From Department of Microbiology, Tumor and Cell Biology
Karolinska Institutet, Stockholm, Sweden

PERKS AND CONSIDERATIONS WHEN TARGETING FUNCTIONAL NON-CODING REGIONS WITH CRISPR/CAS9

Keyi Geng

Karolinska Institutet

Stockholm 2023

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2023

ISBN 978-91-8017-024-6
Cover illustration: “Stubborn and Sneaky DNA”. Artwork by John Svetoft. Permission granted by the artist.
The illustration depicts the unintended genomic alterations associated with the application of CRISPR/Cas9. The targeted DNA sequence (represented by the red fragment, the stubborn DNA) appears resistant to the deletion, while a mixture of unwanted DNA fragments from various sources (represented by the multi-colored fragments, the sneaky DNA) have infiltrated the break sites, leading to integration. Without the aid of appropriate tools (represented by the magnifier), it may be challenging for researchers to accurately detect and analyze these alterations.

Perks and considerations when targeting functional non-coding regions with CRISPR/Cas9

Thesis for Doctoral Degree (Ph.D.)

By

Keyi Geng

The thesis will be defended in public at Air&Fire, Science for Life Laboratory, Tomtebodavägen 23A, Solna, 9th of June 2023, at 13:00

Principal Supervisor:
Dr. Claudia Kutter
Karolinska Institutet
Department of Department of Microbiology, Tumor and Cell Biology

Co-supervisor(s):
Prof. Cecilia Williams
Karolinska Institutet
Department of Biosciences and Nutrition

Dr. Laura Baranello
Karolinska Institutet
Department of Department of Cell and Molecular Biology

Dr. Philip Ewels
Seqera Labs

Opponent:
Dr. Fredrik Wermeling
Karolinska Institutet
Department of Medicine

Examination Board:
Dr. Adam Ameur
Uppsala Universitet
Department of Immunology, Genetics and Pathology

Prof. Klas Wiman
Karolinska Institutet
Department of Oncology-Pathology

Dr. Carsten Daub
Karolinska Institutet
Department of Biosciences and Nutrition

To my family and all I love

“All we have to decide is what to do with the time that is given us”

—J.R.R. Tolkien

Popular science summary of the thesis

CRISPR/Cas9 is a powerful tool used to edit genes. It is like a tiny molecular scissors that can cut and edit genes, but it's not perfect. There are some challenges that come with using it.

First, how can we efficiently deliver the CRISPR/Cas9 system to as many cells as possible? CRISPR/Cas9 components can be delivered in the form of DNA using vectors. Due to their large size, CRISPR/Cas9 vectors often face challenges in passing through cell membranes. In **Paper I**, we showed that small-size vectors helped large-size vectors to get into cells. By adding small-size vectors to CRISPR/Cas9 vectors, we significantly increased the delivery efficiency and decreased cell death. This method can greatly benefit CRISPR/Cas9 users and increase the efficiency of achieving modified cells.

Second, after CRISPR/Cas9 cuts its target, are the two ends of the DNA break directly bound to each other? If not, what would happen at the break site? In **Paper II**, we utilized an advanced and customized approach to investigate the on-target DNA sequence after the CRISPR/Cas9-induced break was repaired. Surprisingly, many undesired scenarios co-occurred at the target site. The targeted DNA sequences could get duplicated and inverted, and DNA fragments belonging to CRISPR/Cas9 vectors or some other regions in the human genome were inserted into the break site.

Lastly, we asked the question when CRISPR/Cas9 is used to cut the target, does the rest of the genome remain unchanged? In **Paper III**, we reported an unexpected large genomic deletion in one commonly used human cell line. This deletion contains many important genes and was found in various cell clones when using CRISPR/Cas9. We observed both molecular and cellular changes in cells carrying this deletion. Although the deleted genomic region was not cleaved by Cas9, the generation of CRISPR/Cas9-modified cells increased the likelihood of cells losing this genomic region. We found that this deletion can be connected to a common deletion event among cancer patients, suggesting the clinical significance of our work.

In summary, this thesis addresses several considerations for targeting the human genome using CRISPR/Cas9, and introduces methods to ensure the efficiency and safety of its application.

Abstract

Since the CRISPR system was discovered as an adaptive immune response in prokaryotic cells, the past decade has witnessed the engineering and deployment of CRISPR/Cas9 as one of the most efficient and powerful molecular tools. By leveraging the nuclease activity of CRISPR/Cas9, researchers are able to probe the biological functions of genetic elements and dissect molecular interactions by disrupting, activating or inactivating genes. In addition to biological research, the CRISPR/Cas9 toolkit has profoundly revolutionized gene therapy and agricultural products. However, there are many challenges regarding its efficiency, specificity and safety. Continuous efforts are being made to advance techniques and characterize the consequences of genome editing. In this thesis, we describe considerations when targeting genomic regions with CRISPR/Cas9 and provide methods to address some concerns related to efficiency and safety.

In **Paper I**, we introduced a non-hazardous method of transfecting human cells with large-size CRISPR/Cas9 vectors. By co-transfecting small-size vectors (3 kb) to cells, the delivery efficiency of CRISPR/Cas9 vectors (15 kb) and cell viability was significantly increased. The performance of the method has been verified in a number of hard-to-transfect human cell lines with both electroporation- and liposome-based transfection.

In **Paper II**, we revealed the complexity of CRISPR/Cas9-induced on-target genomic alterations by combining an advanced droplet-based target enrichment method followed by long-read sequencing and *de novo* assembly-based analysis. This approach enabled us to dissect the on-target sequence content in the order of kilobases, which was very challenging with many other available methods. With this tool, we uncovered the co-occurrence of multiple on-target rearrangements including duplication, inversion, as well as integrations of exogenous DNA and clustered interchromosomal rearrangements in CRISPR/Cas9-modified human cells. Furthermore, our study demonstrated that unintended genomic alterations could lead to the expression of DNA derived from both the target region and exogenous sources, as well as affect cell proliferation.

In **Paper III**, we reported a large unexpected genomic deletion in the HAP1 cell line, which is the one of most popular models used in CRISPR/Cas9-mediated experiments. This 287 kb deletion located on Chromosome 10 contains four widely-expressed protein-coding genes including the *PTEN* gene locus. We detected changes in histone acetylation and transcriptomes in HAP1 cells carrying the deletion. The loss of this genomic locus was not induced by Cas9 off-target nuclease activity. However, the generation of CRISPR/Cas9-modified cells significantly enhanced the frequency of the deletion among cell clones. Furthermore, our analysis indicated that this deletion initially found in HAP1 cells resembled a frequent deletion pattern driven by the *PTEN* gene in cancer patients.

In conclusion, we have presented two methods: one to improve delivery efficiency and another to detect on-target sequence content with higher resolution. Furthermore, we have revealed unintended genomic aberrations at targeted and non-targeted sites. These observations should be taken into consideration when modifying the genome with CRISPR/Cas9, and a comprehensive genomic validation is necessary.

List of scientific papers

I. Jonas Nørskov Søndergaard, **Keyi Geng***, Christian Sommerauer*, Ionut Atanasoai, Xiushan Yin, Claudia Kutter
*Authors contributed equally to this study
Successful delivery of large-size CRISPR/Cas9 vectors in hard-to-transfect human cells using small plasmids
Communications Biology. 2020 Jun 19;3(1):319.
doi: 10.1038/s42003-020-1045-7

II. **Keyi Geng**, Lara G. Merino, Linda Wedemann, Aniek Martens, Małgorzata Sobota, Yerma P. Sanchez, Jonas Nørskov Søndergaard, Robert J. White, Claudia Kutter
Target-enriched nanopore sequencing and *de novo* assembly reveals co-occurrences of complex on-target genomic rearrangements induced by CRISPR-Cas9 in human cells
Genome Research. 2022 Oct 1;32(10):1876-91.
doi: 10.1101/gr.276901.122

III. **Keyi Geng**, Lara G. Merino, Raül G. Veiga, Christian Sommerauer, Janine Epperlein, Eva K. Brinkman, Claudia Kutter
CRISPR-Cas9-mediated genome engineering exaggerates genomic deletion at 10q23.31 including the *PTEN* gene locus mimicking cancer profiles
Pre-print, bioRxiv April 2023
doi: 10.1101/2023.04.05.535680

Scientific papers not included in the thesis

I. Jonas Nørskov Søndergaard, Christian Sommerauer, Ionut Atanasoai, Laura C. Hinte, **Keyi Geng**, Giulia Guiducci, Lars Bräutigam, Myriam Aouadi, Lovorka Stojic, Isabel Barragan, Claudia Kutter
CCT3-*LINC00326* axis regulates hepatocarcinogenic lipid metabolism
Gut. 2022 Oct 1;71(10):2081-92.
doi: 10.1136/gutjnl-2021-325109

II. Keying Zhu, Yang Wang*, Heela Sarlus*, **Keyi Geng***, Erik Nutma, Jingxian Sun, Shin-Yu Kung, Cindy Bay, Jinming Han, Jin-Hong Min, Irene Benito-Cuesta, Harald Lund, Sandra Amor, Jun Wang, Xing-Mei Zhang, Claudia Kutter, André Ortlieb Guerreiro-Cacais, Björn Högberg, Robert A Harris
Myeloid cell-specific topoisomerase 1 inhibition using DNA origami mitigates neuroinflammation
*Authors contributed equally to this study
EMBO reports. 2022 Jul 5;23(7):e54499.
doi: 10.15252/embr.202154499

III. Jente Ottenburghs, **Keyi Geng**, Alexander Suh and Claudia Kutter
Genome size reduction and transposon activity impact tRNA gene diversity while ensuring translational stability in birds
Genome Biology and Evolution. 2021 Apr;13(4):evab016.
doi: 10.1093/gbe/evab016

IV. Donald P. Cameron*, Jan Grosser*, Swetlana Ladigan*, Vladislav Kuzin, Evanthia Iliopoulou, Anika Wiegard, Hajar Benredjem, Sven T. Liffers, Smiths Lueong, Phyllis F. Cheung, Deepak Vangala, Michael Pohl, Richard Viebahn, Christian Teschendorf, Heiner Wolters, Selami Usta, **Keyi Geng**, Claudia Kutter, Marie Arsenian-Henriksson, Jens T. Siveke, Andrea Tannapfel, Wolff Schmiegel, Stephan A. Hahn*, Laura Baranello*
*Authors contributed equally to this study
Co-inhibition of topoisomerase 1 and BRD4-mediated pause release selectively kills pancreatic cancer via readthrough transcription
Pre-print, bioRxiv February 2023
doi: 10.1101/2023.02.10.527824

Contents

List of abbreviations

53BP1	TP53-binding protein 1
AAV	adeno-associated virus
APH	aphidicolin
ATM	ataxia telangiectasia mutated
ATR	ATM and RAD3-related
AV	adenovirus
BLESS	breaks labelling, enrichment on streptavidin, and sequencing
BLISS	breaks labelling *in situ* and sequencing
BLM	Bloom syndrome protein
BRCA1	breast cancer type 1 susceptibility protein
BRCA1-BARD1	BRCA1– BRCA1-associated domain 1
Cas9n	Cas9 nickases
CAST-seq	chromosomal aberrations analysis by single targeted linker-mediated PCR sequencing
CFS	common fragile site
ChIP-seq	chromatin immunoprecipitation followed by sequencing
Chr	chromosome
CIRCLE-seq	circularization for *in vitro* reporting of cleavage effects by sequencing
CRISPR	clustered regularly interspaced short palindromic repeats
CRISPR/Cas9	CRISPR associated proteins
crRNA	CRISPR RNA
CtIP	C-terminal binding protein interacting protein
ctrl	control
DAc	differentially acetylated
DC	daughter cell
dCas9	inactive Cas9
DDR	DNA damage response
DE	differentially expressed
dHJ	double Holliday junction
Digenome-seq	digested genome sequencing

dMDA	droplet multiple displacement amplification
DNA-PK	DNA-dependent protein kinase
dPCR	droplet-PCR
dRNAs	dead-RNAs
DSB	double-stranded break
dsODN	double-stranded oligodeoxynucleotide
eccDNA	extrachromosomal circular DNA
EXO1	exonuclease 1
G4	G-quadruplex
gRNA	guide RNA
GUIDE-seq	genome-wide, unbiased identification of DSBs evaluated by sequencing
H3K27ac	histone 3 lysine 27 acetylation
H3K4me3	histone 3 lysine 4 trimethylation
HR	homologous recombination
HTGTS	high-throughput, genome-wide translocation sequencing
HU	hydroxyurea
IDLV	integrase deficient lentiviral vector
InDels	small insertions or deletions
KO	knockout
LAM-PCR	linear amplification-mediated PCR
LD	large deletion
Lig I/III/IV	DNA ligase I/III/IV
LNP	lipid-based nanoparticles
LRS	long-read sequencing
LV	lentivirus
MCM	minichromosome maintenance
MMEJ	microhomology-mediated end joining
MRN complex	meiotic recombination 11 (MRE11) - RAD50 double strand break repair protein (RAD50) - Nijmegen breakage syndrome 1 (NBS1) complex (MRN complex)
NER	nucleotide excision repair

NHEJ	non-homologous end joining
nt	nucleotides
ONT	Oxford Nanopore Technologies
PacBio	Pacific Biosciences
PAM	protospacer-adjacent motif
PARP1	poly ADP-ribose polymerase 1
PEM-seq	primer-extension-mediated sequencing
Pol III	DNA-dependent RNA polymerase III
Pol θ	DNA polymerase θ
pre-crRNA	precursor CRISPR RNA
RNP	ribonucleoproteins
RPA	replication protein A
SAFE donor	'sequence-ascertained favorable editing' donor
SDSA	synthesis-dependent DNA strand annealing
SITE-seq	selective enrichment and identification of tagged genomic DNA ends by sequencing
spCas9	Cas9 from *Streptococcus pyogenes*
ssDNA	single-stranded DNA
SV	structural variation
tracrRNA	trans-activating CRISPR RNA
tRNA	transfer RNA
tru-gRNA	truncated gRNAs
XLF	XRCC4-like factor
XRCC4	X-ray cross complementing group 4
γH2AX	phosphorylation of the serine-139 of the histone H2A variant

1 Literature review

1.1 CRISPR/Cas9

Genome editing enables precise modification of a chosen sequence by introducing a double-stranded break (DSB) and its subsequent repair. In the past, targeted nucleases such as zinc-finger nucleases and transcription activator-like effector nucleases were engineered to recognize and target desired sequences. However, these methods are demanding because the recombinant proteins require customized DNA binding domains. Since the first demonstration of applying clustered regularly interspaced short palindromic repeats (CRISPR) associated proteins (CRISPR/Cas) as editing tools in 2013[1], CRISPR/Cas has revolutionized genome engineering because of its simple programmability, versatile scalability and low cost.

1.1.1 The origin of CRISPR/Cas

CRISPR/Cas was first found as the adaptive immune response system in many bacteria and Archaea, protecting them from the infection by bacteriophages or viruses[2]. The action of CRISPR/Cas systems can be divided into three steps: adaptation, expression and maturation, and interference[3]. During adaptation, Cas proteins identify and bind to the foreign DNA. A target region that is next to the protospacer-adjacent motif (PAM) is then cleaved and integrated into the CRISPR array as a spacer. During expression and maturation, a precursor CRISPR RNA (pre-crRNA) is transcribed from the CRISPR array and further processed into a smaller mature CRISPR RNA (crRNA), forming an active complex with Cas protein. In the last step of interference, the Cas-crRNA complex recognizes the complementary sequence in the invader DNA, induces cleavage and leads to the degradation of the target.

1.1.2 Application of CRISPR/Cas9

Among all the CRISPR/Cas systems, CRISPR/Cas9 has been the most intensively developed and widely applied genome editing tool. In the original system as a defence response, Cas9 is guided by a crRNA and a trans-activating CRISPR RNA (tracrRNA). tracrRNA facilitates the process of pre-crRNA and forms a crRNA-tracrRNA hybrid to recruit Cas9 nucleases[4]. After adapting CRISPR/Cas9 as an editing tool, crRNA and tracrRNA are now designed to fuse, known as guide RNA (gRNA)[1,5,6]. Once the Cas9 enzyme has been guided to its target in the presence of gRNA, the two nuclease domains HNH and RuvC create a DSB three nucleotides (nt) upstream of the PAM[2,4].

The PAM sequence varies among Cas9 proteins found in different bacteria species[7]. The most commonly used Cas9 is from *Streptococcus pyogenes* (spCas9). Besides, there are variants from *Streptococcus thermophiles* and *Staphylococcus aureus*. The PAM sequence for spCas9 is 3'-NGG[8]. To increase the range of available editing sites, new spCas9 variants were developed, allowing PAMs such as 3'-NG or 3'-NRNH[8–10].

Once a DSB is induced, it will be repaired by cellular mechanisms (discussed in subsection **1.2**). The most frequent repair outcomes are small insertions or deletions (InDels) at the targeted site[11]. Consequently, when targeting a protein-coding gene, a frameshift is introduced, and the loss-of-function of the targeted protein-coding gene is achieved. However, InDels are not sufficient when

knocking out non-coding genes and regulatory elements. Additionally, targeting a multi-copy gene family or a repetitive region itself with a single gRNA poses the risk of a high frequency of off-target editing. In these cases, larger genomic regions can be removed from the genome by using dual gRNAs that flank the target region to introduce two DSBs[12,13]. Through this approach, a DNA sequence of up to 1 Mb can be excised[14]. With the presence of exogenous homologous DNA as a donor template, the DSB can be repaired with specific sequences, enabling precise DNA correction or addition.

1.1.2.1 Beyond single-target editing

Compared to other established nucleases used for gene editing, the primary benefit of the CRISPR/Cas9 system is its superior programmability. It allows large-scale screening since only ~20 nt sequences need to be designed and synthesized, which, together with the scaffold, direct the Cas9 enzyme to different targets. Single gRNA libraries have been generated targeting the coding regions[15,16]. After CRISPR/Cas9 components are delivered to cells, a specific phenotype-based selection, such as proliferation or survival competition, is performed. By sequencing gRNAs in the positively selected pool, genes of interest are identified through enriched or depleted gRNAs[15–17]. Besides screening protein-coding genes, CRISPR/Cas9 is used to reveal regulatory elements such as enhancers[18,19].

The CRISPR/Cas9 toolkit has been expanded for visualization and gene expression regulation. To achieve these goals, a catalytically inactive Cas9 (dCas9) was engineered. dCas9 carries mutations that disrupt the HNH and RuvC domains, i.e., dCas9 binds to its target without the ability to induce any DSBs. dCas9 can be fused with different effector domains or proteins depending on the purpose. For example, eGFP-, mCherry- and SunTag-dCas9 fusion were used to visualize telomeres[20–24]. Optimized dCas9-based live-cell imaging methods such as BIFC-dCas9/gRNA and CRISPR/Casilio increase the specificity, efficiency and resolution allowing the study of the dynamics and conformation of chromatin in real-time[22,25]. dCas9 fused with a transcriptional activation (e.g. VP64) or repression (e.g. KRAB) domain is exploited to up- or down-regulate the targeted gene expression[26–28]. Similarly, transcription can be manipulated through dCas9 fusions for epigenome editing. For example, as the core domain of catalytic histone acetyltransferase, p300 directly catalyses the acetylation of histone (H) 3 lysine (K) 27 (H3K27), of which the enrichment is concomitant with increased gene expression. dCas9–p300 fusions recruited to promoters or enhancers have been shown to activate the expression of corresponding genes[29,30].

1.1.2.2 Efficiency

One of the main challenges when using CRISPR/Cas9 is to identify gRNAs with the maximum on-target activity (efficiency) and minimal off-target risks (specificity). The efficiency depends on not only nuclease activity but also the binding affinity and the accessibility of the targeted region. Due to a lack of knowledge on the exact behaviour of Cas9 interacting with DNA dynamically in living cells, the on-target efficiency of a gRNA still needs to be tested experimentally. Nevertheless, information was extracted from large-scale screenings and analysed to link certain features of the sequence to the on-target activity.

According to the PAM sequence of Cas9, although any nucleotide can precede the two guanines (G), cytosines (C) and thymines (T) are the most and least favoured in this position, respectively[31,32]. Besides

the PAM, other gRNA features also determine the efficiency. For example, poly-N sequences should be avoided in gRNA design due to poor oligo synthesis and many other reasons[33]. GGGG usually indicates poor efficiency since the secondary structure could make the target site less accessible. UUUU would trigger the termination of RNA polymerase III, which transcribes gRNA. The nucleotides next to the PAM exert the greatest impact on Cas9 editing activity. On position 20, a strong preference for G has been observed while C is the last choice[31,32,34]. As to the overall nucleotide usage, conclusions drawn in different studies did not reach a consensus. Some studies showed that high adenine (A) content correlated with high Cas9 loading efficiency[35]. On the contrary, a study in zebrafish indicated that G enrichment and A depletion were linked to high gRNA activity[36]. Factors related to RNA structure such as GC content are also important. gRNA sequences with GC content of 40%-60% are generally considered optimal[32,35]. Notably, the preference of GC content varies across different positions of gRNA[37].

Furthermore, the chromatin state of the targeted site contributes to the CRISPR/Cas9 editing efficiency. Evidence has emerged that the editing can be impeded in heterochromatin[38,39]. This may be explained by nucleosome occupancy. An *In vitro* experiment suggested that nucleosomes could block Cas9 loading onto DNA. However, nucleosome positioning is dynamic in cells. Indeed, Cas9 binding is more resistant to nucleosomes in cells[40].

1.1.2.3 Specificity

Despite the wide use of CRISPR/Cas9, some groups reported that off-target editing occurred at a comparable frequency as the on-target one[41,42]. Generally, the gRNA sequence is considered the major determinant of specificity. Mismatches at the 5' ends often appear to be more tolerant than ones located close to 3' ends[42,43]. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) profiling dCas9 indicated that 10 or even 5 nt adjacent to PAM sequences (3' end of gRNA) determined CRISPR/Cas9 specificity, known as "seed sequence"[44,45]. However, mismatches at the 5' end are not always tolerant. Some mismatches distal to PAM sequences decrease the specificity while the PAM-proximal mismatches exhibit few effects[42].

To assess the CRISPR/Cas9 specificity, many experimental approaches were invented such as GUIDE-seq[46], HTGTS[47], Digenome-seq[48] and BLISS[49]. These methods identified off-targeting sites with up to six mismatches. Besides the region-specific tolerance of mismatch, analysis using GUIDE-seq data suggested that rU–dG or rG–dT were more tolerated[50].

