Developing CRISPR/Cas9 assays to induce multiple controlled double-strand DNA breaks in human cell culture

A thesis submitted to the University of Kent for the degree of

MSc Genetics

School of Biosciences

University of Kent

2019

Ioannis Emmanouilidis

Ioannis Emmanouilidis

I. Declaration

No part of this thesis has been submitted in support of an application for any degree or qualification of the University of Kent or any other University or institute of learning.

Ioannis Emmanouilidis

Thursday 15 August 2019

Ioannis Emmanouilidis

II. Acknowledgements

I want to express my sincere thanks to my family Maria, Glykeria, Eirini and Antonis who supported me throughout my studies. This thesis would not have come into completion without them. You are the reason why I go to the lab every morning, and I do not regret a day of it. Also, I am thankful to my Mangaka Ohana who, all kept me going and made my year fun and memorable. Until next time!

I am indebted to my academic supervisors Dr Peter Ellis and Dr Christopher Toseland, for first enthusing me about the study of the CRISPR/Cas9 system and who guided me to conduct my research on it. Without them, this project would not have been possible. I am also grateful to their teams, consisted but not limited to, Alexander Cook, Ália dos Santos, Yukti Hari-Gupta, Natalia Fili, Rosie Gough, Hannah Reed and Laura Medda for providing their lab protocols, techniques and the necessary training to bring the project idea into fruition. The collaboration with Dr Ben Skinner (University of Cambridge) and Dr Aaron McKenna (University of Washington) have greatly assisted this project by providing their *in silico* expertise for the bioinformatics investigation of the project. A sincere thanks to my former academic supervisor Professor Darren Griffin and Dr Anastasios Tsaousis, for providing me access to their immunofluorescence microscopes and electroporation machines, respectively.

Finally, I would like to take this opportunity to appropriately dedicate this thesis in memory of my late middle school teacher Dionysios Evagelidis, who inspired me to pursue biosciences from a young age with the right mix of chalk and challenges.

III. Contents

I. Declaration....2
II. Acknowledgements....3
III. Contents....4
IV. List of Figures....6
V. List of Tables....8
VI. Abbreviations....9
VII. Abstract....13
1. Introduction....14
1.1 Types of DNA lesions and cellular response....16
1.1.1 Types of DNA damage....16
1.1.2 Methods of repairing DSBs....19
1.2 Chromatin and signalling in the context of DNA DSBs and γH2AX....20
1.2.1 Structure and function of chromatin and histones....20
1.2.2 Histone variants, H2AX and damage sensing....22
1.3 Past and current methods investigating and mediating DNA damage (DSBs). 26
1.3.1 Indirect induction of DSBs....26
1.3.2 Direct induction of DSBs....29
1.4 Targeted genome editing using the CRISPR/Cas9 system....32
1.4.1 The CRISPR system....32
1.4.2 Structure and function of the wild-type Cas9 protein....36
1.4.3 Cas9 protein targeting efficiency and off-targeting....40
1.4.4 Bioinformatics landscape of CRISPR/Cas9....41
1.5 Project Outline....44
2. Materials and methods....45
2.1 Human cell culture for Cas9 transfection....45
2.2 Cell transformation and recombinant expression in *E. coli*....45
2.3 Nickel-ion and gel filtration chromatography....46
2.4 *In silico* gRNA design for multiple genome interventions....48
2.5 Cas9:gRNA preparation for *in vitro* transfection....52
2.5.1 Transfecting Cas9-NLS:gRNA to MCF10A....52

2.5.2 Validation and optimisation of transfection protocol....53
2.6 Cell immunostaining and fixation....54
2.7 Fixed cell microscopy and quantification....55
3. Results....56
3.1 Cas9-NLS-6xHis protein purification....56
3.2 FlashFry crRNA design and genome hits verification....58
3.3 Microscopy Results....64
3.3.1 Transfection of dCas9-Halo protein to assess electroporation efficiency and qualitative validation (widefield microscopy)....64
3.3.2 Widefield microscopy IF qualitative data....66
3.3.3 Widefield microscopy IF particle quantification....72
3.3.4 Widefield microscopy IF nuclear intensity quantification....79
3.3.4 Confocal microscopy IF qualitative and quantitative data....87
4. Discussion....95
4.1 Recombinant Cas9 protein purification....95
4.2 CRISPR bioinformatics....96
4.3 Cas9 *in vitro* delivery....98
4.4 Immunostaining and microscopy....99
4.5 Conclusions....101
5. References....103

IV. List of Figures

Figure 1: Simple schematic of a typical mammalian eukaryotic cell cycle.....14

Figure 2: Cancer and ageing in the context of DNA damage in replicating and non-replicating cells.....15

Figure 3: Differences between SSBs and DSBs in the context of DNA sugar-phosphate backbone.....18

Figure 4: Schematic of DNA's packaging structures.....20

Figure 5: Chromatin remodelling achieving primary, secondary and tertiary structure organisation.....22

Figure 6: Composition of H2A, its variant H2AX, and the nucleosome core model.. 23

Figure 7: Initiation of repair foci upon exogenous damage in the context of DDR….25

Figure 8: Visual comparison of the ZFNs and TALENs systems.....31

Figure 9: Schematic example of a bacterial chromosome containing a CRISPR locus.....32

Figure 10: Native CRISPR pathway commonly found in prokaryotic organisms.....34

Figure 11: Polypeptide sequence and crystal structure of the *S. pyogenes* Cas9 protein.....36

Figure 12: Proposed DNA targeting model based on PAM recognition in Cas9 by Anders *et al.* (2014).....39

Figure 13: Settings used in Ensembl BLASTN search tool to locate where the newly designed crRNAs maps on the human genome (GRCh38.p12).....51

Figure 14: Purification of Cas9-NLS-6xHis recombinant protein using nickel ion affinity chromatography.....57

Figure 15: Gel filtration of Cas9-NLS-6xHis fractions obtained from nickel-ion affinity chromatography.....57

Figure 16: Overview of genomic targeting of 50 breaks using 50A and 50B crRNA sequences based on Ensembl BLASTN g38.....61

Figure 17: Overview of genomic targeting of 100 breaks using 100A and 100B crRNA sequences based on Ensembl BLASTN g38.....62

Figure 18: Overview of genomic targeting of 150 breaks using 150 crRNA sequence based on Ensembl BLASTN g38.....63

Figure 19: CRISPR imaging of transfected MCF10A cells with dCas9-Halo protein ligated with the TMR ligand.....65

Figure 20: Widefield immunofluorescence shows localisation of γH2AX foci in stained MCF10A nuclei utilising the HS17, centromere, 50A, 50B, 100A, 100B and 150 gRNAs with active Cas9 (Merged).....69

Figure 21: Widefield immunofluorescence showing only γH2AX foci without the Hoechst-stained MCF10A nuclei for electroporated only, Cas9 only and Cas9 with HS17, centromere, 50A, 50B, 100A, 100B and 150 gRNAs conditions (Red channel only – TRITC).....71

Figure 22: Overview of the γH2AX foci particle quantification for control, Cas9 only and Cas9 with HS17, centromere, 50A, 50B, 100A, 100B and 150 gRNAs for all of the time course conditions.....73

Figure 23: Normalised percentage distribution frequency of nuclear γH2AX foci number for HS17, centromere, 50A, 50B, 100A, 100B and 150 transfections compared to the control and Cas9 only conditions.....75

Figure 24: Quantification of nuclear γH2AX foci number for HS17, centromere, 50A, 50B, 100A, 100B and 150 transfections in respect to individual time course conditions.....77

Figure 25: Summary of mean γH2AX foci particle count for HS17, 50A, 50B, 100A, 100B and 150 gRNAs based on the predicted *in silico* cuts in respect to individual time course conditions.....78

Figure 26: Overview of the γH2AX foci nuclear intensity quantification for control, Cas9 only and Cas9 with HS17, centromere, 50A, 50B, 100A, 100B and 150 gRNAs for all of the time course conditions.....81

Figure 27: Normalised percentage distribution frequency of nuclear intensity of γH2AX foci (PVI) for HS17, centromere, 50A, 50B, 100A, 100B and 150 transfections compared to the control and Cas9 only conditions.....83

Figure 28: Quantification of nuclear intensity γH2AX foci (PVI) for HS17, centromere, 50A, 50B, 100A, 100B and 150 transfections in respect to individual time course conditions.....84

Figure 29: Summary of mean nuclear intensity γH2AX foci for HS17, 50A, 50B, 100A, 100B and 150 gRNAs based on the predicted *in silico* cuts in respect to individual time course conditions.....86

Figure 30: Confocal immunofluorescence with merged orthogonal Z-stack shows the localisation of γH2AX foci in greater resolve in stained MCF10A nuclei utilising the 50A, 50B, 100A, 100B and 150 gRNAs with active Cas9 (only for 2h time course condition).....92

Figure 31: Overview for the confocal imaging at 2h incubation for particle quantification, total and average foci area and nuclear intensity for the transfected 50A, 50B, 100A, 100B and 150 gRNAs.....93

Figure 32: Normalised percentage distribution frequency of nuclear γH2AX foci number for 50A, 50B, 100A, 100B and 150 transfections.....94

V. List of Tables

Table 1: Individually calculated concentrations of Cas9-NLS-6xHis protein fractions after gel filtration based on absorption spectra defined by Beer-Lambert law.....58

Table 2: Best selected crRNA sequences in terms of hits which used for *in vitro* transfections.....60

Table 3: Scoring metrics for crRNA sequences designed in FlashFry.....60

VI. Abbreviations

53BP1	p53-binding protein 1
6-4PPs	Pyrimidine-pyrimidone (6-4) photoproducts
aa	Amino acid(s)
ATM	Ataxia telangiectasia mutated protein
ATR	Ataxia telangiectasia and Rad3 related protein
BH	Bridge helix
bp	Basepairs
BRCA1	Breast cancer type 1 susceptibility protein
Cas9	CRISPR associated protein 9
CDK	Cyclin-dependent kinase 2
ChIP	Chromatin immunoprecipitation
Chk1	Checkpoint kinase 1
Chk2	Checkpoint kinase 2
CPDs	Cyclobutane pyrimidine dimers
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
dCas9	Nuclease-deficient Cas9

DDK	Dbf4-dependent kinase
DDR	DNA damage response
DNA-Pkcs	DNA-dependent protein kinase
DSBs	Double-strand breaks
EGFP	Enhanced Green Fluorescent Protein
G_1	First gap growth phase
G_1/S	Cell cycle checkpoint between G_1 and S phases
G_2	Second gap growth phase
G_2/M	Cell cycle checkpoint between G_2 and M phases
gRNA	Guide RNA
H2AX	H2A histone variant
HFD	Histone Functional Domain
HNH	Cas9 endonuclease domain named after characteristic asparagine and histidine residues
HP1	Heterochromatin protein 1
HR	Homologous repair
ICL	DNA interstrand crosslink repair
IF	Immunofluorescence

IR	Ionising radiation
LET	Linear energy transfer
M	Mitosis phase
MDC1	Mediator of DNA damage checkpoint protein 1
MeCP2	Methyl-CpG-binding protein 2
MMEJ	Microhomology-mediated end joining
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
NLS	Nuclear localisation signal
PAM	Protospacer adjacent motif
PARP1	Poly (ADP-ribose) polymerase 1
PBS	Phosphate-buffered saline
PI domain	PAM interacting domain
PIKK	Phosphatidylinositol-3 kinase-like protein kinases
RuvC	Cas9 endonuclease domain named after an *E.coli* DNA repair protein
S	Cell's synthesis phase
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis

SSB	Single strand break
ssDNA	Single-strand DNA
TALEN	Transcription activator-like effector nuclease
TBS	Tris-buffered saline
tracrRNA	Trans-activating CRISPR RNA
UV	Ultraviolet
XFV	Xeroderma pigmentosum variant
ZFN	Zinc finger nuclease
γH2AX	H2AX phosphorylated at serine-139

VII. Abstract

The RNA-guided Cas9 (CRISPR associated protein 9) endonuclease enzyme, is a powerful tool to mediate defined genome alteration. Cas9-based genomic intervention such as gene knock-down attained through DNA double-strand breaks (DSBs) in the genomic area of interest. DSBs are one of the most genotoxic lesions. Historically, studies were centred in the identification of genes involved in DNA damage repair and control. However, these approaches include the use of DNA damaging agents such as ionising irradiation or genotoxic drugs. They induce DSBs at random and non-predictive genome sites where damage dosage is difficult to control. Such interventions are unsuitable for studying specific DSBs sites and how the different DNA damage recognition and repair pathways invoke in response to them in the context of the local chromatin state. This project demonstrates that controlled *in vitro* CRISPR/Cas9 assays in conjunction with promiscuous gRNAs capable of targeting multiple sites is possible to inflict multiple DSBs at defined genomic sites in the human breast epithelial cell line MCF10A. Here, CRISPR/Cas9 introduced DSBs at defined quantities and locations across the human genome using custom-designed gRNAs based on *in silico* predictions. The titratable damage is backed by data presented in this project. The current methodology conceived in this project is anticipated to be a point of reference for more sophisticated and optimised *in vitro* transfections. Furthermore, using a well-defined system for controlled DNA aberrations can be a powerful tool to monitor epigenetic events occurring in the context of different drug-resistant cancer cell lines adapted to genotoxic drugs.

1. Introduction

One of the most critical processes for living organisms is to maintain the integrity of its genome to preserve their cellular homeostatic processes and deliver their genetic information, unchanged, to the next generation (1). There are approximately ~10^{13} cells in the human body, in circulation, each containing ~6 billion DNA base pairs and they suffer as much as tens of thousands of DNA lesions (per cell) in a given day (1–3). These lesions are capable of inhibiting genome replication and transcription during the G_1/S, G_2/M (Figure 1) and spindle assembly checkpoints (M phase) (4–6). These surveillance mechanisms have evolutionarily adapted to prevent the cells from progressing to division processes and carry on the defective genome to daughter cells (4). If these repairing processes do not repair the lesions effectively or correctly, mutations can accumulate to the wider-scale genome often leading to cell senescence, ageing and even cancer and neurodegeneration, essentially

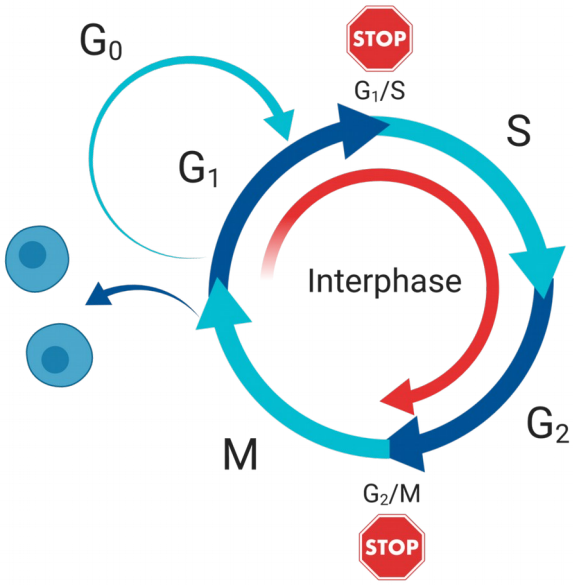


Figure 1: Simple schematic of a typical mammalian eukaryotic cell cycle.

A typical interphase cycle consists of the growth gap (G_1), DNA synthesis (S), growth gap (G_2) phases. Consequently, the cell division takes place at the mitotic phase (M). Resting gap phase (G_0) is only observed in quiescent cells (e.g. neuron cells). Depending on the cell cycle phase at which DNA damage occurs, cells can be blocked in the G_1/S or G_2/M checkpoint borders. Most cancer cells are defective at the G_1 point. Created with BioRender.

endangering cell's and organism's viability in the long term as illustrated in Figure 2 (1, 7–11). A cell must be able to cope with DNA damage, that is tolerable and sustainable by invoking its DNA repair pathways (12). Alternatively, apoptotic mechanisms are invoked in response to excessive DNA damage that is beyond repair (Figure 2) (12). The DNA damage principle is often exploited by current cytotoxic drugs such as bleomycin or cisplatin by inducing significant aberrations within the DNA structure, pushing cancer cells to undergo apoptosis.

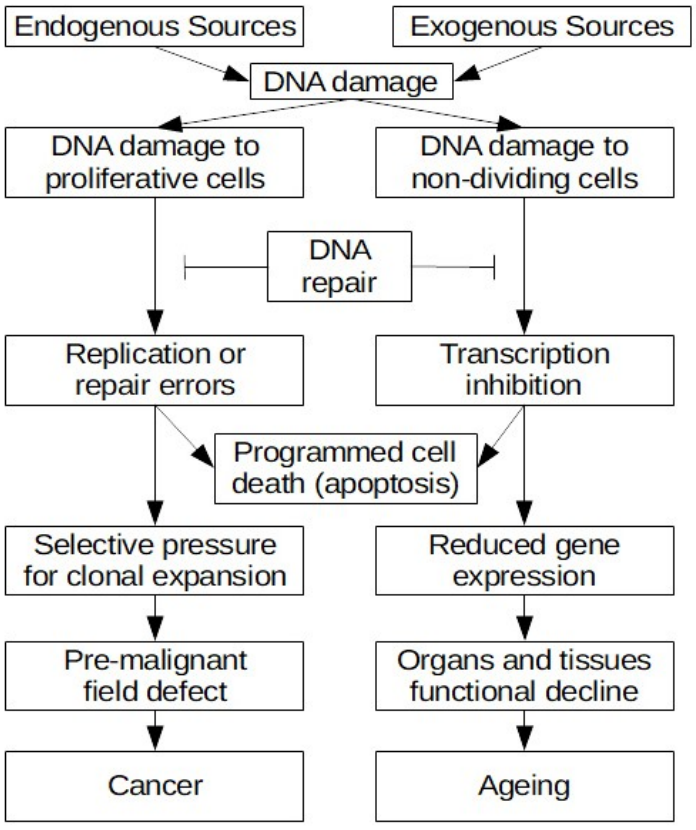


Figure 2: Cancer and ageing in the context of DNA damage in replicating and non-replicating cells.

DNA damage is an inseparable part of the cell experience and an ongoing threat to cell viability and survival. Damage to the genetic material imposes great risk in its faithful transmission to the offspring. DNA repair proteins are activated upon sustainable DNA damage. Excess of unrepaired DNA damage triggers apoptosis to prevent cancer and additional senescence. Ultimately, depending on the cell type, accumulating DNA damage from exogenous or endogenous sources can lead to carcinogenesis or senescence. Adapted from Bernstein *et al.* (2013).

1.1 Types of DNA lesions and cellular response

1.1.1 Types of DNA damage

DNA lesions (or DNA damage) are sites within a given section of a DNA molecule containing alteration in its structure or base-pairing components (13). DNA damage definition includes but not limited to base deletion or alteration, sugar alteration and strand breaks primary caused by endogenous (spontaneous – e.g. reactive oxygen species, replication errors) or exogenous (environmental– e.g. UV-A, UV-B, ionising radiation, genotoxic chemicals) forces (14, 15). DNA lesions often refer to a chemical and physical abnormality in DNA. It prevents the replication machinery performing and functioning correctly to proceed to genome replication. In contrast, mutations is a change within the DNA sequence's base pairs. For example, inaccurate repairs within the DNA sequence can lead to mutations. Both DNA damage and mutation are types of errors that occur in the DNA, but they are fundamentally different in regards to the DNA context (structure vs sequence) and the biological consequences that they produce. Within cells, DNA damage rate is a vital pathology determinant, and if it exceeds the DNA repair rate, the cells become diseased.

In mammalian cells, DNA replication is regulated by the intra S-phase checkpoint (ISC) (16). The detection of DNA aberrations is achieved through Phosphatidylinositol-3 kinase-like protein kinases (PIKKs), composed of ataxia telangiectasia and Rad3 related (ATR) and ataxia telangiectasia mutated (ATM) which they bind to the damaged DNA sites and activating p53, checkpoint kinase 1 (Chk1) and checkpoint kinase 2 (Chk2) (17–21). Both Chk1 and Chk2 can arrest the

cell cycle and trigger the repair pathways and phosphorylate p53, a major transcription factor for bax and p21 to promote cell cycle arrest and pro-apoptotic pathways (22). Then, ISC suppresses cyclin-dependent kinase 2 (CDK) and Dbf4-dependent kinase (DDK) kinase pathways preventing replication machines from copying the damaged DNA template (16, 23, 24). ATM and ATR signalling pathways are critical in radiation-induced damage triggered during cell cycle checkpoints as ATR activates Chk1 and ATM activates both Chk2 and ATR (through crosstalk) (25). In return, these components to recruit non-homologous end joining (NHEJ), homologous repair (HR), microhomology-mediated end joining (MMEJ), DNA interstrand crosslinks (ICL) repair, and nucleotide excision repair (NER) to contribute positively on genome integrity, repair, maintenance and cell survival (26, 27). These assembled networks of cellular components and pathways sensing and responding upon the DNA damage are often named with the umbrella term DNA damage response (DDR).

Different types of damages can occur. Abasic sites are generated by depyrimidination and depurination processes occurring 500-700 times and 2,000-10,000 times per mammalian cell per day, respectively (28–30). These types of damages are often the result of N-glycosidic bonds destabilisation favoured by pH decrease, temperature increase or presence of alkylations within the apurinic/apyrimidinic (AP) site of interest. Additionally, oxidative damage allows chemical modifications in the DNA bases (such as guanine) happening due to the lower electron reduction within a molecule (31, 32). An example of a well researched oxidative lesion is 8-oxo-2'-deoxyguanosine (8-oxo-dG) (32). Reportedly, 10,000-

11,500 oxidative adducts occur per cell per day in humans while specific oxidative damage such as 8-hydroxyguanine (8-oxo-Gua), 5-(hydroxymethyl) uracil (5-HMUra) and 8-hydroxydeoxyguanosine (8-oxo-dG) occur 2,800 times per human cell per day (33–36). Furthermore, cytosine deamination and O6-methylguanines (O^6MeG) alternations occur 192 and 3,120 times mammalian per cell per day (3).

Single-strand breaks (SSB) are one of the most common lesions that can be found in the genome where one helix strand is severed (Figure 3). In contrast, DSBs are considered as one of the most genotoxic DNA damages. They are capable of leading to an unsustainable level of genomic instability within a cell since both helix strands are severed (Figure 3). SSBs and DSBs lesions have reported rate of occurrences at 55,200 SSBs per mammalian cells per day and ~10-50 DSBs per cell cycle in humans, respectively (3, 14, 37). SSBs can be repaired accurately and relatively easy using the undamaged complementary strand under the base excision repair (BER), single-strand break repair (SSBR), nucleotide excision repair (NER) mismatch repair (MMR) cellular mechanisms (38).

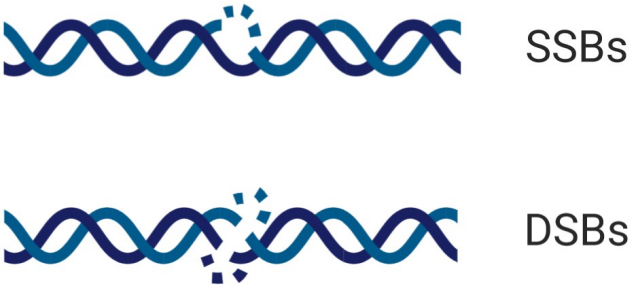


Figure 3: Differences between SSBs and DSBs in the context of DNA sugar-phosphate backbone.

A cell will regularly experience SSBs during its lifetime and is capable of repairing them effectively due to the presence of an undamaged complementary template in the opposite strand. This is not the case for DSBs as such lesions are more difficult to repair and create unrejoined strands causing cytotoxicity and cell death. Created with BioRender.

1.1.2 Methods of repairing DSBs

DSBs lesions are more difficult to repair compared to SSBs since the absence of viable template for repair can lead to insertions, translocations, deletions and chromosome fusions furthering genome instability. HR and NHEJ are the two primal mechanisms exist to repair DSBs in eukaryotic cells (39). However, it is unclear from the current literature which mechanisms or conditions govern the choice to repair the damage through HR or NHEJ (39). While the ATM/ATR-p53-p21-apoptosis/arrest axis is an essential player for DDR and DSBs detection, findings in the last two decades, has witnessed that specialised DNA damage sensing and signalling orchestrate the DDR network capable of detecting chromatin complications (40). Chromatin signalling is an important factor to facilitate recognition of DSBs in eukaryotic organisms that use histone modifications as docking sites to ease effective damage response and repair. This often results in propagating complex protein phosphorylation and ubiquitin pathways to execute appropriate but highly specific DNA damage responses (40).

1.2 Chromatin and signalling in the context of DNA DSBs and γH2AX

1.2.1 Structure and function of chromatin and histones

An unwound DNA from a human diploid cell can span in length as much as 1.8m, which suggests that the genome is tightly packed within its ~10μm nucleus. It is compacted into the chromatin structure consisting of DNA and protein complex forming the nucleosome as its primary structural unit. Each nucleosome comprised of ~100kDa octamer built from histone proteins which are capable of wrapping 146bp of DNA in a left-handed superhelical turn around the core octamer particle ~1.7 times (Figure 4) (41, 42).

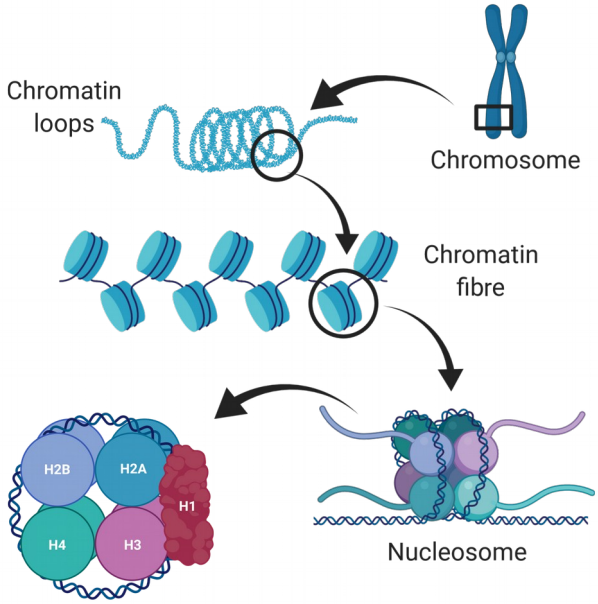


Figure 4: Schematic of DNA's packaging structures.

A total of three levels of chromatin organisation of the DNA can be observed. **1)** DNA is anchored around the nucleosomes creating the "beads on a string" structure which is also known as the euchromatin. **2)** Multiple histones arrays create the histone fibre wrapping around each other to create the 30nm fibre and its most compact form is known as the heterochromatin. **3)** Finally, higher-level supercoiling of the 30nm fibre and chromatin produces the main structure for the mitotic/meiosis metaphase chromosome. Created with BioRender.

In the eukaryotic organisms, five main families of histones exist H2A, H2B, H3, H4, and H1/H5 (Figure 4) (43). The octamer consists of two copies each of the four, H2A, H2B, H3, H4 proteins (Figure 4) (43). The histone H1 is known as the linker histone, which is responsible for maintaining the integrity and stability of the

nucleosome core particle, including, tertiary structures of higher-order (44, 45). The construction of the octamer nucleosome core is based on the interaction of the associated H3-H4 tetramer complex with two H2A-H2B dimer complexes (Figure 4) (43). Helix dipoles in H2B, H3, and H4 cause a net positive charge in the nucleosome that favours the negatively charged phosphate group of DNA to wrap around the histone creating the “beads on a string” structure (43). This allows the chromatin to undergo higher levels of condensation to form the secondary structure comprised of the 30nm fibre (Figure 5) (43). Interactions between neighbouring nucleosomes within their N-terminal domains, allows the 30nm fibre structure to form fibre-fibre interactions creating tertiary chromatin loops and stabilised by H1 (43). Luger, Dechassa and Tremethick suggest that apart from H1 histone, methyl-CpG-binding protein 2 (MeCP2), heterochromatin protein 1 (HP1) and poly (ADP-ribose) polymerase 1 (PARP1) architectural proteins may also contribute to the higher chromatin assembly (43).

The dynamic alteration of the chromatin structure is a barrier-like function which moderates genome replication, transcription and repair based on the open or closed access to the DNA (Figure 5) (43). It is achieved through chromatin remodelling using histone modification enzymes and ATP dependent chromatin remodelling complexes, which are the dominant factors to accomplish this remodelling process in eukaryotes (41, 46).

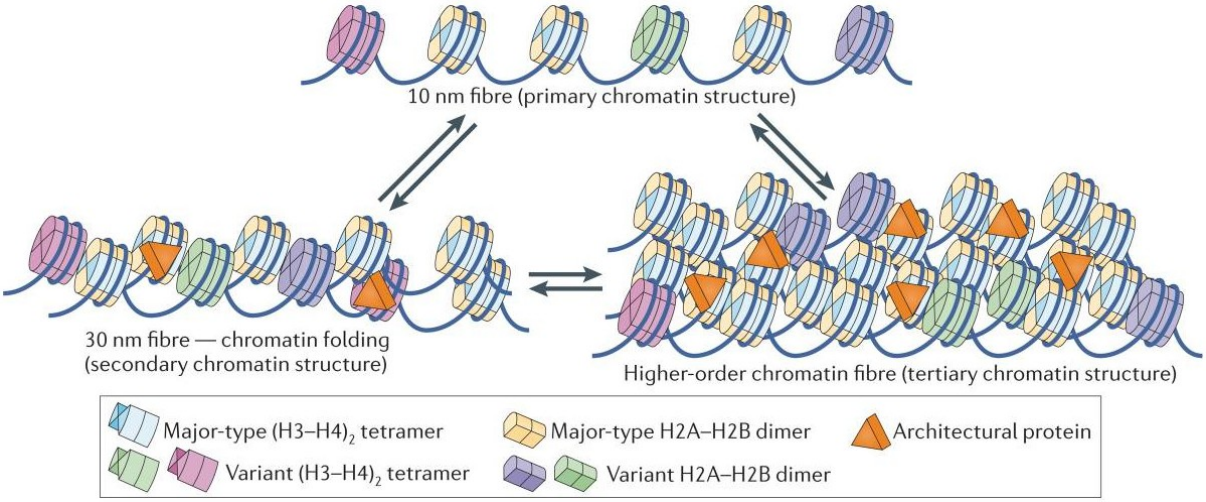


Figure 5: Chromatin remodelling achieving primary, secondary and tertiary structure organisation.

Interaction between adjacent nucleosomes and 30nm fibre-fibre interaction allows chromatin to adjust to higher levels of condensation (primary, secondary and tertiary). This flexible potential allows chromatin to be remodelled for repair, transcription and replication with the aid from ATP dependent chromatin remodelling complexes and histone modification enzymes. From Luger, Dechassa and Tremethick (2012).

1.2.2 Histone variants, H2AX and damage sensing

During the S phase histones are deposited independently to aid for the chromatin stabilisation by replacing the canonical S-phase histones with the histone variants behind the replication forks (47). Two of the nucleosome histone families H2A and H3 have highly conserved and functionally specialised histone variants (48). H2A includes the H2AX and H2AZ variants (47). Specifically, H2AX is encoded by the *H2AFX* gene and represents 2-25% of the H2A histones found in mammals (49). While it maintains the characteristic H2A Histone Functional Domain (HFD) (18-91aa) when compared to H2A, is characterised by its longer 14 amino acid C-terminal and the four amino acid C-terminal SQEY motif whose serine residue can phosphorylate (Figure 6A and 6B) (50). This phosphorylation event is generated in response to the DSBs occurred in the given DNA strand by the activated PIKKs (PI3-

family) of protein kinases ATM, ATR and DNA-dependent protein kinase (DNA-Pkcs) (Figure 6B and 6C) (50). The study from Rogakou *et al.* revealed that exposure to irradiation results in the γH2AX formation as part of the DDR within 20 seconds (49). This activity occurs within minutes following the exposure to irradiation and results in a sequence of events to occur (51).

