemistry of the ETH Zurich, Switzerland). All the work has been carried out in the laboratory of Dr. Joanna Loizou at the Institute of Cancer Research, Medical University of Vienna.

All chapters of this thesis were written by the author. Dr Joanna Loizou provided input to the writing of the thesis.

Reprint permission for chapter 3.1 was obtained according to the reprint and permission policies of the Nature Publishing Group.

Table of Contents

List of Figures	vi
List of Tables	vi
Zusammenfassung	viii
List of Abbreviations	ix
Chapter I – Introduction	1
1. DNA Damage Repair & Genome Stability	1
1.1 DNA Damage	1
1.2 The DNA Damage Response (DDR).....	1
1.3 DDR & Human Disease	11
2. Exploiting the DDR Knowledge to Treat Human Disease	14
2.1 Cancer Therapy	14
2.2 Gene Therapy	16
3. Genome Editing: Bridging DDR & Human Disease	16
3.1 A Historic Perspective of Genome Editing.....	16
3.2 CRISPR-Cas systems	17
3.3 Functionality of the CRISPR-Cas9 System	19
3.4 Cas Diversity	20
3.5 The PAM Motif: Cas9 Search Mechanism and Target Binding	21
3.6 Applications of CRISPR-Cas9-Mediated Genome Engineering	22
3.7 Emerging CRISPR-Based Technologies for Genome Engineering	22
4. Genome Editing to Treat Human Disease	30
4.1 The Promise: Therapeutic Scope for CRISPR Editing	30
4.2 The Challenge: Obstacles to Overcome	32
5. DNA Damage Response: The Foundation of Genome Editing	35
5.1 Template-Free Repair of Cas9-Mediated DSBs	35
5.2 Homology-Directed Repair (HDR).....	36
Chapter 2: Aim of this study	40
Chapter 3: Results	41
3.1 Genome-Scale CRISPR Screens are Efficient in Non-Homologous End-Joining Deficient Cells	41
3.1.1 Prologue	41
3.1.2 PDF of the article.....	42
3.2 Prime Editing Efficiency and Fidelity are Enhanced in the Absence of Mismatch Repair	58
3.2.1 Prologue	58
3.2.1 PDF of the Article.....	59
Chapter 4: Discussion	79
4.1 Contribution of this thesis to the field of genome editing.....	79
4.2 Alternative pathways for the mutagenic repair of Cas9-induced breaks (Project 1)	80
4.3 Exploiting MMEJ in the context of CRISPR-mediated genome editing (Project 1)	81
4.4 Future directions for the study of Cas9-induced mutagenic repair (Project 1) ...	82

4.4 The inhibitory role of MMR for prime editing (Project 2)	83
4.5 Exploiting the inhibitory role of MMR for improved prime editing (Project 2)	88
4.6 Other DNA repair requirements for prime editing (Project 2)	89
4.7 The importance of using adequate cellular models	90
4.8 Future directions	92
<i>References</i>	94
<i>Curriculum Vitae</i>	114
<i>List of Publications</i>	116

List of Figures

Figure 1: Direct Repair and Mismatch Repair Pathways- page 3

Figure 2: Base Excision Repair (BER) and Nucleotide Excision Repair (NER) pathways – page 6

Figure 3: Double-Strand Break (DSB) Repair Pathway Choice – page 10

Figure 4: Illustration of the concept of synthetic lethality – page 16

Figure 5: Schematics of the CRISPR-Cas (type II) system – page 20

Figure 6: Transcriptional control via CRISPR interference (CRISPRi) or CRISPR activation (CRISPRa) – page 24

Figure 7: Base editing strategies – page 26

Figure 8: Schematics of prime editing – page 29

Figure 9: Schematics of PE3 and PE3b - page 30

Figure 10: Repair outcomes after a Cas9-induced DNA double-strand break (DSB) and methods for enhancing precision repair – page 38

Figure 11: MMR counteracts prime editing efficiency in PE2 – page 88

Figure 12: MMR involvement in the PE3 strategy – page 89

Figure 13: Validation of prime editing inhibition by MMR activity, in the human RPE cell line - page 95

List of Tables

Table 1: Summary of DDR mechanism and respective type of damage, prime lesions, key factors and associated diseases – page 14

Abstract

The human genome is constantly exposed to endogenous and exogenous sources of DNA damage that pose a threat to genomic stability. Hence, cells have evolved an intricate network of mechanisms specialised in the repair of several lesions, called the DNA damage response.

The discovery of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and its associated endonuclease Cas9 has revolutionised genome editing and opened new therapeutic opportunities. CRISPR-Cas9 mediated genome editing relies on the generation of a targeted DNA double-strand break (DSB) in a precise region of the genome, directed by a single-guide RNA (sgRNA) and harnesses the cells endogenous DNA repair machinery for its repair. In the last decade, the rapidly evolving CRISPR field has expanded the genome editing toolbox with the aim of making this technology more efficient, precise and less reliant on potentially deleterious DSBs. Notwithstanding, most CRISPR-Cas9-based approaches heavily rely on the DNA damage response.

In this thesis, we have studied the DNA repair mechanisms that underlie CRISPR-based genome editing technologies in human cells. First we studied the mutagenic repair of Cas9-induced DSBs (Project 1), where we discovered that non-homologous end-joining, the pathway previously thought to be the repair mechanism dealing with such lesions, is entirely dispensable and can be fully compensated for by an alternative pathway that generates distinct DNA repair signatures. Next (Project 2), we studied the DNA repair requirements of a recent CRISPR technology, called prime editing. Prime editing is a versatile and precise technology that allows for the installation of any type of edit in the genome, without the generation of DSBs. We identified an inhibitory role of the DNA mismatch repair (MMR) pathway for prime editing, which occurs through MMR directly binding to sites of prime editing and excising the desired edit. Additionally, MMR abrogation was shown to improve prime editing efficiency and fidelity, across different endogenous loci and types of edits, in multiple cell lines.

Overall, these studies advance our knowledge on the DNA repair requirements of CRISPR-based genome editing approaches, providing new ways to further advance the technology. Studies like ours are also important for the implementation of CRISPR in the clinical setting, which ultimately relies on a safe navigation of the DNA repair decision process: from damage to desired editing outcome, with minimal side-effects.

Zusammenfassung

Das menschliche Genom ist ständig endogenen und exogenen Quellen von DNA-Schäden ausgesetzt, die eine Bedrohung für die Stabilität des Genoms darstellen. Daher haben die Zellen ein kompliziertes Netz von Mechanismen entwickelt, die auf die Reparatur verschiedener Läsionen spezialisiert sind und als DNA-Schadensantwort bezeichnet werden. Die Entdeckung von Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) und der dazugehörigen Endonuklease Cas9 hat die Genom-Editierung revolutioniert und neue therapeutische Möglichkeiten eröffnet. CRISPR-Cas9 basierte Genommanipulation basiert auf der Erzeugung eines DNA-Doppelstrangbruchs (DSB) in einer präzisen Region des Genoms, die durch eine Single-Guide-RNA (sgRNA) gesteuert wird. Die zelleigene DNA-Reparaturmaschinerie wird für die Behebung des Schadens genutzt. In den letzten zehn Jahren hat das sich rasch entwickelnde CRISPR-Feld den Werkzeugkasten für die Genom-Editierung mit dem Ziel erweitert, diese Technologie effizienter, präziser und weniger abhängig von potenziell schädlichen DSBs zu machen. Dennoch sind die meisten CRISPR-Cas9-basierten Ansätze stark auf die DNA-Schadensreaktion angewiesen.

In dieser Arbeit haben wir die DNA-Reparaturmechanismen untersucht, die den CRISPR-basierten Genom-Editierungstechnologien in menschlichen Zellen zugrunde liegen. Zunächst untersuchten wir die mutagene Reparatur von Cas9-induzierten DSBs (Projekt 1); dabei stellten wir fest, dass der bisherige postulierte Reparaturmechanismus solcher Läsionen (non-homologous end joining, NHEJ) überflüssig ist und durch einen alternativen Weg, der unterschiedliche DNA-Reparatursignaturen erzeugt, vollständig kompensiert werden kann. Als nächstes (Projekt 2) untersuchten wir die DNA-Reparaturanforderungen einer neuen CRISPR-Technologie, dem so genannten Prime Editing. Prime Editing ist eine vielseitige und präzise Technologie, mit der jede Art von Schnitt im Genom vorgenommen werden kann, ohne dass dabei DSBs entstehen. Wir haben eine hemmende Rolle des DNA-Mismatch-Reparatur-Wegs (MMR) für das Prime-Editing identifiziert, die dadurch entsteht, dass MMR direkt an die Stellen des Prime-Editings bindet und den gewünschten Schnitt herausschneidet. Darüber hinaus konnte gezeigt werden, dass die Ausschaltung der MMR in mehreren Zelllinien die Effizienz und Zuverlässigkeit des Prime Editing in verschiedenen endogenen Loci und Arten von Edits verbessert.

Insgesamt erweitern diese Studien unser Wissen über die Anforderungen von CRISPR-basierten Genome Editing-Ansätzen an die zellulären DNA-Reparaturmechanismen und bieten neue Möglichkeiten, die Technologie weiter zu verbessern. Studien wie unsere sind auch wichtig für die Umsetzung von CRISPR in der klinischen Praxis, die letztlich auf eine sichere Navigation durch den DNA-Reparatur-Entscheidungsbaum angewiesen ist: von der Läsion zum gewünschten Editing-Ergebnis, mit minimalen Nebenprodukten.

List of Abbreviations

γH2AX	Phosphorylated form of histone H2AX
53BP1	p53-Binding Protein 1
6-4PPs	6-4 pyrimidine-pyrimidone photoproducts
A	Adenosine
AAV	Adeno-Associated Virus
8-oxoG	Oxidation of guanine at position 8
ABE	Adenine Base Editor
Alt-EJ	Alternative End-Joining (same as MMEJ)
APEX1	Apurinic-Apyrimidinic Endonuclease 1
APOBEC	Apolipoprotein B-Editing Enzyme Catalytic Polypeptide
ATM	Ataxia-Telangiectasia Mutated
ATP	Adenosine Triphosphate
ATR	Ataxia-Telangiectasia and Rad3-Related
ATRIP	Ataxia-Telangiectasia and Rad3-Related-Interacting Protein
BE	Base Editor
BER	Base Excision Repair
bp	Base Pairs
BRCA1	Breast Cancer Gene 1
BRCA2	Breast Cancer Gene 2
C	Cytosine
CAR	Chimeric Antigen Receptor
Cas	CRISPR-Associated Protein
CBE	Cytidine Base Editor
CCR5	CC Chemokine Receptor type 5
CHK	Checkpoint Kinase
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRa	CRISPR activation
CRISPRi	CRISPR interference
CS	Cockayne Syndrome
CPDs	Cyclobutane-Pyrimidine Dimers
crRNA	CRISPR RNA
dCas9	Catalytically Inactive Cas9
DMD	Duchenne Muscular Dystrophy
DNA	Deoxyribonucleic Acid
DNA2	Dna2-Nuclease-Helicase
DNA-PK	DNA-Dependent Protein Kinase
DNA-PKcs	DNA-Dependent Protein Kinase Catalytic Subunit
DDB	DNA Damage-Binding Protein
DDR	DNA Damage Response
DR	Direct Repair
DSB	Double-Strand Break
DSBR	Double-Strand Break Repair
dsDNA	Double-stranded DNA
<i>E. coli</i>	<i>Escherichia coli</i>
EGFR	Epidermal Growth Factor Receptor
ERCC	Excision Repair Cross-Complementing
ERK	Extracellular Signal-Regulated Kinases
EXO1	Exonuclease 1

FA	Fanconi Anaemia
FANC	Fanconi Anaemia Complementation Group
FDA	(United States) Food and Drug Administration
FEN1	Flap Endonuclease 1
G	Guanosine
GG-NER	Global Genome Nucleotide Excision Repair
HDR	Homology-Directed Repair
HGPS	Hutchinson-Gilford Progeria Syndrome
HIV	Human Immunodeficiency Virus
HR	Homologous Recombination
I	Inosine
ICL	Interstrand-Crosslink
Indels	Insertions and Deletions
IR	Ionizing Radiation
kb	Kilobases
KO	(gene) Knock-Out
KRAB	Krüppel-associated box
LIG	DNA Ligase
LP-BER	Long Patch Base Excision Repair
MGMT	O ⁶ -Methylguanine DNA Methyltransferase
MLH	MutL Homolog
MMC	Mitomycin C
MMEJ	Microhomology-Mediated End-Joining (same as Alt-EJ)
MMR	Mismatch Repair
MRE11	Meiotic Recombination 11
MSH	MutS Homolog
MSI	Microsatellite Instability
MUTYH	MutY DNA Glycosylase
NBS1	Nijmegen breakage syndrome 1
nCas9	Nickase Cas9
NEIL1	Nei Like DNA Glycosylase 1
NER	Nucleotide Excision Repair
NHEJ	Non-Homologous End-Joining
OGG1	8-oxoG DNA Glycosylase 1
PACE	Phage- Assisted Continuous Evolution
PAM	Protospacer Adjacent Motif
PARP1	Poly(ADP-Ribose) Polymerase 1
PCNA	Proliferating Cell Nuclear Antigen
PD1	Programmed Cell Death Protein 1
PDX	Patient-Derived Xenograft
PE	Prime Editing
pegRNA	Prime Editing Guide-RNA
POLQ	DNA Polymerase θ
Poi	DNA Polymerase
PTIP	PAX Transcription Activation-Containing Protein
REPAIR	RNA Editing for Programmable A to I Replacement
RFC	Replication Factor C
RIF1	Replication Timing Regulatory Factor 1
RNP	Ribonucleoprotein
ROS	Reactive Oxygen Species

RPA	Replication Protein A
RT	Reverse Transcriptase
sgRNA	Single Guide RNA
SHERLOCK	Specific High-Sensitivity Enzymatic Reporter Unlocking
SHLD	SHLD – Shieldin Complex Subunit
SP	Selection Phage
SP-BER	Short-patch Base Excision Repair
SpCas9	Cas9 from <i>Streptococcus pyogenes</i>
SSA	Single-Strand Annealing
SSB	Single-Strand Break
SSBR	Single-Strand Break Repair
ssDNA	Single-Stranded DNA
ssODN	Single-Stranded Oligonucleotide
SSTR	Single-Stranded Templated Repair
T	Thymine
TALENs	Transcription Activator-Like Effector Nucleases
TC-NER	Transcription Coupled Nucleotide Excision Repair
TERT	Telomerase Reverse Transcriptase
TFIIH	Transcription Factor II H
TLS	Translesion Synthesis
TP53	Tumour Protein 53
tracrRNA	Trans-Activating crRNA
TTD	Trichothiodystrophy
U	Uracil
UDG	Uracil-DNA Glycosylase
UGI	Uracil Glycosylase Inhibitor
UV	Ultraviolet radiation
UVSS	UV-sensitive syndrome
XLF	X-Ray Cross-Complementing Protein 4- Like-Factor
XRCC	X-Ray Cross-Complementing Protein
XP	Xeroderma pigmentosum
ZFNs	Zinc Finger Nucleases

Chapter I – Introduction

1. DNA Damage Repair & Genome Stability

1.1 DNA Damage

The human genome is constantly exposed to endogenous and exogenous sources of damage that threaten genomic stability. The complex lesions that arise from this exposure to damage can lead to genome aberrations, mutations and ultimately cell death (Jackson & Bartek, 2009).

Examples of endogenous sources of damage are DNA mismatches introduced during DNA replication, alkylation of bases, or DNA double-strand breaks (DSBs) caused by abortive activity of topoisomerases. Reactive-oxygen species (ROS) are also important sources of endogenous DNA damage, arising as by-products of cellular metabolism (Cadet & Richard Wagner, 2013).

The most prevalent environmental DNA-damaging agent is the ultraviolet (UV) radiation. Despite absorbance of this radiation by the ozone layer, residual levels can still induce up to 10^5 DNA lesions per cell, per day (Hoeijmakers, 2009). DNA damaging chemicals (widely used for chemotherapeutics), ionizing radiation (IR) and tobacco products are other examples of exogenous sources of DNA damage. DNA damaging agents produce distinct types of damage, but most of them induce a predominant lesion. Understanding the correlation between damage and damaging agent is essential for the comprehension of the mechanisms by which our genome is guarded (**Table 1**).

1.2 The DNA Damage Response (DDR)

Cells have evolved an intricate network of mechanisms to detect, signal and repair DNA lesions. These mechanisms are collectively called the DNA damage response (DDR). Defective DDR renders cells exquisitely sensitive to DNA damaging agents and causes several human illnesses with distinct clinical features (**Table 1**).

The broad diversity of DNA lesions requires the intervention of widely diverse DNA repair mechanisms (**Table 1**) that are mostly conserved across organisms. Some types of damage can be directly repaired by the excision of the lesion. However, most of the lesions are subjected to a sequential process, catalysed by a complex set of factors which comprise sensors, transducers, mediators and effectors. Sensors detect and signal the DNA lesions, leading to the recruitment of other proteins to the site of damage. Some of these recruited factors are transducers, which amplify the DNA damage signal, whereas others are considered mediators, since they act on mediating the interactions between different proteins.

The last group of proteins, the effectors, can act in the direct repair of the DNA. An example of those would be polymerases, that directly fill the gap between two broken ends. Additionally, other effectors act on the regulation of essential cellular mechanisms that have an impact on DNA repair, such as cell cycle, apoptosis, gene expression and metabolism. This multifaceted interplay highlights the complexity of the DDR, as a unique and complex network that cooperates with widely different pathways within the cell.

1.2.1 Direct Repair (DR)

Direct Repair is the simplest mechanism of DNA repair, since it involves the direct reversal of the lesion without a major sequential catalytic cascade. DR often requires the action of a single, specialised, enzyme that excises the lesion, leading to error-free repair of the DNA. One notorious example is the repair of DNA methylation at the oxygen in position 6 of guanine (O⁶-methylguanine), which is induced by some chemotherapeutic agents, such as temozolomide or dacarbazine (Mitchell & Dolan, 1993). O⁶-methylguanine can be directly repaired by the protein O⁶-methylguanine DNA methyltransferase (MGMT) (Kaina *et al*, 2007) (**Figure 1A**). Unrepaired O⁶-methylguanine is mutagenic since, during DNA replication, it pairs with thymine instead of cytosine, generating G:C > A:T base substitutions. Hence, in MGMT deficient cells, the persistent O⁶-methylguanine alerts a second DNA repair pathway – the mismatch repair (MMR) pathway – which exclusively detects the mispaired thymine reinsertion, excising it. The exhaustive cycles of thymine reinsertion and excision result in extensive DNA end-resection and, ultimately, apoptosis. Temozolomide and similar chemotherapeutics are therefore most effective in cells with low levels of MGMT and intact MMR (Thomas *et al*, 2017).

1.2.2 Mismatch Repair (MMR)

The DNA Mismatch Repair pathway (MMR) recognises errors that have been usually introduced during replication, causing either the incorporation of a wrong nucleotide, or small insertions or deletions (indels) (**Figure 1B**). Therefore, MMR inactivation in human cells has been widely associated with increased mutation rates, genome instability and cancer onset (Jiricny, 2006). These cancers exhibit a hyper-mutable phenotype, called microsatellite instability (MSI)

In human cells, MutS complexes recognise the mismatches. The MutS α complex, an heterodimer formed by MSH2 (MutS homolog 2) and MSH6 (MutS homolog 6) recognises single-base mismatches. The MutS β complex (MSH2-MSH3) (MutS homolog 3) recognises small indels (<10 bp) (Li, 2008). These heterodimers scan the DNA for errors, playing a critical

role in the recognition and initiation of repair. The detection of a lesion is associated with an ATP-dependent conformational change of the complex, leading to the recruitment of MutL complexes, which are the downstream processors: the MLH1 (MutL homolog 1) - MLH3 (MutL homolog 3) heterodimer that also requires intervention of the endonuclease PMS2. PMS2, in the presence of Replication Factor C (RFC) and Proliferating Cell Nuclear Antigen (PCNA), nicks the DNA flanking the mismatch, generating entry points for the exonuclease EXO1 to degrade the strand containing the mismatch. Replication Protein A (RPA) coats the single-stranded DNA molecule and polymerases Pol- ϵ and Pol- δ then fill the missing nucleotides. After the re-synthesis of DNA, the remaining nick is sealed by DNA Ligase I (LIG1), resulting in an error-free repair of the lesion.

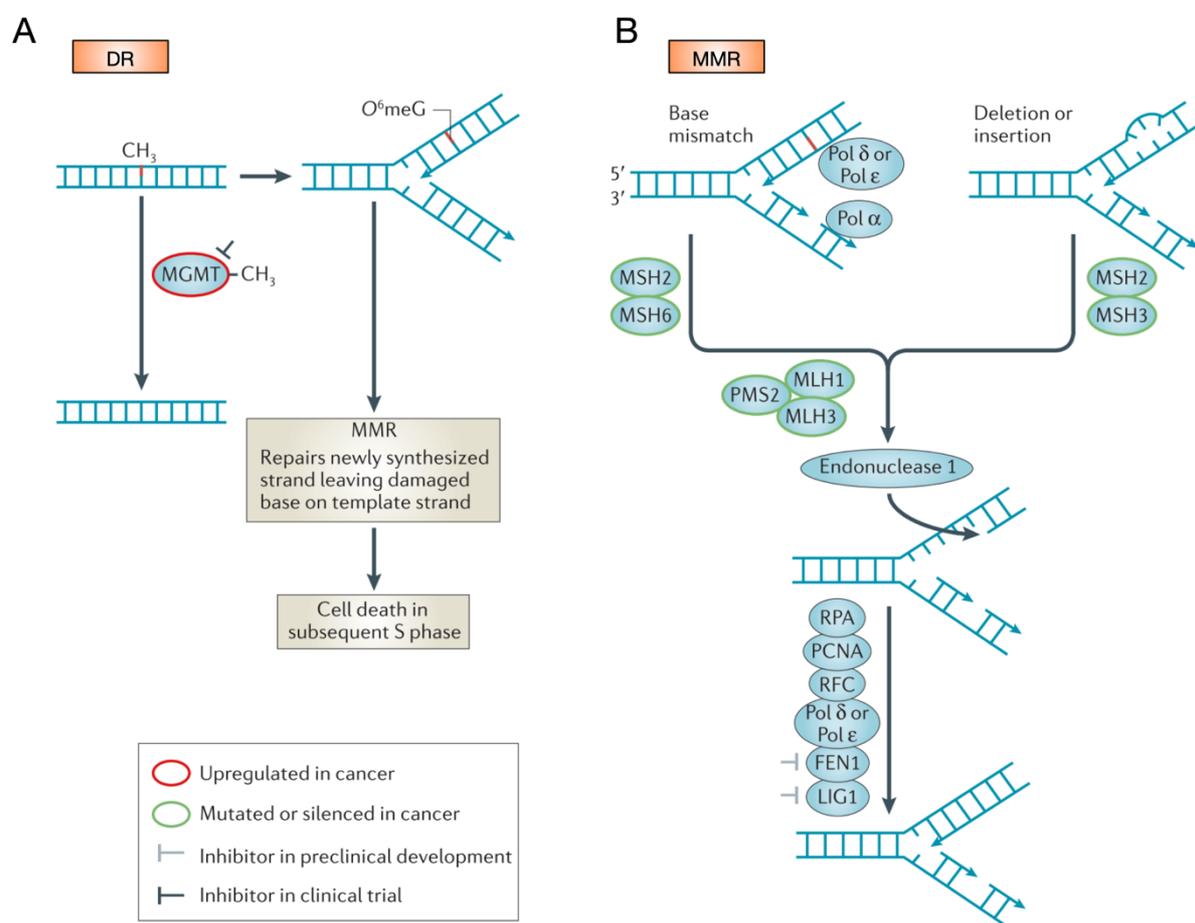


Figure 1: The Direct Repair (DR) and Mismatch Repair (MMR) pathways. A) Recognition and repair of DNA damage by the DR pathway. O⁶-methylguanine (O⁶meG) is recognized and removed by the enzyme MGMT. High expression levels of MGMT are commonly found in cancer and inhibitors are in development. If the damage is not repaired, the persistent O⁶meG alerts the MMR pathway, which excises the mispaired thymine in the newly synthesised strand. This ultimately leads to apoptosis. **B)** The MMR pathway recognises mismatches of 1bp (by MSH2-MSH6), or deletions and insertions (by MSH2-MSH3). Repair of the lesion requires the action of MLH1, MLH3, PMS2, Endonuclease 1 (EXO1), FEN1, PCNA, LIG1 among others. MMR is strand-specific, only correcting the daughter strand. Importantly, mutations in MMR-associated factors are frequently linked to cancer. Figure reprinted and adapted with permission from Springer Nature Reviews Cancer (DNA repair dysregulation from cancer driver to therapeutic target, Curtin NJ (2012)).

1.2.3 Base Excision Repair (BER)

The Base Excision Repair (BER) pathway corrects replication-independent lesions which can be either single-base damage (short patch (SP) - BER) or damage of a few (2-13) bases (long patch (LP) - BER) (**Figure 2A**). The most common lesions repaired by BER are spontaneous deaminations, oxidations (produced by ROS) or alkylations (by SAM). The oxidation of guanine at position 8 (8-oxoG) is one of the most common forms of base oxidation repaired by BER. 8-oxoG is mutagenic, since it mispairs with thymine leading to G:C > A:T mutations. 8-oxoG is found increased in cancer tissues, highlighting the relevance of BER in repairing such lesions (Wiseman & Halliwell, 1996).

To initiate BER, damaged bases are removed by specialised glycosylases (for example, the 8-oxoG DNA glycosylase OGG1), forming abasic sites. Although each glycosylase is specialised for the excision of a specific type of damaged base, they also act in a redundant manner. MUTYH (Kairupan & Scott, 2007) and NEIL1 (Hazra *et al*, 2002) are other DNA glycosylases that act in the repair of distinct oxidised lesions. After the excision, BER endonucleases, such as APE1, induce a nick in the DNA, generating a DNA single-strand break (SSB). This incision is then repaired by either SP- or LP-BER. In SP-BER, the single nucleotide is replaced by Pol- β and the gap is filled by Ligase III (LIG3). LP-BER occurs by the excision of at least 2-13 nucleotides and DNA synthesis is catalysed by Pol- β or by Pol- ϵ/δ (Pascucci *et al*, 1999). Re-joining of the DNA ends is then mediated by Ligase I (LIG1). LP-BER is also dependent on factors such as PCNA, Flap endonuclease 1 (FEN1) and the 9-1-1 clamp complex (Rad9-Hus1-Rad1). Other factors, such as Poly(ADP-ribose) polymerase 1 (PARP1) and XRCC1 (X-ray cross-complementing protein 1) facilitate repair, by recruiting repair enzymes and providing the scaffold for SP- and LP-BER.

1.2.4 Nucleotide Excision Repair (NER)

Nucleotide Excision Repair (NER) is a highly versatile repair pathway that deals with a broad range of structurally distinct helix-distorting adducts on DNA. This pathway is thought to be solely responsible, in humans, for repairing UV-induced lesions that commonly take the form of cyclobutane-pyrimidine dimers (CPDs), where the most frequent are T-T cyclobutane dimers, or 6-4 pyrimidine-pyrimidone photoproducts (6-4PPs) (Marteijn *et al*, 2014). Moreover, NER is known to remove bulky adducts caused by smoking-related carcinogens (for example, benzo[a]pyrene diol epoxide) (Nouspikel, 2009). NER is therefore a fundamental pathway that guards the genome against environmental mutagens.

NER can be divided into two distinct sub-pathways that only differ in how they recognise the lesion: transcription-coupled NER (TC-NER) and global-genome NER (GG-NER) (**Figure 2B**). TC-NER exclusively repairs DNA adducts that have been introduced

during DNA transcription (Bukowska & Karwowski, 2018). Damage is recognised during transcription elongation by RNA polymerase II, upon stalling at sites of lesion in the template strand. Upon stalling, Cockayne syndrome WD repeat protein A (CSA) and B (CSB) are recruited to bind to the lesion. Contrary to TC-NER, GG-NER repairs helix distortions that arise independently of transcription. This is performed by binding of a complex comprised of Xeroderma pigmentosum complementing group C (XPC), RAD23B, DNA damage-binding protein (DDB) and XPE. After the recognition step, TC- and GG-NER converge. The transcription factor II H (TFIIH) unwinds the DNA, through its helicase activity, and permits the binding of the pre-incision complex comprised of XPA and RPA. The damage, now in the form of an oligonucleotide, is then excised by the complex ERCC1 (excision repair cross-complementation group 1) - XPF (5'-end incision) and XPG (3'-end incision). PCNA mediates DNA re-synthesis, catalysed by Pol- ϵ , Pol- δ and Pol- κ , and ligation is performed by DNA ligase III (LIG3).

Mutations in the NER pathway are associated with several disorders, such as Xeroderma pigmentosum (XP), Cockayne syndrome (CS), UV-sensitive syndrome (UVSS) and Trichothiodystrophy (TTD). These diseases have distinct clinical manifestations, but they share an enhanced sensitivity to sunlight exposure and neurological abnormalities. XP is characterised by elevated risk of skin cancer, whereas CS is characterised by progeria syndrome and TTD by cutaneous abnormalities (Bukowska & Karwowski, 2018). Currently, there are no available treatments for NER-deficient disorders.

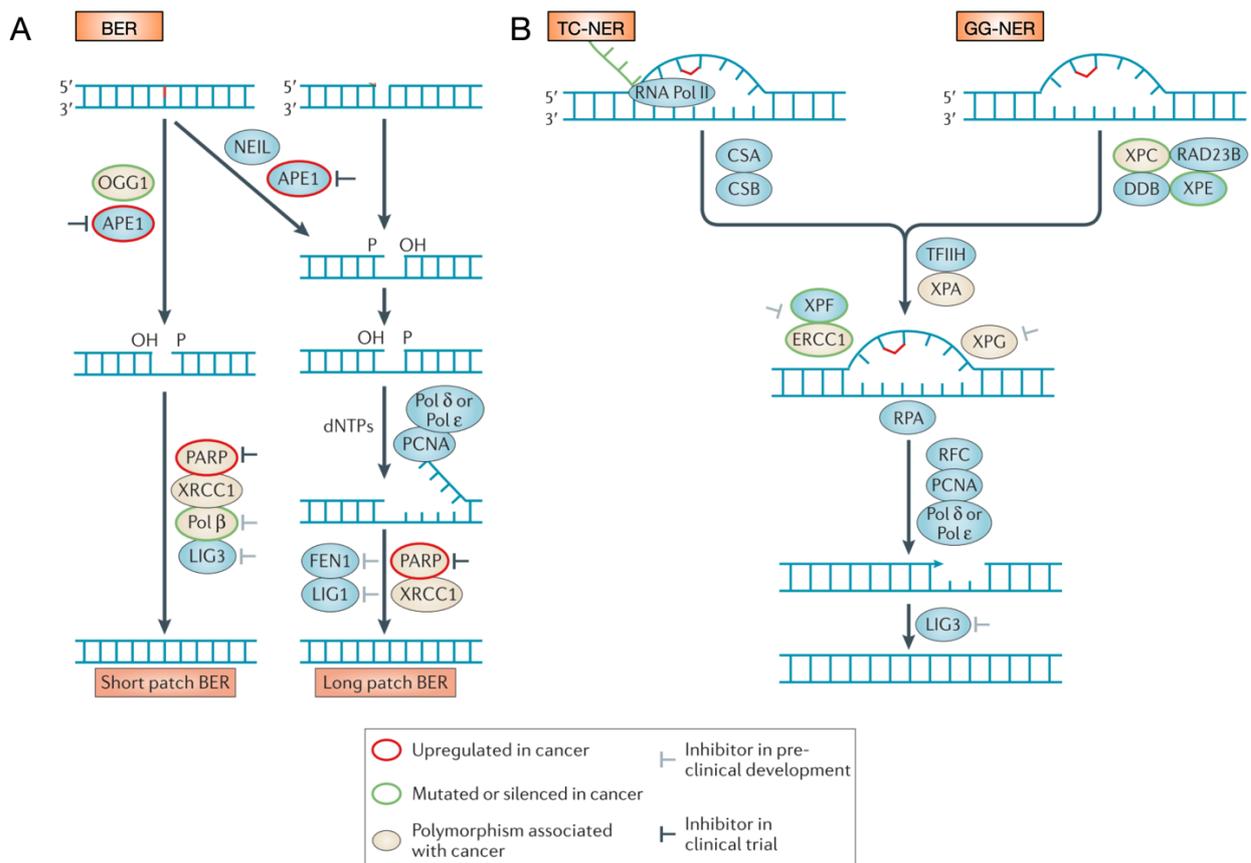


Figure 2: The Base Excision Repair (BER) and Nucleotide Excision Repair (NER) pathways. A) The BER pathway repairs damaged bases, through their excision by specialised glycosylases (such as OGG1 and NEIL1). The resulting abasic site is hydrolysed by the endonuclease APE1, generating a ssDNA nick that is then repaired by short-patch BER (1 nucleotide) or long-patch BER (2-13 nucleotides). In short patch repair, the nucleotide is replaced by Pol-β and the gap is joined by LIG3. For long path repair, PCNA and FEN1 are required for the end-processing. The gap is filled by Pol-β or by Pol-ε/δ and joined by LIG1. PARP1 and XRCC1 are important scaffold factors. **B)** The NER pathway can be sub-divided into TC-NER and GG-NER. TC-NER recognises DNA lesions during DNA transcription, upon RNA polymerase II stalling. It then recruits CSA and CSB. In GG-NER, lesion recognition happens through XPC, DDB, RAD23B and XPE binding. Both pathways converge in the unwinding of the DNA by TFIIH and binding of XPA and RPA. Excision of the damaged oligonucleotide occurs through ERCC1-XPF and XPG. The gap is filled by Pol-ε, Pol-δ and Pol-κ, and ligation is mediated by LIG3. Figure reprinted and adapted with permission from Springer Nature Reviews Cancer (DNA repair dysregulation from cancer driver to therapeutic target, Curtin NJ (2012)).

1.2.5 Translesion synthesis (TLS)

Translesion synthesis (TLS) can be considered more of a DNA damage tolerance mechanism, rather than a DNA repair pathway (Goodman & Roger, 2013). TLS relies on polymerases that are not stalled upon DNA lesions, such as thymine dimers or abasic sites. These polymerases have lower fidelity than normal DNA replication polymerases, hence errors can occur during this process. (Goodman & Roger, 2013). Nonetheless, TLS is important for some processes, such as providing a template for the repair of the DNA double-

strand breaks (DSBs) that are generated after the excision of interstrand-crosslinks (ICLs), as discussed in the next section.

1.2.6 Fanconi Anaemia (FA) repair pathway

The Fanconi Anaemia pathway is named after the rare monogenetic disease that arises when the body is not able to repair interstrand-crosslinks (ICLs). ICLs create a covalent bond between opposite strands of the DNA, blocking cellular essential processes, such as replication and transcription. If not repaired, ICLs can lead to replication fork collapse and apoptosis.

The main ICL-inducing drugs are bifunctional alkylating agents (nitrogen mustards), mitomycin C (MMC) and platinum compounds, such as cisplatin. Due to their high toxicity, most of these compounds are used as chemotherapeutic agents. ICLs can also be induced endogenously by by-products of lipid peroxidation (aldehydes and malonaldehydes) or ethanol metabolism (acetaldehyde and formaldehyde) (Garaycochea et al, 2018; Noll et al, 2006).

Fanconi Anaemia, originally described by Guido Fanconi in 1927, is characterised by bone marrow failure, cancer predisposition (mostly leukaemia) and congenital abnormalities covering a broad range of affected organs (Garaycochea & Patel, 2014). Fanconi Anaemia arises from the loss of one FA or FA-associated protein.

ICLs are recognised by the Fanconi Anaemia Complementation Group (FANC)M, during S-phase, when replication forks are stalled. This leads to the recruitment of the FA-core complex, which consists of 10 proteins including FANCC and FANCL. FANCL, through its ubiquitin ligase activity, activates the downstream effector complex formed by FANCI and FANCD2 (FANCI-D2) (Kottemann & Smogorzewska, 2013). Upon activation of FANCI-D2, repair factors from other pathways are recruited, which cleave the nucleotides flanking the ICL and remove the crosslink from one of the two DNA strands (unhooking). TLS then occurs over the strand from where the crosslink was removed (Klein Douwel *et al*, 2014). Upon ligation, a heteroduplex is formed which serves as template for the homologous recombination (HR) pathway, which will be described later (section 1.2.7.2), to perform templated error-free repair of the crosslink, allowing cell division to proceed.

1.2.7 Double-Strand Break Repair (DSBR)

DSBs (double-strand breaks) are highly cytotoxic lesions that lead to the topological separation of two DNA ends. Studies show that approximately 50 DSBs per cells arise endogenously, per day (Vilenchik & Knudson, 2003). Therapeutically induced DSBs can directly result from the action of topoisomerase inhibitors (such as etoposide or doxorubicine), or ionizing radiation (IR). Moreover, processes of DSBR play fundamental roles in important

cellular processes, such as meiosis or the generation of immune-receptor diversity (Jackson & Bartek, 2009). For this reason, cells have evolved dynamic and complex mechanisms to detect and repair this type of lesions.

The competition between end-joining and templated repair pathways is a key aspect of DSBR. Pioneering studies in the yeast *Saccharomyces cerevisiae* have helped elucidate this mechanism in humans. However, contrary to what is observed in yeast, DSBR in humans depends on a broader diversity of factors and end-joining pathways are favoured over templated repair pathways (Dicarlo *et al*, 2013; Mao *et al*, 2009).

End-joining pathways, such as non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ), repair the DSB by directly re-ligating the broken DNA ends. Templated repair is achieved through the action of the homologous-recombination (HR) pathway or single-strand annealing (SSA) (**Figure 3**). End-joining pathways are error-prone, introducing insertions and deletions (indels) in the DNA, whereas templated repair is error-free. Human cells upregulate end-joining pathways over templated repair through several mechanisms. One key aspect in the competition between end-joining and template-repair is cell cycle. NHEJ is active throughout all cell cycle phases, predominating in G0 and G1 (Shrivastav *et al*, 2008), whereas, HR is favoured when a sister chromatid is present, in G2/S (Chang *et al*, 2017). The processing of DSB ends, through end resection, is another main factor that dictates pathway choice. Unprocessed DSBs, with blunt ends, generally undergo repair through NHEJ, whereas 5'-3' resection of DNA ends directs repair towards HR, SSA, or MMEJ (**Figure 3**).

DSBs are recognised by the MRE11-RAD50-NBS1 (MRN) complex. MRE11 is an exonuclease that performs the initial steps of DNA end-resection. RAD50 binds the two DNA ends, holding them together, and NBS1 recruits the ataxia-telangiectasia mutated (ATM) PI3K-family kinase (Ciccia & Elledge, 2010). Upon activation by auto-phosphorylation, ATM recruits and phosphorylates other factors that are important for the repair of the lesion. One of these substrates is the histone variant H2AX which, upon phosphorylation (becoming γ H2AX), acts as an important signalling marker for DSBs. Other important proteins recruited to DSBs at this stage of repair are the p53-binding protein 1 (53BP1) and breast cancer 1 (BRCA1). 53BP1 is a positive regulator of NHEJ, by binding to the DNA ends and inhibiting end-resection (Daley & Sung, 2014). 53BP1 and BRCA1 compete with each other for repair pathway choice, with BRCA1 favouring HR (Feng *et al*, 2015).

Non-Homologous End-Joining (NHEJ)

In humans, NHEJ is the default pathway by which DSBs are repaired, being active throughout all phases of the cell cycle. NHEJ directly re-joins the broken DNA ends with minimal end-processing, approximately 30 minutes after break induction (Mao *et al*, 2009).

NHEJ is promoted by 53BP1, which prevents end-resection and other pathways that rely on this step. For a long time, the mechanism by which 53BP1 shields DNA-ends was not completely understood. This changed with the discovery of the shieldin-complex (Gupta *et al*, 2018; Noordermeer *et al*, 2018; Dev *et al*, 2018), composed of SHLD1, SHLD2 (FAM35A), SHLD3 (CTC-534A2.2) and REV7. The shieldin-complex binds to DSB sites in a 53BP1- and RIF1-dependent manner. SHLD2 subunits bind to ssDNA, restricting DNA-end-resection and antagonising HR by blocking the binding of BRCA2 and RAD51.

Once the DSB has been committed to NHEJ-mediated repair, the Ku70-Ku80 heterodimer (encoded by the *XRCC6* and *XRCC5* genes, respectively) binds to the broken DNA ends, which leads to the recruitment and activation of the DNA-dependent protein kinase catalytic subunit (DNAPK-cs), forming DNA-PK. DNA-PK recruits a complex assembled between Ligase IV (LIG4), X-ray cross complementing protein 4 (XRCC4) and the XRCC4-like factor, XLF. This complex is responsible for the ligation of the two DNA broken ends (**Figure 3**).

Homologous Recombination (HR)

HR is slower than NHEJ, requiring seven or more hours to complete (Mao *et al*, 2009; Liang *et al*, 1998). HR requires extensive end-resection of approximately 1 kb or more, which is achieved after the short resection by MRN, through the action of the exonuclease 1 (EXO1), Bloom RecQ helicase (BLM) and Dna2-nuclease-helicase, DNA2. The processing of these DNA ends generates long stretches of ssDNA that expose regions of homology between the resected DNA and the template (usually a sister chromatid) (**Figure 3**). BRCA1 mediates the recruitment of RPA, that is responsible for the binding and stabilisation of the ssDNA (Chen & Wold, 2014). This induces an S/G2 cell cycle arrest that is mediated by the phosphorylation of checkpoint kinase 1 (CHK1), as well as RAD51 activation (Jensen *et al*, 2010). RAD51 conducts the homology search for a repair template. Pol- δ mediated extension of the gap and ligation of DNA ends is mainly achieved by LIG1.

Single-Strand Annealing (SSA)

SSA has been described as a sub-type of HR. This pathway requires regions of homology exceeding 200 bp in mammals and 20 bp in yeast (Ren *et al*, 2014; Liskay, R. M., Letsou, A. & Stachelek, 1987). SSA relies on RAD52, instead of RAD51, which binds to the

resected RPA-coated ssDNA (**Figure 3**). SSA generates flaps in the DNA that are resolved by the ERCC1-XPF nuclease complex and the filling of the DNA gaps is performed by unknown players, highlighting how unexplored this pathway is. SSA is conceptually closer to MMEJ rather than HR, but it generates larger deletions than MMEJ.

Microhomology-Mediated End-Joining (MMEJ)

MMEJ, also known as alternative end-joining (alt-EJ), is a recently discovered pathway to repair DSBs sharing characteristics with templated repair (both HR and SSA) and NHEJ. Similar to HR and SSA, MMEJ requires an initial, although shorter, end-resection step, exposing 5-25 bp regions of microhomology that flank the DSB. However, like NHEJ, MMEJ repairs the DSB without requiring a repair template, leading to an error-prone outcome (Seol *et al*, 2018). DNA flaps after resection are removed by the endonuclease complex ERCC1-XPF and DNA polymerase θ (POLQ) fills the gap (**Figure 3**). The break is sealed by LIG1 or LIG3. PARP1 is also required for MMEJ, by inhibiting Ku binding and therefore competing with the NHEJ pathway.

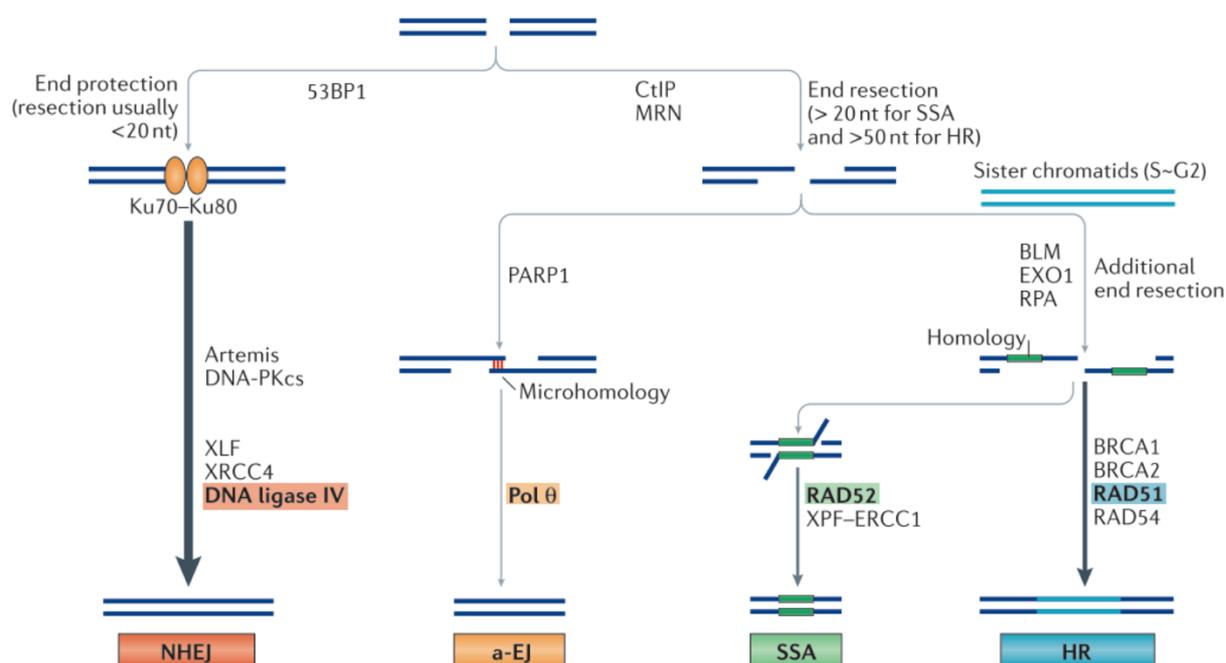


Figure 3: Double-Strand Break (DSB) Repair Pathway Choice. DSBs can be repaired by non-homologous end-joining (NHEJ), alternative end-joining (a-EJ or MMEJ), single-strand annealing (SSA) or homologous recombination (HR). 53BP1 acts as a positive regulator of NHEJ, inhibiting end-resection. NHEJ does not require extensive end-resection and the DNA ends are protected by the Ku70-Ku80 heterodimer. Ligation is performed by LIG4, stabilized by the XRCC4-XLF complex. CtIP and MRN are involved in extensive 5'-3' end resection, generating stretches of ssDNA. A-EJ requires <25 bp microhomology and it depends on PARP1 and Pol θ . SSA requires >20 bp of microhomology regions and is dependent on RAD52 and the XPF-ERCC1 complex. EXO1 and BLM provide the additional end-resection that is necessary for HR to occur. RPA binds to ssDNA for protection. RAD51-mediated strand exchange, in combination with BRCA1 and BRCA2, promotes error-free repair by HR. Figure reprinted with permission from Springer Nature Reviews Molecular Cell Biology (Non-

homologous DNA end joining and alternative pathways to double-strand break repair, Chang H, et al (2017)).

1.3 DDR & Human Disease

The relevance of the DDR to maintain genome integrity and stability is highlighted by the diversity of severe diseases that arise in humans following its impairment (**Table 1**).