To increase the specificity of the CRISPR/Cas9 system, many strategies have been proposed. Truncated gRNAs (tru-gRNAs) have been shown to effectively reduce off-target editing in two different ways. Although it seems counterintuitive, tru-gRNAs which are shortened by 2-3 nt at 5'end decrease the off-target sites by 5,000-fold while maintaining on-target genome editing efficiencies when compared to full-length gRNAs[41] (**Fig. 1A**). On the other hand, when tru-gRNAs are shorter than full-length gRNAs by at least 4 nucleotides, Cas9 is still able to bind to DNA, but no DSBs can be induced. Based on this observation, dead-RNAs (dRNAs) or tru-gRNAs are used to compete with the active gRNA without inducing any DSB. By co-delivering dRNAs to the cell, off-target sites are shielded from the active gRNA without compromising on-target efficiency[51] (**Fig. 1B**). Besides modifying gRNAs, another strategy is to use paired Cas9 nickases (Cas9n). It carries a mutation in the RuvC or HNH domain. Two gRNAs direct

paired nickases to the neighbouring DNA sequences to create offset nicks[52]. The Cas9n system reduces off-target editing by up to 1,500-fold in human cells by requiring two Cas9 binding sites to be in close proximity to generate InDels[52] (**Fig. 1C**).

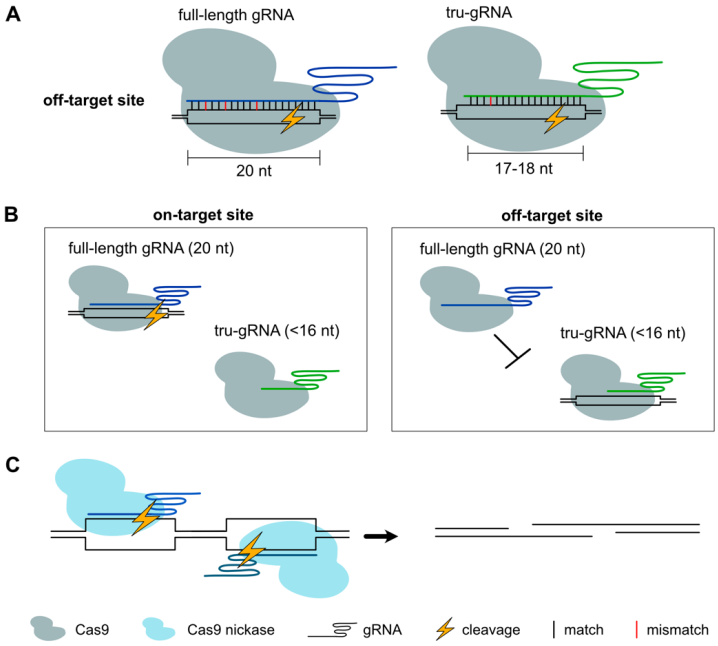

Fig 1. Schematic of various strategies to increase CRISPR/Cas9 specificity
(**A**) Cas9 guided by a full-length gRNA tolerate (dark blue) more mismatches than by a tru-gRNA (green, 2~3 nt shorter than full-length gRNA). (**B**) A tru-gRNA (green) with at least 4 nt shorter guides Cas9 to an off-target site without inducing a DSB, impeding the binding of Cas9 guided by a full-length gRNA (dark blue). Meanwhile, a DSB can be achieved at the on-target site by a Cas9-full-length gRNA complex. (**C**) Cas9 nickases reduce CRISPR/Cas9 off-target editing by requiring two bindings next to each other.

1.1.3 CRISPR/Cas9 delivery system

Continuous efforts have been made to optimize the delivery of CRISPR/Cas9 components into cells. Delivery methods can be generally divided into viral and non-viral approaches, and each of them has its pros and cons.

Adeno-associated viruses (AAVs) are commonly used for gene delivery because of their replicative deficiency and inability to integrate into the genome, which is beneficial in decreasing off-target events and cellular toxicity[52]. It can transport Cas9, gRNAs or homologous donor templates into cells by transduction. However, AAVs are limited by low on-target editing efficiency and low cloning capacity (<4.7 kb) considering the size of the spCas9 gene (~4.1 kb)[53]. Another viral vector is the lentivirus (LV). Unlike AAVs, LVs have a much higher cloning capacity (<8 kb), allowing for the cloning of both Cas9 and gRNA into the same vector. In addition, the transduction of LVs is very efficient[54]. However, the random integration of LVs into the genome leads to unwanted effects such as oncogene activation[55]. This impedes the application of LVs in clinical trials. On the contrary, adenoviruses (AVs) have been broadly used as vectors in clinical settings. AVs do not integrate into the genome, but they are known to trigger high-level innate immune responses[56]. Virus-based delivery is generally more efficient than

non-virus methods, but the safety needs to be thoroughly assessed, and laboratories with high biosafety levels and ethical approval are required.

Non-virus approaches, for example, electroporation and lipid-based nanoparticles (LNPs), are used to deliver either plasmids or Cas9-gRNA ribonucleoproteins (RNP) directly into the cells[57]. Electroporation utilizes pulses of electrical current to increase the permeability of cell membranes temporarily and stimulate the transient opening of pores, enabling CRISPR/Cas9 components to be delivered. Electroporation is efficient, especially for hard-to-transfect cell types, for example, primary cells[58]. The disadvantage is that it requires different stimulation conditions for different cell types. The strong current pulse usually results in low cell viability. Moreover, liposomes consist of lipid bilayers with spherical structures. Cargoes are encapsulated and form a positively changed complex with LNPs, which greatly facilitates the fusion through the negatively charged cell membrane[57]. Among LNPs, lipofectamine is the most popular reagent. It is easy to use without exerting excessive cell stress.

1.1.4 HAP1, a widely-used cell line in CRISPR/Cas9 experiments

Although CRISPR/Cas9 system was first found in prokaryotic organisms, it has been widely applied in mammalian cells as a gene-editing tool. Most mammalian cells are diploid, inheriting one copy from each parent. Additionally, many cell models used by researchers are cancer cells, which are likely to be aneuploid. Loss of function study using genome editing in diploid or polyploid cells is very challenging. Mutations achieved in one copy are usually buffered by the unmodified copy(s) located on the other allele(s). Therefore, the phenotypic consequences of the mutation are often masked, hindering the identification of the gene function[59,60]. Furthermore, in non-haploid cells, different scenarios of on-target alterations can occur on multiple alleles with diverse biological consequences[61]. Therefore, because of its near-haploid nature, HAP1 is one of the most popular cell lines used in CRISPR/Cas9-mediated single-target editing and large-scale screening.

The HAP1 cell line was created from KBM-7 cells, which are near-haploid cells derived from a patient with chronic myeloid leukaemia, carrying two copies of Chromosome (Chr) 8 and a fragment of Chr 15[62,63]. An attempt was made to generate induced pluripotent stem cells from KBM-7 cells through the transfection of four Yamanaka factors (KLF4, c-MYC, OCT4 and SOX2). This resulted in the emergence of HAP1 cells that retain their haploid nature except for two copies of a fragment of Chr 15[64].

It has been reported that HAP1 has a strong tendency to change from a highly unstable haploid status to diploid cells under exposure to stress or long-time culture[65,66]. Nevertheless, it is still beneficial to use HAP1 cells. Duplication of the genome after creating mutations at targets results in modifications at both alleles, enabling the measurement of phenotypic changes.

1.2 DNA damage repair

After a DSB is induced by Cas9, the repair of the break occurs using intrinsic cellular mechanisms. The choice of pathways depends on many factors such as cell type, cell cycle phase and chromatin states.

1.2.1 DNA damage response

DNA damage response (DDR) is a set of cellular processes activated upon damage to DNA molecules, including the recognition of damage, activation of certain signalling pathways, and recruitment of repair proteins. DDR may also trigger programmed cell death if the DNA damage is too severe to repair.

Kinases such as ataxia telangiectasia mutated (ATM) and RAD3-related (ATR) are initially activated and act as major regulators in DDR-signalling pathways. ATM is recruited to damage sites by meiotic recombination 11 (MRE11) – double-strand break repair protein (RAD50) - Nijmegen breakage syndrome 1 (NBS1) complex (MRN complex) and undergoes autophosphorylation to be active[67]. ATR is recruited by replication protein A (RPA) through ATR-interacting protein (ATRIP). ATM/ATR further phosphorylates and activates cell-cycle checkpoint kinases such as CHK1 and CHK2[68,69]. DNA-dependent protein kinase (DNA-PK), consisting of a catalytic subunit DNA-PKcs and a regulatory subunit Ku70-80, is also among the kinases first activated upon DSBs. ATM/ATR/DNA-PK can phosphorylate the serine-139 of the histone H2A variant (γH2AX), flanking damage sites with the size of megabases[70] and serving as seeds for the formation of DDR foci[71]. Next to γH2AX, a large number of proteins are found such as ATM, mediator of DNA damage checkpoint protein 1 (MDC1) and MRN complex[71]. TP53-binding protein 1 (53BP1) and breast cancer type 1 susceptibility protein (BRCA1) are later recruited. The DDR foci are rapidly assembled and established a 'repair-prone' environment. The spread of γH2AX is constrained within topologically associating domains, and γH2AX–53BP1 chromatin domains are correctly established under the guidance of one-sided cohesin-mediated loop extrusion around DSB sites[72].

1.2.2 Non-homologous end joining

The non-homologous end joining (NHEJ) pathway directly re-ligates broken DNA ends with minimal DNA end processing. It starts with the binding of a ring-shaped Ku70/Ku80 protein heterodimer to shield DNA ends from further resection[73] (**Fig. 2A**). Next, DNA-PKcs are bound and recruit other NHEJ factors including DNA ligase IV (LigIV), X-ray cross complementing group 4 (XRCC4), XRCC4-like factor (XLF) and end-processing enzymes[74,75]. Active DNA-PKcs can phosphorylate and activate enzymes recruited[76]. Meanwhile, the autophosphorylation of DNA-PKcs results in conformational changes and the dissociation of DNA ends[77]. These steps enable the access of end-processing enzymes to DNA ends[78,79]. The removal of mismatched nucleotides and re-synthesis require enzymes including polynucleotide kinase 3' phosphatase, Artemis nucleases and DNA polymerases, such as Polμ and Polλ. After end processing, LigIV-XRCC4-XLF ligates DNA ends[80–82].

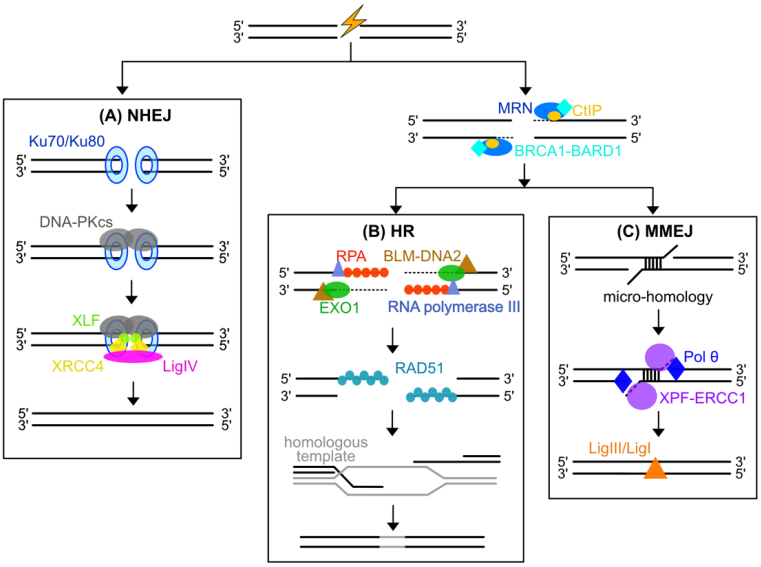

Fig 2. DSB repair pathways
After a DSB is induced, the damage is usually repaired by NHEJ (**A**), HR (**B**) and MMEJ (**C**). The names of proteins involved in each step are labelled and coloured corresponding to symbols representing proteins in the illustration.

1.2.3 Homologous recombination

In the homologous recombination (HR) pathway, DNA damage is repaired using the sister chromatid as the homologous template (**Fig. 2B**). HR only occurs in S and G2 phases while NHEJ can operate throughout cell cycle.

During end resection, two steps are required in HR: short-range and long-range end resections. In the first step, MRN- C-terminal binding protein interacting protein (CtIP) binds at the break site. Both phosphorylated CtIP and BRCA1– BRCA1-associated domain 1 (BRCA1-BARD1) facilitate the short-end resection[83,84]. MRE11 in the MRN complex creates nicks and further extends it to the break site. These short overhangs provide access for endo/exonucleases such as exonuclease 1 (EXO1) and Bloom syndrome protein (BLM)/DNA2[84–87]. During the long-range end resection, the nucleotides are removed from 5' to 3' by EXO1, resulting in long 3' single-stranded DNA (ssDNA) overhangs[86]. The activity of EXO1 is regulated by the MRN complex and CtIP[86]. BLM and DNA2 work together to generate long overhangs to enable the 5' to 3' polarity of resection. BLM serves as a helicase to unwind DNA strands, and DNA2 induces cleavage. CtIP can also interact and stimulate the activity of BLM as helicase and DNA2 as 5' endonuclease[86].

These overhangs are vulnerable to degradation and need protection. The binding of the PRA complex stabilizes the ssDNA[88]. Besides, DNA-dependent RNA polymerase III (Pol III) was shown to give rise to RNA using the ssDNA as a template. As consequence, RNA-DNA hybrids serve as essential intermediates to protect the overhangs and promote end resection[89].

The recombinase RAD51 then removes the RPA complex and likely the RNA-DNA hybrids too, forming RAD51 nucleoprotein filament, which is facilitated by the BRCA2-DSS1 complex and BRCA1-BARD1[90,91]. Afterwards, RAD51 mediates strand invasion through the pairing of homologous sequences, thereby forming a D-loop. During this process, many additional accessory factors are involved such as RAD54 and BRCA1-BARD1[92,93].

The 3' end of the invading strand is extended by DNA polymerase δ, together with proliferating cell nuclear antigen (PCNA) and replication factor C (RFC) subunit 1-5, using the homologous DNA as a template[94–96]. The product can be resolved by either double Holliday junction (dHJ) or the non-crossover synthesis-dependent DNA strand annealing (SDSA), which is applicable in genome editing. When the invasion from both ends occurs on the same template, the DNA can be ligated by two Holliday junctions[95]. Alternatively, in SDSA, the heteroduplex DNA strands are separated by helicases such as BLM so they can be annealed with the other end of the break followed by gap filling and ligation to finish the repair[95].

1.2.4 Microhomology-mediated end joining

Besides NHEJ and HR, microhomology-mediated end joining (MMEJ) can be used to repair CRISPR/Cas9-induced DSBs (**Fig. 2C**). Similar to NHEJ, MMEJ does not require a donor template to repair the damage. During short-range resection of the DSB, ends are realigned using micro-homologous sequences with a length of 5–25 bp near the DSB site[97]. Since the remaining ssDNA at the 3' end is removed, the MMEJ pathway usually leads to deletion with variable sizes[97].

The short-range resection is initiated by the MRN complex and CtIP[85,98]. During the S/G2 phase, CtIP is phosphorylated, which stimulates the endonuclease activity of MRE11. As a consequence, MRE11 generates a nick near the DSB site, impeding the repair by NHEJ[84,85]. Moreover, this nick triggers the exonuclease activity of MRE11 and generates short overhangs, allowing the realignment using micro-homology[99,100].

The mechanism of the annealing process has not been fully understood. Poly ADP-ribose polymerase 1 (PARP1) was suggested to promote the annealing reaction[101–103]. However, conflicting evidence was reported that PARP1 promoted Ku loading, which favoured the choice of NHEJ instead of MMEJ[104]. After annealing, if the micro-homology is distal from DNA ends, XPF-ERCC1 endonuclease removes the ssDNA[105]. DNA polymerase θ (Pol θ) can synthesize under the direction of a template to fill the gap. Once there is no ssDNA, DNA ends are ligated by DNA ligase III (LigIII) or DNA ligase I (LigI)[106].

1.2.5 DNA damage repair after CRISPR/Cas9-mediated DSBs

The NHEJ is dominant and error-prone. Since the InDels introduced by NHEJ are not accurate repairs, many strategies are designed to inhibit NHEJ and favour HR to achieve precise editing. Although MMEJ is not used as frequently as NHEJ or HR, it is still common that CRISPR/Cas-induced DSBs are repaired through MMEJ, even with a frequency of up to ~58% in some cases[107]. Taking advantage of predictable outcomes of MMEJ on given DNA sequences, template-free and precise CRISPR/Cas9 genome editing has been developed[107,108].

The repair time for CRISPR/Cas9-induced DSBs varies, but on average it is a lengthy process with half-life times of up to 10 hours[109]. Persistent Cas9 binding on a DNA end blocks the binding of DNA repair

enzymes. At actively transcribed gene loci, RNA polymerase removes Cas9 from a DSB site in a highly strand-biased manner. However, this behaviour results in increased mutagenesis frequencies[110].

1.3 Challenges in tackling CRISPR/Cas9 editing outcomes

Despite its popularity, the CRISPR/Cas9 system has been associated with a number of unintended outcomes, raising critical safety concerns in both research and clinical application. Many methods were developed to experimentally test the CRISPR/Cas9 experiment design in specific cell models.

1.3.1 Undesired CRISPR/Cas9 editing outcomes

Besides the imperfect specificity, another substantial barrier to CRISPR/Cas9 application is the toxicity and its adverse effects on mutation selection. The initial observation revealed that CRISPR/Cas9 system triggered cell cycle arrest and cell death in human retinal pigment epithelial-1 cells and human pluripotent stem cells. The underlying mechanism was that DSBs induced by CRISPR/Cas9 activated DDR and p53 pathway[111,112]. In other words, CRISPR/Cas9 workflow may lead to the emergence and enrichment of cells with p53-inactivating mutations. Indeed, intensive genetics characterization performed in hundreds of cell lines by independent groups confirmed the selection of *TP53* mutation in CRISPR/Cas9-mediated gene-editing experiments[113–115] (**Fig. 3A**). Cells with mutations located in *KRAS* or some tumour suppressor genes such as *CDKN1A* were enriched as well. Detailed analysis of the selection pattern uncovered a p53-linked interactome, consisting of genes with nonredundant roles in DDR[113]. Additionally, compared with targeting heterochromatin, p53 pathway-associated toxicity is higher when the target resides in euchromatin, where expressed genes and functional regulatory elements are located[116]. These studies urge the necessity of thoroughly assessing and monitoring CRISPR/Cas9-modified cells to avoid incidental selection of DNA repair-deficient cells.

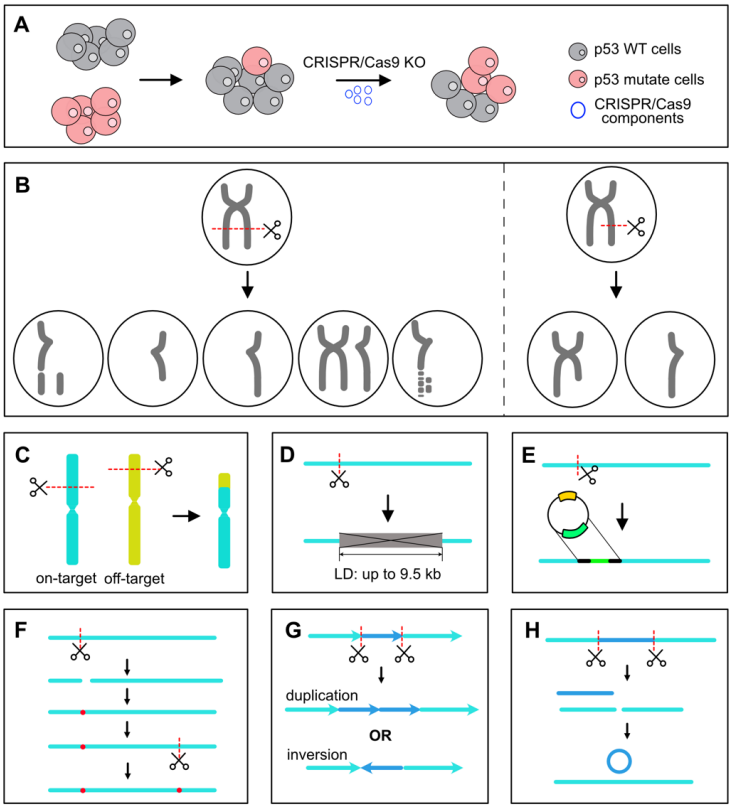

Fig 3. Undesired genomic aberrations induced by CRISPR/Cas9
(**A**) Cells with *p53* mutations (pink) are enriched after CRISPR/Cas9 knockout (KO). (**B**) Aneuploidy, chromothripsis and allele-specific chromosome loss at the targeted chromosome (grey) in CRISPR/Cas9-edited cells. (**C-E**) Translocation between the targeted chromosome (turquoise) and off-target chromosome (lime) (**C**), large deletion (grey box, D) and insertion (**E**) from a plasmid (circle with green and yellow pieces) are reported at the targeted site. In dual gRNA system (**F-G**), unsynchronized cuts (**F**), as well as duplication (**G**), inversion (**G**), or circularization (**H**) of the target region could occur in CRISPR/Cas9-modified cells. CRISPR/Cas-induced cleavage is indicated with scissors and a dashed line in red.

In addition to the mutation that emerged at non-targeting sites, various undesired genomic alterations linked to the on-target DSB have been reported.

Recent work has shown frequent aneuploidy in primary human T cells after being modified by CRISPR/Cas9[117] (**Fig. 3B**). Both the gain and loss of the targeted chromosome were found among the edited cells. Although the aneuploidy activated the p53 pathway and led to cell death, T cells with aneuploidy were detectable 11 days post-transfection. Notably, it is not the first time that chromosome-scale abnormalities after CRISPR/Cas9-mediated editing were reported. An earlier study in human embryos described an allele-specific chromosome loss[118] (**Fig. 3B**). The gRNA was designed to target the paternal chromosome. After mitosis, loss of chromosomal arms on the paternal allele occurred if DSBs were not repaired in time. Besides the targeted chromosome, chromosomal loss at off-target sites was also found. Later, a study in human cell lines confirmed the observation of chromosomal loss and characterized on-target chromothripsis[119] (**Fig. 3B**). Through tracking mother

cells where DSBs were induced, together with their daughter and granddaughter cells, researchers rebuilt the time-serial events after CRISPR/Cas9-induced cleavage including the formation of micronuclei and chromosome bridges. Eventually, extensive fragmentation and rearrangements, known as chromothripsis, as well as chromosomal loss occurred at the targeted arm in granddaughter cells. Another chromosome structure abnormality is translocation, which refers to the phenomenon that a genomic sequence from a different chromosome or 500 kb away from the on-target DSB site is attached to one end of the on-target breakage[120] (**Fig. 3C**). Translocation usually happens at low frequency among CRISPR/Cas9-modified cells, varying from 0.1%-1% in the single gRNA system[121,122]. Low CRISPR/Cas9 editing specificity increases translocation frequency by introducing off-target DSBs. Additionally, recent work suggested that naturally occurring break sites independent of gRNA sequence induced translocation as well[123].