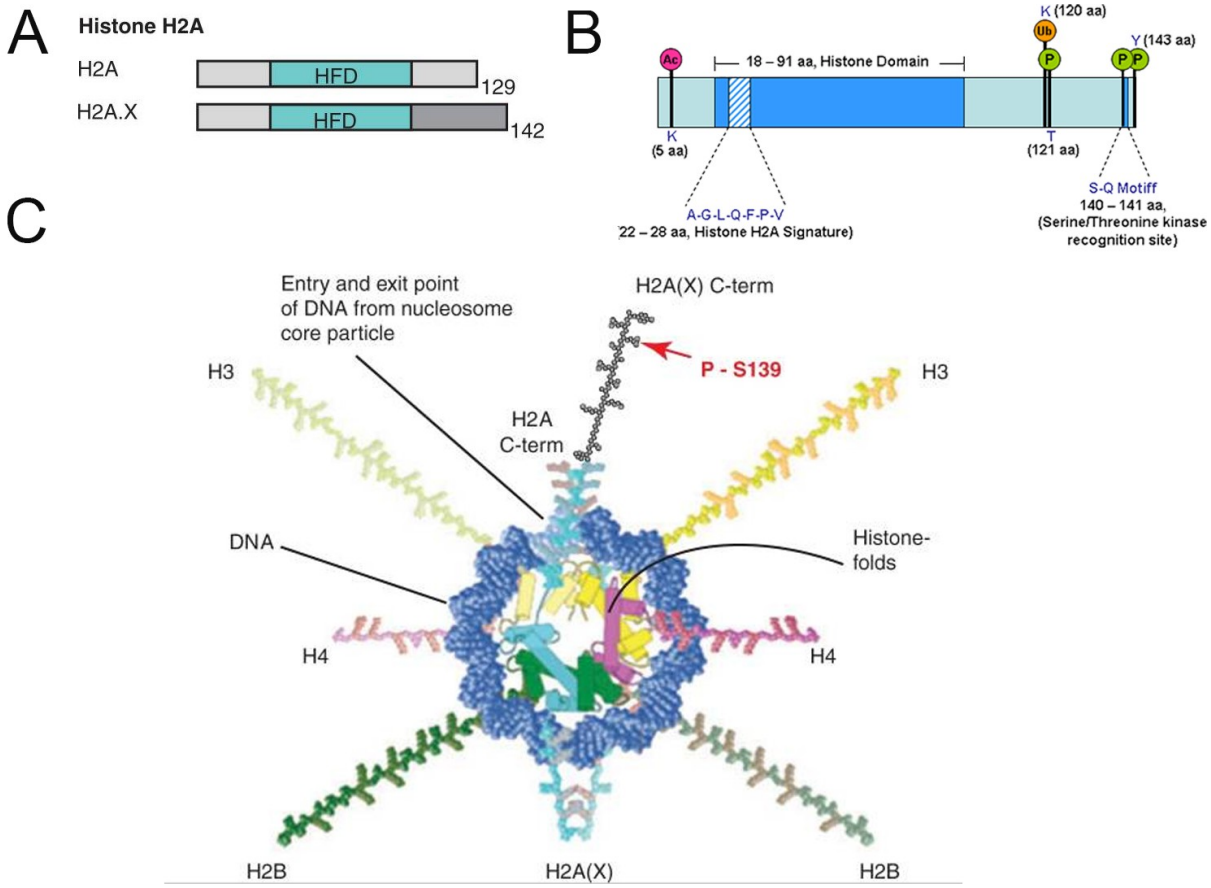


Figure 6: Composition of H2A, its variant H2AX, and the nucleosome core model.

(A) H2AX is one of the evolutionary well-conserved variants of H2A which retained most of the H2A sequence, including the HFD except a small portion of the C-terminus. **(B)** Upon DSBs, kinases of the PI3-family phosphorylate H2AX which is known as γH2AX at Ser139 in the SEQY motif. **(C)** Schematic drawing of the histone octamer including the H2AX variant and the phosphorylation location. Adapted from Henikoff and Smith (2015), Kinner *et al.* (2008) and atlasgeneticsoncology.org

Since its discovery, γH2AX has been utilised in assays to understand the different pathways involved in DNA damage recognition and repair. A sophisticated network of novel proteins is enrolled by interacting and co-localise with γH2AX. Upon extension of the γH2AX domain to several mega-bases in mammalian cells, the mediator of DNA damage checkpoint protein 1 (MDC1) binds to the γH2AX, forming the γH2AX/MDC1 complex (Figure 7) (52, 53). Next, the γH2AX/MDC1 complex recruits the MRN (comprised of Mre11, Rad50 and Nbs1 proteins) complex and ATM proteins to create a positive amplifying feedback cascade to allow γH2AX extension to surrounding chromatin (Figure 7) (54). This propagates a new wave of proteins into the DSBs location such as ubiquitin ligases RNF8 and RNF168 which initiates the poly-ubiquitylation of the H2 histones (H2AX, H2A, H2B) at DSB sites (Figure 7) (54). Further concentration of protein components such as RAD51, RAD52 and RAD54 reversibly interact with the γH2AX site of interest to assist for the DNA repair and recombination (Figure 7) (55, 56). The cascade nexus employs the breast cancer type 1 susceptibility protein (BRCA1) and p53-binding protein 1 (53BP1) to mediate cell-cycle arrest and employ the DNA repair mechanisms such as HR and NHEJ (Figure 7) (57). Additionally, EXPAND1 interacts with 53BP1 to maintain the chromatin structure in a relaxing state to promote cell survival (Figure 7) (58).

While γH2AX is not unequivocal evidence for DSBs, γH2AX-derived foci formation based on the current literature consensus is capable of stimulating checkpoint and DNA repair activation, and sister chromatid cohesion and chromatin remodelling (54, 59).

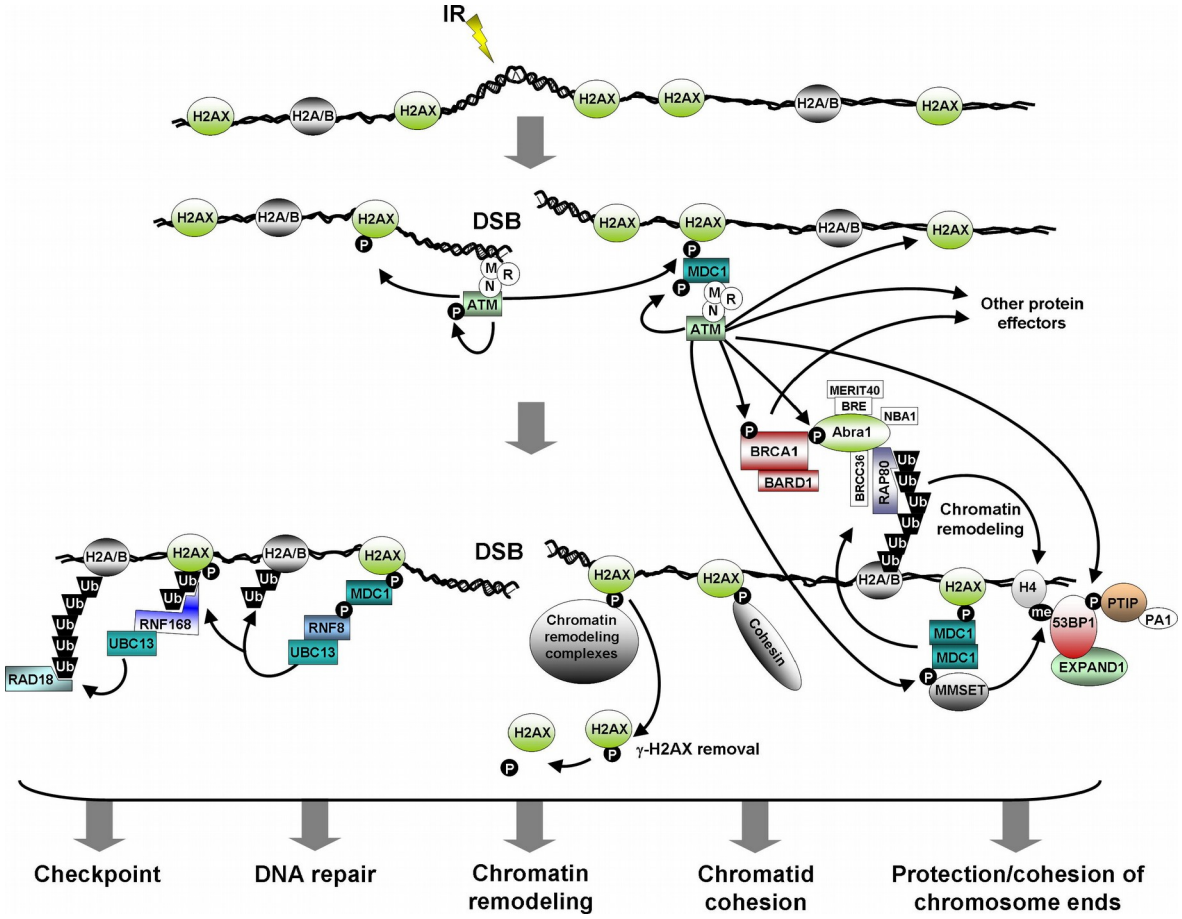


Figure 7: Initiation of repair foci upon exogenous damage in the context of DDR.

Upon DSBs from exogenous sources, MRN complex recruits ATM, ATR and DNA-PK which propagates the phosphorylation of H2AX and in return employs MDC1. MDC1 forms a complex with γH2AX which allows it to spread to surrounding chromatin territory. This promotes the proper concentration of the DDR cascade proteins and other protein effectors to stabilise and repair the DSB. From Ivashkevich *et al.* (2012).

Notably, *H2AX−/−* knock-out causes radiation and damaging agents sensitivity, genomic instability, cell-cycle arrest, immunological deficiency and growth retardation in mice (60–63). The number of chromosome breaks increases between the *H2AX+/+*, to haploinsufficient *H2AX+/−* and further to *H2AX−/−* (63, 64). The *H2AX−/− p53−/−* is the most severe phenotype where the tumorigenesis increases; apoptosis induction is severed and unable to repair spontaneous DSBs effectively and efficiently (64). Therefore, H2AX plays a vital role in the DNA homeostasis and protection of the genetic information by propagating sophisticated chromatin modification to recruit DDR.

1.3 Past and current methods investigating and mediating DNA damage (DSBs)

1.3.1 Indirect induction of DSBs

Past and current studies with the intent to better understand how cells respond upon mediated DNA damage have involved the use of exogenous DNA damaging agents. Such studies aimed to understand which genes are involved in the DNA repair pathways and sensitivity towards DNA damage agents (1, 65). However, the extent and the way these pathways are interconnected in the context of DNA damage investigation and sensing in the different genomic locations based on the local chromatin state is not well understood (66, 67).

Specific frequencies bands within the electromagnetic spectrum can be considered as the most pervasive exogenous sources of DNA damage. Ultraviolet (UV) light (10-400nm) can induce thousand of lesions with the residual UV-A and UV-B spectrum (UV-C absorbed mostly in the ozone layer). In lab, UV-irradiated cells

with induced DSB have been a common practice for mutagenesis and carcinogenesis based studies for a long time (68, 69). DSBs can be produced under heavy UV-B irradiation and potentially from UV-B based reactive oxygen species (ROS) (70–72). However, Rizzo *et al.* show that UV-A may not be as capable as UV-B to form DSBs in primary skin fibroblasts (73). There is not sufficient evidence that UV-A has a pivotal role in DSBs formation. Nonetheless, there is a consensus between these studies that DSBs are not produced directly by the UV wavelength bands itself but from the already generated DNA photoproducts such as cyclobutane pyrimidine dimers (CPDs) and pyrimidine-pyrimidone (6-4) photoproducts (6-4PPs) (74–76). UV-induced cell cycle arrest in the xeroderma pigmentosum variant (XPV) led to the accumulated Mre11/Rad50/Nbs1 complex, prominent nuclear γH2AX foci and stalled replication forks (77, 78). Also, UV is capable of forming interstrand DNA cross-links during the cell's interphase. Specifically, replication forks stalled in the S phase can cause DSBs when they try to replicate the DNA.

Ionising radiation (IR) is defined as the emission of energy capable of ionising atoms or molecules by losing or gaining electrons. In this definition, the upper part of the UV spectrum, X-rays and γ-rays are included. Historically there was an interest in the effects of IR, especially in the aftermaths of World War II since it produces various types of lesions within cells (79). High linear energy transfer (LET) is capable of producing SSBs and DSBs by creating radiolysis radicals that intervene against the DNA backbone (80–82). Nowadays, IR is used by numerous studies to invoke and study repair pathways and their mechanisms. IR is a technique to mediate chromosomal aberrations such as translocations, deletions and dicentrics (83). For

each IR-induced DSB, there are approximately 10 SSBs, and such breaks usually leave terminal nucleotides and phosphoglycolates that cannot be easily ligated by repairing mechanisms (82, 84).

Clastogens are DNA damaging and DSBs inducing capable agents which include but not limited to anticancer chemotherapeutic drugs. Chemotherapeutic drugs such as DNA-alkylating agents have been designed to attach to a DNA alkyl group defined as C_nH_{2n+1}. Such agents intervene against the cancer cells' genome, causing genotoxicity and cytotoxicity to cancer cells leading to apoptosis. Example of such drugs is temozolomide and methanosulfonate (85, 86). Other drugs capable of forming a covalent linkage between adjacent nucleotides include cross-linking agents cisplatin, mitomycin C, and radiomimetic compounds such as phleomycin and bleomycin (85, 86). Topoisomerase inhibitors, such as etoposide and camptothecin, are classes of drugs capable of stalling the DNA cleavage-topoisomerase complexes and generate DSBs and SSBs, respectively (87).

In recent years, studies have utilised hydrogen peroxide (H_2O_2) to mediate ROS related damage through superoxides and hydroxyl radicals to induce DSBs, causing approximately one DSB in every 30 SSBs (88–93). Such studies attempt to generate free radical damage since their mutagenic potency may lead to certain types of cancers (88). The radical hydroxyl mechanism starts by attacking the deoxyribose DNA sugar-phosphate backbone and bases. In both cases, reactivity starts by hydrogen obstructions which leads to strand breakage or by creating new adducted products in π electron-rich bonds at N7-C8 in purines and C5-C6 of pyrimidines (68, 94, 95).

In conclusion, all of the above methods have been important in the study of DNA damage in different cellular contexts. However, these methods suffer from two main disadvantages. They induce many lesions other than DSBs, and they are not capable of inducing targetted DNA damage.

1.3.2 Direct induction of DSBs

Targeted breaks are necessary to investigate how DDR varies in different chromatin regions. Initial studies used restriction enzymes such as the intron-encoded endonucleases I-PpoI and I-Sce1 (96). These homing endonucleases interrogate specific base-pair contents of a given length approximately 15bp and 18bp, respectively. These enzymes cleave the DNA through DSBs, which target a minimal number of sites within the genome of interest (96). Next, DSBs inducible via AsiSI restriction enzyme (DivA) transformed human cell line allows one to mediate hundreds of DSBs located in euchromatin only since DNA methylation is capable of inhibiting AsiSI activity (97). This method is capable only to interrogate the promoter and the gene body only (97). Eventually, there was a need to develop new methods to target all types of chromatin and induce programmed DSBs on demand. Thus, engineered restriction enzymes transcription activator-like effector nucleases (TALENs) and zinc-finger nucleases (ZFNs) were created.

ZFNs are nucleases derived from the fusion of zinc finger DNA-binding domain proteins with the DNA-cleavage domain, FokI type II restriction endonuclease (98). The DNA binding domain contains between three and six individual zinc finger repeats that are capable of recognising specific genomic motifs spanning from 9bp to 18bp in length (98). Each zinc finger in an optimal scenario is capable of identifying

3bp sequences in the form of 5′-$(GNN)_N$-3′ and FokI imposes DSBs in the designated region in either strand (Figure 8) (99, 100). Similarly, TALENs contain the FokI endonuclease which is capable of DNA-cleavage and the TALE proteins derived from *Xanthomonas* bacterium type III secretion system which act as the DNA-binding region (Figure 8) (98, 101, 102). TALE proteins can contain up to 33-35 amino acid domain repeats each capable of recognising a single base pair in the DNA sequence of interest (Figure 8) (98). TALENs mechanism is not well understood, but the TALEN's Repeat Variable Diresidues (RVD) 12th and 13th seem to be associated with nucleotide recognition (103).

While TALENs and ZFNs make more site-specific DSBs compared to irradiation, chemotherapy drugs and previously mentioned restriction enzymes, they are prone to potential problems in regards to undesired cleavage and high engineering costs. To further complicate matters, TALENs and ZFNs are not base sequence addressable. The CRISPR/Cas9 system attempts to solve the above problems with its ability to induce highly specific targeting cleavage of an interrogated DNA sequence guided by a complementary ribonucleotide sequence.

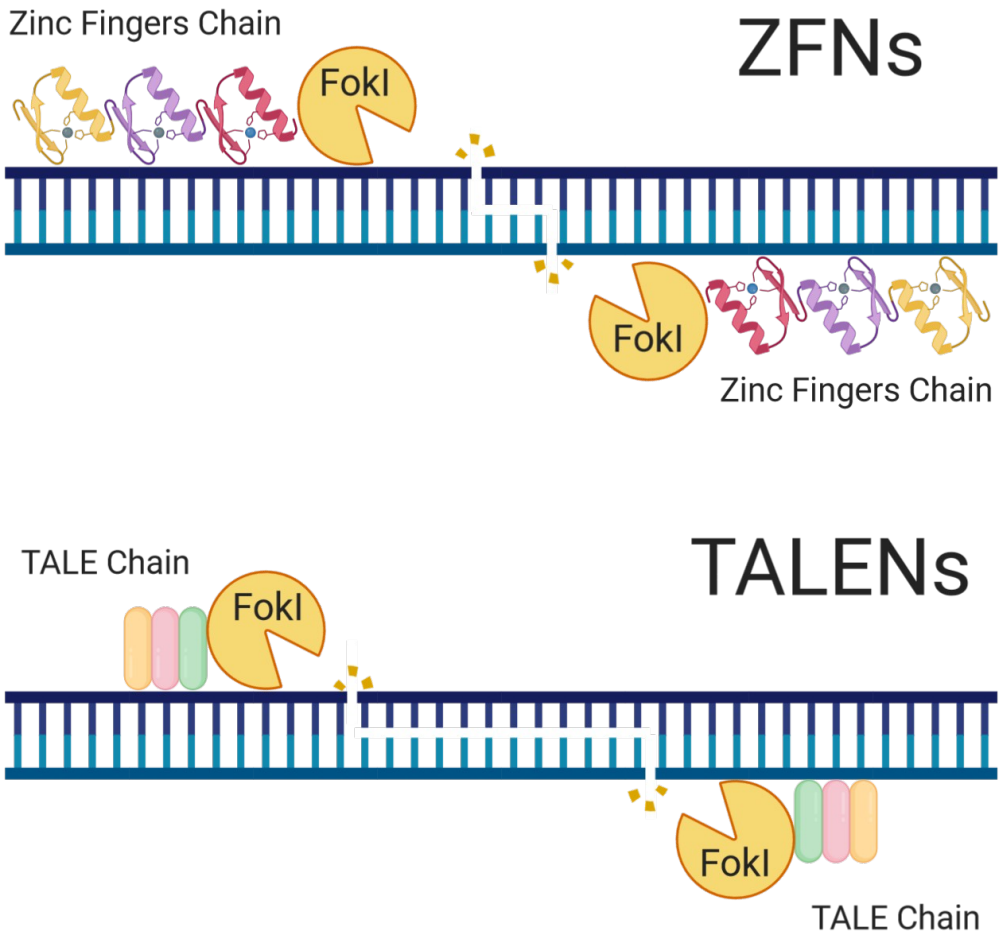


Figure 8: Visual comparison of the ZFNs and TALENs systems.

Both ZFNs and TALENs are capable of exhibiting DSBs with higher precision in desired genomic locations when compared to traditional methods of DSBs induction like chemotherapy drugs, irradiation and other restriction enzymes. The figure demonstrates the ability of the individual compartments of ZFNs and TALENs to recognise three base pairs and single base pair, respectively. Created with BioRender.

1.4 Targeted genome editing using the CRISPR/Cas9 system

1.4.1 The CRISPR system

Clustered regularly interspaced short palindromic repeats (CRISPR) endonucleases have evolved to be an effective adaptive immune system for both bacteria and archaea against foreign genetic elements such as conjugative phages and plasmids (104–108). CRISPR sequences of the prokaryotic chromosomal DNA are well-characterised loci containing identical palindromic repeats ranging from 21-40bp in length (Figure 9) (109, 110). In between them, spacer sequences, also known as protospacers spanning 20-58bp in length (Figure 9) (110). Each of the protospacers at the CRISPR array locus have unique sequences, and they are considered to be an archive of past viral infections that do not match with spacers found in other bacterial species (Figure 9) (109). The numbers of repeats and protospacers vary per locus and may contain as much as 120 adjacent repeats/protospacers and as low as two adjacent repeats/protospacers (109, 110).

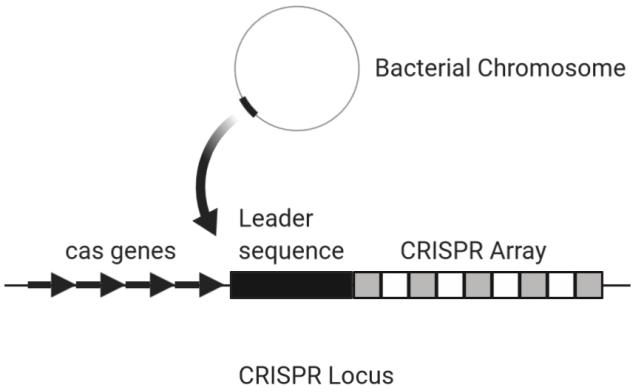


Figure 9: Schematic example of a bacterial chromosome containing a CRISPR locus.

A CRISPR locus consists of CRISPR associated system (Cas) genes, the leader sequence ■, CRISPR repeats ■ (21-40bp) and CRISPR spacers □ (20-58bp). Created with BioRender.

The CRISPR array locus is superseded by the leader sequence and the CRISPR associated system (Cas) genes (Figure 9). Some bacterial species may contain more than one locus (up to 8). *Streptococcus pyogenes* possesses multiple CRISPR loci

with distinctive repeat units. CRISPR sequences can be found in 50% and 90% of the sequenced bacterial and archaea genomes, respectively (111). Depending on their genetic content and structural differences, the CRISPR/Cas system is classified into three main types, type I, type II, and type III and further twelve subtypes (112). CRISPR/Cas prokaryotic systems universally contain the cas1 and cas2 genes, whereas type I, type II, and type III contain the cas3, cas9, and cas10 genes, respectively (112).

The most notable protein from CRISPR is the RNA-guided endonuclease Cas9, a protein complex which relies on RNA cofactor to guide genome intervention catalysis. CRISPR/Cas9 system defends against a plasmid or phage by incorporating the foreign invading DNA into the CRISPR array (Figure 10) (113). Then, a CRISPR unit containing the matching sequence is transcribed to a CRISPR RNA (crRNA) and is joined with trans-activating CRISPR RNA (tracrRNA) to generate the guide RNA (gRNA) sequence (Figure 10). The tracrRNA ribonucleotide sequence is transcribed from the bacterial genome upstream of the cas genes. Next, the gRNA is inserted and bound to the Cas9 protein by forming an active protein/RNA-guided complex (Figure 10). The newly formed activated complex results in a structure conformation activating the two nuclease domains, HNH and RuvC. Upon interrogating the matching sequence of the foreign DNA elements based on the complementary gRNA template, the protein/RNA-guided complex catalyses the foreign DNA by applying DSBs (Figure 10).

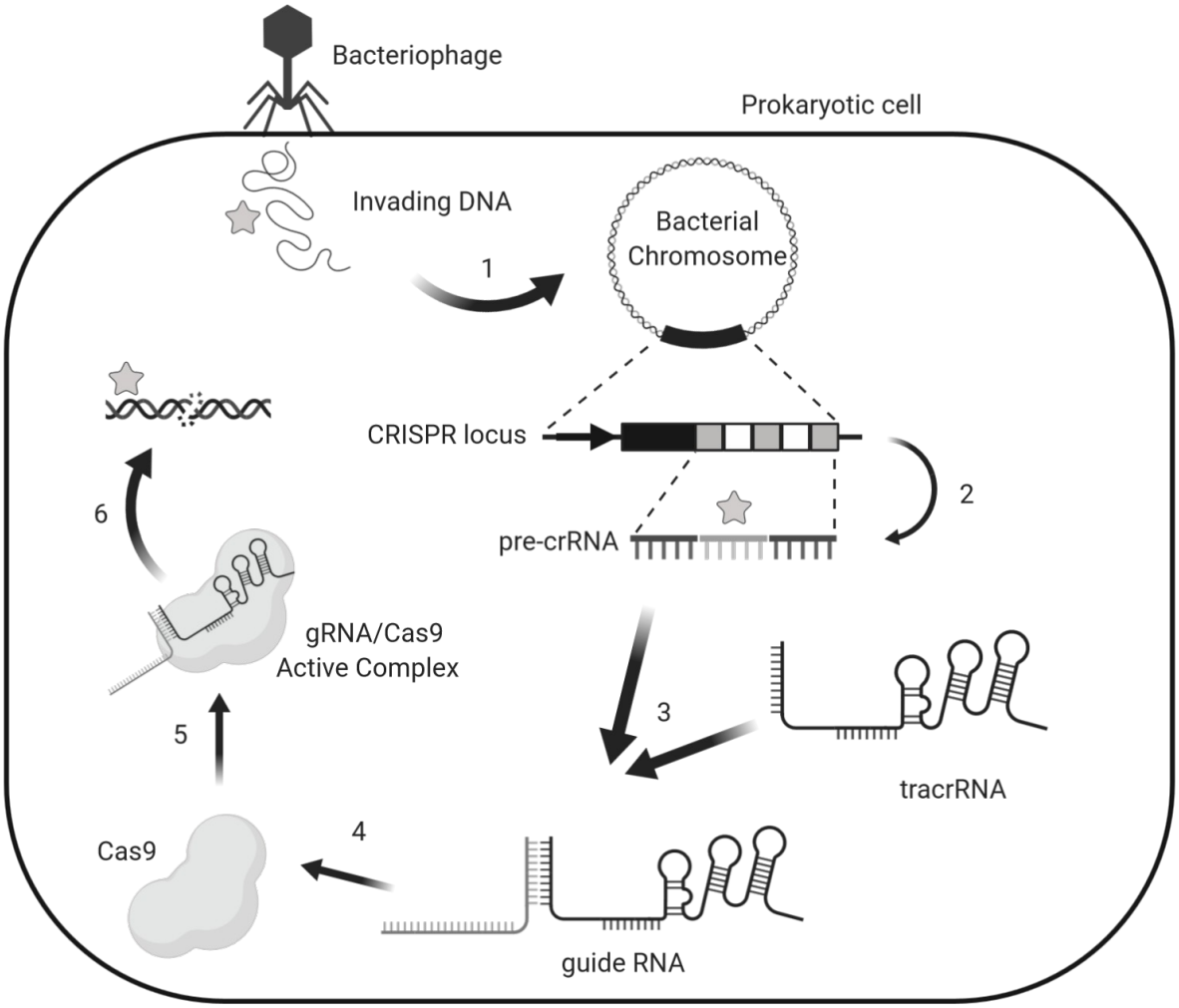


Figure 10: Native CRISPR pathway commonly found in prokaryotic organisms.

The initiation of the CRISPR pathway begins with the invasion of a foreign genetic element (★) (in this case, bacteriophage) into the prokaryotic cell **(1)**. Next, fragments from the foreign genetic element are incorporated into the CRISPR Array, which acts as a storage archive of past infections **(2)**. The cell constitutively transcribes spacer/repeat group containing the foreign element **(3)** which is combined with tracrRNA to form the gRNA molecule. The gRNA is joined into the Cas9 protein **(4)** forming the Cas9:gRNA activated complex **(5)** which binds complementary to the DNA target and cleaves it through DSBs **(6)**. As a result, this process inactivates the invading DNA inhibiting the viral infection. Created with BioRender.

The CRISPR pathway is an essential adaptative system to inhibit viral infection in prokaryotic organisms. Nonetheless, current methods for controlling and manipulating the CRISPR/Cas9 technology has allowed to be emerged as a high precision genome engineering tool. Three Cas9 variants have been adopted for genome editing purposes. Firstly, the wild-type Cas9, which is capable of inactivating genes by cleaving DNA through DSBs induction. This results invoking cellular repair mechanisms either NHEJ which inactivates genes through indels or be repaired by the homology-directed repair (HDR) if a donor template with the relevant homology is present (114, 115). Secondly, Cong *et al.* demonstrated that mutant Cas9D10A is capable of inducing SSBs and invoke only the high-fidelity HDR pathway and not NHEJ allowing for specific and adjacent DNA nicks to be generated (91, 116, 117). Lastly, nuclease-deficient Cas9 (dCas9) due to mutations in D10A in the RuvC and H840A in the HNH nuclease domains allows for specific gene silencing or activation without cleaving the targeting DNA (117–119). Chen *et al.* also demonstrated that imaging repetitive DNA sequences or gene loci can be achieved with dCas9 fused with Enhanced Green Fluorescent Protein (EGFP) (120).

The CRISPR/Cas9 system has seen wide adoption since its initial application in 2012 (117, 121). Cas9 is capable of cleaving any sequence that is complementary to its gRNA cofactor (117). It has been successfully utilised in many cells lines and genomes including but not limited to bacteria, *C. elegans*, zebrafish, *Drosophilia*, *Xenopus tropicalis*, yeast, rabbit, pig, zebrafish, plant, rat, mice, monkey and human genomes (116, 122, 131–133, 123–130).