1.3.1 Cancer

An intrinsic characteristic of cancer development is genome instability (Stratton *et al*, 2009). Additionally, most cancer treatments rely on the generation of DNA damage (Hoeijmakers, 2001). Therefore, it is common to observe germline polymorphisms or mutations in DDR genes predisposing to cancer. Additionally, somatic mutation or epigenetic silencing of DDR genes contribute to tumour development and subsequent malignant progression. In fact, current knowledge suggests that DDR defects are positively selected in cancer cells, as a way to tolerate oncogene-induced replication stress (Pearl *et al*, 2015). An example already mentioned is the predisposition to cancer (mostly colorectal and endometrial carcinomas) observed in MMR deficient tumours. Loss of some elements of one DNA repair pathway might be compensated by another pathway, posing both challenges and opportunities for cancer therapy, as explored in section 2.1.

1.3.2 Heritable Human Diseases of Genetic Instability

Several human diseases are caused by genomic instability triggered by expansions and contractions of unstable DNA sequences. The repetitive nature of these sequences frequently cause aberrant secondary structures in the DNA (Jackson & Bartek, 2009). If there is an impairment in the DDR, these structures cannot be resolved, leading to diseases that commonly present with neuromuscular and neurodegenerative abnormalities. Notable examples of such disorders are fragile X syndrome (expanded CGG repeats in the *FMR1* gene), Friedrich's ataxia (expanded GAA repeats in the *FXN* gene), spinocerebellar ataxias (expanded CAG repeats in various genes) and Huntington's disease (expanded CAG repeat in the *HTT* gene).

1.3.3 Neurodegenerative Disorders

Neurons exhibit high levels of ROS production, which can be damaging for DNA (Weissman *et al*, 2007). Additionally, neurons are heavily dependent on transcription, a process that is blocked by oxidative lesions in the DNA. Consistent with the role of the BER

pathway in the repair of oxidative damage, mutations in this pathway are associated with neurodegenerative disorders, including ataxias, Alzheimer's, Huntington's and Parkinson's diseases (Jackson & Bartek, 2009).

Moreover, by being arrested in G0, neurons are not able to repair DNA DSBs by templated repair mechanisms (such as HR). These cells mostly rely on the error-prone NHEJ pathway, which contributes to genomic instability. This limitation in cell division also explains the potential accumulation of permanently damaged neurons. The DDR dependencies of neuronal cells can potentially explain the neurodegeneration that is observed in diseases caused by defects in genes associated with DSBR and TC-NER, such as ataxias and Cockayne syndrome (CS), respectively.

1.3.4 Ageing

Ageing is, in part, a consequence of the accumulation of DNA damage, being the result of a combined accumulation of DNA lesions over time and an associated decline in the ability to repair them. The link between DNA repair and ageing is further supported by the fact that patients with inherited disorders in DDR genes often display features of progeria (accelerated ageing).

An example of a disease that manifests in progeria is Hutchinson-Gilford Progeria Syndrome (HGPS). HGPS is a rare syndrome with dramatic consequences. Individuals suffering from this disease develop severe growth abnormalities within two years after birth. HGPS patients display features of progeria and generally die of atherosclerosis and cardiovascular complications. HGPS is caused by a *de novo* heterozygous mutation (G608G) in exon 11 of the *LMNA* gene, which activates a cryptic splice site in this gene. This mutation leads to the production and accumulation of a permanently farnesylated and carboxymethylated version of the LMNA protein, called progerin. Progerin accumulates in the nuclei of cells, disrupting nuclear architecture and causing genomic instability which leads to premature senescence (Gonzalo & Kreienkamp, 2015). Moreover, HGPS has been shown to be associated defective DSB repair, with mouse and human fibroblasts displaying higher levels of DSBs, as measured by γ H2AX levels. (Scaffidi & Misteli, 2006; Liu *et al*, 2005). Overall, this evidence highlight that the underlying genome instability observed in progerin syndromes, such as HGPS, often involve impairment of the DDR. Genome editing has recently been shown to be a promising therapeutic approach for HGPS, as described on later chapters.

1.3.5 Cardiovascular Disease and Metabolic Syndrome

Several studies suggest a link between the DDR and cardiovascular disease (Mercer *et al*, 2007). The pro-apoptotic action of p53 can become detrimental in the setting of atherosclerosis and, subsequently, lead to stroke or heart attack. Moreover, enhanced DNA damage and its signalling has been characterised in atherosclerosis. An example of this is a recent study describing the involvement of the NHEJ protein DNA-PK in atherosclerosis development. DNA-PK activity was shown to increase with atherosclerosis progression, playing an important role in the repair of DNA damage observed in the vessel wall (Haemmig *et al*, 2020).

Overall, this and other interactions highlight the role that the DDR plays in modulating metabolic and cardiovascular mechanisms, with relevance to human disease. The interesting correlation between DDR and metabolism is, however, still largely unexplored.

1.3.6 Immune Deficiencies

Fundamental processes for a functional innate immune response rely on DDR mechanisms, particularly those of DSBR. These processes generate the immune-receptor diversity that allows an effective recognition and clearance against antigens and pathogens (Litman *et al*, 2007). Therefore, defects in factors involved in DSBR are frequently associated with immune deficiencies. That is the case, for example, for NHEJ mutations that have been linked to immunodeficiencies (Jackson & Bartek, 2009). Additionally, patients with ataxia telangiectasia (*ATM /ATR* mutations), or Nijmegen breakage syndrome (*NBS1* mutations) show impaired immunity, which manifests in recurrent infections.

Table 1: Summary of DDR mechanism and respective type of damage, prime lesions, key factors and associated diseases

DDR mechanism	Main sources of damage	Prime lesions acted upon	Key protein components	Associated diseases
Direct Repair (DR)	SAM; Nitrosated amines; Temozolomide; Dacarbazine; Alkylating agents	O6-methylguanine	MGMT	
Mismatch Repair (MMR)	DNA replication; SAM; Base deamination	DNA mismatches and indels arising from DNA replication;	MSH2/3/6; MLH1/3; PMS2; RFC; PCNA; EXO1; RPA; POLE/D; LIG1	Hereditary and sporadic cancers
Base Excision Repair (BER)	ROS; SAM; IR; Temozolomide	8oxoG or other abnormal bases; SSBs; simple base adducts	OGG1; MUTYH; NEIL1; APE1; POLB/E/D; LIG3/1; FEN1; PARP1; PCNA; XRCC1	Ataxias; Alzheimer's; Huntington's; Parkinson's
Nucleotide Excision Repair (NER)	UV; Tobacco; ROS	CPDs; 6-4PPs; Bulky adducts	CSA/B; XPA/C/E/G; RAD23B; DDB; TFIIH; ERCC1-XPF; RPA; POLE/D/K; LIG3	Xeroderma pigmentosum; Cockayne syndrome; UVSS; TTD
Translesion Synthesis (TLS)	DNA replication	Base damage blocking replication-fork progression	Translesion DNA polymerases	Cancer
Fanconi Anemia (FA)	Bifunctional alkylating agents; MMC; platinum compounds; aldehydes	Interstrand cross-links	FANCA/C/D1/D2/E/F/G//L/M; BRCA1	Fanconi Anemia
Double-strand Break Repair (DSBR)	IR; Replicative stress	DNA DSBs	MRE11; RAD50; NBS1; 53BP1; BRCA1; ATM; ATR; H2AX; DNAPK; LIG4; XRCC4; XLF; CtIP; EXO1; DNA2; CHK1/2; RPA; RAD51; BRCA2; RAD52; POLQ; PARP1; LIG1/3	Li-Fraumeni syndrome; Cancer; Ataxia telangiectasia; Nijmegen breakage syndrome; RIDDLE syndrome; Seckel syndrome; HGPS; Lig4 syndrome;

2. Exploiting the DDR Knowledge to Treat Human Disease

2.1 Cancer Therapy

Chemotherapy and radiotherapy are two of the most common forms of cancer treatment. They function by generating DNA damage that exceeds the tumour cell capacity for repair. Among others mechanisms, these agents work by either blocking topoisomerases on the DNA (e.g. etoposide and doxorubicin), alkylating bases (e.g. temozolomide) or by inducing ICLs (e.g. cisplatin). Even though chemotherapy and radiotherapy affect the normal tissue to a certain degree, they are often more efficient in damaging the tumour tissue, reflecting the frequency of DDR impairment in cancer cells and the higher rates of proliferation that make them more vulnerable to DNA damage.

On the other hand, DNA repair is also frequently associated with cancer therapy resistance, since upregulated DNA repair pathways can cause resistance to DNA-damaging chemotherapeutic agents. For this reason, DNA repair inhibitors have the potential to sensitize tumour cells to therapy (Curtin, 2012). Currently, there are a number of compounds directly targeting the DDR under clinical evaluation for cancer therapy. These compounds target cell cycle-regulating kinases that are induced by DNA replication stress or damage (CHK1 and WEE1, for example), enzymes involved in BER (such as the apurinic-apyrimidinic endonuclease 1, APEX1), DR (such as MGMT), DSBR (DNA-PK, POLQ) or telomere maintenance (such as the telomerase reverse transcriptase, TERT) (Pearl *et al*, 2015).

The initial rationale for the design of drugs that target DDR-enzymes was to use them as potentiators, inhibiting the mechanisms that repair the damage caused by radio and chemotherapy in cancer cells. However, these inhibitors soon became stand-alone therapies, by targeting the DDR mechanisms that are essential for the cancer cells to survive. This gave rise to the concept of synthetic lethality. Synthetic lethality can be defined as a type of interaction in which the combined loss of function of two genes leads to cell death, whereas loss of function of only one of these genes does not impair cell viability (**Figure 4**).

The best example of a synthetic lethal interaction being exploited in the clinical setting is the one observed between PARP1, a key factor in the repair of SSBs, and the HR factors BRCA1 and BRCA2. PARP1 inhibition is not lethal to normal cells, but highly cytotoxic to cells that have defective HR (Fong *et al*, 2009; Ding *et al*, 2016). This is because PARP inhibition results in the trapping of the PARP-DNA complex at replication forks, causing the conversion of SSBs into DSBs, that need to be effectively repaired by HR. In BRCA-deficient cells, since HR is not functional, these DSBs are not repaired in an error-free manner, contributing to genome instability and apoptosis of the cancer cell.

The most clinically advanced PARP1 inhibitor is olaparib (also known as AZD-2281), which has been approved for use in multiple cancer types associated with germline *BRCA* mutations (Caulfield *et al*, 2019). Several other PARP inhibitors have completed clinical trials to determine their application for cancer treatment (Jiang *et al*, 2019). Since HR defects are observed in several cancer types, such as breast, ovarian, prostate and pancreatic cancer, PARP1 inhibitors have a broad applicability.

Overall, among other avenues, personalised cancer therapy can be achieved by understanding the DDR fingerprint of a cancer cell and how it differs from the one of a normal cell. This allows for the exploitation of vulnerabilities and the subsequent design of intelligent, tailored therapies and more sensitive early malignancy detection methods.

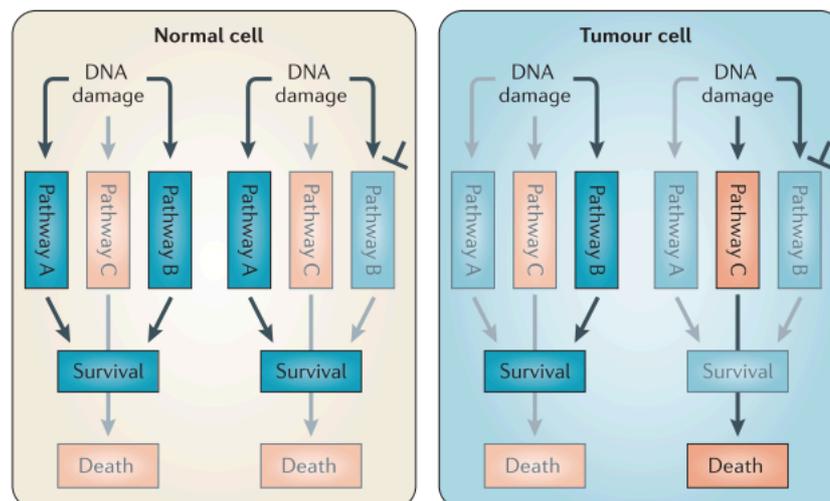


Figure 4: Illustration of the concept of synthetic lethality. Normal cells respond to DNA damage by activating specialised pathways (A or B). There might also be an inappropriate pathway C that, upon activation from the damage, leads to cell death. This pathway is inactivated in normal cells. If normal cells are treated with an inhibitor for pathway B, they are still able to deal with the damage, through activation of pathway A. In a tumour cell where pathway A suffered a silent mutation, survival relies exclusively on pathway B. If an inhibitor for pathway B is applied, the tumour cell recurs to pathway C, which leads to cell death. Figure reprinted with permission from Springer Nature Reviews Cancer (DNA repair dysregulation from cancer driver to therapeutic target, Curtin NJ (2012)).

2.2 Gene Therapy

Correcting gene dysfunction at the DNA level is a promising approach to treat many human diseases, including genetic diseases and cancer. Substantial advances in the field of human genome editing have been recently achieved, through approaches that harness the potential of the DDR. This will be further explored in the next sections.

3. Genome Editing: Bridging DDR & Human Disease

3.1 A Historic Perspective of Genome Editing

Since the discovery of the DNA double helix, in 1953 (Watson & Crick, 1953), several technologies have advanced with the intend of reading and altering genome sequences and gene-expression patterns in cells and organisms. The advent of genomic sequencing and the generation of whole-genome sequencing data have been fundamental for the advances in the genome editing field. The study of endogenous DNA repair pathways in several model organisms throughout the years has highlighted the existence of mechanisms that are involved in the repair of DNA DSBs. This knowledge led to the realisation that inducing DSBs at specific regions of the genome is a valuable strategy for targeted genome engineering.

Early genome editing approaches explored the utility of base-pairing for site-specific genome modification. These included oligonucleotides coupled to chemical cleavage (Strobel

& Dervan, 1990, 1991) self-splicing introns (Yang *et al*, 1996), or RNA (Sullenger & Cech, 1994). However, these methods were never considered robust approaches for genome engineering.

During the early 2000's, endonucleases became increasingly useful tools to target specific regions of the genome. First, zinc finger-mediated DNA binding domains, coupled with the nuclease domain of the restriction enzyme FokI, were exploited for editing of DNA sequences. Zinc finger nucleases (ZFNs) work by delivering a site-specific DNA DSB to the genome, that is in turn repaired by the endogenous DNA repair machinery of the cell, either in an error-prone (by end-joining pathways), or error-free manner (by templated-mediated repair pathways) (Hossain *et al*, 2015). The widespread adoption of ZFNs was however hampered by difficulties inherent to the design of such proteins. For this reason, transcription activator-like effector nucleases (TALENs), which were easier to manipulate than ZFNs, were more widely used. TALENs occur naturally in bacteria and, like ZFNs, induce a targeted DSB mediated by FokI. TALENs were shown, among other applications, to efficiently inactivate the HIV receptor CCR5 in somatic cells (Miller *et al*, 2011) and engineer T-cells for the treatment of childhood B-cell leukaemia (Qasim *et al*, 2017).

Despite these advancements, technical challenges hampered the widespread implementation of these nuclease-based genome editing technologies. This was only solved in the early-2000s with the discovery of a system that truly revolutionised not only genome editing, but the entire field of molecular biology.

3.2 CRISPR-Cas systems

CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats) were first described in 1987, as a series of short repeats, interspaced by short undescribed spacer sequences in the genome of *E.coli* (Ishino *et al*, 1987). It was later discovered that the spacer sequences have extrachromosomal origin and that CRISPRs are present in several bacteria and archae (Bolotin *et al*, 2005; Mojica *et al*, 2005; Pourcel *et al*, 2005). The observation that CRISPRs are transcribed into RNA and that Cas (CRISPR-associated) genes encode proteins with nuclease and helicase functions, led to the postulation that CRISPR-Cas systems act as an antiviral defence mechanism in prokaryotes. These systems would have been conserved across prokaryotic species and selected to provide acquired resistance (Barrangou *et al*, 2007), by using an anti-sense RNA as a memory signature of past virus infections (Brouns *et al*, 2008).

CRISPR-Cas systems have been classified into three major types (type I, II and III) and 12 subtypes, given their genetic and functional differences (Barrangou & Marraffini, 2014). Overall, CRISPR-Cas-mediated adaptive immunity functions through three steps (**Figure 5**):

- 1) Adaptation: new spacers are copied from foreign nucleic acids and integrated into the CRISPR locus;
- 2) CRISPR RNA (crRNA) biogenesis: CRISPR arrays are transcribed and processed into small interfering crRNAs;
- 3) Targeting: cleavage of foreign nucleic acids by Cas proteins is directed at DNA sequences complementary to the crRNA spacer sequence.

Several studies conducted in *Streptococcus thermophilus* have shown that most areas of the viral genome can be targeted, including non-coding sequences (Barrangou & Marraffini, 2014). However, there is a danger that CRISPR systems can target host sequences. Therefore, CRISPRs must recognise and distinguish 'self' from 'non-self'. The sequence in the exogenous nucleic acid that corresponds to the CRISPR spacer has been named the protospacer sequence and is flanked by a highly conserved motif, called protospacer adjacent motif (PAM) (Mojica *et al*, 2009). PAMs have been identified in several type I and type II systems, but not for type-III CRISPR systems (Barrangou & Marraffini, 2014). Interestingly, viruses frequently mutate PAM motifs, as a way to escape CRISPR immunity, highlighting the relevance of these conserved sequences (Sun *et al*, 2013).

Another difference that is observed between the three CRISPR systems is the different cas gene content. Type I and III are closer to each other, as both systems rely on Cas6 to cleave the repeat sequences of the crRNA precursor, to generate small crRNAs. This cleavage typically occurs 8 bp upstream of the 5'-end. Targeting is mediated by a large, multi-subunit ribonucleoprotein complex. In type-I systems, the targeting complex is called Cascade (CRISPR-associated complex for antiviral defence) (Brouns *et al*, 2008) and, in *E.coli*, it is formed by Cse1, Cse2, Cas7, Cas5 and Cas6e subunits. Cleavage of the nucleic acid is performed by Cas3, which is not a member of the Cascade complex, characterised in *Streptococcus thermophilus* (Sinkunas *et al*, 2013). Two sequences are essential for cleavage to happen, the PAM sequence (AAG motif, immediately upstream of the protospacer) and the 'seed' sequence (6-8 bp within the crRNA, strictly complementary to the target sequence) (Semenova *et al*, 2011). The CRISPR type-III system also uses a large ribonucleoprotein targeting complex, with the subunit Cas10 harbouring its catalytic activity (Doudna & Charpentier, 2014).

In contrast to type-I and III, type-II CRISPR systems require minimal Cas machinery for immunity (Jinek *et al*, 2012). This property laid the foundation for the use of CRISPR-Cas systems as useful tools for genome editing.

3.3 Functionality of the CRISPR-Cas9 System

In type-II CRISPR systems, a trans-activating crRNA (tracrRNA) is essential for crRNA maturation by ribonucleoprotein III and Cas9. This is a small RNA trans-encoded upstream of the CRISPR-Cas locus (**Figure 5**). In 2012, a landmark study that conferred the 2020 Chemistry Nobel Prize to Emmanuel Charpentier and Jennifer Doudna was published, describing an application of the CRISPR-Cas9 system in genome engineering (Jinek *et al*, 2012). In this study, by using *Streptococcus pyogenes* (*S.pyogenes*) as model, the authors showed the mature crRNA and the tracrRNA can be engineered into an RNA complex that directs the endonuclease Cas9 to a particular region of the genome, where it induces a precise DSB. This RNA structure, called single-guide RNA (sgRNA), retains both the crRNA 20 bp sequence which allows base pairing to the DNA target site and a duplex RNA structure that binds to Cas9. (**Figure 5**).

Cas9 is a large multifunctional protein with two independent nuclease domains, HNH and RuvC-like, and an α -helix-region with an Arginine-rich bridge helix, called the REC domain. The HNH nuclease domain cleaves the strand complementary to the 20 bp crRNA sequence, whereas the RuvC-like domain cleaves the opposite strand. The combined catalytic activity of these two Cas9 domains is what induces a DNA DSB. Both the base pairing to the crRNA sequence and the recognition of a PAM sequence (NGG, at the 3'-end of the protospacer) are essential for efficient DNA targeting by the sgRNA:Cas9 complex (**Figure 5**) (Jinek *et al*, 2012; Gasiunas *et al*, 2012).

The engineering of the sgRNA created a unique and revolutionary way to harness the CRISPR-Cas9 system to program DSBs at any sequence of interest with an adjacent PAM motif. Contrary to ZFNs and TALENs, the CRISPR-Cas9 system only requires minimal engineering, making it an easy and cost-effective technology to target DNA sequences precisely and efficiently.

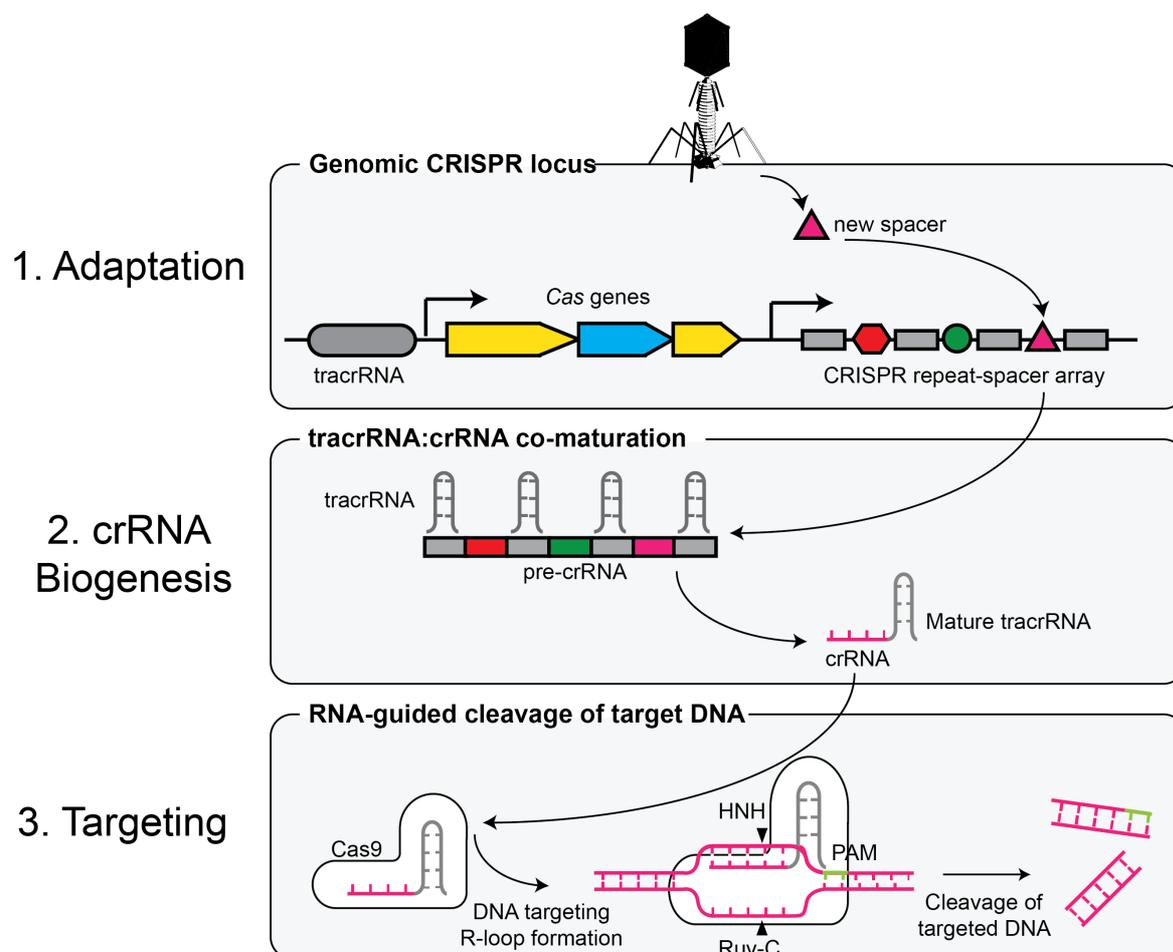


Figure 5: Schematics of the CRISPR-Cas (type II) system. In the adaptation phase, foreign genetic material, from a bacteriophage, is integrated into the CRISPR locus in a form of spacer, between repeat sequences. The CRISPR locus also contains a tracrRNA and Cas genes. This array is subsequently transcribed into the pre-crRNA, that gets further processed into crRNA which anneals with the mature tracrRNA. This complex is recognized by the protein Cas9 (in the targeting phase) that targets the foreign DNA complementary to the crRNA sequence. The protospacer adjacent motif (PAM) sequence, 3' to the crRNA-complementary sequence in the foreign DNA, is responsible for Cas9 recognition and binding. Targeted DNA is then cleaved by the endonuclease activity of Cas9, through the formation of a double-strand break. Bacteria and archaea use this system to acquire memory of past virus infections and cleave the foreign genetic material.

3.4 Cas Diversity

Cas9 is an endonuclease exclusively associated with CRISPR type-II systems. Despite sharing similar domain architecture (HNH, RuvC-like and REC domains), there is a high structural diversity of Cas9 proteins (Hsu *et al*, 2014). Furthermore, Cas9 variations have been engineered, allowing the expansion of the CRISPR-Cas9 toolbox (further discussed in section 3.7). Mutations in either the HNH (H840A), or the RuvC-like domain (D10A), generate variants of Cas9 with ssDNA cleavage capacity, called nickase Cas9 (nCas9). nCas9 H840A cleaves the non-target strand, whereas nCas9 D10A cleaves the target strand. Additionally, simultaneous mutating both domains results in an RNA-guided DNA binding protein that is catalytically 'dead' (dCas9) (Jinek *et al*, 2012; Gasiunas *et al*, 2012).

Biochemical studies have reported that Cas9 recognises DNA target sites by bending the DNA strand at each PAM, flipping the base pairs out of the DNA duplex and toward the sgRNA, for sequence interrogation (Cofsky *et al*, 2021). Moreover, Cas9 has been shown, *in vitro*, to remain bound to the cut DNA for hours (Brinkman *et al*, 2018), but only for minutes *in vivo* suggesting that Cas9 is rapidly displaced from chromatin. The mechanisms that trigger Cas9 to leave the cut DNA are not yet completely understood, but there is evidence suggesting that the histone chaperone FACT might have an important role (Wang *et al*, 2019).

Another CRISPR RNA-guided endonuclease that has been applied to genome editing is Cas12a (also known as Cpf1) (Zetsche *et al*, 2015). In contrast to Cas9, Cas12a generates a staggered cut, which can be particularly useful for applications that require DNA integration in a particular orientation.

More recently, naturally RNA-targeting endonucleases, such as Cas13, have been described and used for applications that include RNA editing, as discussed in section 3.7.

3.5 The PAM Motif: Cas9 Search Mechanism and Target Binding

Besides allowing the distinction between 'self' and 'non-self', PAM recognition activates Cas9 to alter its conformation, from target binding to cleavage (Sternberg *et al*, 2014). Single-molecule studies have shown that the Cas9-sgRNA complex first associates with the PAM, allowing Cas9 to induce the DNA strand separation. The PAM motif lies within a base-paired structure (Anders *et al*, 2014), which indicates that DNA melting starts at the level of PAM recognition, resulting in directional R-loop formation (DNA:RNA hybrid and the associated non-template ssDNA), expanding towards the distal end of the protospacer sequence (**Figure 5**).

There is a high diversity and complexity of PAM motifs. The 5'-NGG PAM motif from *S. pyogenes* Cas9 (SpCas9) allows the targeting, on average, of every 8-12 bp in the human genome (Hsu *et al*, 2013).

Importantly, PAM engineering studies have created Cas9 variants with broader and more flexible PAM motifs. In 2018, David Liu and colleagues (Hu *et al*, 2018a) applied continuous evolution to generate and select for Cas9 mutants with broader PAM compatibility, obtaining xCas9, a Cas9 variant cleaving multiple PAMs at a higher efficiency than SpCas9. Since the generation and characterisation of xCas9, other PAM-flexible Cas9 variants have been engineered and, more recently, a 'PAMless' Cas9 (SpRY) has been developed, through protein rational design (Walton *et al*, 2020). Generally, this flexibility comes at the cost of lower DNA-cleavage activity (Legut *et al*, 2020). Nonetheless, these tools can be extremely useful for other genome-engineering applications for which stricter Cas9 positioning is required, such as single base editing, as discussed later on (section 3.7.3).

3.6 Applications of CRISPR-Cas9-Mediated Genome Engineering

By utilising a 'humanised' version of Cas9, coupled with custom-designed sgRNAs, CRISPR-Cas9 has proven to be an efficient method to edit genomes (Jinek *et al*, 2012; Cong *et al*, 2013; Mali *et al*, 2013). Cas9-induced DSBs induce the endogenous DNA repair machinery of the cell, leading to gene disruption, or knock-out (KO), mostly through the NHEJ pathway, or gene replacement (by HR). Multiplexing (i.e. the use of multiple sgRNAs simultaneously) was also shown to be possible (Cong *et al*, 2013; Mali *et al*, 2013), offering a promising approach to study polygenic diseases.

Moreover, as a genome-engineering tool, CRISPR-Cas9 offers promising applications for agricultural research. CRISPR-Cas9-mediated editing has been used in crops, such as wheat and rice, revealing to be a powerful method to produce variants that are disease-protected (Doudna & Charpentier, 2014).

Importantly, the efficient levels of genome editing achieved by the CRISPR-Cas9 system allow the perturbation of many targets in parallel, making it possible to conduct functional genome-scale screens, in an unbiased and high-throughput manner. Genome-wide or focused lentiviral sgRNA libraries can be designed to introduce loss-of-function mutations into different genes, in each cell (Wang *et al*, 2014; Shalem *et al*, 2014; Sanjana *et al*, 2014; Doench *et al*, 2016). These phenotypes can then be positively or negatively selected, allowing the simultaneous interrogation of multiple perturbations, using a broad range of readouts, such as cell fitness (Shalem *et al*, 2014; Wang *et al*, 2014), fluorescence-activated cell sorting (FACS) (Parnas *et al*, 2015; Brockmann *et al*, 2017; Park *et al*, 2017), RNA-sequencing (Dixit *et al*, 2016; Jaitin *et al*, 2016; Datlinger *et al*, 2017; Replogle *et al*, 2020) or microscopy (Feldman *et al*, 2019). For the past years, functional CRISPR screens have been found to be powerful and widely versatile approaches that have significantly increased our knowledge of biological systems.

Finally, CRISPR-Cas9-mediated genome engineering offers great therapeutic potential for the correction of human diseases. This will be discussed in detail in later chapters.

3.7 Emerging CRISPR-Based Technologies for Genome Engineering

The cleavage of the DNA by Cas9 induces genome editing, following repair by either end-joining or templated-mediated repair pathways. However, it is the intrinsic programmability of this system that makes this technology so unique. Alternatives to cleavage-induced editing have been developed to expand the CRISPR-Cas9 toolbox. Additionally, fusing Cas9 variants with effector proteins has the potential to achieve a wider variety and specificity of genomic alterations.

3.7.1 Transcriptional Control by CRISPR interference (CRISPRi) and activation (CRISPRa)

Transcription regulation is exquisitely complex in eukaryotes. Genes are controlled by activating and repressing transcription factors that act in regulatory elements spanning large regions of the genome. Further levels of regulation include epigenetic modifications or histone acetylation and methylation. The lack of tools for targeting transcription and epigenetic modifiers to specific regions of the genome has been a long-standing challenge.

In 2013, dCas9 was fused to effector proteins with opposing gene regulatory functions, in order to create a flexible RNA-guided platform to modulate transcription (Qi *et al*, 2013; Gilbert *et al*, 2013). The most effective domains fused with dCas9 for CRISPR interference (CRISPRi) and CRISPR activation (CRISPRa) are the heterochromatin inducer KRAB (Krüppel-associated box) and the transcriptional activator VP64 (which consists of four copies of the herpes simplex VP16 activation domain), respectively (**Figure 6**). Both CRISPRi and CRISPRa have been proven to be effective approaches to respectively up- or down-regulate gene transcription in human cells.

Since the advent of CRISPRi/a, the technology has proved its value as a robust platform to modulate transcription at a genome-wide level, showing high reproducibility and efficiency, with low off-targets. More recently, a CRISPRi-based approach has been developed to titrate expression of human genes. This approach is based on sgRNAs containing mismatches to their target site, allowing differential knock-down of gene expression. Mismatched sgRNAs coupled with CRISPRi can have important implications for fine tuning pathways or studying genes whose complete depletion would lead to cell death (Jost *et al*, 2020).

Importantly, CRISPRi and CRISPRa screens for growth phenotypes have yielded complementary insights (Gilbert *et al*, 2014; Horlbeck *et al*, 2016). CRISPRi screens have identified essential genes in different contexts, highlighting cancer-specific vulnerabilities that can be exploited therapeutically (Kampmann, 2018). CRISPRi can also be a useful strategy to interrogate the effects of downregulating essential genes (for which a KO would be lethal), as it is the case with many DDR-associated genes.

On the other hand, CRISPRa has tremendous potential to explore and elucidate mechanisms of drug resistance, which arise frequently from gain-of-function mutations. A notable example was the CRISPRa screen performed in BRAF(V600E) melanoma cells for resistance to a commonly used BRAF inhibitor, that not only recapitulated previously known resistance mechanisms mediated by EGFR and Erk activation, but also revealed novel resistance mechanisms (Konermann *et al*, 2015). CRISPRa can also be a useful approach for cellular reprogramming (Black *et al*, 2016; Liu *et al*, 2018).

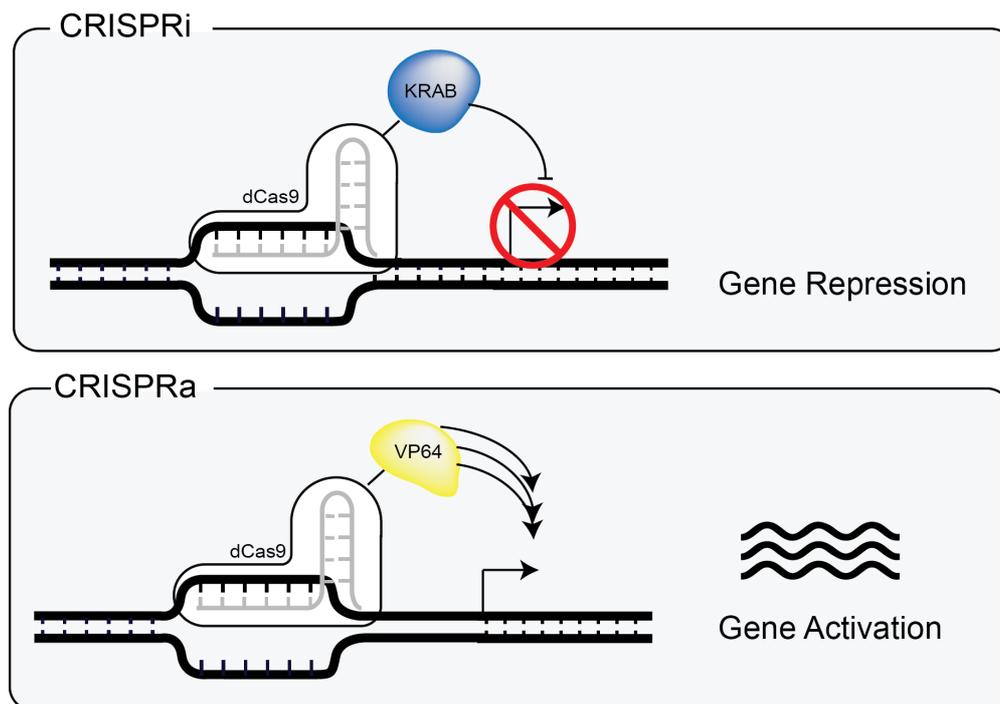


Figure 6: Transcriptional control via CRISPR interference (CRISPRi) or CRISPR activation (CRISPRa). In CRISPRi, a catalytically inactive Cas9 (dCas9) is fused with the Krüppel-associated box (KRAB), an heterochromatin inducer that blocks transcription. In CRISPRa, dCas9 is fused with VP64 that activates transcription, leading to gene activation.

3.7.2 Epigenome Editing

dCas9 fusions can also be employed to target epigenetic modifications, such as acetylation and methylation of histones. CRISPRoff is an example of such an approach, where dCas9 is fused to several epigenome modifiers, has the potential to induce stable and heritable epigenetic reprogramming through specific DNA methylation and subsequent gene silencing (Nuñez *et al*, 2021). This process can be reversed by CRISPRon, which removes the installed DNA methylation and recruits transcription machinery (Nuñez *et al*, 2021).

Despite the potential of these approaches, their application in an *in vivo* therapeutic setting remains to be assessed.

3.7.3 Single-Base Editing

Most common genetic variants that are associated to human diseases are single nucleotide substitutions. Therefore, the ability to precisely edit single-bases is highly relevant. Editing of single bases can be achieved by conventional CRISPR-Cas9 genome editing, providing the cells with a repair template that allows the HR pathway to introduce the desired alteration. However, as explored later (section 5.2), genome editing by homology-directed repair (HDR) can be a challenging approach, as its efficiency remains low. Additionally, this

strategy relies on the generation of a DSB, which besides creating toxicity to the cell, also engages end-joining pathways leading to error-prone repair and subsequent undesired indels.

Single-base CRISPR-mediated genome editing has been developed utilising nickase (nCas9), for targeting enzymes known to alter DNA bases to specific regions of the genome, in a DSB-independent manner (**Figure 7**). The first generation of base editors consisted of fusing nCas9 with cytosine deaminases, such as APOBEC1 (Apolipoprotein B-editing enzyme, catalytic polypeptide 1) (Komor *et al*, 2016; Nishida *et al*, 2016) (**Figure 7A**). APOBEC1 is an RNA-targeting enzyme that deaminates cytosines in a diverse set of pathways (Rosenberg *et al*, 2011). Cytosine (C) deamination, mediated by APOBEC1, converts it into uracil (U). The resulting G:U heteroduplex activates the BER pathway which, through the activity of a uracil-DNA glycosylase (UDG), excises the erroneous nucleotide and restores the DNA to the initial G:C duplex (**Figure 7B**). This endogenous DNA repair process antagonises the desired outcome of base editing. Therefore, a uracil glycosylase inhibitor (UGI) can be fused to nCas9 to inhibit BER and increase the efficiency of base-editing. Upon BER inhibition, the U:G mismatch is instead recognised by the MMR pathway, which utilises the nicked strand as a signal of repair (**Figure 7C**). Hence, by using a variant of nCas9 that nicks the non-edited strand containing the G, the U:G mismatch is favourably resolved into U:A, or T:A, reaching the desired C>T (or G>A) substitution. Therefore, the most advanced forms of cytosine-deamination base editors (CBEs) consist of an APOBEC fused to nCas9(H840A) and a UGI (Komor *et al*, 2016).

The single-base editing toolbox was further expanded in 2017, with the development of adenine base editors (ABEs) which can perform A > G (or T > C) targeted conversion (**Figure 7**). The most efficient reported ABE was developed by directed evolution of a tRNA adenine deaminase that accepts DNA as a substrate, called TadA (Gaudelli *et al*, 2017) (**Figure 7D**). TadA converts adenosine (A) to inosine (I), which is read as guanosine (G) by DNA polymerases (**Figure 7E**). Therefore, an A:T heteroduplex gets converted into I:T, that upon DNA repair and replication gets converted into G:C. As with CBEs, ABEs consist of a fusion between the effector enzyme (TadA) and nCas9 for DNA targeting.

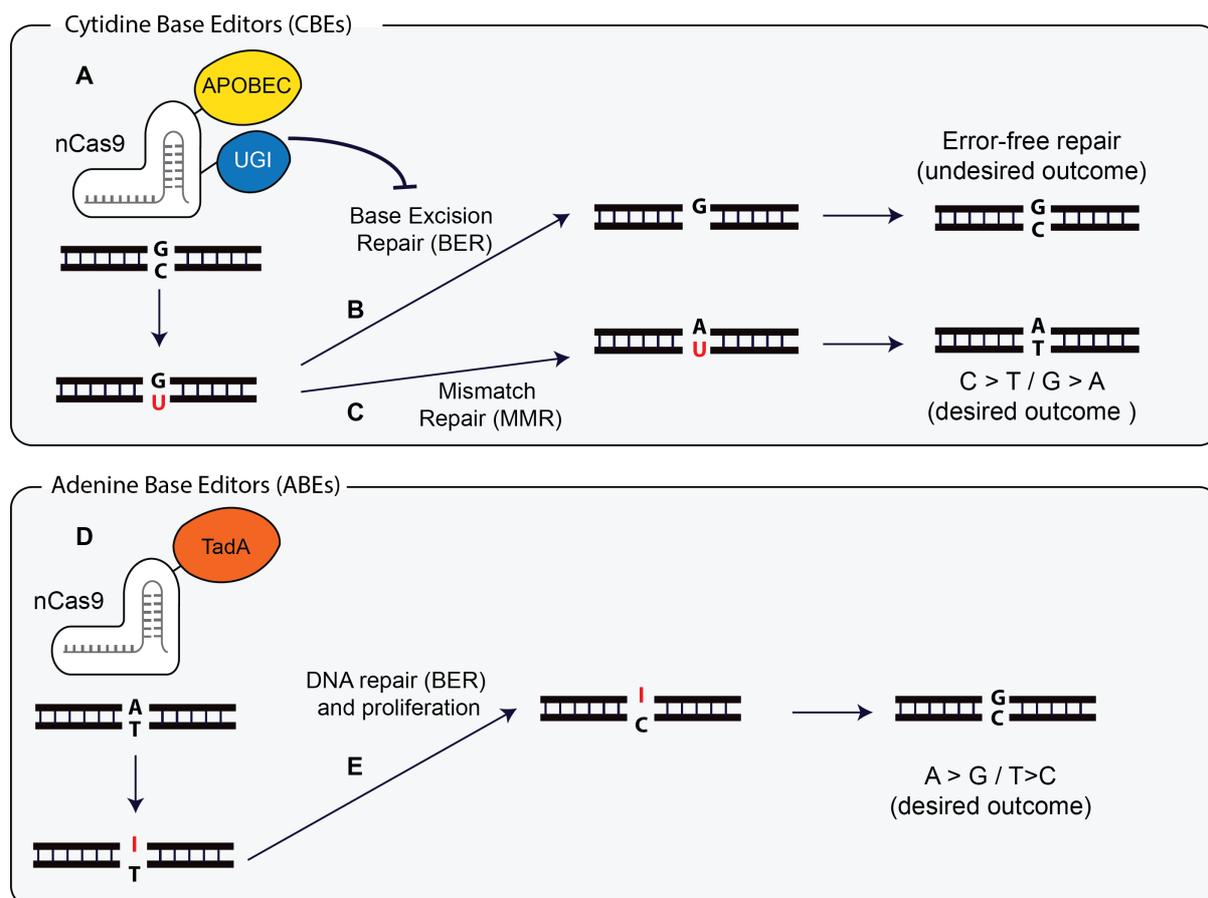


Figure 7: Base editing strategies. **A)** Cytidine base editors (CBEs) work through a fusion of nickase Cas9 (nCas9) with the enzyme APOBEC and a uracil glycosylase inhibitor (UGI). APOBEC deaminates cytosines (C), converting them into uracils (U). **B)** The Base Excision Repair (BER) pathway deals with this heteroduplex by removing the U and replacing it back to C. This error-free repair is not the desired outcome of a base editing strategy. BER can be inhibited by the UGI. **C)** In that case, the Mismatch Repair (MMR) pathway takes over. Since this pathway is strand-biased, it preferentially repairs the nicked strand containing the non-edited guanine (G), converting it to adenine (A), so that is pairs with U. The A:U heterodimer is then converted into A:T during replication, the desired outcome of CBEs. **D)** Adenine base editors (ABEs) are based on the adenine deaminase TadA that converts A into inosine (I). **E)** Inosine is read by the cellular machinery of repair and proliferation as a G, leading to the insertion of a C opposite to it. The I:C heteroduplex is then converted into G:C, the desired outcome of this base editing strategy.

An important feature of base editors is the fact that they act only on ssDNA, and not on dsDNA. This is critical to restrict the enzymatic activity of the base modifier to a small window of nucleotides within the ssDNA-loop created by Cas9. However, several studies reported off-target effects derived from base editors activity, often independently of the DNA:sgRNA interaction (Kim *et al*, 2017). For this reason, further studies are still warranted to study and decrease the off-target effects generated by base editors.

BEs can also be used as an alternative approach to create loss-of-function variants, without the generation of potentially deleterious DSBs. A method called 'CRISPR-STOP' has been developed using CRISPR base editors to efficiently create early stop codons by single base substitutions (Kuscu *et al*, 2017). Moreover, base editors have been successfully utilised

in pooled screens for the functional interrogation of human single-nucleotide variants (Cuella-Martin *et al*, 2021; Hanna *et al*, 2021).

3.7.4 CRISPR-Mediated RNA Editing

Cas orthologs that bind to RNA instead of DNA have been exploited for RNA editing. So far, Cas13 has been in the centre of most RNA editing applications, but a recently described Cas ortholog (Cas7-11) offers a better programmability, which can be harnessed for precise editing of RNA (Özcan *et al*, 2021).

Cas13 cuts uracil bases anywhere in the proximity of the target site. This ability to promiscuously cleave RNA hampers its application for precise editing, but created the basis for the development of an RNA detection platform termed 'SHERLOCK' (Specific High-Sensitivity Enzymatic Reporter Unlocking) (Gootenberg *et al*, 2017). Here, Cas13 activation and subsequent RNA-cleavage releases a reporter signal, which has been applied as a diagnostic test to detect viral RNA (Gootenberg *et al*, 2017). During the COVID-19 pandemic, SHERLOCK has been used as a diagnostic method for the detection of the most recent coronavirus (Zhang F. 2020 v.20200321). This technology was granted emergency-used approval, becoming the first CRISPR-application to be approved by the United States Food and Drug Administration (FDA).

RNA editing offers several applications, such as the study of RNA-protein interactions, the modulation of transcripts, or the visualization of RNA trafficking and localisation with fluorescently-tagged Cas13. Nonetheless, the application of RNA editing to treat disease still remains to be explored.

3.7.5 Prime-Editing

Prime Editing is the most recent promising approach for precise genome editing, using an RNA template for gene alteration (Anzalone *et al*, 2019). Prime editing is considered a 'search-and-replace' tool with the potential to mediate targeted insertions, deletions and all possible combinations of single-base conversions, without DSB generation or the requirement for a donor DNA template. It is therefore considered to be an extremely versatile technology.

Prime editing, similarly to base editors, works with a fusion of a nCas9 (H840A) to an effector protein, this time being a reverse transcriptase (RT) (**Figure 8A**). RT is an enzyme that uses RNA as a template to generate complementary dsDNA, in a process called reverse transcription. RT enzymes can be found for example in retroviruses, to replicate their genomes, but also by eukaryotic cells for telomere extension. During reverse transcription, RT

uses an RNA template and a short primer complementary to the 3'-end of the RNA, to direct the synthesis of cDNA.

Besides the nCas9-RT fusion, prime editing requires a prime editing guide RNA (pegRNA). PegRNAs, similarly to sgRNAs, conserve the ability to recruit nCas9-RT to a particular DNA target site, but they are particular in the sense that pegRNAs encode, in their 3'-extension, new genetic information to be introduced to the target DNA (**Figure 8B**).