Besides these large-scale chromosomal aberrations, many other structural variations (SVs) are found at the on-target sites.

NHEJ is generally the most frequently used pathway to repair, especially with the absence of a donor DNA template, so small InDels (<20 bp) are usually expected at DSB sites[124]. However, large deletion (LD) has been observed in the organisms such as zebrafish and mice, as well as human cells when using single gRNA editing systems to target different loci[120,125–127] (**Fig. 3D**). The sizes of LDs range from a few hundred base pairs to 9.5 kb with various frequencies. Due to the design to validate focal mutations in the vicinity of targets, conventional genotyping methods often fail to detect LDs, leading to the frequent oversight of such events. Long-range PCR and long-read sequencing (LRS) are required to capture LDs. The analysis of the sequence content showed that microhomologies are usually enriched at breakage sites of LDs, indicating that MMEJ could be the pathway used for the repair instead of NHEJ[128].

Large insertions (> 20 bp) can occur at on-target sites, with the most commonly inserted sequences derived from plasmids (**Fig. 3E**). While some components may exhibit higher integration frequency, any sequence present in the plasmid has the potential to be inserted into the genome of cells[120,129]. As a result, peptides generated using the integrated sequences as templates may be produced and display cytotoxicity.

These SVs including LDs and insertions that occur in the F0 generation can be inherited and found in the F1 generation[127]. This finding in zebrafish reinforces the necessity of pre-testing editing outcomes in clinical applications.

To excise a genomic region, dual gRNA systems are used. It is expected that the unintended outcomes observed in a single gRNA system are found in cells edited with dual gRNAs. Additional on-target genomic alterations have been reported in dual gRNA systems.

To achieve a successful cleavage, both DSBs must be induced before any of the breaks are repaired. Otherwise, an unsynchronized fashion of either induction of DSBs or DNA damage repair fails the target deletion using a dual gRNA system[130] (**Fig. 3F**).

After the removal of the target region and the repair of DSBs, a single event of either a duplicated or an inverted target region has been reported in many model systems (**Fig. 3G**). For example, an

inversion occurred with a frequency between 0.71 to 23.28% in HEK293T cells and 0 to 6.9% in mice while a duplication of a deleted region occurred with a frequency between 0.17 to 5.97% in HEK293T cells and 0 to 28.1% in mice[131,132]. Duplication and inversion are both considered the consequences of the MMEJ repair pathway[133]. Disrupting CtIP or FANCD2 was shown to reduce the likelihood of duplication or inversion and thus enhanced the desired deletion. Besides, the cleaved DNA fragment can get circularized and form extrachromosomal circular DNA (eccDNA)[134] (**Fig. 3H**). eccDNAs are mainly induced by NHEJ with sizes from a few hundred base pairs to 200 kb. In some cases, a megabase ring chromosome forms where dual DSBs occur.

Given that the CRISPR-Cas9 system constitutes a prokaryotic self-defence mechanism, eukaryotic cells have not evolved corresponding counteracting mechanisms to prevent its potential deleterious effects. Therefore, due to the complexity of genomic outcomes in CRISPR/Cas9 application, a thorough evaluation covering a broad spectrum of considerations must be carried out so genome editing can be applied in a controllable and safe manner.

1.3.2 Experimental tools to evaluate CRISPR/Cas9 editing outcomes

Enormous efforts are made to develop efficient and sensitive tools, enabling comprehensive assessments of CRISPR/Cas9 editing performances. Although the computational algorithms are user-friendly and effortless, considerable amounts of off-target sites and on-target mutagenesis are missed due to limited knowledge of Cas9 behavioural patterns and DNA damage repair outcomes. As a result, many experimental methods have been developed to identify off-target editing and unintended genomic contents at on-target sites.

1.3.2.1 Cell-free methods

Extracted genomic DNA is directly incubated with CRISPR/Cas9 components in cell-free methods so these assays generally have less background and do not require high-transfection efficiency. *In vitro* methods including digested genome sequencing (Digenome-seq) (**Fig. 4A**), circularization for *in vitro* reporting of cleavage effects by sequencing (CIRCLE-seq) (**Fig. 4B**), and selective enrichment and identification of tagged genomic DNA ends by sequencing (SITE-seq) (**Fig. 4C**) allow in-depth characterization of CRISPR/Cas9-induced cleavages. However, due to a lack of cellular mechanisms, cell-free methods are unable to test editing consequences led by DNA damage repairs such as translocation and LD.

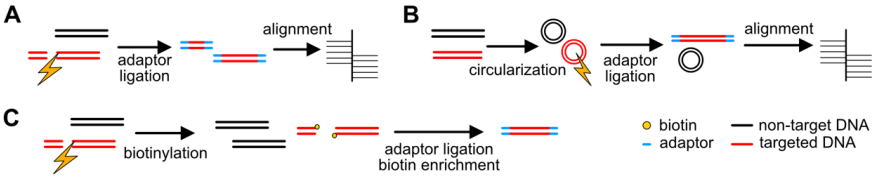

Fig 4. Cell-free methods to detect CRISPR/Cas cleavages
The overview of cell-free methods: Digenome-seq (**A**), CIRCLE-seq (**B**) and SITE-seq (**C**). In these methods, genomic DNA is first extracted and then incubated with Cas9 and gRNA *in vitro*.

Digenome-seq/DIG-seq The experimental part of Digenome-seq is very straightforward[48] (**Fig. 4A**). Genomic DNA is first digested by Cas9 with the presence of gRNA and then prepared for whole-

genome sequencing on Illumina platform. It is highly sensitive and able to detect off-target mutations with a frequency as low as 0.1%. Conversely, Digenome-seq lacks an enrichment step, so it requires high sequencing coverage with around 500 million reads and advanced bioinformatic skills.

Given that chromatin accessibility impacts Cas9 loading, DIG-seq (Digenome-seq using cell-free chromatin DNA) was developed. In DIG-seq, native chromatin is used as input. DIG-seq shortens the list of potential off-target sites compared with other *in vitro* assays.

CIRCLE-seq CIRCLE-seq utilizes DNA circles to exclude DNA molecules lacking nuclease-induced DSBs[135] (**Fig. 4B**). Genomic DNA is first sheared and then self-circularized. During Cas9 treatment, circular DNA with cleavage is linearized, enabling the adaptor ligation and library preparation. By introducing the circularization step prior to Cas9 treatment, CIRCLE-seq only captures DNA molecules targeted by Cas9, which greatly increases the signal-to-background ratio and reduces the demanded sequence depth (~5 million reads).

SITE-seq Compared with Digenome-seq, SITE-seq selectively tags DNA that is cut by Cas9 with biotin[136] (**Fig. 4C**). After DNA fragmentation, subsequent streptavidin pull-down enables efficient profiling with shallow sequencing (>3 million reads).

1.3.2.2 Cell-based methods

Cell-free methods are unable to investigate the genomic consequences led by many other key factors including DNA damage repair mechanisms and dynamics of chromatin environments. The demand for exhaustive evaluation results in the continuous development of cell-based methods.

BLESS/BLISS Breaks labelling, enrichment on streptavidin, and sequencing (BLESS) directly captures genome-wide DSBs by the ligation of biotinylated adaptors and pull-down with streptavidin-coated beads[137] (**Fig. 5A**). A refined version, breaks labelling *in situ* and sequencing (BLISS), has been made for low input[49] (**Fig. 5A**). DSB ends are ligated with double-stranded oligodeoxynucleotides (dsODNs) adapters. *In vitro* transcription mediated by the T7 promoters within adapters further increases the efficiency of capturing DSB sites. Both methods are independent of the cellular repair machinery. On one hand, these methods enable mapping at single-nucleotide resolution. On the other hand, these two methods fail to capture repaired CRISPR/Cas9-mediated cleavages and all the genomic alterations led by DNA damage repair.

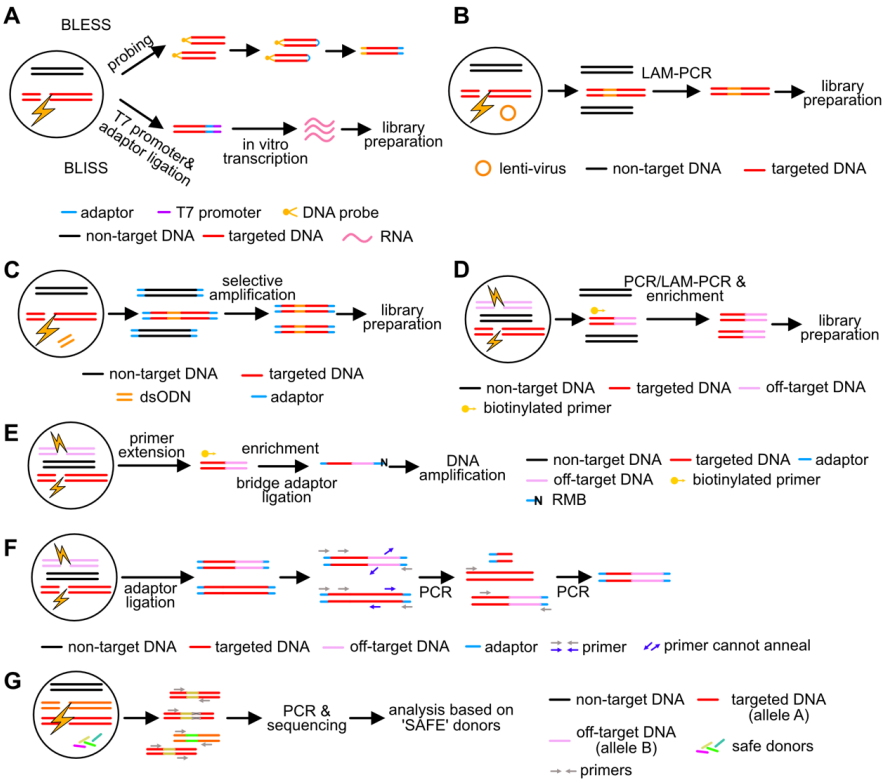

Fig 5. Cell-based methods to assess CRISPR/Cas9 editing outcomes
Cell-based assays enrich genomic regions with DSBs through various approaches: BLESS/BLISS (**A**), IDLV capture (**B**), GUIDE-seq (**C**), HTGTS/LAM-HTGTS (**D**), PEM-seq (**E**), CAST-seq (**F**) and SAFE donor (**G**). These methods take epigenetic state and chromatin architectures into consideration, which impact both Cas9 loading and DSB repair.

IDLV capture Integrase deficient lentiviral vector (IDLV) capture identifies CRISPR/Cas9-mediated DSBs by detecting integrated IDLV sites[138] (**Fig. 5B**). The IDLV-tagged molecules are enriched by linear amplification-mediated PCR (LAM-PCR). IDLV capture is applicable to a wide range of nuclease tools including Cas9, ZFNs and TALENs. However, this method is restricted by low sensitivity and a high false positive rate due to the random insertion of IDLV.

GUIDE-seq Genome-wide, unbiased identification of DSBs evaluated by sequencing (GUIDE-seq) employs the NHEJ-mediated insertion of supplied and blunt dsODNs with defined sequences to capture CRISPR/Cas9 editing sites[46] (**Fig. 5C**). The selective amplification is achieved through the primers annealing to dsODNs and single-tailed adapters, referred to as Single-Tail Adapter/Tag-PCR method. It is considered to have better performance than LAM-PCR in reducing bias and background. GUIDE-seq can detect the editing sites with frequencies of 0.1%. The bottleneck of GUIDE-seq is the efficiency of dsODNs integration, which can be determined by the transfection efficiency, choice of NHEJ and cytotoxicity of dsODNs.

HTGTS/LAM-HTGTS High-throughput, genome-wide translocation sequencing (HTGTS) is designed to detect chromosomal reengagements, especially translocation between two DSBs[47] (**Fig. 5D**). The DSB

end of the on-target site is “bait”, and the off-targeting DSB end is “prey”. DNA molecules with baits are amplified with biotinylated primers and subsequently enriched. HTGTS requires a large amount of input material and deep sequencing. Some modifications such as LAM-PCR and enzyme blocking were introduced in an advanced version, named LAM-HTGTS, to improve the efficiency and reduce the time[139] (**Fig. 5D**).

PEM-seq Because a high cycle number of amplifications is used in LAM-HTGTS, it is difficult to distinguish the original template from PCR duplicates. To quantify translocation events, primer-extension-mediated sequencing (PEM-seq) has adopted primer extension and random molecular barcode (RMB) to generate one copy and label each fragment, respectively[140] (**Fig. 5E**).

CAST-seq To meet the demand of clinical use, chromosomal aberrations analysis by single targeted linker-mediated PCR sequencing (CAST-Seq) was established[123] (**Fig. 5F**). It consists of fragmentation, linker ligation and three PCR steps. Compared with LAM-HTGTS, it has potential as a preclinical test since it has been tested in human primary cells. Additionally, instead of using restriction enzyme blocking, CAST-seq employed decoy primer to reduce background more efficiently, and it can detect aberrations with frequencies as low as 0.006%. However, CAST-seq requires much more input material, which could be a hurdle in clinical application.

SAFE donor An approach using ‘sequence-ascertained favorable editing’ (SAFE) donor has been recently published to detect unintended genomic alterations such as aneuploidy and loss of heterozygosity[141] (**Fig. 5G**). This method relies on the insertion of supplied DNA donor mixtures with the desired nucleotide substitutions. Its data analysis pipeline assumes that different donor sequences are used during the repair occurring on different alleles. As a newly developed method, its performance remains to be tested. For example, how to enhance HR to ensure the efficient knock-in of the SAFE donor? Is there any modification required in polypoid cells? Despite these concerns, the SAFE donor approach broadens the detection capabilities of unintended outcomes.

The choice of method is highly dependent on which unintended outcomes were aimed to examine, the required sensitivity and characteristics of a chosen cell model such as transfection efficiency or the preference towards certain repair pathways. However, a benchmark is still possible by comparing the cleavage sites identified by each method. Comparative assessments showed that in a given CRISPR/Cas9 system (e.g. same Cas9 variant and gRNA design), many cleavage sites could be validated by whole-genome sequencing and were reported by all the methods tested[142]. Meanwhile, distinct off-target sites detected by a specific method but not the others were also found. The results suggest that none of the methods is comprehensive and can be used to set a gold standard. Generally, *in vitro* methods produce longer lists of off-target sites than *in cellulo* methods. For example, when targeting the *VEGFA* gene, GUIDE-seq identified 22 off-target sites while Digenome-seq and SITE-seq detected 69 and 996 off-target cleavages, respectively. 91% of the off-target sites discovered by GUIDE-seq were validated but only 53% and 10% of candidates reported by Digenome-seq and SITE-seq were confirmed using deep whole-genome sequencing.

1.3.2.3 Long-read sequencing-based methods

All the methods described above depend on PCR amplification and a short-read sequencing platform (Illumina). These steps inevitably introduce inherent limitations when amplifying regions with low

complexity such as repetitive or AT/GC-rich sequences and capturing genomic contents with considerable length. The emergence of LRS platforms including Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) shed light on the solutions. LRS technologies have been proven to be optimal choices when resolving SVs and repeats.

LRS In 2018, Kosicki *et al* revealed LDs around kilobases, inversion and *de novo* insertion in CRISPR/Cas9-modified cells using LRS, providing direct evidence of unintended genomic alterations based on kilobase-long reads[125].

SMRT-OTS/ Nano-OTS To detect difficult-to-capture off-target editing sites located in dark genomic regions, *in vitro* assays SMRT-OTS and Nano-OTS were established[143] (**Fig. 6A**). SMRT-OTS is based on PacBio's SMRT sequencing while Nano-OTS is for ONT nanopore platform. In both methods, the enrichment of CRISPR/Cas9 targeted molecules is dependent on the cleavage which allows the ligation of adapters. These methods revealed some validated off-target sites that could not be resolved easily with short-read sequencing.

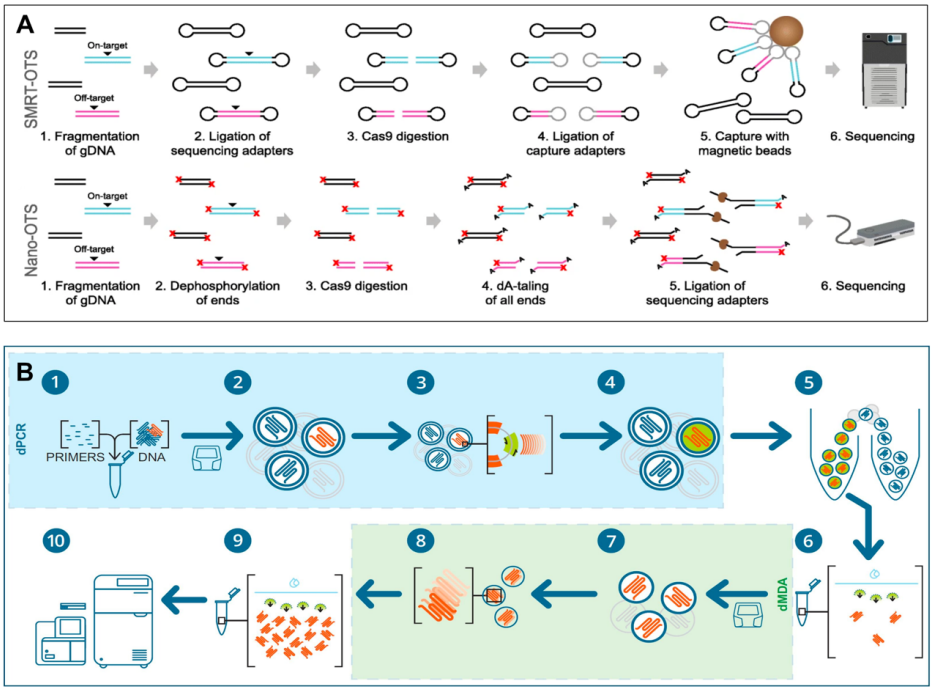

Fig 6. LRS-based methods to investigate CRISPR/Cas9 consequences
(**A**) SMRT-OTS and Nano-OTS were developed to detect CRISPR/Cas9 off-target activity for PacBio Sequel system and ONT MinION system, respectively. In SMRT-OTS, capture adaptors (grey) are ligated to the molecules cleaved by Cas9 and allow the enrichment using magnetic beads (brown). In Nano-OTS, the input fragmented DNA is dephosphorylated (red cross). Therefore, only CRISPR/Cas9 targeted molecules can be added with dA-tails and sequencing adaptors. (**B**) Xdrop workflow includes targeted DNA molecular enrichment in double emulsion droplets (blue box, "dPCR") and multiple displacement amplification (green box, "dMDA"). The pictures of workflows are adapted from Höijer *et al*[143] (**A**) and Blondal *et al*[129] (**B**) with minor changes under Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).

Xdrop Xdrop is a microfluidic target-enrichment approach compatible with an LRS platform[129,144] (**Fig. 6B**). Primers are designed to generate a small amplicon within the region of interest during droplet-PCR (dPCR). Through amplification signals, DNA molecules with the targeted region are enriched. The resulting DNA is made into library using droplet multiple displacement amplification (dMDA) and sequenced on an LRS platform. This method has a broad range of applications including identifying unwanted insertion in CRISPR/Cas9-modified cells[129]. Since the Xdrop approach requires a known region to enrich for, it cannot detect unknown off-target sites. Meanwhile, its enormous potential can be foreseen in revealing other unintended on-target genomic alterations.

1.4 Genome instability

1.4.1 Replication defects and genome instability

There are many cellular mechanisms to ensure the stability of the genome such as DNA damage repair and cell cycle checkpoints. Therefore, genetic variation usually occurs at low frequency. However, genotoxicity induced by external or internal sources can greatly increase the occurrence of genome instability, resulting in various genomic alterations: (I) mutation; (II) chromosome instability caused by frequent chromosome missegregation; (III) SVs including deletion, insertion, inversion, duplication and translocation. Although genome instability can be triggered by the dysfunction of many processes, defects during DNA replication are the leading causes.

In the G1 phase, minichromosome maintenance (MCM) 2-7 complex is recruited to a replication origin and serves as the catalytic subunit of replicative DNA helicase, mediated by the origin recognition complex, cell division cycle 6 (CDC6) ATPase and chromatin licensing and DNA replication factor 1 (CDT1)[145]. The loading process of an inactive MCM2-7 complex, referred to as origin licensing, ensures the accurate duplication of the genome. Only licensed origin can initiate DNA replication forks[145]. Inefficient origin licensing may lead to unfinished replication products and eventually result in genome instability. MCM2-7 depletion in human cells causes hypersensitivity to DNA replication stress and accumulation of DNA lesions[146]. Additionally, a study using mouse embryonic fibroblasts with mutated MCM4 showed that cells with reduced MCM2-7 stability and licensed origins were prone to chromosome breaks under exposure to a replicative polymerase inhibitor[147].

As cells enter the S phase, replication licensing systems are inactivated by the cyclin-dependent kinase (CDK), which, together with the DBF4-dependent kinase (DDK), also facilitates the activation of helicase and the assembly of replisome[148]. Excessive initiation of DNA replication usually contributes to oncogenic transformation and genomic instability[149–151]. For example, the overexpression of c-MYC induces replication stress and DNA damage by interacting with MCM2-7 and increases unscheduled origin firing[152]. Interestingly, unlike the noncancerous cell under replication stress, cancer cells prefer to reactivate the early-replicating origins instead of employing unutilized sites[153]. Over-activation of replication initiation may increase the probability of replication fork collisions, resulting in breakages.

Once replication is initiated, forks may pause, termed as fork stalling. Transient fork stalling is not destructive as the replisome remains associated, and the replication can be restarted once the obstacles are removed. However, persistent fork stalling can lead to DSBs or fork regression which is

able to create a Holliday junction[154,155]. Additionally, unlike the transient pause, the replisome is likely to be disassociated in these cases[156].