1.4.2 Structure and function of the wild-type Cas9 protein

CRISPR type II Cas9 is one of the most well-studied nucleases in terms of selectively targeting a specific DNA sequence based on the complementary gRNA sequence. The activated Cas9:gRNA positively charged heteroduplex structure allows identification and cleavage of the desired genomic sequence. The type II Cas9 polypeptide sequence is composed of six core domains (Figure 11A) (134). Jinek *et al.* and Nishimasu *et al.* show that RuvC nuclease domain is composed of a total of three discontinuous segments, RuvC-I, RuvC-II and RuvC-III (Figure 11A) (134, 135). Additionally, it contains the HNH nuclease domain coexisting between the RuvC-II and RuvC-III domains, the recognition domains REC I, REC II and an arginine-rich

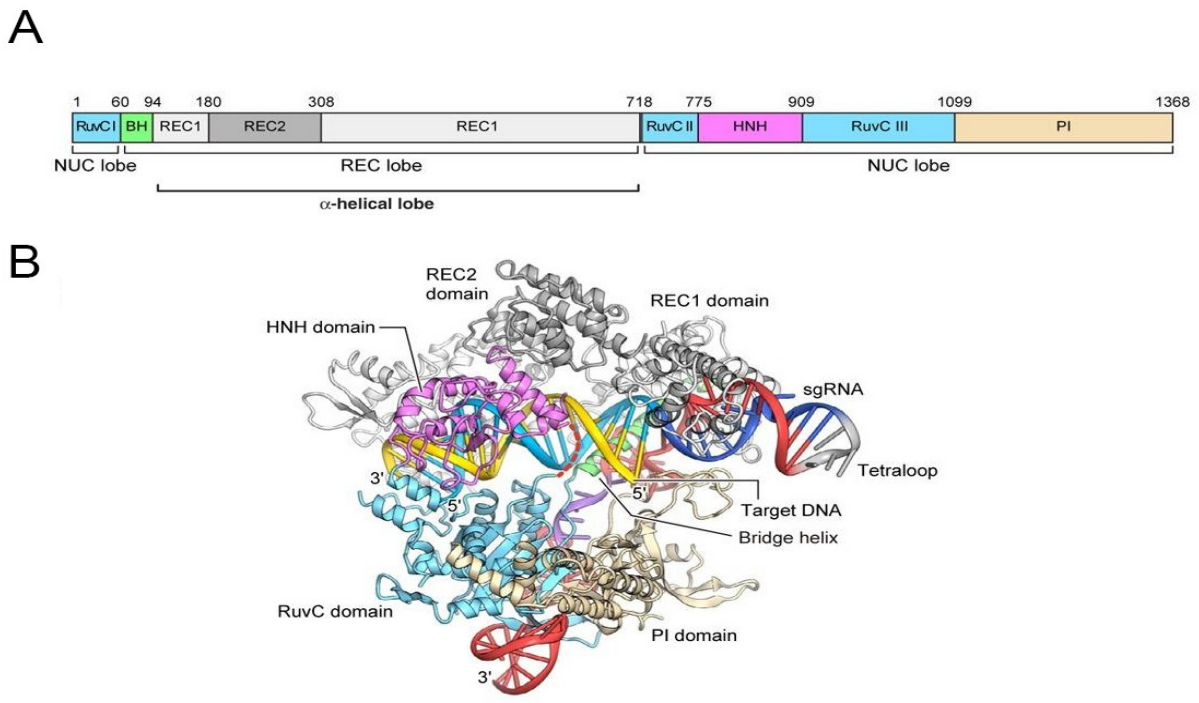


Figure 11: Polypeptide sequence and crystal structure of the *S. pyogenes* Cas9 protein.

(A) Proteomic analysis of the Cas9 sequence suggests that the six domains (HNH, RuvC, Bridge Helix, REC I, REC II and PI domain) can be classified into two categories based on the domain functions, recognition lobe (REC) and nucleases domain (NUC). **(B)** This composition gives rise to a unique nuclease structure where two states can be observed (apo and DNA binding states) due to the flexible conformation interactions taking place between the NUC and REC lobes. Adapted and adjusted from Jinek *et al.* (2014) and Nishimasu *et al.* (2014).

(Arg) region known as the bridge helix (BH) (Figure 11A) (134). Finally, in the sequence's C-terminal (CTD), the protospacer adjacent motif (PAM) interacting domain (PI domain) is located (Figure 11A) (134, 135).

High-resolution crystal structure (2.2 – 2.6Å) of Cas9 suggests a bi-lobed conformation which consists of the α-helical lobe known as the REC lobe and the nuclease domain lobe known as the NUC lobe (Figure 11A and 11B) (134, 135). The two-lobe domains are bridged together with a single bridge helix (BH) rich in Arg, and the given architecture allows the recognition of crRNA-tracrRNA's stem-loop and its integration to the ribonucleoprotein complex (134–136). The gRNA is composed of three stem-loops or one tetraloop (117).

The initiation of the DNA catalysis starts with the recognition of the DNA's PAM sequence, consisting of three nucleotides 5′-NGG-3 or 5′-NGA-3 by the PI domain in the NUC lobe (137, 138). This process initiates the binding of the 20-nucleotide segment of the gRNA 5' end with the complementary annealed single-strand DNA (ssDNA) sequence. The arginine residues Arg1333 and Arg1335 are responsible for interacting with the PAM's major groove GG bases (137). The gRNA:ssDNA complex is anchored in a T-shape like formation promoting tight conformation packing of the NUC lobe with the REC lobe (134, 135). The unpacked and the packed states of the Cas9 are known as the apo state (inactive state) and the DNA bound state (activated complex state). Evidence of these states are the three crystal structures published to date, two representing the DNA bound state (by Anders *et al.* and Nishimasu *et al.*) and one in the apo state (by Jinek *et al.*) (134, 135, 137). Next, under the positively charged and tightened cavity state, the REC I

domain promotes the nucleic acid binding of gRNA (135). The binding is aided by the phosphate lock loop by the Ser1109, which hybridises the PAM and target DNA backbones strands (known as 1-2bp melting) as shown in Figure 12 (137). The function and role of REC II are not well understood. However, molecular dynamics simulations from Palermo *et al.* suggest non-REC I residues contribute to Cas9 plasticity for gRNA recognition and easier switching between the two states for DNA recognition as it undergoes significant conformational changes (139, 140). Eventually, the lock loop propagates the unzipping of the target DNA and gRNA:ssDNA linkage (Figure 12) (137). The above illustrates the high specificity and efficiency of the Cas9 system as a genome-editing tool.

Ultimately, the HNH domain, one of the least packed areas within the Cas9 protein, along with the RuvC domain exerts an invasive conformation against their respective DNA strands which results in their cleavage just after the PAM's upstream third nucleotide base (141). The epitome of the cleavage activity takes place in in the Asp10 and His840 residues of the RuvC and HNH nuclease domains respectively. Both His840 and Asp10 cleaves the 3' - 5' phosphate bond by activating a nearby water molecule which attacks the phosphate group of the DNA backbone promoting the chain of chemical reactions to cleave the DNA strands (Figure 12) (135, 142).

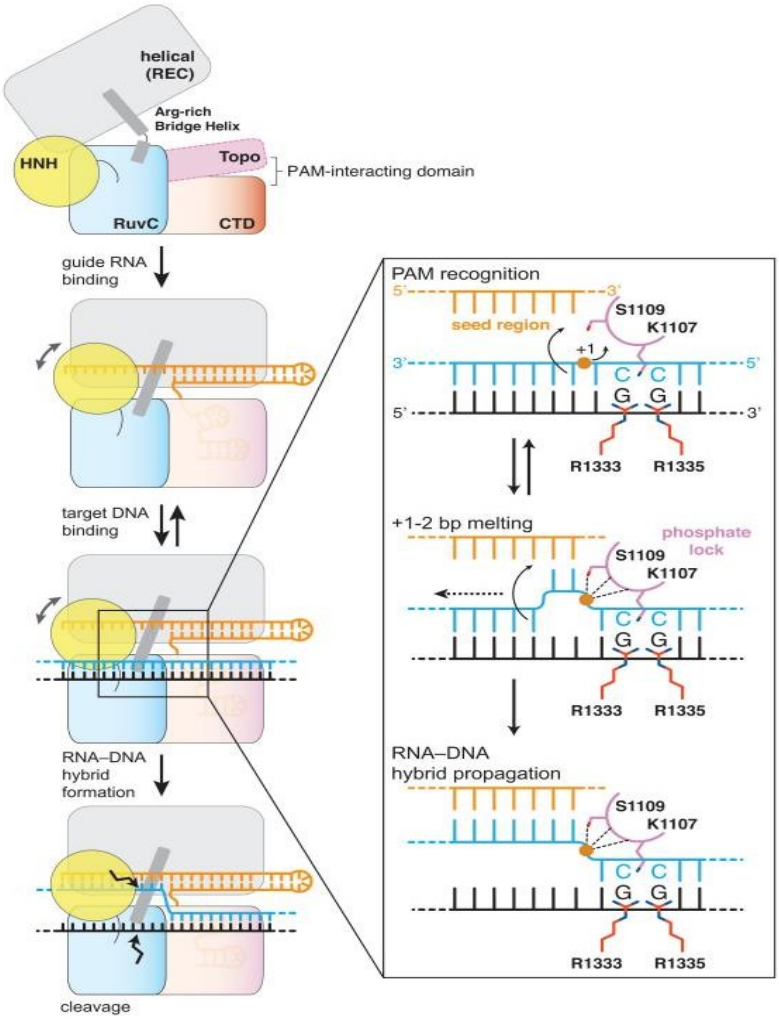


Figure 12: Proposed DNA targeting model based on PAM recognition in Cas9 by Anders *et al.* (2014).

There are four fundamental steps by which the Cas9 protein interrogates targeting DNA. The targeting sequence is recognised by the PI domain through the PAM sequence and forces the Cas9's α-helical REC I and BH to undergo to an ideal conformation to accommodate the phosphate lock loop. This process allows the gRNA:ssDNA to hybridise the ribonucleotide and propagate the cleavage action. From Anders *et al.* (2014).

1.4.3 Cas9 protein targeting efficiency and off-targeting

Targeting efficiency is defined as the percentage of the desired mutation one wants to achieve. It is one of the most crucial parameters when assessing the effectiveness of genome editing tools. When compared to traditional techniques like ZFNs and TALENs, Cas9 is a logistically inexpensive and easier to design alternative with relatively higher targeting efficiency (143–145). Additionally, their widespread use is hindered by time-consuming and engineering costs. In human cells, ZFNs and TALENs have demonstrated efficiencies between 1% and 50% (143–145). However, the active nucleases in both ZFNs and TALENs may lead to unnecessary off-target intervention, leading to undesirable DSBs in the genome of interest, random donor DNA integration, chromosomal rearrangements and even cell death (146–149). Predicting the nuclease activity based on the sequence remains a big challenge for ZFNs and TALENs based genome engineering before the experimental validation in the cellular context (150).

Unlike to ZFNs and TALENs, which relies on a protein/DNA recognition patterns, CRISPR/Cas9 relies on a ribonucleotide to ribonucleotide recognition, essentially allowing gRNAs to target nearly any sequence in the genome of interest. Additionally, the Cas9's PAM sequence increases the specificity of the cutting by acting as a preliminary sequence. The Cas9 efficiency remains largely, a topic of ongoing debate. Reportedly, the Cas9 efficiency in plants and zebrafish genomes can reach up to 70% (127, 151). In humans cells, Cas9 target efficiency depends on the type of cell line is utilised and what repair pathway is utilised (NHEJ vs HDR vs knockout) for each case. In induced pluripotent stem cells (iPSC) as low as 2-4% or

as high as 50% (HDR rate), in 293T cells 10-25%, in K562 cells 8-13%, human mesenchymal stem cells (MSCs) 37.3-80.2% (knockout) (152–154). Zhou *et al.* were able to improve the Cas9's genome target efficiency up to 78% by using dual gRNAs in the mouse genome (155). Also, CRISPR/Cas9 systems are capable to tolerate up to one base difference within the PAM sequence and as many as five bases mismatches in the protospacer region (156, 157). Currently, no studies are investigating the efficiency of desirable DSBs only in any given genome. Identification off-target mutations in order to assess transfection efficiency can be achieved through many means including but not limited to T7 Endonuclease I mutation detection assay, chromatin immunoprecipitation (ChIP), CUT&RUN *in situ* genome-wide profiling and whole-genome sequencing (157–159).

To combat the target inefficiencies, redesigning of crRNA can change the target specificity and location by predicting its genomic intervention using bioinformatics by generating large crRNA libraries based on sequenced genomes (160). High throughput screening, sequencing and especially increased computational power in the last decade have allowed CRISPR target finder tools to increase the easiness and precision of *in silico* designing of genomic targets including on-target and off-target site predictions.

1.4.4 Bioinformatics landscape of CRISPR/Cas9

Recent years have seen the development of sophisticated bioinformatics tools allowing researchers to understand the molecular parameters of the Cas9 system to optimise it for experimental use. The majority of these tools and their main objective

is to design and calculate gRNAs with minimised off-target activity and increased on-target guide efficiency.

Earliest studies concluded that despite the 3bp PAM and the 20bp crRNA sequence specificity, some target sites exhibit different CRISPR/Cas9 activities (116, 117, 161–164). That was problematic for the earliest designing tools since the selection of target sites was based only on simple recognition patterns for the PAM and target sequence motif (165–167). These findings led to large-scale genome screens testing thousand of candidate sequences to identify these causes (156, 162, 168). By limiting sequence GC content, poly-T sequences and including guanine (G) upstream of PAM (e.g. 5' - **G**NGG – 3') one can optimise CRISPR/Cas9 and increase its efficiency across the desired target sites (161, 163, 169). However, each of the studies differs in how they approach the prediction of target sites and what fundamental rules (local and global) included for the gRNA calculation (only 20bp target sequence, 20bp sequence with the PAM, GC content and chromatin state) (170, 171). More sophisticated approaches such as from Doench *et al.* may include additional information rule sets not related to the sequence such as cut site position in regards to the transcription start site (TSS) or gRNA thermodynamic stability (168, 171).

The current standard scores to measure off-target activity include the CFD (considering 20bp target sequence + PAM) and the MIT-Broad score (considers only target sequence) (164). These scoring systems solve some of the long-standing problems with off-targeting false negatives predictions based on experimental data where algorithms model tend to overestimate off-target activity (172, 173). Other

approaches for off-targeting scoring such as CRISTA and Elevation, expand the off-targeting assessment with genomic location, sequence overlap with DNase 1 sensitive regions and gRNA secondary structure criteria (174, 175).

The bioinformatics landscape of the CRISPR/Cas9 research constantly changes as newer approaches and techniques are being developed to target sequences factoring a multitude of parameters to predict on and off-targeting activity and other relevant factors. *In silico* research has made remarkable progress regarding gRNA designing and assessing it against performance metrics. Several CRISPR/Cas9 bioinformatics tools are capable of predicting genome editing based on machine learning models (174, 176). Reintegrating experimental data and key rules such as sequence and non-sequence information into the current models has been and will improve the accuracy of the tools (177).

Further complicating the matter there is no consensus on the current literature on how to approach solving on-targeting or off-targeting activity for either algorithmic or machine training models. Also, there is no agreement on how to measure CRISPR/Cas9 activity. Each of the models makes different assumptions regarding initial conditions, and machine learning approaches use different training datasets. These are just a few of the factors hindering current bioinformatics research. This also suggests that differences in prediction models outcome for CRISPR targeting success are governed by multifactorial features not well understood to this date.

1.5 Project Outline

The ability of histone proteins to undergo chemical modification ensures control of the DNA metabolism within cells to guarantee chromosomal homeostasis. Under abrupt and extremes changes within the histones such as DSBs, histone H2AX is phosphorylated at Ser139 to form γH2AX, which recruits epigenetic responses to repair and stabilise the damaged DNA strands (178). However, the majority of studies are not able to control the DSBs dosage though the use of genotoxic drugs, irradiation or topoisomerases (179).

The specificity and efficiency of the CRISPR/Cas9 system in regards to DSBs induction has been demonstrated by Berg *et al.* which they designed an crRNA:tracrRNA duplex capable of cutting the genome as much as 17 times (180). Nevertheless, the 17 genome cuts gRNA is capable of a limited number of DSBs, which delays the progression of the cell cycle. It may not be capable of producing the number of lesions typically found in studies utilising non-Cas9 interventions.

The aim of this project was to therefore to form a new pipeline to design and transfect a Cas9:gRNA complex into a non transformed human cell line and produced a given dosage of DSBs, at will, at known genomic locations. It is crucial to demonstrate the feasibility of this project in order to study epigenetic responses in a controlled amount of DSBs for future experiments.

2. Materials and methods

2.1 Human cell culture for Cas9 transfection

Human mammary epithelial cells (MCF10A) were maintained in MCF10A cells APC (-/-) medium (Sigma-Aldrich – CLLS1069). The complete baseline of this solution consists of Dulbecco's Modified Eagle's Medium (DME)/Ham's F-12 Nutrient Mixture (Sigma-Aldrich – 51448C) supplemented with horse serum with 5% (v/v) final concentration (Sigma-Aldrich – H1270), cholera toxin 10ng/ml (Sigma-Aldrich – C8052), hydrocortisone 0.5mg/ml (Sigma-Aldrich – H6909), hEGF 20ng/ml (Sigma-Aldrich – E9644) and human insulin 10 µg/ml (Sigma-Aldrich – I9278). Cells were cultured at 37°C in 5% CO_2 incubation and grown to 90-95% confluent before splitting for passaging (maximum 20) occurred every 3 to 4 days (assuming doubling time of approximately 20 hours). Cells were trypsinised with 0.25% trypsin/EDTA.

2.2 Cell transformation and recombinant expression in *E. coli*

Plasmid pET-Cas9-NLS-6xHis responsible for encoding SpCas9 (wild type Cas9 derived from *S. pyogenes*) containing nuclear localisation signal (NLS) sequence and 6x-Histidine tag fused at the C-terminal, was obtained from plasmid repository Addgene (#62933). The pET-Cas9-NLS-6xHis was a gift from David Liu (181).

Plasmid purification achieved using QIAprep Spin Miniprep Kit capable of purifying plasmid up to 20µg. Cell transformation achieved by adding as low as

~200ng of plasmid DNA to 50μL in thawed *E. coli* BL21 (DE3) competent cells (182). Cells-plasmid solution incubated on ice for 30 min and heat-socked at 42°C for 45s and then placed back on ice for an additional 2 min. A total of 250μL of LB (Luria Broth) medium was added to BL21 (DE3)-pET-Cas9-NLS-6xHis cells and incubated in a shaker for 1h at 37°C. Then, 50μL of bacterial culture plated on LB agar containing 50μg/ml ampicillin and 35μg/ml chloramphenicol and subsequently incubated at 37°C overnight (182). A single colony was inoculated in 100mL LB solution with 1:1,000 dilution of 100 mg/ml ampicillin or 50 mg/ml kanamycin placed at 37°C in a shaking incubator (250 rpm) overnight.

Then, 25mL of the overnight culture was added in 1L of LB (total 4 x 1L expression) and 1:1,000 dilution of 100mg/ml ampicillin or 50mg/ml kanamycin was put together into the 1L LB incubated at 37°C in a shaking incubator (~13.3 rpm) (182). When cell growth reached 0.5-0.7 at 600nm (OD600), protein expression induction was initiated by adding 1mL of 1mM M isopropyl-β-D-1-thiogalactopyranoside (IPTG) in each 1L flask. Cultures left to grow overnight at 18°C. Expression cultures were centrifuged at 4°C, 4000rpm in swing-bucket JA-10 rotor for 20 min and resuspended in resuspension buffer (50mM Tris.HCl pH 7.5, 40mM Imidazole, 200mM NaCl, 1mM DTT, 20% sucrose). PMSF inhibitor is added in 1 mM final concentration before storing in -20°C until purification.

2.3 Nickel-ion and gel filtration chromatography

Freshly unfrozen resuspension solution was sonicated for 5 min at 30s intervals at 14 microns to lysate cells. Then, sonicated samples were centrifuged in

swing-bucket JA-25.5 rotor at 18,000 rpm 4°C for 30 min. Then, the supernatant was removed to use for affinity chromatography by loading the lysate to the injection loop system of the chromatography.

Nickel-ion exclusion chromatography achieved using prepacked chromatography column HisTrap 5mL (GE Healthcare – 17-5248-02) in ÄKTA pure protein purification system at 4°C. Buffer A (50mM Tris:HCl at pH 7.5, 40mM Imidazole, 500mM NaCl, 1mM DTT) was utilised to elute unwanted proteins and other debris from the lysate solution and switched to buffer B (50mM Tris:HCl at pH 7.5, 400mM Imidazole) to release Cas9-NLS-6xHis from the nickel-ion column and eluted into 1mL fractions. The sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) gel was run by loading 1:1 eluted sample with running stock solution (2:1:1 SDS Invitrogen NuPAGE LDS Sample Buffer (4X) - NP0007:1M DTT: PBS) at 120V and 0.03A (for 1h) and scaled up to 180V. Any SDS-Page gel involving Cas9-NLS-6xHis investigation utilised PageRuler Plus Prestained Protein Ladder 10 to 250 kDa (Thermo Scientific – 26620). Gels undergo a coomassie treatment with staining (0.1% coomassie (W/V), 50% methanol and 10 % acetic acid) and destaining (40% methanol and 10 % acetic acid) solutions.

If SDS-PAGE gel bands show band shadowing, it is recommended to re-run the eluted protein through size exclusion chromatography to increase the degree of purification. Gel filtration chromatography performed by using HiLoad 16/600 Superdex 200 pg column (GE Healthcare – 28989335) in ÄKTA system at 4°C utilising gel filtration buffer (50mM Tris:HCl at pH 7.5, 1mM DTT, 150mM NaCl). Once again, SDS-PAGE electrophoresis 8% resolving gel was run with mixed 15μL

eluted sample and 5µL SDS NuPAGE (Invitrogen – NP0007) at 120V and 0.03A using Bio-Rad PowerPac Basic Power supply to check if gel filtration was successful. Protein concentration quantification utilised LVis plate from the CLARIOstar High-Performance Monochromator Multimode Microplate Reader by checking the peak spectra at 280nm and calculated using Beer-Lambert Law. Extinction coefficient value for CRISPR/Cas9 (Entry: Q99ZW2) was adopted from ProtParam - ExPASy (assuming disulphide bonds are present – 120575 M^{-1} cm^{-1}) in order to calculate the protein concentration for a given fraction.

2.4 *In silico* gRNA design for multiple genome interventions

The gRNAs used in this project were designed using FlashFry, a command-line tool to discover and score viable CRISPR targets within large sets of genomes developed by McKenna, A. and Shendure, J. (183). The programme was configured on Xubuntu 18.04 LTS Linux system terminal and was run on Java 8 (1.8.0_201) configured as follows:

```
sudo add-apt-repository ppa:webupd8team/java   // add relevant PPA
sudo apt update                                // update system package index
sudo apt upgrade                               // upgrade to latest packages
sudo apt install oracle-java8-installer        // install Java installer script
javac -version                                 // check Java version installed
sudo apt install oracle-java8-set-default      // set default variables on Java
```

FlashFry was downloaded from Github - https://github.com/mckennalab/FlashFry and configured according to the author's recommendations. The binary database was created based on the latest human genome (hg38 build) in FASTA format from UCSC:

```
wget https://github.com/aaronmck/FlashFry/releases/download/1.9.0/FlashFry-
assembly-1.9.0.jar
wget http://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/hg38.fa.gz

//Commands to download FlashFry programme and human genome files


mkdir tmp                                                    //Create folder directory
                                                             named  tmp

java -Xmx4g -jar FlashFry-assembly-1.9.0.jar \               //Execute FlashFry tool

 index \                                                     //Invoke indexing module
 --tmpLocation ./tmp \                                       //Hold content temporarily at
                                                             tmp while database file is
                                                             generated
 --database hg38_cas9_database \                             //Give binary database name
 --reference hg38.fa.gz \                                    //Source from hg38
 --enzyme spcas9                                             //Index 23bp targets for
                                                             SpCas9 for NAG or NGG
```

From version 1.9.0 and onwards, FlashFry allows target extraction capable of cutting multiple times in the genome.

```
java -Xmx4g -jar FlashFry-assembly-1.9.0.jar \               //Execute FlashFry tool

 extract \                                                   //Invoke extract module
 --minInGenome 150 \                                         //Minimum cuts variable in
                                                             genome (e.g. 150 times)
 --maxInGenome 150 \                                         //Maximum cuts variable in
                                                             genome (e.g. 150 times)
 --subsampleProportion 0.1 \                                 //Show competent targets at
                                                             p=0.1 population portion
 --binaryOTFile hg38_cas9_database \                         //Source from binary database
 --outputFile targets.output                                 //Save results in file named
                                                             targets.output
```

The project generated crRNAs capable of inducing 50, 100 and 150 cuts across the genome. Two crRNA versions for the 50 and 100 targets were designed capable of interrogating GC rich and TA rich contents in the human genome. FlashFry scoring annotations include measuring DNA targeting specificity (hsu2013), on-target efficiency (doench2014ontarget), cutting frequency determination (doench2016cfd) and crRNA efficiency (moreno2015) (156, 164, 168, 184).

Subsequently, viable crRNAs were selected based on on-target and off-target scoring metrics generated by FlashFry:

```
java -Xmx4g -jar FlashFry-assembly-1.9.0.jar \       //Execute FlashFry tool

 score \                                             //Invoke score module
 --input targets.output \                            //Read source file to score
                                                       named as targets.output
 --output targets.output.scored \                    //Name output file
 --scoringMetrics doench2014ontarget,doench2016cfd,dangerous,hsu2013,minot \

//Use scoring annotation methods hsu2013, doench2014ontarget, doench2016cfd,
moreno2015 and GC, Poly Thymine dangerous sequences.

 --database hg38_cas9_database                       //Source from binary database
```

The verification of the newly designed crRNA hits across the human genome was done in BLAST/BLAT search from Ensembl with adjusted option to report the maximum number of hits to report to 5000, E-value for alignment report at 1.0, match/mismatch scores equal to 1,-1 with filtering low complexity regions and query sequences options enabled as shown in Figure 13.

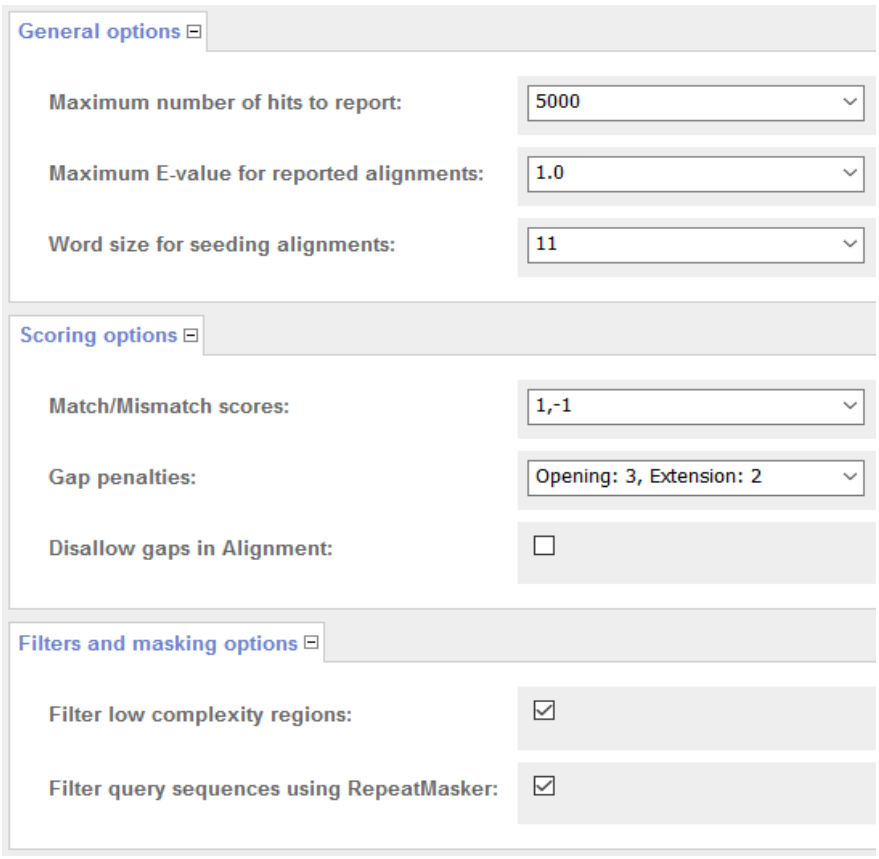


Figure 13: Settings used in Ensembl BLASTN search tool to locate where the newly designed crRNAs maps on the human genome (GRCh38.p12).

The above BLASTN configuration used to verify on-target hits only in an attempt to establish a consensus with the FlashFry discovery and hit metrics.

2.5 Cas9:gRNA preparation for *in vitro* transfection

2.5.1 Transfecting Cas9-NLS:gRNA to MCF10A

When MCF10A cell line grown to complete confluency (~1x10^6 for a complete timecourse condition) in a petri dish in standard antibiotic-free cell culture reagents (including serum and supplements), cells were harvested by a quick wash of 37°C pre-warm PBS or TBS and trypsinised with 0.25% trypsin/EDTA. Cells were centrifuged at 500 rpm for 5min at 4°C and counted using Bio-Rad TC20 automated cell counter. Then, cells resuspended in 37°C pre-warm Opti-MEM I Reduced Serum Medium (Gibco – 31985070). The protocol baseline for the active Cas9:gRNA was based on Dharmacon's Cas9 nuclease protein NLS electroporation protocol. Both crRNA and tracrRNA (Dharmacon Edit-R CRISPR-Cas9 Synthetic tracrRNA – U-002005-20) stock solutions (200μM each) prepared by adding the appropriate volume of RNase-free water. Then, the 100μM solution of crRNA:tracrRNA duplex was created by combining 200μM stock solutions in a 1:1 ratio. The solution was gently mixed for 10 min and stored at -20°C for future experiments.

The Cas9 ribonucleoprotein (RNP) complex was created by combining 1.5μM Cas9 protein and 3μM gRNA final concentration and kept in ice until mixed with the cell-Opti-MEM medium. The delivery of Cas9:gRNA was done through electroporation using Gene Pulser/MicroPulser Electroporation Cuvettes with 0.2cm gap cuvettes at in Gene Pulser Xcell Electroporation System. Transfection achieved using exponential pulse at 300V and 300μF. Cells were left to rest for 5min in the corresponding well, and complete cell culture media added on top of the Opti-MEM in 1:1 manner. Transfected MCF10A cells were incubated at 37°C in 5% CO_2 on

Greiner CELLSTAR 24 well culture plates (Greiner bio-one – 662160). Menzel-Gläser glass coverslips 13mm diametre (Thermo Scientific – 15737602) pre-coated with 50 µg/ml poly-D-lysine (Sigma – P7280) for 1h, 2h, 4h and 6h.

The project utilised HS17 crRNA (5'-CAGACAGGCCCAGATTGAGG-3') from Berg *et al.* which cuts 17 times in the human genome based on the *RPL12* pseudogenes homology sequence and Centromeric crRNA (5'-TTTCTTGCCATATTCCACGT-3') that targets repetitive sequences within chromosomal centromeres (180). Centromeric crRNA allows for multiple breaks in the same ~1Mb region. These sequences used as a baseline comparison with the crRNAs sequences designed *in silico*, as explained in section 2.4 of the materials and methods chapter.

2.5.2 Validation and optimisation of transfection protocol

Previously purified dCas9-Halo transfected into the MCF10A cells prior using the functional Cas9 and gRNA to verify that protein can be electroporated inside the cytoplasm. This step is optional but essential to adjust electroporation protocol for untested cell lines and assess transfection efficiency. Purified fraction was 6µg/µl and a total of 20µg (4µL) and 40µg (8µL) dCas9-Halo protein was added in 400µL resuspended pellet in Opti-MEM. Transfection protocol adjusted and optimised accordingly with the primary objective to use the minimum possible quantity of Cas9 to be electroporated successfully. The expressed protein was derived from pET302-6His-dCas9-Halo plasmid and was a gift from Timothée Lionnet and obtained from Addgene repository (#72269) (185). For Halo labelling, 100nM final concentration of

TMR ligand (Promega – G8252) used on the MCF10A cells and incubated at 37°C in 5% CO_2.

2.6 Cell immunostaining and fixation

Coverslips containing MCF10A cells were washed twice with TBS prior fixing in 4% (w/v) in paraformaldehyde (PFA) for 10min. PFA was quenched with 50mM NH_4Cl/TBS solution for 15min after a round of three washes with TBS beforehand. Once again, cells were washed thrice and stored at 4°C overnight.