After nicking of the genomic DNA by nCas9, reverse transcription, mediated by the RT enzyme, is initiated at the exposed 3'-hydroxyl group (**Figure 8C**). The 3'-extension of the pegRNA is utilised as template to synthesise new DNA, leading to the introduction of the desired sequence (**Figure 8D**). The result of these initial steps is a DNA structure with two redundant ssDNA flaps, a 5' flap containing the original DNA sequence and a 3' flap containing the edited sequence copied from the pegRNA (**Figure 8E**). What follows is a process of flap equilibration that is still not fully understood. 5' flaps are favourably cleaved by structure-specific endonucleases, such as FEN1, a factor of the BER pathway. The preferential excision of the 5' flap leads to integration of the edited 3' flap, driving the introduction of the desired edit. The result of this integration is a heteroduplex DNA containing an edited and a non-edited strand (**Figure 8F**).

The resolution of this heteroduplex is the basis of the development of more advanced prime editing strategies (PE3 and PE3b). In these strategies, an additional nick is induced in the non-edited strand, directing DNA repair to use the edited strand as a template. MMR has been speculated as the pathway responsible for this repair (Scholefield & Harrison, 2021). This nick is induced by the same nCas9(H840A)-RT fusion, this time coupled with a regular sgRNA (instead of a pegRNA) and leads to a fully edited dsDNA (**Figure 9**). PE3b differs from PE3 since the nicking of the non-edited strand only happens after the resolution of the edited strand. This is achieved through the design of the second sgRNA with a spacer that matches the edited strand, hence guaranteeing that the second nick only occurs after editing of the first strand. Compared to PE3, PE3b considerably reduces the occurrence of indels by 13-fold, which frequently arise from simultaneous nicks (Anzalone *et al*, 2019).

Prime editing has been described as remarkably versatile, capable of installing a broad range of alterations with high efficiency and low off-target effects. In humans, prime editing has the potential to correct up to 89% of the known pathogenic human genetic variants (Anzalone *et al*, 2019). Moreover, by potentially being cell-cycle-independent (contrary to HDR), one of the biggest advantages offered by this technology is the potential to edit postmitotic cells, such as neurons. Prime editing has been broadly applied to genome editing studies in multiple organisms, such as rice and wheat (Lin *et al*, 2020), zebrafish (Petri *et al*, 2021), mice (Liu *et al*, 2020), human stem cells (Sürün *et al*, 2020) and patient-derived organoids (Schene *et al*, 2020). Despite its versatility, the efficiency of prime editing widely

varies across cell types, edits and target loci (Anzalone *et al*, 2019) and further studies are still warranted to determine its application in the clinics.

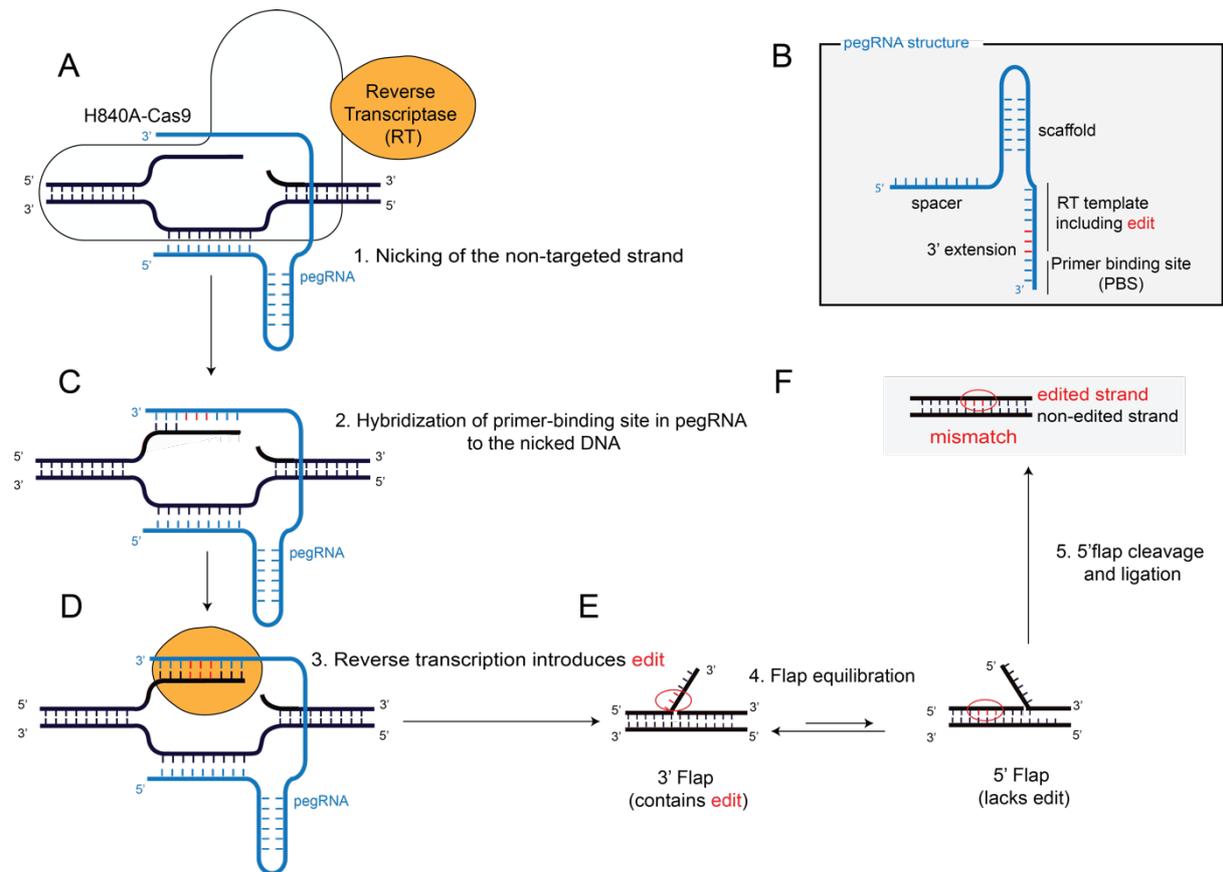


Figure 8: Schematics of prime editing. **A)** Prime editing works via a nickase Cas9 (H840A-nCas9) that targets the non-targeted strand, coupled with a reverse transcriptase enzyme (RT). **B)** The nCas9-RT complex is directed to the genomic DNA through the prime editing sgRNA (pegRNA), that besides the spacer and scaffold sequences, also contains a 3' extension that includes a primer binding site (PBS) for the RT and a template, comprising the desired alteration to introduce in the genome (depicted in red). **C)** The pegRNA anneals to the genomic DNA complementary to the spacer sequence, directing the nick of the non-target strand, which hybridises with the PBS at the 3' extension of the pegRNA. **D)** RT uses the pegRNA 3' extension as template to synthesize DNA, introducing the desired alteration into the genomic DNA. **E)** The process of flap equilibration leads to installation of the edit after 5' flap removal and 3' flap ligation. **F)** An heteroduplex DNA containing an edited a non-edited strand is the final result.

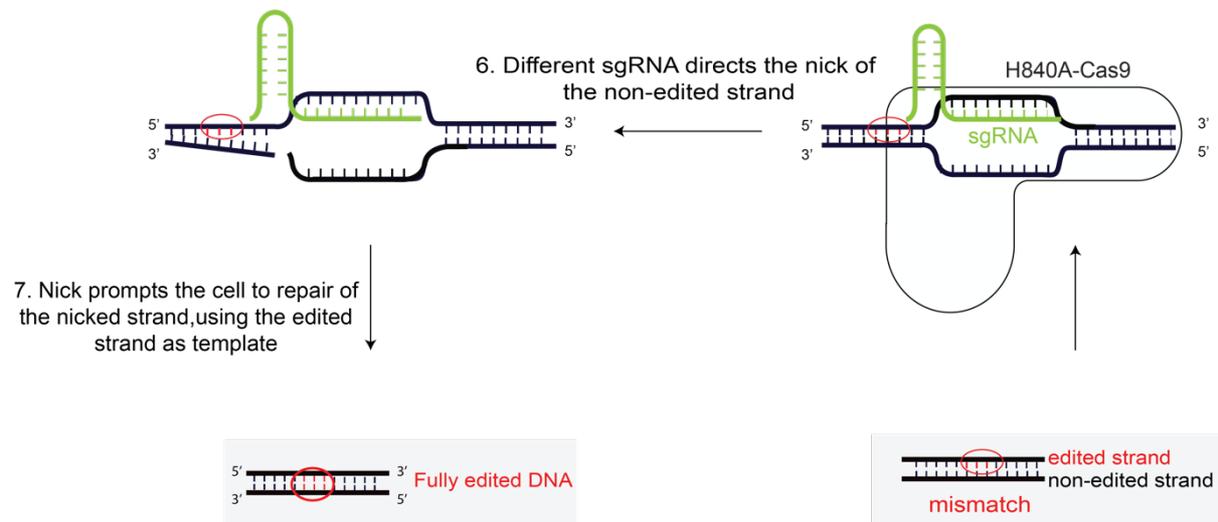


Figure 9: Schematics of PE3 and PE3b. The heteroduplex DNA that results from PE is resolved by the nicking of the non-edited strand (using a sgRNA), which prompts the cell to repair this break using the edited strand as template.

4. Genome Editing to Treat Human Disease

4.1 The Promise: Therapeutic Scope for CRISPR Editing

Despite the complexity inherent to the genetics of human disease, some of the most common genetic disorders are caused by specific mutations in particular genes, or regions of the genome. Sickle cell disease and muscular dystrophy are examples of two common human diseases that can be treated or cured by genome editing in a near future.

Sickle cell disease occurs in individuals who have two defective copies of the *HBB* gene, encoding β -globin. In these patients, a single A-to-T mutation results in a glutamate-to-valine substitution in β -globin, causing the defective protein to form chain-like polymers of haemoglobin, inducing a sickled shape in red blood cells that leads to obstructed blood vessels and, eventually, organ failure. Bone marrow transplantation can be used to treat this disease, but it requires the use of cells from a suitable donor. Gene editing solves this restriction, since it can be achieved by removing hematopoietic stem cells from the patient to correct the disease-causing mutation *ex vivo*. The corrected edited stem cells could then be transplanted back into the patient to produce healthy red blood cells.

Overall, blood disorders are particular good candidates for genome editing, since cells are easy to extract. In fact, several genome editing studies have shown promising outcomes for the correction of pathogenic mutations in blood disorders (Wu et al, 2019; Baik et al, 2019). Moreover, clinical trials, developed for sickle cell disease and β -thalassaemia have used CRISPR-Cas9 to engineer hematopoietic stem cells to boost the production of foetal

haemoglobin, a type of haemoglobin that is present at birth and is afterwards replaced by the adult form. This was achieved by disabling the *BCL11A* gene, which encodes a transcription factor that otherwise represses foetal haemoglobin synthesis. Results from this clinical trial have shown that two patients were cured following this therapy, building excitement around somatic targeted CRISPR-Cas9-based therapy (Zipkin, 2019; Frangoul *et al*, 2021).

Moving away from blood disorders, most genetic diseases would require the correction of the causing mutation in the tissue. Duchenne muscular dystrophy (DMD) is an example of such disease, causing weakening of skeletal muscles over time. DMD arises due to a frameshift mutation in exon 51 of the gene that encodes dystrophin, a protein necessary to maintain the integrity of striated muscles. Importantly, the restoration of a small percentage of normal dystrophin expression is sufficient to provide a beneficial clinical outcome, making DMD a good candidate disease for genome editing. Several studies have reported the *in vivo* restoration of the open-reading frame of *DMD*, and the synthesis of a partially functional dystrophin, following adeno-associated virus (AAV)-delivery of CRISPR-Cas9 components (Nelson *et al*, 2016; Long *et al*, 2016; Tabebordbar *et al*, 2016; Amoasii *et al*, 2018).

Moreover, therapeutic genome editing approaches are currently under development for genetic ocular diseases, such as retinitis pigmentosa, which causes progressive retinal degeneration and may result in blindness. In mice, CRISPR-Cas9 mediated knockdown of the *Nrl* (Neural retina-specific leucine zipper) gene, responsible for specifying rod cell fate during retinal development, led to loss of rod features and preservation of cone-like features in three distinct mouse models of retinal degradation (Yu *et al*, 2017).

Besides the correction of genetic diseases, genome editing has also achieved promising results for cancer therapeutics, such as via the enhancement of engineered autologous T-cells (via chimeric antigen receptor (CAR)-T-cells), for T-cell immunotherapy (Eyquem *et al*, 2017). Additionally, CRISPR-Cas9 has been employed to target programmed cell death protein 1 (PD1), blocking inhibitory signals that prevent tumour recognition by the immune system (Fellmann *et al*, 2017).

Multiple *in vivo* studies have indicated that base editing is a promising approach for therapeutic genome editing. Base editors are advantageous over CRISPR approaches that rely on SpCas9, since DSBs are not generated. An example of an *in vivo* study utilising base editors for therapeutic genome editing was the generation of a loss-of-function variant in the *PCSK9* gene in cynomolgus monkeys. The animals receiving the base-editing machinery showed reduced LDL cholesterol, indicating a viable approach for the treatment for hypercholesteremia (Musunuru *et al*, 2021). Additionally, base editing has been successfully used for the correction sickle cell disease and Hutchinson Gilford progeria, in mice (Newby *et al*, 2021; Koblan *et al*, 2021b).

Several interventional clinical trials involving CRISPR-Cas9-based gene editing are currently on-going, all over the world (Wang *et al*, 2020). However, despite the tremendous potential of genome-editing applications for curing human disease and the considerable advancement in the field, further studies are still necessary before the definitive introduction of this technology in the clinics. Preclinical results are promising, but safety still needs to be addressed. Moreover, for each disease, important considerations need to be made, as the underlying mutation needs to be matched with the best gene editing method, cells or tissue delivery and the extend of gene correction that would have a therapeutic value (Ferreira da Silva *et al*, 2021). Being a technology that is heavily dependent on the endogenous DNA repair machinery of the cell, DNA repair impairment might compromise the amenability of certain diseases to be treated by CRISPR-Cas9 genome editing, a knowledge that would be essential for its implementation in the clinic.

4.2 The Challenge: Obstacles to Overcome

4.2.1 Delivery

Delivery remains one of the biggest challenges for somatic genome editing, motivating the continuous emergence of new strategies to improve it. The currently favoured form of *ex vivo* or *in vitro* delivery of CRISPR-Cas9 components to primary cells is electroporation of Cas9 as a pre-formed complex with sgRNA, in the form of a protein-RNA complex (ribonucleoprotein, RNP). *In vivo* delivery, however, is more challenging and it usually requires viral vectors, typically adeno-associated virus (AAVs), or lipid nanoparticles bearing mRNA Cas9 and a sgRNA.

Viral delivery methods have the potential to target a broad range of tissues in an efficient manner. However, viral vectors offer limited cargo size. The maximum size for a transgene cassette to be inserted into an AAV is approximately 4.7 kb, which is not a lot considering that *S. pyogenes* Cas9 alone is 4.2 kb long (Wang *et al*, 2020). Moreover, additional viral vectors are necessary to express the sgRNA and a template sequence for homology-directed repair (HDR). This reduces the efficiency of editing, as both viral vectors need to be acquired simultaneously (Yang *et al*, 2016; Lau & Suh, 2017). The identification of smaller orthologs of Cas9, more compatible with AAV delivery, can be an approach to overcome the cargo limitation (Wang *et al*, 2020). Another caveat associated with viral-delivery is the long-term exposure of genome-editing factors, which may increase the exposure of patients to off-target effects, or immune reactions (Doudna, 2020).

Nanoparticles are an alternative approach to viral-delivery of CRISPR-Cas9 components and they have been shown to be efficient in a variety of studies (Zuris *et al*, 2015; Cheng *et al*, 2020). In contrast to viral delivery, nanoparticles do not lead to genomic

integration and offer low immunogenicity. However, disadvantages of this approach include high toxicity and limited tissue-tropism.

Electroporation has been widely used to deliver CRISPR-Cas9 components *ex vivo*, but it has also been successful for delivery to animal zygotes (Qin & Wang, 2019) and to introduce CRISPR-Cas9 constructs directly into skeletal muscle of mice with DMD. However, this strategy is not easy to implement *in vivo*, since it requires specific reagents and engineering.

4.2.2 Immunogenicity

Another factor that affects the implementation of CRISPR-Cas9 in the clinics is the potential immune response that might be triggered by the delivery of editing components (such as Cas9), which are derived from bacteria. This can lead to inflammation and compromised genomic stability. Cas9 antibodies and reactive antibody T-cells have been detected in humans exposed to bacteria containing CRISPR systems (Wagner *et al*, 2019). However, it is not clear if these detected levels would be enough to trigger an immune response against CRISPR editing factors. These caveats can be overcome by transient methods of delivery in an *ex vivo* setting, since the natural decay of Cas9 in edited cells would minimize Cas9 exposure.

4.2.3 Off-Target Effects

Off-target effects occur when CRISPR-Cas9-induced DSB and subsequent repair occurs at genomic sites that are not intended for modification. This usually happens at sites with a similar sequence to the target site, as it has been described that there is a tolerance for sgRNA mismatches in the binding of Cas9 (Hsu *et al*, 2013).

Several strategies to detect and mitigate off-target effects have been developed over the years (Tsai *et al*, 2015; Wienert *et al*, 2019). First, off-target effects can be minimised by carefully selecting target sequences that lack homology with other regions of the genome (Cho *et al*, 2014). Reduced off-target effects can also be achieved by the generation of paired nicks (by nCas9) instead of a single DSB generated by Cas9 (Cho *et al*, 2014). This creates the additional requirement of having two target sites close together for the generation of the DSB, adding an additional layer of control. Delivery also seems to influence off-target effects, with direct delivery of the RNP complex Cas9-sgRNA showing reduced off-target effects compared to viral delivery methods (Kim *et al*, 2014). Finally, 'high-fidelity' variants of Cas9 have been engineered to reduce nonspecific interactions, resulting in increased DNA specificity (Slaymaker *et al*, 2016; Kleinstiver *et al*, 2016).

4.2.4 Precision

Imprecise genome editing happens when the editing outcome is not as desired, even though it occurred in the correct genomic location. This is usually the result of different mechanisms of DNA repair functioning after a Cas9-induced DSB. Moreover, large deletions and complex rearrangements have been described to occur in a diversity of cellular models (Kosicki *et al*, 2018). Even though these are rare events, they are relevant in a clinical setting, since these translocations might lead to cancer (Maddalo *et al*, 2014; Buechele *et al*, 2015). Understanding the DNA repair mechanisms that act on a Cas9-induced break is therefore essential to reduce or eliminate undesired events, therefore improving CRISPR's precision.

4.2.5 Considerations for Germline Genome Editing

Germline editing differs from somatic genome editing in the sense that it results in genetic changes that are heritable. Even though germline editing has been broadly applied in animals and plants, it is still an important matter of debate in humans. In 2018, during the Second International Summit on Human Genome Editing (held in Hong Kong), a controversial study was presented describing the editing of human embryos that had resulted in the birth of twin girls, in China. This presentation triggered widespread discussions on human germline editing and its ethical and scientific regulations, raising important questions regarding the application of the CRISPR-Cas9 technology in human embryos. Several consensual points aimed at regulating the application of genome editing technologies for germline editing were drafted (Doudna, 2020):

- i. It is considered inappropriate to perform germline genome editing that culminates in human pregnancy;
- ii. *In vitro* germline genome editing in humans (both embryos and gametes) should be allowed and there should be no prohibition on public funding for this type of research, provided that there is appropriate informed consent from the donors;
- iii. Future clinical applications of human germline genome editing should not proceed, unless there is (1) a compelling medical reason, (2) evidence supporting its clinical use, (3) an ethical justification and (4) a transparent public process to solicit input.

Finally, a strong point raised during this discussion was the fact that there is not enough knowledge about the DNA repair mechanisms and the development pathways that operate in early human embryos to predict editing outcome with certainty. Cell cycle control in human embryos appears to differ greatly from what is observed in somatic cells (Bazrgar *et al*, 2014) and the mechanisms by which DSBs are repaired in embryos are still under debate (A. Lea & K. Niakan, 2019; Zuccaro *et al*, 2020). Further research is necessary to resolve these

remaining questions, once again highlighting the importance of understanding the DNA repair mechanisms for a safe implementation of the CRISPR-Cas9 technology.

5. DNA Damage Response: The Foundation of Genome Editing

CRISPR-Cas9-mediated genome editing is a technology that mostly relies on the DDR (**Figure 10A**). This is true for the conventional repair of Cas9-mediated DSBs - which will be the focus of this chapter - but also for emergent technologies, such as base editing or prime editing. The ultimate goal of genome editing is therefore a safe steering of the DDR-decision process: from lesion to desired outcome, with very few side-products. Despite the longstanding study of the DDR in model organisms such as yeast, its control can only be achieved by a deep understanding about the repair mechanisms that act in human cells upon CRISPR-Cas9-mediated lesions, a question that still warrants further investigation.

5.1 Template-Free Repair of Cas9-Mediated DSBs

Following Cas9 cleavage, template-free DNA repair is generally considered to be error-prone, leading to small indels (< 10 bp) that culminate in gene disruption (gene KO) (**Figure 10A**). This ability to generate loss-of-function variants has been attributed to the NHEJ pathway, which directly re-ligates DNA ends following cleavage (Bothmer *et al*, 2017). However, several studies have uncovered an important role for the MMEJ pathway in the repair of Cas9-induced DSBs (**Figure 10A**) (van Overbeek *et al*, 2016; Brinkman *et al*, 2018).

Despite the intrinsic error-prone nature of end-joining pathways, the mutational profile generated by a sgRNA was shown to be highly reproducible, predictable and mainly dependent on the targeted DNA sequence (van Overbeek *et al*, 2016; Shou *et al*, 2018). Following these observations, several recent studies have systematically analysed repair products generated by distinct sgRNAs and correlated them with the regions flanking the DSB (Allen *et al*, 2018; Shen *et al*, 2018). This led to the conclusion that end-joining mediated Cas9 editing can be harnessed to achieve a particular desired editing outcome. These studies have uncovered general patterns of repair, showing that most sgRNAs have a single outcome that contributes at least 20% of the observed repair profiles. When a consensus exists, it is almost always a single nucleotide insertion (most likely of the nucleotide that flanks the DSB, distal from the PAM sequence), a microhomology-mediated deletion of at least 3 nucleotides, or a deletion of 1-2 nucleotides (Allen *et al*, 2018).

The ability to predict template-free repair outcomes has highlighted the relevance of the MMEJ pathway, which has been harnessed in different contexts. One strategy makes use of microhomologies between an exogenous DNA donor, to drive precise genomic integration

(Nakade *et al*, 2014). Moreover, alt-EJ has been explored for therapeutic gene correction (Iyer *et al*, 2019), by repairing a DSB induced near the centre of a disease-causing microduplication.

5.2 Homology-Directed Repair (HDR)

HDR is an error-free type of DNA repair that utilises a DNA strand as template (**Figure 10A**). It can be further categorized based on the nature of the nucleic acid acting as template into homologous recombination (HR), if the template is a dsDNA (from a plasmid, or a sister chromatid), or single-stranded templated repair (SSTR), if a ssDNA donor (e.g. a synthetic oligonucleotide) is used as repair template. Due to its meticulousness, HDR is more relevant for therapeutic purposes. Its efficiency, however, remains problematically low, especially in post-mitotic cells like myofibers and neurons. Multiple efforts have been put into improving the efficiency of HDR and many of these are based on the control of the DDR.

5.2.1 NHEJ Inhibition

Because NHEJ is the default pathway to deal with Cas9-breaks, pharmacological inhibition of NHEJ has been explored to increase HDR (**Figure 10B**). DNA-PK inhibitors (NU7441 or NU7026) have been employed to increase HDR (Robert *et al*, 2015; Riesenberger & Maricic, 2018). Additionally the SCR7 small molecule has been shown to increase HDR, by inhibiting LIG4, in human and mouse cells (Srivastava *et al*, 2012; Chu *et al*, 2015; Hu *et al*, 2018b). The utilisation of this compound has, however, been disputed as it has been shown to also target other DNA ligases, such as LIG1 and LIG3, which participate in alt-EJ. The inhibition of 53BP1, for example through the use of engineered ubiquitin-variants (Canny *et al*, 2017) or the expression of a dominant negative form (Paulsen *et al*, 2017), has also shown promising results for the improvement of HDR.

5.2.2 Cell Cycle Control

Several approaches to control cell cycle have been employed to manipulate repair outcomes following a Cas9-mediated break (**Figure 10C**). Fusing Cas9 with CtIP (Charpentier *et al*, 2018), for example, bypasses the requirement for cell-cycle dependent activation of CtIP, which is necessary for HDR to occur. Hence, it allows HDR to happen outside of its permissive cell cycle phases.

Using small-molecules to arrest cells in cell cycle phases in which HDR is more active has also been shown to improve precise editing (Lin *et al*, 2014; Wienert *et al*, 2020). Another cell-cycle based strategy that avoids undesired indels generated by end-joining, is the restriction of Cas9 expression to S/G2 phases of the cell cycle. This can be achieved by fusing

Cas9 to Geminin, a substrate for proteasome degradation during G1 (Lomova *et al*, 2019). Overall, either through bypassing the HDR cell cycle requirements, or by accumulating cells in HDR-permissive cell cycle phases, these approaches can improve precise editing, while minimising undesirable mutagenic outcomes.

5.2.3 Single-Stranded Oligonucleotide (ssODN) Donor Templates

Several studies have shown striking differences in editing efficiency depending on the type of donor DNA used. Single-stranded oligonucleotides (ssODN) promote the highest levels of HDR when used as templates (Yeh *et al*, 2019). Additionally, HDR frequencies can be further improved by rationally designing the orientation, polarity and length of the ssODN donor to match the properties of the DNA-Cas9 complex. Biochemical studies have shown that the Cas9-DNA interaction is asymmetric, indicating that Cas9 locally releases the PAM-distal side of the nontarget strand (Richardson *et al*, 2016). Therefore, increased HDR rates can be achieved by designing ssODNs donors that are complementary to the strand that is released first. (**Figure 10D**)

While HDR mediated by dsDNA donors largely resembles the HR pathway, the repair mechanisms for SSTR are still not fully understood. In humans, depending on the cellular genetic background, SSTR efficiencies vary greatly, reflecting genetic and transcriptional differences that up or down-regulate genome editing by this method (Richardson *et al*, 2018). The FA pathway, for example, has been shown to be essential for SSTR (**Figure 10D**) (Richardson *et al*, 2018). The involvement of a pathway known to repair ICLs in the repair of Cas9-generated DSBs highlights how relevant it is to study the repair mechanisms triggered by Cas9-mediated lesions. Additionally, this finding has important therapeutic implications, as it indicates that FA patients might only be amenable to treatment by CRISPR-Cas9, upon the temporary reactivation of the pathway. Other diseases might have special requirements for genome editing, depending on their underlying genetic defects.

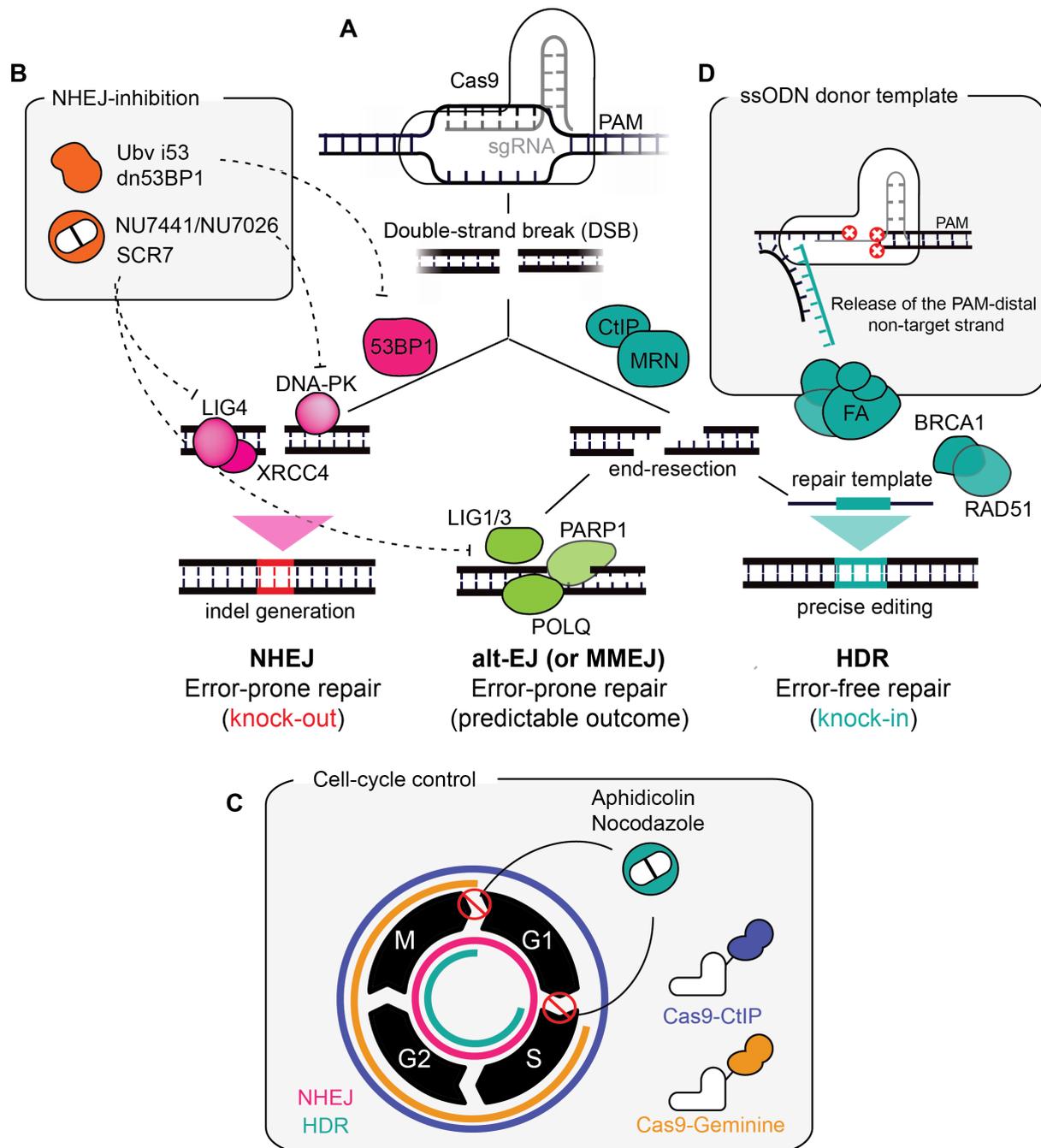


Figure 10: Repair of Cas9-induced DNA double-strand break (DSB) along with approaches to improve fidelity of repair. A) Cas9 cuts DNA to induce a DSB, which induces NHEJ as the major DNA repair pathway. A key mediator that functions upstream in the pathway is 53BP1. Ligation of the DNA ends depends on LIG4 that is complexed with DNA-PK, and XRCC4. This is an error-prone DNA repair pathway that leads to insertions and deletions (indels), cumulating in loss-of-function variants. End-resection would promote alternative end-joining (alt-EJ or MMEJ) or homology-directed repair (HDR) that would depend on the enzymatic activities of CtIP and MRN. Alt-EJ leads to highly predictable outcomes, as it relies on microhomology. HDR is error-free and thus will install the desired edit, provided in the form of the repair template. **B)** NHEJ can be inhibited to increase HDR. 53BP1 can be inhibited with ubiquitin variants (i53) and dominant negative versions of the protein (dn53BP1). Pharmacological inhibition using NU7441 and NU7026 for DNA-PK or SCR7 for LIG4 (NHEJ) and LIG1/LIG3 (alt-EJ) have also been used to improve HDR. **C)** Perturbing cell cycle progression in S phase can also increase HDR since HDR (green) functions in S/G2/M phases, while NHEJ (pink) functions in all phase of the cell cycle. Thus, S-phase blocking compounds such as aphidicolin and nocodazole have been used to increase HDR. Also, a Cas9-CtIP fusion that promotes end-resection increases HDR through the cell cycle as it promotes resection. A Cas9-Geminine fusion that dampens

error-prone repair, functions in this way as it is not expressed in G1. **D)** ssDNA oligonucleotides (ssODN) function as donor templates to increase HDR. If the ssODN is complementary to the strand that is released first, this promotes HDR by exploiting the asymmetry of the Cas9:DNA complex. NHEJ: non-homologous end-joining; MMEJ: microhomology-mediated end-joining; LIG4: DNA ligase IV; DNA-PK: DNA-dependent protein kinase; XRCC4: X-ray cross-complementing factor 4; LIG1: DNA ligase I; LIG3: DNA ligase III; POLQ: DNA polymerase θ ; PARP1: Poly (ADP-ribose) polymerase 1; FA: Fanconi anemia; PAM: protospacer adjacent-motif. Figure adapted from (Meyenberg *et al*, 2021) and re-printed with permission from *Frontiers in Genetics*.

Chapter 2: Aim of this study

The discovery of the CRISPR-Cas9 system and its application for human genome editing has opened new avenues for functional genetic studies, but also therapeutic options for a broad range of human diseases. CRISPR-Cas9 heavily relies on mechanisms of DNA damage repair that must be understood in the context of the technology.

Considering that a patient's genome, or transcriptome, influences editing outcome, it is still not possible to rationalise which genome editing application would be more efficient for a particular disease background. Additionally, emerging technologies for precise genome editing, such as base editors and prime editing, are sought to engage DNA repair pathways that are yet to be explored.

The therapeutic implementation of CRISPR-Cas9 technology deeply relies on a fundamental understanding of the DNA repair mechanisms and pathways that are engaged and lead to specific editing outcomes, as well as the activity of these pathways in particular cell and tissue types. This combined knowledge will ultimately lead to improvements of the current technology, as well as the generation of biomarkers with the potential to correlate disease with therapeutic editing strategy.

This thesis is divided into studying the mutagenic repair of Cas9-induced breaks (Project 1) and precise genome editing by the prime editing technology (Project 2). The aims of these projects are the following:

Project 1:

To understand and characterise the role of NHEJ and alt-EJ in the mutagenic repair of Cas9-induced breaks.

Project 2:

To identify DDR factors and/or pathway dependencies for precise genome-editing mediated by prime editing.

Chapter 3: Results

3.1 Genome-Scale CRISPR Screens are Efficient in Non-Homologous End-Joining Deficient Cells

3.1.1 Prologue

For this project, we sought to study the mutagenic repair of Cas9-induced DSBs, which is frequently attributed to the action of the NHEJ repair pathway. Here, by using genetic human cellular models of NHEJ deficiency, as well as genome-scale approaches, we found that NHEJ is dispensable for the repair of Cas9-induced breaks and it can be fully compensated by alt-EJ, in a POLQ dependent manner. The repair of Cas9-induced breaks by alt-EJ gives rise to a distinct repair signature, characterised by larger deletions, compared to the profile generated by NHEJ. Additionally, we showed that cells that are deficient for both NHEJ and alt-EJ are still able to repair Cas9-mediated DSBs, indicating the existence of an additional mechanism able to deal with these lesions.

This manuscript has been featured in:

Ferreira da Silva *et al*, **Genome-scale CRISPR screens are efficient in non-homologous end-joining deficient cells**, *Sci Reports* (2019)

3.1.2 PDF of the article

OPEN

Genome-scale CRISPR screens are efficient in non-homologous end-joining deficient cells

Joana Ferreira da Silva, Sejla Salic, Marc Wiedner, Paul Datlinger, Patrick Essletzbichler, Alexander Hanzl, Giulio Superti-Furga , Christoph Bock , Georg Winter & Joanna I. Loizou*

The mutagenic repair of Cas9 generated breaks is thought to predominantly rely on non-homologous end-joining (NHEJ), leading to insertions and deletions within DNA that culminate in gene knock-out (KO). In this study, by taking focused as well as genome-wide approaches, we show that this pathway is dispensable for the repair of such lesions. Genetic ablation of NHEJ is fully compensated for by alternative end joining (alt-EJ), in a POLQ-dependent manner, resulting in a distinct repair signature with larger deletions that may be exploited for large-scale genome editing. Moreover, we show that cells deficient for both NHEJ and alt-EJ were still able to repair CRISPR-mediated DNA double-strand breaks, highlighting how little is yet known about the mechanisms of CRISPR-based genome editing.

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) - Cas9 -mediated gene editing has become a powerful approach for efficient genome editing in eukaryotic cells, where it is used to either generate loss-of-function alleles or introduce precise alterations¹⁻³. The protein Cas9, together with an engineered single-guide RNA (sgRNA), forms a complex that directs the cleavage of a specific locus, by introducing a DNA double-strand break (DSB) at the DNA sequence complementary to the 23 bp protospacer-PAM (5'-NGG protospacer adjacent motif) sequence⁴⁻⁶. In human cells, DSBs are mostly repaired by the error-prone non-homologous end-joining (NHEJ) pathway that induces insertions and deletions (indels), hence disrupting gene function. In contrast, the less efficient homology directed repair (HDR) pathway makes use of a provided DNA template hence allowing for the generation of desired alterations⁷⁻⁹.

Despite the widespread use of CRISPR-Cas9 genome editing, there is still a lack of understanding about the DNA repair pathways that resolve Cas9-mediated cleavage. Supported by studies based on pharmacologic inhibition, it is widely accepted that NHEJ is the major DNA repair pathway that deals with Cas9 lesions¹⁰⁻¹². However, confounders such as incomplete inhibition, off-target effects and dominant-negative patterns can skew the results of such studies, prompting us to develop genetic tools to investigate the mutagenic repair of Cas9 generated DNA breaks, using isogenic cell line models fully deficient in NHEJ. Surprisingly, our results show that NHEJ is dispensable for the repair of Cas9-induced breaks both at specific loci and using genome-scale CRISPR approaches. Moreover, we observed a differential indel signature with larger deletions in the absence of NHEJ, as well as residual editing in cells deficient for both NHEJ and alt-EJ, suggesting the existence of an alternative mechanism for the repair of Cas9-generated breaks.

Results

In order to address the NHEJ dependency of mutagenic repair of Cas9-breaks, a NHEJ-deficient cell line was generated in the human HAP1 cell line, by knocking-out DNA Ligase IV (LIG4), an essential factor for the ligation of the two DNA ends¹³ (Supplementary Fig. 1A). In line with the function of NHEJ, Δ LIG4 cells were hypersensitive to the DNA DSB-inducing agents neocarzinostatin (NCS), doxorubicin and etoposide¹⁴, but not to the alkylating agent methyl methanesulfonate (MMS), providing a specific phenotypic confirmation of NHEJ abrogation in this cell line (Supplementary Fig. 1B). To investigate the role of LIG4 in the repair of Cas9 generated breaks, we developed a cellular assay to measure the kinetics of genomic disruption (Fig. 1A). This consisted of expressing GFP tagged doxycycline-inducible Cas9¹⁵, together with a construct expressing mCherry with a sgRNA targeting the mCherry site required for fluorescence. To ensure rapid turnover of the mCherry protein, its sequence was modified to consist of a PEST sequence, hence reducing its intracellular half-life¹⁶. As confirmed by immunoblotting, Cas9 expression was achieved 24 hours after doxycycline treatment (Fig. 1B and Supplementary Fig. 1C),

CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Lazarettgasse 14, AKH BT 25.3, 1090, Vienna, Austria. *email: jloizou@cemm.oeaw.ac.at

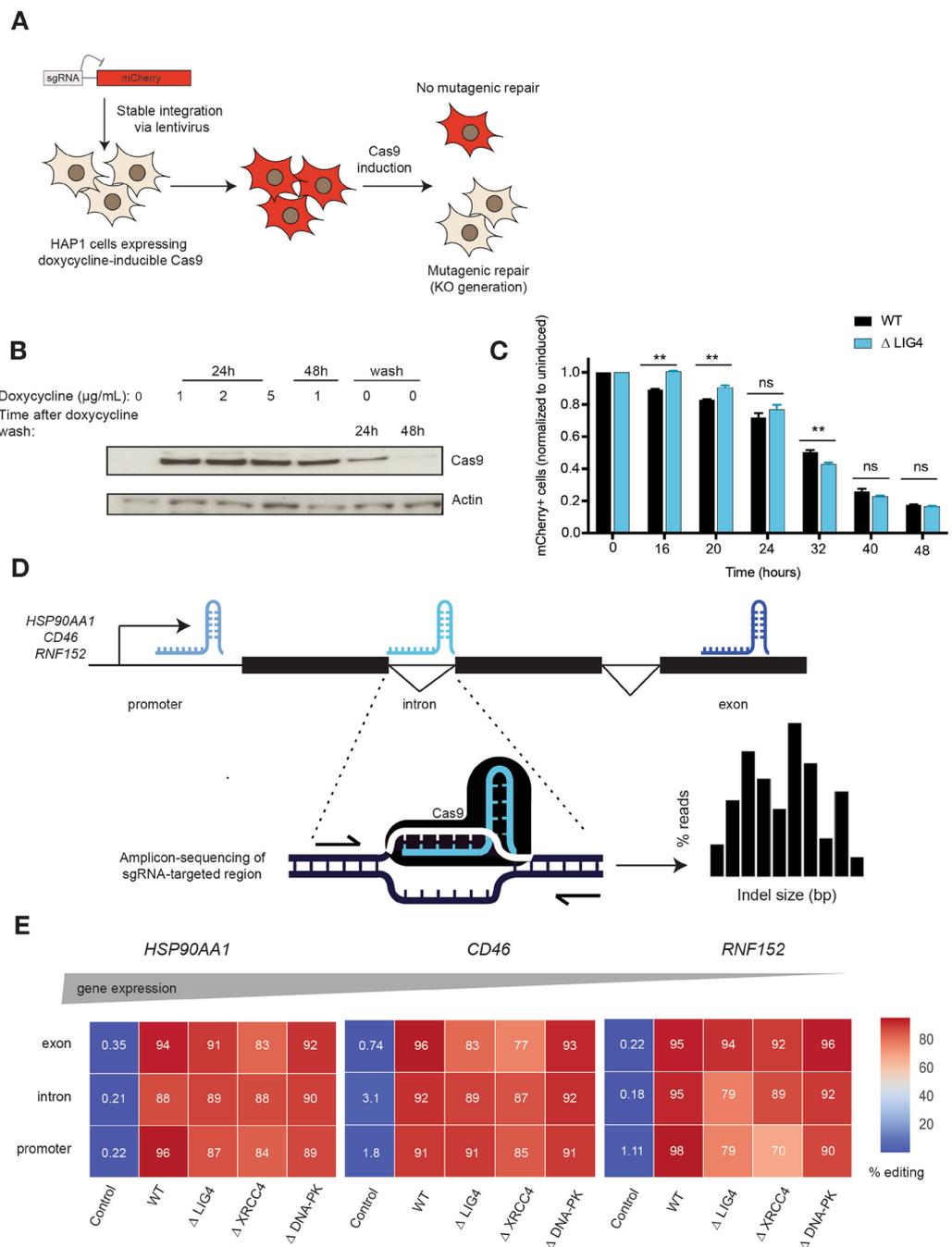


Figure 1. Mutagenic repair of CRISPR-Cas9-mediated DNA breaks is efficient in the absence of non-homologous end-joining. **(A)** Scheme of the cellular assay to determine the kinetics of indel generation, within the mCherry site required for fluorescence. Cells were transduced with a doxycycline-inducible Cas9-GFP and a mCherry plasmid, coupled with a sgRNA targeting the mCherry fluorescence site. Following Cas9-induction, the loss of mCherry fluorescence was used as a readout of mutagenic repair. **(B)** Immunoblot for Cas9 and β -actin in HAP1 cells expressing doxycycline-inducible Cas9 tagged with GFP, with or without doxycycline treatment, as indicated. Figure represents cropped parts of the same gel (entire gel can be found in Supplementary Fig. 1C). **(C)** Kinetics of indel generation within the mCherry locus (measured by gating on GFP-positive cells) after Cas9-induction with doxycycline, at the indicated time points. Each time point was normalized to the uninduced (0 h) time point. The assay was performed in WT and Δ LIG4 HAP1 cells ($n = 3$). Statistical significance was calculated by Student's t -test. ns = not significant, $**p$ -value ≤ 0.01 . **(D)** Scheme of the cellular assay used to measure Cas9-induced indel formation in differentially expressed genes (*HSP90AA1*, *CD46* and *RNF152*) within different genomic regions (promoters, exons and introns). Cells were transfected with Cas9 and the respective sgRNAs following which the targeted regions were PCR-amplified. Amplicon sequencing was used to determine the efficiency of editing, as well as the distribution of indel profiles. **(E)** Percentage of edited reads following Cas9 activity at promoters, exons and introns within *HSP90AA1*, *CD46* and *RNF152* in wild-type (WT) cells and knock-out cells for the NHEJ factors LIG4, XRCC4 and DNA-PKcs (Δ LIG4, Δ XRCC4 and Δ DNA-PK, respectively).

generating a DSB within the mCherry sequence that was subsequently repaired in an error-prone manner, leading to loss of fluorescence (Supplementary Fig. 1D). This system was used to assess error-prone repair leading to indel generation in wild-type (WT) and Δ LIG4 cells, of which the latter lack functional NHEJ. These results unexpectedly revealed that mutagenic repair occurs with equal efficiency in NHEJ abrogated cells as in WT cells, with 50–60% editing at 32 hours and 80% editing at 48 hours after Cas9 induction (Fig. 1C).

Following the observation that NHEJ is dispensable for Cas9-mediated editing of an exogenous locus, we designed a strategy to assess editing of endogenous loci, by testing different genomic regions (promoters, introns and exons) across genes selected to range in expression levels in the human HAP1 cell line¹⁷ (*HSP90AA1*, *CD46* and *RNF152*) (Fig. 1D). Upon genomic amplification of the edited region, sequencing was used to determine editing efficiency and frequency of indel size generated in the targeted loci (Fig. 1D). Moreover, we extended our investigations to include other NHEJ genes by knocking out the core component X-ray repair cross-complementing protein 4 (XRCC4) and the signaling kinase DNA-PKcs (Δ XRCC4 and Δ DNA-PK, respectively) (Supplementary Fig. 1E). We phenotypically confirmed that these cell lines were defective in NHEJ, by assessing their hypersensitivity to DSB-inducing agents (Supplementary Fig. 1F). Although all NHEJ deficient cell lines were exquisitely sensitive to the tested DNA DSB-inducing agents, amplicon sequencing of the Cas9 targeted regions revealed that editing was comparable to WT cells, ranging from 70–98% across all genomic regions tested, regardless of gene expression (Fig. 1E).

So as not to limit our investigations to a single locus, we next performed genome-wide loss-of-function CRISPR-Cas9 screens, using the GeCKO v2.0 library that targets 19,052 genes with 122,417 sgRNAs^{18,19} in both WT and Δ LIG4 cells. This library allows the generation of functional null alleles at endogenous loci, in a highly multiplexed fashion, via comparative measurements of drop-outs of sgRNAs targeting 683 genes that were recently shown to be pan-essential²⁰. Thus, if NHEJ would be required for CRISPR-Cas9 mediated disruption, we would expect for LIG4 deficiency to prevent the identification of essential genes. To allow for depletion of sgRNAs targeting essential genes, we analyzed sgRNA representation 20 days after puromycin selection, in both WT and LIG4 deficient backgrounds (Fig. 2A). Next-generation sequencing (NGS) was used to assess sgRNA abundance (Supplementary Table S1) and the fold-change of each gene was calculated by comparing to the sequenced library for three biological replicates (Supplementary Fig. 2A, Supplementary Table S2). This led to a Spearman's correlation of 0.66 between gene enrichment in WT and Δ LIG4 cells (Fig. 2B). Importantly, genes annotated as core essential²⁰ were depleted similarly in both WT and Δ LIG4 cells (Fig. 2B,C and Supplementary Fig. 2B). Furthermore, screens performed in both genetic backgrounds distinguish essential and non-essential²¹ genes with equal efficiency (Fig. 2D). Gene ontology (GO) analysis of core essential genes²⁰ (Supplementary Fig. 2C) revealed an enrichment of essential fundamental molecular processes, including the constitution of ribosomes, rRNA binding and purine NTP-dependent helicase activity. Genes annotated for the top three enriched GO terms of the 'essentialome' were found depleted in both WT and Δ LIG4 cells, with a high intersection between the genetic backgrounds (Fig. 2E). In summary, comparative identification of core essential genes using an unbiased, genome-wide approach revealed that mutagenic repair of Cas9-generated breaks can be efficiently achieved in the absence of NHEJ.