During replication fork progression, chromosome fragility and SV may occur due to mutations affecting fork progression such as mutations in genes encoding RPA subunits[157]. Additionally, errors in lesion repair or DNA adduct block fork progression, which might explain the genome instability led by defects in exposure to UV or carcinogens, as well as the defects in nucleotide excision repair (NER) or base excision repair. For example, in budding yeast, Rad3 is the subunit of TFIIH and is involved in NER. Rad3 mutant causes TFIIH retention and makes gap filling less efficient. As consequence, replication fork breakage and DSBs occur[158].

1.4.2 Fragile site and replication stress

Fragile sites are genomic regions exhibiting gaps or breakages on metaphase chromosomes under replication stress. They are considered to be associated with genome instability, especially large genomic deletions in cancer cells. Fragile sites are usually stable, but gaps or breaks are likely to occur when exposed to certain chemicals inducing replication stress such as aphidicolin (APH) and hydroxyurea (HU). To date, more than 200 fragile sites have been identified[159]. Based on the frequency of the breakage in the population, these loci can be divided into common fragile sites (CFSs) and rare fragile sites (<5% of individuals)[160]. Some common characteristics have been found among CFSs[161]. First, during DNA replication, the fragile sites are replicated in the very late stage of the S phase. Chemicals such as APH cause further delays in replication, resulting in some unreplicated loci even in the G2 phase[162]. Second, CFSs are enriched in large genes. For example, fragile site FRA3B encompasses the *FHIT* gene with a length of 1.5 Mb while the median length of human genes is around 23 kb[163,164]. Last, fragile sites are cell-type specific. Studies in different cell lines with various tissue origins characterised many cell-type-specific fragile sites although some were found across all cell types[164,165].

The vulnerability of fragile sites under replication stress can be attributed to the secondary structures that tend to form at these sites. DNA commonly adopts a canonical right-handed double helix structure known as B-DNA in most genomic regions. However, at fragile sites, DNA usually forms alternative structures that differ from the classic B-DNA helix[166]. Non-B DNA structures occur preferentially in genomic regions with certain repeats during DNA replication, damage repair, or transcription processes that involve the production of ssDNA. For example, after DNA strands are separated, CAG/CTG repeats can form hairpin structures based on intra-strand base pairing[167]. Although the barrier formed by a CAG/CTG repeat is weak, a gap is likely to be made during replication bypass. What exacerbates the situation is that the ligation step of nick repair is less efficient due to the interference of the hairpins on 5' flaps, leading to an unligated nick. Indeed, evidence of CAG/CTG-induced genome instability has been shown *in situ* and in culture[167–170]. There are many other non-B DNA structures linked to fragile sites besides hairpins. Mononucleotide stretches are the simplest repetitive sequences, which are able to trigger slipped-strand DNA. Slipped-strand DNA refers to a structure where DNA is mispaired and annealed, resulting in InDels during replication[166]. Furthermore, A or T repeats proximal to replication origins are associated with fork stalling and fork collapse under exposure to HU[171]. GC and GT repeats are preferable sites for left-handed Z-DNA[172,173]. The presence of Z-DNA and triplexes in a plasmid was reported to reduce the replication rate dramatically[174].

Moreover, DSB hotspots in the c-MYC proto-oncogene exhibit capability to form Z-DNA in B-cell precursors derived from acute lymphoblastic leukaemia patients[175,176]. Another secondary structure related to the breakage at fragile sites is G-quadruplex (G4). G4 contains stacked guanine tetrads of Hoogsteen-hydrogen bonded guanines[177]. G4 structure is enriched at the breakage sites of translocation events in human cells, indicating the role of G4 in genomic fragility[178,179]. In addition, the G4 structure is found within telomeric DNA which contains TTAGGG repeats. Under replication stress or the depletion of the helicases that are recruited to resolve the G4 structure, the telomeric sequence appears to be fragile[180,181]. The evidence suggests that G4 is also a preferable structure at fragile sites.

To sum up, fragile sites are enriched with repeat sequences that are able to form non-B DNA structures, resulting in fragility through many different mechanisms.

1.5 Transfer RNA

Transfer RNAs (tRNAs) are small non-coding RNA molecules that are highly abundant in cells. As fundamental components of the translation machinery, tRNAs deliver amino acids to ribosomes. Although tRNAs have long been considered purely physical adapters, recent findings reveal a broader spectrum of roles in many biological processes. For instance, the deregulation of tRNA transcripts exhibits effects on mRNA translation globally or selectively. Consequently, altered transcriptional outputs contribute to many diseases including cancers.

1.5.1 The transcription of tRNA genes

tRNA genes are transcribed by Pol III. Within tRNA genes, there are internal promoters, comprising A- and B-box elements. The promoters are recognized by the basal transcription factor TFIIIC, followed by the recruitment of TFIIIB. The Pol III complex, composed of 17 subunits, then binds and gives rise to tRNAs[182,183].

tRNA genes represent one of the largest multi-copy gene families in the vertebrate genome, and they are annotated computationally by tRNAscan-SE[184,185]. This algorithm scans for the sequences of promoters, Pol III termination and predicts the likelihood of folding into proper secondary structures[185]. Around 600 tRNA genes are annotated in the human genome[186]. After Pol III transcription, tRNAs fold into cloverleaf-like secondary structures, including several hairpin loops. One of these loops contains the anticodon, which is complementary to the codon of the mRNA. Based on the anticodons, tRNA genes can be divided into isoacceptor families, and according to the amino acids charged, they can be further classified into isotype classes[187]. An isotype class may contain one (e.g. $tRNA^{Met}$) or up to six (e.g. $tRNA^{Ser}$) isoacceptor families. Although there are 62 possible isoacceptors, not all are present in the genome, which can be explained by wobble base pairing. Within one isoacceptor family, there can be 1 (e.g. $tRNA^{SelCys}$ (TCA)) to 29 (e.g. $tRNA^{Ala}$(GCA)) tRNA genes[186].

Not all tRNA genes are actively used in mammalian genomes, and the usage is cell type-specific[187–189]. The active tRNA gene loci are enriched for euchromatic histone marks for example, histone (H) 3 lysine (K) 4 trimethylation (H3K4me3) and H3K27ac[182,189,190]. Although Pol III bindings to tRNA genes diverge, there is strong functional conservation at the levels of isoacceptor families and isotype classes across mammalian livers[187]. Interestingly, codon triplets are also conversed across these mammals, which suggests that the tRNA anticodon pool and mRNA codon pool are well-balanced. A similar observation

was found during the development of mice[190]. Although protein-coding gene expression is highly dynamic, the codon pool remains stable. The tRNA anticodon pool shows few changes and a high correlation with the codon pool. This indicates that tRNA gene transcription is tightly controlled to meet the demand during mRNA translation[190]. Furthermore, a study in human and mouse cell lines reported that the codon-driven translation efficiency was highly stable, regardless of cell type[191].

What will happen if this tight balance between the tRNA anticodon and the mRNA codon pool is disrupted? tRNA deletion studies in yeast unravelled a complex regulatory network of tRNA transcription and its phenotypic impacts. The cell growth rate was decreased in the challenging medium but not in the standard medium. Furthermore, closer inspection revealed that cells with deleted tRNA genes from single-copy isoacceptor families were highly susceptible to the stress condition[192]. Another work in human cells identified a subset of tRNA genes acting in cell proliferation and cell cycle arrest by sgRNA CRISPR/Cas9 screening[193].

These findings imply that tRNA gene transcription is tightly controlled. The disruption of transcription leads to phenotypic changes in different organisms.

1.5.2 Dysregulation of tRNA transcripts in cancers

Several oncogenic signalling pathways can impact not only Pol II-transcribed genes but also Pol III machinery directly or indirectly (**Fig. 7A**). For example, the extracellular signal-regulated kinase (ERK) and the proto-oncogene product c-MYC are able to directly bind to and activate TFIIIB, enhancing Pol III transcription[194,195]. Another example is mammalian target of rapamycin (mTOR), a protein kinase regulating cell growth. It increases the global transcription of Pol III by directly suppressing the activity of MAF1, a Pol III complex suppressor[196]. In contrast, tumour suppressors, retinoblastoma tumour suppressor protein (RB) and TP53, suppress Pol III activity through the repression of TFIIIB[197,198]. The increased transcription of tRNAs by Pol III in cancer is closely linked to the high demand for protein synthesis due to rapid proliferation[199].

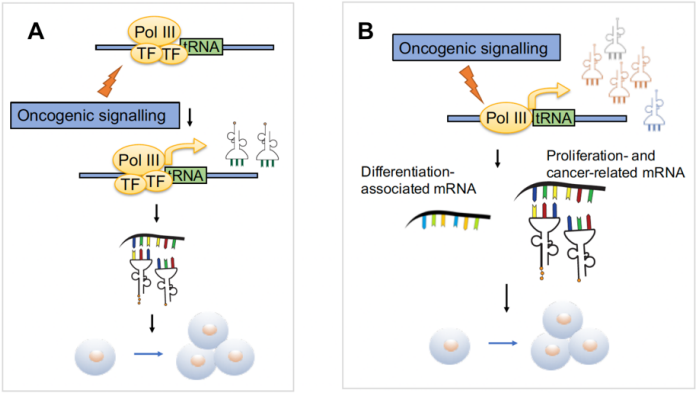

Fig 7. Oncogenic signalling regulates Pol III machinery in cancers
(**A**) Oncogenic signalling pathways stimulate Pol III machinery globally by directly or indirectly interacting with transcription factors (labelled as "TF") such as TFIIIB. (**B**) Some oncogenic signalling can drive selective changes in the transcription of tRNA genes, which benefits the translation of pro-tumorigenic mRNAs.

Besides impacts on global transcription, oncogenic signalling drives the selective overexpression of certain tRNA isoacceptor families. The altered abundance in certain tRNA molecules benefits the translation of pro-tumorigenic mRNAs based on the differences in the usage of synonymous codons[200] (**Fig. 7B**). In breast cancer, $tRNA^{Ser}$(GCU), $tRNA^{Thr}$(CGU) and $tRNA^{Tyr}$(GUA) are overexpressed in tumour compared with normal tissue[201]. $tRNA^{Glu}$(UUC) and $tRNA^{Arg}$(CCG) are significantly upregulated in metastatic breast cancer cell lines compared to poorly metastatic cells. These tRNAs enhance the abundance of Exosome Component 2 and activate the Glutamate receptor-interacting protein 1-associated protein 1 pathway by increasing the stability and ribosome occupancy of these transcripts[202]. Additionally, overexpression of initiator $tRNA^{iMet}$ genes significantly affects the global tRNA expression pattern, which leads to increased metabolic activity and proliferation rate in a breast cancer-derived epithelial cell line[203].

In sum, the transcription regulation and functions of tRNAs are more complicated than we previously thought.

2 Research aims

The overall aim of this thesis is to reveal obstacles during CRISPR/Cas9-mediated genome editing and provided solutions to achieve desired genomic outcomes in modified cells.

Aims of Paper I

Successful delivery of large-size CRISPR/Cas9 vectors in hard-to-transfect human cells using small plasmids

- To develop a non-viral delivery method, increasing the transfection efficiency of large CRISPR/Cas9 vectors and cell viability.

Aims of Paper II

Target-enriched nanopore sequencing and *de novo* assembly reveals co-occurrences of complex on-target genomic rearrangements induced by CRISPR-Cas9 in human cells

- To investigate the co-occurrences of undesired on-target genomic alterations that were overlooked due to technical limitations.
- To study biological consequences led by on-target genomic rearrangements.
- To provide a powerful and data-driven workflow to detect the on-target sequence content at the kilobase scale.

Aims of Paper III

CRISPR-Cas9-mediated genome engineering exaggerates genomic deletion at 10q23.31 including the *PTEN* gene locus mimicking cancer profiles

- To report an unintended large genomic deletion in a widely-used cell line.
- To characterize the impact and contributors of this large genomic deletion.
- To explore the significance of this deletion in human cancers.

3 Materials and methods

A detailed description can be found in the "Materials and Methods" sections of constituent papers I-III.

CRISPR/Cas9 single cell-derived clone generation

gRNAs were designed using the tools from the Zhang lab (https://zlab.bio/guide-design-resources). Each gRNA was cloned into CRISPR/Cas9 vector pSpCas9(BB)-2A-Puro (px459). The small-size plasmid pBlueScript was co-delivered with CRISPR/Cas9 vector into cells to increase transfection efficiency. Two px459 vectors (one with gRNA-1 and the other one with gRNA-2 cloned) and empty px459 (without gRNA cloned) were transfected into cells to generate modified and control clones, respectively. HAP1 cells were transfected using Turbofectin 8.0 (OriGene), and HepG2 were electroporated with NEON system (Invitrogen). Cells were selected with puromycin for two days and recovered in normal medium. Around 100-500 cells were seeded into 10 cm dishes. After the formation of single cell-derived clones, they were hand-picked and propagated separately.

Genomic DNA extraction and PCR genotyping

Cells were lysed and incubated overnight in 400 µL lysis buffer (0.5% SDS, 0.1M NaCl, 0.05M EDTA, 0.01M Tris-HCl, 200 µg/mL proteinase K) at 55°C. 200 µL of 5M NaCl were added and incubated on ice for 10 min. After centrifugation, 400 µL of the supernatant was mixed with 800 µL of 100% ethanol. The mixture was incubated on ice for at least 10 min. Genomic DNA was pelleted, washed once with 70% ethanol and dissolved in nuclease-free water.

Primers were designed using NCBI Primer-BLAST with default parameters. PCR was performed using Taq polymerase PCR (New England Biolabs) with 25 µL reaction system, according to the manufacturer's instructions. The PCR products were assessed by electrophoresis.

RNA extraction and purification

Cells were harvested in 700 µL Qiazol (QIAGEN). The cell lysis was mixed with 140 µL chloroform, shaken and incubated at room temperature. The upper phase was transferred after phase separation, and an equal volume of isopropanol was added. After thorough mix and incubation, RNA was pelleted, washed once and dissolved in nuclease-free water. Genomic DNA in RNA was removed using TurboDNase (Invitrogen) according to the manufacturer's manual. Resulting RNA samples were purified with RNA Clean & Concentrator Kit (Zymo Research) according to the manufacturer's protocol.

RNA-seq

The concentration and the integrity of RNA was measured with Qubit™ RNA HS Assay Kit and Agilent RNA 6000 Nano kit on Bioanalyzer. RNA molecules with PolyA tails were first enriched with NEBNext® Poly(A) mRNA Magnetic Isolation Module. The enriched molecules were used as input for library preparation with NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina. Library was assessed and quantified with Agilent High Sensitivity DNA chips on Agilent Bioanalyzer and KAPA SYBR® FAST qPCR kit. Sequencing was performed on Illumina platform Nextseq500 with pair-end mode using NextSeq 500/550 High Output v2 kit for 150 cycles.

RNA-seq data analysis

Sequencing reads were assessed with FastQC. Reads with low quality and adaptor sequences were trimmed using Trimmomatic. Reads first filtered for ribosomal RNA and then aligned to human reference genome hg38 with HiSAT2. Bedgraph files without soft-clipped reads were generated for visualization. The aligned reads were further quantified for each transcript. The raw count tables were used as input for differential expression analysis with DESeq2.

ChIP-seq

ChIP-seq was performed as previously described[187,191]. Briefly, 15-20M HAP1 or HepG2 cells were fixed, lysed and sonicated. The lysis with fragmented DNA was incubated with H3K4me3 (05-1339, 194 Millipore), H3K27ac (ab4729, abcam), or Pol III antibodies recognizing antigen POLR3 antibodies[187]. The enriched DNA was used to perform library preparation with Takara SMARTer® ThruPLEX® DNA-seq Kit. Library quality was assessed with Agilent Bioanalyzer High Sensitivity DNA chips and quantified with KAPA SYBR® FAST qPCR kit. Sequencing runs were performed on Illumine Nextseq 500 with the NextSeq 500/550 High Output v2 kit for 75 cycles, single-end mode.

ChIP-seq data analysis

ChIP-seq reads were assessed by FastQC and then aligned to the human reference genome hg38 using BWA. PCR duplicates and reads assigned to the ENCODE exclusion list were removed with SAMtools and NGSUtils. Afterwards, bam files were indexed and sorted. Peak calling was performed using MACS2. The differential enrichment analysis was carried out to identify differentially acetylated regions using DiffiBind. Bedgraph files were generated with filtered reads using deepTools and visualized in IGV.

Xdrop-LRS data analysis

The analysis was summarized (**Fig. 8**). FASTQ reads were first corrected using Canu and then SACRA to process chimeric reads. Reads were mapped to the region of interest using minimap2. The sequences of mapped reads were extracted and used as input for *de novo* sequence assembly using Canu or Raven. The output contigs were assessed in several ways: (I) pairwise alignment using Needle; (II) MegaBLAST against NCBI database of nucleotide collection; (III) and manual inspection of the read alignment to the assembled contigs. If the assembled contig needed to be extended, reads aligned to the 5′ or 3′ end of the contig were extracted and used for the second round of assembly. If the read coverage was too low to perform *de novo* assembly, a manual extension was performed by visualizing, comparing and summarizing in IGV. The success in an assembled contig was confirmed by the visualization of aligned reads (raw reads and corrected reads) spanning breakpoints of the contig using IGV.

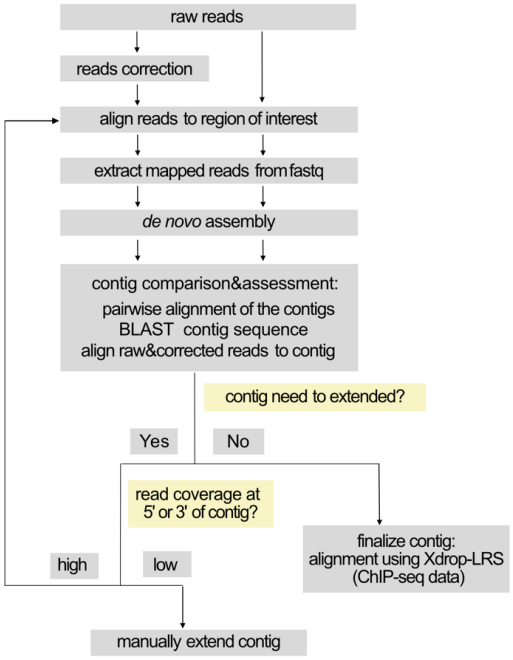

Fig 8. Customized *de novo* assembly-based analysis
The analysis consists of three steps: read correction and extraction; *de novo* assembly; contig comparison and assessment.

All the scripts for sequencing data analysis can be found under these two Github repositories:

https://github.com/KeyiG/Cas9_ontarget_alteration.git

https://github.com/KeyiG/HAP1_10q23_P-Pdel.git

4 Results

4.1 Paper I

4.1.1 An optimized method to deliver CRISPR/Cas9 vectors into human cells

Efficient delivery of the CRISPR/Cas9 system into human cells is very challenging. Due to the large size of Cas9 and other required components such as selection markers, the transfection of the CRISPR/Cas9 vector (~9-19 kb) is usually inefficient and causes cell death. To overcome these limitations, we co-transfected cells with a small-size vector during electroporation (**Fig. 9**).

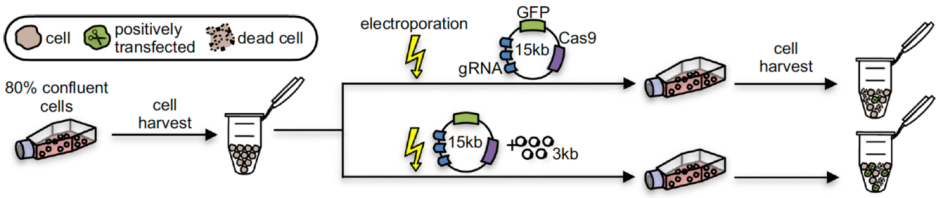

Fig 9. Illustration of cell transfection
Delivery of large CRISPR-GFP vectors (15 kb) without (upper track) and with (lower track) small vectors (3 kb) to human cells using electroporation. Co-transfection results in high delivery (shown as green cells in the tube) efficiency and low cell death rate (shown as dead cells in the tube). The figure is adapted from Søndergaard *et al*[58] with minor changes under Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).

Co-transfection of 3 kb vectors with CRISPR-GFP vectors (15 kb) increased the transfection efficiency from 4.2% to 40% and decreased the cell death from 91% to 45% in a human lung cancer cell line, A549. Furthermore, the robustness of this method has been verified in numerous cell types including many hard-transfect human cell lines such as Huh7 and HepG2.

Importantly, the small-size vectors used here (pBlueScript) are designed for bacterial expression without mammalian promoters so genes carried by vectors are unable to utilize transcription machinery in human cells. Therefore, the application of this vector for research in human cells should be safe.

4.2 Paper II

CRISPR/Cas9 has revolutionized genome editing. However, many undesired and unpredictable genomic consequences have raised safety concerns in both research and clinical applications. In this paper, we described the co-occurrence of on-target genomic alterations in CRISPR/Cas9-modified cells. The complexity of such rearrangements was underestimated before due to technical limitations. Therefore, we introduced an advanced workflow, allowing thorough assessment of on-target sequence content.

4.2.1 Target regions were detected in validated deletion clones modified by CRISPR/Cas9

After optimizing the transfection efficiency (**Paper I**), we proceeded with deleting a genomic region (870 bp) containing two active tRNA genes on Chr 17 with CRISPR/Cas9 dual-gRNA system (Δt) in two human cancer cell lines, HepG2 and HAP1 (**Fig. 10A**). By amplifying with primers annealing to sequences flanking binding sites of two gRNAs (flanking primers) and Sanger sequencing of amplicons, single cell-derived clones with homozygous deletion were identified (**Fig. 10A**). Using a similar approach, deletion clones with the genomic region between the two tRNA genes removed were generated (Δi) (**Fig. 10B**).