Then, cells were permeabilised and simultaneously blocked in 2% (w/v) BSA and 0.1% (w/v) Triton X-100 TBS solution (approximately 0.5ml per well in a 24-well plate) for 15min. Nucleus stained with Hoechst 33342 (Invitrogen – H3570) in a 1/12300 dilution of 20mM of the Hoechst stock in 2% (w/v) BSA and 0.1% (w/v) Triton X-100 TBS solution for 1h. Without washing, cells were permeabilised in the primary γH2AX Ser139 mouse antibody in 1:500 dilution (Sigma-Aldrich – JBW301) for 1h. The coverslips were inverted on a 50μL drop of the antibody/probe placed on a parafilm layer. Then, coverslips washed three times with TBS in a beaker and re-inverted with fluorophore-conjugated (such as Alexa Fluor 555) anti-mouse secondary antibody (Abcam – ab150114) for 1h as before. Coverslips washed one by one with dH_2O three times and mounted on Menzel-Gläser SuperFrost microscope slides (Thermo Scientific – 12372098) with anti-fading agent 2.5% (w/v) DABCO supplemented in mowiol (Sigma-Aldrich – D2522). The slides incubated at 37°C overnight and then stored at 4°C.

2.7 Fixed cell microscopy and quantification

Widefield immunofluorescence images were obtained using CytoVision Olympus BX61 microscope equipped with Olympus UPlanFI 100 X 1.30 Oil Japan numerical aperture objective lens, Olympus WH10X-H/22 eyepiece ocular lens and Hamamatsu Photonics Digital CCD Camera ORCA-R2 C10600-10B-H. Confocal Z stacking imaging was performed on ZEISS Elyra 7 equipped with the Axio Observer 7 inverse strand microscope, using Plan-Apochromat 40x/1.4 Oil (DIC) numerical aperture and Andor iXon 897 EM-CCD camera (SMLM) using polarised laser modules.

DNA damage foci were evaluated and quantified in Fiji (Fiji Is Just ImageJ) by splitting the RGB channels (186). Next, the images were converted to binary to segment foci particles digitally. Despeckle function was used to remove CCD camera artefacts from the images such as salt-and-pepper noise. The area of the nucleus was selected with the Freehand Selection Tool, and the Analyze Particles function used to calculate the number of foci in a given nucleus. In this project, a foci is defined as a particle that has a total area of approximately 15 pixels after filtering out the background and artefact noise in the widefield images. Foci pixel value intensity (PVI) was measured by first, splitting the channels and then using the Measure function in the Analyze menu of Fiji. No binary conversion was used for this measurement. The confocal images analysis was done by adjusting the intensity threshold of the conditions to the controls (electroporated only cells) and then merging all the layers of the orthogonal projection.

3. Results

3.1 Cas9-NLS-6xHis protein purification

The first step in this project was to purify recombinant Cas9 protein for use on the transfections. Eluted protein samples from the sharp peaks were collected using high imidazole buffer and analysed using SDS-PAGE (Figure 14). A sharp peak (at approximately 100mL) containing the Cas9-NLS-6xHis protein was observed in the cation exchange HisTrap chromatogram (Figure 14A). While major contaminants have been eluted in the first 50mL of the elution programme (Figure 14A), SDS-PAGE showed band shadowing, indicating that some contaminants were remaining in the expressed protein (Figure 14B). The eluted fractions from the nickel ion chromatography (1mL each) were combined and eluted through the HiLoad 16/600 Superdex 200 pg column, and the collected peak (Figure 15A) was analysed in SDS-PAGE (Figure 15B). The eluted samples after the size exclusion chromatography show minimal contaminants. This two-step purification protocol allows purifying the Cas9-NLS-6xHis at its highest degree possible, excluding any unwanted contaminants effectively. Absorption spectra at 280nm show that the protocol is capable of producing as much as 45.91mg of purified Cas9 protein for a single 1mL fraction (Table 1).

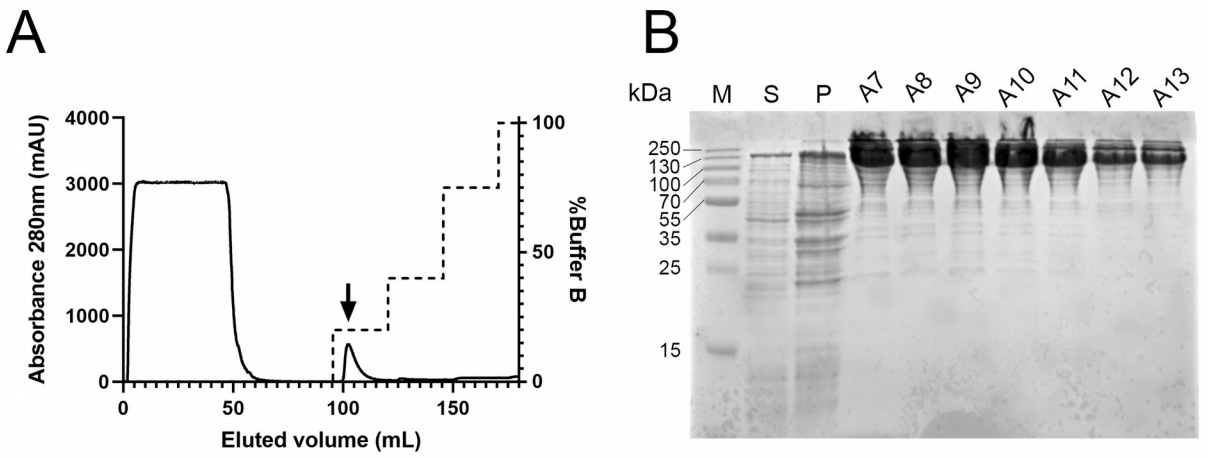


Figure 14: Purification of Cas9-NLS-6xHis recombinant protein using nickel ion affinity chromatography.

(A) Chromatography profile of Cas9-NLS-6xHis protein using HisTrap chelating HP histidine-tagged protein purification column at 4°C in ÄKTA pure protein purification system shown as a straight line (—). The elution gradient of high imidazole (Buffer B) is indicated by the dashed line (---). Elution of Cas9-NLS-6xHis occurred at 100mL elution indicated by the arrow (↓). The fractions collected were 1mL each. **(B)** SDS-PAGE gel (8% resolve) analysis of the peak stained in NuPAGE. M: Thermo Scientific - PageRuler Plus Prestained Protein Ladder (10 - 250 kDa) S: Bacterial supernatant solution P: Bacterial pellet. A7-13: Fractions of Cas9 peak.

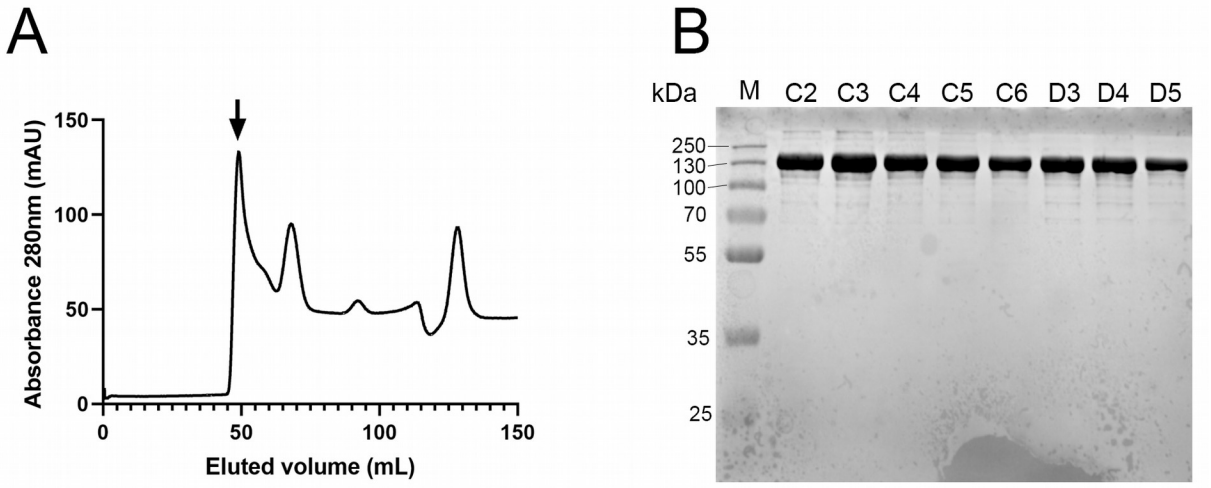


Figure 15: Gel filtration of Cas9-NLS-6xHis fractions obtained from nickel-ion affinity chromatography.

(A) Chromatography profile of Cas9-NLS-6xHis protein using HiLoad 16/600 Superdex 200 pg column at 4°C in ÄKTA pure protein purification system. Elution and collection of the Cas9-NLS-6xHis occurred in at 47mL at the first absorbance peak indicated by the arrow (↓). **(B)** Eluted samples were subsequently analysed in SDS-PAGE 8% resolve gel generating a total of 8 high concentration and viable Cas9 protein fractions. Each of the eluted fractions was 1mL each. M: Thermo Scientific - PageRuler Plus Prestained Protein Ladder (10 - 250 kDa).

Table 1: Individually calculated concentrations of Cas9-NLS-6xHis protein fractions after gel filtration based on absorption spectra defined by Beer-Lambert law.

Fractions	Absorbance at 280nm	Concentration (M)	Concentration (mg/mL)
C2	1.744	2.89×10^4	45.83
C3	1.747	2.90×10^4	45.91
C4	1.075	1.78×10^4	28.25
C5	0.456	7.56×10^5	11.98
C6	0.318	5.27×10^5	8.36
D3	1.285	2.13×10^4	33.77
D4	0.115	1.91×10^5	3.02
D5	0.238	3.95×10^5	6.25

3.2 FlashFry crRNA design and genome hits verification

The FlashFry tool compresses the genome into an organised index for easier and faster discovery of target sequences for the computer. This index essentially contains lookup tables for the enzymatic 23bp (PAM + sequence) target sites in regards to the desired genome. Construction of the index binary for the human genome using FlashFry for targeting both 5′-NGG-3 and 5′-NGA-3 preliminary sequences took approximately 5 hours. Next, a total of five sequences were selected from the generated output list. Sequences were selected based on the predicted cuts 50, 100 and 150 times across the human genome in either PAM sequence NGG or NGA (Table 2). Specifically, there are two versions for each of the 50 and 100 crRNA sequences. These variants that target GC-rich sequences 50A (60.9% GC) and 100A (56.5% GC), and AT-rich regions 50B (69.6% AT) and 100B (60.9%). All of the sequences intervene orientationally forward in the DNA strand. Neither of the strands contain any dangerous sequences features that would be difficult to work with, except for the 100B, which shows multiple thymine (T) bases in a row (Poly T).

FlashFry allows scoring metrics into any newly designed sequences list in order to characterise the sequences *in silico*. One of them, the on-target activity score (doench2014ontarget) ranges from 0 (not active) to 1 (more active guide sequence) which is not applicable for 50A, 50B, 100A, 100B and 150 crRNA sequences (Table 3). These crRNAs are designed for multiple sequence interventions and not for single cleavage (knock out). This multiple target property of these sequences does not create enough sequence context on either side of a guide for the scoring metric to be applicable. Cutting frequency determination (CFD) score based on doench2016cfd predict the likelihood of off-target activity from 0 (not active) to 1 (active) for a given sequence. All of the sequences are highly active, cutting at the off-target based on the doench2016cfd metric score (Table 3). The Hsu2013 scoring is also known as as the crispr.mit.edu score which ranks the specificity of the sequence in either NGG and NAG Cas9 targets from 0 to 100 (higher score are better) (Table 3). All of the sequences showed expected low scores in Hsu2013 (Table 3). A high Hsu2013 score is anticipated in a sequence designed to cut a single time in a designated target in the human genome (knock out) (Table 3). Predicted on count hits from FlashFry and Ensembl show that there is consensus in the number predicted cuts (Table 3). Visual and quantified overview of the hits is summarised in Figures 16, 17 and 18 including hits relative to the number of bases for a given chromosome. Additionally, there are 847 off-target hits for 50A and 452 off-target hits for 50B crRNA sequence (Table 3). Last but not least, 100A, 100B and 150 crRNA demonstrated 1200, 753 and 1142 predicted off-target hit count, respectively (Table 3).

Table 2: Best selected crRNA sequences in terms of hits which used for *in vitro* transfections.

Name	Sequence	Orientation	Dangerous GC	Dangerous PolyT
50A	5'-ACCCCTGGCAGCTGCGGTTCAGG-3'			
50B	5'-TATAATAAGCAAATTGCAATGGG-3'			None
100A	5'-GGGGCTTCCAGGTCACAGGTAGG-3'	Forward	None	
100B	5'-ACTTTAAGTTTTAGGGTACATGG-3'			Yes
150	5'-GTGCCAGAAATCTGGCCACCAGG-3'			None

Table 3: Scoring metrics for crRNA sequences designed in FlashFry

Name	Doench2014	Doench2016	Hsu2013	FlashFry Hits	Ensembl Hits	Off-target Hits[1]
50A		1	5.077	54	53	847
50B		0.941	5.291	52	50	452
100A	N/A	1	2.772	111	100	1200
100B		1	6.161	100	100	753
150		1	3.798	158	154	1142

[1] Off-target defined as having up to four base pairs mismatches.

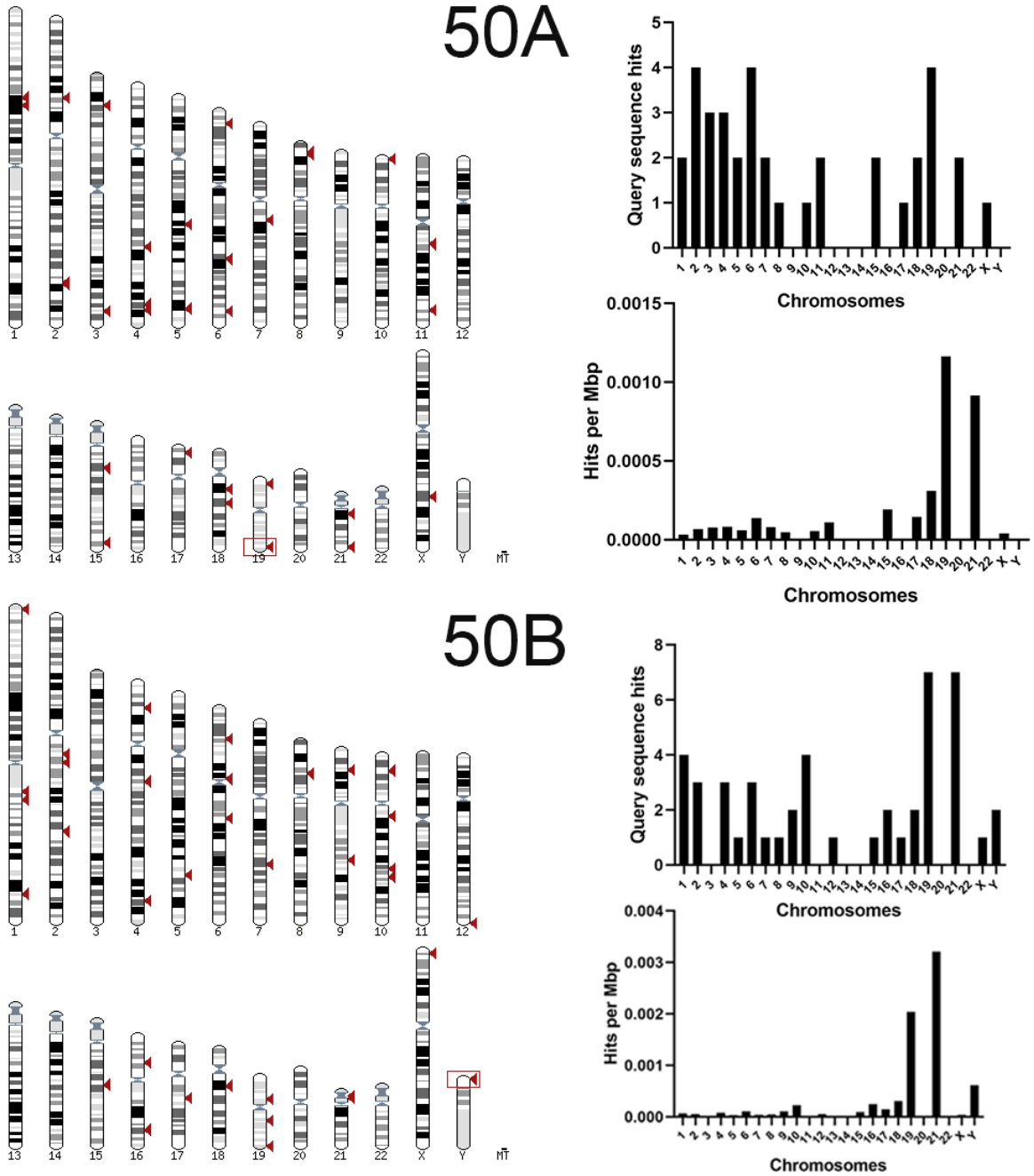


Figure 16: Overview of genomic targeting of 50 breaks using 50A and 50B crRNA sequences based on Ensembl BLASTN g38.

Introducing a defined number of breaks *in silico* with the 50A and 50B set of crRNAs. Red arrows represent only the overlapping genes matched the query crRNA sequence, which may suggest a toxic intervention. Bar charts depict all the hits across individual chromosomes, including hits per Mbp for the given crRNA sequence. Sequences generated on FlashFry.

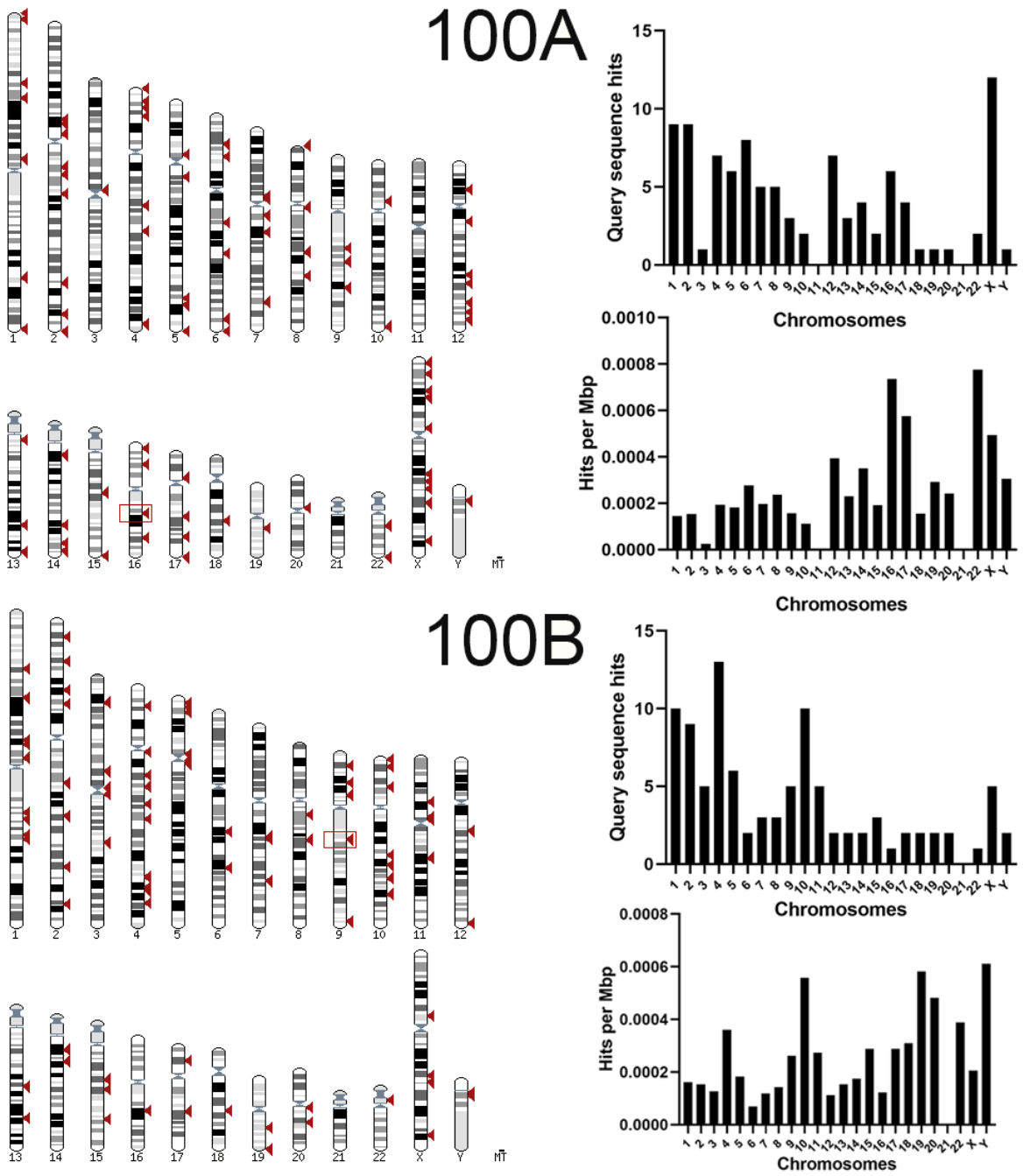


Figure 17: Overview of genomic targeting of 100 breaks using 100A and 100B crRNA sequences based on Ensembl BLASTN g38.

Introducing a defined number of breaks *in silico* with the 100A and 100B set of crRNAs. Red arrows represent only the overlapping genes matched the query crRNA sequence, which may suggest a toxic intervention. Bar charts depict all the hits across individual chromosomes, including hits per Mbp for the given crRNA sequence. Sequences generated on FlashFry.

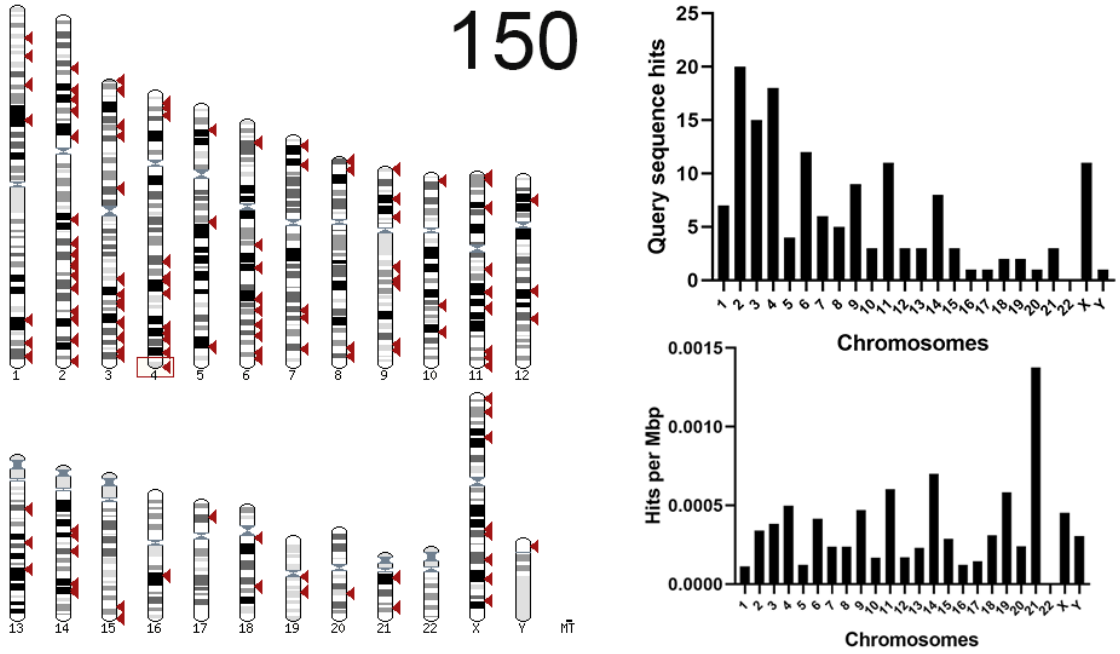


Figure 18: Overview of genomic targeting of 150 breaks using 150 crRNA sequence based on Ensembl BLASTN g38.

Introducing a defined number of breaks *in silico* with the 150 crRNA. Red arrows represent only the overlapping genes matched the query crRNA sequence, which may suggest a toxic intervention. Bar charts depict all the hits across individual chromosomes, including hits per Mbp for the given crRNA sequence. Sequences generated on FlashFry.

3.3 Microscopy Results

3.3.1 Transfection of dCas9-Halo protein to assess electroporation efficiency and qualitative validation (widefield microscopy)

To optimise and validate the feasibility of the electroporation protocol, TMR ligated dCas9-Halo protein was electroporated *in vitro* into confluent MCF10A cell culture (Figure 19). The transfection was assessed in widefield immunofluorescence microscopy by transfecting 20µg (Figure 19B) and 40µg (Figure 19C) of dCas9-Halo to MCF10A cells. The transfection efficiency for 20µg and 40µg was 38.3% and 33.5% on average, respectively. In some instances, the delivery of the dCas9-Halo system achieved above 50% (Figure 19D). The control (electroporation without dCas9-Halo) did not provide any immunofluorescence feedback as expected (Figure 19A and 19D). This part of the pipeline did not involve any gRNA integration to the dCas9-Halo system as this part of the project focuses on creating a starting point for the transfection protocol, which has been optimised in subsequent trials.

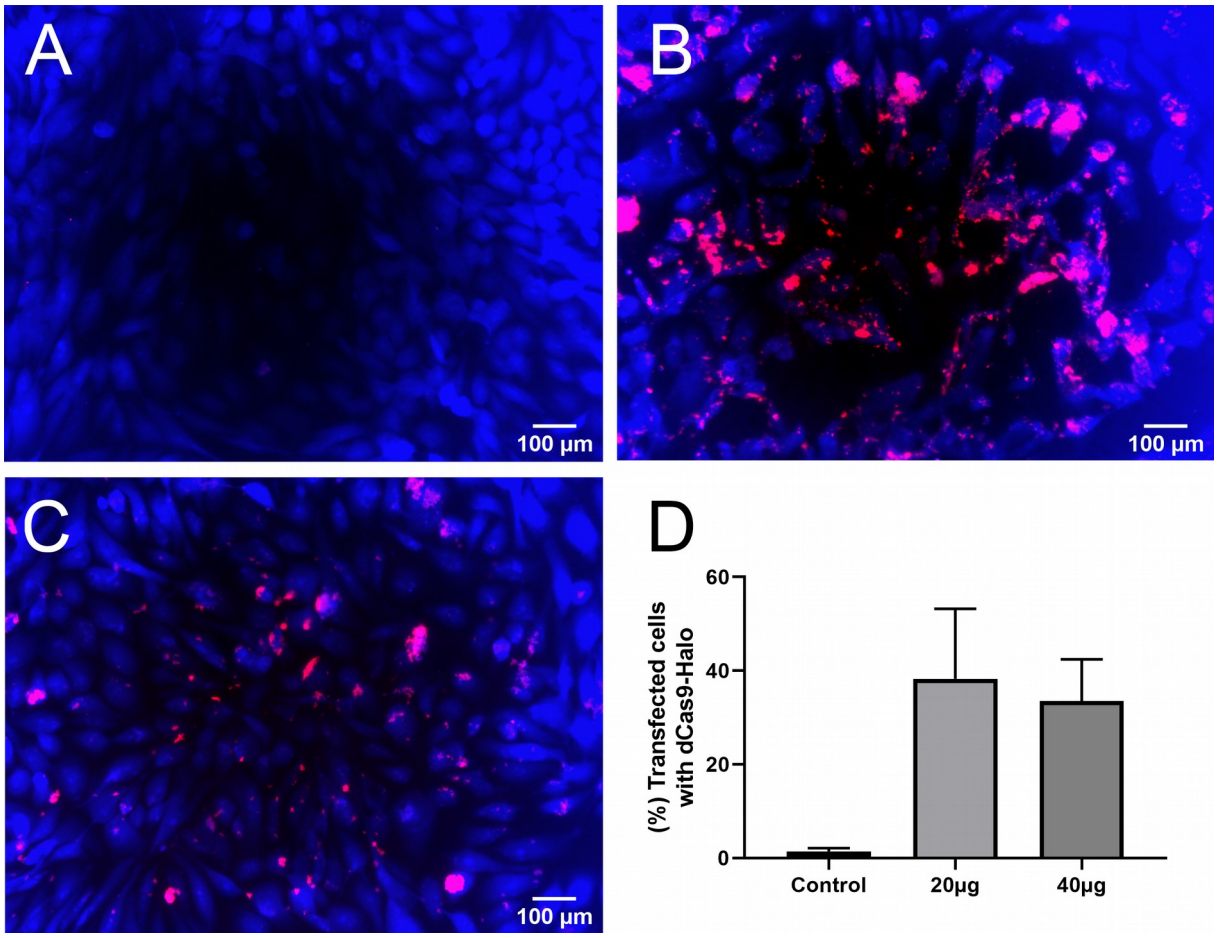


Figure 19: CRISPR imaging of transfected MCF10A cells with dCas9-Halo protein ligated with the TMR ligand.

Widefield immunofluorescence imaging with MCF10A cells electroporated without **(A)** dCas9-Halo, **(B)** with 20µg electroporated dCas9-Halo and **(C)** electroporated 40µg dCas9-Halo. **(D)** Quantification of the dCas9-Halo transfection widefield images for control (n=1727), 20µg (n=3246) and 40µg (n=1977). Transfection efficiency at 300V and 300µF reaches approximately less than 40% on average. Bars indicate the mean with SD error bar. Nuclei fixed with 4% (w/v) PFA and quenched with with 50mM NH_4Cl/TBS solution. Nuclei blocked with 2% (w/v) BSA and stained with Hoechst 33342 (1:12300) and 100nM final concentration of TMR ligand. Imaging was done on Olympus BX61 microscope with DAPI and Cy5 filters.

3.3.2 Widefield microscopy IF qualitative data

The delivery of the active Cas9 into the MCF10A cells is necessary in order to carry the desired genetic modifications. For this part of the project, electroporation of Cas9:gRNA ribonucleoprotein complex is needed to investigate a single aspect of DNA damage through γH2AX foci fluorescence feedback and assess the success of the assay. Apart from the testing the newly designed gRNAs 50A, 50B, 100A, 100B and 150, the HS17 and centromeric gRNAs were used as a point of reference to assess the damage both quantitatively and qualitatively.

To visualise the nuclear location of the DSBs based on the γH2AX foci, MCF10A cells were grown to complete confluency and electroporated with the designated Cas9:gRNA complex at 300V and 300μF. Following the PFA fixation, cells were quenched with NH_4Cl/TBS, permeabilised with 2% (w/v) BSA and immunofluorescence was performed with anti-γH2AX (1:500), which then labelled with Alexa Fluor 555 red (ex:555nm, em:580nm) conjugated secondary antibody (1:500).