It is well documented that different sgRNAs lead to specific indel outcomes, displaying a single predominant repair outcome^{11,12,22}. Following this observation, and since these predictions have important applications for template-free genome editing²³, we sought to determine whether indel signatures would be altered in the absence of NHEJ. Besides providing the possibility of manipulating the predicted outcome of a sgRNA, this approach additionally has the potential to reveal which pathway compensates for NHEJ in the mutagenic repair of Cas9-breaks. By investigating the spectrum of indels generated upon exon targeting of three distinct genes (*HSP90AA1*, *CD46* and *RNF152*), we observed a striking increase in the frequency of larger deletions in all three NHEJ deficient cell lines, in comparison to WT cells (Fig. 3A, Supplementary Table S3). For example, the sgRNA used to target *HSP90AA1* predominantly generated 1 bp insertions (>50%) in WT cells (Fig. 3A). In NHEJ deficient cell lines, the same sgRNA generated 1 bp insertions in only 19–0.1% of the editing outcomes. Instead, 10–30 bp deletions (42–47%) were the dominant mutation pattern in these genetic backgrounds. Moreover, for sgRNAs that prominently generated deletions, we observed an increase in the size of these deletions in NHEJ-abrogated cells. For the *CD46*-targeting sgRNA (Fig. 3A), deletions smaller than 5 bp in WT cells (39%) were considerably decreased in NHEJ abrogated cell lines (2.5–5.8%), giving rise to larger deletions. A similar trend was observed for the *RNF152*-targeting sgRNA (Fig. 3A) and for other sgRNAs targeting different exonic regions, introns or promoters of these genes (Supplementary Fig. 3A, Supplementary Table S3). Importantly, our study is limited to the analysis of indels < 80 bp. Even though this range covers the majority of Cas9-proximal editing events, it does not allow speculation on larger rearrangements.

The observed shift in indel size suggested the activity of a distinct DNA repair pathway that is able to fully compensate for the loss of NHEJ, leading to the mutagenic repair of Cas9-generated DNA breaks. Since alt-EJ (also known as microhomology-mediated end joining) is known to generate larger rearrangements, we hypothesized that this might be the pathway responsible for the editing observed. The first step in alt-EJ involves 5'-end resection, to expose and allow the base-pairing of flanking regions of microhomology (MH), across the border of the DSB²⁴. To test if alt-EJ is the pathway active at such lesions, we generated cells lacking the proofreading-deficient A-family DNA polymerase theta (POLQ; Δ POLQ), the polymerase that functions in alt-EJ, following MH annealing¹³ (Supplementary Fig. 3B). Additionally, a cell line defective in both NHEJ and alt-EJ was generated by knocking-out POLQ, in the previously generated Δ LIG4 cell line (Supplementary Fig. 3B). The abrogation of alt-EJ was phenotypically confirmed in Δ POLQ and Δ LIG4/POLQ cells, by measuring their sensitivity to several DSB-inducing agents (Supplementary Fig. 3C). As expected, Δ POLQ cells were more sensitive to DSB-inducing agents than WT cells, but more resistant than Δ LIG4 cells, since alt-EJ is not considered to be the main pathway by which DSBs are repaired²⁵. Δ LIG4/POLQ cells displayed an additive sensitivity to DSB-inducing agents. Indel signature analysis of cells, transfected with sgRNAs targeting exonic regions of *HSP90AA1* and *RNF152*

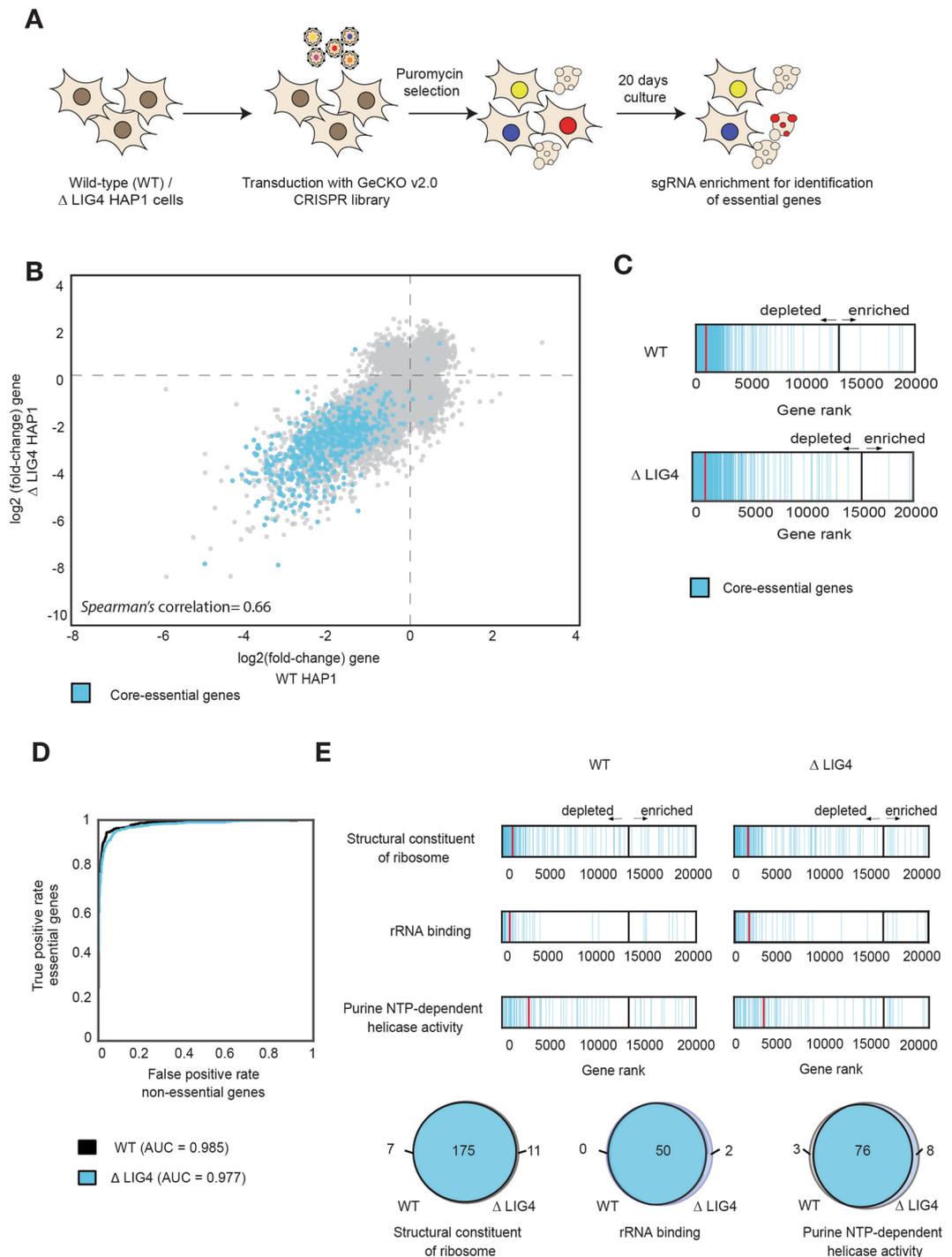


Figure 2. Genome-wide CRISPR-Cas9 knockout screens for global gene disruption are efficient in the absence of non-homologous end-joining. **(A)** Schematic overview of the CRISPR screens. HAP1 WT and Δ LIG4 cells were infected at a low MOI (0.3) with the GeCKO v2.0 genome-wide CRISPR knockout library. After a period of puromycin selection, cells were kept in culture for 20 days, allowing essential genes to drop out of the population. Cells were harvested and sgRNAs were sequenced to determine relative abundances. **(B)** Scatter plot representing the $\log_2(\text{fold-change})$ enrichment of each gene, after culturing WT or Δ LIG4 cells transduced with the GeCKO v2.0 CRISPR library for 20 days (see *Methods* for details on the calculation of gene fold-change enrichment). Blue colored nodes represent core essential genes. Data shown for three independent experiments ($n = 3$). Spearman's correlation (0.66) between WT and Δ LIG4 screens is depicted. **(C)** Density plot representing the position of core essential genes in the gene rank, based on $\log_2(\text{fold-change})$. Red lines represent the median $\log_2(\text{fold-change})$ of the depicted genes. Black lines represent the threshold between depleted and enriched genes. Data shown for 3 independent experiments ($n = 3$). **(D)** Receiver operating characteristic (ROC) analysis of depleted genes in WT and Δ LIG4 cells. False positive rates are calculated for non-essential genes and plotted against true positive rates for essential genes. Area under the curve (AUC) for each ROC curve is represented. Data shown for 3 independent experiments ($n = 3$). **(E)** Density plot representing the gene rank position of

genes annotated for the top three enriched GO terms in the core ‘essentialome’, based on their $\log_2(\text{fold-change})$. Red lines represent the median $\log_2(\text{fold-change})$ of the depicted genes. Black lines represent the threshold between depleted and enriched genes. Venn diagrams represent the intersection of depleted genes for the annotated GO terms in WT and ΔLIG4 cells. Data shown for 3 independent experiments ($n = 3$).

(Fig. 3B), showed that WT and ΔPOLQ cells have a very similar indel profile that differed from the larger deletions observed in ΔLIG4 cells. This finding indicates that NHEJ is the default pathway that repairs Cas9-generated breaks and illustrates that the editing observed in ΔLIG4 cells is the product of a distinct and alternative pathway. This observation was extended to the targeting of an additional exon sequence, as well as intronic and promoter regions of these genes (Supplementary Fig. 3D).

We observed that 72–77% of the analyzed reads in $\Delta\text{LIG4}/\text{POLQ}$ cells corresponded to the unedited genomic sequence, indicating that mutagenic repair of Cas9-breaks was largely impaired in these double deficient cells (Fig. 3B, Supplementary Fig. 3D). This observation was further confirmed by assessing the kinetics of mCherry editing (Supplementary Fig. 3E) and indicates that alt-EJ, via POLQ, is indeed responsible for the mutagenic repair observed in NHEJ deficient cells, leading to the generation of larger indels. Surprisingly, however, in the absence of both pathways, indel generation was still achieved in 10–20% of reads (Fig. 3B and Supplementary Fig. 3D).

Discussion

The repair of DSBs has been widely studied by using the nuclease *I-SceI*. However, considering that the structure of DNA termini affects repair outcome, it is important to highlight that the *I-SceI* nuclease generates a staggered cut leaving a 3' overhang, whereas Cas9 generates blunt ends^{4–6}. In line with this, NHEJ is predominantly precise when functioning on DNA breaks introduced by the *I-SceI* nuclease²⁶, but largely error-prone when functioning on blunt ends^{27,28}. Moreover, the Cas9-sgRNA complex has been shown to adhere to DNA for several hours post-cutting, a phenomenon that impacts the outcome and fidelity of DSB repair^{9,11}. Hence, the repair of Cas9-induced DSBs is not representative of *I-SceI* induced breaks and therefore further research is warranted to elucidate these repair mechanisms. This is particularly relevant in light of the therapeutic potential of CRISPR-Cas9 based technologies and especially considering that error-prone pathways are being explored to correct disease-relevant mutations^{23,29}.

Following the generation of a DSB, different pathways engage in its repair, with NHEJ and other end-joining pathways, such as alt-EJ and single-strand annealing (SSA), contributing to different amounts^{30,31}. When NHEJ is absent, due to the lack of one key protein, the activity of other end-joining pathways becomes apparent. Alt-EJ pathways typically require larger regions of microhomology, with the POLQ-dependent alt-EJ pathway requiring between 2–20 bp of microhomology, compared to the NHEJ microhomology requirement of ≤ 4 bp. Alternatively, SSA requires > 20 bp homology¹³. End-resection is therefore the first barrier that needs to be overcome in order to enable alt-EJ pathways to function, with NHEJ factors such as Ku70/80³² and the p53-binding protein 1 (53BP1)³³, present at high concentrations, preventing this from happening. Additionally, extensive end-resection is also dependent on cell cycle, as factors that promote end-resection are more active during S and G2 phases³³. Hence, in G1 phase, DSBs are preferentially repaired by NHEJ and even during S and G2 phases, when extensive end resection can take place, the resection machinery must still overcome the presence of NHEJ factors at DNA ends. This is well represented by the 4:1 estimated ratio of NHEJ to HDR in WT mammalian somatic cells in S/G2 phases³⁴. If NHEJ is absent, alt-EJ may be favored over SSA in G1 phase, owing to the limited amount of resection that alt-EJ requires compared to SSA. However, it is still not clear what dictates the use of alt-EJ as opposed to SSA in S/G2 phases. Time can be an important factor, as the longer a DSB remains unrepaired, the more end processing can occur to favor SSA. In our study, contrary to NHEJ deficiency, alt-EJ deficiency led to an indel profile that was very similar to that observed in WT cells. This confirms the current view that NHEJ is the main pathway by which Cas9-breaks are repaired and that alt-EJ plays only a minor role. However, the high efficiency of editing in NHEJ deficient cells, together with the almost complete abrogation of mutagenic repair in $\Delta\text{LIG4}/\text{POLQ}$ cells, indicates that alt-EJ, in a POLQ-dependent manner, can fully compensate for the absence of NHEJ. This can have important applications for improving error-free repair, as the simultaneous transient inhibition of LIG4 and POLQ might increase HDR efficiency. Additionally, our results indicate that, in the absence of both NHEJ and alt-EJ, editing is still possible albeit with reduced efficiency (10–20%). This observation suggests the possible existence of an additional DNA repair mechanism that deals with Cas9-generated lesions. We speculate that SSA might be a potential pathway for the residual repair observed in the absence of both NHEJ and alt-EJ.

Taken together, our results show that mutagenic repair of DSBs generated by Cas9 can occur efficiently in the absence of NHEJ. We draw this conclusion utilizing genetic models where NHEJ has been abrogated, as opposed to chemical inhibitors that might lead to off-target effects and incomplete inhibition³⁵. While it is theoretically possible that, in the absence of NHEJ, the repair is predominantly error-free, leading to a continuous Cas9 cutting until the target site is no longer homologous to the sgRNA, we do not favour this hypothesis based on the kinetics of mCherry-editing. Here, several cycles of Cas9 cleavage and repair would result in a considerable delay in repair of NHEJ deficient cells compared to WT cells. Since we observed similar kinetics of editing between WT and ΔLIG4 cells, coupled with the low rates of HDR in the absence of a provided repair template³⁶, we conclude that NHEJ is dispensable for the efficient mutagenic repair of Cas9-breaks. This is further confirmed by the efficiency of genome-wide CRISPR loss-of-function approach in the absence of NHEJ, which indicates for the first time, and in a holistic approach, that global-gene disruption by CRISPR-Cas9 is independent of NHEJ.

In the context of CRISPR-Cas9-mediated editing, it has been described that the repair outcomes are predominantly determined by the sgRNA sequence, rather than genomic context¹². Following this observation, several

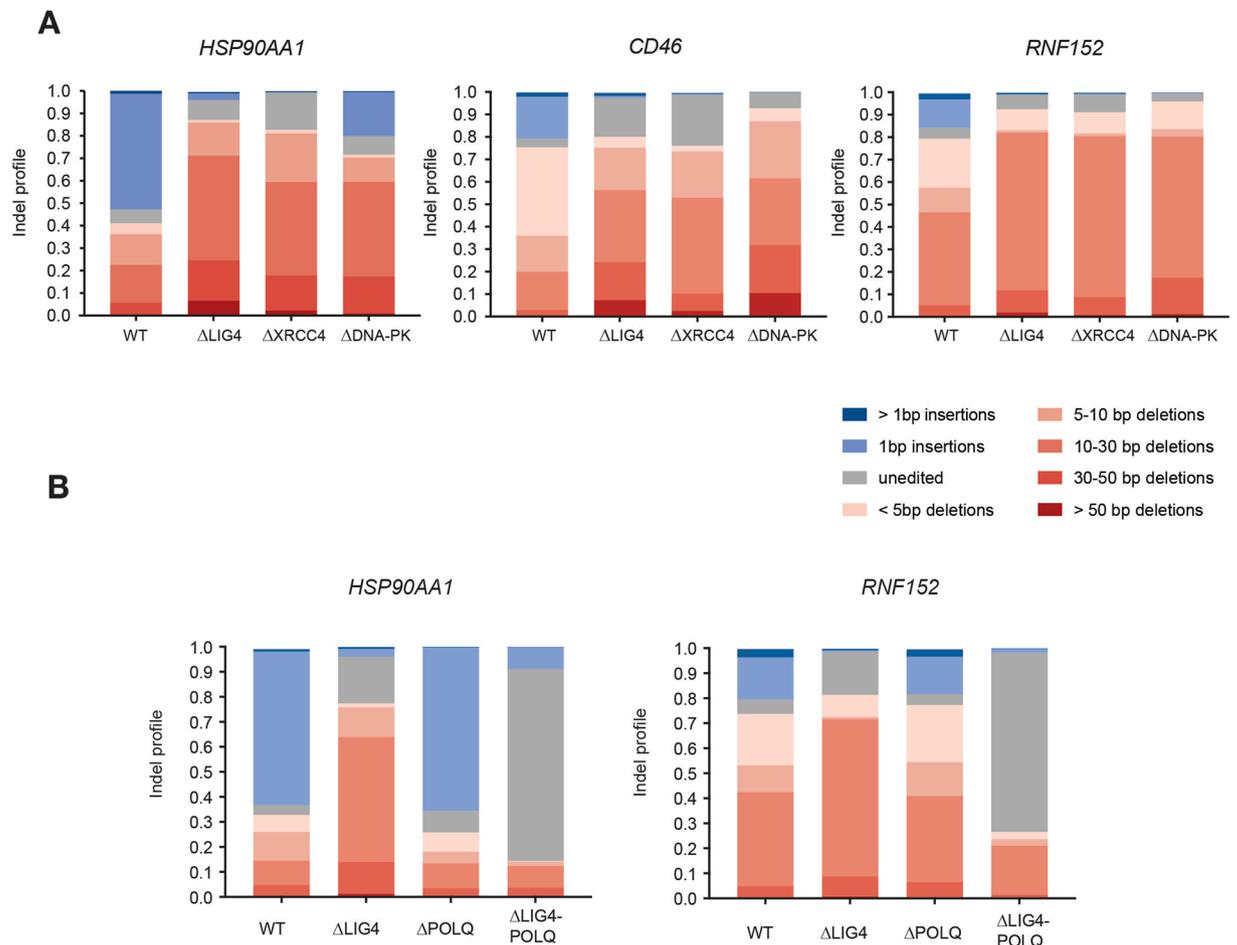


Figure 3. Genetic dissection of DNA repair pathway contribution to mutagenic repair of Cas9 generated lesions. (A) HAP1 WT, Δ LIG4, Δ XRCC4 and Δ DNA-PK cells were transfected with Cas9 and sgRNAs targeting exonic regions of 3 different genes (*HSP90AA1*, *CD46* and *RNF152*). After selection, genomic DNA was extracted and sgRNA-targeted regions were PCR-amplified. Amplicon sequencing was used to determine the indel size distribution, following editing. (B) Indel size distribution resulting from editing of exonic regions of *HSP90AA1* and *RNF152* in WT, Δ LIG4, Δ POLQ and Δ LIG4/POLQ cells, following the same procedure as described in A.

studies have shown that each sgRNA generates a preferential editing outcome, a prediction that might have important applications for template-free genome editing²³. By amplicon sequencing of several targeted regions, we were able to confirm the individual editing biases of sgRNAs. Moreover, we observe that these outcomes can be manipulated by the abrogation of NHEJ. NHEJ deficiency led to a distinct indel profile, characterized by the absence of small insertions and the predominance of larger deletions (10–30 bp). We hypothesize that this observation may have important applications for mutagenizing non-coding regions of the genome to disrupt, for example, the binding of transcription factors. As the implementation of functional genetic screens for non-coding transcriptional regulatory elements has been hampered by the small indel size generated by NHEJ (<10 bp)^{37,38}, we speculate that the larger indels produced upon inhibition of this pathway might accelerate the development of CRISPR-Cas9 approaches for the identification of active functional enhancers, in a high-throughput manner.

Materials and Methods

Cell lines and culture conditions. Human HAP1 cells were obtained from Horizon Discovery and were grown in Iscove's Modified Dulbecco's Medium (IMDM) from GIBCO®, containing L-Glutamine and 25 mM HEPES and supplemented with 10% Fetal Bovine Serum (FBS) and 1% Penicillin/Streptomycin (P/S). All cell lines were diploid at the time of the experiments. HEK293T cells were obtained from the CRUK Cell Facility and were used for virus production, by culturing in Dulbecco's modified Eagle medium (DMEM) and supplemented with 10% FBS.

Plasmids. The human GeCKO v2.0 CRISPR knockout pooled library was a gift from Feng Zhang (Addgene #1000000048). Lenti-iCas9-neomycin was a gift from Qin Yan (Addgene # 85400). For the pCROP-mCherry-PEST plasmid, CROPseq-Guide-Puro vector (Addgene #86708) was initially digested with

BsiWI/MluI and the puromycin resistance was replaced with PCR-amplified mCherry. The obtained plasmid was digested using BsrGI/MluI and a gene block (IDT) containing homology overhangs and a PEST sequence was inserted via Gibson-assembly (NEB HiFi Assembly), according to the manufacturer's protocol. The plasmid was then digested with BsmBI and a sgRNA targeting the active site of mCherry was inserted in the place of the filler. sgRNA mCherry:

forward: 5'-CACCGTTGGAGCCGTACATGAACTG-3'
reverse: 5'-AAACCAGTTCATGTACGGCTCCAAC-3'

BsrGI/MluI gene block (IDT):

5'-TCCACCGGCGGCATGGACGAGCTGTACAAGAAGCTTAGCCATGGCTTCCCGCCGGAGGTGG
AGGAGCAGGATGATGGCAGCTGCCCATGTCTTGTGCCAGGAGAGCGGGATGGACCGTCACCC
TGCAGCCTGTGCTTCTGCTAGGATCAATGTGTAGTAAACGCGTTAAGTCGACAATCAACCTCTG-3'

CRISPR-Cas9-mediated gene editing. CRISPR-Cas9 knockouts of LIG4, DNA-PKcs, XRCC4, POLQ and LIG4/POLQ were generated in collaboration with Horizon Genomics. Sequences for sgRNAs were designed by Horizon Genomics or with the use of <http://chopchop.cbu.uib.no/>. Sequences of sgRNAs used were:

LIG4: 5'-AAGGTCGTTTACTTGCTGTA-3'
XRCC4: 5'-TTACTGATGGTCATTCAGCA-3'
DNA-PK: 5'-ATAGAGCTGGTACATGGGTG-3'
POLQ: 5'-GATTCGTTCTCGGAAGCGG-3'

Sanger sequencing. Genomic DNA was extracted using the QIAGEN DNeasy Blood & Tissue Kit, according to the manufacturer's protocol. Genomic regions around the sgRNA-targeted sequences were amplified using the following primer pairs:

LIG4-forward: 5'-GTAGTGACATTATGCAACTCAGCAG-3'
LIG4-reverse: 5'-TAGAGATGGAAAAGATGCCCTCAA-3'
XRCC4-forward: 5'-TGAGAGGCCAGTACAGAAAACATTA-3'
XRCC4-reverse: 5'-ACCTGTGTATAAATTTGACAGCAAT-3'
DNA-PK-forward: 5'-CTGCTGACCACTGAATTAGACAAAC-3'
DNA-PK-reverse: 5'-TTGCAGCCTGTGAACCTTTACATAG-3'
POLQ-forward: 5'-AGTAGAAGCCCAATGGGGTATG-3'
POLQ-reverse: 5'-GAGGTTTGAGTTTGAAGACTGGC-3'

PCR amplification conditions were as follows: heat lid 110 °C; 94 °C 2 min; loop 35 × (94 °C 30 s; 55 °C 30 s; 68 °C 1 min) 68 °C 7 min. Frameshift mutations were confirmed using Nucleotide BLAST against the reference genome GCF_000001405.33.

Dose-response curves. Dose-response curves for neocarzinostatin (NCS), doxorubicin, etoposide and methyl methanesulfonate (MMS) were performed in 96-well plates, by seeding 1,000 HAP1 cells per well, the day before treatment. The following day, compounds were added at twofold serial dilutions, from the highest dose (NCS: 500 mg/mL; doxorubicin: 125 nM; etoposide: 2 μM; MMS: 750 nM). Four days after treatment, cell viability was measured using Cell Titer-Glo (Promega).

Immunoblotting and antibodies. Cell extracts were prepared in RIPA lysis buffer (NEB) supplemented with protease inhibitors (Sigma) and phosphatase inhibitors (Sigma, NEB). Immunoblots were performed using standard procedures. Protein samples were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (3–8% gradient gels, Invitrogen) and subsequently transferred onto nitrocellulose membranes. Primary antibodies for Cas9 (7A9-3A3, Cell Signaling Technology #14697) and β-Actin (A5060, Sigma) were used at 1:1,000. Secondary antibodies were used at 1:5,000 (HRP-conjugated goat anti-mouse or anti-rabbit IgG from Jackson Immunochemicals). Immunoblots were imaged using a Curix 60 (AGFA) table-top processor.

Kinetics of indel generation for mCherry active site. *Virus production.* HEK293T cells were seeded in 6-well plates at 200,000 cells per well and transfection was performed the following day with 0.3 μg per well of the VSG and 0.5 μg per well of the psPAX2 packaging vectors, together with 1 μg per well of either the iCas9-GFP vector, or the pCROP-mCherry-PEST vector. The Effectene Transfection Reagent (QIAGEN) was used at 20 μL per well. Supernatant containing the viral particles was harvested two- and three-days post-transfection and filtered with a 0.45 μm filter (Milipore Steriflip HV/PVDF). Viral supernatants were stored at –80 °C.

Generation of iCas9-GFP, pCROP-mCherry-PEST cell lines. Cells were first transduced with the iCas9-GFP vector, using a virus dilution of 1:12 in a 24 well-plate and 8 μg/mL of polybrene. Spin-infection was performed at 2,000 rpm, 30 minutes, at 30 °C. In order to enrich for Cas9-expressing cells, transduced cells were treated with doxycycline (2 μg/mL) for 24 hours and the GFP positive population was sorted, using a SH800S Cell Sorter (Sony Biotechnology). Sorted cells were kept in culture, in the absence of doxycycline, for at least one week. After this period, cells were transduced with the pCROP-mCherry-PEST plasmid, following the same spin-infection protocol, and then sorted for mCherry-positive cells.

Kinetics of mCherry editing. To assess the kinetics of indel generation, 20,000 cells per well were plated in triplicate in a 12-well plate. Doxycycline (2 µg/mL) was added to the medium at the indicated time points and cells were analyzed in a BD LSRFortessa flow-cytometer. mCherry fluorescence was assessed upon gating on GFP-positive cells.

Genome-wide CRISPR-Cas9 screen. Virus production. The GeCKO v2.0 CRISPR library virus was produced as reported by the distributor (Addgene #1000000048) using both library A and B in a one-production step. HEK-293T cells were seeded at 40% confluency in T-225 flasks and transfected, 24 hours later, with the GeCKO library A and B, pVSVG and psPAX2 packaging plasmids, using Lipofectamine® 2000 (Invitrogen, ThermoFisher Scientific), according to the manufacturer's protocol. After 6 hours, medium was changed to DMEM (10% FBS) and after 60 hours, supernatant-containing virus was harvested and filtered through a 0.45 µm filter (Milipore Steriflip HV/PVDF).

Screen setup. Three biological replicates were performed for each screen. HAP1 cells were infected at a multiplicity of infection (MOI) between 0.3–0.5. For each cell line (WT and ΔLIG4), 100 million cells were spin-infected. Day 1: 6 12-well plates were seeded with 1.5 million cells per well, supplemented with viral supernatant and IMDM (10% FBS, 1% P/S) to reach a volume of 1 mL per well. Polybrene was added at the final concentration of 8 µg/mL. Cells were spin-infected at 2,000 rpm, 37 °C, for 3 hours, pooled and transferred to 15 cm dishes. Day 2: Cells were exposed to 2 µg/mL of puromycin to select for infected cells. Day 7: Medium was replaced with IMDM (10% FBS, 1% P/S). Cells were kept in culture for 20 days after puromycin selection and split every 2–3 days to avoid confluency, re-seeding > 100 million cells each time. After this period, cells were harvested and genomic DNA was extracted using the QIAGEN Blood & Cell Culture Maxi Kit, according to the manufacturer's protocol.

sgRNA amplification and sequencing. Amplification of the sgRNA sequences was performed in a two-step PCR, using PCR1- and barcoded PCR2-primers, as described by the distributor (Addgene). Primer sequences were obtained from <http://genome-engineering.org/gecko/wp-content/uploads/2013/12/GeCKO-plasmid-readout-primers-July2014.xlsx>. PCR1 amplified the sgRNA sequences, using 130 µg genomic DNA in 13 × 100 µL reactions per sample and GoTaq G2 DNA Polymerase (Promega). PCR1 program: Heat lid 110 °C; 94 °C 2 min; loop 20 × (94 °C 30 s; 55 °C 30 s; 68 °C 1 min) 68 °C 7 min. PCR1 reaction tubes were pooled for each sample. PCR2 added Illumina sequencing adapters by using 2 µL of input from PCR1. A test PCR with different amplification cycles was conducted and products were ran on a 0.8% agarose gel. The number of cycles for which a band with approximately 380 bp was visible, but not saturated, was selected (*n*). PCR2 was then performed following the program: Heat lid 110 °C; 94 °C 2 min; loop *n* × (94 °C 30 s; 55 °C 30 s; 68 °C 1 min) 68 °C 7 min. PCR2 product was purified by size-exclusion, using magnetic AMPure XP DNA beads (NEB), using a 1:0.45 ratio to remove fragments > 1,000 bp, followed by a 1:2 ratio clean-up. Barcoded samples were pooled and sequenced using 61 base-pair single-end sequencing. Sequencing of the GeCKO plasmids (library A and B) was performed in the same way, using 200 ng of plasmid per reaction for PCR1.

Screen analysis. sgRNA sequences were retrieved by trimming all sequences 5' relative to the adapter sequence (CGAAACACCG) and 20 nucleotides 3' following this. MAGeCK³⁹ was used to generate the sgRNA counts, using a pre-made index of the GeCKO v2.0 library. sgRNA counts were normalized to million counts, for each sequencing sample and averaged across the three biological replicates. Gene log₂(fold-change) was calculated by selecting a best representative sgRNA for each gene, as following: 1) The log₂(fold-change) of each sgRNA was calculated by comparing to the sequenced GeCKO library; 2) The average of the log₂(fold-change) for all sgRNAs targeting the same gene was calculated. Genes with less than 3 sgRNAs were excluded from this analysis; 3) If the average was positive, it was assumed that the gene had a tendency to be enriched in the screen, in comparison to the sequenced library. Therefore, the sgRNA with the 2nd highest log₂(fold-change) was selected as the best representative for that particular gene. If the average was negative, it was assumed that the gene had a tendency to be depleted in the screen. Therefore, the sgRNA with the 2nd lowest log₂(fold-change) was selected as the best-representative sgRNA. By excluding the highest and lowest sgRNAs, we prevent the introduction of biases. Significance of the enrichment analysis (assessed by *p*-value) was calculated using MAGeCK, comparing the screens (WT and ΔLIG4) with the sequenced library.

Receiver operating characteristic (ROC) analysis of cell viability was calculated by filtering the 683 genes annotated to be core-essential²⁰ and 927 genes annotated as non-essential²¹. The ability of each screen (WT and ΔLIG4) to distinguish between these essential (true positives) and non-essential genes (false positives) was assessed by plotting their ROC curves (False Positive Rate [FPR] vs True Positive Rate [TPR]) and calculating the respective Area Under the Curve (AUC). Values used for the ROC curve were based on the gene $-\log_{10}(p\text{-value})$.

Gene Ontology (GO) enrichment analysis of core-essential genes²⁰ for molecular processes, was performed by extracting the GO annotations from the Gene Ontology Consortium [www.geneontology.org]. For every GO term, the fold-enrichment was computed over a background comprising the entire human genome. *p*-value was calculated by Fisher's exact test and adjusted by *Bonferroni* correction.

Indel analysis by next generation sequencing. Amplicons were designed to have the sgRNA target site at the center of the products. sgRNAs were designed to target different genomic regions within different genes, using <http://chopchop.cbu.uib.no/>. sgRNA sequences can be found in Supplementary Table S4.

HAP1 cells were seeded in 6-well plates at a confluency of 40% and transfected with the respective sgRNA and Cas9 constructs the following day, using Effectene as the transfection reagent (QIAGEN), according to the manufacturer's protocol. Transfected cells were then selected with blasticidine (20 µg/mL) for 2 days and harvested

as soon as confluent (4–7 days). Genomic DNA was extracted using the QIAGEN DNeasy 96 Blood & Tissue Kit, according to the manufacturer's protocol. PCR reactions were set in a reaction volume of 50 μ L. The DNA polymerase Q5 High Fidelity (NEB) was used to amplify 100 ng of genomic DNA, using the following program: Heat lid at 110 °C; 98 °C for 30 s; loop 35 \times (98 °C for 30 s, Annealing temperature (primer dependent) for 30 s, 72 °C for 1 min); 72 °C for 2 min. Primer sequences can be found in Supplementary Table S4. PCR products were purified by a 2.0 \times AMPure XP bead clean-up (NEB), measured using a Qubit HS assay (Invitrogen), and used as input for Nextera XT (Illumina cat. no 15032350) library preparations, performed according to protocols provided by the supplier. Libraries were sequenced on the Illumina MiSeq platform, using a 150-cycle v3 flow-cell with dual indexing. The machine was set to read lengths of 159 (read1) + 8 (i7) + 8 (i5) bases. The analysis of the data was performed by defining two 10 bp 'anchor' sequences on both sides of the sgRNA, at a fixed distance of 80 bp. Reads spanning the sgRNA target site were extracted from the BAM file, via a grep operation for the pattern '<anchor_left>.*<anchor_right>' on the BAM file, using the -o option to return only the matching part of the sequence. For unedited fragments, this sequence equals 10 bp anchor_left + 30 bp + 20 bp (sgRNA) + 30 bp + 10 bp anchor_right (total of 100 bp). The size of the indels were calculated as the deviation from the unedited fragment length, summarized and plotted.

Statistical analysis and data visualization. All simulations and visualizations used the Python programming language, version 2.0 (Python Software Foundation, <https://www.python.org/>).

Ethical approval. The methods described in this manuscript were carried out in accordance with the relevant guidelines and regulations.

Data and Code availability

All data generated or analyzed during this study is included in this published article and its supplementary information. Sequencing of sgRNA cassettes in the 6 genome-wide CRISPR-Cas9 screens (associated with Fig. 2 and Supplementary Fig. 2) have been deposited in the NCBI Sequence Read Archive (SRA) with the final SRA accession code: PRJNA565227.

Received: 30 May 2019; Accepted: 7 October 2019;

Published online: 31 October 2019

References

- Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* (80-). **346**, 1258096–1258096 (2014).
- Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32**, 347–355 (2014).
- Hsu, P. D., Lander, E. S. & Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157**, 1262–1278 (2014).
- Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* (80-). **339**, 819–823 (2013).
- Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, (2013).
- Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–6 (2013).
- Hustedt, N. & Durocher, D. The control of DNA repair by the cell cycle. *Nat. Cell Biol.* **19**, 1–9 (2017).
- Canny, M. D. *et al.* Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR–Cas9 genome-editing efficiency. *Nat. Biotechnol.* **36**, (2017).
- Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR–Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).
- Bothmer, A. *et al.* Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nat. Commun.* **8**, 13905 (2017).
- Brinkman, E. K. *et al.* Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks. *Mol. Cell* **0**, (2018).
- van Overbeek, M. *et al.* DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol. Cell* **63**, 633–646 (2016).
- Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* **18**, 495–506 (2017).
- Adachi, N., Suzuki, H., Iizumi, S. & Koyama, H. Hypersensitivity of Nonhomologous DNA End-joining Mutants to VP-16 and ICRF-193. *J. Biol. Chem.* **278**, 35897–35902 (2003).
- Cao, J. *et al.* An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting. *Nucleic Acids Res.* **44**, gkw660 (2016).
- Rogers, S., Wells, R. & Rechsteiner, M. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* (80-). **234**, 364–368 (1986).
- Brockmann, M. *et al.* Genetic wiring maps of single-cell protein states reveal an off-switch for GPCR signalling. *Nature*. <https://doi.org/10.1038/nature22376> (2017).
- Shalem, O. *et al.* Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* (80-). **343**, 84–87 (2014).
- Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
- Hart, T. *et al.* Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. *G3 (Bethesda)*. **7**, 2719–2727 (2017).
- Hart, T., Brown, K. R., Sircoulomb, F., Rottapel, R. & Moffat, J. Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. *Mol. Syst. Biol.* **10**, 733–733 (2014).
- Allen, F. *et al.* Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotechnol.* **37**, 64–72 (2018).
- Shen, M. W. *et al.* Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646–651 (2018).
- Seol, J. H., Shim, E. Y. & Lee, S. E. Microhomology-mediated end joining: Good, bad and ugly. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **809**, 81–87 (2018).
- Saito, S., Maeda, R. & Adachi, N. Dual loss of human POLQ and LIG4 abolishes random integration. *Nat. Commun.* **8**, 16112 (2017).
- Guirouilh-Barbat, J., Rass, E., Plo, I., Bertrand, P. & Lopez, B. S. Defects in XRCC4 and KU80 differentially affect the joining of distal nonhomologous ends. *Proc. Natl. Acad. Sci.* **104**, 20902–20907 (2007).
- Boulton, S. J. & Jackson, S. P. *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J.* **15**, 5093–103 (1996).

28. Schär, P., Herrmann, G., Daly, G. & Lindahl, T. A newly identified DNA ligase of *Saccharomyces cerevisiae* involved in RAD52-independent repair of DNA double-strand breaks. *Genes Dev.* **11**, 1912–24 (1997).
29. Román-Rodríguez, F. J. *et al.* NHEJ-Mediated Repair of CRISPR-Cas9-Induced DNA Breaks Efficiently Corrects Mutations in HSPCs from Patients with Fanconi Anemia. *Cell Stem Cell.* <https://doi.org/10.1016/j.stem.2019.08.016> (2019).
30. Ceccaldi, R. *et al.* Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. *Nature* **518**, 258–262 (2015).
31. Mateos-Gomez, P. A. *et al.* Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature* **518**, 254–257 (2015).
32. Mimitou, E. P. & Symington, L. S. Ku prevents Exo1 and Sgs1-dependent resection of DNA ends in the absence of a functional MRX complex or Sae2. *EMBO J.* **29**, 3358–3369 (2010).
33. Escribano-Díaz, C. *et al.* A Cell Cycle-Dependent Regulatory Circuit Composed of 53BP1-RIF1 and BRCA1-CtIP Controls DNA Repair Pathway Choice. *Mol. Cell* **49**, 872–883 (2013).
34. Beucher, A. *et al.* ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. *EMBO J.* **28**, 3413–27 (2009).
35. Jiang, W. *et al.* Differential Phosphorylation of DNA-PKcs Regulates the Interplay between End-Processing and End-Ligation during Nonhomologous End-Joining. *Mol. Cell* **58**, 172–185 (2015).
36. Richardson, C. D. *et al.* CRISPR-Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. *Nat. Genet.* <https://doi.org/10.1038/s41588-018-0174-0> (2018).
37. Korkmaz, G. *et al.* Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.* **34**, 192–198 (2016).
38. Lopes, R., Korkmaz, G. & Agami, R. Applying CRISPR-Cas9 tools to identify and characterize transcriptional enhancers. *Nat. Rev. Mol. Cell Biol.* **17**, 597–604 (2016).
39. Li, W. *et al.* MAGECK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).

Acknowledgements

We thank the Biomedical Sequencing Facility (BSF) at CeMM for all next generation sequencing and related bioinformatics analyses. J.F.S. was partially funded by an FWF stand-alone grant to J.I.L. (FWF29763) and by a DOC Fellowship from the Austrian Academy of Sciences (OAW25035). CeMM is funded by the Austrian Academy of Sciences.

Author contributions

J.F.S., G.W. and J.I.L. designed and conceptualized the project. J.F.S. planned and conducted all the experiments and also analyzed and visualized the data. S.S. optimized the protocol and prepared the samples for NGS indel analysis, together with J.F.S. M.W. generated the knockout cell lines. P.D. and C.B. contributed to the analysis of indels. A.H. and P.E. cloned the pCROP plasmid and contributed to genetic manipulations. G.S.-F. contributed to the project design. J.F.S. and J.I.L. wrote the original draft of the manuscript. J.I.L. supervised the project. All authors commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-52078-9>.

Correspondence and requests for materials should be addressed to J.I.L.