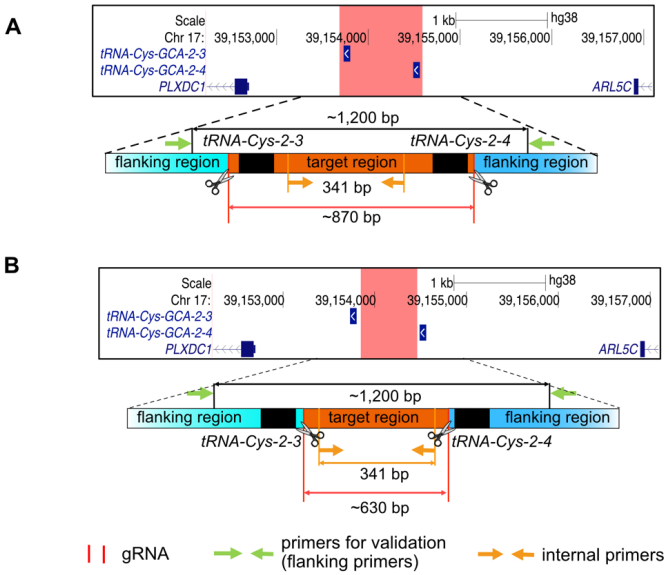

Fig 10. Genomic regions on Chr 17 were targeted by CRISPR/Cas9
(**A-B**) The UCSC genome browser views (hg38) and illustrations of CRISPR/Cas9 deletion strategies show the genomic location containing our two targets: the genomic region with two tRNA genes (**A**) and the genomic region between two tRNA genes (**B**). Two sets of primers (flanking primers: green; internal primers: yellow) were used for deletion validation (flanking primers) and target region detection (internal primers). The figure is adapted from Geng *et al*[61] with permission granted.

Although the sequence content of the amplicon obtained from each validated clone suggested a homozygous deletion, the targeted regions were still detectable in some of the validated clones by ChIP-seq or PCR with primers annealing within the target region (internal primers) (**Fig. 10**).

To explain this observation, we considered many reported factors leading to the failure of homozygous editing including mutation in the PAM sequence, unsynchronized cleavage, inversion or duplication. However, we ruled out all of these possibilities based on our ChIP-seq and PCR results. We therefore hypothesized that the detection of the target region in some validated deletion clones are consequences of complex genomic alterations that have not been comprehensively documented.

4.2.2 Customized workflow revealed complex on-target genomic alterations

The Xdrop technology has been recently developed and used for detecting insertion in CRISPR/Cas9-modified cells[129]. To investigate the sequence proximal to the DSB site in an unbiased manner, we developed a *de novo* assembly-based approach. With this workflow, we were able to obtain the on-target sequence content with a much longer length compared with other available methods on Illumina platform, and no prior knowledge or assumption is required since reference sequence for alignment is unnecessary here.

With this data-driven tool, we revealed the co-occurrence of complex genomic alterations in the HAP1 deletion clone Δt72, as well as the HepG2 deletion clones Δt15, Δt8 and Δi50 (**Fig. 11**). In the HAP1 deletion clone Δt72, besides an allele with desired deletion, duplication and inversion of the target region were found on the other allele. In the HepG2 deletion clone Δt15, in addition to duplication and inversion of the target region, exogenous DNA fragments from the *E. coli* genome and CRISPR/Cas9 vector were inserted at DSB sites. Two scenarios of on-target alterations were identified in the HepG2 deletion clone Δt8, which could represent the rearrangements of two alleles. In this first scenario, inversion of the target region and DNA fragments from non-targeted chromosomes (Chr 21 and Chr 8) were found at the on-target site. No evidence indicated that these clustered interchromosomal rearrangements were induced by CRISPR/Cas9 off-target editing. In the second scenario, an inverted target region ligated to sequences derived from CRISPR/Cas9 vector was inserted at DSB sites. In the HepG2 deletion clone Δi50, a 331 bp-long deletion occurred at the 3' DSB site. Although MMEJ is considered to be the mechanism for LDs, we did not detect micro-homology around the breakage.

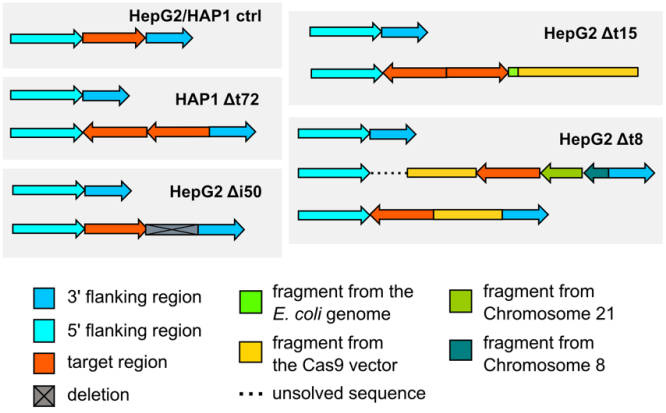

Fig 11. Multiple on-target genomic aberrations co-occurred and were captured by our workflow
Duplication, inversion of the targeted region, as well as exogenous DNA insertion and interchromosomal rearrangements, were identified in HAP1 deletion clone Δt72 and the HepG2 deletion clones Δt15, Δt8 and Δi50. The arrows indicate the genomic orientation based on the coordinates of the sequence in the human genome. The figure is adapted from Geng *et al*[61] with permission granted.

4.2.3 On-target genomic rearrangements exerted biological consequences

We further evaluated the biological consequences of these genomic alterations.

First, the inserted and fragmented CRISPR/Cas9 vector components in the HepG2 deletion clone Δt15 and Δt8 were expressed, which could be toxic to cells.

Second, proliferation assays showed that in the HAP1 Δt deletion clones, on-target genomic alterations and the resulting functional DNA brought significant growth advantages to cells.

Last, based on the genomic rearrangements identified with our workflow, we utilized a two-colour TaqMan-qPCR assay to determine the ratio between alleles with on-target alterations and the desired deletion. The fractions of alleles with on-target aberrations varied among the tested clones (**Table. 1**). We reason that it could be caused by the proliferation advantage or disadvantage elicited by the on-target rearrangements.

sample	formula*	observed *F*	alleles with genomic alterations : alleles with expected deletion (estimated)
HAP1 Δt72	$\frac{2\times F_{target}}{1}=\frac{target}{flank}=2^{-(\Delta Cq_s-\Delta Cq_c)}=1.12$	0.56	1:1
HepG2 Δt15	$\frac{2\times F_{target}}{1}=\frac{target}{flank}=2^{-(\Delta Cq_s-\Delta Cq_c)}=0.04$	0.02	1:49
HepG2 Δt8	$\frac{F_{target}}{1}=\frac{target}{flank}=2^{-(\Delta Cq_s-\Delta Cq_c)}=0.72$	0.72	2:1
HepG2 Δi50	$\frac{F_{target}}{1}=\frac{target}{flank}=2^{-(\Delta Cq_s-\Delta Cq_c)}=0.39$	0.39	1:2

$*\ \Delta Cq_{s/c} = Cq\text{target} - Cq\text{flank}$ s: sample; c: ctrl
F: fraction of alleles with on-target genomic alterations

Table 1. The ratios between alleles with on-target alterations and desired deletion were determined
The table summarizes the formula used to calculate the fractions of the alleles carrying intact target region (*F*) from the cycle number (Cq) obtained from qPCR; *F* values were obtained from the experiment, and the estimated ratios (last column) were based on the *F* values and the reasonable number of alleles one cell may carry. The table is from Geng *et al*[61] with permission granted.

4.3 Paper III

In addition to undesired and elusive on-target aberrations, CRISPR/Cas9-associated non-target mutagenesis and SVs are major impediments. In this study, we reported an unexpected genomic deletion at 10q23.31.

4.3.1 Unexpected 10q23.31 deletion was found in CRISPR/Cas9-modified HAP1 cells independent of gRNA sequences

We previously targeted genomic regions located on Chr 17 (described in **Paper II**) and generated desired homozygous deletion HAP1 clones (Δt and Δi) (**Fig. 10**). When inspecting the genomic occupancy of H3K4me3 and H3K27ac, we observed an abnormal ChIP-seq read depletion on Chr 10 (10q23.31) in Δt1 and Δi17 clones but not in the control (ctrl) clone (**Fig. 12**). The undetectable ChIP-seq enrichment and background suggested a genomic deletion. To confirm it, we designed a pair of primer annealing within the deleted region. The absence of PCR amplicons in Δt1 and Δi17 verified the deletion, and the PCR screening in our other CRISPR/Cas9-modified HAP1 cell clones implied the frequent losses of 10q23.31.

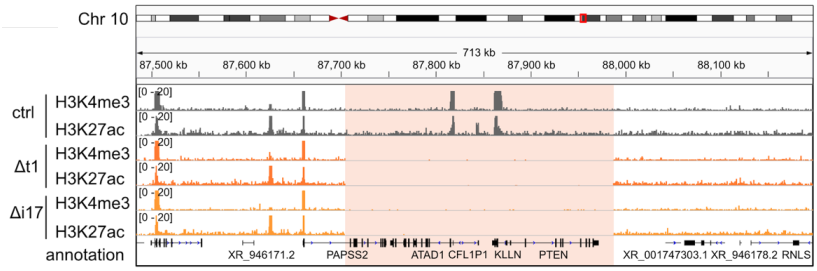

Fig 12. Abnormal ChIP-seq read distribution in modified HAP1 clones indicated an unintended and large genomic deletion at 10q23.31
Coverage tracks of H3K4me3 and H3K27ac ChIP-seq performed in control (ctrl, dark grey), Δt1 (dark orange) and Δi17 (orange) clones at the *PAPSS2-PTEN* locus on Chr 10 (highlighted in red).

According to our ChIP-seq read distribution in Δt1 and Δi17 clones, the deleted region is around 287 kb, starting from the first intron of *PAPSS2* to the downstream of the *PTEN* gene body. We, therefore, refer to this deleted region as the *PAPSS2-PTEN* locus. It encompasses four protein-coding genes: *PAPSS2*, *ATAD1*, *KLLN* and *PTEN*, which are widely expressed in various commonly-used cell lines.

Since the sequences of gRNAs used to generate Δt and Δi clones were highly diverse, the *PAPSS2-PTEN* locus deletion found in CRISPR/Cas9-modified HAP1 cell clones was gRNA-independent.

4.3.2 The *PAPSS2-PTEN* locus deletion was associated with gene expression changes and differential acetylation of H3K27

We next examined whether the transcriptome was altered in the clones with the *PAPSS2-PTEN* locus deletion. In alignment with our ChIP-seq data, the RNA-seq data showed no signal at the *PAPSS2-PTEN* locus in the Δt1 clone while transcripts derived from *PAPSS2-PTEN* locus were captured in the ctrl clone (**Fig. 13A**).

Detailed inspection revealed unique and abnormal transcript signatures in Δ*PAPSS2-PTEN* cells at 10q23.31. Since the deleted *PAPSS2-PTEN* locus does not include the first exon and promoter region of *PAPSS2*, there were reads mapping to the first exon of the *PAPSS2* gene in Δt1 and ctrl clones (**Fig. 13B**). Furthermore, we found reads located downstream of the *PAPSS2-PTEN* locus in Δt1 but not in the ctrl clone (**Fig. 13C**). This is likely caused by Pol II readthrough. In Δ*PAPSS2-PTEN* cells, Pol II could still be recruited to the transcription-start site of *PAPSS2*. However, due to the deletion, Pol II might lose its termination signal and give rise to a new transcript.

Surprisingly, we identified CRISPR/Cas9-modified HAP1 cell clones in three independent published RNA-seq datasets which shared the exact transcript signatures at the *PAPSS2-PTEN* locus as found in our Δt1 clone (**Fig. 13A-C**). These HAP1 cell clones carried the intended CRISPR/Cas9-mediated deletion for *METAP1* (ΔM in dataset1), *SREBF2* (ΔS in dataset2) and *SMARCC1* (ΔSM1 in dataset3), respectively (**Fig. 13A-C**).

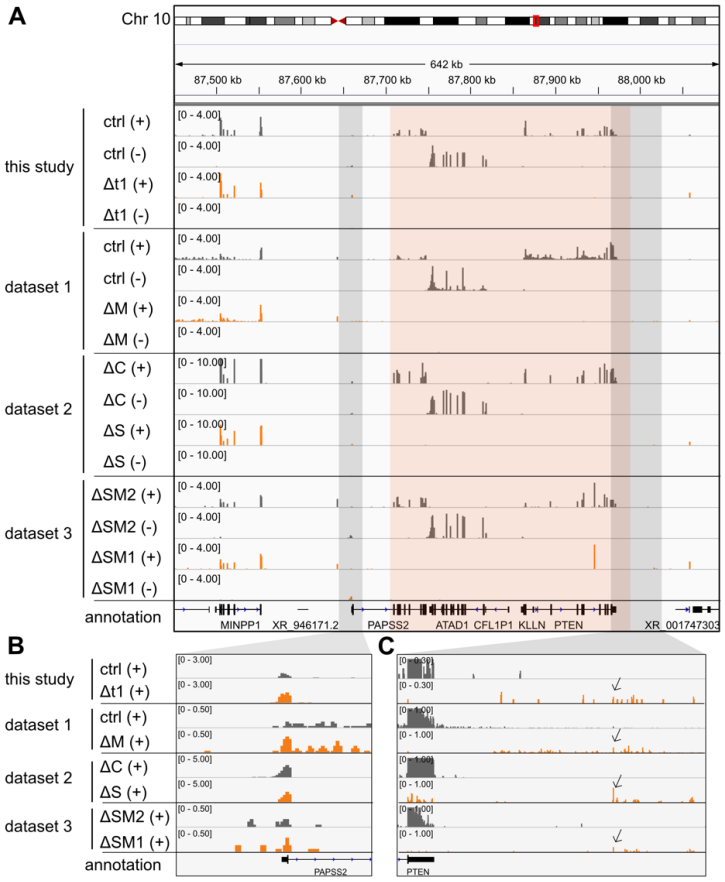

Fig 13. The transcripts signature of *PAPSS2-PTEN* locus deletion was shared across independent datasets
(**A**) Coverage tracks of the various genotypic HAP1 clones with (dark grey) and without (orange) *PAPSS2-PTEN* locus (highlight in red) in our own (labelled as this study) and three published RNA-seq datasets. The coverage is separated by the strand (labelled in brackets). (**B-C**) Magnified views of read coverage at the up- (**B**) and downstream (**C**) regions (highlight in light grey in **A**) of the deleted genomic locus. The arrows show the unique transcripts in HAP1 clones without the *PAPSS2-PTEN* locus.

To assess the alteration of gene expression patterns linked to the deletion of the *PAPSS2-PTEN* locus, we combined our and the three publicly available datasets. Differential expression analysis revealed 1489 downregulated and 1429 upregulated genes. The downregulated genes were enriched for processes such as DNA replication, DNA damage repair, cell cycle, as well as DNA- and histone binding. The functions of the upregulated genes were broader, including development, cell size, vesicle organization, transportation and catalytic activities.

Since our differential expression analysis indicated changes in molecular processes of DNA- and histone binding, we took advantage of the H3K27ac ChIP-seq data from two independent sources: our data in Δt1 and ctrl clones, as well as a published dataset in which one of the CRISPR/Cas9-edited HAP1 clones carried *PAPSS2-PTEN* locus deletion.

Differential enrichment analysis discovered 852 differentially acetylated (DAc) sites, among which around 70% were located in promoter regions or gene bodies, indicating their regulatory potential on gene expression. We next focused on the DAc peaks that were proximal to differentially expressed (DE) genes (≤ 5 kb) and extracted 131 DAcDE pairs. The fold changes of H3K27ac enrichment were highly correlated with changes in their paired DE gene expression, implying that changes at the chromatin level affected the transcriptome.

4.3.3 Generation of CRISPR/Cas9-modified cells exaggerated 10q23.31 deletion

To investigate determinants of 10q23.31 deletion, we tested the following factors: the presence of gRNA with Cas9 protein and the delivery of a small vector as a transfection-efficiency enhancer. Single-cell derived HAP1 clones were generated and assessed by qPCR or PCR, amplifying with primers annealing within the *PAPSS2-PTEN* locus.

First, we transfected cells with CRISPR/Cas9 vectors with or without gRNA and selected using puromycin. The difference in the frequency of losing the *PAPSS2-PTEN* locus (83% with gRNA; 67% without gRNA) was insignificant. It confirmed our previous observation that this deletion was gRNA-independent.

Second, we compared cells transfected with the small vector (pBlueScript) to cells without transfection followed by puromycin selection. The deletion of the *PAPSS2-PTEN* locus was detected in 67% of clones transfected with pBlueScript and 71% of clones receiving no plasmid. Therefore, co-transfection of small vectors (pBlueScript) did not contribute to the deletion.

Third, there were 29% of clones without transfection or puromycin selection harbouring the deletion of the *PAPSS2-PTEN* locus. This finding confirmed that the deletion was not the direct consequence of CRISPR/Cas9 off-target activity. However, the generation of CRISPR/Cas9-modified cells dramatically increased the frequency of the deletion.

Furthermore, our cell cycle analysis revealed that under exposure to excessive cellular toxicity, losing the *PAPSS2-PTEN* locus helped cells to escape cell cycle arrest.

4.3.4 The *PAPSS2-PTEN* locus deletion in HAP1 cells resembled a collateral deletion at 10q23.31 in human cancers

Since we detected the deletion at the *PAPSS2-PTEN* locus in a human cancer cell line, we then explored the occurrence of this deletion in patients across various cancer types using publicly available data.

We found that among 466 patients with the *PTEN* gene deletion, many patients also carried the deletion at *KLLN* (67%), *ATAD1* (59%), and *PAPSS2* (43%) gene loci. This observation suggested that collateral deletion at the *PTEN* locus was more prevalent than single-gene deletion in cancer patients.

By performing PCA on the transcriptomic profiles, we observed that patient samples exhibited distinct grouping patterns based on the presence or absence of *PAPSS2-PTEN* locus deletion. It is striking since the patient samples selected for the analysis were not controlled by the factors which usually distinguish individual samples such as gender and tumour grade. Additionally, many DE genes associated with *PAPSS2-PTEN* locus deletion identified in HAP1 cells were also altered in patient samples with 10q23.31 deletion.

To conclude, our study identified a deletion at 10q23.31 with significant implications for CRISPR/Cas9 application in HAP1 cells, as well as its potential clinical relevance (**Fig. 14**).

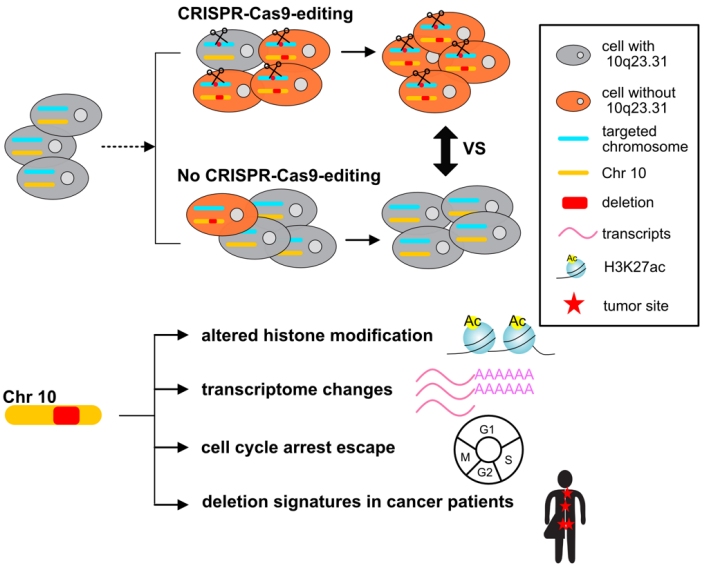

Fig 14. Graphic summary of Paper III
The generation of CRISPR/Cas9-modified HAP1 cell clones increases the occurrence of the unexpected *PAPSS2-PTEN* deletion on Chr 10. The deletion is associated with changes at multiple levels in HAP1 cells and can be found in various human cancer types.

5 Discussion

5.1 Paper I

There are many methods to deliver CRISPR/Cas9 components. LV-based delivery is very efficient, but it causes many safety concerns such as high off-targeting frequency and the random integration of lentiviral sequences[55]. Besides LV, the CRISPR/Cas9 system can be delivered as pre-incubated RNP complex or CRISPR/Cas9 vectors with gRNA sequences cloned. RNP complex delivery is generally considered to have the least safety concerns[204]. However, the RNP system is vulnerable to degradation mechanisms in cells and much more expensive than vector-based delivery[205]. A CRISPR/Cas9 vector is more stable than the RNP complex and shows much fewer safety concerns than LVs. Therefore, the easy-to-use and economical method to assure an efficient transfection reported in this study carries significance in the application of CRISPR/Cas9.

The underlying mechanism through which small vectors serve as transfection enhancers remains unclear. It is well established that small plasmids exhibit higher efficiency in cellular uptake when compared to larger plasmids. Based on this observation, we hypothesize that upon passing through the cell and nuclear membranes, small vectors may temporarily increase the permeability of these membranes. The negatively charged small vectors (such as pBlueScript) help to maintain the membrane pores created by electric fields open, allowing for efficient delivery of CRISPR/Cas9 vectors into the cells.

Furthermore, this co-transfection pattern may prevent the CRISPR/Cas9 vector from becoming entangled in multiple membrane pores, thus promoting increased cell viability[206]. Alternatively, the high survival rate observed may be attributed to the lower cytotoxicity exhibited by small vectors when compared to larger vectors. Some DNA sensors/receptors present on cell membranes can trigger programmed cell death[207]. During co-transfection, small vectors travel towards the cell membrane at a higher velocity than larger vectors and thus reach the cell membrane first, saturating the capacity of the DNA sensors/receptors. This, in turn, shields the sensors from the larger vectors and reduces cell death.

It is important to note that these explanations are highly speculative and further work is required to elucidate the exact mechanisms involved.

In sum, our simple, non-hazardous and cost-effective method has been proven to increase the transfection efficiency of large-size CRISPR/Cas9 vectors in many cell lines. It has shown great potential in clinical and industrial applications.

5.2 Paper II

In this study, we reported the co-occurrence of multiple on-target genomic alterations in CRISPR/Cas9-edited cells using an advanced target enrichment-LRS technology (Xdrop) followed by our customized analysis.

5.2.1 Superior performance of Xdrop-LRS with *de novo* assembly workflow

In the era of CRISPR/Cas9, many undesired on-target aberrations have been documented such as chromothripsis, off-target mediated translocation, LDs, plasmid integration as well as duplication or inversion in dual gRNA system (discussed in subsection **1.3.1**). However, most work either failed to detect allelic resolution or only captured these alterations as a single event in one specific allele. Consequently, the complexity of altered on-target genomic events was underestimated. It is largely due to technical limitations.