The results are seen in Figure 20 for all the transfection conditions for all time courses. Upon delivery of the Cas9:gRNA complex into MCF10A cells through electroporation, the formation of γH2AX foci can be observed from the first hour of incubation (Figure 20). This suggests that within an hour of incubation, Cas9-NLS is imported to MCF10A nuclei from the cytoplasm. Between two and four hours following transfection, the highest level of broken DNA is observed (Figure 20). At six hours post electroporation, these breaks are decreased in numbers or no longer present (Figure 20). This reveals that several of the MCF10A cells are in various

stages of DNA damage repair, ranging from small and distinct γH2AX particles to large and diffused foci. The exception is the transfection involving the first hour Cas9+Centromere gRNA post-transfection, which does not show any or minimal foci formation possibly correlating to chromatin accessibility (Figure 20). Transfections involving the control (electroporation without any Cas9 or gRNA) and Cas9 without any gRNA attached showed minimal to none foci formation. Figure 21 shows only the red channel, which depicts the γH2AX foci (TRITC) without the Hoechst (DAPI). Additional repeats were conducted in order to perform quantification and statistical analysis.

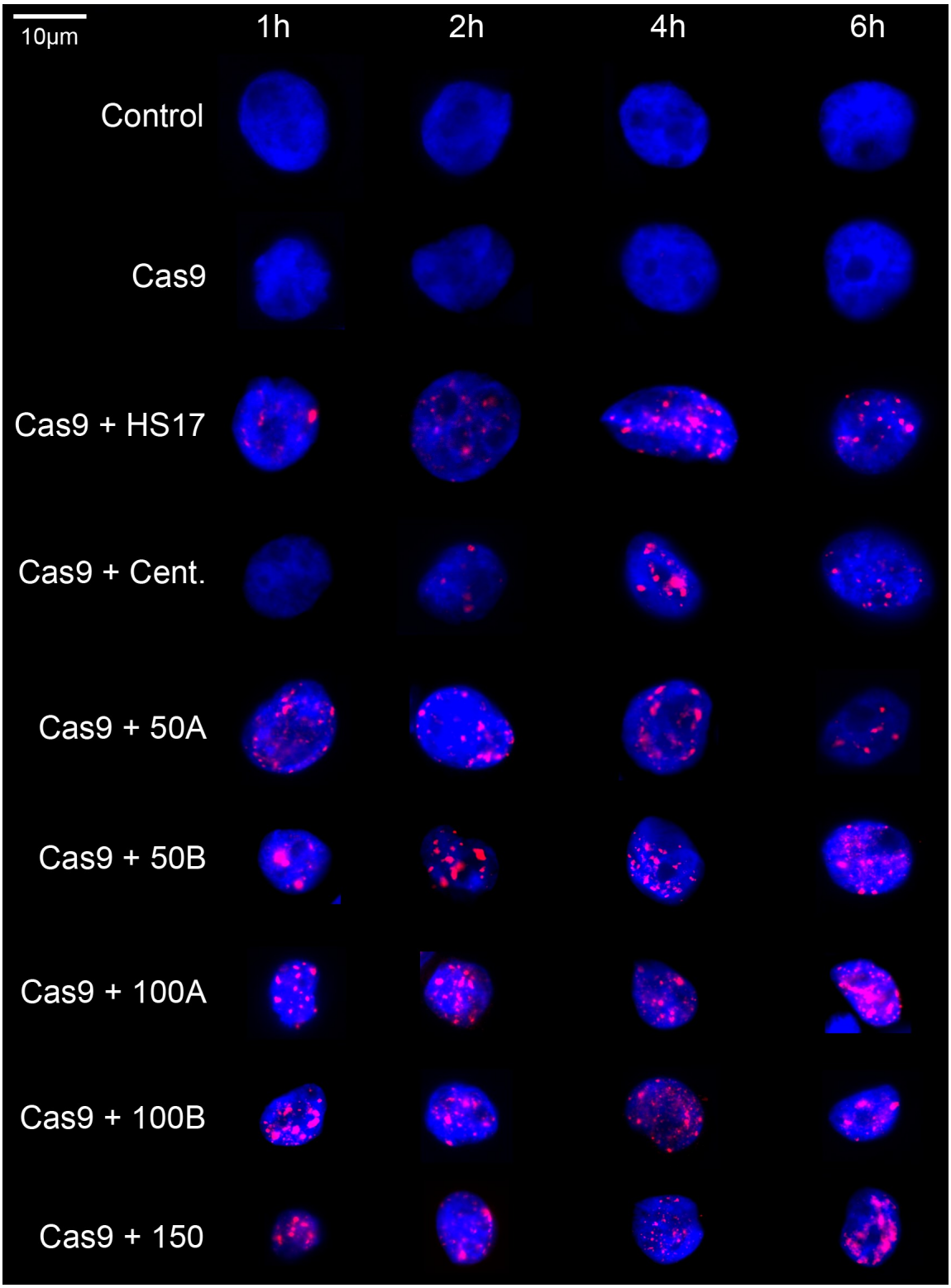

10μm
1h
2h
4h
6h
Control
Cas9
Cas9 + HS17
Cas9 + Cent.
Cas9 + 50A
Cas9 + 50B
Cas9 + 100A
Cas9 + 100B
Cas9 + 150

Figure 20: Widefield immunofluorescence shows localisation of γH2AX foci in stained MCF10A nuclei utilising the HS17, centromere, 50A, 50B, 100A, 100B and 150 gRNAs with active Cas9 (Merged).

Nuclei staining reveal MCF10A cells recruiting γH2AX DSB damage repair foci in four different time courses 1h, 2h, 4h and 6h post-transfection with the respective Cas9:gRNA complex. Control involves MCF10A cells that were electroporated without any Cas9 or gRNA. Cas9 only involves transfections without any gRNAs involved. Utilisation of HS17, centromere, 50A, 50B, 100A, 100B and 150 shows formation of γH2AX foci. To form the gRNAs the HS17, centromere, 50A, 50B, 100A, 100B and 150 crRNAs combined with the Dharmacon Edit-R CRISPR-Cas9 Synthetic tracrRNA. Nuclei fixed with 4% (w/v) PFA and quenched with with 50mM NH_4Cl/TBS solution. Nuclei blocked with 2% (w/v) BSA and stained with Hoechst 33342 (1:12300) anti-γH2AX-Ser139 (1:500) primary antibody. Anti-γH2AX-Ser139 labelled with Alexa Fluor 555 dye-conjugated secondary mouse antibody (1:500). Imaging was done on Olympus BX61 microscope with DAPI and TRITC filters.

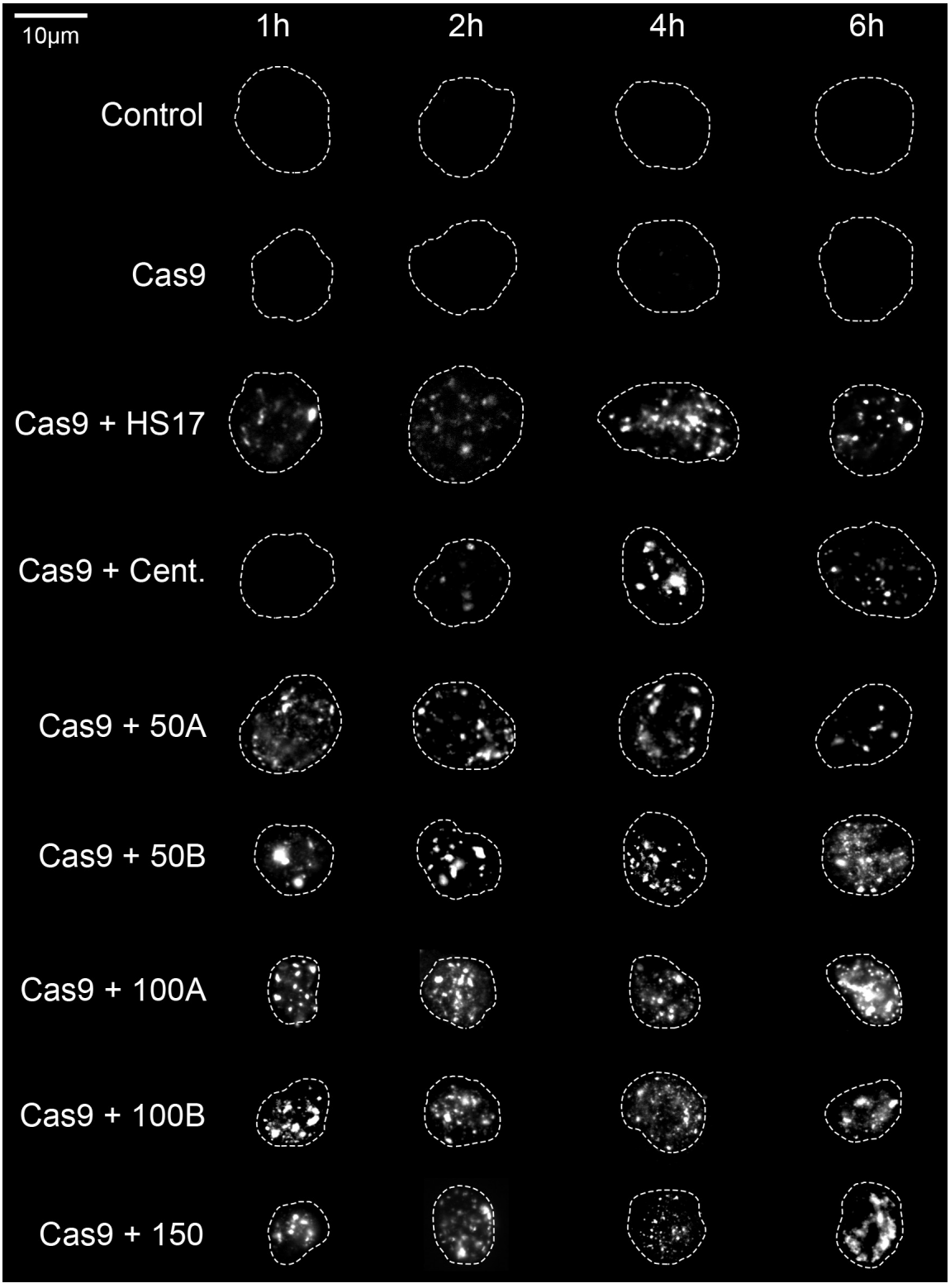

10μm
1h
2h
4h
6h
Control
Cas9
Cas9 + HS17
Cas9 + Cent.
Cas9 + 50A
Cas9 + 50B
Cas9 + 100A
Cas9 + 100B
Cas9 + 150

Figure 21: Widefield immunofluorescence showing only γH2AX foci without the Hoechst-stained MCF10A nuclei for electroporated only, Cas9 only and Cas9 with HS17, centromere, 50A, 50B, 100A, 100B and 150 gRNAs conditions (Red channel only – TRITC).

The red channel shows none or minimal γH2AXactivity occurs for all time courses for control (electroporation only) and Cas9 only transfections. Additionally, Cas9+Cenromere gRNA at 1h incubation shows no γH2AX activity as well. In certain cases, Cas9+HS17, Cas9+Centromere and Cas9+50A show a visual reduction of foci count, especially in the longest incubation 6h post-transfection cases. Utilisation of HS17, centromere, 50A, 50B, 100A, 100B and 150 shows formation of γH2AX foci. To form the gRNAs the HS17, centromere, 50A, 50B, 100A, 100B and 150 crRNAs combined with the Dharmacon Edit-R CRISPR-Cas9 Synthetic tracrRNA. Nuclei fixed with 4% (w/v) PFA and quenched with with 50mM NH_4Cl/TBS solution. Nuclei blocked with 2% (w/v) BSA and stained with anti-γH2AX-Ser139 (1:500) primary antibody. Anti-γH2AX-Ser139 labelled with Alexa Fluor 555 dye-conjugated secondary mouse antibody (1:500). Imaging was done on Olympus BX61 microscope with DAPI and TRITC filters. Dotted lines represent the extent of the nuclei for each condition based on Figure 20.

3.3.3 Widefield microscopy IF particle quantification

To investigate the induction of DSBs quantitatively, γH2AX foci were detected using antibody immunofluorescence (Figure 20 and 21) and quantified by image analysis using Fiji. This was achieved by converting the red channel (TRITC) containing the γH2AX foci of all of the immunofluorescence cell images to binary mask and filtering out the artefact noise.

Similarly, when compared to Figure 20 and 21, the induction of DSBs within the MCF10A nuclei is accumulated at increasing DSBs dosages for HS17, centromere, 50A, 50B, 100A, 100B and 150 gRNAs. The majority of the untreated cells (control) and transfected only with the active Cas9 protein show minimal DNA damage with few exceptions for all time courses (Figure 22). Quantification in Figure 22 also verifies that Cas9+Centromere gRNA formation is minor or none for the first hour of incubation. The single most conspicuous observation to emerge from the data comparison from the overview Figure 22 is that each of the transfections conditions has different DNA damage trends over 6 hours. For example, DNA damage in 100B transfected MCF10A cells remains the same across the incubation time course, while 100A transfected cells demonstrate a downtrend in the DNA damage (Figure 22). In contrary, 50B shows an upward DNA damage trend compared to the 50A transfected cells (Figure 22). However, all of the conditions are capable of retaining a large number of γH2AX foci within the 6h time frame.

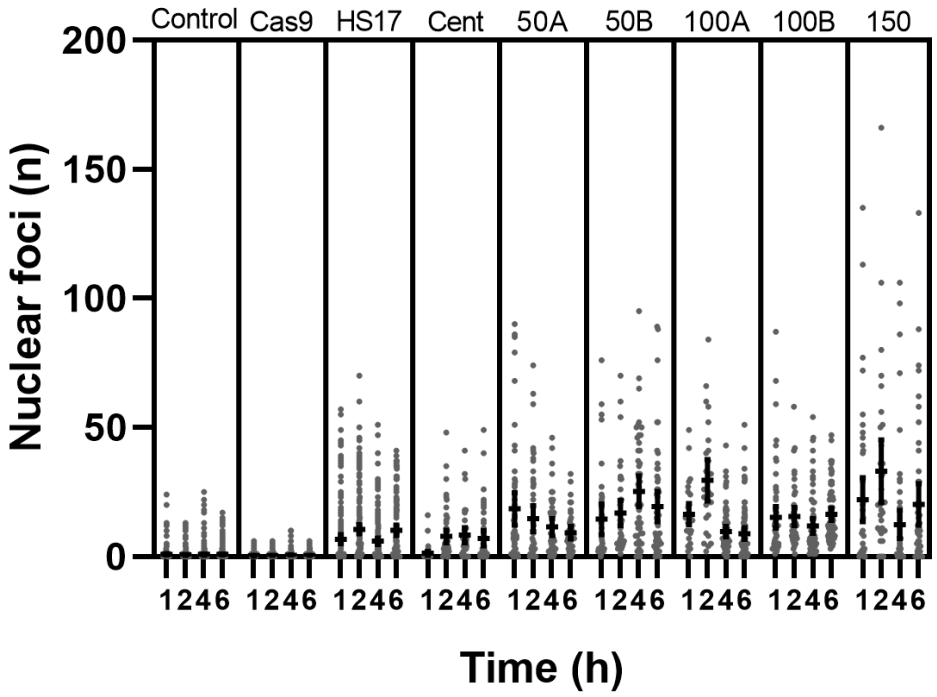


Figure 22: Overview of the γH2AX foci particle quantification for control, Cas9 only and Cas9 with HS17, centromere, 50A, 50B, 100A, 100B and 150 gRNAs for all of the time course conditions.

Quantification of the number of nuclear foci for γH2AX from combined data from independent experiments for each condition and relevant time course. The figure demonstrates the accumulation of γH2AX for the conditions utilised the HS17, centromere, 50A, 50B, 100A, 100B and 150 gRNAs. Error bars represent 95% CI, and middle points represent the mean. Quantification was done in Fiji.

Control: 1h **(n=516)**, 2h **(n=458)**, 4h **(n=412)**, 6h **(n=449)**
Cas9: 1h **(n=301)**, 2h **(n=260)**, 4h **(n=288)**, 6h **(n=291)**
Cas9+HS17: 1h **(n=187)**, 2h **(n=212)**, 4h **(n=197)**, 6h **(n=145)**
Cas9+Centromere: 1h **(n=51)**, 2h **(n=62)**, 4h **(n=44)**, 6h **(n=51)**
Cas9+50A: 1h **(n=57)**, 2h **(n=53)**, 4h **(n=48)**, 6h **(n=43)**
Cas9+50B: 1h **(n=39)**, 2h **(n=45)**, 4h **(n=52)**, 6h **(n=53)**
Cas9+100A: 1h **(n=33)**, 2h **(n=28)**, 4h **(n=76)**, 6h **(n=79)**
Cas9+100B: 1h **(n=65)**, 2h **(n=54)**, 4h **(n=73)**, 6h **(n=75)**
Cas9+150: 1h **(n=48)**, 2h **(n=33)**, 4h **(n=69)**, 6h **(n=48)**

Note: The number of nuclei shown here also apply to the data shown in subsequent Figures 23, 24 and 25. Data presented here are from one experiment each except for Cas9+HS17 which was repeated 3 times.

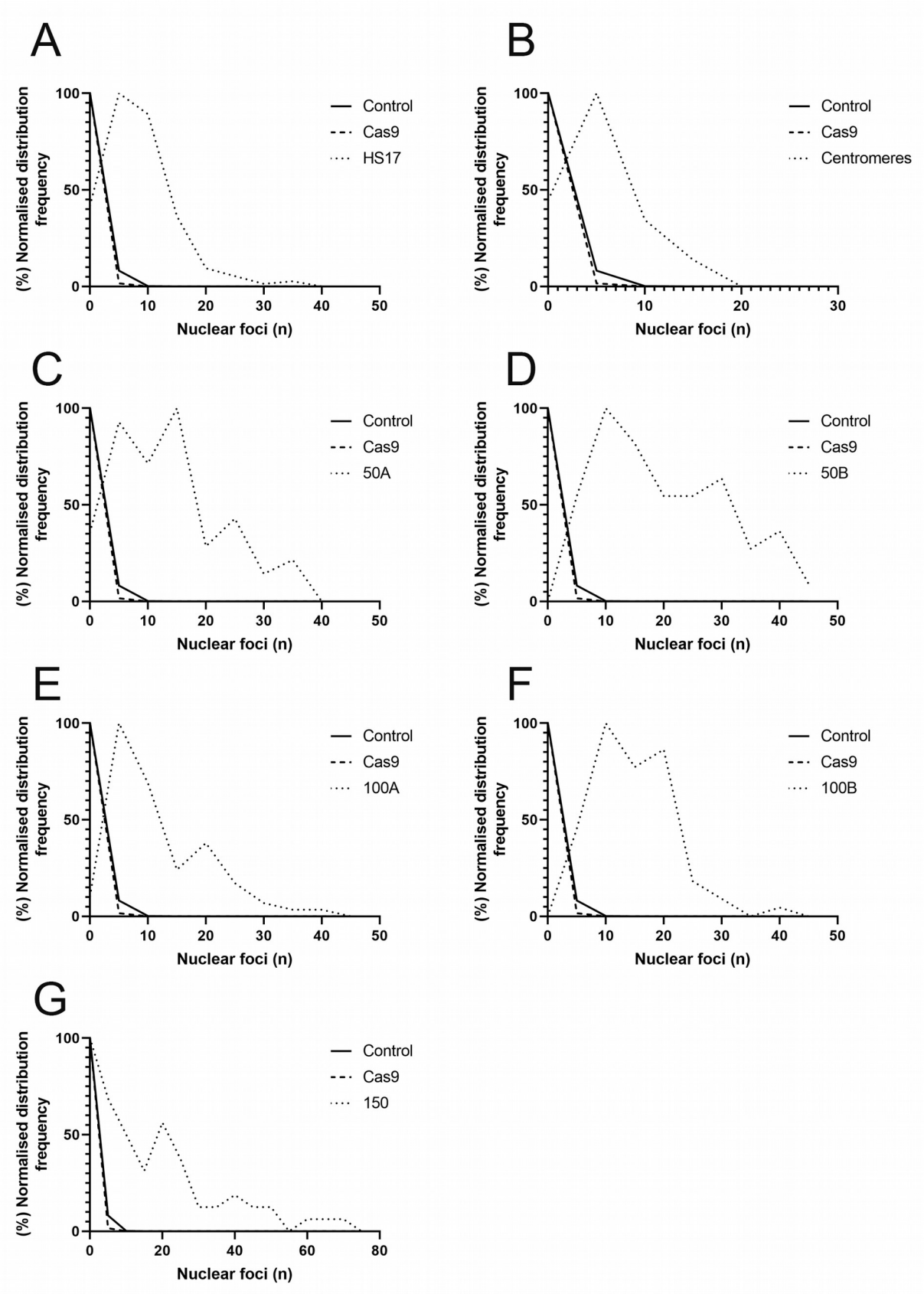

A
(%) Normalised distribution frequency
Nuclear foci (n)
Control
Cas9
HS17
B
Centromeres
C
50A
D
50B
E
100A
F
100B
G
150

Figure 23: Normalised percentage distribution frequency of nuclear γH2AX foci number for HS17, centromere, 50A, 50B, 100A, 100B and 150 transfections compared to the control and Cas9 only conditions.

All time courses have been pooled together to assess the distribution of the data compared to the control and Cas9 only transfection for **(A)** HS17, **(B)** Centromeres, **(C)** 50A, **(D)** 50B, **(E)** 100A, **(F)** 100B, **(G)** 150. The figure shows that in average 0 to 6 foci in a given nucleus may be considered background noise based on the control and Cas9 repetitions (histogram skewed to the right). The bin size for histograms was configured to five.

To asses the distribution of the measurements between the different gRNA conditions used, normalised frequency distribution was utilised (Figure 23). The distributions were created by stacking all of the time course conditions for a given transfection together. Based on the sample population gathered, the bin width was configured to five (Figure 23). These distributions revealed that the majority of the electroporated only MCF10A cells (control) and the ones transfected only with Cas9 demonstrate between zero to six foci on average. The techniques utilised from electroporation to immunofluorescence demonstrated that zero to six foci could be considered interference artefacts from background noise. Figure 22 further suggests that gRNAs with increasing predictive cuts highlights broader distribution in the number of γH2AX foci (Figure 23). The overall cell response to the cuts shows that only a particular population is capable of reaching the predicted number of cuts like HS17 and few in 50A and 50B (Figure 22A, 22C and 22D). However, extreme caution must be exercised in interpreting these findings as a high number of foci are challenging to measure, especially under a single focal plane. The frequency distribution for the centromeric gRNA transfection is similar to the HS17 despite its ability to involve multiple breaks in the same ~1Mb sequence region within the

chromosomal compartments. This comes in consensus when comparing HS17 and centromeres in the overview Figure 22 as they reveal the same number of cuts on average within the context of 95% CI. An explanation for this observation is that widefield microscopy is unable to resolve structures less than 200nm thus potentially merging distinct γH2AX foci into a single and larger one.

Further analysis confirms that under the one focal plane, only certain individual transfected cells display the predictive number of γH2AX foci (Figure 24). Despite that, all of the transfection conditions carrying the relevant Cas9:gRNA are highly significant when compared to the control based on the one-way ANOVA test, which utilises Tukey's multiple comparison (Figure 24). The majority of the cases express strong significance, either $P \leq 0.001$ or $P \leq 0.0001$ (Figure 24). Occasionally, there are cells like in HS17, 50A and 50B that display twice the number of predictive cuts, suggesting arrest in G_2/M with duplicated genome (Figure 24). As expected, non-gRNA bearing Cas9 conditions are not significant when compared to the control for any time course experiment, as well as the 1h condition of Cas9+Centromere, as shown in Figure 24A.

Based on the above, the correlation between the average measured foci count and predicted *in silico* cuts were tested, as shown in Figure 25. All of the time course conditions show a positive correlation of the average number of foci with the predicted number of cuts. Despite the positive correlation, the conditions do not agree with the mean number of measured foci with the predicted number of cuts (Figure 25). Apart from this non-alignment, only the 4h hour post-transfection period show no correlation (Figure 25C).

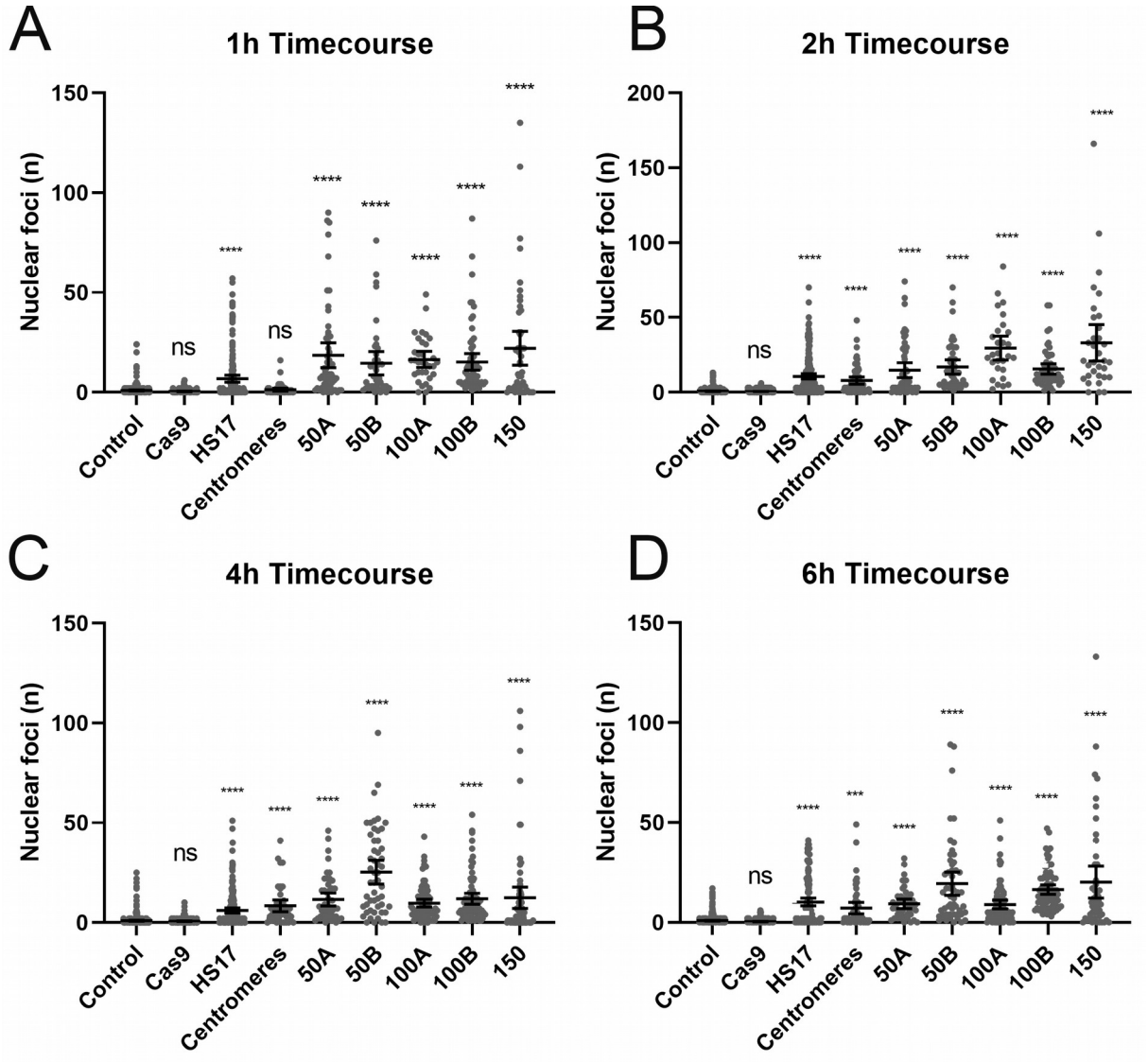


Figure 24: Quantification of nuclear γH2AX foci number for HS17, centromere, 50A, 50B, 100A, 100B and 150 transfections in respect to individual time course conditions.

Traced DNA damage foci formation for the **(A)** 1h, **(B)** 2h, **(C)** 4h and **(D)** 6h time course conditions for each of the transfection conditions. Cas9 without gRNA is not significant in all time points. Additionally, Cas9+Centromere does not show any significant foci formation compared to the control. Error bars represent 95% CI, and middle points represent the mean. One way ANOVA with Tukey's multiple comparisons test was performed to asses the significance of the transfection condition compared to the control (electroporation only). *ns*=not significant, *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$, **** $P \leq 0.0001$

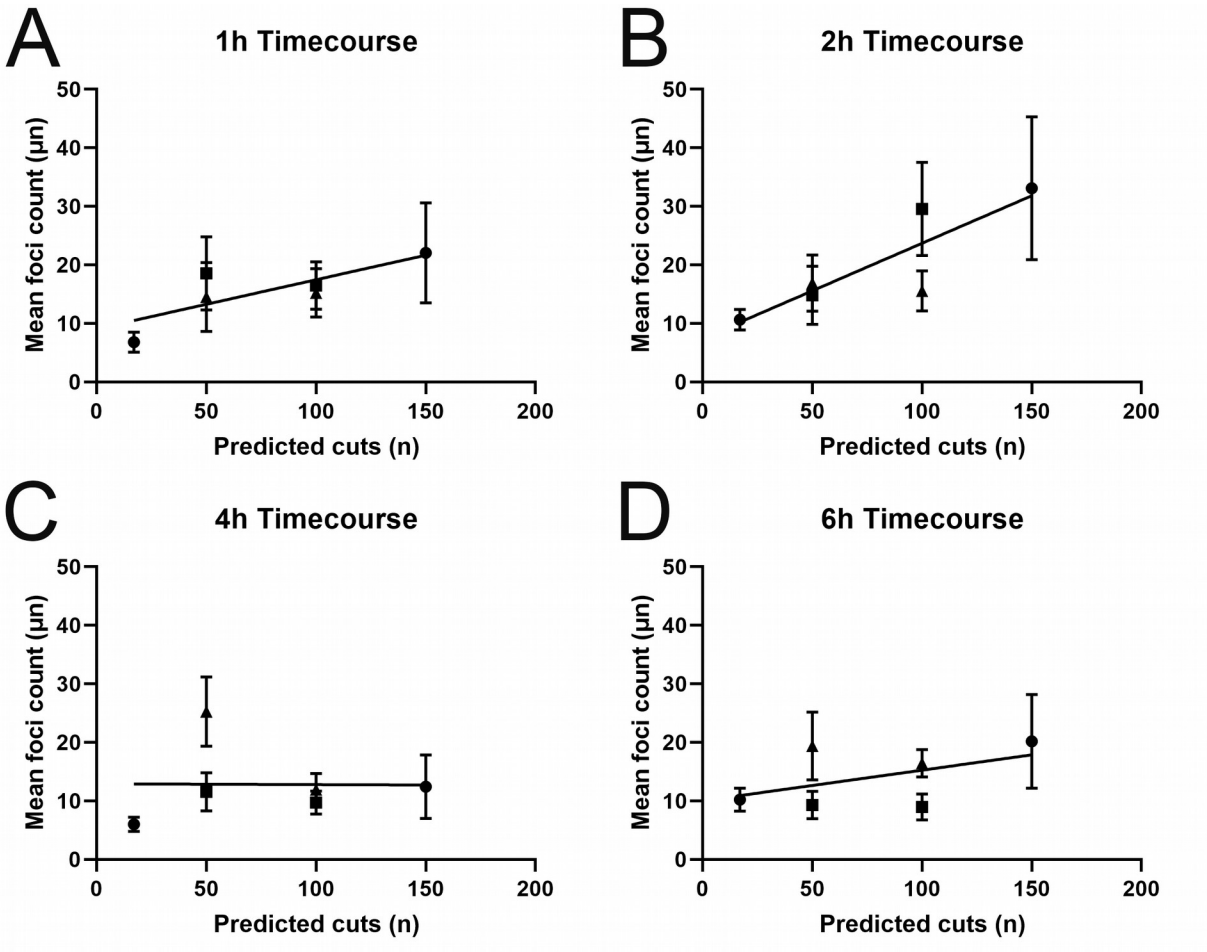


Figure 25: Summary of mean γH2AX foci particle count for HS17, 50A, 50B, 100A, 100B and 150 gRNAs based on the predicted *in silico* cuts in respect to individual time course conditions.