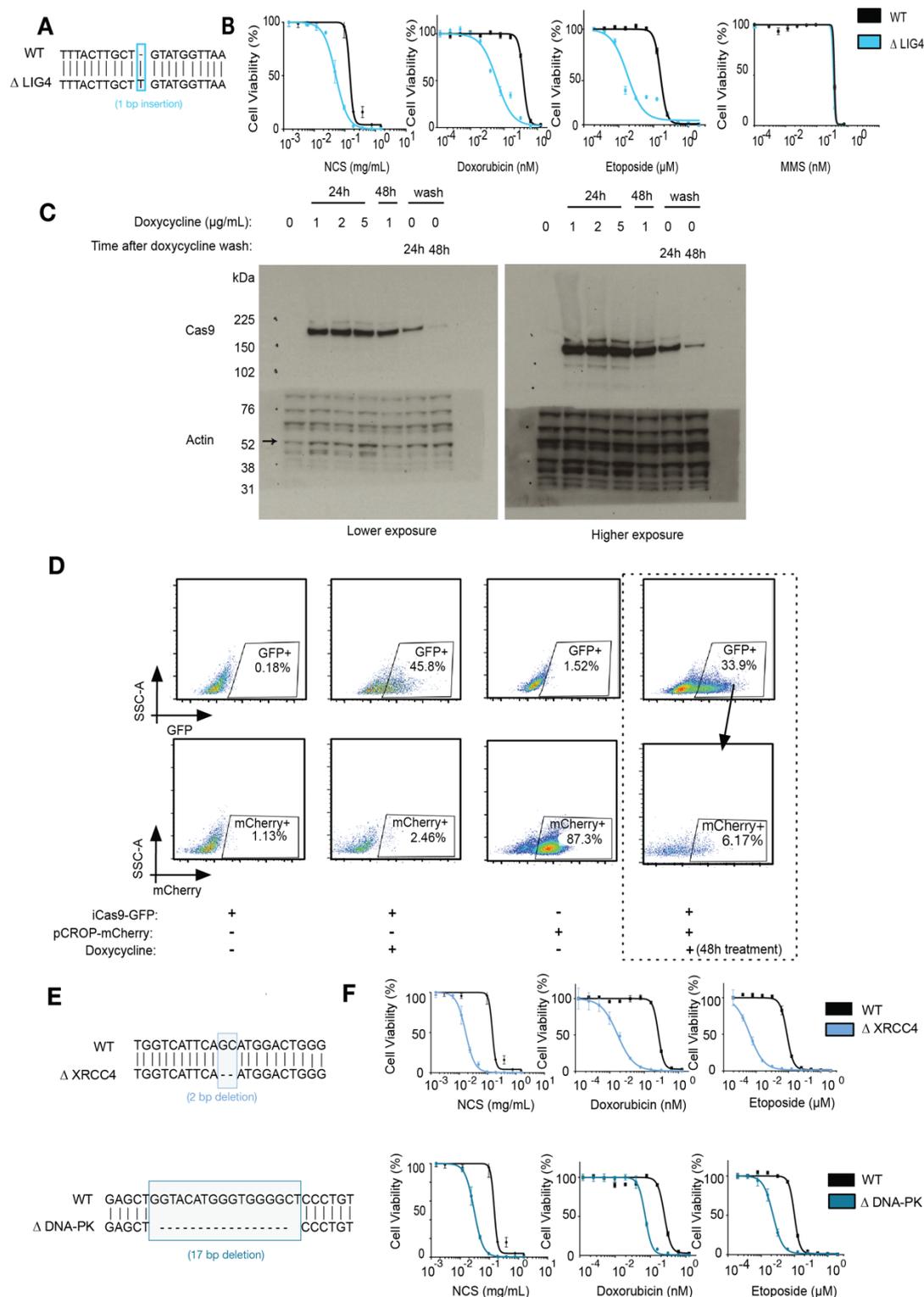
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

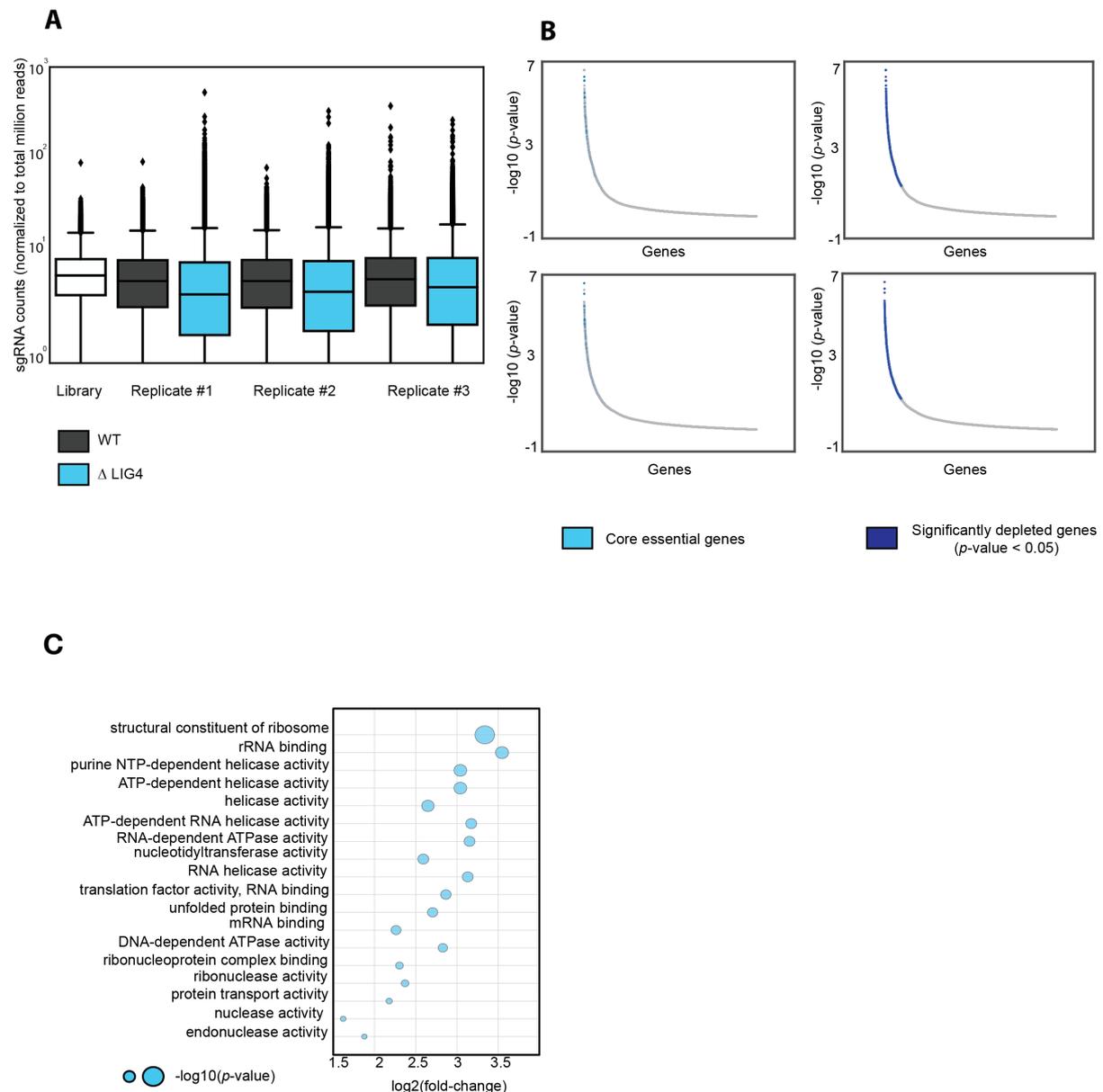
© The Author(s) 2019



Supplementary Figure 1

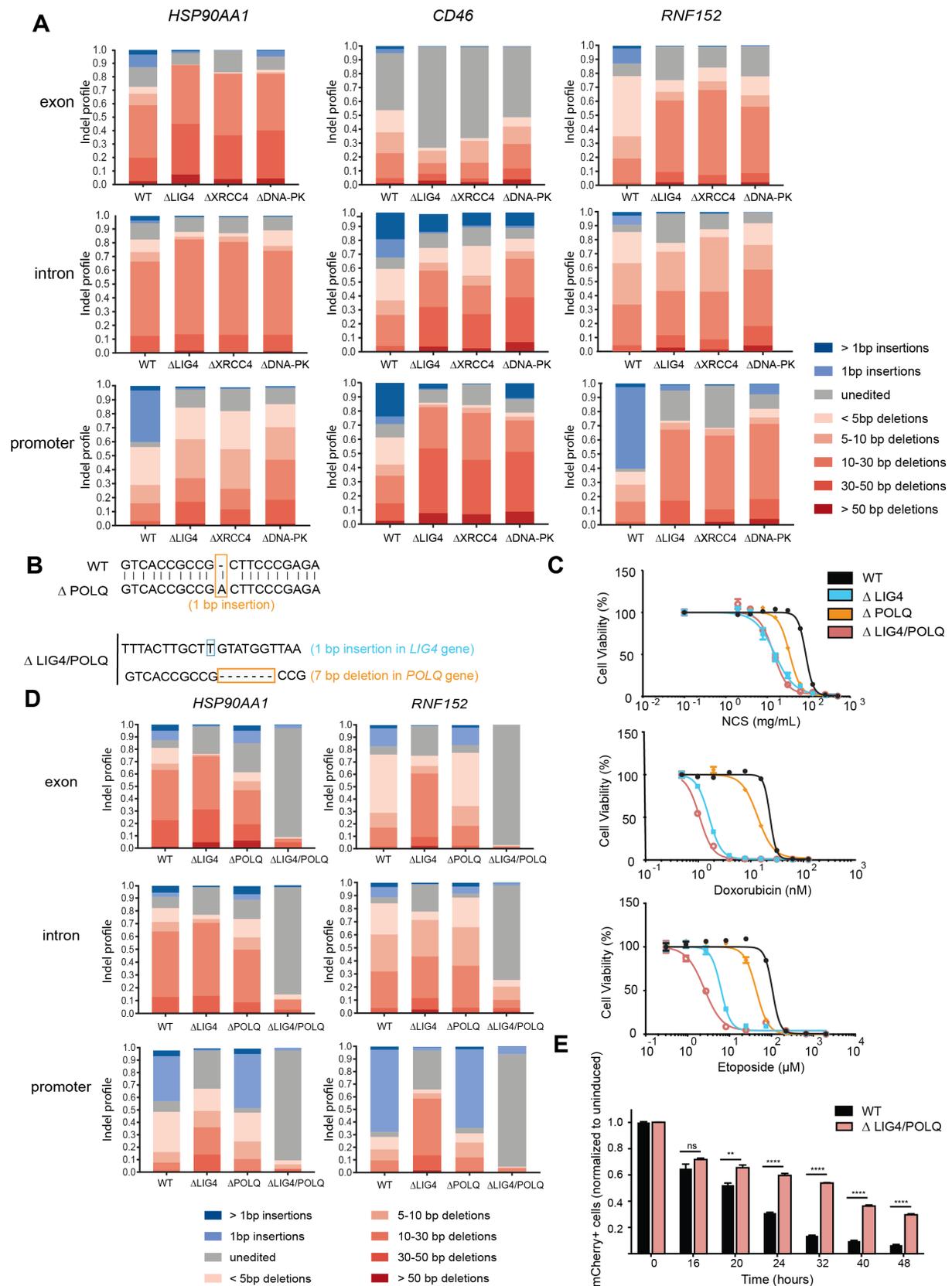
Supplementary Figure 1, related to Figure 1: Generation of a relevant cellular system to assess the kinetics of Cas9-activity. **A.** CRISPR-Cas9 generated clonal HAP1 cell line carrying a 1bp insertion in the *LIG4* gene. **B.** Dose-response curves for WT and Δ LIG4 HAP1 cell lines to the DNA double-strand break-inducing agents neocarzinostatin (NCS), doxorubicin and etoposide and to the alkylating agent methyl methanesulfonate (MMS). Cells were treated with the indicated compounds for 3 days and viability was measured by Cell Titer

Glo. **C.** Immunoblot for Cas9 and β -actin in HAP1 cells expressing doxycycline-inducible Cas9 tagged with GFP, with or without doxycycline treatment, as indicated. Figure represents the entire gel, with different exposure times. **D.** Representative FACS-plots for WT HAP1 cells expressing the indicated integrated constructs, with or without doxycycline treatment for Cas9 induction. mCherry editing was assessed by gating GFP-positive cells. **E.** Generation of Δ XRCC4 and Δ DNA-*PK* clonal HAP1 cell lines, with a 2bp and a 17bp deletion, respectively. **F.** Dose response curves for WT, Δ XRCC4 and Δ DNA-*PK* HAP1 cells to the DNA double-strand break-inducing agents neocarzinostatin (NCS), doxorubicin and etoposide. Cells were treated with the indicated compounds for 3 days and viability was measured using Cell Titer Glo.



Supplementary Figure 2

Supplementary Figure 2, related to Figure 2: Genome-wide CRISPR-Cas9-KO screens identify essential genes efficiently in WT and ΔLIG4 cells. **A.** Depiction of sgRNA representation, normalized to total millions of reads for the GeCKO v2.0 library (depicted as 'Library'), as well as each biological replicate of HAP1 WT and ΔLIG4 cells. **B.** Genes ranked by $-\log_{10}(p\text{-value})$, calculated with MAGeCK. Light blue colored nodes represent genes identified as core essential in the WT and ΔLIG4 screens (*left-hand side*). Dark blue colored nodes represent genes considered to be significantly depleted in each screen ($p\text{-value} < 0.05$) (*right hand-side*). **C.** Gene ontology (GO) enrichment analysis for molecular processes of essential genes plotted by $\log_2(\text{fold-change})$. Size of the node represents statistical significance.



Supplementary Figure 3

Supplementary Figure 3, related to Figure 3: NHEJ and alt-EJ-mediated repair generate different types of indel signatures. A. HAP1 WT, Δ LIG4, Δ XRCC4 and Δ DNA-PK cells were transfected with Cas9 and sgRNAs targeting exonic, intronic or promoter regions of 3 different

genes (*HSP90AA1*, *CD46* and *RNF152*). After selection, genomic DNA was extracted and sgRNA-targeted regions were PCR-amplified. Amplicon sequencing was used to determine the indel size distribution, following editing. **B.** Generation of a clonal Δ POLQ HAP1 cell line with a 1bp insertion and a clonal Δ LIG4/ Δ POLQ double mutant HAP1 cell line (+1bp/-7bp, for LIG4 and POLQ respectively). **C.** Dose-response curves for WT, Δ LIG4, Δ POLQ and Δ LIG4/POLQ HAP1 cell lines to the DNA double-strand break-inducing agents neocarcinostatin (NCS), doxorubicin and etoposide. Cells were treated with the indicated compounds for 3 days and viability was measured by Cell Titer Glo. **D.** Indel size distribution in WT, Δ LIG4, Δ POLQ and Δ LIG4/POLQ HAP1 cells, following the repair of Cas9-induced breaks within exonic, intronic and promoter regions of *HSP90AA1* and *RNF152*, following the same procedure described in A. **E.** Kinetics of indel generation within the mCherry locus (measured by gating on GFP-positive cells) after Cas9-induction with doxycycline, at the indicated time points. Each time point was normalized to the uninduced (0h) time point. The assay was performed in WT and Δ LIG4/POLQ HAP1 cells (n=3). Statistical significance was calculated by the Student's t-test. ns=not significant, ** p -value ≤ 0.01 , **** p -value ≤ 0.0001 .

3.2 Prime Editing Efficiency and Fidelity are Enhanced in the Absence of Mismatch Repair

3.2.1 Prologue

Prime editing is a powerful and versatile genome engineering approach that allows for the precise introduction of base substitutions, insertions and deletions, into any given genomic locus. However, the efficiency of prime editing widely varies across different cell backgrounds, even for the same genomic locus. In this project, we sought to address the cause of this variation. Through a focused genetic screen targeting DNA repair factors, we have showed that the efficiency of prime editing is hampered by the mismatch repair (MMR) pathway. We have shown that the MMR factor MLH1 directly binds to sites of prime editing, indicating that MMR directly counteracts the insertion of the edit. Consequently, ablation of MMR increases prime editing efficiency across different human cell lines, several types of edits and multiple genomic loci.

The results from this project have been published in a pre-print server and are currently under editorial consideration for publication in 2021/2022.

Ferreira da Silva *et al*, **Prime Editing Efficiency and Fidelity are Enhanced in The Absence of Mismatch Repair**, bioRxiv 2021

<https://www.biorxiv.org/content/10.1101/2021.09.30.462548v1.full.pdf>

3.2.1 PDF of the Article

1 Prime Editing Efficiency and Fidelity are Enhanced in the Absence of 2 Mismatch Repair

3
4 Ferreira da Silva J^{1,2#}, Oliveira GP^{1#}, Arasa-Verge EA¹, Kagiou C^{1,2}, Moretton A^{1,2},
5 Timelthaler G¹, Jiricny J³, Loizou JI^{1,2*}

6
7 ¹Institute of Cancer Research, Department of Medicine I, Comprehensive Cancer Centre,
8 Medical University of Vienna, Borschkegasse 8a, 1090 Vienna, Austria

9 ²CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences,
10 Vienna, Austria

11 ³Institute of Biochemistry of the ETH Zurich, Otto-Stern-Weg 3, 8093 Zurich, Switzerland

12 #Equal contribution

13 *Correspondence

14
15 Correspondence: joanna.loizou@meduniwien.ac.at (JIL)

16 17 Running title

18 Enhancement of Prime Editing Efficiency and Fidelity by Ablation of Mismatch Repair

19 20 Keywords

21 Prime editing; mismatch repair; genome engineering; DNA repair

22 23 Abstract

24 Prime editing (PE) is a powerful genome engineering approach that enables the introduction
25 of base substitutions, insertions and deletions into any given genomic locus. However, the
26 efficiency of PE varies widely and depends not only on the genomic region targeted, but also
27 on the genetic background of the edited cell. To determine which cellular factors affect PE
28 efficiency, we carried out a focused genetic screen targeting 32 DNA repair factors, spanning
29 all reported repair pathways. We show that, depending on cell line and type of edit, ablation
30 of mismatch repair (MMR) affords a 2-17 fold increase in PE efficiency, across several human
31 cell lines, types of edits and genomic loci. The accumulation of the key MMR factors MLH1
32 and MSH2 at PE sites argues for direct involvement of MMR in PE control. Our results shed
33 new light on the mechanism of PE and suggest how its efficiency might be optimised.

37 **Introduction**

38 CRISPR-Cas9-based genome editing technologies are powerful new tools of functional
39 genomics, with considerable potential as future therapeutics (Jinek et al., 2012). However, the
40 efficiency of currently available genome editing protocols is limited. Moreover, the process
41 gives rise to undesirable side products that hinder the implementation of this technology in
42 clinical settings. To overcome these hurdles, there is need to identify the DNA metabolic
43 pathways and molecular mechanisms that govern editing outcomes, as well as the activities
44 of these pathways in different cellular and tissue contexts (Richardson et al., 2018; Yeh et al.,
45 2019; Ferreira da Silva et al., 2021; Hussmann et al., 2021). The first generation of Cas9-
46 based genome engineering tools used nucleases that could be directed to any desired region
47 of the genome by a single-guide RNA (sgRNA). Following the targeting of a site-specific DNA
48 double-strand break (DSB), the endogenous DNA end-joining pathways frequently repair this
49 lesion in an error prone manner, leading to insertions or deletions (indels) that give rise to loss-
50 of-function alleles (Bothmer et al., 2017). This approach was further adapted to include either
51 a single- or double-stranded donor template containing the desired edit. Here, the DSB is
52 processed by homology-directed repair (HDR), which catalyses the insertion of the donor
53 template that includes the edit. Unlike the former approach, which generates random indels,
54 the latter method permits the introduction of desired indels, as well as point mutations, into the
55 genome (Yeh et al., 2019). However, since HDR is inefficient, depends on potentially
56 deleterious DSBs and requires cell division, alternative approaches were needed.

57

58 Amongst such alternative approaches are Base Editing (BE) and Prime Editing (PE). The
59 former uses nucleobase modification chemistry to efficiently and precisely incorporate single
60 nucleotide variants into the genome of cells (Komor et al., 2016; Gaudelli et al., 2017; Gu et
61 al., 2021), but its scope is limited to single-base substitutions. This led to the development of
62 PE as a highly versatile genome editing approach that allows for the targeted insertion of
63 indels, point mutations and combinations thereof into the genome (Anzalone et al., 2019). PE
64 utilises a fusion of a Cas9(H840A) nickase (Jinek et al., 2012) and reverse transcriptase (RT)
65 that is targeted to a precise genomic region by a PE guide RNA (pegRNA). The pegRNA
66 includes the desired sequence change, as well as a short 3' terminal extension complementary
67 to the 5' sequence upstream from the nick within the target site. Annealing of the 3' terminus
68 of the pegRNA to the 3' segment of the nicked DNA strand generates a substrate for the RT,
69 which copies the RNA template and thus incorporates the desired edit into the 3' extension of
70 the nick. Dissociation of the RNA and annealing of the DNA strands generates a 3' flap
71 containing the edit. Transient melting and reannealing of the nicked target site give rise to a
72 mixture of molecules containing either 3' or 5' flaps. Successful installation of the desired edit
73 requires removal of the 5' flap and ligation of the resulting nick to yield a DNA heteroduplex

74 containing the edit in the RT-synthesised strand. The editing outcome of this method, referred
75 to as PE2, depends on the resolution of this heteroduplex. Utilising an additional sgRNA that
76 directs nicking to the original DNA strand, either concurrently to the edit installation (PE3), or
77 subsequently (PE3b), increases PE efficiency (Anzalone et al., 2019). The increased
78 efficiency in PE3 strategies has been suggested to require the DNA repair pathway known as
79 DNA Mismatch Repair (MMR) that would function in the repair of the nicked, non-edited
80 strand, by utilising the edited strand as template (Petri et al., 2021; Scholefield and Harrison,
81 2021).

82

83 Due to its versatility, PE has been used in a wide variety of models, such as zebrafish (Petri
84 et al., 2021), rice and wheat (Lin et al., 2020), mouse (Liu et al., 2020) and human stem
85 cells (Sürün et al., 2020). A notable feature of PE is its highly variable rates across different
86 genetic backgrounds, even within the same genomic locus and using the same
87 pegRNA (Anzalone et al., 2019). To address whether this could be explained by different DNA
88 repair capacities, we performed a targeted genetic screen aimed at identifying DNA repair
89 factors involved in PE. Here, we uncover an inhibitory role for MMR pathway in PE and show
90 that MMR proteins localise to sites of PE to directly counteract edit installation, rather than
91 promote it. Thus, deletion or transient depletion of MMR factors increase PE efficiency and
92 fidelity across different edit sites, types and cell lines.

93 **Results**

94 **A targeted genetic screen identifies the DNA repair pathway mismatch repair as** 95 **inhibitory for prime editing**

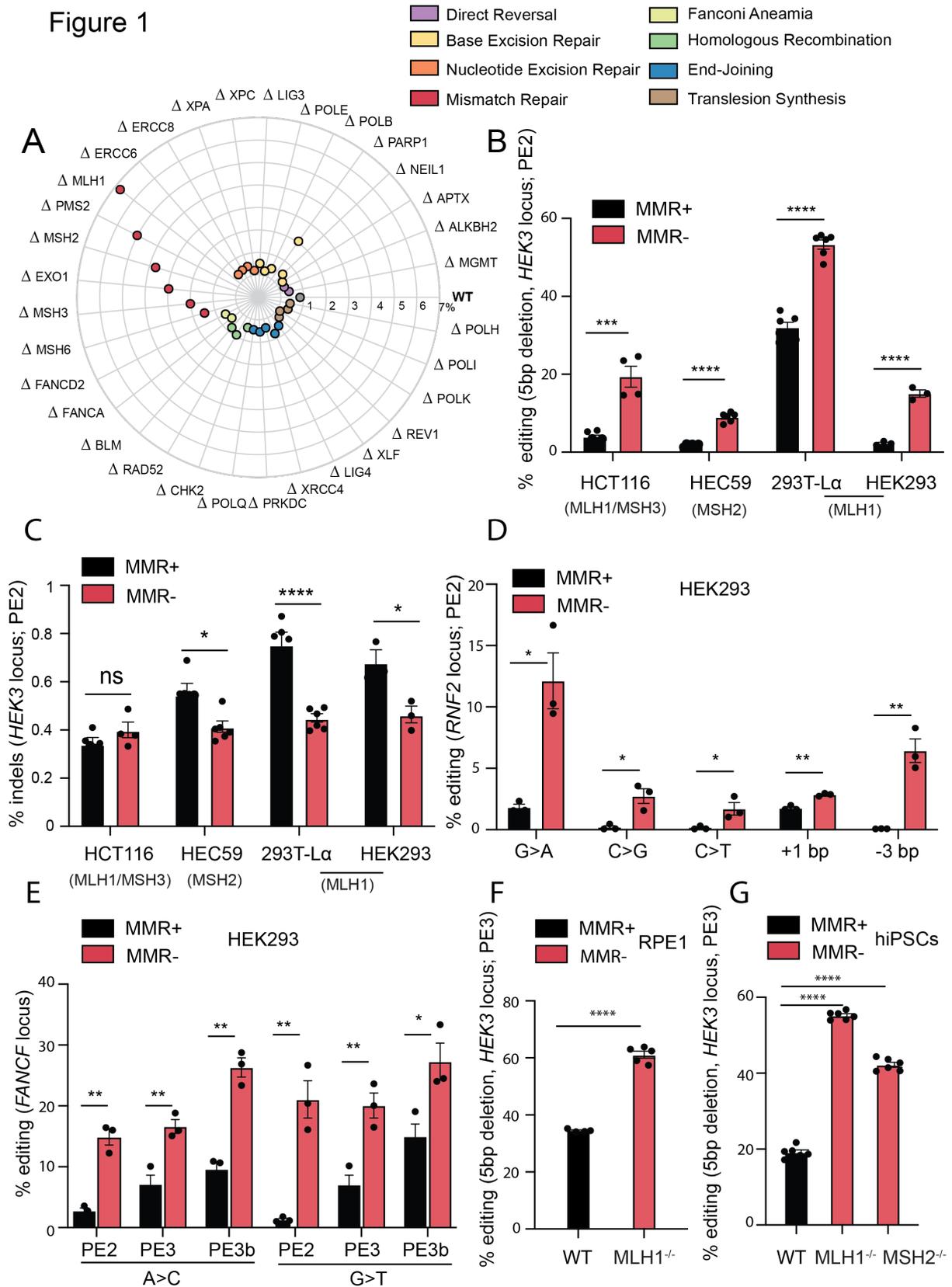
96 To investigate the DNA repair requirements for PE, we conducted a targeted genetic screen,
97 utilising a collection of isogenic knockouts in the human near-haploid HAP1 cell line
98 (**Supplementary Data 1**). The 32 targeted genes were selected to represent divergent
99 functions within all known human DNA repair pathways. The library thus provided a
100 comprehensive coverage of the DNA damage response. The cell lines received the PE
101 machinery, including the Cas9(H840A)-RT and a pegRNA encoding a 5-base pair (bp)
102 deletion in the *HEK3* locus. PE efficiency was determined by amplicon sequencing of the
103 genomic locus.

104 Wild-type HAP1 cells were remarkably inefficient at PE (<1% alleles edited). In contrast,
105 isogenic HAP1 cell lines mutated at the *MLH1*, *PMS2*, *MSH2*, *EXO1* and *MSH3* loci displayed
106 higher PE levels, ranging from 2 to 6.8-fold (**Figure 1A**). Disruption of other DNA repair
107 pathways had little or no impact on PE efficiency. This finding clearly indicates that MMR

108 functions to inhibit PE. Of all MMR genes targeted in the screen, only the loss of MSH6 failed
109 to increase PE efficiency.

110 The MMR pathway evolved to correct base/base mispairs and small indels arising in DNA
111 during replication and recombination. To initiate repair, these lesions are recognised by the
112 heterodimers MutS α (MSH2-MSH6) or MutS β (MSH2-MSH3). Whereas MutS α recognises
113 base/base mismatches and indels of 1-2 nucleotides, larger indels are recognised by MutS β
114 (Drummond et al., 1995; Palombo et al., 1995, 1996; Acharya et al., 1996; Gradia et al., 1997).
115 Substrate binding brings about an ATP-dependent conformational change of the MutS
116 complexes and recruitment of the MutL α (MLH1-PMS2) (Li and Modrich, 1995) or MutL
117 β (MLH1-MLH3) (Lipkin et al., 2000) heterodimers. Assembly of the MutL complex together
118 with RFC and PCNA (Pluciennik et al., 2010), bound at a pre-existing strand discontinuity
119 (either a nick or a free 3' terminus), activates cryptic endonucleases of the PMS2 or MLH3
120 proteins, which then introduce additional DNA single-strand breaks (SSBs) into the
121 discontinuous DNA strand, in the vicinity of the mismatch. These SSBs act as entry points for
122 EXO1, which degrades the discontinuous strand in a 5' to 3' direction up to, and some distance
123 past, the misincorporated nucleotide(s) (Kadyrov et al., 2006). The resulting gap is filled-in by
124 DNA polymerase δ and the remaining nick is ligated by DNA ligase I (Stojic et al., 2004; Iyer
125 et al., 2006; Fishel, 2015). Since the edit introduced in our screen is a 5 bp deletion, this makes
126 it a substrate of MutS β , but not MutS α (Palombo et al., 1996), which explains the lack of an
127 effect on editing upon the loss of MSH6 (**Figure 1A**). This result highlights the highly
128 specialised nature of the DNA damage response that functions on different substrates.

Figure 1



129

130 **Figure 1: Mismatch repair inhibits prime editing in human cells. A)** Genetic screen in 32
 131 HAP1 isogenic knockout cell lines covering different DNA damage repair pathways, as well as
 132 their wild-type (WT) counterpart, showing the efficiency of installation of a 5 base pair (bp)

133 deletion in the *HEK3* locus, using PE2. Values correspond to editing efficiency, measured by
134 amplicon sequencing analysis in two independent biological replicates, with two technical
135 replicates each. Each radial line represents an increment of 1%. **B)** PE2 of a 5 bp deletion in
136 the *HEK3* locus in the indicated mismatch repair-deficient cell lines (MMR-), and their
137 respective complemented counterparts (MMR+). For each cell line, the mutated MMR genes
138 are represented. Values correspond to editing efficiency measured by amplicon sequencing
139 analysis in at least three independent biological replicates. **C)** Percentage of indels after
140 inducing a 5 bp deletion in the *HEK3* locus using PE2 in varying mismatch repair-deficient,
141 and their respective complemented, cell lines, in at least three independent biological
142 replicates. Percentage of indels = number of reads containing indels that do not correspond
143 to the desired edit/total number of aligned reads. For each cell line, the correspondent mutated
144 MMR gene is indicated. **D)** PE2 of the indicated types of edits (*RNF2* locus) in HEK293 cells
145 wild-type (MMR+), or knockout for MLH1 (MMR-). Values correspond to editing efficiency,
146 measured by amplicon sequencing analysis in three independent biological replicates. **E)**
147 Efficiency of PE2, PE3 and PE3b after inducing A>C or G>T mutations in the *FANCF* locus.
148 HEK293 wild-type (MMR+) and MLH1 knockout cells (MMR-) were used. Values correspond
149 to editing efficiency, measured by amplicon sequencing analysis in three independent
150 biological replicates. **F)** PE3 efficiency of a 5 bp deletion in the *HEK3* locus in RPE1 wild-type
151 (WT) cells and an isogenic knockout MLH1 cell line (RPE1-MLH1^{-/-}), determined by Sanger
152 sequencing and TIDE analysis, for at least three independent biological replicates. RPE1 cells
153 express Cas9(H840A)-RT in a constitutive manner (RPE1 PE2-BSD). **G)** PE3 efficiency in
154 wild-type (WT) human induced-pluripotent stem cells (hiPSCs), as well as isogenic knockouts
155 for MLH1 and MSH2 (MLH1^{-/-}, MSH2^{-/-}), in three independent biological replicates, with two
156 technical replicates each. Statistical analysis using multiple unpaired t tests. Error bars reflect
157 mean and SEM. Ns p-value non-significant; * p-value < 0.05; ** p-value < 0.01; *** p-
158 value<0.001; **** p-value < 0.0001.

159

160 **Mismatch repair hinders PE2 and PE3 across several human cell lines, genomic loci** 161 **and edit types**

162 To further explore the inhibitory role of MMR in PE, we expanded our investigations to a panel
163 of MMR-deficient human cell lines, alongside their complemented counterparts, in which we
164 measured the editing efficiency and fidelity of the *HEK3* locus. We used the colorectal cancer
165 line HCT116, which is mutated in both *MSH3* and *MLH1*, alongside the MMR-proficient
166 HCT116 cell line complemented with chromosomes 5 and 3 that house the wild-type copies
167 of the two genes, respectively (**Supplementary Figure 1A**) (Koi et al., 1994; Haugen et al.,

168 2008). The endometrial adenocarcinoma cell line HEC59, which is mutated at
169 the *MSH2* locus, was used together with its MMR-proficient counterpart complemented with
170 chromosome 2 that carries the wild-type *MSH2* gene (**Supplementary Figure 1B**) (Umar et
171 al., 1997). Additionally, we used a doxycycline-inducible model of MLH1 deficiency in the
172 embryonic kidney cell line HEK293T (293T-L α) (**Supplementary Figure 1C**) (Cejka et al.,
173 2003). Finally, we generated an isogenic pair of MLH1 wild-type and knockout HEK293 cells
174 (**Supplementary Figure 1D**).

175

176 We controlled for the transfection efficiencies of all the matched MMR-deficient and proficient
177 cell line pairs and showed that these were comparable, as measured by the percentage of
178 cells transfected with a GFP expressing plasmid (**Supplementary Figure 1E**). We then
179 performed PE2 editing by deleting 5 bp within the *HEK3* locus. All cell lines showed
180 significantly increased PE2 editing (ranging from 1.7 to 6.6-fold) when MMR was ablated,
181 compared to their MMR-proficient counterparts (**Figure 1B**). Importantly, even though PE
182 efficiencies were increased by MMR deficiency, this did not come at the expense of higher
183 indel frequencies within the amplicon region (**Figure 1C**). Indeed, we observed that loss of
184 MMR prevented unwanted indels at the *HEK3* locus in the HEC59, 293T-L α and HEK293 cell
185 lines (**Figure 1C**).

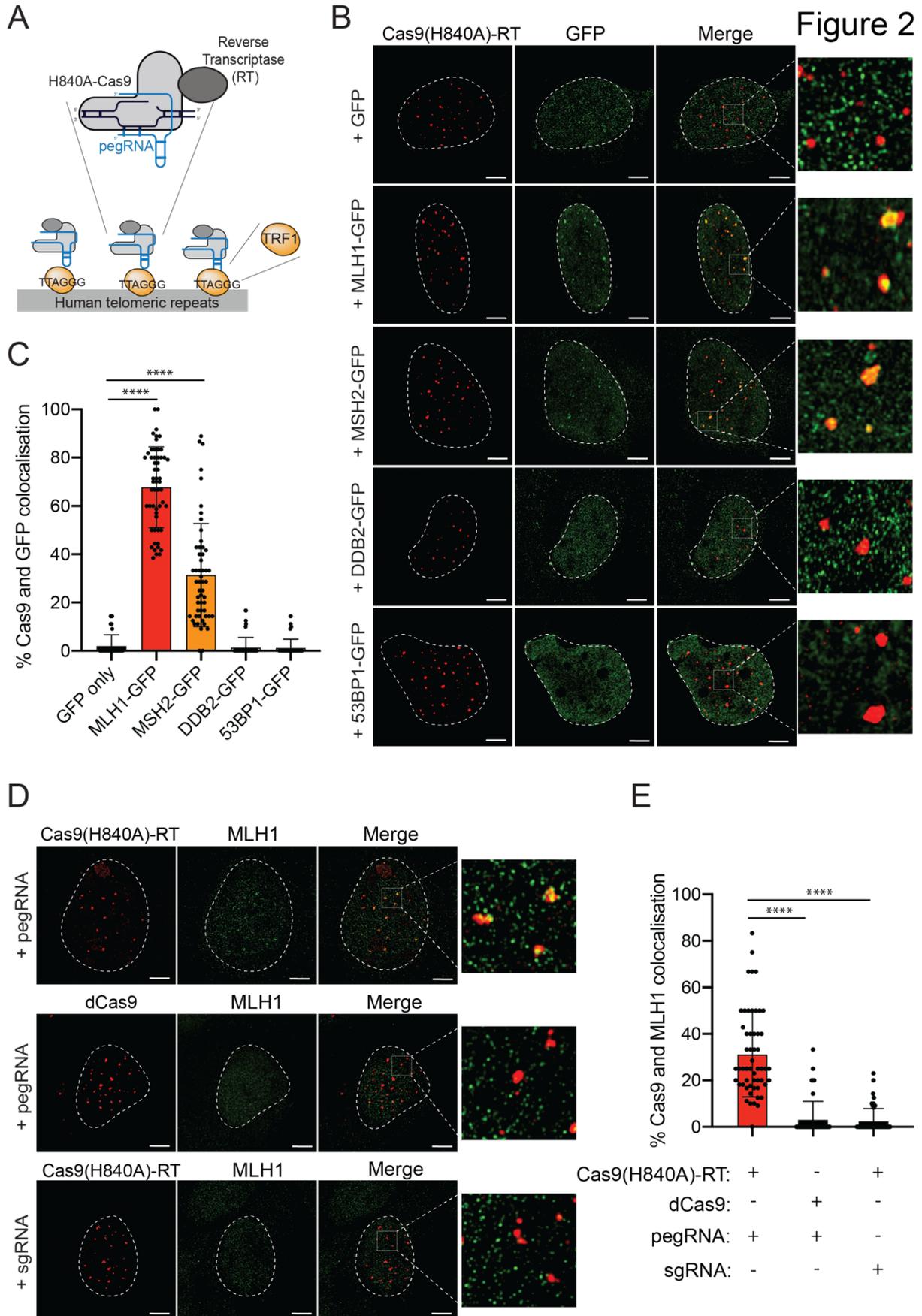
186 To further investigate the substrates of MMR in PE, we measured the editing efficiencies of a
187 transition (G>A), two transversions (C>G and C>T), a 1 bp insertion and a 3 bp deletion, all
188 within a different endogenous locus, the *RNF2* locus. We found that active MMR significantly
189 diminished the efficiency of all these edits, ranging from 1.6 to 14-fold, using HEK293 cells
190 that lack MLH1, a factor that is part of both the MutL α and MutL β heterodimers, which together
191 repair base/base mismatches, indels of 1-2 nucleotides and larger indels (**Figure 1D**). These
192 findings were also corroborated in the MLH1/*MSH3*-deficient HCT116 cell line
193 (**Supplementary Figure 1F**). To test the inhibitory role of MMR on different PE strategies, we
194 measured the efficiency of PE2, PE3 and PE3b on the *FANCF* locus, via the installation of
195 either an A>C or a G>T substitution in HEK293 and HCT116 cells. Editing efficiency was
196 improved by MMR deficiency for all types of PE (1.8 -16-fold), albeit to a lesser extent for PE3
197 (**Figure 1E, Supplementary Figure 1G**). Overall, these results show that MMR counteracts
198 PE efficiency across different edits and different genomic loci, in various human cell lines.

199 Since both HCT116 and HEC59 are cancer-derived cell lines that display MMR deficiency, it
200 is possible that the higher levels of PE efficiency are due to cellular adaptation. The human
201 retinal pigmental cell line RPE1 is a non-cancer derived cell line, thus we utilised this for
202 corroborating our findings. PE efficiencies are generally very low in RPE1 wild-type cells

203 **(Supplementary Figure 1H)**. To overcome this shortcoming, we developed a lentivirus
204 system for stable delivery of the PE3 system, where RPE1 cells constitutively
205 express Cas9(H840A)-RT (denoted RPE1 PE2-BSD). We generated a CRISPR genetic
206 knockout for the MLH1 factor in this cell line (**Supplementary Figure 1I**) and performed PE3,
207 by transducing both the pegRNA and the nicking sgRNA installing a 5 bp deletion within the
208 *HEK3* locus. We observed an editing efficiency of approximately 35% in WT RPE1 that was
209 further increased to 60% in RPE1-MLH1^{-/-} (**Figure 1F**). We additionally extended our findings
210 to human induced-pluripotent stem cells (hiPSCs), engineered to be deficient for either MLH1
211 or MSH2 (Zou et al., 2021) (**Supplementary Figure 1J**). Wild-type hiPSCs demonstrated
212 20% editing efficiency of a 5 bp deletion in the *HEK3* locus, while the MLH1 and MSH2
213 deficient counterparts displayed an increased efficiency of approximately 55% and 40%,
214 respectively (**Figure 1G**). Overall, these results confirm that the MMR pathway specifically
215 plays a role in counteracting PE.

216 **Mismatch repair factors are recruited to sites of prime editing**

217 To confirm that the MMR proteins are directly involved in the processing of PE intermediates,
218 we determined if they are recruited to sites of ongoing editing marked by Cas9(H840A)-RT.
219 Cas9(H840A)-RT was directed to human repetitive telomeric regions, a strategy that has
220 proven efficient for imaging Cas9 (Chen et al., 2013) (**Figure 2A**). Using this experimental
221 approach, we were able to colocalize TRF1 (an essential component of the telomeric shelterin
222 complex) with catalytically inactive Cas9 (dCas9), as previously described (Wang et al., 2008)
223 and also with Cas9(H840A)-RT (**Supplementary Figure 2A-B**). Therefore, this setup allows
224 for the visualisation of genomic loci undergoing PE, in a pegRNA-dependent manner. Next,
225 we used this system in U2OS cells to express Green Fluorescent Protein (GFP), or GFP-
226 tagged MMR proteins, as well as two additional DNA repair proteins that do not function in
227 MMR (DDB2 that functions in nucleotide excision repair and 53BP1 that promotes non-
228 homologous end-joining). We observed that 65% of MLH1-GFP foci and 25% of MSH2-GFP
229 foci colocalised with Cas9(H840A)-RT foci (**Figure 2B-C**). Importantly, we did not observe
230 colocalisation of either DDB2-GFP or 53BP1-GFP foci and Cas9(H840A)-RT foci (**Figure 2B-**
231 **C**). Furthermore, by using an antibody against MLH1 (**Supplementary Figure 2C-D**) we
232 confirmed the localisation of endogenous MLH1 to sites of PE (**Figure 2D-E**). We found that
233 30% of MLH1 foci colocalised with Cas9, while we did not observe colocalisation when a
234 dCas9, or a sgRNA, were used (**Figure 2D-E**). These findings reveal that intermediates of PE
235 are substrates of MMR and we propose that MMR functions to degrade the invading
236 heterologous strand and thus restore the original DNA sequence.



237

238

239 **Figure 2: The mismatch repair protein MLH1 localises to sites of active prime editing.**

240 **A)** Scheme of the setup used for imaging. Cas9(H840A)-RT is targeted, through a pegRNA,
241 to human telomeric repetitive regions. TRF1 is a telomeric protein that binds to these regions.
242 **B)** Representative super-resolution images of Cas9(H840A)-RT and the indicated GFP-
243 tagged DNA repair proteins, in U2OS cells 24 hours after reverse transfection with GFP or
244 GFP-tagged MLH1, MSH2, DDB2 or 53BP1, as well as a pegRNA targeting telomeric repeats.
245 Data from three biological replicates and at least 50 cells per condition. **C)** Quantification of B
246 indicating colocalization of Cas9(H840A)-RT foci with GFP foci. **D)** Representative super-
247 resolution images of Cas9(H840A)-RT, or dCas9, and MLH1 with a pegRNA or a sgRNA
248 targeting telomeric repeats. Data from three biological replicates and at least 50 cells per
249 condition. **E)** Quantification of D, indicating colocalization of Cas9(H840A)-RT foci with MLH1
250 foci. Statistical analysis using multiple unpaired t tests. Error bars reflect mean and SEM. ****,
251 p-value < 0.0001. All scale bars, 5 μ m.

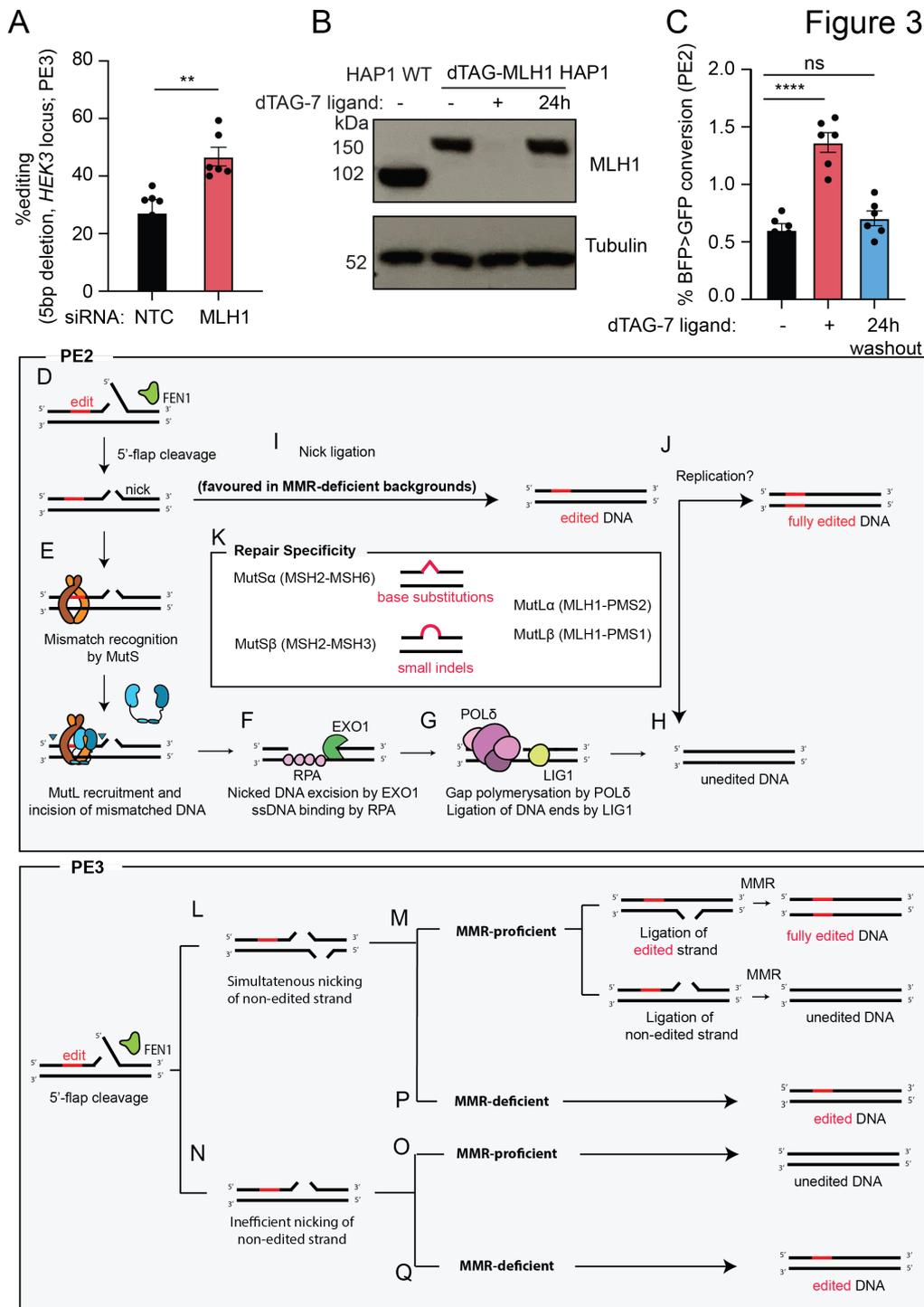
252

253 **Reversible mismatch repair depletion can be exploited to increase prime editing**
254 **efficiency**

255 We next sought to transiently deplete MLH1 as a strategy to improve PE efficiency. Since loss
256 of MMR leads to increased mutational burden and genome instability (Jiricny, 2006), long-
257 term inhibition of MMR is not desirable. Thus, to achieve transient MMR ablation, we depleted
258 MLH1 in HEK293 cells with a pool of siRNAs (**Supplementary Figure 3A**), and subsequently
259 showed that this effectively increased PE efficiency by approximately 2-fold through the
260 generation of a 5 bp deletion in the *HEK3* locus (**Figure 3A**).

261 An alternative approach for achieving transient loss-of-function is through targeted protein
262 degradation. The degradation tag (dTAG) system has proved to be an efficient strategy for
263 rapid and transient ligand-induced targeted protein degradation (Nabet et al., 2018). Using
264 CRISPR-mediated knock-in, we introduced the dTAG into the *MLH1* locus of HAP1 cells,
265 which allowed for the targeted degradation of MLH1 after treatment with the dTAG ligand.
266 Importantly, the protein levels of MLH1 were restored to those found in wild-type cells after 24
267 hours of removal of the ligand (**Figure 3B**). Using a flow cytometry-based readout, in which
268 the pegRNA encodes a 1 bp substitution that converts the Blue-Fluorescent Protein (BFP) to
269 GFP, we observed a 3-fold increase in BFP to GFP conversion upon treatment of the cells
270 with the dTAG ligand and subsequent endogenous degradation of MLH1 (**Figure 3C**). PE
271 efficiency at the *HEK3* locus through a 5 bp deletion, as measured by sequencing genomic
272 DNA, was also significantly increased by 3-fold upon treatment with the dTAG-ligand

273 **(Supplementary Figure 3B)**. Importantly, this effect could be rescued to wild-type levels by
 274 removing the dTAG-ligand from the medium, thus restoring MLH1 levels (**Figure 3B-C**). Taken
 275 together, these results indicate that transient ablation of MMR represents a promising strategy
 276 that can be used to increase PE efficiency.



277

278 **Figure 3: Reversible ablation of MLH1 can be exploited to increase prime editing**
 279 **efficiency** A) PE efficiency of a 5 bp deletion in the *HEK3* locus in HEK293 cells transfected

280 with non-targeting control (NTC) or MLH1 siRNA pools, two days prior to PE vector delivery.
281 Values correspond to editing efficiency, measured by Sanger sequencing and analysed by
282 TIDE (Brinkman et al., 2014), in three independent biological replicates, with two technical
283 replicates each. **B)** Immunoblot for MLH1 and tubulin in HAP1 cell extracts with ('+') or without
284 ('-') a 1-hour treatment with 500 nM dTAG-7 ligand. Recovery of dTAG-MLH1 expression was
285 measured 24 hours after ligand removal ('24h'). **C)** PE2 efficiency of BFP>GFP conversion in
286 dTAG-MLH1 HAP1 cells, untreated ('-'), or treated with 500 nM of dTAG-7 ligand for 4 days
287 ('+'), or treated with dTAG-7 ligand for 24 hours, followed by its removal for 3 days ('24h
288 washout'). Data measured by flow cytometry for three biological replicates with two technical
289 replicates each. **D-Q)** Schematic model of MMR activity in counteracting PE efficiency. After
290 the cleavage of the non-edited 5'-flap by the flap endonuclease (FEN1), a nick is installed in
291 the edited strand (D). This nick is recognised as a mismatch by the MutS complex, after which
292 MutL is recruited and catalyses incisions that flank the mismatch (E). Exonuclease 1 (EXO1)
293 degrades the incised DNA and Replication protein A (RPA) coats the single-stranded DNA
294 (ssDNA) (F). Polymerase δ fills the gap and Ligase 1 (LIG1) ligates the nick (G). This repair
295 culminates in an unedited DNA molecule (H). If the nick is ligated before mismatch recognition,
296 an heteroduplex DNA is generated, containing the edit in one of the strands (I). The resolution
297 of this heteroduplex potentially relies on replication (J). In PE3, the non-edited strand is
298 simultaneously nicked (L), which directs MMR to repair the mismatch depending on which
299 strand is ligated first (M). However, it is possible that the nicking of the non-edited strand is
300 inefficient, leading to the MMR-mediated removal of the edit (O). In MMR-deficient
301 backgrounds, there is ligation of the heteroduplex without removal of the edit (P,Q). Statistical
302 analysis using unpaired t tests. Error bars reflect mean and SEM. Ns, p-value non-significant;
303 **, p-value < 0.01;****, p-value < 0.0001.

304 **Discussion**

305 Here we show that the MMR pathway counteracts PE efficiency and fidelity, across different
306 human immortalised and induced pluripotent stem cell lines, genomic loci and edit types.
307 Although the role of MMR in PE had not been addressed experimentally, it was hypothesised
308 to be required for the resolution of the heteroduplex DNA, thus promoting repair of the non-
309 edited strand by utilising the edited strand as template (Petri et al., 2021; Scholefield and
310 Harrison, 2021). Our results provide clear evidence to the contrary, namely that the MMR
311 system functions on the PE intermediate by degrading the invading, RT-synthesised strand
312 to restore the original sequence. This outcome conforms to our understanding of the molecular
313 mechanism of MMR as gleaned from in vitro systems that made use of circular heteroduplex
314 substrates and extracts of human cells (Holmes et al., 1990; Thomas et al., 1991). On these

315 substrates, activation of MMR was strictly dependent on the presence of two factors: a
316 mismatch and a pre-existing nick in one strand that was less than 1 kb distant. The repair
317 process was then directed to the nicked strand. Following the extrapolation of these insights
318 into a cellular setting, MMR, activated by misincorporated nucleotides during replication, would
319 be initiated either by the mismatch/indel and either end of an Okazaki fragment in the lagging
320 strand, or the 3'-end of the primer of the leading strand. During recombination of homologous
321 but non-identical fragments, MMR would be initiated by the heterology (mismatch or indel)
322 between the invading donor and the recipient DNA strands, with the 3'-terminus of the invading
323 strand acting as the signal required to activate the MutL endonucleases.