For example, duplication and inversion were initially reported using a conventional PCR approach in cultured human cells and mouse models[131,132]. The choice of this method impedes the identification of whether these two events co-occur on the same DNA molecule. Additionally, the standard PCR approach requires prior knowledge to design primers, which inevitably leads to oversight of unexpected rearrangements. Similarly, investigation relying on fluorescence *in situ* hybridization is also biased by the probe design[208]. To overcome these limitations, many powerful tools (discussed in subsection **1.3.2**) are developed to characterize on-target sequence content in high-throughput and less biased manners. However, these methods generally rely on steps including DNA fragmentation, PCR-based enrichment and sequencing on Illumina platform, which limited the length of on-target sequences detected to hundreds of base pairs.

On the contrary, Xdrop-LRS can reveal sequences in the order of kilobases, allowing the identification of multiple genomic alterations located on the same allele. Compared with other target enrichment strategies compatible with LRS, Xdrop requires less DNA input and shows a higher recovery rate[129]. Furthermore, unlike the classic alignment approach, our *de novo* assembly-based analysis has no requirement for reference sequence, which further reduces the bias introduced during data analysis.

Xdrop-LRS combined with customized *de novo* assembly-based analysis outcompetes many available methods in revealing co-occurred unexpected genomic alterations, but it is unable to detect large-scale rearrangements such as aneuploidy or CRISPR/Cas9 off-targeting events.

5.2.2 Hypothetical models of on-target genomic rearrangements

Although we cannot reconstruct or track all events after the induction of DSBs in cells, we propose hypothetical models in our deletion clones according to our data.

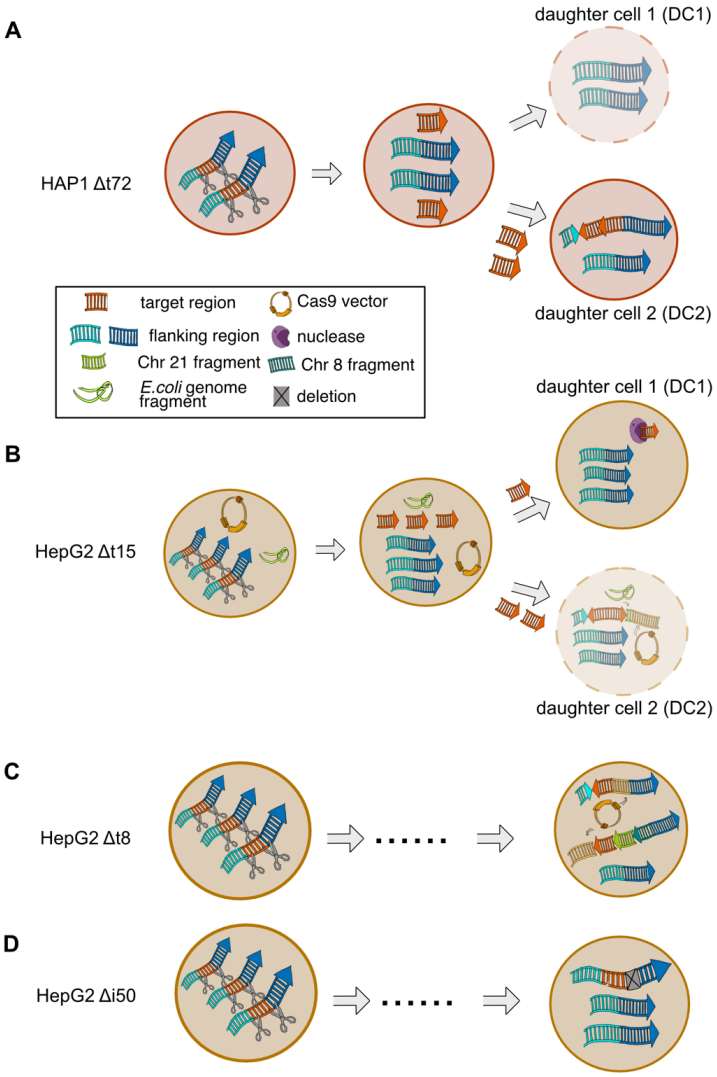

Fig 15. Hypothetical model in HAP1 and HepG2 deletion clones with on-target aberrations
(**A-D**) schematic of our models in the deletion clone HAP1 Δt72 (**A**), HepG2 Δt15 (**B**), Δt8 (**C**) and Δi50 (**D**). Cells with growth disadvantages are shown with lighter colours and dashed outer lines. The figure is adapted from Geng *et al*[61] with permission granted.

In the deletion clone HAP1 Δt72, HAP1 cells turned diploid before the targeted region was cleaved. The resulting DSBs were not repaired before cell division. One daughter cell (DC, DC2) received two fragments derived from targets, and the other DC (DC1) did not receive any target-derived fragments. NHEJ repaired the DSBs in DC1 while DC2 employed MMEJ, leading to duplication and inversion of the target. Cells with detectable targets (derived from DC2) become dominant because of proliferation benefits (**Fig. 15A**). Similar to the HAP1 deletion clone Δt72, we reason that HepG2 deletion clone Δt15 was heterogeneous. In DC2, two target-derived fragments were inserted at the targeted site on only one of the alleles, together with exogenous DNA fragments from the *E. coli* genome and CRISPR/Cas vectors. These complex rearrangements might bring proliferation disadvantages to cells, which could

explain why only 2% of alleles in HepG2 deletion clone Δt15 carried the on-target alterations (**Table 1**, **Fig. 15B**). In the HepG2 deletion clone Δt8, we postulate that there were two alleles representing the two contigs assembled using Xdrop-LRS data, and the third allele carried the desired deletion (**Fig. 15C**). In the HepG2 deletion clone Δi50, a large on-target genomic deletion was located on one allele while there were two alleles with the expected deletion (**Fig. 15D**).

In conclusion, with Xdrop-LRS followed by a customized *de novo* assembly-based analysis workflow, we documented the complexity of on-target genomic rearrangements in human cells.

5.3 Paper III

5.3.1 Mechanism of the *PAPSS2-PTEN* locus deletion

Although the mechanism initiating the *de novo* emergence of *PAPSS2-PTEN* locus deletion is unclear, we postulate that it is linked to genome fragility.

Genome-wide screening identified more than two hundred fragile sites including FRA10A where the *PAPSS2-PTEN* locus resides[159,209]. Fragile sites are usually enriched with DNA sequences prone to form non-canonical DNA structures (discussed in subsection **1.4.2**). We have found several pieces of evidences suggesting that non-B DNA structures might form in the *PAPSS2-PTEN* locus, resulting in the fragility of this locus: (I) Sequences in *PTEN* exon 1 were reported to form a stable alternative secondary structure *in vitro*[210]; (II) using publicly available data in human cells, we identified two G-quadruplex sites located within *PAPSS2-PTEN* locus[211]; (III) our analysis on sequences at the break boundaries of *PAPSS2-PTEN* locus revealed A-mononucleotide and GT-repeats, which can lead to slipped-strand DNA[166] or left-handed Z-DNA, respectively[175].

Additionally, the generation of HAP1 cells likely makes them prone to DNA breaks. HAP1 cells were derived from KBM-7 cells through the transfection of Yamanaka factors, which are known to trigger replication stress and genome instability[212,213].

5.3.2 Clinical implication of the *PAPSS2-PTEN* locus deletion

The frequent loss of 10q23.31 and its association with prostate-specific antigen reoccurrence in patients have been reported. Besides impairing the function of PTEN as a tumour suppressor, the deletion of *PAPSS2* is considered a contributor to the reoccurrence[214]. Additionally, a recent study showed that cancer cells carrying the co-deletion of *ATAD1* and *PTEN* genes were prone to BIM-mediated apoptosis[215]. These studies reinforce the significance of collateral deletion in cancers.

Furthermore, large genomic deletion could create collateral lethality in cancer cells, which makes cancer cells more responsive to treatment. For example, in glioma, cells carrying a large deletion including the *ENO1* gene were unable to survive when the paralogue ENO2 was depleted or inhibited. On contrary, the suppression of ENO2 showed no effect on the proliferation of cells with intact *ENO1* gene[216]. The vulnerability led by collateral deletion grants an auspicious possibility to the specific and efficient treatment. Future studies can exploit the possible collateral lethality created by the deletion of the *PAPSS2-PTEN* locus. For example, PAPSS2 is involved in sulfur metabolism. Does inhibiting the paralogue of PAPSS2 (PAPSS1) decrease the survival of cancer cells with 10q23.31 deletion but exhibit no effect on non-cancerous cells with intact 10q23.31?

Although we first identified the loss of the *PAPSS2-PTEN* locus in HAP1 cells, this deletion showed significant clinical relevance and enormous potential in providing targeted and effective treatment options through collateral lethality.

6 Conclusions

In **Paper I**, we have shown an easy and non-hazardous method to increase the transfection efficiency when delivering large-size CRISPR/Cas9 vectors into hard-to-transfect human cells. By co-transfecting small-size vectors with large-size vectors, we successfully enhanced the efficiency and decreased cell death. Furthermore, this method has been intensively tested in many human cell lines.

In **Paper II**, we have documented the co-occurrence of complex on-target genomic alterations in CRISPR/Cas9-modified human cells. These unintended on-target aberrations resulted in functional DNA derived from the targeted region in the genome despite the successful induction of on-target DSBs. This adverse effect could bring biological consequences to cells such as the expression of exogenous DNA fragments and affecting cell growth. In addition, we introduced an advanced workflow, coupling Xdrop technology with our customized *de novo* assembly-based approach. This powerful and data-driven tool enables to dissect on-target sequences in the order of kilobases, expanding the ability to detect on-target rearrangements.

In **Paper III**, we have reported an undesired genomic deletion of 287 kb in HAP1 cells which are commonly used as cell models in CRISPR/Cas9 screening or single-target editing experiments. This genomic region encompasses four widely-expressed protein-coding genes including *PTEN*. This deletion was associated with changes at chromatin and transcript levels. The deletion was not led by CRISPR/Cas9 off-target editing but the generation of CRISPR/Cas9-modified cells greatly increased the frequency of this deletion. Furthermore, our data suggested that this deletion initially identified in HAP1 cells resembled a common deletion pattern in cancer patients.

7 Points of perspective

Numerous clinical trials are underway to evaluate the potential of CRISPR/Cas9-based therapy of genetic diseases and disorders such as Sickle Cell Disease, β-Thalassemia and Duchenne muscular dystrophy[217]. While CRISPR/Cas9 gene therapy shows great promise, safety and efficiency concerns, such as those discussed in this thesis, need to be addressed instead of rushing headlong into the clinical use.

Given the uncertainties and risks surrounding the use of CRISPR/Cas9, further work should focus on reducing unintended genomic rearrangements and off-target cleavages. In addition to the approaches discussed in subsection **1.1.2.3**, there are many other promising strategies to improve CRISPR/Cas9 technology: (I) Continuous development of high-fidelity Cas variants. For example, variants HiFiCas9[218] and Cas9TX[219] minimize the off-target editing and chromosomal translocation, respectively. (II) Reducing the duration of Cas9 activity in cells. Off-target cleavages increase under the prolonged presence of Cas9. Therefore, delivering the Cas9-gRNA complex directly to cells (RNP system) is considered to be optimal in minimizing the off-target sites. (III) Decreasing the binding time of Cas9 at DNA. Excessive binding duration could impede the accessibility of DNA damage repair mechanisms. (IV) Increasing the usage of HR to repair the DSB induced by CRISPR/Cas9. Through HR, precise editing can be achieved while NHEJ and MMEJ can lead to many undesired consequences such as chromothripsis and LDs. Efforts have been made to suppress NHEJ by inhibiting a key enzyme[220] and to favour the choice of HR by controlling the cell cycle stage[221,222].

Currently, most available assays to evaluate the CRISPR/Cas9 genomic outcomes have their inherent technical limitations. A large amount of input material, laborious work and high cost also hinder the translation of these methods into clinically feasible and scalable approaches. Therefore, it is urgent to develop tools that enable comprehensive assessments and are suitable for clinical application.

In conclusion, the use of CRISPR/Cas9 must be approached with utmost caution and prudence. CRISPR/Cas9-based gene therapy should be limited to patients with life-threatening conditions. Enthusiasm must not cloud judgement, especially when it comes to matters of public health.

8 Acknowledgements

One does not simply walk through the journey of PhD study alone. This degree is guarded by more than just intellectual challenges. There are doubts and uncertainties that do not rest. The critical eye of the science community is ever watchful. Clearly, I would not have made it without all the support from everyone around me.

First of all, my main supervisor, **Claudia Kutter**. It will easily take several pages if I list everything you have done for me that I really appreciate. To summarize all into one sentence, I say you offered me the best PhD candidate training. You are always so enthusiastic and supportive and never shoot down my ideas even if some of them are not very feasible and are sort of stupid. You have shown me an example of being a great researcher, doing good science and always choosing the right path instead of the easy one. Being your PhD student is definitely among the best decisions I have made.

I am also very lucky to have great co-supervisors. **Philip Ewels**, thank you for always being around when I need help, especially with data analysis. **Laura Baranello**, thanks for being very caring and for all the valuable feedback which always pushes me to think deeper about my project. **Cecilia Williams**, thanks for all the suggestions and encouragement given to me. My advisor from NBIS advisory program, **Erik Fasterius**. Thanks for helping me get hands-on with data analysis. Thanks to my mentor, **Shuijie Li**.

My opponent **Fredrik Wermeling**, examination board coordinator **Adam Ameur**, examination board members **Klas Wiman** and **Carsten Daub**, and chairperson **Itziar Martinez Gonzalez**, thank you for agreeing to be an important part of my defence. I appreciate your time and work on it. I also want to thank **Edmund Loh** who has been a very responsible chairperson in reviewing my annual progress.

And of course, all the people from Kutter, Pelechano and Friedländer lab!

Jonas, thank you for being a very cool friend and an amazing coworker! Helped me set up experiments & analysis, answered my smart (rarely)/OK-ish (sometimes)/stupid (quite often) questions, fixed all kinds of things that did not work and organized fun after-work hangouts. Thanks for everything! **Christian**, an amazing and enthusiastic person, a proofreader with high-resolution/throughput eyes. Knows everything and is ready to help me and others all the time. I really enjoy being your friend and working with you! **Lara**, I am so happy to have you on our project team. Thank you for all your hard work, encouragement and caring both as a talented teammate and a great friend! **Aniek**, **Linda**, **Raül**, **Janine**, thank you for your contributions and for being great teammates working with me on the projects. I also learnt a lot from all of you, both your expertise and the spirit! **Xueli** and **Quim**, thanks for your work too. **Carlos**, our bioinformatician expert, thank you for your help and support. You are always so nice, patient and calm. A cool friend to have! **Riccardo**, I enjoyed hanging out with you a lot! And thank you for always making us tasty pizza :) Thanks to **Jente**, **Ionut**, **Siddharth**, your skills helped me a lot. **Pablo**, **William**, **Sofia**, **Natalie**, and all the rest of the current and previous KutterLab members, thanks for being part of my PhD journey!

Thanks, **Vicent**, for your helps on my project and the kind words to reduce my stress on thesis writing. **Yujie**, never too many hugs to you if I want to express how grateful I am to have you as my friend. So lucky to know you even before we came to Sweden. You helped me so much in many ways, and I

would not make it this far without your friendship! **Eva**, I always admire your ability and attitude towards science. Everyone would want to have a knowledgeable person with a warm heart around, like you! **Yerma**, my climbing instructor and collaborator, your hugs and caring bring so much warmth to my heart. You are such a sweet person just like those candies you like to eat. **Jingwen**, thank you for all the helps you gave to me, helping me settle down in Sweden, sharing nice meals with me and advising me on data analysis. **Xiushan** and **Bingnan,** thanks for helping me with my experiments. **Alisa**, thank you for your help when I set up ChIPseq (I bothered you a lot just like in your graduation video, haha), and so much fun to hang out with you. **Marcel**, always nice to have a cool and knowledgeable person like you around. **Mengjun**, **Susanne**, **Pepe**, **Mandy**, **Honglian**, **Shujing**, **Ryan**, **Donal**, **Guido** and **Lilit**, thanks for your help/suggestions/company.

Marc, thank you for feedback on my project. **Vaishnovi**, we spent so much great time together, exploring nice food, playing games, hanging out, talking and laughing. Thanks for the joy that cheered me up! **Wenjing**, thanks for all the nice hangouts! **Inna**, thank you for your help in the lab and for always lending us instruments :) **Laura**, I enjoy talking to you and hanging out with you. **Morteza**, thanks for sharing really useful info with me. Hope we all have good luck at Migration Agency! **Panos**, thanks for being a kind and cool office mate.

Thanks to **Christoffer** and **Peter** at Samplix for support in one of my projects.

Keying, thanks for letting me contribute to your cool project and all the suggestions on my own manuscript submission. **Jing**, thank you for all your helps on my experiments and for sharing your experience with me.

Jyoti and **Maria**, thanks for all the nice brunches, dinners and parties. I feel happy and relaxed with you. So good that we live so close!

ShanShan, **Jieyu**, **Junjie**, **Jingyan**, **Xiaomeng**, **Ruining**, **Dandan**, **Qing**, **Danyang**, **Dang**, **Yachao**, **Hui**, **Yangjun**, **Can** and many other friends, I am glad that I met you in Sweden. You make my life here much more colorful. If it was not you, I would feel so homesick. Thanks, **Chenggui**, our dear friend.

I would like to thank China Scholarship Council for providing financial support for my study.

I would like to convey my gratitude to **Prof. Wen Wang** at Shanghai Jiaotong University for his great help, support and encouragement.

Thanks to **Ingrid**, **Jan**, **Urika**, **Erik** and **Louise**, my warm and sweet family in Sweden.

To my **parents** and **grandparents**. Without you, there will be no “me”, biologically and personally. You give me the greatest love, trust and freedom. You showed me what selfless love and a good heart are, how to be brave and fight, …… The best parents and grandparents in the world!

And lastly, to **John**, my husband, best friend, psychiatrist, chef, lifetime teammate…….

9 References

1. Cong L, Ran FA, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science (80-)*. 2013;339(6121):819-823. doi:10.1126/science.1231143

2. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-)*. 2012;337(6096):816-821. doi:10.1126/SCIENCE.1225829/SUPPL_FILE/JINEK.SM.PDF

3. Amitai G, Sorek R. CRISPR–Cas adaptation: insights into the mechanism of action. *Nat Rev Microbiol 2016 142*. 2016;14(2):67-76. doi:10.1038/nrmicro.2015.14

4. Jiang F, Doudna JA. CRISPR–Cas9 Structures and Mechanisms. *https://doi.org/101146/annurev-biophys-062215-010822*. 2017;46:505-529. doi:10.1146/ANNUREV-BIOPHYS-062215-010822

5. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8(11):2281-2308. doi:10.1038/nprot.2013.143

6. Zalatan JG, Lee ME, Almeida R, et al. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell*. 2015;160(1-2):339-350. doi:10.1016/j.cell.2014.11.052

7. Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods 2013 1011*. 2013;10(11):1116-1121. doi:10.1038/nmeth.2681

8. Hu JH, Miller SM, Geurts MH, et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nat 2018 5567699*. 2018;556(7699):57-63. doi:10.1038/nature26155

9. Nishimasu H, Shi X, Ishiguro S, et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science (80-)*. 2018;361(6408):1259-1262. doi:10.1126/SCIENCE.AAS9129/SUPPL_FILE/PAPV2.PDF

10. Miller SM, Wang T, Randolph PB, et al. Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nat Biotechnol 2020 384*. 2020;38(4):471-481. doi:10.1038/s41587-020-0412-8

11. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*. 2014;32(4):347-350. doi:10.1038/nbt.2842

12. Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, Jaenisch R. XOne-step generation of mice carrying reporter and conditional alleles by CRISPR/cas-mediated genome engineering. *Cell*. 2013;154(6):1370. doi:10.1016/j.cell.2013.08.022

13. Maddalo D, Manchado E, Concepcion CP, et al. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature*. 2014;516(7531):423-428. doi:10.1038/nature13902

14. Canver MC, Bauer DE, Dass A, et al. Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/cas9 nuclease system in mammalian cells. *J Biol Chem*. 2014;289(31):21312-21324. doi:10.1074/jbc.M114.564625

15. Wang T, Birsoy K, Hughes NW, et al. Identification and characterization of essential genes in the human genome. *Science (80-)*. 2015;350(6264):1096-1101. doi:10.1126/SCIENCE.AAC7041/SUPPL_FILE/WANG-SM.PDF

16. Evers B, Jastrzebski K, Heijmans JPM, Grernrum W, Beijersbergen RL, Bernards R. CRISPR

knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat Biotechnol 2016 346*. 2016;34(6):631-633. doi:10.1038/nbt.3536

17. Chen S, Sanjana NE, Zheng K, et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell*. 2015;160(6):1246-1260. doi:10.1016/j.cell.2015.02.038

18. Lopes R, Korkmaz G, Agami R. Applying CRISPR–Cas9 tools to identify and characterize transcriptional enhancers. *Nat Rev Mol Cell Biol 2016 179*. 2016;17(9):597-604. doi:10.1038/nrm.2016.79

19. Rajagopal N, Srinivasan S, Kooshesh K, et al. High-throughput mapping of regulatory DNA. *Nat Biotechnol 2016 342*. 2016;34(2):167-174. doi:10.1038/nbt.3468

20. Ma H, Naseri A, Reyes-Gutierrez P, Wolfe SA, Zhang S, Pederson T. Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc Natl Acad Sci U S A*. 2015;112(10):3002-3007. doi:10.1073/PNAS.1420024112/-/DCSUPPLEMENTAL

21. Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell*. 2014;159(3):635-646. doi:10.1016/j.cell.2014.09.039

22. Hong Y, Lu G, Duan J, Liu W, Zhang Y. Comparison and optimization of CRISPR/dCas9/gRNA genome-labeling systems for live cell imaging. *Genome Biol*. 2018;19(1):1-10. doi:10.1186/S13059-018-1413-5/FIGURES/6

23. Chen B, Gilbert LA, Cimini BA, et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*. 2013;155(7):1479-1491. doi:10.1016/j.cell.2013.12.001

24. Chen B, Hu J, Almeida R, et al. Expanding the CRISPR imaging toolset with Staphylococcus aureus Cas9 for simultaneous imaging of multiple genomic loci. *Nucleic Acids Res*. 2016;44(8):e75-e75. doi:10.1093/NAR/GKV1533

25. Clow PA, Du M, Jillette N, Taghbalout A, Zhu JJ, Cheng AW. CRISPR-mediated multiplexed live cell imaging of nonrepetitive genomic loci with one guide RNA per locus. *Nat Commun 2022 131*. 2022;13(1):1-10. doi:10.1038/s41467-022-29343-z