Linear correlation summary plots of mean foci count and predicted *in silico* cuts for the **(A)** 1h, **(B)** 2h, **(C)** 4h and **(D)** 6h time course conditions based on Figure 22. All conditions show positive correlation except the 4h, which shows zero correlation. Error bars represent 95% CI and points represent the mean foci count. Circle (●)= Non GC/AT-rich targeting gRNAs (HS17,150), Square (■) = GC-rich targeting gRNAs (50A, 100A), Triangle (▲) = AT-rich targeting gRNAs (50B, 100B).

3.3.4 Widefield microscopy IF nuclear intensity quantification

The first set of immunofluorescence widefield microscopy analysis investigated the quantitative aspect of DNA damage through particle analysis in regards to γH2AX. Unfortunately, it is not possible to take into account all of the possible γH2AX foci in a given cell due to the technical limitations of the widefield microscopy. Optical microscopy is limited to one focal plane of focus, which is prone to out-of-focus blur. This may potentially blur out foci not accounted during the particle quantification process. Acknowledging this limitation, this project made use of measuring the pixel value intensity (PVI) of the TRITC channel that includes the γH2AX in greyscale (8-bit integer) by designating an overall mean value based on the brightness of the selected pixels from 0 (black) to 255 (white). The main objective of this measurement is to take into consideration the blurred background fluorescence feedback from γH2AX. This is because nuclear intensity measurement can work better since γH2AX foci are designated units of damage capable of luminous emission of light (due to fluorescence) which are additive.

In a similar manner to the nuclear foci particle quantification, PVI also produced γH2AX overview plot (Figure 26), normalised percentage distribution frequency (Figure 27) and quantification (Figure 28) and summary plots (Figure 29). There were no significant differences between foci particle quantification and the foci PVI plots in regards to overall trends and comparisons, with certain exceptions. The Cas9+Centromeres transfection is not statistically significant for both the 1h and 4h post-transfection incubation periods (Figure 28) instead of the 1h only (Figure 24). Nonetheless, the linear correlations between the mean PVI and predicted cuts are

positive for all time course conditions (Figure 29) except for the 4h, which demonstrates a negative correlation (Figure 29C). In contrast, the linear correlation for the 4h foci particle condition shows no correlation (Figure 25C).

However, a closer inspection revealed that the 150 gRNA transfection could be considered an outlier in the majority of the nuclear intensity quantification cases since it lies in an abnormal distance from the other values of the dataset (Figure 29). For that reason, the correlation line was computed not involving any outliers for any time course (Figure 29E, 29F 29G and 29H). Comparing Figure 25 and Figure 29 that possible outliers happen in certain conditions (i.e. 100's and 150 gRNAs) because γH2AX foci formation occurred relatively rapidly, peaking shortly after MCF10A cells incubated with Cas9:gRNA for the first 2h, followed by a decline during the 4h and 6h of incubation. This could be a result of extensive initial repair foci formation at 2h.

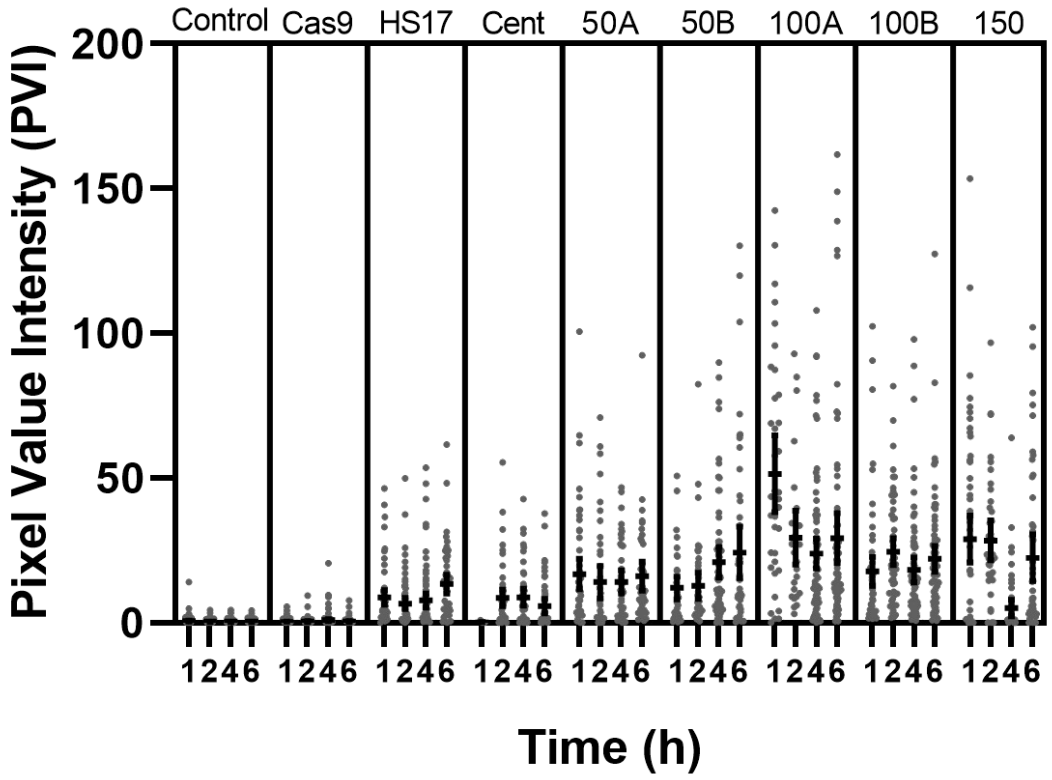


Figure 26: Overview of the γH2AX foci nuclear intensity quantification for control, Cas9 only and Cas9 with HS17, centromere, 50A, 50B, 100A, 100B and 150 gRNAs for all of the time course conditions.

Quantification of the overall pixel value intensity of the γH2AX nuclear foci derived from combined data of independent experiments for each condition and relevant time course which they show DNA damage accumulation. The figure demonstrates the accumulation of γH2AX for the conditions utilised the HS17, centromere, 50A, 50B, 100A, 100B and 150 gRNAs. Nuclear intensity measurement was utilised to measure the γH2AX fluorescence intensity, including the ones outside the desired single focal plane of the widefield microscope. Error bars represent 95% CI, and middle points represent the mean. Nuclear intensity measurement was done in Fiji.

Control: 1h **(n=198)**, 2h **(n=193)**, 4h **(n=159)**, 6h **(n=222)**
Cas9: 1h **(n=216)**, 2h **(n=201)**, 4h **(n=216)**, 6h **(n=194)**
Cas9+HS17: 1h **(n=70)**, 2h **(n=83)**, 4h **(n=82)**, 6h **(n=55)**
Cas9+Centromere: 1h **(n=54)**, 2h **(n=56)**, 4h **(n=48)**, 6h **(n=51)**
Cas9+50A: 1h **(n=57)**, 2h **(n=45)**, 4h **(n=44)**, 6h **(n=45)**
Cas9+50B: 1h **(n=43)**, 2h **(n=45)**, 4h **(n=64)**, 6h **(n=50)**
Cas9+100A: 1h **(n=36)**, 2h **(n=28)**, 4h **(n=84)**, 6h **(n=70)**
Cas9+100B: 1h **(n=69)**, 2h **(n=59)**, 4h **(n=74)**, 6h **(n=77)**
Cas9+150: 1h **(n=66)**, 2h **(n=40)**, 4h **(n=57)**, 6h **(n=48)**

Note: The number of nuclei shown here also apply to the data shown in subsequent Figures 27, 28 and 29. Data presented here are from one experiment each except for Cas9+HS17 which was repeated 3 times.

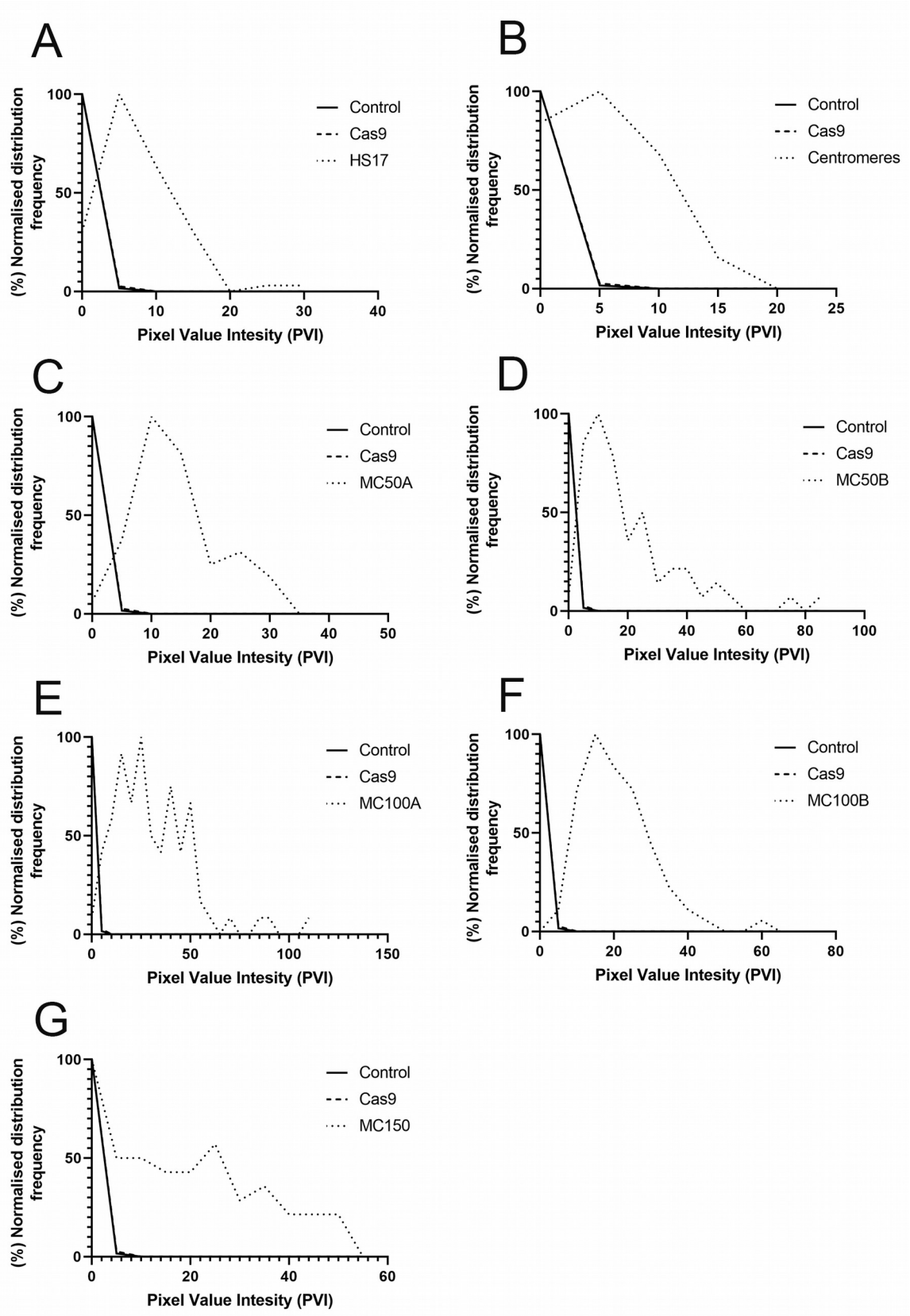

A
(%) Normalised distribution frequency
Pixel Value Intesity (PVI)
Control
Cas9
HS17
B
Centromeres
C
MC50A
D
MC50B
E
MC100A
F
MC100B
G
MC150

Figure 27: Normalised percentage distribution frequency of nuclear intensity of γH2AX foci (PVI) for HS17, centromere, 50A, 50B, 100A, 100B and 150 transfections compared to the control and Cas9 only conditions.

All time courses have been pooled together to assess the distribution of the data compared to the control and Cas9 only transfection for **(A)** HS17, **(B)** Centromeres, **(C)** 50A, **(D)** 50B, **(E)** 100A, **(F)** 100B, **(G)** 150. The figure shows that in average 5 PVI in a given nucleus may be considered background noise based on the control and Cas9 repetitions (histogram skewed to the right). This come in consensus with the particle quantification histogram in regards to the background noise. The bin size for histograms was configured to five.

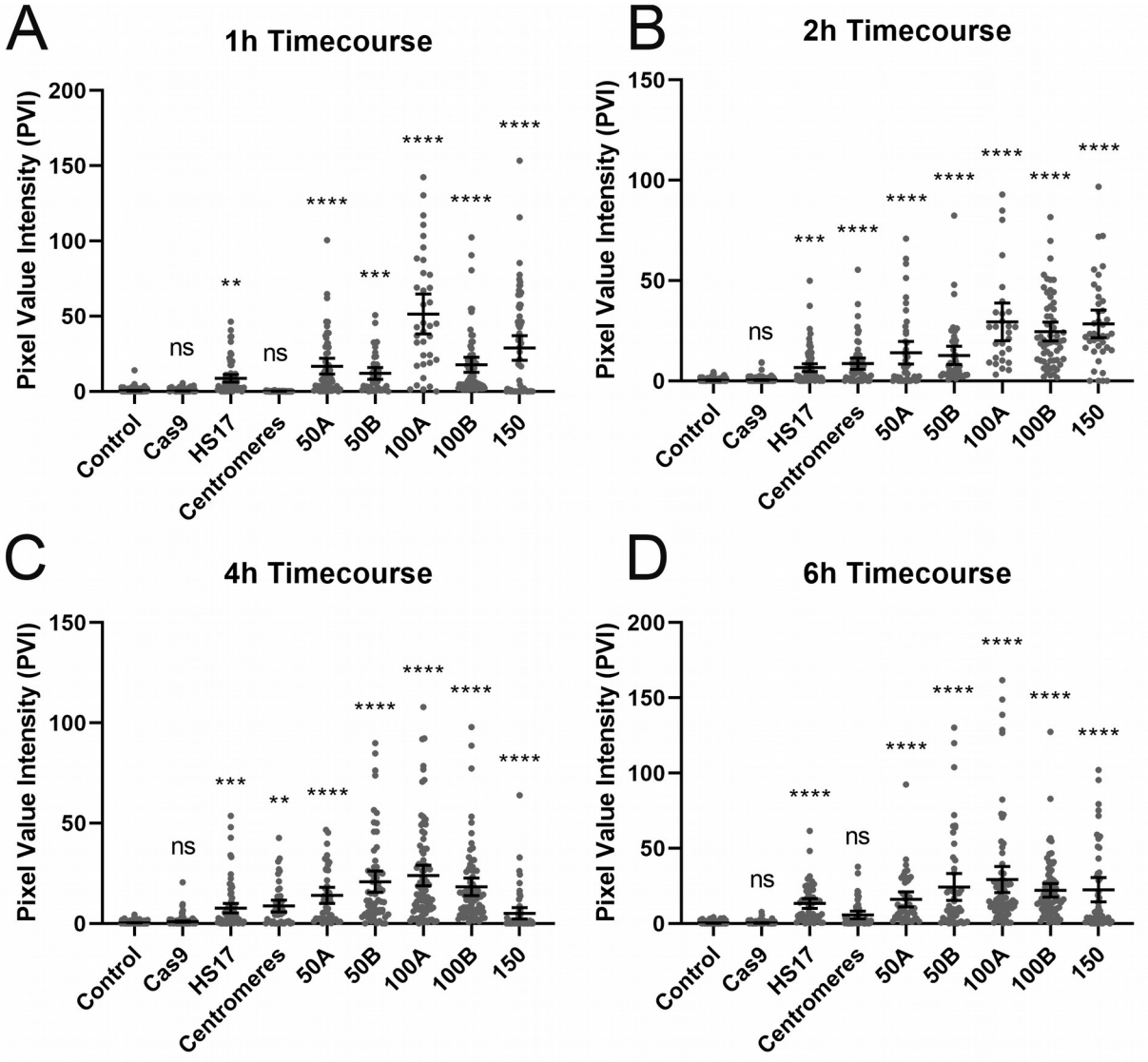


Figure 28: Quantification of nuclear intensity γH2AX foci (PVI) for HS17, centromere, 50A, 50B, 100A, 100B and 150 transfections in respect to individual time course conditions.

Traced DNA damage foci formation for the **(A)** 1h, **(B)** 2h, **(C)** 4h and **(D)** 6h time course conditions for each of the transfection conditions. Cas9 without gRNA is not significant in all time points. Similarly to the particle quantification in Figure 24, Cas9+Centromere does not show any significant foci formation compared to the control. Error bars represent 95% CI, and middle points represent the mean. One way ANOVA with Tukey's multiple comparisons test was performed to asses the significance of the transfection condition compared to the control (electroporation only). *ns*=not significant, *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$, ****$P \leq 0.0001$

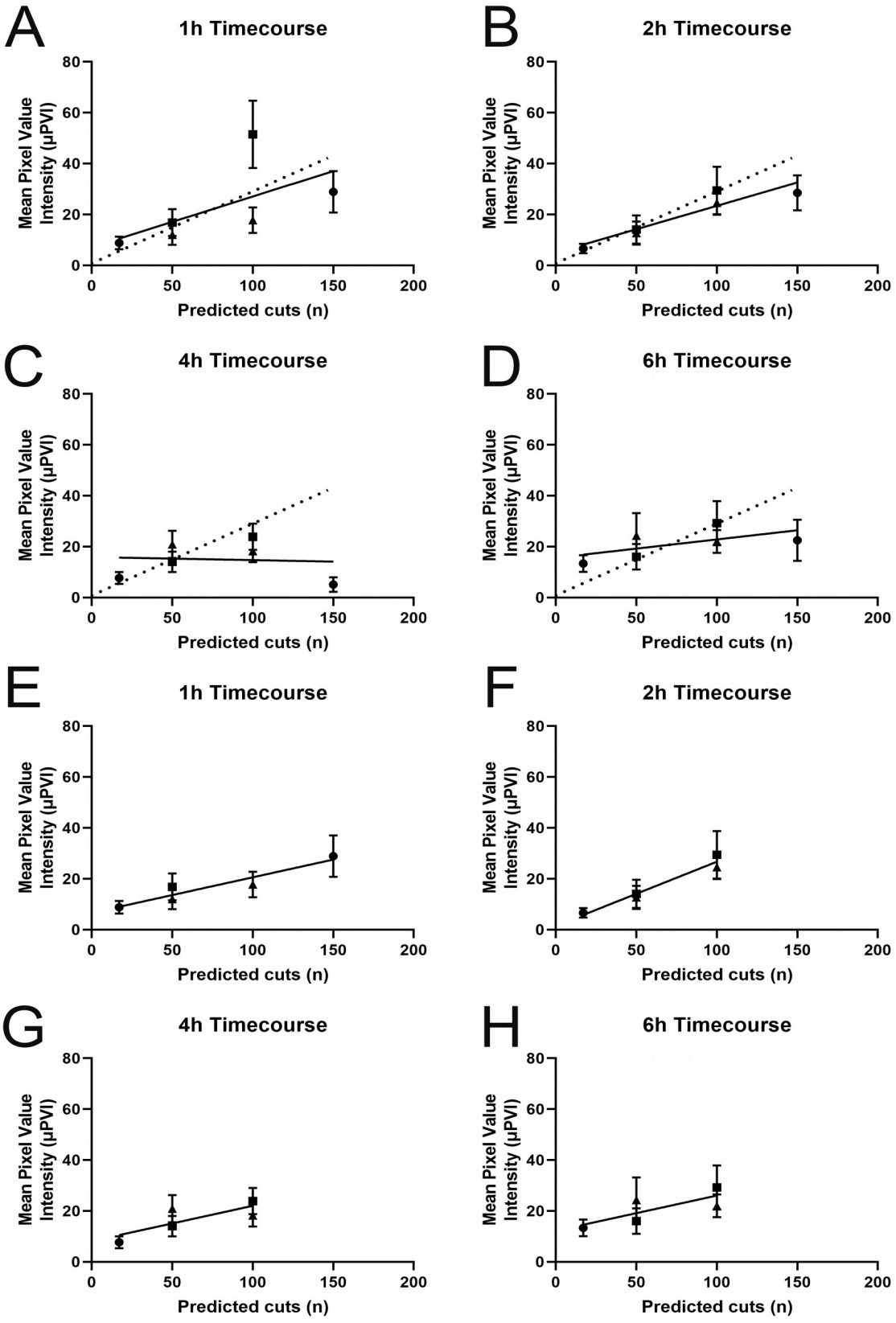

A
1h Timecourse
B
2h Timecourse
C
4h Timecourse
D
6h Timecourse
E
1h Timecourse
F
2h Timecourse
G
4h Timecourse
H
6h Timecourse
Mean Pixel Value Intensity (µPVI)
Predicted cuts (n)
0
20
40
60
80
0
50
100
150
200

Figure 29: Summary of mean nuclear intensity γH2AX foci for HS17, 50A, 50B, 100A, 100B and 150 gRNAs based on the predicted *in silico* cuts in respect to individual time course conditions.

Linear correlation summary plots of mean nuclear intensity and predicted *in silico* cuts for the **(A)** 1h, **(B)** 2h, **(C)** 4h and **(D)** 6h time course conditions based on Figure 22. All conditions show positive correlation except the 4h, which shows a negative correlation. In addition, **(E)** 1h, **(F)** 2h, **(G)** 4h and **(H)** 6h shows adjusted linear correlations not taking into consideration the outlier data points. Linear correlation for a given dataset is indicated as a solid straight line (**—**). Straight linear start trend lines are indicated as (····) to assess possible outliers on the dataset. Error bars represent 95% CI and points represent the mean foci count. Circle (●) = Non GC/AT-rich targeting gRNAs (HS17,150), Square (■) = GC-rich targeting gRNAs (50A, 100A), Triangle (▲) = AT-targeting gRNAs (50B, 100B).

3.3.4 Confocal microscopy IF qualitative and quantitative data

Induction of DSBs under widefield microscopy leads to the formation of micrometre-sized DNA-repair foci unable to take into account the organisation on the nanometre scale due to its diffraction limitation at ~200nm. This project provides considerable insight into the DNA damage generation in the MCF10A nuclei from the FlashFry designed sequences with the use of laser confocal microscopy (Figure 30). Based on the widefield microscopy results, the maximum extent of γH2AX foci number is well conserved within the 2h hour incubation. Additionally, the blur formation from widefield microscope obfuscates potential γH2AX foci out of focus due to the single focal plane. For these reasons, confocal was chosen to visualise (Figure 30) and quantify the DNA damage at 2h to address the limitations of widefield microscopy (Figure 31). Both fixation and immunostaining protocols remained the same as the widefield.

The most remarkable result to emerge from the data is that the greater resolution and increased signal-to-noise ratio of the super-resolution microscopy show that the control demonstrates minimal fluorescence feedback from the Alexa Fluor 555 red γH2AX conjugated antibody activity (Figure 30). There are two possible explanations for this observation. Either the antibody cross-reacts with other nuclear structures, or there is small H2AX phosphorylation activity taking place for any of the transfection conditions (Figure 30). Additionally, in Figure 30, MCF10A cells were transfected with 50A, 50B, 100A, 100B and 150 gRNAs and foci can be observed with the higher resolution under the z-stack (Figure 30). Both large and small foci (nanofoci) particles can be observed within the MCF10A nuclei for all of the

transfected conditions (Figure 30). Both merged and individual DAPI (nucleus - blue), and TRITC (γH2AX - red) channels are provided to contrast the induced DNA damage (Figure 30) better. No significant visual difference was observed between the transfected conditions in regards to the number H2AX foci formation, and thus quantification of the merged z-stack was employed.

The damage was quantified in regards to nuclear foci number (Figure 31A), total foci area covered within the MCF10A nuclei measured (Figure 31B), average individual foci area (Figure 31C) and nuclear intensity (Figure 31D). Statistical significance refers to the comparisons among conditions using one-way ANOVA test that makes use of Tukey's multiple comparison method. In the confocal microscopy, the number of γH2AX foci is much more numerically close to the predicted cuts compared to the widefield and only one instance it matches exactly. The 50A and 50B transfections have on average 50.4, and 31.6 γH2AX foci, respectively, whereas 100A and 100B have 58.3 and 63.5 γH2AX foci on average (Figure 31A). The 150 gRNA transfection produces 62.1 foci on average (Figure 31A). The statistical test states foci count significance only between 50B vs 100A, 50B vs 100B, and 50B vs 150 conditions in Figure 31A. All conditions except for the AT-rich targeting gRNAs 50B and 100B show bimodality in regards to foci count (Figure 31A). This may suggest that the rest of the cells are in different stages of arrest during DNA damage repair.

Likewise, wide dataset distributions can also be observed for the quantified total foci area covered within the MCF10 nuclei (Figure 31B). Specifically, only 50A and 50B are significant to the 100A in regards to total foci area (Figure 31B). Apart

from that, no significant difference was found among the different conditions in terms of the predicted cuts, i.e. 50A/50B vs.100A/100B vs 150 (Figure 31B). Despite the higher predicted cuts, the total foci area within the nuclei remains mostly the same for the merged z-stacks at the 2h interval. In contrast, the average foci area per MCF10A nuclei at increasing targeted cuts do not show bimodality distribution, and the conditions do not express any statistical significance with each other (Figure 31C). No statistical significance among the transfection conditions can also be seen in the context of nuclear intensity, but bimodality can be observed among conditions in 2h post-transfection (Figure 31D).

Overall, the experiment in this project is in good agreement with the current literature (such the one from Pilch *et al.*, Bewersdorf *et al.*, Sak *et al.*) as the diameter of γH2AX foci in super-resolution fluorescence microscopy is within 0.5-1 μm (51, 187, 188). However, nanofoci particles have also been detected that are less than 0.5 μm in diameter correlating favourably with the research done by Sisario *et al.* (189).

Summary of the distribution of the nuclear foci data from the confocal microscopy is shown in Figure 32. Interestingly, higher than predicted values of nuclear foci detected only for 50A, which may suggest these particular cells arrested in the G_2/M checkpoint with the duplicated genome. This forms the characteristic “dual peaks” in the normalised percentage distribution in Figure 32A, and creating the characteristic bimodal necks in the violin plots in Figure 31A. Similar observations can be seen in 100A and 150 transfection conditions (Figure 31A). Only AT-rich targeting gRNAs do not demonstrate distribution dual peaks (Figure 30) and have a

much more refined non-bimodal data distribution as per Figure 31A. Figure 32 was configured with a bin width of 10.

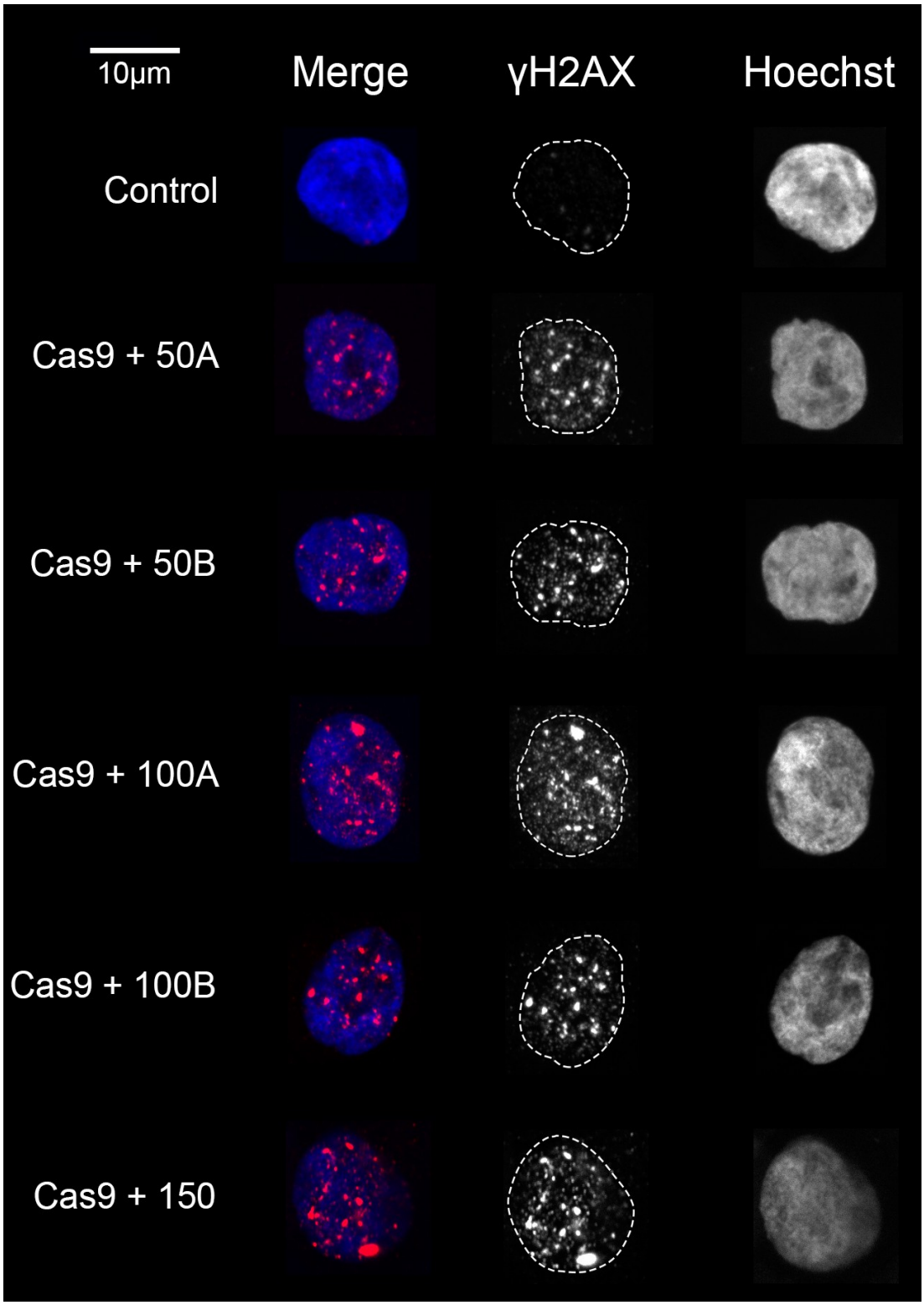

10μm
Merge
γH2AX
Hoechst
Control
Cas9 + 50A
Cas9 + 50B
Cas9 + 100A
Cas9 + 100B
Cas9 + 150

Figure 30: Confocal immunofluorescence with merged orthogonal Z-stack shows the localisation of γH2AX foci in greater resolve in stained MCF10A nuclei utilising the 50A, 50B, 100A, 100B and 150 gRNAs with active Cas9 (only for 2h time course condition).

The utilisation of 50A, 50B, 100A, 100B and 150 gRNAs shows the formation of γH2AX foci in merged confocal planes only for the 2h condition. Low activity of γH2AX can be observed in the control in the split red channel. A total of 6-10 z-stack series have been captured per image. To form the gRNAs, the 50A, 50B, 100A, 100B and 150 crRNAs combined with the Dharmacon Edit-R CRISPR-Cas9 Synthetic tracrRNA. Nuclei fixed with 4% (w/v) PFA and quenched with with 50mM NH_4Cl/TBS solution. Nuclei blocked with 2% (w/v) BSA and stained with Hoechst 33342 (1:12300) anti-γH2AX-Ser139 (1:500) primary antibody. Anti-γH2AX-Ser139 labelled with Alexa Fluor 555 dye-conjugated secondary mouse antibody (1:500). Imaging was done on ZEISS Elyra 7 using polarised laser modules. Imaging processing was done in Fiji and Zeiss ZEN blue. Dotted lines represent the extent of the nuclei for each transfection condition based on the nuclei periphery.