324

325 We speculate that PE2 resembles the latter mechanism, whereby the RT-synthesised 3' flap
326 would displace the 5' terminus of the Cas9(H840A)-RT-generated nick. This would give rise
327 to a 5' flap, which could be cleaved off by one of several structure-specific endonucleases
328 (SLX1, FAN1, DNA2) and finally trimmed by flap endonuclease 1 (FEN1) (**Figure 3D**). Binding
329 of MutS at the mismatch and its interaction with MutL and RFC/PCNA bound at the 3'-terminus
330 (not shown) would activate the MutL nickase to generate additional incisions flanking the edit
331 (**Figure 3E**). Exonuclease 1 (EXO1) would then degrade the discontinuous strand to generate
332 a long single-stranded gap bound by Replication Protein A (RPA) (**Figure 3F**). Finally,
333 polymerase δ (POL δ) would fill the gap and Ligase 1 (LIG1) ligate the nick (**Figure 3G**). This
334 process would result in the removal of the edit (**Figure 3H**). Ligation of the nick (**Figure 3D**)
335 prior to MMR activation (**Figure 3I**) would generate an heteroduplex with one edited and one
336 non-edited strand (**Figure 3I**), which would be refractory to MMR, and would persist until
337 replication, which would give rise to 50% progeny carrying the edit and 50% non-edited
338 (**Figure 3J**). The path in **Figure 3 D-I-J** would be favoured in the absence of MMR, thus
339 accounting for the increased yield of edited alleles in MMR-deficient backgrounds.

340

341 Importantly, our data confirm the results of biochemical characterisations of the substrate
342 specificities of MutS α and MutS β (Drummond et al., 1995; Palombo et al., 1995, 1996;
343 Acharya et al., 1996; Gradia et al., 1997), which showed that the former recognises
344 preferentially base/base mismatches and indels of 1-2 nucleotides, whereas the latter binds
345 to larger indels (**Figure 3K**). Based on these findings, deletion of *MSH6* should have failed to
346 affect the outcome of PE using a 5 bp deletion, which was indeed the case, as we report here
347 (**Figure 1A**).

348

349 Besides types of edits, different PE strategies are likely to be impacted by MMR activity to
350 different degrees. PE3 was developed as a more efficient PE strategy, in which both edited
351 and non-edited DNA strands are nicked (**Figure 3L**). When nicks are present in both strands,

352 the nick nearer the mismatch will be preferentially deployed by the MMR system, but the
353 excision will destabilize the duplex and may lead to a DNA DSB, which explains the increased
354 presence of indels in the final outcome of PE3 (Anzalone et al., 2019). If one strand of the
355 heteroduplex is ligated first, MMR is directed to the nicked strand. Therefore, ligation of the
356 edited strand directs MMR to repair the non-edited strand leading to editing on both strands
357 of the DNA heteroduplex, whereas ligation of the non-edited strand results in an unedited DNA
358 molecule (**Figure 3M**). These outcomes rely on the assumption that the nicking sgRNA acts
359 as efficiently as the pegRNA. This might not always be the case as some DNA molecules
360 might have been edited and nicked by the pegRNA only (**Figure 3N**). This would lead to the
361 same outcome as PE2, which is removal of the edit by MMR (**Figure 3O**). In PE3, as well as
362 PE2, cells that lack functional MMR ligate the heteroduplex DNA without removal of the edit
363 (**Figure 3P-Q**). Thus, we propose that MMR activity counteracts PE3 efficiency, as well as
364 PE2, albeit to a lesser extent. This difference is due to the loss of a clear discrimination signal
365 of which strand to repair, created by the nick. In PE3b, the nicking of the non-edited strand is
366 designed to happen only after the integration of the edit. Hence, we propose that the MMR
367 dependency of PE3b is the same as PE2.

368
369 It remains to be seen what the size limitation of PE-generated indels is, that are addressed by
370 MMR. While our work was under revision, two other studies described the suppression of PE
371 efficiency by MMR activity and extensively characterised the types of edits that are efficiently
372 repaired by this pathway. Chen and colleagues showed that MMR involvement decreases with
373 increasing indel size and that G/C to C/G edits, which form C:C mismatches, are less
374 frequently removed by MMR factors (Chen et al., 2021). Koepfel and colleagues
375 systematically measured the insertion efficiency of indels ranging from 1-69 bp in length
376 (Koepfel et al., 2021). The authors observed an overall increase of insertion efficiency upon
377 MMR depletion, with the greatest difference seen for indels 1-4 bp long. These results agree
378 with the known substrate specificities of MutS α and MutS β (Drummond et al., 1995; Palombo
379 et al., 1995, 1996; Acharya et al., 1996; Gradia et al., 1997). Given that one of the most
380 promising applications of PE includes insertion, deletion or replacement of large sequences
381 of DNA, for example to tag endogenous loci within the genome (Anzalone et al., 2019; Ioannidi
382 et al., 2021), how these lesions are processed and how their insertion efficiency can be
383 augmented, remains of substantial future interest.

384
385 Our findings suggest that the improvement in PE efficiencies in the absence of MMR does not
386 come at the cost of generation of undesirable indels around the edit site (**Figure 1C**). However,
387 MMR deficiency brings about a mutator phenotype, which will severely limit the utility of PE
388 protocols that make use of long-term MMR inactivation. This deleterious outcome might be

389 substantially reduced by interfering with MMR transiently. Our results suggest that targeting
390 MMR factors with siRNA or protein degradation technologies, such as proteolysis targeting
391 chimeras (PROTAC), represent promising approaches to improve PE efficiencies. Another
392 exciting approach would be to interfere with MMR solely at the edit site, similarly to what had
393 been described for improving the efficiency of HDR (Charpentier et al., 2018; Rees et al.,
394 2019). While our article was under revision, two new PE strategies were described, PE4 and
395 PE5, that rely on co-expressing dominant negative MLH1 fragments with the PE2 and PE3
396 machineries, respectively (Chen et al., 2021). The authors also reported that pegRNAs
397 encoding contiguous silent or benign mutations around the intended edit function to evade
398 recognition and repair by MMR (Chen et al., 2021). This strategy has the potential to improve
399 PE efficiency without the increase in mutational burden that is associated with long-term MMR
400 loss.

401

402 Together with recent reports (Chen et al., 2021; Koepfel et al., 2021), our data shed new light
403 on the molecular mechanism of a new and highly promising genome editing technology. We
404 have shown that the MMR pathway inhibits PE efficiency by physically localising to edit sites
405 and promoting their reversion to non-edited sequences. However, the variability in PE
406 observed across cell lines cannot be explained solely by the involvement of MMR and other
407 factors such as cell cycle stage (Wang et al., 2021) or cellular metabolism might also be
408 contributing factors. Hence, further studies are warranted to identify alternative cellular
409 determinants that might limit or promote the use of this technology. The advancement in
410 knowledge reported here can be applied to further the development of prime editors, as well
411 as in the design of novel therapeutic strategies.

412

413 **Methods**

414

415 **Plasmids and oligos**

416 DNA oligos were obtained from Integrated DNA Technologies (IDT) unless otherwise noted.
417 pCMV-PE2 was a gift from David Liu (Addgene plasmid # 132775). pLenti PE2-BSD was a
418 gift from Hyongbum Kim (Addgene plasmid # 161514) (Kim et al., 2021). pU6-pegRNA-GG-
419 acceptor was a gift from David Liu (Addgene plasmid # 132777). PegRNAs were cloned into
420 the pU6-pegRNA-GG-acceptor using BsaI Golden Gate assembly (NEB), following the
421 manufacturer's instructions. sgRNAs utilised in PE3 and PE3b experiments were cloned in the
422 lenti-sgRNA puro vector, using BsmBI Golden Gate assembly (NEB), following the
423 manufacturer's instructions. lenti-sgRNA puro was a gift from Brett Stringer (Addgene
424 plasmid# 104990) (Stringer et al., 2019). lenti-sgRNA neo was a gift from Brett Stringer

425 (Addgene plasmid # 104992) and it was used to clone the sgRNA utilised in PE3 experiments
426 in RPE1 PE2-BSD cells.

427 For immunofluorescence experiments, the dCas9 plasmid was a gift from David Segal
428 (Addgene plasmid # 100091) (O'Geen et al., 2017) and the pSLQ1651-sgTelomere(F+E) was
429 a gift from Bo Huang & Stanley Qi (Addgene plasmid # 51024) (Chen et al., 2013). Additionally,
430 the following plasmids were used: pCMVTet-eGFP-MLH1, pCMVTet-eGFP-MSH2, pEGFP-
431 C1-53BP1. pLenti6.3 WT GFP-DDB2 was also used and it was a gift from Dr. A. Pines (Pines
432 et al., 2012). pmaxGFPTM (Lonza) was used for immunofluorescence experiments, as well as
433 to test transfection efficiency. pCRIS-PITChv2-BSD-dTAG (BRD4), used for the generation of
434 dTAG expressing cells, was a gift from Dr. Georg Winter.

435 BFP-positive cells were generated using the BFP dest clone plasmid. BFP dest clone was a
436 gift from Jacob Corn (Addgene plasmid # 71825) (Richardson et al., 2016).

437 Sequences of sgRNA, pegRNA constructs, as well as primers for genomic DNA amplification
438 are listed in **Supplementary Data 2**. The pegRNA targeting telomeres included a stem loop
439 extension as described in (Chen et al., 2013). All plasmids for mammalian cell experiments
440 were purified using the Plasmid Plus Midi Kit (Qiagen) or the Spin Miniprep Kit (Qiagen), both
441 including endotoxin removal steps.

442 For virus production, the psPAX2 and VSV.G packaging virus were used. psPAX2 was a gift
443 from Didier Trono (Addgene plasmid # 12260). VSV.G was a gift from Tannishtha Reya
444 (Addgene plasmid # 14888).

445

446 **Construction of plentipegRNAPuro vector**

447 The plentipegRNAPuro vector was generated as follows. The lenti-sgRNA puro vector was
448 digested with EcoRI for 2h at 37°C followed by digestion with BsmBI for 2 hours at 55°C and
449 treatment with 4 µl of rSAP (NEB) for 1 hour at 37°C. The mRFP and terminator sequence
450 present in the pU6-pegRNA-GG-acceptor was PCR amplified with a forward primer converting
451 the Bsal cut site to BsmBI and with the reverse primer containing an EcoRI cut site. The PCR
452 product was digested with BsmBI and EcoRI as above. The vector and digest were both
453 purified using gel extraction using the Wizard® SV Gel and PCR Clean-Up System (Promega)
454 and ligated using T4 ligase (NEB) for 1h at room temperature. In order to allow for Golden
455 gate cloning using BsmBI, the Bsal cut site present in the newly assembled vector was
456 converted to a BsmBI cut site using the Q5 Site- Directed Mutagenesis kit (NEB).

457

458 **Lentiviral production and transduction**

459 Lentiviral production was achieved by plating 5×10^6 xLentiTM cells (Oxgene) in a 10-cm dish
460 transfected one day post seeding with packaging plasmids (1 µg VSV.G, 2 µg psPAX2 and 4

461 µg of transfer plasmid using PEI (Sigma-Aldrich). Virus containing supernatant was collected
462 72 hours post transfection, cleared by centrifugation and stored at -80°C.

463 Cell transduction was performed using spin-infection as follows. 0.5×10^6 cells were mixed in a
464 well of a 12-well plate with varying concentrations of supernatant containing viral particles and
465 8 µg/ml of polybrene (Sigma) which was then centrifuged at 2,000 rpm for 30 minutes at 30°C.

466

467 **Mammalian cell culture**

468 All cells were grown at 3% oxygen at 37°C and routinely checked for possible mycoplasma
469 contamination. Human HAP1 cells were obtained from Horizon Discovery and were grown in
470 Iscove's Modified Dulbecco's Medium (IMDM) (Gibco), containing L-glutamine and 25 nM
471 HEPES and supplemented with 10% Fetal Bovine Serum (FBS) (Gibco) and 1%
472 Penicillin/Streptomycin (P/S) (Sigma-Aldrich). U2OS and HEK293 cells were purchased from
473 ATCC cell repository and cultured in DMEM (Gibco), supplemented with 10% FBS and 1%
474 P/S. HEC59, wild-type and complemented with chromosome 2, were cultured in F12 DMEM
475 with 10% FBS and 1% P/S. HEC59 complemented cells were cultured with 400 µg/mL of
476 geneticin (G418, Gibco). HCT116 cells, wild-type and complemented with both chromosomes
477 3 and 5, were cultured with McCoy's 5A medium (Gibco), with 10% FBS and 1% P/S. HCT116
478 cells complemented with chromosomes 3 and 5 were cultured with 400 µg/mL geneticin
479 (G418, Gibco) and 6 µg/mL blasticidin (Invivogen). 293T-Lα were cultured in DMEM medium
480 (Gibco) supplemented with 10% FBS or Tet-system approved FBS (Takara Bio), 1% P/S, 100
481 µg/mL zeocin (Gibco) and 300 µg/mL hygromycin (Gibco). 293T-Lα were grown in doxycycline
482 (1 µg/mL) for 7 days before any experiment, to completely deplete MLH1 expression.
483 Doxycycline was replenished in the medium every 2 days. RPE1 cells were a gift from the
484 Jackson lab (Gurdon Institute, Cambridge, UK) and cultured in F12 DMEM with 10% FBS and
485 1% P/S. iPSCs (WT, MLH1 and MSH2-deficient) were a gift from the Nik-Zainal lab (University
486 of Cambridge, UK) and cultured on non-tissue culture treated plates (Stem Cell Technologies)
487 pre-coated with 10 µg/mL Vitronectin XF (Stem Cell Technologies) in TeSR-E8 medium (Stem
488 Cell Technologies). The medium was changed daily and the cells were passaged every 4–8
489 days depending on confluency using Gentle Cell Dissociation Reagent (Stem Cell
490 Technologies). 10 µM of ROCKi (Stem Cell Technologies) was added to the medium
491 whenever passaging or thawing iPSCs.

492

493 **Generation of PE2-BSD RPE1 and HEK293T cells**

494 RPE1 and HEK293T cells were transduced at a low multiplicity of infection with the pLenti-
495 PE2-BSD vector and selected two days post transduction with blasticidin (10 µg/ml).
496 Transduced cells were then single cell sorted into 96 well plates and single colonies isolated

497 following 2-3 weeks of clonal expansion. Cas9(H840A)-RT expression was confirmed by
498 immunoblotting.

499

500 **Generation of MLH1 isogenic knockout cell lines**

501 MLH1 knockouts were generated in RPE1 PE2-BSD and HEK293 cell lines by nucleofection
502 of *S.pyogenes* Cas9 together with an in-vitro transcribed sgRNA. Recombinant Cas9
503 containing a nuclear localization sequence and a C-terminal 6-His tag was purchased from
504 Integrated DNA Technologies (#1081059). The sgRNA targeting MLH1 (**Supplementary Data**
505 **2**) was designed utilizing the VBC score tool (<https://www.vbc-score.org/>). T7 in vitro
506 transcription was performed using HiScribe (NEB E2050S), using PCR-generated DNA as
507 template, as previously described here: [dx.doi.org/10.17504/protocols.io.bqjbmuin](https://doi.org/10.17504/protocols.io.bqjbmuin).

508 The 4D-Nucleofector System X-Unit (Lonza) was used for nucleofection. A mixture of 30 pmol
509 of Cas9 and 60 pmol of in vitro transcribed sgRNA was prepared in a final volume of 5 μ L of
510 Cas9 buffer (20 mM HEPES-KOH pH 7.5, 150 mM KCl, 10% glycerol) and incubated for 20
511 minutes, room-temperature. 200,000 HEK293 or RPE1 cells were centrifuged (800 g, 8
512 minutes), washed with PBS and resuspended in 15 μ L of SF Cell-Line Solution (V4XC-2032,
513 Lonza) or P3 Primary Cell Solution (V4XP-3032, Lonza), respectively. The Cas9-sgRNA
514 mixture was added to the cells to a final volume of 20 μ L and transferred to 16-well
515 NucleocuvetteTM strips (Lonza). Pulse was applied utilizing the CM-130 program for HEK293
516 cells and EA-104 for RPE1 cells. After nucleofection, cells were left to recover for 10 minutes
517 at room temperature, after which they were resuspended in 80 μ L of pre-warmed medium,
518 transferred to appropriate dishes and kept in culture.

519 Confirmed of knock-out cell lines was performed by Sanger sequencing, through amplification
520 of genomic DNA with appropriate primers (**Supplementary Data 2**). Tracking of indels by
521 decomposition was performed by the tool TIDE (Brinkman et al., 2014). For RPE1 cells, more
522 than 90% of alleles contained an out-of-frame (+1bp) mutation, which allowed for the use of
523 the pooled population. HEK293 cells showed a lower frequency of out-of-frame indels, hence
524 single cell clones were seeded by limiting dilutions into 96-well plates and a clone containing
525 a +1 bp mutation was selected, 2-3 weeks after clonal expansion, for further studies.
526 Abrogation of MLH1 expression was confirmed in both cell lines by immunoblotting.

527

528 **Focused DNA repair genetic screen**

529 CRISPR-Cas9 knockouts of DNA repair genes were generated in collaboration with Horizon
530 Genomics. Sequences of sgRNAs were designed by Horizon Genomics or with the use of
531 <http://chopchop.cbu.uib.no/>. sgRNA sequences and frameshift mutations can be found in
532 **Supplementary Data 1**.

533 For the genetic screen, 80,000 cells were seeded in technical duplicates in 12-well plates.
534 Cells were transfected the day after with 636ng of pCMV-PE2 and 159ng of the *HEK3* pegRNA
535 inducing a 5 bp deletion, per well. 1.6 μ L Lipofectamine 2000 (ThermoFisher Scientific) were
536 used per well, following the manufacturer's instructions. A separate transfection control was
537 performed using 795 ng of the pmaxGFPTM vector (Lonza). Medium containing transfection
538 reagents was removed 16 hours post-transfection. Transfection efficiency was measured 48
539 hours after transfection, by determining the percentage of GFP positive cells by flow
540 cytometry. Genomic DNA was harvested 96 hours post-transfection, using the QUIAmp DNA
541 Blood Mini kit (Quiagen), following the manufacturer's instructions.

542

543 **Transfection and genomic DNA preparation of mismatch repair-deficient cell lines**

544 HEC59, HCT116 and 293T-L α and HEK293 cells were seeded in 48-well plates in duplicates
545 (50,000 cells/well). Transfections were performed the next day, using 1 μ L Lipofectamine 2000
546 (ThermoFisher Scientific) per well, following the manufacturer's instructions. Cells were
547 transfected with 320 ng of the pCMV-PE2 vector, 80 ng of the respective pegRNA and, for
548 PE3 and PE3b, 33.2 ng of the nicking sgRNA, per well. A transfection control was performed
549 in parallel, by transfecting 400 ng per well of the pmaxGFPTM vector (Lonza).

550 iPSCs were seeded in 48-well plates in duplicates (50,000 cells/well). Transfections were
551 performed the next day, using 1 μ L Lipofectamine Stem (ThermoFisher Scientific) per well,
552 following the manufacturer's instructions. Cells were transfected with 320 ng of the pCMV-
553 PE2 vector, 80 ng of the respective pegRNA and, for PE3 and PE3b, 33.2 ng of the nicking
554 sgRNA, per well. A transfection control was performed in parallel, by transfecting 400 ng per
555 well of the pmaxGFPTM vector (Lonza).

556 Genomic DNA was extracted 96 hours after transfection, by removing the medium,
557 resuspending the cells in a lysis solution (100 μ L DirectPCR Lysis Reagent (Cell) (Viagen
558 Biotech), 76 μ L of water and 4 μ L Proteinase K) and incubating 45 minutes at 55°C and 45
559 minutes at 85°C.

560

561 **Prime editing in RPE1 cells**

562 Wild-type and MLH1-knockout RPE1 PE2-BSD cells were transduced at a high multiplicity of
563 infection with the plentiPEGRNAPuro encoding a 5bp deletion in the *HEK3* locus, together with
564 a nicking sgRNA for PE3 cloned in the lenti-sgRNA neo vector (**Supplementary Data 2**). Spin-
565 infection was performed with 500,000 cells/well in a 12-well plate with 8 μ g/mL polybrene
566 (2,000 rpm, 90 minutes, 32°C). Cells were selected the day after transduction with blasticidine
567 (10 μ g/ml), puromycin (2 μ g/ml) and G418 (400 μ g/ml). Genomic DNA was extracted as
568 described in the section 'Transfection and genomic DNA preparation of mismatch repair-
569 deficient cell lines', 96 hours post-transduction. Antibiotic selection was maintained throughout

570 the entire duration of the experiment. Prime editing efficiency was measured by Sanger
571 sequencing, after amplification of the genomic DNA with appropriate primers (**Supplementary**
572 **Data 2**). Editing efficiency was calculated by sequence decomposition, using TIDE (Brinkman
573 et al., 2014).

574

575 **High-throughput DNA sequencing of genomic samples**

576 Genomic sites of interest were amplified from genomic DNA samples and sequenced on an
577 Illumina Miseq or NextSeq, depending on the number of pooled samples. Amplification primers
578 containing Illumina forward and reverse primers (**Supplementary Data 2**) were used for a first
579 round of PCR (PCR1) to amplify the genomic region of interest. A mixture of staggered forward
580 primers was used to create complexity. PCR1 reactions were performed in a final volume of
581 25 μ L, using 0.5 μ M of each forward and reverse primers, 1 μ L genomic DNA and 12.5 μ L of
582 Phusion U Multiplex PCR 2x Master Mix (ThermoFisher Scientific). PCR1 was carried as
583 following: 98°C 2 min, 30 cycles [98°C 10 seconds, 61°C 20 seconds, 72°C 30 seconds],
584 followed by a final extension of 72 °C for 7 minutes. Unique Illumina dual index barcode primer
585 pairs were added to each sample in a second PCR reaction (PCR2). PCR2 was performed in
586 a final volume of 25 μ L, using 0.5 μ M of each unique forward and reverse Illumina barcoding
587 primer pair, 1 μ L of unpurified PCR1 reaction and 12.5 μ L of of Phusion U Multiplex PCR 2x
588 Master Mix. PCR2 was carried as following: 98°C 2 minntes, 12 cycles [98°C 10 seconds,
589 61°C 20 seconds, 72°C 30 seconds], followed by a final extension of 72°C for 7 minutes. PCR
590 products were analysed by electrophoresis in a 1% (w/v) agarose gel and purified using
591 magnetic AMPure XP beads (Beckman Coulter), using a ratio of beads:PCR product of 2:1.
592 DNA concentration was measured by fluorometric quantification (Qubit, ThermoFisher
593 Scientific) and sequenced on an Illumina instrument, according to manufacturer's instructions.
594 Sequencing reads were demultiplexed using MiSeq Reporter (Illumina) and alignment of
595 amplicon sequences to a reference sequence was performed using CRISPResso 2 (Clement
596 et al., 2019). CRISPResso2 was ran in standard mode and prime editing yield was calculated
597 as: number of aligned reads containing the desired edit/total aligned reads. Percentage of
598 indels was calculated as: number of aligned reads containing indels that are not the desired
599 edit/ total number of aligned reads.

600

601 **siRNA transfections**

602 The following siRNAs from Dharmacon (used at a final concentration of 100 nM) were used in
603 this study: MLH1 SMARTpool ON-TARGETplus (L-003906-00-0005) and Non-targeting
604 control SMARTpool ON-TARGETplus (D-001810-10-05). siRNA transfections in HEK293 cells
605 were performed using Dharmafect 1 following manufacturer's instructions. siRNA delivery was
606 performed 48 hours prior to transfection of prime editing vectors.

607

608 **Generation of dTAG-MLH1 HAP1 cell line**

609 A targeting vector encoding for the BSD-dTAG sequence (amplified from pCRIS-PITChv2-
610 BSD-dTAG (BRD4)) surrounded by two 1kb-long homology arms upstream and downstream
611 of the start codon of MLH1 was generated using Gibson assembly (NEB). An in vitro
612 transcribed sgRNA targeting the region spanning the start codon of MLH1 was generated as
613 previously described (Richardson et al., 2018). Cas9 protein (IDT) together with the targeting
614 vector and in vitro transcribed sgRNA were nucleofected into 200,000 haploid cells in 16-well
615 strips, using a 4D Nucleofector (Lonza) and the program DS-118. Three days after
616 nucleofection, 10 µg/ml blasticidin (Invivogen) were added to the culture medium for one week,
617 after which single and haploid clones were sorted into 96-well plates. Clonal haploid
618 populations were grown and validated for correct homology-directed repair by LR-PCR and
619 immunoblot analysis. dTAG-7 (R&D Systems), at the final concentration of 500 nM, was used
620 to test target degradation in the generated clones and all further targeted protein degradation
621 experiments.

622

623 **Generation of dTAG-MLH1 HAP1 BFP-positive cell line**

624 dTAG-MLH1 HAP1 cells were transduced at a low multiplicity of infection with the BFP dest
625 clone plasmid. Spin-infection was performed with 500,000 cells per well in a 12-well plate
626 (2,000 rpm, 30 minutes, 30°C). Cells were cell sorted by fluorescence (BD FACSMelody), one
627 week after transduction.

628

629 **Prime editing in dTAG-MLH1 HAP1 cell line**

630 25,000 dTAG-MLH1 HAP1 BFP-positive cells were seeded in two technical replicates and
631 three biological replicates in 48-well plates, treated or not with dTAG-7 (R&D Systems) at the
632 final concentration of 500 nM, as indicated. The day after seeding, cells were transfected with
633 200 ng of the pCMV-PE2 vector and 50 ng of a pegRNA cloned into the pU6-pegRNA-GG-
634 acceptor vector, encoding a 1bp substitution in BFP, converting it to GFP (**Supplementary**
635 **Data 2**). dTAG-ligand was replenished in the '+' condition and removed from the '24h'
636 condition. Medium was replaced every 24 hours for the entire course of the experiment (96
637 hours), always replenishing dTAG-7 in the '+' condition. Prime editing efficiency was
638 determined by percentage of GFP positive cells, measured by flow-cytometry.

639 Prime editing efficiency of the *HEK3* locus was measured by seeding 25,000 dTAG-MLH1
640 HAP1 cells in two technical replicates and three biological replicates in 48-well plates, treated
641 ('+') or not ('-') with dTAG-7 (R&D Systems) at the final concentration of 500 nM. The day after
642 seeding, cells were transfected with 200 ng of the pCMV-PE2 vector, 50 ng of the *HEK3*
643 pegRNA and 33.5 ng of the *HEK3* nicking sgRNA for PE3 (**Supplementary Data 2**), using 0.5

644 μ L of Lipofectamine 2000 (Thermo Fisher Scientific) and following the manufacturer's
645 instructions. Medium was replaced every 24 hours for the entire course of the experiment,
646 always replenishing dTAG-7 in the '+' condition. Genomic DNA was extracted 96 hours after
647 transfection, by removing the medium, resuspending the cells in a lysis solution (100 μ L
648 DirectPCR Lysis Reagent (Cell) (Viagen Biotech), 76 μ L of water and 4 μ L Proteinase K) and
649 incubating 45 minutes at 55 °C and 45 minutes at 85 °C. Prime editing efficiency was
650 determined by Sanger sequencing, after amplifying genomic DNA with appropriate primers
651 (**Supplementary Data 2**) and measured by sequence decomposition using TIDE (Brinkman
652 et al., 2014).

653

654 **Immunoblotting**

655 Cell extracts were prepared in RIPA lysis buffer (NEB) supplemented with protease inhibitors
656 (Sigma) and phosphatase inhibitors (Sigma, NEB). Immunoblots were performed using
657 standard procedures. Protein samples were separated by sodium dodecyl sulfate-
658 polyacrylamide gel electrophoresis (SDS-PAGE) (3–8% gradient gels, Invitrogen) and
659 subsequently transferred onto nitrocellulose membranes. Primary antibodies for MLH1
660 (554073, BD Pharmigen), MSH2 (ab52266, Abcam), MSH3 (ab69619, Abcam), Tubulin (3873,
661 Cell Signaling) and β -Actin (A5060, Sigma) were used at 1:1,000. Secondary antibodies were
662 used at 1:5,000 (HRP-conjugated goat anti-mouse or anti-rabbit IgG from Jackson
663 Immunochemicals). Immunoblots were imaged using a Curix 60 (AGFA) table-top processor.

664

665 **Immunofluorescence**

666 U2OS cells were reverse transfected using PEI (Sigma-Aldrich). 50,000 cells were seeded
667 per well of μ -Slide 8 well (Ibidi) chambered coverslip plates. Pre-extraction was performed
668 using 0.1% Tween in PBS 24 hours after reverse transfection. Cells were then fixed with 4%
669 para-formaldehyde and fixed cells were processed for immunofluorescence using the
670 following antibodies: anti-Cas9 (Cell Signalling, 14697), anti-TRF1 (Abcam, ab1423), anti-
671 MLH1 (ThermoFisher, A300-015A), anti-MSH2 (Bethyl, A300-452A), anti-GFP (Abcam,
672 ab6556). Primary antibodies were diluted 1:500 and secondary antibodies (Alexa Fluor® 568
673 goat anti-mouse and Alexa Fluor® 488 goat anti-rabbit, LifeTechnologies) were diluted
674 1:2,000.

675

676 **Imaging**

677 16-bit fluorescence images were acquired using an Olympus IXplore spinning disk confocal
678 microscope (equipped with the Yokogawa CSU-W1 with 50 μ m pinhole disk and a Hamamatsu
679 ORCA Fusion CMOS camera). A 60X oil immersion objective (NA 1.42) in combination with a
680 3.2X magnification lens (equalling 192X total magnification) was used for super-resolution

681 imaging of fixed cells and z-stacks with a 0.24 μ m slice interval were acquired. These z-stacks
682 were then processed using the Olympus 3D deconvolution software (constrained iterative
683 deconvolution, using automatic background removal and noise reduction, filter using
684 advanced maximum likelihood algorithm and 5 iterations). Finally, “maximum-z” projection
685 images of the deconvoluted z-stacks were generated. For further data analysis, the ImageJ
686 (NIH) distribution FIJI was used. Nuclear foci were counted manually and at least 50 cells per
687 condition were imaged in each experiment. Quantification of the foci was performed manually
688 based on maximum intensity projections.

689

690 **References**

- 691 Acharya, S., Wilson, T., Gradia, T., Kane, M., Guerrette, S., Marsischky, G., et al. (1996).
692 hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. *Proc. Natl.*
693 *Acad. Sci. U. S. A.* 93, 13629–13634. doi:10.1073/PNAS.93.24.13629.
- 694 Anzalone, A. V, Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., et al.
695 (2019). *Search-and-replace genome editing without double-strand breaks or donor DNA.*
696 doi:10.1038/s41586-019-1711-4.
- 697 Bothmer, A., Phadke, T., Barrera, L. A., Margulies, C. M., Lee, C. S., Buquicchio, F., et al.
698 (2017). Characterization of the interplay between DNA repair and CRISPR/Cas9-induced
699 DNA lesions at an endogenous locus. *Nat. Commun.* 8, 13905.
700 doi:10.1038/ncomms13905.
- 701 Brinkman, E. K., Chen, T., Amendola, M., and Van Steensel, B. (2014). Easy quantitative
702 assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* 42,
703 e168–e168. doi:10.1093/nar/gku936.
- 704 Cejka, P., Stojic, L., Mojas, N., Russell, A. M., Heinemann, K., Cannavó, E., et al. (2003).
705 Methylation-induced G2/M arrest requires a full complement of the mismatch repair
706 protein hMLH1. *EMBO J.* 22, 2245. doi:10.1093/EMBOJ/CDG216.
- 707 Charpentier, M., Khedher, A. H. Y., Menoret, S., Brion, A., Lamribet, K., Dardillac, E., et al.
708 (2018). CtIP fusion to Cas9 enhances transgene integration by homology-dependent
709 repair. *Nat. Commun.* 9, 1–11. doi:10.1038/s41467-018-03475-7.
- 710 Chen, B., Gilbert, L., Cimini, B., Schnitzbauer, J., Zhang, W., Li, G., et al. (2013). Dynamic
711 imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*
712 155, 1479–1491. doi:10.1016/J.CELL.2013.12.001.
- 713 Chen, P. J., Hussmann, J. A., Yan, J., Knipping, F., Ravisankar, P., Chen, P.-F., et al. (2021).
714 Enhanced prime editing systems by manipulating cellular determinants of editing
715 outcomes. *Cell* 184, 5635-5652.e29. doi:10.1016/J.CELL.2021.09.018.
- 716 Clement, K., Rees, H., Canver, M. C., Gehrke, J. M., Farouni, R., Hsu, J. Y., et al. (2019).

- 717 CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat.*
718 *Biotechnol.* 2019 373 37, 224–226. doi:10.1038/s41587-019-0032-3.
- 719 Drummond, J., Li, G., Longley, M., and Modrich, P. (1995). Isolation of an hMSH2-p160
720 heterodimer that restores DNA mismatch repair to tumor cells. *Science* 268, 1909–1912.
721 doi:10.1126/SCIENCE.7604264.
- 722 Ferreira da Silva, J., Meyenberg, M., and Loizou, J. I. (2021). Tissue specificity of DNA repair:
723 the CRISPR compass. *Trends Genet.* 0. doi:10.1016/J.TIG.2021.07.010.
- 724 Fishel, R. (2015). Mismatch repair. *J. Biol. Chem.* 290, 26395–26403.
725 doi:10.1074/jbc.R115.660142.
- 726 Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., et al.
727 (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA
728 cleavage. *Nature*. doi:10.1038/nature24644.
- 729 Gradia, S., Acharya, S., and Fishel, R. (1997). The Human Mismatch Recognition Complex
730 hMSH2-hMSH6 Functions as a Novel Molecular Switch. *Cell* 91, 995–1005.
731 doi:10.1016/S0092-8674(00)80490-0.
- 732 Gu, S., Bodai, Z., Cowan, Q. T., and Komor, A. C. (2021). Base editors: Expanding the types
733 of DNA damage products harnessed for genome editing. *Gene Genome Ed.* 1, 100005.
734 doi:10.1016/J.GGEDIT.2021.100005.
- 735 Haugen, A. C., Goel, A., Yamada, K., Marra, G., Nguyen, T. P., Nagasaka, T., et al. (2008).
736 Genetic instability caused by loss of MutS homologue 3 in human colorectal cancer.
737 *Cancer Res.* 68, 8465–8472. doi:10.1158/0008-5472.CAN-08-0002.
- 738 Holmes, J., Clark, S., and Modrich, P. (1990). Strand-specific mismatch correction in nuclear
739 extracts of human and *Drosophila melanogaster* cell lines. *Proc. Natl. Acad. Sci. U. S. A.*
740 87, 5837–5841. doi:10.1073/PNAS.87.15.5837.
- 741 Hussmann, J. A., Ling, J., Ravisankar, P., Yan, J., Cirincione, A., Xu, A., et al. (2021). Mapping
742 the genetic landscape of DNA double-strand break repair. *Cell* 184, 5653-5669.e25.
743 doi:10.1016/J.CELL.2021.10.002.
- 744 Ioannidi, E. I., Yarnall, M. T. N., Schmitt-Ulms, C., Krajeski, R. N., Lim, J., Villiger, L., et al.
745 (2021). Drag-and-drop genome insertion without DNA cleavage with CRISPR-directed
746 integrases. *bioRxiv*, 2021.11.01.466786. doi:10.1101/2021.11.01.466786.
- 747 Iyer, R. R., Pluciennik, A., Burdett, V., and Modrich, P. L. (2006). DNA mismatch repair:
748 Functions and mechanisms. *Chem. Rev.* 106, 302–323. doi:10.1021/cr0404794.
- 749 Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A
750 programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.
751 *Science (80-.)*. 337, 816–821. doi:10.1126/science.1225829.
- 752 Jiricny, J. (2006). The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol.* 2006 75
753 7, 335–346. doi:10.1038/nrm1907.

- 754 Kadyrov, F., Dzantiev, L., Constantin, N., and Modrich, P. (2006). Endonucleolytic function of
755 MutL α in human mismatch repair. *Cell* 126, 297–308.
756 doi:10.1016/J.CELL.2006.05.039.
- 757 Kim, H. K., Yu, G., Park, J., Min, S., Lee, S., Yoon, S., et al. (2021). Predicting the efficiency
758 of prime editing guide RNAs in human cells. *Nat. Biotechnol.* 39, 198–206.
759 doi:10.1038/S41587-020-0677-Y.
- 760 Koepfel, J., Madli Peets, E., Weller, J., Pallaseni, A., and Liberante, F. (2021). Predicting
761 efficiency of writing short sequences 1 into the genome using prime editing 2 3. *bioRxiv*.
762 doi:10.1101/2021.11.10.468024.
- 763 Koi, M., Umar, A., Chauhan, D. P., Cherian, S. P., Carethers, J. M., Kunkel, T. A., et al. (1994).
764 Human Chromosome 3 Corrects Mismatch Repair Deficiency and Microsatellite
765 Instability and Reduces N-Methyl-N'-nitro-N-nitrosoguanidine Tolerance in Colon Tumor
766 Cells with Homozygous hMLH1 Mutation. *Cancer Res.* 54, 4308–4312.
- 767 Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable
768 editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*
769 533, 420–424. doi:10.1038/nature17946.
- 770 Li, G. M., and Modrich, P. (1995). Restoration of mismatch repair to nuclear extracts of H6
771 colorectal tumor cells by a heterodimer of human MutL homologs. *Proc. Natl. Acad. Sci.*
772 *U. S. A.* 92, 1950–1954. doi:10.1073/pnas.92.6.1950.
- 773 Lin, Q., Zong, Y., Xue, C., Wang, S., Jin, S., Zhu, Z., et al. (2020). Prime genome editing in
774 rice and wheat. *Nat. Biotechnol.* 38, 582–585. doi:10.1038/s41587-020-0455-x.
- 775 Lipkin, S. M., Wang, V., Jacoby, R., Banerjee-Basu, S., Baxevanis, A. D., Lynch, H. T., et al.
776 (2000). MLH3: A DNA mismatch repair gene associated with mammalian microsatellite
777 instability. *Nat. Genet.* 24, 27–35. doi:10.1038/71643.
- 778 Liu, Y., Li, X., He, S., Huang, S., Li, C., Chen, Y., et al. (2020). Efficient generation of mouse
779 models with the prime editing system. *Cell Discov.* 6. doi:10.1038/s41421-020-0165-z.
- 780 Nabet, B., Roberts, J. M., Buckley, D. L., Paulk, J., Dastjerdi, S., Yang, A., et al. (2018). The
781 dTAG system for immediate and target-specific protein degradation. *Nat. Chem. Biol.*
782 2018 145 14, 431–441. doi:10.1038/s41589-018-0021-8.
- 783 O'Geen, H., Ren, C., Nicolet, C. M., Perez, A. A., Halmai, J., Le, V. M., et al. (2017). dCas9-
784 based epigenome editing suggests acquisition of histone methylation is not sufficient for
785 target gene repression. *Nucleic Acids Res.* 45, 9901–9916. doi:10.1093/NAR/GKX578.
- 786 Palombo, F., Gallinari, P., Iaccarino, I., Lettieri, T., Hughes, M., D'Arrigo, A., et al. (1995).
787 GTBP, a 160-kilodalton protein essential for mismatch-binding activity in human cells.
788 *Science (80-)*. 268, 1912–1914. doi:10.1126/SCIENCE.7604265.
- 789 Palombo, F., Iaccarino, I., Nakajima, E., Ikejima, M., Shimada, T., and Jiricny, J. (1996).
790 hMutS β , a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA.

- 791 *Curr. Biol.* 6, 1181–1184. doi:10.1016/S0960-9822(02)70685-4.
- 792 Petri, K., Zhang, W., Ma, J., Schmidts, A., Lee, H., Horng, J. E., et al. (2021). CRISPR prime
793 editing with ribonucleoprotein complexes in zebrafish and primary human cells. *Nat.*
794 *Biotechnol.* 2021, 1–5. doi:10.1038/s41587-021-00901-y.
- 795 Pines, A., Vrouwe, M. G., Marteiijn, J. A., Typas, D., Luijsterburg, M. S., Cansoy, M., et al.
796 (2012). PARP1 promotes nucleotide excision repair through DDB2 stabilization and
797 recruitment of ALC1. *J. Cell Biol.* 199, 235–249. doi:10.1083/JCB.201112132.
- 798 Pluciennik, A., Dzantiev, L., Iyer, R. R., Constantin, N., Kadyrov, F. A., and Modrich, P. (2010).
799 PCNA function in the activation and strand direction of MutL α endonuclease in mismatch
800 repair. *Proc. Natl. Acad. Sci. U. S. A.* 107, 16066–16071. doi:10.1073/pnas.1010662107.
- 801 Rees, H. A., Yeh, W. H., and Liu, D. R. (2019). Development of hRad51–Cas9 nickase fusions
802 that mediate HDR without double-stranded breaks. *Nat. Commun.* 10, 1–12.
803 doi:10.1038/s41467-019-09983-4.
- 804 Richardson, C. D., Kazane, K. R., Feng, S. J., Zelin, E., Bray, N. L., Schäfer, A. J., et al.
805 (2018). CRISPR–Cas9 genome editing in human cells occurs via the Fanconi anemia
806 pathway. *Nat. Genet.* doi:10.1038/s41588-018-0174-0.
- 807 Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L., and Corn, J. E. (2016). Enhancing
808 homology-directed genome editing by catalytically active and inactive CRISPR–Cas9
809 using asymmetric donor DNA. *Nat. Biotechnol.* 34, 339–344. doi:10.1038/NBT.3481.
- 810 Scholefield, J., and Harrison, P. T. (2021). Prime editing – an update on the field. *Gene Ther.*
811 2021 287 28, 396–401. doi:10.1038/s41434-021-00263-9.
- 812 Stojic, L., Brun, R., and Jiricny, J. (2004). Mismatch repair and DNA damage signalling. *DNA*
813 *Repair (Amst)*. 3, 1091–1101. doi:10.1016/j.dnarep.2004.06.006.
- 814 Stringer, B. W., Day, B. W., D'Souza, R. C. J., Jamieson, P. R., Ensbey, K. S., Bruce, Z. C.,
815 et al. (2019). A reference collection of patient-derived cell line and xenograft models of
816 proneural, classical and mesenchymal glioblastoma. *Sci. Rep.* 9. doi:10.1038/S41598-
817 019-41277-Z.
- 818 Sürün, D., Schneider, A., Mircetic, J., Neumann, K., Lansing, F., Paszkowskirogacz, M., et al.
819 (2020). Efficient generation and correction of mutations in human iPS cells utilizing
820 mRNAs of CRISPR base editors and prime editors. *Genes (Basel)*. 11.
821 doi:10.3390/genes11050511.
- 822 Thomas, D. C., Roberts, J. D., and Kunkel, T. A. (1991). Heteroduplex repair in extracts of
823 human HeLa cells. *J. Biol. Chem.* 266, 3744–3751. doi:10.1016/s0021-9258(19)67858-
824 0.
- 825 Umar, A., Koi, M., Risinger, J. I., Glaab, W. E., Tindall, K. R., Kolodner, R. D., et al. (1997).
826 Correction of hypermutability, N-Methyl-N'-nitro-N-nitrosoguanidine resistance, and
827 defective dna mismatch repair by introducing chromosome 2 into human tumor cells with

- 828 mutations in MSH2 and MSH6. *Cancer Res.* 57, 3949–3955.
- 829 Wang, Q., Liu, J., Janssen, J. M., Tasca, F., Mei, H., and Gonç Alves, M. A. F. V (2021).
830 Broadening the reach and investigating the potential of prime editors through fully viral
831 gene-deleted adenoviral vector delivery. *Nucleic Acids Res.* 49, 11986–12001.
832 doi:10.1093/NAR/GKAB938.
- 833 Wang, X., Kam, Z., Carlton, P. M., Xu, L., Sedat, J. W., and Blackburn, E. H. (2008). Rapid
834 telomere motions in live human cells analyzed by highly time-resolved microscopy.
835 *Epigenetics Chromatin* 2008 11 1, 1–19. doi:10.1186/1756-8935-1-4.
- 836 Yeh, C. D., Richardson, C. D., and Corn, J. E. (2019). Advances in genome editing through
837 control of DNA repair pathways. *Nat. Cell Biol.* 21(12), 1468–1478. doi:10.1038/s41556-
838 019-0425-z.
- 839 Zou, X., Koh, G. C. C., Nanda, A. S., Degasperi, A., Urgo, K., Roumeliotis, T. I., et al. (2021).
840 A systematic CRISPR screen defines mutational mechanisms underpinning signatures
841 caused by replication errors and endogenous DNA damage. *Nat. Cancer* 2, 643–657.
842 doi:10.1038/s43018-021-00200-0.

843

844 **Data availability**

845 All sequencing data have been deposited in the European Nucleotide Archive (EMBL-EBI;
846 ENA) with the study accession number PRJEB47501.

847

848 **Competing Interests**

849 The authors declare no commercial or financial relationships that could be construed as
850 potential conflict of interest.

851

852 **Author Contributions**

853 JFdaS, GO, JJ and JIL conceptualized the study. JFdaS and JIL obtained funding. JFdaS,
854 GO, EA, CK and AM carried out investigations. GT contributed to the confocal microscopy.
855 JFdaS performed analysis and visualization. JIL supervised the study. JFsaS with input from
856 JIL wrote the original draft and all authors reviewed and edited the final manuscript.

857

858 **Funding**

859 JFdaS is supported by a DOC fellowship from the Austrian Academy of Sciences
860 (ÖAW25035). A.M. is funded by the Austrian Science Fund, grant number P 33024 awarded
861 to J.I.L. The Loizou lab is funded by an ERC Synergy Grant (DDREAMM Grant agreement ID:
862 855741). This work was funded, in part, by a donation from Mr Benjamin Landesmann. CeMM
863 is funded by the Austrian Academy of Sciences.

864

865 **Acknowledgments**

866 We would like to thank the VBCF (Vienna, Austria) for all next generation sequencing. We
867 would like to acknowledge Dr Jacob Corn (ETHZ, Zurich, Switzerland) and Dr Richard
868 Sherwood (Harvard Medical School, Boston, USA) for critically reading the manuscript. We
869 would like to acknowledge Dr Markus Schröder (Corn Lab, ETHZ, Zurich, Switzerland) for
870 bioinformatic analysis. We would like to thank Mr Chris Fell (LBI-RUD/CeMM, Vienna, Austria)
871 and Mr Marc Wiedner (CeMM, Vienna, Austria) for technical support. We are thankful to Prof
872 KJ Patel (Weatherall Institute of Molecular Medicine, Oxford, UK) for providing the FANCD2
873 deficient HAP1 cell line. Prof Nik-Zainal (University of Cambridge, UK) generously provided
874 the MLH1 and MSH2-deficient human induced pluripotent stem cells. The Jackson lab (The
875 Gurdon Institute, University of Cambridge, UK) kindly provided the RPE cells. The Winter Lab
876 (CeMM, Vienna, Austria) provided valuable tools for protein degradation. We are grateful to
877 members of the Loizou lab for helpful discussions and feedback.