26. Perez-Pinera P, Kocak DD, Vockley CM, et al. RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nat Methods 2013 1010*. 2013;10(10):973-976. doi:10.1038/nmeth.2600

27. Konermann S, Brigham MD, Trevino AE, et al. Optical control of mammalian endogenous transcription and epigenetic states. *Nat 2013 5007463*. 2013;500(7463):472-476. doi:10.1038/nature12466

28. Gilbert LA, Larson MH, Morsut L, et al. XCRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*. 2013;154(2):442. doi:10.1016/j.cell.2013.06.044

29. Hilton IB, D'Ippolito AM, Vockley CM, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol 2015 335*. 2015;33(5):510-517. doi:10.1038/nbt.3199

30. Brocken DJW, Tark-Dame M, Dame RT. dCas9: A Versatile Tool for Epigenome Editing. *Curr Issues Mol Biol 2018, Vol 26, Pages 15-32*. 2017;26(1):15-32. doi:10.21775/CIMB.026.015

31. Wong N, Liu W, Wang X. WU-CRISPR: Characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol*. 2015;16(1):1-8. doi:10.1186/S13059-015-0784-0/FIGURES/4

32. Doench JG, Hartenian E, Graham DB, et al. Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nat Biotechnol 2014 3212*. 2014;32(12):1262-1267. doi:10.1038/nbt.3026

33. Konstantakos V, Nentidis A, Krithara A, Paliouras G. CRISPR–Cas9 gRNA efficiency prediction: an overview of predictive tools and the role of deep learning. *Nucleic Acids Res*. 2022;50(7):3616-3637. doi:10.1093/NAR/GKAC192

34. Chari R, Mali P, Moosburner M, Church GM. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat Methods 2015 129*. 2015;12(9):823-826. doi:10.1038/nmeth.3473

35. Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic screens in human cells using the CRISPR-Cas9 system. *Science (80-)*. 2014;343(6166):80-84. doi:10.1126/SCIENCE.1246981/SUPPL_FILE/WANG.SM.PDF

36. Moreno-Mateos MA, Vejnar CE, Beaudoin JD, et al. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat Methods 2015 1210*. 2015;12(10):982-988. doi:10.1038/nmeth.3543

37. Labuhn M, Adams FF, Ng M, et al. Refined sgRNA efficacy prediction improves large- and small-scale CRISPR–Cas9 applications. *Nucleic Acids Res*. 2018;46(3):1375-1385. doi:10.1093/NAR/GKX1268

38. Fujita T, Yuno M, Fujii H. Allele-specific locus binding and genome editing by CRISPR at the p16INK4a locus. *Sci Reports 2016 61*. 2016;6(1):1-11. doi:10.1038/srep30485

39. Kallimasioti-Pazi EM, Thelakkad Chathoth K, Taylor GC, et al. Heterochromatin delays CRISPR-Cas9 mutagenesis but does not influence the outcome of mutagenic DNA repair. *PLOS Biol*. 2018;16(12):e2005595. doi:10.1371/JOURNAL.PBIO.2005595

40. Verkuijl SA, Rots MG. The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies. *Curr Opin Biotechnol*. 2019;55:68-73. doi:10.1016/J.COPBIO.2018.07.005

41. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol 2014 323*. 2014;32(3):279-284. doi:10.1038/nbt.2808

42. Fu Y, Foden JA, Khayter C, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol 2013 319*. 2013;31(9):822-826. doi:10.1038/nbt.2623

43. Hsu PD, Scott DA, Weinstein JA, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol 2013 319*. 2013;31(9):827-832. doi:10.1038/nbt.2647

44. Kuscu C, Arslan S, Singh R, Thorpe J, Adli M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol 2014 327*. 2014;32(7):677-683. doi:10.1038/nbt.2916

45. Wu X, Scott DA, Kriz AJ, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol 2014 327*. 2014;32(7):670-676. doi:10.1038/nbt.2889

46. Tsai SQ, Zheng Z, Nguyen NT, et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol 2014 332*. 2014;33(2):187-197. doi:10.1038/nbt.3117

47. Frock RL, Hu J, Meyers RM, Ho YJ, Kii E, Alt FW. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat Biotechnol 2014 332*. 2014;33(2):179-

186. doi:10.1038/nbt.3101

48. Kim D, Bae S, Park J, et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods 2015 123*. 2015;12(3):237-243. doi:10.1038/nmeth.3284

49. Yan WX, Mirzazadeh R, Garnerone S, et al. BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat Commun 2017 81*. 2017;8(1):1-9. doi:10.1038/ncomms15058

50. Tsai SQ, Joung JK. Defining and improving the genome-wide specificities of CRISPR–Cas9 nucleases. *Nat Rev Genet 2016 175*. 2016;17(5):300-312. doi:10.1038/nrg.2016.28

51. Rose JC, Popp NA, Richardson CD, et al. Suppression of unwanted CRISPR-Cas9 editing by co-administration of catalytically inactivating truncated guide RNAs. *Nat Commun 2020 111*. 2020;11(1):1-11. doi:10.1038/s41467-020-16542-9

52. Ran FA, Hsu PD, Lin CY, et al. XDouble nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell*. 2013;154(6):1380-1389. doi:10.1016/j.cell.2013.08.021

53. Yang Y, Wang L, Bell P, et al. A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nat Biotechnol 2016 343*. 2016;34(3):334-338. doi:10.1038/nbt.3469

54. Kotterman MA, Chalberg TW, Schaffer D V. Viral Vectors for Gene Therapy: Translational and Clinical Outlook. *https://doi.org/101146/annurev-bioeng-071813-104938*. 2015;17:63-89. doi:10.1146/ANNUREV-BIOENG-071813-104938

55. Popescu NC, Zimonjic D, DiPaolo JA. Viral integration, fragile sites, and proto-oncogenes in human neoplasia. *Hum Genet*. 1990;84(5):383-386. doi:10.1007/BF00195804

56. Muruve DA. The innate immune response to adenovirus vectors. *Hum Gene Ther*. 2004;15(12):1157-1166. doi:10.1089/HUM.2004.15.1157

57. Yip BH. Recent Advances in CRISPR/Cas9 Delivery Strategies. *Biomolecules*. 2020;10(6). doi:10.3390/BIOM10060839

58. Søndergaard JN, Geng K, Sommerauer C, Atanasoai I, Yin X, Kutter C. Successful delivery of large-size CRISPR/Cas9 vectors in hard-to-transfect human cells using small plasmids. *Commun Biol*. 2020;3(1):1-6. doi:10.1038/s42003-020-1045-7

59. Essletzbichler P, Konopka T, Santoro F, et al. Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. *Genome Res*. 2014;24(12):2059. doi:10.1101/GR.177220.114

60. Sun S, Zhao Y, Shuai L. The milestone of genetic screening: Mammalian haploid cells. *Comput Struct Biotechnol J*. 2020;18:2471-2479. doi:10.1016/J.CSBJ.2020.09.006

61. Geng K, Merino LG, Wedemann L, et al. Target-enriched nanopore sequencing and de novo assembly reveals cooccurrences of complex on-target genomic rearrangements induced by CRISPR-Cas9 in human cells. *Genome Res*. September 2022:gr.276901.122. doi:10.1101/GR.276901.122

62. Kotecki M, Reddy PS, Cochran BH. Isolation and characterization of a near-haploid human cell line. *Exp Cell Res*. 1999;252(2):273-280. doi:10.1006/EXCR.1999.4656

63. Bürckstümmer T, Banning C, Hainzl P, et al. A reversible gene trap collection empowers haploid genetics in human cells. *Nat Methods*. 2013;10(10):965. doi:10.1038/NMETH.2609

64. Carette JE, Raaben M, Wong AC, et al. Ebola virus entry requires the cholesterol transporter Niemann–Pick C1. *Nat 2011 4777364*. 2011;477(7364):340-343. doi:10.1038/nature10348

65. Beigl TB, Kjosås I, Seljeseth E, Glomnes N, Aksnes H. Efficient and crucial quality control of HAP1 cell ploidy status. *Biol Open*. 2020;9(11). doi:10.1242/BIO.057174/225361

66. Olbrich T, Mayor-Ruiz C, Vega-Sendino M, et al. A p53-dependent response limits the viability of mammalian haploid cells. *Proc Natl Acad Sci U S A*. 2017;114(35):9367-9372. www.pnas.org/cgi/doi/10.1073/pnas.1705133114. Accessed May 15, 2021.

67. Bakkenist CJ, Kastan MB. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nat 2003 4216922*. 2003;421(6922):499-506. doi:10.1038/nature01368

68. Falck J, Petrini JHJ, Williams BR, Lukas J, Bartek J. The DNA damage-dependent intra-S phase checkpoint is regulated by parallel pathways. *Nat Genet*. 2002;30(3):290-294. doi:10.1038/NG845

69. Tibbetts RS, Cortez D, Brumbaugh KM, et al. Functional interactions between BRCA1 and the checkpoint kinase ATR during genotoxic stress. *Genes Dev*. 2000;14(23):2989-3002. doi:10.1101/GAD.851000

70. Rogakou EP, Pilch DR, Orr AH, Ivanova VS, Bonner WM. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem*. 1998;273(10):5858-5868. doi:10.1074/JBC.273.10.5858

71. Bekker-Jensen S, Mailand N. Assembly and function of DNA double-strand break repair foci in mammalian cells. *DNA Repair (Amst)*. 2010;9(12):1219-1228. doi:10.1016/J.DNAREP.2010.09.010

72. Arnould C, Rocher V, Finoux AL, et al. Loop extrusion as a mechanism for formation of DNA damage repair foci. *Nat 2021 5907847*. 2021;590(7847):660-665. doi:10.1038/s41586-021-03193-z

73. Spagnolo L, Rivera-Calzada A, Pearl LH, Llorca O. Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA DSB repair. *Mol Cell*. 2006;22(4):511-519. doi:10.1016/J.MOLCEL.2006.04.013

74. Mari PO, Florea BI, Persengiev SP, et al. Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. *Proc Natl Acad Sci U S A*. 2006;103(49):18597-18602. doi:10.1073/PNAS.0609061103/SUPPL_FILE/IMAGE530.GIF

75. Andres SN, Modesti M, Tsai CJ, Chu G, Junop MS. Crystal structure of human XLF: a twist in nonhomologous DNA end-joining. *Mol Cell*. 2007;28(6):1093-1101. doi:10.1016/J.MOLCEL.2007.10.024

76. Ma Y, Pannicke U, Schwarz K, Lieber MR. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell*. 2002;108(6):781-794. doi:10.1016/S0092-8674(02)00671-2

77. Wang H, Shi LZ, Wong CCL, et al. The interaction of CtIP and Nbs1 connects CDK and ATM to regulate HR-mediated double-strand break repair. *PLoS Genet*. 2013;9(2). doi:10.1371/JOURNAL.PGEN.1003277

78. Ding Q, Reddy YVR, Wang W, et al. Autophosphorylation of the catalytic subunit of the DNA-dependent protein kinase is required for efficient end processing during DNA double-strand break repair. *Mol Cell Biol*. 2003;23(16):5836-5848. doi:10.1128/MCB.23.16.5836-5848.2003

79. Dobbs TA, Tainer JA, Lees-Miller SP. A structural model for regulation of NHEJ by DNA-PKcs autophosphorylation. *DNA Repair (Amst)*. 2010;9(12):1307-1314. doi:10.1016/J.DNAREP.2010.09.019

80. Bernstein NK, Williams RS, Rakovszky ML, et al. The molecular architecture of the mammalian DNA repair enzyme, polynucleotide kinase. *Mol Cell*. 2005;17(5):657-670. doi:10.1016/J.MOLCEL.2005.02.012

81. Niewolik D, Peter I, Butscher C, Schwarz K. Autoinhibition of the Nuclease ARTEMIS Is Mediated by a Physical Interaction between Its Catalytic and C-terminal Domains. *J Biol Chem*. 2017;292(8):3351-3365. doi:10.1074/JBC.M116.770461

82. Niewolik D, Pannicke U, Lu H, et al. DNA-PKcs dependence of Artemis endonucleolytic activity, differences between hairpins and 5' or 3' overhangs. *J Biol Chem*. 2006;281(45):33900-33909. doi:10.1074/JBC.M606023200

83. Yun MH, Hiom K. CtIP-BRCA1 modulates the choice of DNA double-strand-break repair pathway throughout the cell cycle. *Nat 2009 4597245*. 2009;459(7245):460-463. doi:10.1038/nature07955

84. Anand R, Ranjha L, Cannavo E, Cejka P. Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection. *Mol Cell*. 2016;64(5):940-950. doi:10.1016/J.MOLCEL.2016.10.017

85. Sartori AA, Lukas C, Coates J, et al. Human CtIP promotes DNA end resection. *Nat 2007 4507169*. 2007;450(7169):509-514. doi:10.1038/nature06337

86. Nimonkar A V., Genschel J, Kinoshita E, et al. BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. *Genes Dev*. 2011;25(4):350-362. doi:10.1101/GAD.2003811

87. Sturzenegger A, Burdova K, Kanagaraj R, et al. DNA2 cooperates with the WRN and BLM RecQ helicases to mediate long-range DNA end resection in human cells. *J Biol Chem*. 2014;289(39):27314-27326. doi:10.1074/jbc.M114.578823

88. Chen H, Lisby M, Symington LS. RPA coordinates DNA end resection and prevents formation of DNA hairpins. *Mol Cell*. 2013;50(4):589-600. doi:10.1016/J.MOLCEL.2013.04.032

89. Liu S, Hua Y, Wang J, et al. RNA polymerase III is required for the repair of DNA double-strand breaks by homologous recombination. *Cell*. 2021;184(5):1314-1329.e10. doi:10.1016/j.cell.2021.01.048

90. Xu J, Zhao L, Xu Y, Zhao W, Sung P, Wang HW. Cryo-EM structures of human RAD51 recombinase filaments during catalysis of DNA-strand exchange. *Nat Struct Mol Biol*. 2017;24(1):40-46. doi:10.1038/NSMB.3336

91. Shrivastav M, De Haro LP, Nickoloff JA. Regulation of DNA double-strand break repair pathway choice. *Cell Res*. 2008;18(1):134-147. doi:10.1038/CR.2007.111

92. Zhao W, Steinfeld JB, Liang F, et al. BRCA1-BARD1 promotes RAD51-mediated homologous DNA pairing. *Nature*. 2017;550(7676). doi:10.1038/NATURE24060

93. Crickard JB, Moevus CJ, Kwon Y, Sung P, Greene EC. Rad54 Drives ATP Hydrolysis-Dependent DNA Sequence Alignment during Homologous Recombination. *Cell*. 2020;181(6):1380-1394.e18. doi:10.1016/J.CELL.2020.04.056

94. Kowalczykowski SC. An Overview of the Molecular Mechanisms of Recombinational DNA Repair. *Cold Spring Harb Perspect Biol*. 2015;7(11):a016410.

doi:10.1101/CSHPERSPECT.A016410

95. Scully R, Panday A, Elango R, Willis NA. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat Rev Mol Cell Biol 2019 2011*. 2019;20(11):698-714. doi:10.1038/s41580-019-0152-0

96. Sung P, Klein H. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nat Rev Mol Cell Biol 2006 710*. 2006;7(10):739-750. doi:10.1038/nrm2008

97. Seol JH, Shim EY, Lee SE. Microhomology-mediated end joining: Good, bad and ugly. *Mutat Res Mol Mech Mutagen*. 2018;809:81-87. doi:10.1016/J.MRFMMM.2017.07.002

98. Lee JH, Paull TT. Direct activation of the ATM protein kinase by the Mre11/Rad50/Nbs1 complex. *Science*. 2004;304(5667):93-96. doi:10.1126/SCIENCE.1091496

99. Truong LN, Li Y, Shi LZ, et al. Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proc Natl Acad Sci U S A*. 2013;110(19):7720-7725. doi:10.1073/PNAS.1213431110/SUPPL_FILE/PNAS.201213431SI.PDF

100. Paull TT, Gellert M. The 3' to 5' exonuclease activity of Mre11 facilitates repair of DNA double-strand breaks. *Mol Cell*. 1998;1(7):969-979. doi:10.1016/S1097-2765(00)80097-0

101. Deng SK, Gibb B, De Almeida MJ, Greene EC, Symington LS. RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. *Nat Struct Mol Biol 2014 214*. 2014;21(4):405-412. doi:10.1038/nsmb.2786

102. Audebert M, Salles B, Calsou P. Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. *J Biol Chem*. 2004;279(53):55117-55126. doi:10.1074/jbc.M404524200

103. Wang M, Wu W, Wu W, et al. PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic Acids Res*. 2006;34(21):6170-6182. doi:10.1093/NAR/GKL840

104. Caron MC, Sharma AK, O'Sullivan J, et al. Poly(ADP-ribose) polymerase-1 antagonizes DNA resection at double-strand breaks. *Nat Commun 2019 101*. 2019;10(1):1-16. doi:10.1038/s41467-019-10741-9

105. Ahmad A, Robinson AR, Duensing A, et al. ERCC1-XPF Endonuclease Facilitates DNA Double-Strand Break Repair. *https://doi.org/101128/MCB00293-08*. 2023;28(16):5082-5092. doi:10.1128/MCB.00293-08

106. Liang L, Deng L, Nguyen SC, et al. Human DNA ligases I and III, but not ligase IV, are required for microhomology-mediated end joining of DNA double-strand breaks. *Nucleic Acids Res*. 2008;36(10):3297-3310. doi:10.1093/NAR/GKN184

107. Shen MW, Arbab M, Hsu JY, et al. Predictable and precise template-free CRISPR editing of pathogenic variants. *Nat 2018 5637733*. 2018;563(7733):646-651. doi:10.1038/s41586-018-0686-x

108. Shou J, Li J, Liu Y, Wu Q. Precise and Predictable CRISPR Chromosomal Rearrangements Reveal Principles of Cas9-Mediated Nucleotide Insertion. *Mol Cell*. 2018;71(4):498-509.e4. doi:10.1016/J.MOLCEL.2018.06.021

109. Brinkman EK, Chen T, de Haas M, Holland HA, Akhtar W, van Steensel B. Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks. *Mol Cell*. 2018;70(5):801-813.e6.

doi:10.1016/j.molcel.2018.04.016

110. Clarke R, Heler R, MacDougall MS, et al. Enhanced Bacterial Immunity and Mammalian Genome Editing via RNA-Polymerase-Mediated Dislodging of Cas9 from Double-Strand DNA Breaks. *Mol Cell*. 2018;71(1):42-55.e8. doi:10.1016/j.molcel.2018.06.005

111. Ihry RJ, Worringer KA, Salick MR, et al. p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nat Med 2018 247*. 2018;24(7):939-946. doi:10.1038/s41591-018-0050-6

112. Haapaniemi E, Botla S, Persson J, Schmierer B, Taipale J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat Med 2018 247*. 2018;24(7):927-930. doi:10.1038/s41591-018-0049-z

113. Jiang L, Ingelshed K, Shen Y, et al. CRISPR/Cas9-Induced DNA Damage Enriches for Mutations in a p53-Linked Interactome: Implications for CRISPR-Based Therapies. *Cancer Res*. 2021;82(1):36-45. doi:10.1158/0008-5472.CAN-21-1692/670650/AM/CRISPR-CAS9-INDUCED-DNA-DAMAGE-ENRICHES-FOR

114. Sinha S, Barbosa K, Cheng K, et al. A systematic genome-wide mapping of oncogenic mutation selection during CRISPR-Cas9 genome editing. *Nat Commun 2021 121*. 2021;12(1):1-13. doi:10.1038/s41467-021-26788-6

115. Enache OM, Rendo V, Abdusamad M, et al. Cas9 activates the p53 pathway and selects for p53-inactivating mutations. *Nat Genet*. 2020. doi:10.1038/s41588-020-0623-4

116. Álvarez MM, Biayna J, Supek F. TP53-dependent toxicity of CRISPR/Cas9 cuts is differential across genomic loci and can confound genetic screening. *Nat Commun 2022 131*. 2022;13(1):1-14. doi:10.1038/s41467-022-32285-1

117. Nahmad AD, Reuveni E, Goldschmidt E, et al. Frequent aneuploidy in primary human T cells after CRISPR–Cas9 cleavage. *Nat Biotechnol 2022 4012*. 2022;40(12):1807-1813. doi:10.1038/s41587-022-01377-0

118. Zuccaro M V., Xu J, Mitchell C, et al. Allele-Specific Chromosome Removal after Cas9 Cleavage in Human Embryos. *Cell*. 2020;183(6):1650-1664.e15. doi:10.1016/J.CELL.2020.10.025

119. Leibowitz ML, Papathanasiou S, Doerfler PA, et al. Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nat Genet 2021 536*. 2021;53(6):895-905. doi:10.1038/s41588-021-00838-7

120. Liu M, Zhang W, Xin C, et al. Global detection of DNA repair outcomes induced by CRISPR–Cas9. *Nucleic Acids Res*. 2021;49(15):8732-8742. doi:10.1093/NAR/GKAB686

121. Yin J, Lu R, Xin C, et al. Cas9 exo-endonuclease eliminates chromosomal translocations during genome editing. *Nat Commun 2022 131*. 2022;13(1):1-14. doi:10.1038/s41467-022-28900-w

122. Wu J, Zou Z, Liu Y, et al. CRISPR/Cas9-induced structural variations expand in T lymphocytes in vivo. *Nucleic Acids Res*. 2022;50(19):11128-11137. doi:10.1093/NAR/GKAC887

123. Turchiano G, Andrieux G, Klermund J, et al. Quantitative evaluation of chromosomal rearrangements in gene-edited human stem cells by CAST-Seq. *Cell Stem Cell*. 2021;28(6):1136-1147.e5. doi:10.1016/J.STEM.2021.02.002

124. Lee H, Kim JS. Unexpected CRISPR on-target effects. *Nat Biotechnol 2018 368*. 2018;36(8):703-704. doi:10.1038/nbt.4207

125. Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR–Cas9

leads to large deletions and complex rearrangements. *Nat Biotechnol*. 2018. doi:10.1038/nbt.4192

126. Shin HY, Wang C, Lee HK, et al. CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nat Commun 2017 81*. 2017;8(1):1-10. doi:10.1038/ncomms15464

127. Höijer I, Emmanouilidou A, Östlund R, et al. CRISPR-Cas9 induces large structural variants at on-target and off-target sites in vivo that segregate across generations. *Nat Commun 2022 131*. 2022;13(1):1-10. doi:10.1038/s41467-022-28244-5