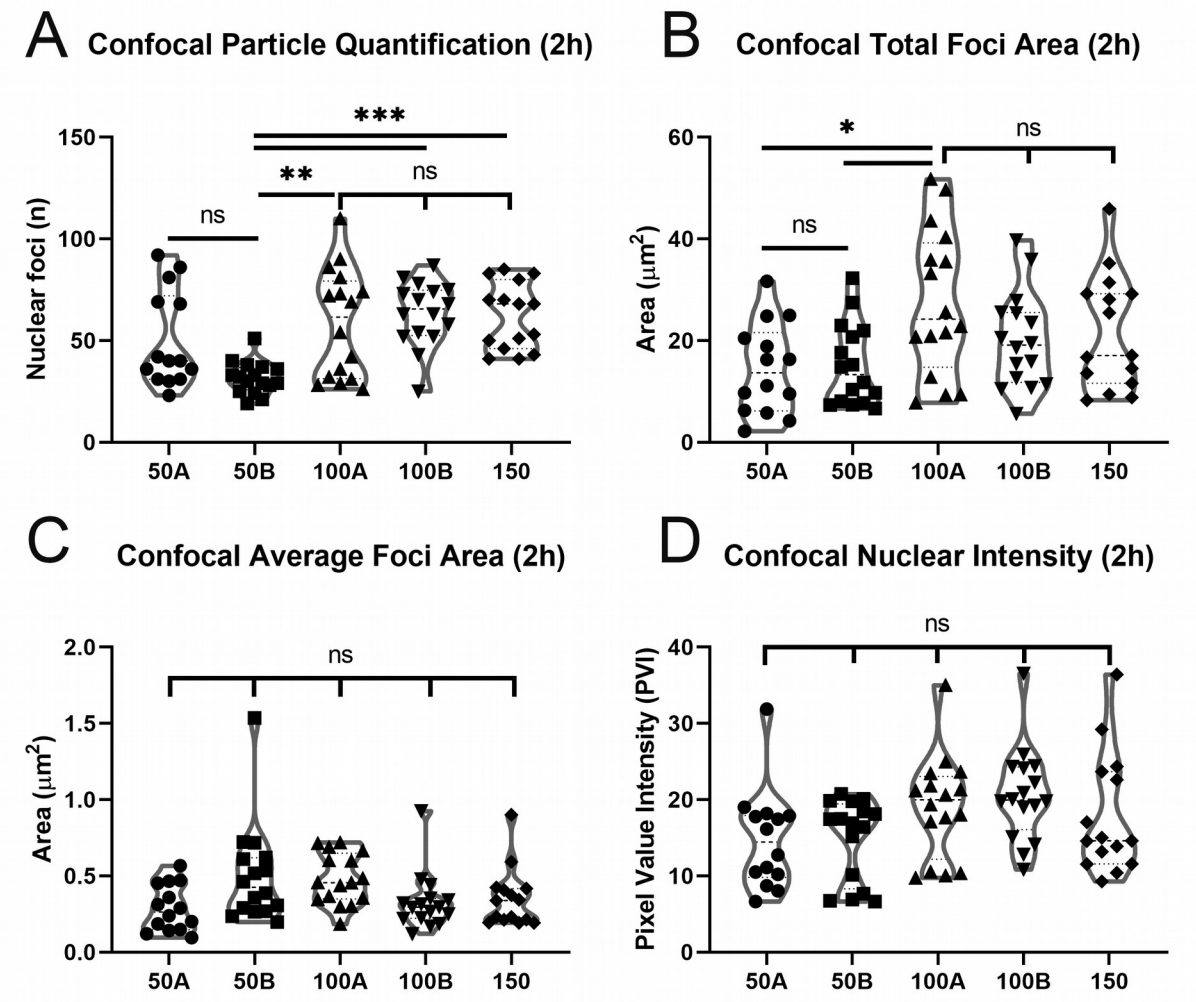


Figure 31: Overview for the confocal imaging at 2h incubation for particle quantification, total and average foci area and nuclear intensity for the transfected 50A, 50B, 100A, 100B and 150 gRNAs.

The confocal microscopy and its higher resolve allow the quantification to take into consideration the multiple focal planes from the merged z-stack. **(A)** Accumulation of DNA damage is observed per increasing prediction gRNA hits but reaches a stage where there is no further change on the measured foci number for the 100's and 150 cuts gRNAs. All conditions except 50A and 100A show bimodality. **(B)** Despite the increase of targeted cuts and the number of foci in the MCF10A nuclei, the total area of MCF10A foci per cell did not change during the 2h interval post Cas9:gRNA transfections **(C-D)** When comparing all conditions with each other in regards to **(C)** average individual foci area and **(D)** nuclear intensity, they are not significant. However, only the latter shows bimodal distribution. One way ANOVA with Tukey's multiple comparisons test was performed to asses the significance of the transfection conditions with each other. *ns*=not significant, *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$, ****$P \leq 0.0001$. Solid dashed lines (---) represents the median, and grey dotted lines (...) shows the 25th and 75th percentiles in the violin plots. 50A **(n=14)**, 50B **(n=16)**, 100A **(n=16)**, 100B **(n=16**), 150 **(n=15)**.

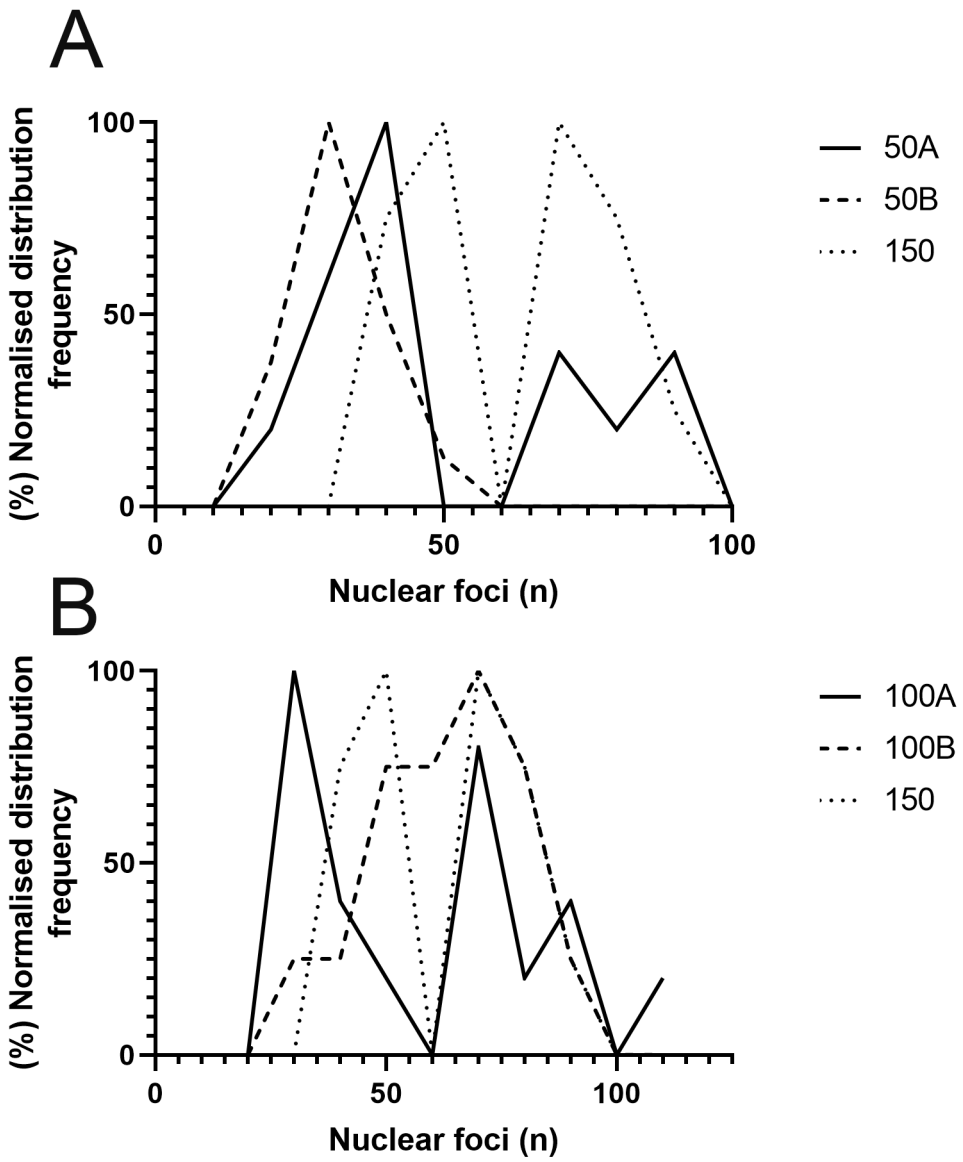


Figure 32: Normalised percentage distribution frequency of nuclear γH2AX foci number for 50A, 50B, 100A, 100B and 150 transfections.

The distribution shows the maximum extent of the nuclear foci counts for any of the conditions is no more than 100. All transfections except for AT-rich targeting 50B and 100B gRNAs show double peaks. The bin size for histograms was configured to ten.

4. Discussion

This project describes that by using a programmable Cas9 system along with tractable γH2AX immunostaining is possible to induce titratable DNA damage (DSBs) which is backed by a multitude of data, including but not limited to bioinformatics, widefield and confocal microscopy. This project is an expansion upon the work of Berg *et al.* in an attempt to establish a functional proof of concept for designing gRNAs that have multiple recognitions points for DSBs intervention at lower logistical cost. Although the performance was optimal in regards testing and verifying the programmable DNA damage induction, this project is governed by limitations that need to be addressed for future improvements.

4.1 Recombinant Cas9 protein purification

Protein samples were collected at every stage of the purification process and subsequently analysed using SDS-PAGE. Both chromatograms from ion exchange and then size exclusion chromatography shows efficient Cas9 purification with high protein yields. SDS-PAGE also highlights that major contaminants have been excluded from the main protein successfully. This suggests the protocols followed here can express and purify the recombinant protein in high yields. Most importantly, the two-step purification protocol followed in this project does not seem to affect the protein in any way according to γH2AX qualitative data gathered from the microscopy. Nevertheless, potential improvements for this part of the project can utilise an independent *in vitro* nuclease assays to further confirm the functionality of

the purified Cas9 protein by using known gRNAs against *in vitro* DNA target in a similar manner to Rajagopalan *et al.* (190). Then, the competent functionality of Cas9 can be confirmed by analysing the DNA products using agarose gel electrophoresis (190).

4.2 CRISPR bioinformatics

FlashFry was the primary computational tool used to assist for the design and scoring of the crRNA sequences capable of multiple DNA damage interventions. FlashFry allows one to have more granular control on how to design the CRISPR targets compared to online web interfaces such as CRISPOR. This project was able to compute competent crRNA sequences with multiple recognition points. However, despite the advancement in the CRISPR bioinformatics landscape and the verified controlled DNA damage induced in the MCF10A cells' nuclei, the designing of efficient multiple targeting gRNAs remains an area for further improvement.

Firstly, it is crucial to refine crRNA sequences to increase their on-targeting and limit off-targeting activity. Future work can focus on further refining the crRNA sequences by imposing more strict filtering rules in the scoring metrics of the produced set of sequences. Specifically, future investigations can take into consideration the flanking bases and calculate the on-targeting activity in regards to multiple points of intervention. In other words, investigate whether the multiple targeting site itself varies or it is part of a bigger sequence repeat. This was something not able to calculate during this project with the doench2014ontarget

metric. Future work should also take into consideration the broad distribution of cuts per chromosome. In this project, 50A and 50B gRNAs end up favouring cuts into smaller chromosomes in terms of cuts per Mbp. Design effort should look at a wider variety of gRNA sequences in the context applying DSBs uniformly and in different regions levels (intronic or exonic).

Additionally, Bowtie and BWA tools can be used instead of BLAST in order to verify the number of off-target hits and assess the off-targeting activity (like in doench2016 score metric) (177). These tools are better suited at handling short sequences compared to other traditional tools such as BLAST. BLAST was designed to assess the similarity between longer sequences adequately. Sequences with a lower number of base pairs are prone to hit random background noise in BLAST algorithms. However, it is unknown how Bowtie and BWA tools will handle gRNAs with many recognition points and how they will score the newly designed sequences.

Another strategy to utilise in future work is to use dual gRNAs to facilitate multiple DNA damage in an additive way in a similar manner to Zhou *et al.* which could increase the efficiency of DNA damage induction (155). This way, by designing gRNAs with lesser cuts and thus with lesser off-targeting activity, it might be possible to make more efficient DNA damage inductions.

It is important to state that certain limitations govern the bioinformatics aspect of this project in regards to our current understanding of driver variables of CRISPR-Cas9 activity and technical ones, as most of the tools today have been designed with the perspective of single gene sequence knockouts.

4.3 Cas9 *in vitro* delivery

Originally, dCas9 was used to assess its delivery into the MCF10A cells and subsequently optimise the transfection protocol. This allowed to deliver the functional Cas9:gRNA complex through electroporation at 300V and 300µF effectively to carry out the anticipated multiple DNA damage. The electrotransfer delivery of large proteins such as Cas9 has high versatility and can be applied to almost any cell line without any prior cell line engineering. While this method is quick, cheap and capable to transfect eukaryotic cells, there may be potential limitations using it. The protocols followed in the project achieved 40% transfection rates of dCas9 in the MCF10A cell line.

According to the distribution frequency of the widefield microscopy both the controls (electroporated only cells) and Cas9 only conditions exhibit between 0 to 6 foci within the MCF10A nuclei, albeit the 40% transfection efficiency. These DSBs foci may be yielded due to electroporation apart from the γH2AX anti-body cross-reaction. Electroporation could stress the cells through the application of current and voltage in the Opti-MEM solution containing the MCF10A cells. Additionally, limited γH2AX activity can be observed in the cells of the control group in the confocal microscopy. Potentially, electroporation protocol causes some unwanted DNA damage in the MCF10A cells. However, based on the data presented in this project, the unwanted DNA damage may be limited. This could be a problem if future work wants to quantify and verify the DNA damage or study specific protein-DNA interactions with much more refined gRNAs. Such quantification can be achieved through ChIP-seq or with its cheaper alternative, CUT&RUN sequencing (191–194).

Additionally, only a small and sporadic number of cells were apoptotic (whole MCF10A nucleus was positive to γH2AX staining) using the electroporation protocol method used here.

Future work could utilise the iCut system used in Berg *et al.* study to induce multipoint DNA damage without electroporation (180). The iCut system allows one to tune the expression levels of the Cas9 (through pCW-Cas9 plasmid such as the one from Eric Lander and David Sabatini addgene plasmid repository #50661) in both the transcription and protein level through doxycycline and SHIELD-1 agonists respectively inside the cell line of interest (180, 195). This utilisation of this system will give control of the amount of Cas9 residing inside the cells, something which is not possible through electroporation. The application of defined electrical field in a cell solution does not guarantee a uniform amount of protein transfection. Nevertheless, the engineering of each cell line with the iCut system can increase lab costs.

4.4 Immunostaining and microscopy

Immunostaining was an integral part of this project to visualise DNA damage by detecting γH2AX repair foci by Alexa Fluor 555 secondary antibody fluorescence feedback. The majority of qualitative and quantitative data of transfected MCF10A cells derived from widefield microscopy. They are consistent with DSBs observation by previous studies using either non-specific induction (i.e. ionising radiation) or specific DNA damage induction using the Cas9 system in a similar manner like in the

study from Berg *et al.* (1, 180). This project also shows the first demonstration of targeted induction of DSBs using confocal microscopy, which shows a clearer depiction of γH2AX nuclear localisation, including nanofoci.

Based on the data from widefield microscopy, the kinetics of Ser139 H2AX phosphorylation occurs following at least an hour post Cas9:gRNA transfection except for Cas9 transfection utilising the centromere gRNA. This condition showed its first foci formation during the 2h incubation at 37ºC which may occur due to the dense packing of this chromatin. The maximal γH2AX foci induction is observed within the 2h followed by a progressive decrease from the four hours and onwards post Cas9:gRNA complex transfection. Data comparison from the mean number of foci and mean intensity over time suggest a tendency towards γH2AX foci diffusion as a result of extensive foci repair and fusion at the 2h incubation.

These γH2AX kinetic observations bear close resemblance on the results from Ding *et al.* which they irradiated (X-rays) in human peripheral blood lymphocytes (HPBLs) and showed maximal γH2AX foci at 2h post-exposure (196). Other studies using UVB induce maximal γH2AX foci formation at 30min after exposure (50). In this project's results is important to take into consideration the kinetics of Cas9 importation to the nucleus through the NLS signalling. Future work could utilise more time course conditions, including half-hours, to better understand the kinetics of Cas9 based DNA damage inductions.

Additionally, future work could benefit from scanning additional images using confocal microscopy in different post-transfection time points and not only 2h. This is

useful to study the kinetics of inducible DNA damage at higher resolution images. This can be aided by performing 3D foci counting utilising tools such as FociPicker3D which do not require merging the z-stack and thus collapsing underneath foci (197). Collapsing foci beneath each other through z-stack merging could be a source of error for the quantification process. The third dimension could also help for imaging the chromatin experiencing specific DNA damage induced by the Cas9 assays. Confocal microscopy data can be backed up by semi quantifying the γH2AX through western blotting at increasing DNA damage from the corresponding gRNAs similarly to Revet *et al.* and Bouquet *et al.* studies (198, 199).

4.5 Conclusions

To recapitulate, while this project clearly has some limitations such as in CRISPR designing and the γH2AX quantification of multiple foci, the evidence presented in this report has demonstrated its feasibility and sets a springboard for further development and improvement of the overall pipeline.

Yet the most intriguing finding of this project has been found through the γH2AX immunostaining in conjunction with the confocal microscopy. Active Cas9 with custom-designed gRNA for multiple genome interventions is capable of producing clusters of nanofoci present in the majority of the transfected MCF10A cells. This observation has been only shown in cells induced by non-Cas9 methods such as irradiation. Nonetheless, the designing of gRNAs with unique properties such

as with multiple recognition points as presented here is becoming more accessible and granular than ever before.

This project has led to the formation of a new pipeline that is capable of designing Cas9 assays inducing titratable DSBs at defined positions in the human genome. The amount of DNA damage can be regulated with the proper designing of the crRNA sequence by using computational tools such as FlashFry. Taken together, such procedure under the proper improvements and optimisations can be suitable to study the epigenetic response in the context of chromatin.

5. References

1. Jackson, S.P., and J. Bartek. 2009. The DNA-damage response in human biology and disease. Nature. 461: 1071–8.
2. Lindahl, T., and D.E. Barnes. 2000. Repair of endogenous DNA damage. Cold Spring Harb. Symp. Quant. Biol. 65: 127–33.
3. Bernstein, C., A. R., V. Nfonsam, and H. Bernstei. 2013. DNA Damage, DNA Repair and Cancer. In: New Research Directions in DNA Repair. InTech.
4. Hustedt, N., and D. Durocher. 2016. The control of DNA repair by the cell cycle. Nat. Cell Biol. 19: 1.
5. Taylor, S.S., M.I.F. Scott, and A.J. Holland. 2004. The spindle checkpoint: a quality control mechanism which ensures accurate chromosome segregation. Chromosome Res. 12: 599–616.
6. Meraldi, P., V.M. Draviam, and P.K. Sorger. 2004. Timing and Checkpoints in the Regulation of Mitotic Progression. Dev. Cell. 7: 45–60.
7. Sedelnikova, O.A., I. Horikawa, D.B. Zimonjic, N.C. Popescu, W.M. Bonner, and J.C. Barrett. 2004. Senescing human cells and ageing mice accumulate DNA lesions with unrepairable double-strand breaks. Nat. Cell Biol. 6: 168–170.
8. Rübe, C.E., A. Fricke, T.A. Widmann, T. Fürst, H. Madry, M. Pfreundschuh, and C. Rübe. 2011. Accumulation of DNA Damage in Hematopoietic Stem and Progenitor Cells during Human Aging. PLoS One. 6: e17487.
9. Hoeijmakers, J.H.J. 2009. DNA Damage, Aging, and Cancer. N. Engl. J. Med. 361: 1475–1485.
10. Collado, M., M.A. Blasco, and M. Serrano. 2007. Cellular Senescence in Cancer and Aging. Cell. 130: 223–233.
11. Best, B.P. 2009. Nuclear DNA Damage as a Direct Cause of Aging. Rejuvenation Res. 12: 199–208.
12. Nowsheen, S., and E.S. Yang. 2012. The intersection between DNA damage response and cell death pathways. Exp. Oncol. 34: 243–54.
13. Kaufmann, W.K., and R.S. Paules. 1996. DNA damage and cell cycle checkpoints. FASEB J. 10: 238–47.

14. Vilenchik, M.M., and A.G. Knudson. 2003. Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. Proc. Natl. Acad. Sci. U. S. A. 100: 12871–6.

15. Swenberg, J.A., K. Lu, B.C. Moeller, L. Gao, P.B. Upton, J. Nakamura, and T.B. Starr. 2011. Endogenous versus exogenous DNA adducts: their role in carcinogenesis, epidemiology, and risk assessment. Toxicol. Sci. 120 Suppl: S130-45.

16. Diffley, J.F.X. 2004. Regulation of early events in chromosome replication. Curr. Biol. 14: R778-86.

17. Bartek, J., and J. Lukas. 2003. Chk1 and Chk2 kinases in checkpoint control and cancer. Cancer Cell. 3: 421–429.

18. Ou, Y.-H., P.-H. Chung, T.-P. Sun, and S.-Y. Shieh. 2005. p53 C-terminal phosphorylation by CHK1 and CHK2 participates in the regulation of DNA-damage-induced C-terminal acetylation. Mol. Biol. Cell. 16: 1684–95.

19. Maréchal, A., and L. Zou. 2013. DNA damage sensing by the ATM and ATR kinases. Cold Spring Harb. Perspect. Biol. 5.

20. Pabla, N., S. Huang, Q.-S. Mi, R. Daniel, and Z. Dong. 2008. ATR-Chk2 Signaling in p53 Activation and DNA Damage Response during Cisplatin-induced Apoptosis. J. Biol. Chem. 283: 6572–6583.

21. Awasthi, P., M. Foiani, and A. Kumar. 2015. ATM and ATR signaling at a glance. J. Cell Sci. 128: 4255–4262.

22. Giglia-Mari, G., A. Zotter, and W. Vermeulen. 2011. DNA Damage Response. Cold Spring Harb. Perspect. Biol. 3: a000745–a000745.

23. Flynn, R.L., and L. Zou. 2011. ATR: a master conductor of cellular responses to DNA replication stress. Trends Biochem. Sci. 36: 133–140.

24. Woods, D., and J.J. Turchi. 2013. Chemotherapy induced DNA damage response: convergence of drugs and pathways. Cancer Biol. Ther. 14: 379–89.

25. Sancar, A., L.A. Lindsey-Boltz, K. Unsal-Kaçmaz, and S. Linn. 2004. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. Annu. Rev. Biochem. 73: 39–85.

26. Ciccia, A., and S.J. Elledge. 2010. The DNA damage response: making it safe to play with knives. Mol. Cell. 40: 179–204.

27. Yan, S., M. Sorrell, and Z. Berman. 2014. Functional interplay between ATM/ATR-mediated DNA damage response and DNA repair pathways in oxidative stress. Cell. Mol. Life Sci. 71: 3951–67.

28. Nakamura, J., V.E. Walker, P.B. Upton, S.Y. Chiang, Y.W. Kow, and J.A. Swenberg. 1998. Highly sensitive apurinic/apyrimidinic site assay can detect spontaneous and chemically induced depurination under physiological conditions. Cancer Res. 58: 222–5.

29. Lindahl, T., and B. Nyberg. 1972. Rate of depurination of native deoxyribonucleic acid. Biochemistry. 11: 3610–3618.

30. Lindahl, T. 1993. Instability and decay of the primary structure of DNA. Nature. 362: 709–715.

31. Burrows, C.J., and J.G. Muller. 1998. Oxidative Nucleobase Modifications Leading to Strand Scission. Chem. Rev. 98: 1109–1152.

32. Cooke, M.S., M.D. Evans, M. Dizdaroglu, and J. Lunec. 2003. Oxidative DNA damage: mechanisms, mutation, and disease. FASEB J. 17: 1195–214.

33. Ames, B.N., M.K. Shigenaga, and T.M. Hagen. 1993. Oxidants, antioxidants, and the degenerative diseases of aging. Proc. Natl. Acad. Sci. U. S. A. 90: 7915–22.

34. Helbock, H.J., K.B. Beckman, M.K. Shigenaga, P.B. Walter, A.A. Woodall, H.C. Yeo, and B.N. Ames. 1998. DNA oxidation matters: the HPLC-electrochemical detection assay of 8-oxo-deoxyguanosine and 8-oxo-guanine. Proc. Natl. Acad. Sci. U. S. A. 95: 288–93.

35. Foksinski, M., R. Rozalski, J. Guz, B. Ruszkowska, P. Sztukowska, M. Piwowarski, A. Klungland, and R. Olinski. 2004. Urinary excretion of DNA repair products correlates with metabolic rates as well as with maximum life spans of different mammalian species. Free Radic. Biol. Med. 37: 1449–54.

36. Tudek, B., A. Winczura, J. Janik, A. Siomek, M. Foksinski, and R. Oliński. 2010. Involvement of oxidatively damaged DNA and repair in cancer development and aging. Am. J. Transl. Res. 2: 254–84.

37. Haber, J.E. 1999. DNA recombination: the replication connection. Trends Biochem. Sci. 24: 271–5.

38. Reardon, J.T., and A. Sancar. 2006. Purification and characterization of Escherichia coli and human nucleotide excision repair enzyme systems. Methods Enzymol. 408: 189–213.

39. Hsu, D.-W., R. Kiely, C.A.-M. Couto, H.-Y. Wang, J.J.R. Hudson, C. Borer, C.J. Pears, and N.D. Lakin. 2011. DNA double-strand break repair pathway choice in Dictyostelium. J. Cell Sci. 124: 1655–63.

40. Harper, J.W., and S.J. Elledge. 2007. The DNA Damage Response: Ten Years After. Mol. Cell. 28: 739–745.

41. Bao, Y. 2011. Chromatin response to DNA double-strand break damage. Epigenomics. 3: 307–21.

42. Luger, K., A.W. Mäder, R.K. Richmond, D.F. Sargent, and T.J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature. 389: 251–260.

43. Luger, K., M.L. Dechassa, and D.J. Tremethick. 2012. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nat. Rev. Mol. Cell Biol. 13: 436–47.

44. Ramakrishnan, V., J.T. Finch, V. Graziano, P.L. Lee, and R.M. Sweet. 1993. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. Nature. 362: 219–223.

45. Misteli, T., A. Gunjan, R. Hock, M. Bustin, and D.T. Brown. 2000. Dynamic binding of histone H1 to chromatin in living cells. Nature. 408: 877–881.

46. Liu, B., R.K. Yip, and Z. Zhou. 2012. Chromatin remodeling, DNA damage repair and aging. Curr. Genomics. 13: 533–47.

47. Henikoff, S., and M.M. Smith. 2015. Histone variants and epigenetics. Cold Spring Harb. Perspect. Biol. 7: a019364.

48. Redon, C., D. Pilch, E. Rogakou, O. Sedelnikova, K. Newrock, and W. Bonner. 2002. Histone H2A variants H2AX and H2AZ. Curr. Opin. Genet. Dev. 12: 162–9.

49. Rogakou, E.P., D.R. Pilch, A.H. Orr, V.S. Ivanova, and W.M. Bonner. 1998. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J. Biol. Chem. 273: 5858–68.

50. Kinner, A., W. Wu, C. Staudt, and G. Iliakis. 2008. Gamma-H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin. Nucleic Acids Res. 36: 5678–94.

51. Pilch, D.R., O.A. Sedelnikova, C. Redon, A. Celeste, A. Nussenzweig, and W.M. Bonner. 2003. Characteristics of γ-H2AX foci at DNA double-strand breaks sites. Biochem. Cell Biol. 81: 123–129.

52. Stucki, M., J.A. Clapperton, D. Mohammad, M.B. Yaffe, S.J. Smerdon, and S.P. Jackson. 2005. MDC1 Directly Binds Phosphorylated Histone H2AX to Regulate Cellular Responses to DNA Double-Strand Breaks. Cell. 123: 1213–1226.

53. Rogakou, E.P., C. Boon, C. Redon, and W.M. Bonner. 1999. Megabase chromatin domains involved in DNA double-strand breaks in vivo. J. Cell Biol. 146: 905–16.

54. Ivashkevich, A., C.E. Redon, A.J. Nakamura, R.F. Martin, and O.A. Martin. 2012. Use of the γ-H2AX assay to monitor DNA damage and repair in translational cancer research. Cancer Lett. 327: 123–133.

55. Essers, J., A.B. Houtsmuller, L. van Veelen, C. Paulusma, A.L. Nigg, A. Pastink, W. Vermeulen, J.H.J. Hoeijmakers, and R. Kanaar. 2002. Nuclear dynamics of RAD52 group homologous recombination proteins in response to DNA damage. EMBO J. 21: 2030–7.

56. Sugawara, N., X. Wang, and J.E. Haber. 2003. In vivo roles of Rad52, Rad54, and Rad55 proteins in Rad51-mediated recombination. Mol. Cell. 12: 209–19.

57. Rothkamm, K., I. Kruger, L.H. Thompson, and M. Lobrich. 2003. Pathways of DNA Double-Strand Break Repair during the Mammalian Cell Cycle. Mol. Cell. Biol. 23: 5706–5715.

58. Huen, M.S.Y., J. Huang, J.W.C. Leung, S.M.-H. Sy, K.M. Leung, Y.-P. Ching, S.W. Tsao, and J. Chen. 2010. Regulation of chromatin architecture by the PWWP domain-containing DNA damage-responsive factor EXPAND1/MUM1. Mol. Cell. 37: 854–64.

59. Rybak, P., A. Hoang, L. Bujnowicz, T. Bernas, K. Berniak, M. Zarębski, Z. Darzynkiewicz, and J. Dobrucki. 2016. Low level phosphorylation of histone H2AX on serine 139 (γH2AX) is not associated with DNA double-strand breaks. Oncotarget. 7: 49574–49587.

60. Georgoulis, A., C.E. Vorgias, G.P. Chrousos, and E.P. Rogakou. 2017. Genome Instability and γH2AX. Int. J. Mol. Sci. 18.

61. Celeste, A., S. Petersen, P.J. Romanienko, O. Fernandez-Capetillo, H.T. Chen, O.A. Sedelnikova, B. Reina-San-Martin, V. Coppola, E. Meffre, M.J. Difilippantonio, C. Redon, D.R. Pilch, A. Olaru, M. Eckhaus, R.D. Camerini-Otero, L. Tessarollo, F. Livak, K. Manova, W.M. Bonner, M.C. Nussenzweig, and A. Nussenzweig. 2002. Genomic instability in mice lacking histone H2AX. Science. 296: 922–7.