Chapter 4: Discussion

CRISPR-Cas9 technology has revolutionised molecular biology due to its unprecedentedly easy implementation, versatility, and precision. CRISPR allows the possibility to explore the genetic space at scale, but also to perform precise functional studies. Moreover, from a therapeutic point of view, CRISPR-mediated genome editing offers the possibility to cure multiple human genetic diseases. This unprecedented potential is highlighted by the fact that the technology has transferred from bench-to-bedside in less than 10 years.

Despite all the exciting possibilities, a successful implementation of the CRISPR technology relies on accurately predicting the editing outcomes in different cells and tissue types. This is not possible until a deep understanding of the DNA damage and repair mechanisms involved is achieved. CRISPR applications that rely on the generation of DSBs mediated by the endonuclease Cas9 engage DSBR pathways and frequently lead to the generation of unwanted mutagenic outcomes. Additionally, emerging CRISPR technologies such as base editors and prime editing, which are generally considered to be safer and more precise, rely on other DNA repair mechanisms that are yet to be explored (Gu *et al*, 2021). These cellular requirements have the potential to limit or determine the application of such approaches. With this work, we have attempted to elucidate the role of DNA repair mechanisms for Cas9-mediated DSBs, as well as prime editing.

4.1 Contribution of this thesis to the field of genome editing

Strategies that modulate DNA repair outcomes following a Cas9-mediated DSB have further developed the genome editing field. However, it was not until recently that attention was given to the intricate and complex network of DNA repair mechanisms that underly genome editing approaches.

When we started the investigation of mutagenic repair of Cas9-induced breaks, in 2017, only NHEJ had been implicated as a pathway that repairs such lesions, leading to error-prone products that can be exploited when gene disruption is the desired outcome. Our study has shed new light on the redundancy of the NHEJ pathway for the error-prone repair of Cas9 lesions at the genome-scale level, showing that it can be fully compensated for by another repair pathway, MMEJ or alt-EJ. The role of MMEJ in the repair of Cas9 lesions has been further addressed in other studies that comprehensively characterise the involvement of this pathway across multiple target sites and genomic contexts (Shen *et al*, 2018; Allen *et al*, 2019). Importantly, our study was based on the use of isogenic models of NHEJ deficiency in

human cell lines, contrary to pharmacological inhibition. The use of small molecule inhibitors is often undesirable due to possible off-target effects, as it has been described for example for some NHEJ-inhibitors (Greco *et al*, 2016).

Furthermore, our study was conducted at the time when the genome editing field had come to the realisation that the indel products generated by a unique sgRNA are highly reproducible across different cell lines and experiments (van Overbeek *et al*, 2016), an observation that we have also made. With reproducibility comes predictability, a feature that is highly relevant in the field, as it allows a safer navigation of repair outcomes. Expanding on this, we observed that this 'predictable' indel profile gets dramatically altered when MMEJ is the pathway engaged in repair. Hence, in cells or tissues that have impaired or reduced NHEJ activity, the repair outcomes will be shifted towards MMEJ indel signatures, altering the expected indel pattern.

Despite the discussion around the predictability of repair outcomes following the generation of Cas9-breaks, as well as the development of strategies to mitigate unintended off-target DSBs, the genome editing field has identified potential detrimental consequences that arise from on-target DSBs. These include induction of a p53-mediated DNA damage response, which might promote tumorigenesis through selection against cells with a functional p53 pathway, large DNA rearrangements and deletions, chromosomal loss, or chromothripsis (Kosicki *et al*, 2018; Leibowitz *et al*, 2021; Enache *et al*, 2020). For all these reasons, the field is moving towards precise DSB-independent strategies, such as base editors and prime editing. These strategies are highly promising for therapeutic application, but they rely on other DNA repair mechanisms that are still unexplored.

Our work on the identification of the MMR pathway as inhibitory for prime editing not only constitutes the first observation of the cellular requirements for this technology, but also indicates a potential avenue for its improvement. While our work was under revision, the group of David Liu (Broad Institute, Harvard MIT), the inventor of prime editing and base editing, described similar findings in the *Journal Cell* (Chen *et al*, 2021). This highlights how relevant the findings on the DNA repair requirements for prime editing are for the entire genome editing community.

4.2 Alternative pathways for the mutagenic repair of Cas9-induced breaks (Project 1)

In the context of NHEJ deficiency, created by the loss of a key protein for this pathway (such as DNA-PK, LIG4 or XRCC4), the activity of other end-joining pathways for the repair

of Cas9-mediated DSBs becomes apparent. Our kinetic studies (**Figure 1C, Article 1 PDF**) showed similar levels of editing of the mCherry reporter in Δ LIG4 and WT cell lines as fast as 24 hours after Cas9-induction. This result supports a model where NHEJ is unnecessary for the repair of Cas9-mediated DSBs and is replaced by an alternative pathway. Alternatively, one could argue that the observed mutagenic editing is a consequence of a series of error-free events that lead to the continuous cutting by Cas9 until the target DNA is no longer recognisable. However, such scenario would lead to a considerable delay in editing, which is not the case. The redundancy of the NHEJ pathway was further confirmed by the ability to recapitulate essentiality in genome-wide CRISPR screens performed in Δ LIG4 cells (**Figure 2, Article 1 PDF**).

In order to assess alternative repair mechanisms, we have generated different double-knockout cellular models in human cell lines, of which Δ LIG4/POLQ showed the most striking effect on editing abrogation. Contrary to NHEJ, POLQ-dependent MMEJ typically requires 2-20bp of microhomology and a certain degree of end-resection. These extra steps of DNA-end processing explain the generation of a distinct indel profile in Δ NHEJ cell lines, characterised by larger deletions, but they also provide an explanation for the slower onset of repair, as evidenced by our kinetic studies (**Figure 1C, Article 1 PDF**).

Interestingly, the abrogation of editing in Δ LIG4/POLQ cells was not complete and residual mutagenic repair was still observed (10-20% reads) (**Figure 3B, Article 1 PDF**). Further work is warranted to clarify this. One possible explanation would be that some residual activity of either pathway is still present. However, it is also plausible that repair is the result of a distinct DSB-repair pathway. One possible pathway that we did not address is SSA, which requires longer regions of microhomology and end-resection as compared to MMEJ (>20bp). In order to test this hypothesis, editing in the presence of RAD52 inhibitors could be tested in Δ LIG4/POLQ cell lines. Alternatively, genome-wide screens could be performed in Δ LIG4/POLQ cells, using a reporter system for editing. One would expect mutagenic editing to be fully abrogated in cells that have received sgRNAs targeting the pathway in question.

4.3 Exploiting MMEJ in the context of CRISPR-mediated genome editing (Project 1)

As already described in the Introduction chapter, NHEJ inhibition is a possible approach to increase the rates of precise CRISPR-Cas9 editing mediated by HR, exploiting

the competition between these two pathways. Our data on the redundancy of NHEJ for the repair of Cas9-breaks, as well as the identification of MMEJ as a main player in this process, suggests that HDR efficiency can be further increased by combined inhibition of NHEJ and MMEJ. POLQ inhibitors are currently in clinical trials for targeted cancer therapy (Schrempf *et al*, 2021). Their relevance for genome editing, however, remains to be addressed.

Finally, one of the limitations of the CRISPR-Cas9 technology is the study of non-coding regulatory regions, such as transcription factor binding sites (Lopes *et al*, 2016). These regions frequently remain functional despite the generation of small indels, a problem that has been circumvented using staggered sgRNAs spanning the entire regulatory region. This approach, however, amplifies other problems such as sgRNA design, potential off-targets and triggering of the DNA damage response. We hypothesise that, by transiently inhibiting NHEJ, and thereby diverging repair towards MMEJ, larger deletions will be induced (>10 bp), facilitating the disruption of regulatory regions using a single sgRNA.

4.4 Future directions for the study of Cas9-induced mutagenic repair (Project 1)

Whilst our genome-scale data proves the dispensability of NHEJ for the repair of Cas9-induced DSBs, this study does not cover the process of pathway choice when both NHEJ and MMEJ are present. Indel profiles of edited WT and Δ POLQ cells were identical (**Figure 3B**, **Article 1 PDF**), confirming the current view that NHEJ is the default pathway for the repair of Cas9 lesions. However, the number of sgRNAs tested is not enough to draw such a broad conclusion and a more systematic approach will be needed to address this question. Several factors, such as cell cycle and microhomology content of the target site, could dictate pathway choice.

Despite extensive research into the mechanism of DNA repair, predicting which pathways are activated by Cas9-induced breaks at specific sites and the resulting editing outcomes of repair remains a challenge. More recently, systematic approaches that consider not only the genomic sequence to be targeted but also its context, have been developed to address these questions. One study measured the activity of NHEJ and MMEJ, relative to chromatin context (Schep *et al*, 2021). Whereas NHEJ was shown to be more active in regions of euchromatin, the contribution of MMEJ was more apparent in heterochromatin. Additionally,

while this thesis was being written, a high-throughput approach called Repair-Seq has provided a new way of tackling the challenge of studying DNA repair in the context of CRISPR (Husmann *et al*, 2021). Repair-seq uses a CRISPRi-based approach, in which the site of DNA damage (i.e. the sgRNA target site) is coupled with a sgRNA that silences a DNA repair gene. Once this construct is introduced into the cells and the repair genes are downregulated, a break is induced by a programmable nuclease (Cas9 or Cas12). NGS can then be used to couple specific editing outcomes to the silenced repair gene. Repair-seq allows the simultaneous interrogation of thousands of indel profiles across numerous DNA repair profiles (Husmann *et al*, 2021). With this vast potential, it is likely that future studies on alternative mechanisms of repair of Cas9-breaks, as well as prediction of different editing outcomes, will arise from the use of this high-throughput technology.

4.4 The inhibitory role of MMR for prime editing (Project 2)

In this project, we have described an inhibitory role of the MMR pathway during prime editing, across different human cell lines, genomic loci and types of edits. These results, together with a recent manuscript published during the preparation of this thesis (Chen *et al*, 2021), consist of the first description on the cellular DNA repair determinants for this technology.

As described in the introduction of this thesis (chapter 1.2.2), MMR is a specialised pathway that repairs base substitutions, as well as small indels. Although the role of MMR in prime editing had not been addressed experimentally, it was speculated that MMR would be required for the resolution of the heteroduplex DNA that arises after edit installation (Petri *et al*, 2021; Scholefield & Harrison, 2021). Strategies to engage the MMR machinery, for example by nicking the non-edited strand in PE3, have been developed to improve editing efficiency. However, our results suggest a model that opposes this proposition, in which MMR factors directly bind to the heteroduplex DNA, recognising the installed edit as a mismatch to be removed.

We have observed that the efficiency of prime editing, across several edits and genomic loci, is increased in MMR-deficient cell lines, compared with their complemented counterparts (**Figure 1, Article 2 PDF**). Therefore we propose that, in MMR proficient cell lines, the nick present after the 5' flap removal, and before ligation, stimulates targeted repair of the edited strand and subsequent excision of the edit (**Figure 11A**). This happens through

the recognition of the lesion by the MutS complexes (either MSH2-MSH6 or MSH2-MSH3, depending on the edit) (**Figure 11B**), followed by the recruitment of MutL (**Figure 11C**), which incises the mismatched DNA, by creating nicks flanking the edit. The exonuclease EXO1 excises the heteroduplex from these incisions, targeting it for degradation, while RPA coats the ssDNA (Figure 11D). Finally, polymerase δ re-synthesises the gap and Ligase 1 (LIG1) seals the remaining nick (Figure 11E). The outcome of MMR activity is removal of the edit (Figure 11F).

If ligation of the nick occurs before MMR recognition, the signal that dictates which strand should be repaired is lost. This scenario is the most frequent outcome when MMR is abrogated and leads to repair of either strand, culminating in 50% of edited and non-edited outcomes (**Figure 11G**). An important question that still remains is how the heteroduplex DNA that is formed after the ligation of the nick is resolved in the absence of MMR (**Figure 11H**). One possibility is that DNA replication would be required for the introduction of the edit in both DNA strands. This being true, it would not only contradict the claim that prime editing is a cell-cycle independent technology, but it would also mean that, in the absence of MMR, not more than 50% of the DNA molecules would be edited. Future work, for example through the use of small molecule inhibitors that block cells in specific phases of the cell-cycle, is necessary to elucidate the cell cycle requirements for prime editing. Editing in post-mitotic MMR-deficient cells could also help elucidate the replication requirements for resolving the heteroduplex intermediate.

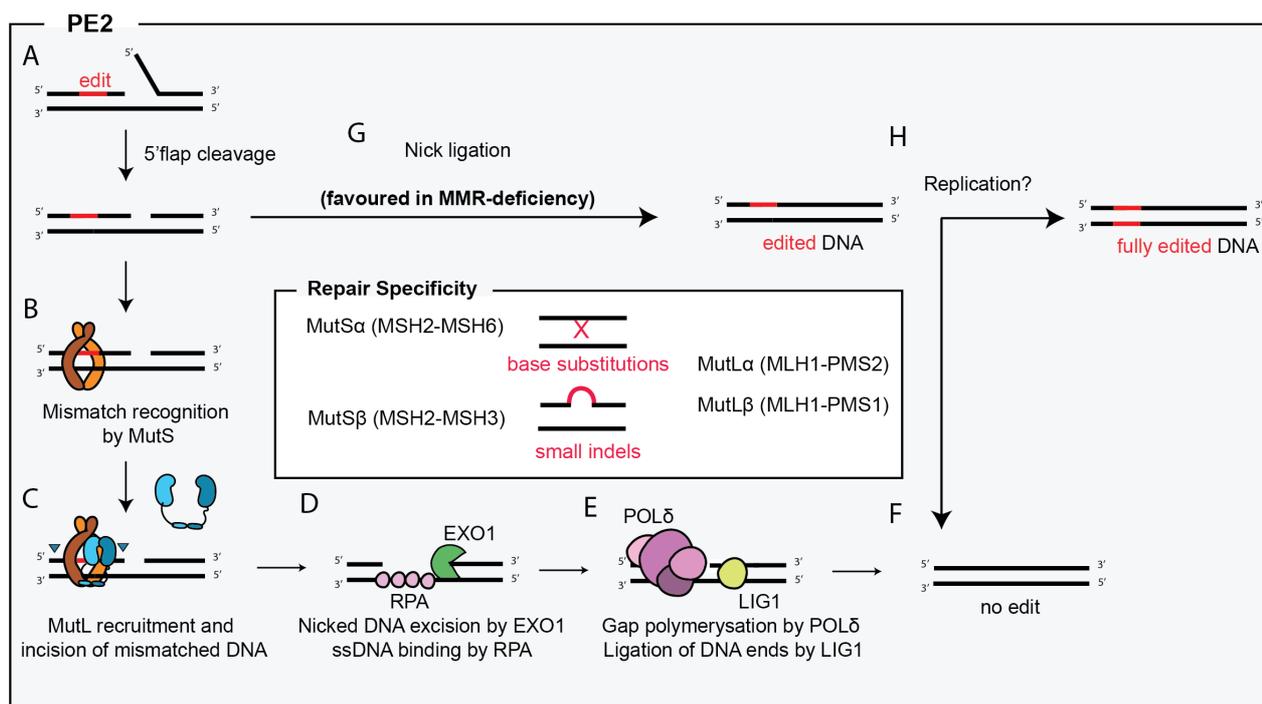


Figure 11: MMR counteracts prime editing efficiency in PE2. **A)** The DNA intermediate that results from the cleavage of the non-edited 5'flap and integration of the edited 3'flap is a substrate for the MMR pathway. **B)** The edit is recognised as a mismatch by the MutS complex that, upon undergoing a conformational change, recruits the MutL complex. **C)** MutL creates incisions that surround the edit, leading to its excision. **D)** EXO1 degrades the DNA strand containing the edit and RPA coats the fragile ssDNA. **E)** Polymerase δ polymerises the gap, utilising the non-edited strand as template. LIG1 seals the nick. **F)** The outcome of this repair is a non-edited DNA molecule. **G)** If ligation of the nick happens before the recognition of by the MutS complex, the outcome that is favoured in MMR-deficient backgrounds, the recognition signal for which strand to be repaired is lost, leading to a heteroduplex containing an edited and a non-edited DNA strand. **H)** It is not clear what leads to the complete integration of the edit, but it is possible that DNA replication would be required for this, leading to a 50% outcome of fully edited DNA, as well non-edited. **I)** An important feature of the MMR pathway is its specificity. While the MutS α complex (MSH2-MSH6) recognises mismatches created by base substitutions, MutS β (MSH2-MSH3) recognises small indels. MutL are heterodimers composed of MLH1 complexed with either PMS2 (MutL α) or PMS1 (MutL β). However, it is not clear in which occasions a MutL complex is favoured over the other.

The impact of MMR activity on prime editing efficiency differs depending on which strategy is applied. PE3, in which the non-edited strand is also nicked, is less affected by MMR activity since nicks on both strands can direct MMR to replace either strand (**Figure 12A**). Ligation of the nick on the edited strand would guide MMR to install edit, whereas ligation of the edited strand would drive removal of the edit, the same outcome observed in PE2. In MMR-deficient backgrounds, the mismatch is not recognised and ligation of both strands happens without any repair. This leads to edit installation (**Figure 12B**). The prime editing dependency on MMR for PE3b is thought to be the same as PE2, as the nicking sgRNA is designed to only work after installation of the edit. If the edit has been removed by MMR, PE3b will not work and it will be, essentially, PE2.

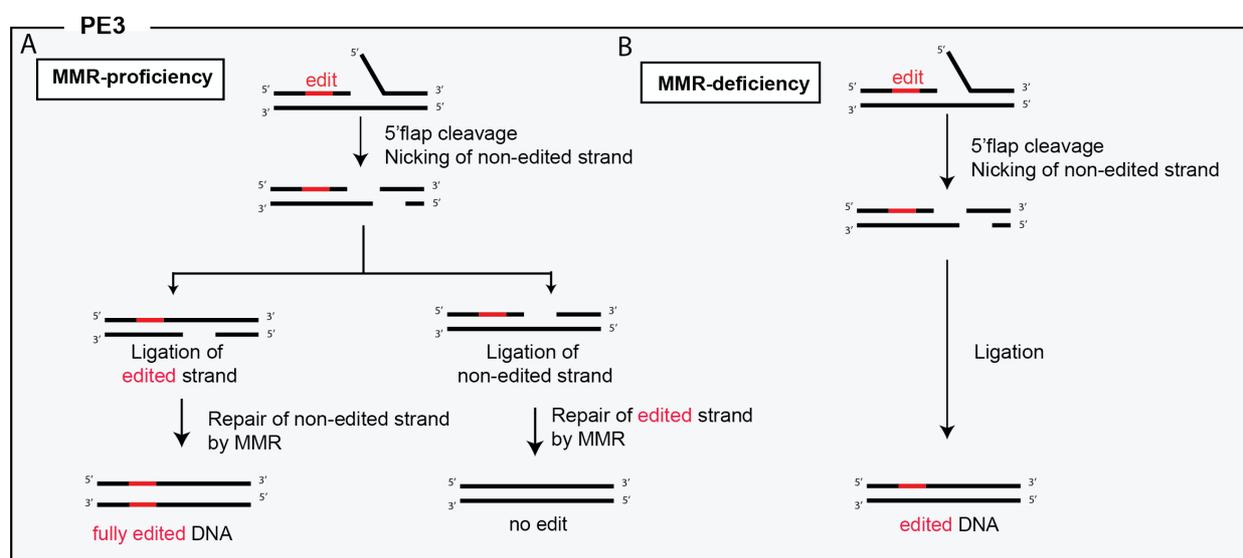


Figure 12: MMR involvement in the PE3 strategy. A) In PE3, the non-edited strand is nicked in addition to the edited strand. Depending on which strand is ligated first, this directs MMR to either install, or remove the edit. If the edited strand is ligated first, it is used as template for repair, leading to a fully edited DNA outcome. If the non-edited strand is ligated first, the edit is excised by MMR similarly to what happens in the strategy PE2. **B)** When MMR is absent, the ligation of both strands happens without any mismatch repair. This leads to a single outcome of edit installation. Therefore, MMR deficiency improves PE3, albeit to a lesser extent compared with PE2.

Besides different prime editing strategies, the highly specialised nature of the DDR suggests that other factors dictate the extent to which MMR can counteract prime editing. First, MMR is known to repair base substitutions and small indels, hence larger edits are less likely to be substrates of this pathway. Indeed, Chen and colleagues have extensively characterised the types of edits that are counteracted by MMR activity, showing that MMR involvement decreased as the indel length increased (Chen *et al*, 2021). Moreover, G/C to C/G edits, which form G:C mismatches, were shown to be less frequently removed by MMR factors, corroborating what has been previously described for this pathway (Lahue *et al*, 1989). Another recent study systematically measured the efficiency of insertion of 2 600 sequences by prime editing in the human cell lines HEK293T and HAP1 (Koeppel *et al*, 2021). These insertions ranged from 1 to 69 bp in length and varied in G/C content. Interestingly, the authors found that, in HEK293T, 1-4 bp sequences were inserted with high efficiency, compared to longer sequences. However, insertion frequency of these same sequences was considerably lower in HAP1 cells (Koeppel *et al*, 2021). This discrepancy could be attributed to the activity of the MMR pathway since HAP1 cells are MMR-proficient, contrary to HEK293T cells which are MMR-deficient due to methylation of the MLH1 promoter (Trojan *et al*, 2002). In fact, the authors tested this hypothesis by using HAP1 cells that are knockout for MLH1 and confirmed that average insertion rates generally increased in the mutant background comparing with wild-type cells (3.9-5.1 fold), with the insertion rates of sequences 1-4 bp long being the most dramatically altered (43-66 fold) (Koeppel *et al*, 2021). Overall this study consists of an additional and independent validation of the role of MMR in counteracting prime editing efficiency and confirms that longer sequences are less efficiently recognised and repaired by MMR.

In addition to type of edit, a certain degree of specialisation also occurs within the MMR pathway, as small base mismatches and small indels are signalled by different complexes (**Figure 11I**). This specificity is evidenced by our genetic screen using a collection of HAP1 knockout cell lines (**Figure 1, Article 2 PDF**). Here, we have shown that MMR factors counteract the installation of a 5bp deletion to different degrees, with the MSH6 deficient cell line showing the least impact on editing efficiency. This result can be explained by the fact that a 5bp deletion is preferentially signalled by the MSH2-MSH3 heterodimer, but not the MSH2-MSH6 heterodimer (Palombo *et al*, 1996), thus explaining the lack of an effect on editing upon loss of MSH6.

Finally, additional factors can determine the requirement of MMR in prime editing, such as sequence context of the target site. MMR has been shown to be more efficient in early replication euchromatin (Supek & Lehner, 2015) and lagging strand DNA, during replication

(Lujan *et al*, 2014). Therefore, a more systematic studies are necessary to dissect the prime editing dependency on MMR, across multiple edit types and sequence contexts.

4.5 Exploiting the inhibitory role of MMR for improved prime editing (Project 2)

MMR deficiency has been extensively associated with increased mutational burden and genome instability (Jiricny, 2006). However, our data on the frequency of undesired indels after prime editing in MMR-deficient and -proficient cell lines indicate that loss of MMR increases outcome fidelity (**Supplementary Figure 1D, Article 2, PDF**). This was also observed by Chen and colleagues (Chen *et al*, 2021) who have shown that depletion of MMR factors reduced the frequency of most categories of unintended prime editing outcomes.

Several approaches based on DDR modulation at the edit site have proved useful for improving the efficiency of HDR (Charpentier *et al*, 2018; Rees *et al*, 2019), suggesting that prime editing could also benefit from such developments. Since MMR loss increases both efficiency and fidelity of prime editing, the transient inhibition of this pathway is a promising approach for harnessing the technology.

By using siRNAs against the MMR factor MLH1, we have shown a significant increase in prime editing efficiency in the *HEK3* locus (**Figure 2F-E, Article 2, PDF**). Additionally, we have employed targeted protein degradation, through the dTAG system, as a strategy to endogenously deplete MLH1. Here, we have generated a clonal knock-in cell line by introducing the mutant FKBP12^{F36V} N-terminally to MLH1. Upon treatment of this cell line with a dTAG ligand, composed of an E3 ligase ligand linked to a selective FKBP12^{F36V} ligand, a ternary complex is formed between the MLH1 tagged protein and the E3 ligase, causing MLH1 polyubiquitination and subsequent degradation. This degradation occurs very rapidly and in a transient way, constituting a safe approach to inhibit MMR (**Figure 2G, Article 2, PDF**). Indeed, we have observed significantly increased prime editing efficiency upon MLH1 transient degradation (**Figure 2H, Article 2, PDF**).

An approach that has been used to increase HDR efficiency in the context of genome editing is the expression of dominant-negative forms of DDR factors. For example, the ectopic expression of a dominant-negative form of 53BP1 was shown to improve HDR (Canny *et al*, 2017). Similarly, the ectopic expression of dominant-negative forms of MMR proteins could improve prime editing, by counteracting the activity of the native pathway in edit removal. In fact, this has been explored by Chen and colleagues in their study (Chen *et al*, 2021), where

an endonuclease-dead MLH1 variant, lacking amino acids 754-756 (dnMLH1), was designed. Co-expression of this variant, together with the prime editing machinery, yielded considerable increases in editing efficiency across multiple endogenous loci and types of edits. The authors took this approach one step further and established a new generation of prime editors: PE4 (PE2 + dnMLH1) and PE5 (PE3 + dnMLH1), highlighting how relevant the discovery of the role of MMR is, for the prime editing field.

As elegant and efficient as the previously described approaches are, they only solve the problem of increasing efficiency and fidelity of prime editing in the laboratory setting. The co-expression of a dominant-negative form of MMR, as well as siRNA delivery, cannot be applied to the clinical setting. The dTAG approach is also not feasible, as it requires the generation of a knock-in model. Therefore, better approaches need to be devised to transiently inhibit MMR during prime editing. Chemical inhibition could be useful, although no small molecules that selectively target MMR have been yet identified. An alternative strategy would be to fuse dominant-negative forms of MMR factors to Cas9 allowing a competition with the native MMR machinery to occur locally, only at sites of editing. This would be a safer approach, circumventing the problems that might arise from systemic MMR inhibition. Nonetheless, the implementation of these approaches in a therapeutic context is reliant on an in-depth analysis of genome-wide mutations and microsatellite instability that might arise from MMR depletion.

A distinct strategy to exploit the inhibitory role of MMR in prime editing that would not require its inhibition would be through the exploitation of the DDR specificity. By designing pegRNAs that contain contiguous substitutions around the intended edit, encoding benign or silent mutations, the generated heteroduplex would evade MMR recognition and repair. This strategy has already been applied successfully (Chen *et al*, 2021), but more systematic studies need to be conducted in order to address design strategies to create silent edits that optimally evade MMR activity.

4.6 Other DNA repair requirements for prime editing (Project 2)

The involvement of MMR in resolving the heteroduplex that arises from prime editing is only a small piece of a much larger puzzle of DDR requirements. Even though the technology is independent of DSB-generation, the intricate response to DNA damage goes beyond the generation of breaks and there are several steps during prime editing that require, or are limited by, DDR intervention. These include not only factors that directly interact with the DNA, but also potential chromatin remodellers, or cell-cycle regulators.

To give a few examples, the generation of single-stranded DNA, upon nicking by the nCas9-RT could recruit factors that signal these types of lesions, such as PARP1 (**Figure 8A**). This signalling might be required for the recruitment of downstream factors involved in the subsequent steps of editing. Additionally, the resolution of the DNA:RNA hybrid formed by the complexed pegRNA might require intervention of endogenous factors, such as RNaseH (**Figure 8D**).

The process of flap equilibration that culminates in the integration of an edited 3' flap and the cleavage of a non-edited 5' flap (**Figure 8E**) relies on nucleases. FEN1, a structure-specific 5' flap endonuclease has been described as required for prime editing (Chen *et al*, 2021), but other nucleases remain unidentified.

Since MMR is a pathway that repairs small indels and base substitutions, its involvement in counteracting prime editing is restricted to such types of edits (Koeppel *et al*, 2021; Chen *et al*, 2021). It therefore remains to be determined what other pathways and DDR factors might be involved in the repair of larger edits. LP-BER, a pathway that has been shown to excise larger fragments of DNA is a potential pathway that would function on such substrates. This is particularly relevant since one of the most promising applications of prime editing is the insertion of large fragments of DNA that can be used, for example, as a strategy to tag proteins endogenously. During the writing of this thesis, two approaches that combine prime editing with integrases and transposases to harness the insertion, editing or deletion of large fragments of DNA have been described (Ioannidi *et al*, 2021; Anzalone *et al*, 2021). Identifying DDR mechanisms that promote, or counteract, these particular types of edits is therefore crucial.

4.7 The importance of using adequate cellular models

Even though DDR mechanisms are highly conserved across organisms, fundamental differences exist that might lead to distinct editing outcomes. For example, whereas in humans the error-prone NHEJ pathway is the default mechanism by which DSBs are repaired, in yeast HR is favoured over NHEJ (Mao *et al*, 2009). Therefore, it is crucial to understand the repair mechanisms that are associated with CRISPR-Cas9-based approaches in models that are relevant for the application of the technology. This becomes particularly relevant in light of the possibility of germline editing as the DDR mechanisms, as well as cell cycle regulators, in human embryos are still poorly understood and might differ from the mechanisms studied in somatic cells.

In the context of studies based on human cell lines, genetic differences might also obscure, or reveal, DDR requirements for genome editing. For example, in the original study describing prime editing, most experiments were performed in HEK293T cells (Anzalone *et al*, 2019). As mentioned before, HEK293T cells (but not HEK293, curiously) are MMR-deficient due to the hypermethylation of the MLH1-promoter (Trojan *et al*, 2002). Even though the authors did not choose this cellular model because of its MMR-status, they inadvertently conducted their studies in a cell line that promoted higher efficiencies of prime editing. This example highlights how relevant it is to understand the mutational background of the cellular models utilised.

In our study we have performed prime editing in a colorectal cancer cell line deficient for the MMR factors MLH1 and MSH3 (HCT116), as well as a cell line derived from a carcinoma of the endometrium, deficient for MSH2 (HEC59). The MMR-proficient counterparts of these cell lines have been generated by complementation of the chromosomes that encode the missing genes (chromosome 3 and 5, for MLH1 and MSH3 respectively and chromosome 2 for MSH2) (Koi *et al*, 1994; Haugen *et al*, 2008; Umar *et al*, 1997). This is an unusual approach comparing with cDNA complementation, but it is a particularity from studying the MMR pathway. Historically, overexpression of MMR factors through their cDNA did not lead to successful generation of stable MMR-proficient clonal cell lines and usually led to silencing of the exogenously expressed MMR cDNA. Alternatively, chromosome-complemented cell lines have been used over decades of MMR research. The exception to this has been the generation of 293T-L α cells, also used in our study. These cells consist of an inducible-system of MMR-deficiency, in which the overexpression of MLH1 is silenced upon doxycycline treatment (Cejka *et al*, 2003).

Despite the widespread utilisation of chromosome-complemented cell lines for MMR studies, it would be possible that the differences observed in prime editing efficiencies derive, not solely from reconstitution of MMR activity, but from expression of other factors encoded in the exogenous chromosomes. To confirm that this is not the case, we have recapitulated our findings in isogenic knockout models of MLH1 in the human cell line RPE, derived from human retinal pigment epithelium. RPE cells are considered to be 'wild-type' models to study DDR mechanisms, as they have no mutations associated with DDR factors.

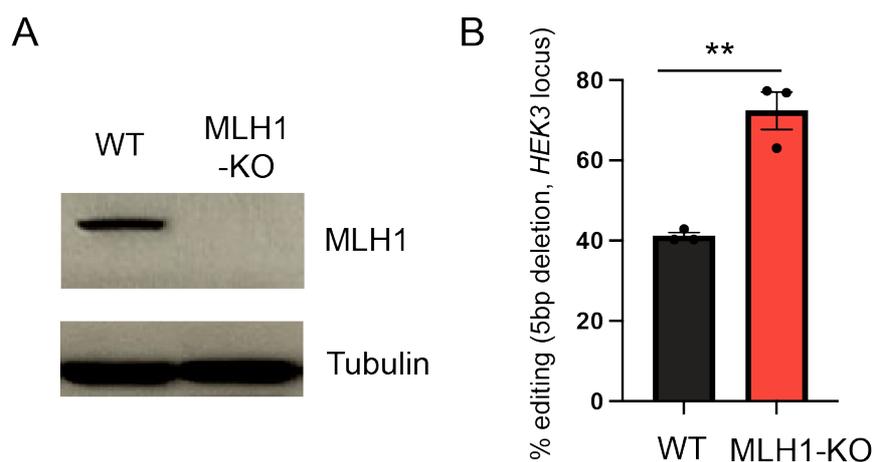


Figure 13: Validation of prime editing inhibition by MMR activity, in the human RPE cell line. **A)** Immunoblot confirming absence of the MLH1 protein in the generated RPE MLH1 knockout (KO) cell line. **B)** Prime editing efficiency of a 5bp deletion in the HEK3 locus, in RPE WT and MLH1-KO cell lines, as measured by Sanger sequencing (p -value=0.0028).

4.8 Future directions

In the last decade, CRISPR technology, in the shape of an incredible and vastly expanding toolbox, has revolutionised the field of molecular biology and opened unprecedented new possibilities for gene therapy. The complex and intricate mechanisms that guard the integrity of our genomes have been studied in detail for decades. These mechanisms constitute the foundation by which every CRISPR-based approach relies on. Nonetheless, the synergy between these two fields has only recently been acknowledged.

The implementation of genome editing technologies in a clinical setting, through DSB-dependent or DSB-independent approaches, relies on overcoming some of the big hurdles of the technology, such as delivery of the editing components and mitigation of off-target effects. But a big part of this implementation relies on a deep understanding of the DDR mechanisms that are crucial for these technologies, allowing a safe navigation of the DNA repair decision tree. The future of these studies relies on high-throughput approaches, such as Repair-Seq, offering the possibility to interrogate hundreds of perturbations across different sequence contexts. Repair-seq has already proved itself invaluable, dramatically expanding our knowledge on the DDR requirements for Cas9- and Cas12-mediated DSBs (Hussmann *et al*, 2021), base editors (Koblan *et al*, 2021a) and prime editing (Chen *et al*, 2021).

Notwithstanding these impressive developments, DDR mechanisms are also cell- and tissue-specific. Hence, integrating this knowledge in the study of genome editing technologies,

as well as utilising more relevant organism and cellular systems is crucial. This will ultimately permit the compilation of a comprehensive therapeutic map allowing for a specific genome editing approach, that is most suitable for each genetic disease, to be utilised in accordance with the tissue to be targeted and its DNA repair profile.

References

- A. Lea R & K. Niakan K (2019) Human germline genome editing. *Nat. Cell Biol.* **21**: 1479–1489
- Allen F, Crepaldi L, Alsinet C, Strong AJ, Kleshchevnikov V, De Angeli P, Páleníková P, Khodak A, Kiselev V, Kosicki M, Bassett AR, Harding H, Galanty Y, Muñoz-Martínez F, Metzakopian E, Jackson SP & Parts L (2018) Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotechnol.* **37**: 64–72 Available at: <http://www.nature.com/doi/10.1038/nbt.4317> [Accessed April 10, 2019]
- Allen F, Crepaldi L, Alsinet C, Strong AJ, Kleshchevnikov V, De Angeli P, Páleníková P, Khodak A, Kiselev V, Kosicki M, Bassett AR, Harding H, Galanty Y, Muñoz-Martínez F, Metzakopian E, Jackson SP & Parts L (2019) Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotechnol.* **37**: 64–82
- Amoasii L, Hildyard JCW, Li H, Sanchez-Ortiz E, Mireault A, Caballero D, Harron R, Stathopoulou TR, Massey C, Shelton JM, Bassel-Duby R, Piercy RJ & Olson EN (2018) Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. *Science (80-.)*. **362**: 86–91
- Anders C, Niewoehner O, Duerst A & Jinek M (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**: 569–573
- Anzalone A, Gao X, Podracky C, Nelson A, Koblan L, Raguram A, Levy J, Mercer J & Liu D (2021) Programmable large DNA deletion, replacement, integration, and inversion with twin prime editing and site-specific recombinases. *bioRxiv*: 2021.11.01.466790 Available at: <https://www.biorxiv.org/content/10.1101/2021.11.01.466790v1> [Accessed November 4, 2021]
- Anzalone A V, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A & Liu DR (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/31634902>
- Baik R, Wyman SK, Kabir S & Corn JE (2019) Genome editing to model and reverse a prevalent mutation associated with myeloproliferative neoplasms. *bioRxiv*: 855452
- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA & Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science (80-.)*. **315**: 1709–1712
- Barrangou R & Marraffini LA (2014) CRISPR-cas systems: Prokaryotes upgrade to adaptive immunity. *Mol. Cell* **54**: 234–244 Available at: <http://dx.doi.org/10.1016/j.molcel.2014.03.011>
- Bazrgar M, Gourabi H, Yazdi PE, Vazirinasab H, Fakhri M, Hassani F & Valojerdi MR (2014)

- DNA repair signalling pathway genes are overexpressed in poor-quality pre-implantation human embryos with complex aneuploidy. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **175**: 152–156 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24485984> [Accessed April 2, 2020]
- Black JB, Adler AF, Wang HG, D'Ippolito AM, Hutchinson HA, Reddy TE, Pitt GS, Leong KW & Gersbach CA (2016) Targeted Epigenetic Remodeling of Endogenous Loci by CRISPR/Cas9-Based Transcriptional Activators Directly Converts Fibroblasts to Neuronal Cells. *Cell Stem Cell* **19**: 406–414 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27524438> [Accessed April 1, 2020]
- Bolotin A, Quinquis B, Sorokin A & Dusko Ehrlich S (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**: 2551–2561 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16079334> [Accessed April 1, 2020]
- Bothmer A, Phadke T, Barrera LA, Margulies CM, Lee CS, Buquicchio F, Moss S, Abdulkerim HS, Selleck W, Jayaram H, Myer VE & Cotta-Ramusino C (2017) Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nat. Commun.* **8**: 13905 Available at: <http://www.nature.com/doi/10.1038/ncomms13905>
- Brinkman EK, Chen T, de Haas M, Holland HA, Akhtar W & van Steensel B (2018) Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks. *Mol. Cell* **0**: Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1097276518303125> [Accessed May 26, 2018]
- Brockmann M, Blomen VA, Nieuwenhuis J, Stickel E, Raaben M, Bleijerveld OB, Altelaar AFM, Jae LT & Brummelkamp TR (2017) Genetic wiring maps of single-cell protein states reveal an off-switch for GPCR signalling. *Nature* **546**: 307–311 Available at: <http://www.nature.com/articles/nature22376> [Accessed April 10, 2019]
- Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, Dickman MJ, Makarova KS, Koonin E V. & Van Der Oost J (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science (80-.).* **321**: 960–964
- Buechele C, Breese EH, Schneidawind D, Lin CH, Jeong J, Duque-Afonso J, Wong SHK, Smith KS, Negrin RS, Porteus M & Cleary ML (2015) MLL leukemia induction by genome editing of human CD34+ hematopoietic cells. *Blood* **126**: 1683–1694 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26311362> [Accessed April 2, 2020]
- Bukowska B & Karwowski BT (2018) Actual state of knowledge in the field of diseases related with defective nucleotide excision repair. *Life Sci.* **195**: 6–18 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/29305302> [Accessed March 31, 2020]
- Cadet J & Richard Wagner J (2013) DNA base damage by reactive oxygen species, oxidizing

- agents, and UV radiation. *Cold Spring Harb. Perspect. Biol.* **5**:
- Canny MD, Moatti N, Wan LCK, Fradet-Turcotte A, Krasner D, Mateos-Gomez PA, Zimmermann M, Orthwein A, Juang Y-C, Zhang W, Noordermeer SM, Seclen E, Wilson MD, Vorobyov A, Munro M, Ernst A, Ng TF, Cho T, Cannon PM, Sidhu SS, et al (2017) Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR–Cas9 genome-editing efficiency. *Nat. Biotechnol.* **36**: Available at: <http://www.nature.com/doi/10.1038/nbt.4021>
- Caulfield SE, Davis CC & Byers KF (2019) Olaparib: A Novel Therapy for Metastatic Breast Cancer in Patients With a BRCA1/2 Mutation. *J. Adv. Pract. Oncol.* **10**: 167–174 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/31538027> [Accessed April 1, 2020]
- Cejka P, Stojic L, Mojas N, Russell AM, Heinimann K, Cannavó E, Pietro M di, Marra G & Jiricny J (2003) Methylation-induced G2/M arrest requires a full complement of the mismatch repair protein hMLH1. *EMBO J.* **22**: 2245 Available at: </pmc/articles/PMC156088/> [Accessed August 31, 2021]
- Chang HHY, Pannunzio NR, Adachi N & Lieber MR (2017) Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* **18**: 495–506
- Charpentier M, Khedher AHY, Menoret S, Brion A, Lamribet K, Dardillac E, Boix C, Perrouault L, Tesson L, Geny S, De Cian A, Itier JM, Anegon I, Lopez B, Giovannangeli C & Concordet JP (2018) CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nat. Commun.* **9**: 1–11
- Chen PJ, Hussmann JA, Yan J, Knipping F, Ravisankar P, Chen P-F, Chen C, Nelson JW, Newby GA, Sahin M, Osborn MJ, Weissman JS, Adamson B & Liu DR (2021) Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184**: 5635–5652.e29 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/34653350> [Accessed October 31, 2021]
- Chen R & Wold MS (2014) Replication protein A: Single-stranded DNA's first responder: Dynamic DNA-interactions allow replication protein A to direct single-strand DNA intermediates into different pathways for synthesis or repair Prospects & Overviews R. Chen and M. S. Wold. *BioEssays* **36**: 1156–1161 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25171654> [Accessed March 31, 2020]
- Cheng Q, Wei T, Farbiak L, Johnson LT, Dilliard SA & Siegwart DJ (2020) Selective organ targeting (SORT) nanoparticles for tissue-specific mRNA delivery and CRISPR–Cas gene editing. *Nat. Nanotechnol.*: 1–8 Available at: <http://www.nature.com/articles/s41565-020-0669-6> [Accessed April 7, 2020]
- Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S & Kim JS (2014) Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24**:

132–141

- Chu VT, Weber T, Wefers B, Wurst W, Sander S, Rajewsky K & Kühn R (2015) Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* **33**: 543–548
- Ciccio A & Elledge SJ (2010) The DNA Damage Response: Making It Safe to Play with Knives. *Mol. Cell* **40**: 179–204 Available at: <http://dx.doi.org/10.1016/j.molcel.2010.09.019>
- Cofsky JC, Soczek KM, Knott GJ, Nogales E & Doudna JA (2021) CRISPR-Cas9 bends and twists DNA to read its sequence. *bioRxiv*: 2021.09.06.459219 Available at: <https://www.biorxiv.org/content/10.1101/2021.09.06.459219v1%0Ahttps://www.biorxiv.org/content/10.1101/2021.09.06.459219v1.abstract>
- Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA & Zhang F (2013) Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* (80-.). **339**: 819–823 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23287718> [Accessed April 10, 2019]
- Cuella-Martin R, Hayward SB, Fan X, Chen X, Huang JW, Tagliatela A, Leuzzi G, Zhao J, Rabadan R, Lu C, Shen Y & Ciccio A (2021) Functional interrogation of DNA damage response variants with base editing screens. *Cell* **184**: 1081-1097.e19 Available at: <https://pubmed.ncbi.nlm.nih.gov/33606978/> [Accessed October 28, 2021]
- Curtin NJ (2012) DNA repair dysregulation from cancer driver to therapeutic target. *Nat. Rev. Cancer* **12**: 801–817 Available at: <http://dx.doi.org/10.1038/nrc3399>
- Daley JM & Sung P (2014) 53BP1, BRCA1, and the Choice between Recombination and End Joining at DNA Double-Strand Breaks. *Mol. Cell. Biol.* **34**: 1380–1388
- Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, Schuster LC, Kuchler A, Alpar D & Bock C (2017) Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**: 297–301
- Dev H, Chiang TWW, Lescale C, de Krijger I, Martin AG, Pilger D, Coates J, Sczaniecka-Clift M, Wei W, Ostermaier M, Herzog M, Lam J, Shea A, Demir M, Wu Q, Yang F, Fu B, Lai Z, Balmus G, Belotserkovskaya R, et al (2018) Shieldin complex promotes DNA end-joining and counters homologous recombination in BRCA1-null cells. *Nat. Cell Biol.* **20**: 954–965
- Dicarlo JE, Norville JE, Mali P, Rios X, Aach J & Church GM (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* **41**: 4336–4343
- Ding X, Chaudhuri AR, Callen E, Pang Y, Biswas K, Klarmann KD, Martin BK, Burkett S, Cleveland L, Stauffer S, Sullivan T, Dewan A, Marks H, Tubbs AT, Wong N, Buehler E, Akagi K, Martin SE, Keller JR, Nussenzweig A, et al (2016) Synthetic viability by BRCA2 and PARP1/ARTD1 deficiencies. *Nat. Commun.* **7**: 12425 Available at:

- <http://www.nature.com/doi/10.1038/ncomms12425> [Accessed January 19, 2017]
- Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Aron L, Marjanovic ND, Dionne D, Burks T, Raychowdhury R, Adamson B, Norman TM, Lander ES, Weissman JS, Friedman N & Regev A (2016) Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**: 1853-1866.e17
- Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, Virgin HW, Listgarten J & Root DE (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**: 184–91 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26780180> [Accessed April 12, 2017]
- Doudna JA (2020) The promise and challenge of therapeutic genome editing. *Nature* **578**: 229–236 Available at: <http://dx.doi.org/10.1038/s41586-020-1978-5>
- Doudna JA & Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science* (80-.). **346**: 1258096–1258096 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25430774> [Accessed April 10, 2019]
- Enache OM, Rendo V, Abdusamad M, Lam D, Davison D, Pal S, Currimjee N, Hess J, Pantel S, Nag A, Thorner AR, Doench JG, Vazquez F, Beroukhim R, Golub TR & Ben-David U (2020) Cas9 activates the p53 pathway and selects for p53-inactivating mutations. *Nat. Genet.* **52**: Available at: <http://dx.doi.org/10.1038/s41588-020-0623-4>
- Eyquem J, Mansilla-Soto J, Giavridis T, Van Der Stegen SJC, Hamieh M, Cunanan KM, Odak A, Gönen M & Sadelain M (2017) Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection. *Nature* **543**: 113–117
- Feldman D, Singh A, Schmid-Burgk JL, Carlson RJ, Mezger A, Garrity AJ, Zhang F & Blainey PC (2019) Optical Pooled Screens in Human Cells. *Cell* **179**: 787-799.e17
- Fellmann C, Gowen BG, Lin PC, Doudna JA & Corn JE (2017) Cornerstones of CRISPR-Cas in drug discovery and therapy. *Nat. Rev. Drug Discov.* **16**: 89–100
- Feng L, Li N, Li Y, Wang J, Gao M, Wang W & Chen J (2015) Cell cycle-dependent inhibition of 53BP1 signaling by BRCA1. *Cell Discov.* **1**: 15019
- Ferreira da Silva J, Meyenberg M & Loizou JI (2021) Tissue specificity of DNA repair: the CRISPR compass. *Trends Genet.* **0**: Available at: <http://www.cell.com/article/S0168952521002006/fulltext> [Accessed August 31, 2021]
- Fong PC, Boss DS, Yap TA, Tutt A, Wu P, Mergui-Roelvink M, Mortimer P, Swaisland H, Lau A, O'Connor MJ, Ashworth A, Carmichael J, Kaye SB, Schellens JHM & De Bono JS (2009) Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N. Engl. J. Med.* **361**: 123–134
- Frangoul H, Altshuler D, Cappellini MD, Chen Y-S, Domm J, Eustace BK, Foell J, de la Fuente J, Grupp S, Handgretinger R, Ho TW, Kattamis A, Kernytsky A, Lekstrom-Himes J, Li