128. Owens DDG, Caulder A, Frontera V, et al. Microhomologies are prevalent at Cas9-induced larger deletions. *Nucleic Acids Res*. 2019;47(14):7402-7417. doi:10.1093/NAR/GKZ459

129. Blondal T, Gamba C, Møller Jagd L, et al. Verification of CRISPR editing and finding transgenic inserts by Xdrop indirect sequence capture followed by short- and long-read sequencing. *Methods*. February 2021. doi:10.1016/j.ymeth.2021.02.003

130. Bosch-Guiteras N, Uroda T, Guillen-Ramirez HA, et al. Enhancing CRISPR deletion via pharmacological delay of DNA-PKcs. *Genome Res*. 2021;31(3):461-471. doi:10.1101/GR.265736.120

131. Kraft K, Geuer S, Will AJ, et al. Deletions, inversions, duplications: Engineering of structural variants using CRISPR/Cas in mice. *Cell Rep*. 2015;10(5):833-839. doi:10.1016/j.celrep.2015.01.016

132. Li J, Shou J, Guo Y, et al. Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. *J Mol Cell Biol*. 2015;7(4):284-298. doi:10.1093/jmcb/mjv016

133. Shou J, Li J, Liu Y, Wu Q. Precise and Predictable CRISPR Chromosomal Rearrangements Reveal Principles of Cas9-Mediated Nucleotide Insertion. *Mol Cell*. 2018;71(4):498-509.e4. doi:10.1016/j.molcel.2018.06.021

134. Møller HD, Lin L, Xiang X, et al. CRISPR-C: circularization of genes and chromosome by CRISPR in human cells. *Nucleic Acids Res*. 2018;46(22):e131-e131. doi:10.1093/NAR/GKY767

135. Lazzarotto CR, Nguyen NT, Tang X, et al. Defining CRISPR–Cas9 genome-wide nuclease activities with CIRCLE-seq. *Nat Protoc 2018 1311*. 2018;13(11):2615-2642. doi:10.1038/s41596-018-0055-0

136. Cameron P, Fuller CK, Donohoue PD, et al. Mapping the genomic landscape of CRISPR–Cas9 cleavage. *Nat Methods 2017 146*. 2017;14(6):600-606. doi:10.1038/nmeth.4284

137. Crosetto N, Mitra A, Silva MJ, et al. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat Methods 2013 104*. 2013;10(4):361-365. doi:10.1038/nmeth.2408

138. Wang X, Wang Y, Wu X, et al. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat Biotechnol*. 2015;33(2):175-179. doi:10.1038/NBT.3127

139. Hu J, Meyers RM, Dong J, Panchakshari RA, Alt FW, Frock RL. Detecting DNA double-stranded breaks in mammalian genomes by linear amplification–mediated high-throughput genome-wide translocation sequencing. *Nat Protoc 2016 115*. 2016;11(5):853-871. doi:10.1038/nprot.2016.043

140. Yin J, Liu M, Liu Y, et al. Optimizing genome editing strategy by primer-extension-mediated

sequencing. *Cell Discov 2019 51*. 2019;5(1):1-11. doi:10.1038/s41421-019-0088-8

141. Lackner M, Helmbrecht N, Pääbo S, Riesenberg S. Detection of unintended on-target effects in CRISPR genome editing by DNA donors carrying diagnostic substitutions. *Nucleic Acids Res*. 2023;51(5):e26-e26. doi:10.1093/NAR/GKAC1254

142. Kim D, Luk K, Wolfe SA, Kim JS. Evaluating and Enhancing Target Specificity of Gene-Editing Nucleases and Deaminases. *https://doi.org/101146/annurev-biochem-013118-111730*. 2019;88:191-220. doi:10.1146/ANNUREV-BIOCHEM-013118-111730

143. Höijer I, Johansson J, Gudmundsson S, et al. Amplification-free long-read sequencing reveals unforeseen CRISPR-Cas9 off-target activity. *Genome Biol*. 2020;21(1):1-19. doi:10.1186/S13059-020-02206-W/FIGURES/6

144. Madsen EB, Höijer I, Kvist T, Ameur A, Mikkelsen MJ. Xdrop: Targeted sequencing of long DNA molecules from low input samples using droplet sorting. *Hum Mutat*. 2020;41(9):1671-1679. doi:10.1002/humu.24063

145. Blow JJ, Hodgson B. Replication Licensing – Defining the Proliferative State? *Trends Cell Biol*. 2002;12(2):72. doi:10.1016/S0962-8924(01)02203-6

146. Ibarra A, Schwob E, Mé J. Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. *PNAS*. 2008;105(26):8956-8961.

147. Shima N, Alcaraz A, Liachko I, et al. A viable allele of Mcm4 causes chromosome instability and mammary adenocarcinomas in mice. *Nat Genet 2006 391*. 2006;39(1):93-98. doi:10.1038/ng1936

148. Masai H, Matsumoto S, You Z, Yoshizawa-Sugata N, Oda M. Eukaryotic Chromosome DNA Replication: Where, When, and How? *https://doi.org/101146/annurev.biochem052308103205*. 2010;79:89-130. doi:10.1146/ANNUREV.BIOCHEM.052308.103205

149. Abbas T, Keaton MA, Dutta A. Genomic instability in cancer. *Cold Spring Harb Perspect Biol*. 2013;5(3). doi:10.1101/cshperspect.a012914

150. Lin JJ, Milhollen MA, Smith PG, Narayanan U, Dutta A. NEDD8-Targeting Drug MLN4924 Elicits DNA Rereplication by Stabilizing Cdt1 in S Phase, Triggering Checkpoint Activation, Apoptosis, and Senescence in Cancer Cells. *Cancer Res*. 2010;70(24):10310-10320. doi:10.1158/0008-5472.CAN-10-2062

151. Vaziri C, Saxena S, Jeon Y, et al. A p53-Dependent Checkpoint Pathway Prevents Rereplication. *Mol Cell*. 2003;11(4):997-1008. doi:10.1016/S1097-2765(03)00099-6

152. Dominguez-Sola D, Ying CY, Grandori C, et al. Non-transcriptional control of DNA replication by c-Myc. *Nat 2007 4487152*. 2007;448(7152):445-451. doi:10.1038/nature05953

153. Fu H, Redon CE, Thakur BL, et al. Dynamics of replication origin over-activation. *Nat Commun 2021 121*. 2021;12(1):1-15. doi:10.1038/s41467-021-23835-0

154. Seigneur M, Bidnenko V, Ehrlich SD, Michel B. RuvAB acts at arrested replication forks. *Cell*. 1998;95(3):419-430. doi:10.1016/S0092-8674(00)81772-9

155. Sogo JM, Lopes M, Foiani M. Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science*. 2002;297(5581):599-602. doi:10.1126/SCIENCE.1074023

156. Cobb JA, Bjergbaek L, Shimada K, Frei C, Gasser SM. DNA polymerase stabilization at stalled

replication forks requires Mec1 and the RecQ helicase Sgs1. *EMBO J*. 2003;22(16):4325-4336. doi:10.1093/EMBOJ/CDG391

157. Wang Y, Putnam CD, Kane MF, et al. Mutation in Rpa1 results in defective DNA double-strand break repair, chromosomal instability and cancer in mice. *Nat Genet*. 2005;37(7):750-755. doi:10.1038/NG1587

158. Moriel-Carretero M, Aguilera A. A Postincision-Deficient TFIIH Causes Replication Fork Breakage and Uncovers Alternative Rad51- or Pol32-Mediated Restart Mechanisms. *Mol Cell*. 2010;37(5):690-701. doi:10.1016/j.molcel.2010.02.008

159. K M, C S, AC T, et al. Global screening and extended nomenclature for 230 aphidicolin-inducible fragile sites, including 61 yet unreported ones. *Int J Oncol*. 2010;36(4). doi:10.3892/IJO_00000572

160. Sutherland GR, Baker E, Richards RI. Fragile sites still breaking. *Trends Genet*. 1998;14(12):501-506. http://www.cell.com/article/S016895259801628X/fulltext. Accessed April 24, 2023.

161. Glover TW, Wilson TE, Arlt MF. Fragile sites in cancer: more than meets the eye. *Nat Rev Cancer 2017 178*. 2017;17(8):489-501. doi:10.1038/nrc.2017.52

162. Le Beau MM, Rassool F V., Neilly ME, et al. Replication of a common fragile site, FRA3B, occurs late in S phase and is delayed further upon induction: implications for the mechanism of fragile site induction. *Hum Mol Genet*. 1998;7(4):755-761. doi:10.1093/HMG/7.4.755

163. Ohta M, Inoue H, Cotticelli MG, et al. The FHIT gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers. *Cell*. 1996;84(4):587-597. doi:10.1016/S0092-8674(00)81034-X

164. LeTallec B, Millot GA, Blin ME, Brison O, Dutrillaux B, Debatisse M. Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. *Cell Rep*. 2013;4(3):420-428. doi:10.1016/j.celrep.2013.07.003

165. Le Tallec B, Dutrillaux B, Lachages AM, Millot GA, Brison O, Debatisse M. Molecular profiling of common fragile sites in human fibroblasts. *Nat Struct Mol Biol 2011 1812*. 2011;18(12):1421-1423. doi:10.1038/nsmb.2155

166. Kaushal S, Freudenreich CH. The role of fork stalling and DNA structures in causing chromosome fragility. *Genes Chromosomes Cancer*. 2019;58(5):270-283. doi:10.1002/GCC.22721

167. Usdin K, House NCM, Freudenreich CH. Repeat instability during DNA repair: Insights from model systems. *Crit Rev Biochem Mol Biol*. 2015;50(2):142-167. doi:10.3109/10409238.2014.999192

168. Miret JJ, Pessoa-Brandão L, Lahue RS. Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*. 1998;95(21):12438-12443. doi:10.1073/PNAS.95.21.12438/ASSET/2EDE1EB0-1168-48C4-955E-6B817ABBC4C5/ASSETS/GRAPHIC/PQ2183239004.JPEG

169. Lahiri M, Gustafson TL, Majors ER, Freudenreich CH. Expanded CAG repeats activate the DNA damage checkpoint pathway. *Mol Cell*. 2004;15(2):287-293. doi:10.1016/J.MOLCEL.2004.06.034

170. Lin Y, Dent SYR, Wilson JH, Wells RD, Napierala M. R loops stimulate genetic instability of

CTG.CAG repeats. *Proc Natl Acad Sci U S A*. 2010;107(2):692-697. doi:10.1073/PNAS.0909740107

171. Tubbs A, Sridharan S, van Wietmarschen N, et al. Dual Roles of Poly(dA:dT) Tracts in Replication Initiation and Fork Collapse. *Cell*. 2018;174(5):1127-1142.e19. doi:10.1016/J.CELL.2018.07.011

172. Vasquez KM, Wang G. The yin and yang of repair mechanisms in DNA structure-induced genetic instability. *Mutat Res*. 2013;743-744:118-131. doi:10.1016/J.MRFMMM.2012.11.005

173. Mirkin E V., Mirkin SM. Replication Fork Stalling at Natural Impediments. *Microbiol Mol Biol Rev*. 2007;71(1):13. doi:10.1128/MMBR.00030-06

174. Brinton BT, Caddle MS, Heintz NH. Position and orientation-dependent effects of a eukaryotic Z-triplex DNA motif on episomal DNA replication in COS-7 cells. *J Biol Chem*. 1991;266(8):5153-5161. doi:10.1016/s0021-9258(19)67768-9

175. Wang G, Christensen LA, Vasquez KM. Z-DNA-forming sequences generate large-scale deletions in mammalian cells. *Proc Natl Acad Sci U S A*. 2006;103(8):2677-2682. doi:10.1073/PNAS.0511084103/SUPPL_FILE/11084FIG7.PDF

176. Rimokh R, Rouault J -P, Wahbi K, et al. A chromosome 12 coding region is juxtaposed to the MYC protooncogene locus in a t(8;12)(q24;q22) translocation in a case of B-cell chronic lymphocytic leukemia. *Genes Chromosomes Cancer*. 1991;3(1):24-36. doi:10.1002/GCC.2870030106

177. Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev*. 2004;18(13):1618-1629. doi:10.1101/GAD.1200804

178. Nambiar M, Srivastava M, Gopalakrishnan V, Sankaran SK, Raghavan SC. G-Quadruplex Structures Formed at the HOX11 Breakpoint Region Contribute to Its Fragility during t(10;14) Translocation in T-Cell Leukemia. *Mol Cell Biol*. 2013;33(21):4266. doi:10.1128/MCB.00540-13

179. Katapadi VK, Nambiar M, Raghavan SC. Potential G-quadruplex formation at breakpoint regions of chromosomal translocations in cancer may explain their fragility. *Genomics*. 2012;100(2):72-80. doi:10.1016/J.YGENO.2012.05.008

180. Bosco N, De Lange T. A TRF1-controlled common fragile site containing interstitial telomeric sequences. *Chromosoma*. 2012;121(5):465-474. doi:10.1007/S00412-012-0377-6

181. Sfeir A, Kosiyatrakul ST, Hockemeyer D, et al. Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell*. 2009;138(1):90-103. doi:10.1016/J.CELL.2009.06.021

182. White RJ. Transcription by RNA polymerase III: More complex than we thought. *Nat Rev Genet*. 2011;12(7):459-463. doi:10.1038/nrg3001

183. White RJ. RNA polymerases I and III, growth control and cancer. *Nat Rev Mol Cell Biol*. 2005;6(1):69-78. doi:10.1038/nrm1551

184. Chan PP, Lowe TM. GtRNAdb: A database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res*. 2009;37(SUPPL. 1):D93-D97. doi:10.1093/nar/gkn787

185. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA genes in genomic sequences. In: *Methods in Molecular Biology*. ; 2019. doi:10.1007/978-1-4939-9173-0_1

186. Chan PP, Lowe TM. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in

complete and draft genomes. *Nucleic Acids Res*. 2016;44(D1):D184-D189. doi:10.1093/NAR/GKV1309

187. Kutter C, Brown GD, Gonçalves Å, et al. Pol III binding in six mammals shows conservation among amino acid isotypes despite divergence among tRNA genes. *Nat Genet*. 2011;43(10):948-957. doi:10.1038/ng.906

188. Moqtaderi Z, Wang J, Raha D, et al. Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. 2010;17(5):635-640. https://pubmed.ncbi.nlm.nih.gov/20418883/. Accessed January 24, 2022.

189. Oler AJ, Alla RK, Roberts DN, et al. Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. 2010;17(5):620-628. https://pubmed.ncbi.nlm.nih.gov/20418882/. Accessed January 24, 2022.

190. Schmitt BM, Rudolph KLM, Karagianni P, et al. High-resolution mapping of transcriptional dynamics across tissue development reveals a stable mRNA-tRNA interface. *Genome Res*. 2014;24(11):1797-1807. doi:10.1101/gr.176784.114

191. Rudolph KLM, Schmitt BM, Villar D, et al. Codon-Driven Translational Efficiency Is Stable across Diverse Mammalian Cell States. *PLoS Genet*. 2016;12(5). doi:10.1371/journal.pgen.1006024

192. Bloom-Ackermann Z, Navon S, Gingold H, Towers R, Pilpel Y, Dahan O. A Comprehensive tRNA Deletion Library Unravels the Genetic Architecture of the tRNA Pool. *PLoS Genet*. 2014. doi:10.1371/journal.pgen.1004084

193. Aharon-Hefetz N, Frumkin I, Mayshar Y, Dahan O, Pilpel Y, Rak R. Manipulation of the human trna pool reveals distinct trna sets that act in cellular proliferation or cell cycle arrest. *Elife*. 2020;9:1-28. doi:10.7554/ELIFE.58461

194. Sriskanthadevan-Pirahas S, Deshpande R, Lee B, Grewal SS. Ras/ERK-signalling promotes tRNA synthesis and growth via the RNA polymerase III repressor Maf1 in Drosophila. *PLoS Genet*. 2018. doi:10.1371/journal.pgen.1007202

195. Gomez-Roman N, Grandori C, Eisenman RN, White RJ. Direct activation of RNA polymerase III transcription by c-Myc. *Nature*. 2003;421(6920):290-294. doi:10.1038/nature01327

196. Kantidakis T, Ramsbottom BA, Birch JL, Dowding SN, White RJ. mTOR associates with TFIIIC, is found at tRNA and 5S rRNA genes, and targets their repressor Maf1. *Proc Natl Acad Sci U S A*. 2010. doi:10.1073/pnas.1005188107

197. Gjidoda A, Henry RW. RNA polymerase III repression by the retinoblastoma tumor suppressor protein. *Biochim Biophys Acta - Gene Regul Mech*. 2013. doi:10.1016/j.bbagrm.2012.09.011

198. Cairns CA, White RJ. p53 is a general repressor of RNA polymerase III transcription. *EMBO J*. 1998. doi:10.1093/emboj/17.11.3112

199. Haurie V, Durrieu-Gaillard S, Dumay-Odelot H, et al. Two isoforms of human RNA polymerase III with specific functions in cell growth and transformation. *Proc Natl Acad Sci U S A*. 2010. doi:10.1073/pnas.0914980107

200. Gingold H, Tehler D, Christoffersen NR, et al. A Dual Program for Translation Regulation in Cellular Proliferation and Differentiation. *Cell*. 2014;158(6):1281-1292. doi:10.1016/j.cell.2014.08.011

201. Pavon-Eternod M, Gomes S, Geslain R, Dai Q, Rosner MR, Pan T. tRNA over-expression in breast cancer and functional consequences. *Nucleic Acids Res*. 2009. doi:10.1093/nar/gkp787

202. Goodarzi H, Nguyen HCB, Zhang S, Dill BD, Molina H, Tavazoie SF. Modulated expression of specific tRNAs drives gene expression and cancer progression. *Cell*. 2016;165(6):1416-1427. doi:10.1016/j.cell.2016.05.046

203. Pavon-Eternod M, Gomes S, Rosner MR, Pan T. Overexpression of initiator methionine tRNA leads to global reprogramming of tRNA expression and increased proliferation in human epithelial cells. *RNA*. 2013. doi:10.1261/rna.037507.112

204. Kim S, Kim D, Cho SW, Kim J, Kim JS. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res*. 2014;24(6):1012-1019. doi:10.1101/GR.171322.113

205. Lin Y, Wagner E, Lächelt U. Non-viral delivery of the CRISPR/Cas system: DNA versus RNA versus RNP. *Biomater Sci*. 2022;10(5):1166-1192. doi:10.1039/D1BM01658J

206. Stewart MP, Langer R, Jensen KF. Intracellular delivery by membrane disruption: Mechanisms, strategies, and concepts. *Chem Rev*. 2018;118(16):7409-7531. doi:10.1021/ACS.CHEMREV.7B00678/ASSET/IMAGES/MEDIUM/CR-2017-00678N_0045.GIF

207. Maelfait J, Liverpool L, Rehwinkel J. Nucleic Acid Sensors and Programmed Cell Death. *J Mol Biol*. 2020;432(2):552-568. doi:10.1016/J.JMB.2019.11.016

208. Rayner E, Durin M-A, Thomas R, et al. CRISPR-Cas9 Causes Chromosomal Instability and Rearrangements in Cancer Cell Lines, Detectable by Cytogenetic Methods. *Cris J*. 2019;2(6):406-416. doi:10.1089/CRISPR.2019.0006/SUPPL_FILE/SUPP_FIG5.PDF

209. Debacker K, Frank Kooy R. Fragile sites and human disease. *Hum Mol Genet*. 2007;16(R2):R150-R158. doi:10.1093/HMG/DDM136

210. Dillon LW, Pierce LCT, Ng MCY, Wang YH. Role of DNA secondary structures in fragile site breakage along human chromosome 10. *Hum Mol Genet*. 2013;22(7):1443-1456. doi:10.1093/HMG/DDS561

211. Lyu J, Shao R, Kwong Yung PY, Elsässer SJ. Genome-wide mapping of G-quadruplex structures with CUT&Tag. *Nucleic Acids Res*. 2022;50(3):e13-e13. doi:10.1093/NAR/GKAB1073

212. Halazonetis TD, Gorgoulis VG, Bartek J. An oncogene-induced DNA damage model for cancer development. *Science*. 2008;319(5868):1352-1355. doi:10.1126/SCIENCE.1140735

213. Blasco MA, Serrano M, Fernandez-Capetillo O. Genomic instability in iPS: time for a break. *EMBO J*. 2011;30(6):991. doi:10.1038/EMBOJ.2011.50

214. Ibeawuchi C, Schmidt H, Voss R, et al. Exploring Prostate Cancer Genome Reveals Simultaneous Losses of PTEN, FAS and PAPSS2 in Patients with PSA Recurrence after Radical Prostatectomy. *Int J Mol Sci*. 2015;16(2):3856. doi:10.3390/IJMS16023856

215. Winter JM, Fresenius HL, Keys HR, et al. Co-deletion of ATAD1 with PTEN primes cells for BIM-mediated apoptosis. *bioRxiv*. July 2021:2021.07.01.450781. doi:10.1101/2021.07.01.450781

216. Muller FL, Colla S, Aquilanti E, et al. Passenger deletions generate therapeutic vulnerabilities in cancer. *Nat 2012 4887411*. 2012;488(7411):337-342. doi:10.1038/nature11331

217. Sharma G, Sharma AR, Bhattacharya M, Lee SS, Chakraborty C. CRISPR-Cas9: A Preclinical and Clinical Perspective for the Treatment of Human Diseases. *Mol Ther*. 2021;29(2):571. doi:10.1016/J.YMTHE.2020.09.028

218. Vakulskas CA, Dever DP, Rettig GR, et al. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and

progenitor cells. *Nat Med 2018 248*. 2018;24(8):1216-1224. doi:10.1038/s41591-018-0137-0

219. Yin J, Lu R, Xin C, et al. Cas9 exo-endonuclease eliminates chromosomal translocations during genome editing. *Nat Commun 2022 131*. 2022;13(1):1-14. doi:10.1038/s41467-022-28900-w

220. Vartak S V., Raghavan SC. Inhibition of nonhomologous end joining to increase the specificity of CRISPR/Cas9 genome editing. *FEBS J*. 2015;282(22):4289-4294. doi:10.1111/FEBS.13416

221. Lin S, Staahl BT, Alla RK, Doudna JA. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife*. 2014;3:e04766. doi:10.7554/ELIFE.04766

222. Gutschner T, Haemmerle M, Genovese G, Draetta GF, Chin L. Post-translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair. *Cell Rep*. 2016;14(6):1555-1566. doi:10.1016/j.celrep.2016.01.019