62. Bassing, C.H., H. Suh, D.O. Ferguson, K.F. Chua, J. Manis, M. Eckersdorff, M. Gleason, R. Bronson, C. Lee, and F.W. Alt. 2003. Histone H2AX: a dosage-dependent suppressor of oncogenic translocations and tumors. Cell. 114: 359–70.

63. Celeste, A., O. Fernandez-Capetillo, M.J. Kruhlak, D.R. Pilch, D.W. Staudt, A. Lee, R.F. Bonner, W.M. Bonner, and A. Nussenzweig. 2003. Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks. Nat. Cell Biol. 5: 675–9.

64. Celeste, A., S. Difilippantonio, M.J. Difilippantonio, O. Fernandez-Capetillo, D.R. Pilch, O.A. Sedelnikova, M. Eckhaus, T. Ried, W.M. Bonner, and A. Nussenzweig. 2003. H2AX haploinsufficiency modifies genomic stability and tumor susceptibility. Cell. 114: 371–383.

65. Jeggo, P.A., L.H. Pearl, and A.M. Carr. 2016. DNA repair, genome stability and cancer: a historical perspective. Nat. Rev. Cancer. 16: 35–42.

66. Chapman, J.R., M.R.G. Taylor, and S.J. Boulton. 2012. Playing the end game: DNA double-strand break repair pathway choice. Mol. Cell. 47: 497–510.

67. Luijsterburg, M.S., and H. van Attikum. 2011. Chromatin and the DNA damage response: the cancer connection. Mol. Oncol. 5: 349–67.

68. Harris, R.S. 2013. Cancer mutation signatures, DNA damage mechanisms, and potential clinical implications. Genome Med. 5: 87.

69. Basu, A.K. 2018. DNA Damage, Mutagenesis and Cancer. Int. J. Mol. Sci. 19.

70. Baumstark-Khan, C., U. Hentschel, Y. Nikandrova, J. Krug, and G. Horneck. 2007. Fluorometric Analysis of DNA Unwinding (FADU) as a Method for Detecting Repair-induced DNA Strand Breaks in UV-irradiated Mammalian Cells. Photochem. Photobiol. 72: 477–484.

71. Rastogi, R.P., S.P. Singh, D.-P. Häder, and R.P. Sinha. 2010. Detection of reactive oxygen species (ROS) by the oxidant-sensing probe 2′,7′-dichlorodihydrofluorescein diacetate in the cyanobacterium Anabaena variabilis PCC 7937. Biochem. Biophys. Res. Commun. 397: 603–607.

72. Slieman, T.A., and W.L. Nicholson. 2000. Artificial and Solar UV Radiation Induces Strand Breaks and Cyclobutane Pyrimidine Dimers in Bacillus subtilis Spore DNA. Appl. Environ. Microbiol. 66: 199–205.

73. Rizzo, J.L., J. Dunn, A. Rees, and T.M. Rünger. 2011. No Formation of DNA Double-Strand Breaks and No Activation of Recombination Repair with UVA. J. Invest. Dermatol. 131: 1139–1148.

74. de Lima-Bessa, K.M., M.G. Armelini, V. Chiganças, J.F. Jacysyn, G.P. Amarante-Mendes, A. Sarasin, and C.F.M. Menck. 2008. CPDs and 6-4PPs play different roles in UV-induced cell death in normal and NER-deficient human cells. DNA Repair (Amst). 7: 303–12.

75. Garinis, G.A., J.R. Mitchell, M.J. Moorhouse, K. Hanada, H. de Waard, D. Vandeputte, J. Jans, K. Brand, M. Smid, P.J. van der Spek, J.H.J. Hoeijmakers, R. Kanaar, and G.T.J. van der Horst. 2005. Transcriptome analysis reveals cyclobutane pyrimidine dimers as a major source of UV-induced DNA breaks. EMBO J. 24: 3952–62.

76. Rastogi, R.P., Richa, A. Kumar, M.B. Tyagi, and R.P. Sinha. 2010. Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. J. Nucleic Acids. 2010: 592980.

77. Limoli, C.L., E. Giedzinski, W.M. Bonner, and J.E. Cleaver. 2002. UV-induced replication arrest in the xeroderma pigmentosum variant leads to DNA double-strand breaks, -H2AX formation, and Mre11 relocalization. Proc. Natl. Acad. Sci. 99: 233–238.

78. Takahashi, A., and T. Ohnishi. 2005. Does γH2AX foci formation depend on the presence of DNA double strand breaks? Cancer Lett. 229: 171–179.

79. Ward, J.F. 1988. DNA damage produced by ionizing radiation in mammalian cells: identities, mechanisms of formation, and reparability. Prog. Nucleic Acid Res. Mol. Biol. 35: 95–125.

80. Mehta, A., and J.E. Haber. 2014. Sources of DNA double-strand breaks and models of recombinational DNA repair. Cold Spring Harb. Perspect. Biol. 6: a016428.

81. Ward, J.F. 1994. The complexity of DNA damage: relevance to biological consequences. Int. J. Radiat. Biol. 66: 427–32.

82. Thompson, L.H. 2012. Recognition, signaling, and repair of DNA double-strand breaks produced by ionizing radiation in mammalian cells: The molecular choreography. Mutat. Res. Mutat. Res. 751: 158–246.

83. Hagiwara, Y., T. Oike, A. Niimi, M. Yamauchi, H. Sato, S. Limsirichaikul, K.D. Held, T. Nakano, and A. Shibata. 2019. Clustered DNA double-strand break

formation and the repair pathway following heavy-ion irradiation. J. Radiat. Res. 60: 69–79.

84. Ma, W., C.J. Halweg, D. Menendez, and M.A. Resnick. 2012. Differential effects of poly(ADP-ribose) polymerase inhibition on DNA break repair in human cells are revealed with Epstein-Barr virus. Proc. Natl. Acad. Sci. U. S. A. 109: 6590–5.

85. Wyrobek, A.J., T.E. Schmid, and F. Marchetti. 2005. Relative susceptibilities of male germ cells to genetic defects induced by cancer chemotherapies. J. Natl. Cancer Inst. Monogr. : 31–5.

86. Chen, J., and J. Stubbe. 2005. Bleomycins: towards better therapeutics. Nat. Rev. Cancer. 5: 102–12.

87. Koster, D.A., K. Palle, E.S.M. Bot, M.-A. Bjornsti, and N.H. Dekker. 2007. Antitumour drugs impede DNA uncoiling by topoisomerase I. Nature. 448: 213–7.

88. Driessens, N., S. Versteyhe, C. Ghaddhab, A. Burniat, X. De Deken, J. Van Sande, J.-E. Dumont, F. Miot, and B. Corvilain. 2009. Hydrogen peroxide induces DNA single- and double-strand breaks in thyroid cells and is therefore a potential mutagen for this organ. Endocr. Relat. Cancer. 16: 845–856.

89. Ye, B., N. Hou, L. Xiao, Y. Xu, H. Xu, and F. Li. Dynamic monitoring of oxidative DNA double-strand break and repair in cardiomyocytes. Cardiovasc. Pathol. 25: 93–100.

90. Mahaseth, T., and A. Kuzminov. 2016. Prompt repair of hydrogen peroxide-induced DNA lesions prevents catastrophic chromosomal fragmentation. DNA Repair (Amst). 41: 42–53.

91. Davis, L., and N. Maizels. 2014. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. Proc. Natl. Acad. Sci. U. S. A. 111: E924–E932.

92. Valverde, M., J. Lozano-Salgado, P. Fortini, M.A. Rodriguez-Sastre, E. Rojas, and E. Dogliotti. 2018. Hydrogen Peroxide-Induced DNA Damage and Repair through the Differentiation of Human Adipose-Derived Mesenchymal Stem Cells. Stem Cells Int. 2018: 1–10.

93. Namas, R., P. Renauer, M. Ognenovski, P.-S. Tsou, and A.H. Sawalha. 2016. Histone H2AX phosphorylation as a measure of DNA double-strand breaks and

a marker of environmental stress and disease activity in lupus. Lupus Sci. Med. 3: e000148.

94. Steenken, S. 1989. Purine bases, nucleosides, and nucleotides: aqueous solution redox chemistry and transformation reactions of their radical cations and e- and OH adducts. Chem. Rev. 89: 503–520.

95. Balasubramanian, B., W.K. Pogozelski, and T.D. Tullius. 1998. DNA strand breaking by the hydroxyl radical is governed by the accessible surface areas of the hydrogen atoms of the DNA backbone. Proc. Natl. Acad. Sci. U. S. A. 95: 9738–43.

96. Polo, S.E., and S.P. Jackson. 2011. Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. Genes Dev. 25: 409–33.

97. Aymard, F., M. Aguirrebengoa, E. Guillou, B.M. Javierre, B. Bugler, C. Arnould, V. Rocher, J.S. Iacovoni, A. Biernacka, M. Skrzypczak, K. Ginalski, M. Rowicka, P. Fraser, and G. Legube. 2017. Genome-wide mapping of long-range contacts unveils clustering of DNA double-strand breaks at damaged active genes. Nat. Struct. Mol. Biol. 24: 353–361.

98. Gaj, T., C.A. Gersbach, and C.F. Barbas. 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. 31: 397–405.

99. Segal, D.J., B. Dreier, R.R. Beerli, and C.F. Barbas. 1999. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5’-GNN-3’ DNA target sequences. Proc. Natl. Acad. Sci. U. S. A. 96: 2758–63.

100. Dreier, B., D.J. Segal, and C.F. Barbas. 2000. Insights into the molecular recognition of the 5’-GNN-3’ family of DNA sequences by zinc finger domains. J. Mol. Biol. 303: 489–502.

101. Boch, J., and U. Bonas. 2010. Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function. Annu. Rev. Phytopathol. 48: 419–436.

102. Boch, J., H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt, and U. Bonas. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. Science. 326: 1509–12.

103. Moscou, M.J., and A.J. Bogdanove. 2009. A simple cipher governs DNA recognition by TAL effectors. Science. 326: 1501.

104. Barrangou, R. 2015. The roles of CRISPR-Cas systems in adaptive immunity and beyond. Curr. Opin. Immunol. 32: 36–41.

105. Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D.A. Romero, and P. Horvath. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 315: 1709–12.

106. Marraffini, L.A., and E.J. Sontheimer. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science. 322: 1843–5.

107. Redman, M., A. King, C. Watson, and D. King. 2016. What is CRISPR/Cas9? Arch. Dis. Child. Educ. Pract. Ed. 101: 213–5.

108. Bolotin, A., B. Quinquis, A. Sorokin, and S.D. Ehrlich. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology. 151: 2551–61.

109. Jansen, R., J.D.A. van Embden, W. Gaastra, and L.M. Schouls. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. 43: 1565–75.

110. Mojica, F.J., C. Díez-Villaseñor, E. Soria, and G. Juez. 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Mol. Microbiol. 36: 244–6.

111. Hille, F., H. Richter, S.P. Wong, M. Bratovič, S. Ressel, and E. Charpentier. 2018. The Biology of CRISPR-Cas: Backward and Forward. Cell. 172: 1239–1259.

112. Makarova, K.S., and E. V Koonin. 2015. Annotation and Classification of CRISPR-Cas Systems. Methods Mol. Biol. 1311: 47–75.

113. Marraffini, L.A., and E.J. Sontheimer. 2010. Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature. 463: 568–71.

114. Overballe-Petersen, S., K. Harms, L.A.A. Orlando, J.V.M. Mayar, S. Rasmussen, T.W. Dahl, M.T. Rosing, A.M. Poole, T. Sicheritz-Ponten, S. Brunak, S. Inselmann, J. de Vries, W. Wackernagel, O.G. Pybus, R. Nielsen, P.J. Johnsen, K.M. Nielsen, and E. Willerslev. 2013. Bacterial natural transformation by highly fragmented and damaged DNA. Proc. Natl. Acad. Sci. 110: 19860–19865.

115. Gong, C., P. Bongiorno, A. Martins, N.C. Stephanou, H. Zhu, S. Shuman, and M.S. Glickman. 2005. Mechanism of nonhomologous end-joining in

mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. Nat. Struct. Mol. Biol. 12: 304–12.

116. Cong, L., F.A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P.D. Hsu, X. Wu, W. Jiang, L.A. Marraffini, and F. Zhang. 2013. Multiplex genome engineering using CRISPR/Cas systems. Science. 339: 819–23.

117. Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, and E. Charpentier. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 337: 816–21.

118. Gasiunas, G., R. Barrangou, P. Horvath, and V. Siksnys. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. U. S. A. 109: E2579-86.

119. Qi, L.S., M.H. Larson, L.A. Gilbert, J.A. Doudna, J.S. Weissman, A.P. Arkin, and W.A. Lim. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 152: 1173–83.

120. Chen, B., L.A. Gilbert, B.A. Cimini, J. Schnitzbauer, W. Zhang, G.-W. Li, J. Park, E.H. Blackburn, J.S. Weissman, L.S. Qi, and B. Huang. 2013. Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. Cell. 155: 1479–1491.

121. Mojica, F.J.M., and R.A. Garrett. 2013. Discovery and Seminal Developments in the CRISPR Field. In: CRISPR-Cas Systems. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 1–31.

122. Yang, D., J. Xu, T. Zhu, J. Fan, L. Lai, J. Zhang, and Y.E. Chen. 2014. Effective gene targeting in rabbits using RNA-guided Cas9 nucleases. J. Mol. Cell Biol. 6: 97–9.

123. Li, D., Z. Qiu, Y. Shao, Y. Chen, Y. Guan, M. Liu, Y. Li, N. Gao, L. Wang, X. Lu, Y. Zhao, and M. Liu. 2013. Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. Nat. Biotechnol. 31: 681–3.

124. Wang, H., H. Yang, C.S. Shivalila, M.M. Dawlaty, A.W. Cheng, F. Zhang, and R. Jaenisch. 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell. 153: 910–8.

125. Oh, J.-H., and J.-P. van Pijkeren. 2014. CRISPR-Cas9-assisted recombineering in Lactobacillus reuteri. Nucleic Acids Res. 42: e131.

126. Pyne, M.E., M. Moo-Young, D.A. Chung, and C.P. Chou. 2015. Coupling the CRISPR/Cas9 System with Lambda Red Recombineering Enables Simplified

Chromosomal Gene Replacement in Escherichia coli. Appl. Environ. Microbiol. 81: 5103–5114.

127. Hwang, W.Y., Y. Fu, D. Reyon, M.L. Maeder, P. Kaini, J.D. Sander, J.K. Joung, R.T. Peterson, and J.-R.J. Yeh. 2013. Heritable and Precise Zebrafish Genome Editing Using a CRISPR-Cas System. PLoS One. 8: e68708.

128. DiCarlo, J.E., J.E. Norville, P. Mali, X. Rios, J. Aach, and G.M. Church. 2013. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Res. 41: 4336–43.

129. Niu, Y., B. Shen, Y. Cui, Y. Chen, J. Wang, L. Wang, Y. Kang, X. Zhao, W. Si, W. Li, A.P. Xiang, J. Zhou, X. Guo, Y. Bi, C. Si, B. Hu, G. Dong, H. Wang, Z. Zhou, T. Li, T. Tan, X. Pu, F. Wang, S. Ji, Q. Zhou, X. Huang, W. Ji, and J. Sha. 2014. Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. Cell. 156: 836–43.

130. Mali, P., J. Aach, P.B. Stranges, K.M. Esvelt, M. Moosburner, S. Kosuri, L. Yang, and G.M. Church. 2013. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. 31: 833–8.

131. Jiang, W., D. Bikard, D. Cox, F. Zhang, and L.A. Marraffini. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. 31: 233–9.

132. Su, Y.-H., T.-Y. Lin, C.-L. Huang, C.-F. Tu, and C.-K. Chuang. 2015. Construction of a CRISPR-Cas9 System for Pig Genome Targeting. Anim. Biotechnol. 26: 279–88.

133. Chen, C., L.A. Fenk, and M. de Bono. 2013. Efficient genome editing in Caenorhabditis elegans by CRISPR-targeted homologous recombination. Nucleic Acids Res. 41: e193.

134. Jinek, M., F. Jiang, D.W. Taylor, S.H. Sternberg, E. Kaya, E. Ma, C. Anders, M. Hauer, K. Zhou, S. Lin, M. Kaplan, A.T. Iavarone, E. Charpentier, E. Nogales, and J.A. Doudna. 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science. 343: 1247997.

135. Nishimasu, H., F.A. Ran, P.D. Hsu, S. Konermann, S.I. Shehata, N. Dohmae, R. Ishitani, F. Zhang, and O. Nureki. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell. 156: 935–49.

136. Wiedenheft, B., S.H. Sternberg, and J.A. Doudna. 2012. RNA-guided genetic silencing systems in bacteria and archaea. Nature. 482: 331–8.

137. Anders, C., O. Niewoehner, A. Duerst, and M. Jinek. 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature. 513: 569–73.

138. Sternberg, S.H., S. Redding, M. Jinek, E.C. Greene, and J.A. Doudna. 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature. 507: 62–7.

139. Palermo, G., Y. Miao, R.C. Walker, M. Jinek, and J.A. McCammon. 2016. Striking Plasticity of CRISPR-Cas9 and Key Role of Non-target DNA, as Revealed by Molecular Simulations. ACS Cent. Sci. 2: 756–763.

140. Palermo, G., Y. Miao, R.C. Walker, M. Jinek, and J.A. McCammon. 2017. CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. Proc. Natl. Acad. Sci. 114: 7260–7265.

141. Sternberg, S.H., B. LaFrance, M. Kaplan, and J.A. Doudna. 2015. Conformational control of DNA target cleavage by CRISPR-Cas9. Nature. 527: 110–3.

142. Górecka, K.M., W. Komorowska, and M. Nowotny. 2013. Crystal structure of RuvC resolvase in complex with Holliday junction substrate. Nucleic Acids Res. 41: 9945–55.

143. Miller, J.C., S. Tan, G. Qiao, K.A. Barlow, J. Wang, D.F. Xia, X. Meng, D.E. Paschon, E. Leung, S.J. Hinkley, G.P. Dulay, K.L. Hua, I. Ankoudinova, G.J. Cost, F.D. Urnov, H.S. Zhang, M.C. Holmes, L. Zhang, P.D. Gregory, and E.J. Rebar. 2011. A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol. 29: 143–8.

144. Maeder, M.L., S. Thibodeau-Beganny, A. Osiak, D.A. Wright, R.M. Anthony, M. Eichtinger, T. Jiang, J.E. Foley, R.J. Winfrey, J.A. Townsend, E. Unger-Wallace, J.D. Sander, F. Müller-Lerch, F. Fu, J. Pearlberg, C. Göbel, J.P. Dassie, S.M. Pruett-Miller, M.H. Porteus, D.C. Sgroi, A.J. Iafrate, D. Dobbs, P.B. McCray, T. Cathomen, D.F. Voytas, and J.K. Joung. 2008. Rapid “open-source” engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol. Cell. 31: 294–301.

145. Mussolino, C., R. Morbitzer, F. Lütge, N. Dannemann, T. Lahaye, and T. Cathomen. 2011. A novel TALE nuclease scaffold enables high genome

editing activity in combination with low toxicity. Nucleic Acids Res. 39: 9283–93.

146. Lee, C.M., T.J. Cradick, E.J. Fine, and G. Bao. 2016. Nuclease Target Site Selection for Maximizing On-target Activity and Minimizing Off-target Effects in Genome Editing. Mol. Ther. 24: 475–487.

147. Wood, A.J., T.-W. Lo, B. Zeitler, C.S. Pickle, E.J. Ralston, A.H. Lee, R. Amora, J.C. Miller, E. Leung, X. Meng, L. Zhang, E.J. Rebar, P.D. Gregory, F.D. Urnov, and B.J. Meyer. 2011. Targeted genome editing across species using ZFNs and TALENs. Science. 333: 307.

148. Pattanayak, V., C.L. Ramirez, J.K. Joung, and D.R. Liu. 2011. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. Nat. Methods. 8: 765–70.

149. Durai, S., M. Mani, K. Kandavelou, J. Wu, M.H. Porteus, and S. Chandrasegaran. 2005. Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. Nucleic Acids Res. 33: 5978–90.

150. Zheng, Q., X. Cai, M.H. Tan, S. Schaffert, C.P. Arnold, X. Gong, C.-Z. Chen, and S. Huang. 2014. Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. Biotechniques. 57.

151. Feng, Z., B. Zhang, W. Ding, X. Liu, D.-L. Yang, P. Wei, F. Cao, S. Zhu, F. Zhang, Y. Mao, and J.-K. Zhu. 2013. Efficient genome editing in plants using a CRISPR/Cas system. Cell Res. 23: 1229–32.

152. Li, X.-L., G.-H. Li, J. Fu, Y.-W. Fu, L. Zhang, W. Chen, C. Arakaki, J.-P. Zhang, W. Wen, M. Zhao, W. V Chen, G.D. Botimer, D. Baylink, L. Aranda, H. Choi, R. Bechar, P. Talbot, C.-K. Sun, T. Cheng, and X.-B. Zhang. 2018. Highly efficient genome editing via CRISPR–Cas9 in human pluripotent stem cells is achieved by transient BCL-XL overexpression. Nucleic Acids Res. 46: 10195–10215.

153. Mali, P., L. Yang, K.M. Esvelt, J. Aach, M. Guell, J.E. DiCarlo, J.E. Norville, and G.M. Church. 2013. RNA-guided human genome engineering via Cas9. Science. 339: 823–6.

154. Xu, X., D. Gao, P. Wang, J. Chen, J. Ruan, J. Xu, and X. Xia. 2018. Efficient homology-directed gene editing by CRISPR/Cas9 in human stem and primary cells using tube electroporation. Sci. Rep. 8: 11649.

155. Zhou, J., J. Wang, B. Shen, L. Chen, Y. Su, J. Yang, W. Zhang, X. Tian, and X. Huang. 2014. Dual sgRNAs facilitate CRISPR/Cas9-mediated mouse genome targeting. FEBS J. 281: 1717–25.

156. Hsu, P.D., D.A. Scott, J.A. Weinstein, F.A. Ran, S. Konermann, V. Agarwala, Y. Li, E.J. Fine, X. Wu, O. Shalem, T.J. Cradick, L.A. Marraffini, G. Bao, and F. Zhang. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. 31: 827–32.

157. Fu, Y., J.A. Foden, C. Khayter, M.L. Maeder, D. Reyon, J.K. Joung, and J.D. Sander. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. 31: 822–6.

158. Skene, P.J., J.G. Henikoff, and S. Henikoff. 2018. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. Nat. Protoc. 13: 1006–1019.

159. Fu, Y., J.D. Sander, D. Reyon, V.M. Cascio, and J.K. Joung. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat. Biotechnol. 32: 279–284.

160. Mashiko, D., S.A.M. Young, M. Muto, H. Kato, K. Nozawa, M. Ogawa, T. Noda, Y.-J. Kim, Y. Satouh, Y. Fujihara, and M. Ikawa. 2014. Feasibility for a large scale mouse mutagenesis by injecting CRISPR/Cas plasmid into zygotes. Dev. Growth Differ. 56: 122–9.

161. Wong, N., W. Liu, and X. Wang. 2015. WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. Genome Biol. 16: 218.

162. Chari, R., P. Mali, M. Moosburner, and G.M. Church. 2015. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. Nat. Methods. 12: 823–6.

163. Shalem, O., N.E. Sanjana, E. Hartenian, X. Shi, D.A. Scott, T. Mikkelson, D. Heckl, B.L. Ebert, D.E. Root, J.G. Doench, and F. Zhang. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science. 343: 84–87.

164. Doench, J.G., N. Fusi, M. Sullender, M. Hegde, E.W. Vaimberg, K.F. Donovan, I. Smith, Z. Tothova, C. Wilen, R. Orchard, H.W. Virgin, J. Listgarten, and D.E. Root. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. 34: 184–191.

165. Xie, S., B. Shen, C. Zhang, X. Huang, and Y. Zhang. 2014. sgRNAcas9: a software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites. PLoS One. 9: e100448.

166. Zhu, L.J., B.R. Holmes, N. Aronin, and M.H. Brodsky. 2014. CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. PLoS One. 9: e108424.

167. Upadhyay, S.K., and S. Sharma. 2014. SSFinder: high throughput CRISPR-Cas target sites prediction tool. Biomed Res. Int. 2014: 742482.

168. Doench, J.G., E. Hartenian, D.B. Graham, Z. Tothova, M. Hegde, I. Smith, M. Sullender, B.L. Ebert, R.J. Xavier, and D.E. Root. 2014. Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. Nat. Biotechnol. 32: 1262–1267.

169. Ren, X., Z. Yang, J. Xu, J. Sun, D. Mao, Y. Hu, S.-J. Yang, H.-H. Qiao, X. Wang, Q. Hu, P. Deng, L.-P. Liu, J.-Y. Ji, J.B. Li, and J.-Q. Ni. 2014. Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in Drosophila. Cell Rep. 9: 1151–62.

170. Isaac, R.S., F. Jiang, J.A. Doudna, W.A. Lim, G.J. Narlikar, and R. Almeida. 2016. Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. Elife. 5.

171. Horlbeck, M.A., L.B. Witkowsky, B. Guglielmi, J.M. Replogle, L.A. Gilbert, J.E. Villalta, S.E. Torigoe, R. Tjian, and J.S. Weissman. 2016. Nucleosomes impede Cas9 access to DNA in vivo and in vitro. Elife. 5.

172. Cameron, P., C.K. Fuller, P.D. Donohoue, B.N. Jones, M.S. Thompson, M.M. Carter, S. Gradia, B. Vidal, E. Garner, E.M. Slorach, E. Lau, L.M. Banh, A.M. Lied, L.S. Edwards, A.H. Settle, D. Capurso, V. Llaca, S. Deschamps, M. Cigan, J.K. Young, and A.P. May. 2017. Mapping the genomic landscape of CRISPR-Cas9 cleavage. Nat. Methods. 14: 600–606.

173. Tsai, S.Q., Z. Zheng, N.T. Nguyen, M. Liebers, V. V Topkar, V. Thapar, N. Wyvekens, C. Khayter, A.J. Iafrate, L.P. Le, M.J. Aryee, and J.K. Joung. 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. 33: 187–197.

174. Abadi, S., W.X. Yan, D. Amar, and I. Mayrose. 2017. A machine learning approach for predicting CRISPR-Cas9 cleavage efficiencies and patterns underlying its mechanism of action. PLoS Comput. Biol. 13: e1005807.

175. Listgarten, J., M. Weinstein, B.P. Kleinstiver, A.A. Sousa, J.K. Joung, J. Crawford, K. Gao, L. Hoang, M. Elibol, J.G. Doench, and N. Fusi. 2018. Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. Nat. Biomed. Eng. 2: 38–47.

176. Lin, J., and K.-C. Wong. 2018. Off-target predictions in CRISPR-Cas9 gene editing using deep learning. Bioinformatics. 34: i656–i663.

177. Wilson, L.O.W., A.R. O'Brien, and D.C. Bauer. 2018. The Current State and Future of CRISPR-Cas9 gRNA Design Tools. Front. Pharmacol. 9.

178. Kuo, L.J., and L.-X. Yang. 2008. Gamma-H2AX - a novel biomarker for DNA double-strand breaks. In Vivo. 22: 305–9.

179. Doench, J.G., E. Hartenian, D.B. Graham, Z. Tothova, M. Hegde, I. Smith, M. Sullender, B.L. Ebert, R.J. Xavier, and D.E. Root. 2014. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat. Biotechnol. 32: 1262–7.

180. van den Berg, J., A. G Manjón, K. Kielbassa, F.M. Feringa, R. Freire, and R.H. Medema. 2018. A limited number of double-strand DNA breaks is sufficient to delay cell cycle progression. Nucleic Acids Res. 46: 10132–10144.

181. Zuris, J.A., D.B. Thompson, Y. Shu, J.P. Guilinger, J.L. Bessen, J.H. Hu, M.L. Maeder, J.K. Joung, Z.-Y. Chen, and D.R. Liu. 2015. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. Nat. Biotechnol. 33: 73–80.

182. Anders, C., and M. Jinek. 2014. In vitro enzymology of Cas9. Methods Enzymol. 546: 1–20.

183. McKenna, A., and J. Shendure. 2018. FlashFry: a fast and flexible tool for large-scale CRISPR target design. BMC Biol. 16: 74.

184. Moreno-Mateos, M.A., C.E. Vejnar, J.-D. Beaudoin, J.P. Fernandez, E.K. Mis, M.K. Khokha, and A.J. Giraldez. 2015. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat. Methods. 12: 982–8.

185. Deng, W., X. Shi, R. Tjian, T. Lionnet, and R.H. Singer. 2015. CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. Proc. Natl. Acad. Sci. U. S. A. 112: 11870–5.

186. Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D.J.

White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona. 2012. Fiji: an open-source platform for biological-image analysis. Nat. Methods. 9: 676–82.

187. Bewersdorf, J., B.T. Bennett, and K.L. Knight. 2006. H2AX chromatin structures and their response to DNA damage revealed by 4Pi microscopy. Proc. Natl. Acad. Sci. U. S. A. 103: 18137–42.

188. Sak, A., and M. Stuschke. 2010. Use of γH2AX and other biomarkers of double-strand breaks during radiotherapy. Semin. Radiat. Oncol. 20: 223–31.

189. Sisario, D., S. Memmel, S. Doose, J. Neubauer, H. Zimmermann, M. Flentje, C.S. Djuzenova, M. Sauer, and V.L. Sukhorukov. 2018. Nanostructure of DNA repair foci revealed by superresolution microscopy. FASEB J. 32: 6469–6477.

190. Rajagopalan, N., S. Kagale, P. Bhowmik, and H. Song. 2018. A Two-Step Method for Obtaining Highly Pure Cas9 Nuclease for Genome Editing, Biophysical, and Structural Studies. Methods Protoc. 1: 17.

191. Skene, P.J., and S. Henikoff. 2017. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife. 6.

192. He, C., and R. Bonasio. 2017. A cut above. Elife. 6.

193. Meyer, C.A., and X.S. Liu. 2014. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. Nat. Rev. Genet. 15: 709–721.

194. Iacovoni, J.S., P. Caron, I. Lassadi, E. Nicolas, L. Massip, D. Trouche, and G. Legube. 2010. High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. EMBO J. 29: 1446–57.

195. Wang, T., J.J. Wei, D.M. Sabatini, and E.S. Lander. 2014. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. Science (80-.). 343: 80–84.

196. Ding, D., Y. Zhang, J. Wang, X. Zhang, Y. Gao, L. Yin, Q. Li, J. Li, and H. Chen. 2016. Induction and inhibition of the pan-nuclear γH2AX response in resting human peripheral blood lymphocytes after X-ray irradiation. Cell Death Discov. 2: 16011.

197. Ivashkevich, A.N., O.A. Martin, A.J. Smith, C.E. Redon, W.M. Bonner, R.F. Martin, and P.N. Lobachevsky. 2011. γH2AX foci as a measure of DNA damage: a computational approach to automatic analysis. Mutat. Res. 711: 49–60.

198. Revet, I., L. Feeney, S. Bruguera, W. Wilson, T.K. Dong, D.H. Oh, D. Dankort, and J.E. Cleaver. 2011. Functional relevance of the histone H2Ax in the response to DNA damaging agents. Proc. Natl. Acad. Sci. 108: 8663–8667.

199. Bouquet, F., C. Muller, and B. Salles. 2006. The Loss of γH2AX Signal is a Marker of DNA Double Strand Breaks Repair Only at Low Levels of DNA Damage. Cell Cycle. 5: 1116–1122.