- AM, Locatelli F, Mapara MY, de Montalembert M, Rondelli D, Sharma A, et al (2021) CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β -Thalassemia. *N. Engl. J. Med.* **384**: 252–260 Available at: <https://pubmed.ncbi.nlm.nih.gov/33283989/> [Accessed February 21, 2021]
- Garaycochea JI, Crossan GP, Langevin F, Mulderrig L, Louzada S, Yang F, Guilbaud G, Park N, Roerink S, Nik-Zainal S, Stratton MR & Patel KJ (2018) Alcohol and endogenous aldehydes damage chromosomes and mutate stem cells. *Nature* **553**: 171–177
- Garaycochea JI & Patel KJ (2014) Why does the bone marrow fail in Fanconi anemia? *Blood* **123**: 26–34
- Gasiunas G, Barrangou R, Horvath P & Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **109**:
- Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI & Liu DR (2017) Programmable base editing of T to G C in genomic DNA without DNA cleavage. *Nature* **551**: 464–471
- Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, Qi LS, Kampmann M & Weissman JS (2014) Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**: 647–661 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25307932> [Accessed April 1, 2020]
- Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS & Qi LS (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**: 442
- Gonzalo S & Kreienkamp R (2015) DNA repair defects and genome instability in Hutchinson-Gilford Progeria Syndrome. *Curr. Opin. Cell Biol.* **34**: 75–83
- Goodman MF & Roger W (2013) Translesion DNA Polymerases. *Cold Spring Harb. Perspect. Biol.* **5(10)**:
- Gootenberg JS, Abudayyeh OO, Lee JW, Essletzbichler P, Dy AJ, Joung J, Verdine V, Donghia N, Daringer NM, Freije CA, Myhrvold C, Bhattacharyya RP, Livny J, Regev A, Koonin E V., Hung DT, Sabeti PC, Collins JJ & Zhang F (2017) Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science (80-.).* **356**: 438–442
- Greco GE, Matsumoto Y, Brooks RC, Lu Z, Lieber MR & Tomkinson AE (2016) SCR7 is neither a selective nor a potent inhibitor of human DNA ligase IV. *DNA Repair (Amst).* **43**: 18–23 Available at: <https://pubmed.ncbi.nlm.nih.gov/27235626/> [Accessed October 31, 2021]
- Gu S, Bodai Z, Cowan QT & Komor AC (2021) Base editors: Expanding the types of DNA damage products harnessed for genome editing. *Gene Genome Ed.* **1**: 100005

- Gupta R, Somyajit K, Narita T, Maskey E, Stanlie A, Kremer M, Typas D, Lammers M, Mailand N, Nussenzweig A, Lukas J & Choudhary C (2018) DNA Repair Network Analysis Reveals Shieldin as a Key Regulator of NHEJ and PARP Inhibitor Sensitivity. *Cell* **173**: 972-988.e23 Available at: <https://doi.org/10.1016/j.cell.2018.03.050>
- Haemmig S, Yang D, Sun X, Das D, Ghaffari S, Molinaro R, Chen L, Deng Y, Freeman D, Moullan N, Tesmenitsky Y, Wara AKMK, Simion V, Shvartz E, Lee JF, Yang T, Sukova G, Marto JA, Stone PH, Lee WL, et al (2020) Long noncoding RNA SNHG12 integrates a DNA-PK-mediated DNA damage response and vascular senescence. *Sci. Transl. Med.* **12**:
- Hanna RE, Hegde M, Fagre CR, DeWeirdt PC, Sangree AK, Szegletes Z, Griffith A, Feeley MN, Sanson KR, Baidi Y, Koblan LW, Liu DR, Neal JT & Doench JG (2021) Massively parallel assessment of human variants with base editor screens. *Cell* **184**: 1064-1080.e20 Available at: <https://pubmed.ncbi.nlm.nih.gov/33606977/> [Accessed October 28, 2021]
- Haugen AC, Goel A, Yamada K, Marra G, Nguyen TP, Nagasaka T, Kanazawa S, Koike J, Kikuchi Y, Zhong X, Arita M, Shibuya K, Oshimura M, Hemmi H, Boland CR & Koi M (2008) Genetic instability caused by loss of MutS homologue 3 in human colorectal cancer. *Cancer Res.* **68**: 8465–8472 Available at: <https://cancerres.aacrjournals.org/content/68/20/8465> [Accessed September 1, 2021]
- Hazra TK, Izumi T, Boldogh I, Imhoff B, Kow YW, Jaruga P, Dizdaroglu M & Mitra S (2002) Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proc. Natl. Acad. Sci. U. S. A.* **99**: 3523–3528
- Hoeijmakers JHJ (2001) Genome maintenance mechanisms for preventing cancer. *Nature* **411**: 366–374
- Hoeijmakers JHJ (2009) DNA damage, aging, and cancer. *N. Engl. J. Med.* **361**: 1475–1485
- Horlbeck MA, Gilbert LA, Villalta JE, Adamson B, Pak RA, Chen Y, Fields AP, Park CY, Corn JE, Kampmann M & Weissman JS (2016) Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife* **5**: 1–20
- Hossain MA, Barrow JJ, Shen Y, Haq MI & Bungert J (2015) Artificial Zinc Finger DNA Binding Domains: Versatile Tools for Genome Engineering and Modulation of Gene Expression. *J. Cell. Biochem.* **116**: 2435–2444
- Hsu PD, Lander ES & Zhang F (2014) Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157**: 1262–1278 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24906146> [Accessed April 10, 2019]
- Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G & Zhang F (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**: 827–832

- Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, Zeina CM, Gao X, Rees HA, Lin Z & Liu DR (2018a) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**: 57–63 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/29512652> [Accessed April 1, 2020]
- Hu Z, Shi Z, Guo X, Jiang B, Wang G, Luo D, Chen Y & Zhu YS (2018b) Ligase IV inhibitor SCR7 enhances gene editing directed by CRISPR-Cas9 and ssODN in human cancer cells. *Cell Biosci.* **8**: 12 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/29468011> [Accessed April 2, 2020]
- Hussmann JA, Ling J, Ravisankar P, Yan J, Cirincione A, Xu A, Simpson D, Yang D, Bothmer A, Cotta-Ramusino C, Weissman JS & Adamson B (2021) Mapping the genetic landscape of DNA double-strand break repair. *Cell* **184**: 5653-5669.e25
- Ioannidi EI, Yarnall MTN, Schmitt-Ulms C, Krajeski RN, Lim J, Villiger L, Zhou W, Jiang K, Roberts N, Zhang L, Vakulskas CA, Walker JA, Kadina AP, Zepeda AE, Holden K, Gootenberg JS & Abudayyeh OO (2021) Drag-and-drop genome insertion without DNA cleavage with CRISPR-directed integrases. *bioRxiv*: 2021.11.01.466786 Available at: <https://www.biorxiv.org/content/10.1101/2021.11.01.466786v1> [Accessed November 4, 2021]
- Ishino Y, Shinagawa H, Makino K, Amemura M & Nakamura A (1987) Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isoenzyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**: 5429–5433
- Iyer S, Suresh S, Guo D, Daman K, Chen JCJ, Liu P, Zieger M, Luk K, Roscoe BP, Mueller C, King OD, Emerson CP & Wolfe SA (2019) Precise therapeutic gene correction by a simple nuclease-induced double-stranded break. *Nature* **568**: 561–565
- Jackson SP & Bartek J (2009) The DNA-damage response in human biology and disease. *Nature* **461**: 1071–1078 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19847258> [Accessed December 11, 2016]
- Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, Salame TM, Tanay A, van Oudenaarden A & Amit I (2016) Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167**: 1883-1896.e15 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27984734> [Accessed April 1, 2020]
- Jensen RB, Carreira A & Kowalczykowski SC (2010) Purified human BRCA2 stimulates RAD51-mediated recombination. *Nature* **467**: 678–683
- Jiang X, Li W, Li X, Bai H & Zhang Z (2019) Current status and future prospects of PARP inhibitor clinical trials in ovarian cancer. *Cancer Manag. Res.* **11**: 4371–4390
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA & Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-.).* **337**: 816–821

- Jiricny J (2006) The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol.* 2006 7: 335–346 Available at: <https://www.nature.com/articles/nrm1907> [Accessed August 31, 2021]
- Jost M, Santos DA, Saunders RA, Horlbeck MA, Hawkins JS, Scaria SM, Norman TM, Hussmann JA, Liem CR, Gross CA & Weissman JS (2020) Titrating gene expression using libraries of systematically attenuated CRISPR guide RNAs. *Nat. Biotechnol.* 38: 355–364
- Kaina B, Christmann M, Naumann S & Roos WP (2007) MGMT: Key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. *DNA Repair (Amst)*. 6: 1079–1099
- Kairupan C & Scott RJ (2007) Base excision repair and the role of MUTYH. *Hered. Cancer Clin. Pract.* 5: 199–209
- Kampmann M (2018) CRISPRi and CRISPRa Screens in Mammalian Cells for Precision Biology and Medicine. *ACS Chem. Biol.* 13: 406–416 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/29035510> [Accessed April 1, 2020]
- Kim D, Lim K, Kim ST, Yoon SH, Kim K, Ryu SM & Kim JS (2017) Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat. Biotechnol.* 35: 475–480 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28398345> [Accessed April 1, 2020]
- Kim S, Kim D, Cho SW, Kim J & Kim JS (2014) Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* 24: 1012–1019
- Klein Douwel D, Boonen RACM, Long DT, Szybowska AA, Räschele M, Walter JC & Knipscheer P (2014) XPF-ERCC1 Acts in Unhooking DNA Interstrand Crosslinks in Cooperation with FANCD2 and FANCP/SLX4. *Mol. Cell* 54: 460–471 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24726325> [Accessed March 31, 2020]
- Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z & Joung JK (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529: 490–495
- Koblan LW, Arbab M, Shen MW, Hussmann JA, Anzalone A V., Doman JL, Newby GA, Yang D, Mok B, Replogle JM, Xu A, Sisley TA, Weissman JS, Adamson B & Liu DR (2021a) Efficient C•G-to-G•C base editors developed using CRISPRi screens, target-library analysis, and machine learning. *Nat. Biotechnol.* 2021: 1–12 Available at: <https://www.nature.com/articles/s41587-021-00938-z> [Accessed November 4, 2021]
- Koblan LW, Erdos MR, Wilson C, Cabral WA, Levy JM, Xiong ZM, Tavarez UL, Davison LM, Gete YG, Mao X, Newby GA, Doherty SP, Narisu N, Sheng Q, Krilow C, Lin CY, Gordon LB, Cao K, Collins FS, Brown JD, et al (2021b) In vivo base editing rescues Hutchinsonson–

- Gilford progeria syndrome in mice. *Nature* **589**: 608–614 Available at: <https://pubmed.ncbi.nlm.nih.gov/33408413/> [Accessed October 28, 2021]
- Koeppel J, Madli Peets E, Weller J, Pallaseni A & Liberante F (2021) Predicting efficiency of writing short sequences 1 into the genome using prime editing 2 3. *bioRxiv* Available at: <https://doi.org/10.1101/2021.11.10.468024> [Accessed November 15, 2021]
- Koi M, Umar A, Chauhan DP, Cherian SP, Carethers JM, Kunkel TA & Boland CR (1994) Human Chromosome 3 Corrects Mismatch Repair Deficiency and Microsatellite Instability and Reduces N-Methyl-N'-nitro-N-nitrosoguanidine Tolerance in Colon Tumor Cells with Homozygous hMLH1 Mutation. *Cancer Res.* **54**: 4308–4312
- Komor AC, Kim YB, Packer MS, Zuris JA & Liu DR (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**: 420–424 Available at: <http://www.nature.com/articles/nature17946> [Accessed May 15, 2019]
- Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O & Zhang F (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**: 583–588
- Kosicki M, Tomberg K & Bradley A (2018) Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**: 765–771
- Kottemann MC & Smogorzewska A (2013) Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nature* **493**: 356–63 Available at: <http://dx.doi.org/10.1038/nature11863>
- Kuscu C, Parlak M, Tufan T, Yang J, Szlachta K, Wei X, Mammadov R & Adli M (2017) CRISPR-STOP: Gene silencing through base-editing-induced nonsense mutations. *Nat. Methods* **14**: 710–712
- Lahue RS, Au KG & Modrich P (1989) DNA mismatch correction in a defined system. *Science* (80-.). **245**: 160–164 Available at: <https://pubmed.ncbi.nlm.nih.gov/2665076/> [Accessed November 6, 2021]
- Lau CH & Suh Y (2017) In vivo genome editing in animals using AAV-CRISPR system: Applications to translational research of human disease. *F1000Research* **6**: 2153 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/29333255> [Accessed April 2, 2020]
- Legut M, Daniloski Z, Xue X, McKenzie D, Guo X, Wessels HH & Sanjana NE (2020) High-Throughput Screens of PAM-Flexible Cas9 Variants for Gene Knockout and Transcriptional Modulation. *Cell Rep.* **30**: 2859-2868.e5
- Leibowitz ML, Papathanasiou S, Doerfler PA, Blaine LJ, Sun L, Yao Y, Zhang C-Z, Weiss MJ & Pellman D (2021) Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nat. Genet.*: 1–11 Available at: <https://doi.org/10.1038/s41588-021->

00838-7 [Accessed April 22, 2021]

- Li GM (2008) Mechanisms and functions of DNA mismatch repair. *Cell Res.* **18**: 85–98
- Liang F, Han M, Romanienko PJ & Jasin M (1998) Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **95**: 5172–5177
- Lin Q, Zong Y, Xue C, Wang S, Jin S, Zhu Z, Wang Y, Anzalone A V., Raguram A, Doman JL, Liu DR & Gao C (2020) Prime genome editing in rice and wheat. *Nat. Biotechnol.* **38**: 582–585 Available at: <https://pubmed.ncbi.nlm.nih.gov/32393904/> [Accessed October 28, 2021]
- Lin S, Staahl BT, Alla RK & Doudna JA (2014) Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**: e04766
- Liskay, R. M., Letsou, A. & Stachelek JL (1987) Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. - PubMed - NCBI. *Genetics* **115**: 161–167 Available at: <https://www.ncbi.nlm.nih.gov/pubmed/3557108> [Accessed March 31, 2020]
- Litman GW, Dishaw LJ, Cannon JP, Haire RN & Rast JP (2007) Alternative mechanisms of immune receptor diversity. *Curr. Opin. Immunol.* **19**: 526–534
- Liu B, Wang J, Chan KM, Tjia WM, Deng W, Guan X, Huang JD, Li KM, Chau PY, Chen DJ, Pei D, Pendas AM, Cadiñanos J, López-Otín C, Tse HF, Hutchison C, Chen J, Cao Y, Cheah KSE, Tryggvason K, et al (2005) Genomic instability in laminopathy-based premature aging. *Nat. Med.* **11**: 780–785
- Liu P, Chen M, Liu Y, Qi LS & Ding S (2018) CRISPR-Based Chromatin Remodeling of the Endogenous Oct4 or Sox2 Locus Enables Reprogramming to Pluripotency. *Cell Stem Cell* **22**: 252-261.e4 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/29358044> [Accessed April 1, 2020]
- Liu Y, Li X, He S, Huang S, Li C, Chen Y, Liu Z, Huang X & Wang X (2020) Efficient generation of mouse models with the prime editing system. *Cell Discov.* **6**: Available at: <https://pubmed.ncbi.nlm.nih.gov/33907184/> [Accessed October 28, 2021]
- Lomova A, Clark DN, Campo-Fernandez B, Flores-Bjurström C, Kaufman ML, Fitz-Gibbon S, Wang X, Miyahira EY, Brown D, DeWitt MA, Corn JE, Hollis RP, Romero Z & Kohn DB (2019) Improving Gene Editing Outcomes in Human Hematopoietic Stem and Progenitor Cells by Temporal Control of DNA Repair. *Stem Cells* **37**: 284–294 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/30372555> [Accessed April 2, 2020]
- Long C, Amoasii L, Mireault AA, McAnally JR, Li H, Sanchez-Ortiz E, Bhattacharyya S, Shelton JM, Bassel-Duby R & Olson EN (2016) Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. *Science* (80-.). **351**: 400–403

- Lopes R, Korkmaz G & Agami R (2016) Applying CRISPR–Cas9 tools to identify and characterize transcriptional enhancers. *Nat. Rev. Mol. Cell Biol.* **17**: 597–604 Available at: <http://www.nature.com/articles/nrm.2016.79> [Accessed July 9, 2018]
- Lujan SA, Clausen AR, Clark AB, MacAlpine HK, MacAlpine DM, Malc EP, Mieczkowski PA, Burkholder AB, Fargo DC, Gordenin DA & Kunkel TA (2014) Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. *Genome Res.* **24**: 1751–1764 Available at: <https://genome.cshlp.org/content/24/11/1751.full> [Accessed November 6, 2021]
- Maddalo D, Machado E, Concepcion CP, Bonetti C, Vidigal JA, Han YC, Ogrodowski P, Crippa A, Rekhtman N, Stanchina E De, Lowe SW & Ventura A (2014) In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* **516**: 423–428 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25337876> [Accessed April 2, 2020]
- Maeder ML, Stefanidakis M, Wilson CJ, Baral R, Barrera LA, Bounoutas GS, Bumcrot D, Chao H, Ciulla DM, DaSilva JA, Dass A, Dhanapal V, Fennell TJ, Friedland AE, Giannoukos G, Gloskowski SW, Glucksmann A, Gotta GM, Jayaram H, Haskett SJ, et al (2019) Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nat. Med.* **25**: 229–233
- Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE & Church GM (2013) RNA-guided human genome engineering via Cas9. *Science* **339**: 823–6 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23287722> [Accessed April 10, 2019]
- Mao Z, Bozzella M, Seluanov A & Gorbunova V (2009) Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst)*. **7**: 1765–1771 Available at: <http://www.biochemj.org/bj/imps/abs/BJ20141236.htm>
- Marteijn JA, Lans H, Vermeulen W & Hoeijmakers JHJ (2014) Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Cell Biol.* **15**: 465–481 Available at: <http://dx.doi.org/10.1038/nrm3822>
- Mercer J, Mahmoudi M & Bennett M (2007) DNA damage, p53, apoptosis and vascular disease. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **621**: 75–86 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17382357> [Accessed March 31, 2020]
- Meyenberg M, Ferreira da Silva J & Loizou JI (2021) Tissue Specific DNA Repair Outcomes Shape the Landscape of Genome Editing. *Front. Genet.* **12**: 728520 Available at: </pmc/articles/PMC8446275/> [Accessed November 12, 2021]
- Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, Dulay GP, Hua KL, Ankoudinova I, Cost GJ, Urnov FD, Zhang HS, Holmes MC, Zhang L, Gregory PD & Rebar EJ (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29**: 143–150

- Mitchell RB & Dolan ME (1993) Effect of temozolomide and dacarbazine on O6-alkylguanine-DNA alkyltransferase activity and sensitivity of human tumor cells and xenografts to 1,3-bis(2-chloroethyl)-1-nitrosourea. *Cancer Chemother. Pharmacol.* **32**: 59–63 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8462125> [Accessed March 31, 2020]
- Mojica FJM, Díez-Villaseñor C, García-Martínez J & Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**: 733–740
- Mojica FJM, Díez-Villaseñor C, García-Martínez J & Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**: 174–182
- Musunuru K, Chadwick AC, Mizoguchi T, Garcia SP, DeNizio JE, Reiss CW, Wang K, Iyer S, Dutta C, Clendaniel V, Amaonye M, Beach A, Berth K, Biswas S, Braun MC, Chen HM, Colace T V., Ganey JD, Gangopadhyay SA, Garrity R, et al (2021) In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates. *Nature* **593**: 429–434 Available at: <https://www.nature.com/articles/s41586-021-03534-y> [Accessed October 28, 2021]
- Nakade S, Tsubota T, Sakane Y, Kume S, Sakamoto N, Obara M, Daimon T, Sezutsu H, Yamamoto T, Sakuma T & Suzuki KIT (2014) Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat. Commun.* **5**: 1–8
- Nelson CE, Hakim CH, Ousterout DG, Thakore PI, Moreb EA, Castellanos Rivera RM, Madhavan S, Pan X, Ran FA, Yan WX, Asokan A, Zhang F, Duan D & Gersbach CA (2016) In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science (80-.)*. **351**: 403–407
- Newby GA, Yen JS, Woodard KJ, Mayuranathan T, Lazzarotto CR, Li Y, Sheppard-Tillman H, Porter SN, Yao Y, Mayberry K, Everette KA, Jang Y, Podracky CJ, Thaman E, Lechavue C, Sharma A, Henderson JM, Richter MF, Zhao KT, Miller SM, et al (2021) Base editing of haematopoietic stem cells rescues sickle cell disease in mice. *Nature* **595**: 295–302 Available at: <https://pubmed.ncbi.nlm.nih.gov/34079130/> [Accessed October 28, 2021]
- Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, Tabata M, Mochizuki M, Miyabe A, Araki M, Hara KY, Shimatani Z & Kondo A (2016) Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science (80-.)*. **353**:
- Nishimasu H, Shi X, Ishiguro S, Gao L, Hirano S, Okazaki S, Noda T, Abudayyeh OO, Gootenberg JS, Mori H, Oura S, Holmes B, Tanaka M, Seki M, Hirano H, Aburatani H, Ishitani R, Ikawa M, Yachie N, Zhang F, et al (2018) Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science (80-.)*. **361**: 1259–1262 Available at: <http://www.sciencemag.org/lookup/doi/10.1126/science.aas9129> [Accessed September 21, 2018]

- Noll DM, McGregor Mason T & Miller PS (2006) Formation and Repair of Interstrand Cross-Links in DNA. *ChemInform* **37**: 277–301
- Noordermeer SM, Adam S, Setiাপutra D, Barazas M, Pettitt SJ, Ling AK, Olivieri M, Álvarez-Quilón A, Moatti N, Zimmermann M, Annunziato S, Krastev DB, Song F, Brandsma I, Frankum J, Brough R, Sherker A, Landry S, Szilard RK, Munro MM, et al (2018) The shieldin complex mediates 53BP1-dependent DNA repair. *Nature* **560**: 117–121
- Nospikel T (2009) Nucleotide excision repair: Variations on versatility. *Cell. Mol. Life Sci.* **66**: 994–1009 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19153657> [Accessed March 31, 2020]
- Nuñez JK, Chen J, Pommier GC, Cogan JZ, Replogle JM, Adriaens C, Ramadoss GN, Shi Q, Hung KL, Samelson AJ, Pogson AN, Kim JYS, Chung A, Leonetti MD, Chang HY, Kampmann M, Bernstein BE, Hovestadt V, Gilbert LA & Weissman JS (2021) Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell* **184**: 2503-2519.e17
- van Overbeek M, Capurso D, Carter MM, Thompson MS, Frias E, Russ C, Reece-Hoyes JS, Nye C, Gradia S, Vidal B, Zheng J, Hoffman GR, Fuller CK & May AP (2016) DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol. Cell* **63**: 633–646 Available at: <http://dx.doi.org/10.1016/j.molcel.2016.06.037>
- Özcan A, Krajcski R, Ioannidi E, Lee B, Gardner A, Makarova KS, Koonin E V., Abudayyeh OO & Gootenberg JS (2021) Programmable RNA targeting with the single-protein CRISPR effector Cas7-11. *Nat. 2021 5977878* **597**: 720–725 Available at: <https://www.nature.com/articles/s41586-021-03886-5> [Accessed November 13, 2021]
- Palombo F, Iaccarino I, Nakajima E, Ikejima M, Shimada T & Jiricny J (1996) hMutS β , a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. *Curr. Biol.* **6**: 1181–1184 Available at: <http://www.cell.com/article/S0960982202706854/fulltext> [Accessed August 31, 2021]
- Park RJ, Wang T, Koundakjian D, Hultquist JF, Lamothe-Molina P, Monel B, Schumann K, Yu H, Krupczak KM, Garcia-Beltran W, Piechocka-Trocha A, Krogan NJ, Marson A, Sabatini DM, Lander ES, Hacohen N & Walker BD (2017) A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors. *Nat. Genet.* **49**: 193–203
- Parnas O, Jovanovic M, Eisenhaure TM, Herbst RH, Dixit A, Ye CJ, Przybylski D, Platt RJ, Tirosch I, Sanjana NE, Shalem O, Satija R, Raychowdhury R, Mertins P, Carr SA, Zhang F, Hacohen N & Regev A (2015) A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* **162**: 675–686
- Pascucci B, Stucki M, Jónsson ZO, Dogliotti E & Hübscher U (1999) Long patch base excision repair with purified human proteins. DNA ligase I as patch size mediator for DNA polymerases δ and ϵ . *J. Biol. Chem.* **274**: 33696–33702

- Paulsen BS, Mandal PK, Frock RL, Boyraz B, Yadav R, Upadhyayula S, Gutierrez-Martinez P, Ebina W, Fasth A, Kirchhausen T, Talkowski ME, Agarwal S, Alt FW & Rossi DJ (2017) Ectopic expression of RAD52 and dn53BP1 improves homology-directed repair during CRISPR–Cas9 genome editing. *Nat. Biomed. Eng.*: 1 Available at: <http://www.nature.com/articles/s41551-017-0145-2>
- Pearl LH, Schierz AC, Ward SE, Al-Lazikani B & Pearl FMG (2015) Therapeutic opportunities within the DNA damage response. *Nat. Rev. Cancer* **15**: 166–180 Available at: <http://www.nature.com/doi/10.1038/nrc3891> [Accessed January 5, 2017]
- Petri K, Zhang W, Ma J, Schmidts A, Lee H, Horng JE, Kim DY, Kurt IC, Clement K, Hsu JY, Pinello L, Maus M V., Joung JK & Yeh J-RJ (2021) CRISPR prime editing with ribonucleoprotein complexes in zebrafish and primary human cells. *Nat. Biotechnol.* **2021**: 1–5 Available at: <https://www.nature.com/articles/s41587-021-00901-y> [Accessed August 31, 2021]
- Pourcel C, Salvignol G & Vergnaud G (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**: 653–663
- Qasim W, Zhan H, Samarasinghe S, Adams S, Amrolia P, Stafford S, Butler K, Rivat C, Wright G, Somana K, Ghorashian S, Pinner D, Ahsan G, Gilmour K, Lucchini G, Inglott S, Mifsud W, Chiesa R, Peggs KS, Chan L, et al (2017) Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells. *Sci. Transl. Med.* **9**: Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28123068> [Accessed April 1, 2020]
- Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP & Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**: 1173–1183
- Qin W & Wang H (2019) Delivery of CRISPR-Cas9 into mouse zygotes by electroporation. In *Methods in Molecular Biology* pp 179–190. Humana Press Inc. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/30353514> [Accessed April 2, 2020]
- Rees HA, Yeh WH & Liu DR (2019) Development of hRad51–Cas9 nickase fusions that mediate HDR without double-stranded breaks. *Nat. Commun.* **10**: 1–12 Available at: <https://www.nature.com/articles/s41467-019-09983-4> [Accessed September 2, 2021]
- Ren CH, Yan Q & Zhang ZY (2014) Minimum length of direct repeat sequences required for efficient homologous recombination induced by zinc finger nuclease in yeast. *Mol. Biol. Rep.* **41**: 6939–6948 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25024047> [Accessed March 31, 2020]
- Replogle JM, Norman TM, Xu A, Hussmann JA, Chen J, Cogan JZ, Meer EJ, Terry JM, Riordan DP, Srinivas N, Fiddes IT, Arthur JG, Alvarado LJ, Pfeiffer KA, Mikkelsen TS, Weissman JS & Adamson B (2020) Combinatorial single-cell CRISPR screens by direct

- guide RNA capture and targeted sequencing. *Nat. Biotechnol.*: 1–8 Available at: <http://www.nature.com/articles/s41587-020-0470-y> [Accessed April 1, 2020]
- Richardson CD, Kazane KR, Feng SJ, Zelin E, Bray NL, Schäfer AJ, Floor SN & Corn JE (2018) CRISPR–Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. *Nat. Genet.* Available at: <http://www.nature.com/articles/s41588-018-0174-0> [Accessed July 28, 2018]
- Richardson CD, Ray GJ, DeWitt MA, Curie GL & Corn JE (2016) Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**: 339–344 Available at: <http://dx.doi.org/10.1038/nbt.3481>
- Riesenberg S & Maricic T (2018) Targeting repair pathways with small molecules increases precise genome editing in pluripotent stem cells. *Nat. Commun.* **9**: 1–9
- Robert F, Barbeau M, Éthier S, Dostie J & Pelletier J (2015) Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. *Genome Med.* **7**: 93 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26307031> [Accessed April 2, 2020]
- Rosenberg BR, Hamilton CE, Mwangi MM, Dewell S & Papavasiliou FN (2011) Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3' UTRs. *Nat. Struct. Mol. Biol.* **18**: 230–238 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21258325> [Accessed April 1, 2020]
- Sanjana NE, Shalem O & Zhang F (2014) Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**: 783–784 Available at: <http://www.nature.com/articles/nmeth.3047> [Accessed April 10, 2019]
- Scaffidi P & Misteli T (2006) Lamin A-dependent nuclear defects in human aging. *Science* (80-). **312**: 1059–1063 Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1855250&tool=pmcentrez&rendertype=abstract> [Accessed March 31, 2020]
- Schene IF, Joore IP, Oka R, Mokry M, van Vugt AHM, van Boxtel R, van der Doef HPJ, van der Laan LJW, Verstegen MMA, van Hasselt PM, Nieuwenhuis EES & Fuchs SA (2020) Prime editing for functional repair in patient-derived disease models. *Nat. Commun.* **11**: Available at: <https://pubmed.ncbi.nlm.nih.gov/33097693/> [Accessed October 28, 2021]
- Schep R, Brinkman EK, Leemans C, Vergara X, van der Weide RH, Morris B, van Schaik T, Manzo SG, Peric-Hupkes D, van den Berg J, Beijersbergen RL, Medema RH & van Steensel B (2021) Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. *Mol. Cell*
- Scholefield J & Harrison PT (2021) Prime editing – an update on the field. *Gene Ther.* **2021** 287 **28**: 396–401 Available at: <https://www.nature.com/articles/s41434-021-00263-9> [Accessed November 10, 2021]

- Schrempf A, Slyskova J & Loizou JI (2021) Targeting the DNA Repair Enzyme Polymerase θ in Cancer Therapy. *Trends in Cancer* **7**: 98–111 Available at: <http://www.cell.com/article/S2405803320302636/fulltext> [Accessed November 2, 2021]
- Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, Van Der Oost J, Brouns SJJ & Severinov K (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U. S. A.* **108**: 10098–10103
- Seol JH, Shim EY & Lee SE (2018) Microhomology-mediated end joining: Good, bad and ugly. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **809**: 81–87
- Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG & Zhang F (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**: 84–7 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24336571> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4089965>
- Shen MW, Arbab M, Hsu JY, Worstell D, Culbertson SJ, Krabbe O, Cassa CA, Liu DR, Gifford DK & Sherwood RI (2018) Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* Available at: <http://www.nature.com/articles/s41586-018-0686-x>
- Shou J, Li J, Liu Y & Wu Q (2018) Precise and Predictable CRISPR Chromosomal Rearrangements Reveal Principles of Cas9-Mediated Nucleotide Insertion. *Mol. Cell* **71**: 498-509.e4 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/30033371> [Accessed April 2, 2020]
- Shrivastav M, De Haro LP & Nickoloff JA (2008) Regulation of DNA double-strand break repair pathway choice. *Cell Res.* **18**: 134–147 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18157161> [Accessed March 31, 2020]
- Sinkunas T, Gasiunas G, Waghmare SP, Dickman MJ, Barrangou R, Horvath P & Siksnys V (2013) In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J.* **32**: 385–394 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23334296> [Accessed April 1, 2020]
- Slymaker IM, Gao L, Zetsche B, Scott DA, Yan WX & Zhang F (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science (80-.)*. **351**: 84–88 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26628643> [Accessed April 2, 2020]
- Srivastava M, Nambiar M, Sharma S, Karki SS, Goldsmith G, Hegde M, Kumar S, Pandey M, Singh RK, Ray P, Natarajan R, Kelkar M, De A, Choudhary B & Raghavan SC (2012) An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression. *Cell* **151**: 1474–1487
- Sternberg SH, Redding S, Jinek M, Greene EC & Doudna JA (2014) DNA interrogation by the

- CRISPR RNA-guided endonuclease Cas9. *Nature* **507**: 62–67
- Stratton MR, Campbell PJ & Futreal PA (2009) The cancer genome. *Nature* **458**: 719–724
- Strobel SA & Dervan PB (1990) Site-specific cleavage of a yeast chromosome by oligonucleotide-directed triple-helix formation. *Science (80-)*. **249**: 73–75 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2195655> [Accessed April 1, 2020]
- Strobel SA & Dervan PB (1991) Single-site enzymatic cleavage of yeast genomic DNA mediated by triple helix formation. *Nature* **350**: 172–174
- Sullenger BA & Cech TR (1994) Ribozyme-mediated repair of defective mRNA by targeted trans-splicing. *Nature* **371**: 619–622
- Sun CL, Barrangou R, Thomas BC, Horvath P, Fremaux C & Banfield JF (2013) Phage mutations in response to CRISPR diversification in a bacterial population. *Environ. Microbiol.* **15**: 463–470 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23057534> [Accessed April 1, 2020]
- Supek F & Lehner B (2015) Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* **521**: 81–84 Available at: <https://www.nature.com/articles/nature14173> [Accessed November 6, 2021]
- Sürün D, Schneider A, Mircetic J, Neumann K, Lansing F, Paszkowskirogacz M, Hänchen V, Leekirsch MA & Buchholz F (2020) Efficient generation and correction of mutations in human iPS cells utilizing mRNAs of CRISPR base editors and prime editors. *Genes (Basel)*. **11**: Available at: <https://pubmed.ncbi.nlm.nih.gov/32384610/> [Accessed October 28, 2021]
- Tabebordbar M, Zhu K, Cheng JKW, Chew WL, Widrick JJ, Yan WX, Maesner C, Wu EY, Xiao R, Ran FA, Cong L, Zhang F, Vandenberghe LH, Church GM & Wagers AJ (2016) In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science (80-)*. **351**: 407–411
- Thomas A, Tanaka M, Trepel J, Reinhold WC, Rajapakse VN & Pommier Y (2017) Temozolomide in the era of precision medicine. *Cancer Res.* **77**: 823–826 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28159862> [Accessed March 31, 2020]
- Trojan J, Zeuzem S, Randolph A, Hemmerle C, Brieger A, Raedle J, Plotz G, Jiricny J & Marra G (2002) Functional analysis of hMLH1 variants and HNPCC-related mutations using a human expression system. *Gastroenterology* **122**: 211–219 Available at: <https://pubmed.ncbi.nlm.nih.gov/11781295/> [Accessed November 16, 2021]
- Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar V V., Thapar V, Wyvekens N, Khayter C, Iafrate AJ, Le LP, Aryee MJ & Joung JK (2015) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**: 187–198
- Umar A, Koi M, Risinger JI, Glaab WE, Tindall KR, Kolodner RD, Boland CR, Barrett JC & Kunkel TA (1997) Correction of hypermutability, N-Methyl-N'-nitro-N-nitrosoguanidine

- resistance, and defective dna mismatch repair by introducing chromosome 2 into human tumor cells with mutations in MSH2 and MSH6. *Cancer Res.* **57**: 3949–3955
- Vilenchik MM & Knudson AG (2003) Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer. *Proc. Natl. Acad. Sci. U. S. A.* **100**: 12871–12876
- Wagner DL, Amini L, Wendering DJ, Burkhardt LM, Akyüz L, Reinke P, Volk HD & Schmueck-Henneresse M (2019) High prevalence of *Streptococcus pyogenes* Cas9-reactive T cells within the adult human population. *Nat. Med.* **25**: 242–248
- Walton RT, Christie KA, Whittaker MN & Kleinstiver BP (2020) Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science (80-.)*: eaba8853 Available at: <https://www.sciencemag.org/lookup/doi/10.1126/science.aba8853> [Accessed April 1, 2020]
- Wang AS, Chen L, Wu RA, Richardson CD, Gowen BG, Kazane KR, Vu JT, Wyman SK, Shin J, Walter JC & Corn JE (2019) The histone chaperone FACT induces Cas9 multi-turnover behavior and modifies genome manipulation in human cells. *bioRxiv*: 705657 Available at: <https://www.biorxiv.org/content/10.1101/705657v1> [Accessed April 1, 2020]
- Wang D, Zhang F & Gao G (2020) CRISPR-Based Therapeutic Genome Editing: Strategies and In Vivo Delivery by AAV Vectors. *Cell* **181**: 136–150
- Wang T, Wei JJ, Sabatini DM & Lander ES (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**: 80–4 Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3972032&tool=pmcentrez&rendertype=abstract>
- Watson JD & Crick FHC (1953) Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* **171**: 737–738
- Weissman L, de Souza-Pinto NC, Stevnsner T & Bohr VA (2007) DNA repair, mitochondria, and neurodegeneration. *Neuroscience* **145**: 1318–1329 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17092652> [Accessed March 31, 2020]
- Wienert B, Nguyen DN, Guenther A, Feng SJ, Locke MN, Wyman SK, Shin J, Kazane KR, Gregory GL, Carter MAM, Wright F, Conklin BR, Marson A, Richardson CD & Corn JE (2020) Timed inhibition of CDC7 increases CRISPR-Cas9 mediated templated repair. *Nat. Commun.* **11**: 2109 Available at: <http://www.nature.com/articles/s41467-020-15845-1> [Accessed May 4, 2020]
- Wienert B, Wyman SK, Richardson CD, Yeh CD, Akcakaya P, Porritt MJ, Morlock M, Vu JT, Kazane KR, Watry HL, Judge LM, Conklin BR, Maresca M & Corn JE (2019) Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science (80-.)*. **364**: 286–289

- Wiseman H & Halliwell B (1996) Damage to DNA by reactive oxygen and nitrogen species: Role in inflammatory disease and progression to cancer. *Biochem. J.* **313**: 17–29
- Wu Y, Zeng J, Roscoe BP, Liu P, Yao Q, Lazzarotto CR, Clement K, Cole MA, Luk K, Baricordi C, Shen AH, Ren C, Esrick EB, Manis JP, Dorfman DM, Williams DA, Biffi A, Brugnara C, Biasco L, Brendel C, et al (2019) Highly efficient therapeutic gene editing of human hematopoietic stem cells. *Nat. Med.* **25**: 776–783
- Yang J, Zimmerly S, Perlman PS & Lambowitz AM (1996) Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature* **381**: 332–335 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8692273> [Accessed April 1, 2020]
- Yang Y, Wang L, Bell P, McMenamin D, He Z, White J, Yu H, Xu C, Morizono H, Musunuru K, Batshaw ML & Wilson JM (2016) A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nat. Biotechnol.* **34**: 334–338 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26829317> [Accessed April 2, 2020]
- Yeh CD, Richardson CD & Corn JE (2019) Advances in genome editing through control of DNA repair pathways. *Nat. Cell Biol.* **21(12)**: 1468–1478 Available at: <http://dx.doi.org/10.1038/s41556-019-0425-z>
- Yu W, Mookherjee S, Chaitankar V, Hiriyan S, Kim JW, Brooks M, Ataeijannati Y, Sun X, Dong L, Li T, Swaroop A & Wu Z (2017) Nrl knockdown by AAV-delivered CRISPR/Cas9 prevents retinal degeneration in mice. *Nat. Commun.* **8**: 1–15
- Zetsche B, Gootenberg JS, Abudayeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, Van Der Oost J, Regev A, Koonin E V. & Zhang F (2015) Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163**: 759–771
- Zipkin M (2019) CRISPR’s “magnificent moment” in the clinic. *Nat. Biotechnol.*
- Zuccaro M V., Xu J, Mitchell C, Marin D, Zimmerman R, Rana B, Weinstein E, King RT, Palmerola KL, Smith ME, Tsang SH, Goland R, Jasin M, Lobo R, Treff N & Egli D (2020) Allele-Specific Chromosome Removal after Cas9 Cleavage in Human Embryos. *Cell* **183**: 1650-1664.e15 Available at: <https://pubmed.ncbi.nlm.nih.gov/33125898/> [Accessed February 12, 2021]
- Zuris JA, Thompson DB, Shu Y, Guilinger JP, Bessen JL, Hu JH, Maeder ML, Joung JK, Chen ZY & Liu DR (2015) Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* **33**: 73–80 Available at: <http://www.nature.com/articles/nbt.3081> [Accessed April 2, 2020]

Curriculum Vitae

Name: Joana Ferreira da Silva

Date of Birth: 10.10.1992

Nationality: Portuguese

Academic Degree: MSc

Email: joana.tfsilva@gmail.com

Education

- 2017-current **PhD in Molecular Medicine (DOC Fellowship, from the Austrian Academy of Sciences)**
CeMM – Research Center for Molecular Medicine (Vienna, Austria)
Medical University of Vienna (MUW) (Vienna, Austria)
Supervisor: Joanna Loizou
Project: Fundamental Insights of DNA Repair for CRISPR-Cas9-Mediated Genome Editing
- 2014-2016 **MSc in Molecular Genetics and Biomedicine**
NOVA University of Lisbon, (Lisbon, Portugal)
Grade: 19/20
Project: New insights into IL-7R α mutational activation in the context of T-Cell Acute Lymphoblastic Leukaemia (T-ALL)
- 2010-2013 **BSc in Biology, with major in Molecular Biology & Genetics**
University of Lisbon, (Lisbon, Portugal)
Grade: 17/20

Research Experience

- 2016 **Laboratory Rotation**
CeMM – Research Center for Molecular Medicine (Vienna, Austria)
Supervisor: Joerg Menche
Project: Development of a network-based approach for the identification of synthetic rescue interactions for Fanconi Anaemia
- 2015-2016 **Master Thesis Internship**
Institute of Molecular Medicine (iMM) (Lisbon, Portugal)
Supervisor: João Barata
Project: New Insights into IL-7R α mutational activation in the context of T-Cell Acute Lymphoblastic Leukaemia (T-ALL)

- 2014 **Scholarship Researcher PARSUK Xperience**
University of Hull (Hull, UK)
 Supervisor: Isabel Pires
 Project: Investigation of hypoxia-mediated cell invasion using 3D cancer models
- 2012-2013 **Scholarship Researcher (Amadeus Dias Foundation)**
BioISI, Biosystems and Integrative Sciences Institute (Lisbon, Portugal)
 Supervisor: Sandra Chaves
 Project: Isolation and Characterisation of UV-resistant yeast strains

Presentation at Meetings & Conferences

- 2019 EU-LIFE: Signalling & Gene Regulation in Health & Disease. Cambridge, UK (poster presentation);
- 2018 Precision CRISPR. London, UK (poster presentation);
- 2017 The DNA Damage Response in Cell Physiology and Disease. Athens, Greece (poster presentation);
- 2016 Associacao Portuguesa de Investigacao em Cancro: International Congress;
- 2013 MICROBIOTEC'13: Portuguese Congress of Microbiology. and Biotechnology. Aveiro, Portugal (poster presentation).

Fellowships

- 2018-2021 DOC Fellowship from the Austrian Academy of Sciences;
- 2014 PARSUK Xperience Fellowship;
- 2012 Amadeu Dias Foundation Fellowship.

Honours & Awards

- 2021 Scholarship to attend KeyStone eSymposium on Precision Engineering of the Genome, Epigenome and Transcriptome;
- 2020 Nominee of the Austrian Academy of Sciences and selected participant to attend the 70th LINDAU Nobel Laureate Meeting;
- 2016 Professor Luis Archer Award: Best classification 2014/2015 for the Masters of Molecular Medicine and Genetics;
- 2014 Graduated with Academic Honours: honour mention attributed for the 1st cycle studies and Integrated Masters with best classification in 2012/2013;
- 2011 FCUL Foundation Award: Best students from 1st year, 2010/2011;
- 2010 Aquilino Ribeiro Revelation Award: creative literary writing.

List of Publications

(* shared first author)

1. **Ferreira da Silva J***, Oliveira GP*, Arasa-Verge EA, Moretton A, Timelthaler G, Jiricny J, Loizou JI. Prime Editing Efficiency and Fidelity are Enhanced in the Absence of Mismatch Repair. *bioRxiv* (2021), *in press in Nature Communications*;
2. Vulliard L, Hancock J, Kamnev A, Fell CW, **Ferreira da Silva J**, Loizou JI, Nagy V, Dupre L, Menche J. BioProfiling.jl: Profiling biological perturbations with high-content imaging in single cells and heterogeneous populations. *Bioinformatics* (2021);
3. Meyenberg M*, **Ferreira da Silva J***, Loizou JI. Tissue Specific DNA Repair Outcomes Shape the Landscape of Genome Editing, *Frontiers in Genetics* (2021);
4. **Ferreira da Silva J***, Meyenberg M*, Loizou JI. Tissue Specificity of DNA Repair: the CRISPR Compass. *Trends in Genetics* (2021);
5. **Ferreira da Silva J**, Salic S, Wiedner M, Datlinger P, Essletzbichler P, Hanzl A, Superti-Furga G, Bock C, Winter G, Loizou JI. Genome-scale CRISPR Screens are Efficient in Non-Homologous End-Joining Deficient Cells. *Sci Reports* (2019);
6. Robinson-Garcia L*, **Ferreira da Silva J***, Loizou JI. Synthetic Lethal Interactions for Kinase Deficiencies to DNA Damage Chemotherapeutics. *AACR Cancer Research* (2019);
7. Owusu M, Bannauer P, **Ferreira da Silva J**, Mourikis TP, Jones A, Majek P, Caldera M, Wieder M, Lardeau CH, Mueller AC, Menche J, Kubicek S, Ciccarelli FD, Loizou JI. Mapping the Human Kinome in Response to DNA Damage, *Cell Reports* (2019)
8. Velimezi G, Robinson-Garcia L, Munoz-Martinez F, Wiegant WW, **Ferreira da Silva J**, Owusu M, Moder M, Wiedner M, Brin Rosenthal S, Fisch KM, Menche J, Van Attikum H, Jackson SP, Loizou JI, Map of Synthetic Rescue Interactions for the Fanconi Anemia DNA Repair Pathway Identifies USP48. *Nature Communications* (2018);
9. Mazouzi A, Battistini F, Moser S, **Ferreira da Silva J**, Wiedner M, Lardeau CH, Ringler A, Weil B, Sapparbaev MK, Neesen J, Kubicek S, Loizou JI. Repair of UV-Induced DNA Damage Independent of Nucleotide Excision Repair is Masked by MUTYH. *Molecular Cell* (2017);
10. Moder M*, Velimezi G*, Owusu M, Mazouzi A, Wiedner M, **Ferreira da Silva J**, Robinson-Garcia L, Schischlik F, Slavkovsky R, Kralovics R, Schuster M, Bock C, Ideker T, Jackson SP, Menche J, Loizou JI. Parallel Genome-Wide Screens Identify a Synthetic Viable Interaction Between the BLM Helicase complex and Fanconi Anemia. *Nature Communications* (2017).