UNIVERSITY OF CALIFORNIA SAN DIEGO

A Human-Optimized Damage Suppressor Protein as a Potential Tool to Improve DNA Damage Protection in Human Cells

A Thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Biology

by

Ginevra Caswell

Committee in charge:

Professor James Kadonaga, Chair
Professor Cornelis Murre
Professor Lorraine Pillus

2023

The Thesis of Ginevra Caswell is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

DEDICATION

This thesis is dedicated to Grisel Cruz Becerra whose daily mentorship shaped me into becoming a biologist focused on problem solving and integrity. I am forever grateful for her encouragement and patience throughout my time in the Kadonaga Laboratory.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Dsup	Damage suppressor protein
NBD	Nucleosome binding domain
HMGN	High mobility group nucleosomal binding domain
ROS	Reactive oxygen species
HDR	Homology directed repair
crRNA	CRISPR RNA
sgRNA	Single-guide RNA
mEGFP	Monomeric enhanced green fluorescent protein
HA-R	Right homology arm
HA-L	Left homology arm

ABSTRACT OF THE THESIS

A Human-Optimized Damage Suppressor Protein as a Potential Tool to Improve DNA Damage Protection in Human Cells

by

Ginevra Caswell

Master of Science in Biology

University of California San Diego, 2023

Professor James Kadonaga, Chair

The tardigrade's ability to survive in extreme environments intrigues the scientific community. The mechanistic basis for its durable physiology was partially elucidated with the identification of the tardigrade-specific damage suppressor (Dsup) protein, which is a nucleosome-binding factor that protects against DNA damage when expressed in human cells. Because Dsup is only present in certain tardigrades species, however, it is not optimized to

function in human cells. Therefore, this study was conducted to determine if a modified Dsup protein containing the nucleosome binding domain (NBD) of the human high mobility group nucleosomal binding domain 2 (HMGN2) protein could improve protection of human cells against reactive oxygen species (ROS), such as hydroxyl radicals, which damage DNA. Two Dsup-HMGN2 hybrid proteins were designed: Dsup containing the entire NBD of human HMGN2 (Dsup-NBD), and Dsup containing only the core NBD region of HMGN2 (Dsup-coreNBD). Dsup wildtype and the two Dsup-HMGN2 hybrids were integrated into the genome of MCF10A cells (human non-tumorigenic epithelial cells), via homology directed repair (HDR) using the CRISPR/Cas9 system. Following exposure to hydrogen peroxide (which generates hydroxyl radicals), the amount of DNA damage in Dsup containing cells was quantified using the alkaline comet assay. Under the conditions of my experiment, I observed significantly more DNA damage in Dsup-expressing cells when compared to the wildtype MCF10A cells lacking Dsup. Despite this conclusion, the optimization of the Dsup protein for DNA damage protection in human cells could have therapeutic potential when used in conjunction with existing technologies such as gene therapy.

INTRODUCTION

While DNA damage is an intrinsic byproduct of internal cellular processes, research efforts to understand and increase protection against DNA damage are critical in an environment surrounded by an increasing amount of external DNA damage-causing agents (Jackson & Bartek, 2009). Human cells are susceptible to DNA damage by reactive oxygen species (ROS), such as hydroxyl radicals. Hydroxyl radicals, which act as the intermediate species for DNA damaging exposures to agents such as hydrogen peroxide and ionizing radiation, remove a hydrogen atom from the DNA backbone (Balasubramanian et al., 1998). This reaction can result in single- and double-stranded DNA breaks that the cellular machinery must work to repair. DNA damage protection and repair mechanisms are essential to prevent diseases such as cancer, that can develop due to uncontrolled proliferation of damaged cells (Cooke et al., 2003, Balasubramanian et al., 1998).

Extremotolerant organisms refer to those that can survive in extreme environments (such as particularly high or low temperature, pressure, and acidity levels) by entering a protective metabolic state (Veling et al., 2022). This metabolic state is induced through the up- or down-regulation of key factors that help mitigate DNA damage (Hashimoto & Kunieda, 2017). The specific factors that have evolved within extremotolerant organisms to help them survive in a variety of harsh environments serve as a genetic trove for biotechnological and biomedical applications. For example, the extremotolerant nematode *Aphelenchus avenae* contains a late embryonic abundant (LEA) protein which was shown to alleviate muscular dystrophy-associated protein aggregates when expressed in human cells (Veling et al., 2022). Research efforts harnessing the potential of proteins found within extremotolerant organisms could expand human disease prevention and treatment.

A phylum known as tardigrades (also referred to as water bears or moss piglets) consists of organisms with eight legs and whose size ranges from about 0.1 to 1 mm in length. Tardigrades are functionally aquatic organisms; however, certain tardigrade species can enter an anhydrobiotic state in the absence of water. This state allows the tardigrade to tolerate extreme conditions, such as a wide temperature range, high pressure, and even exposure to the outer atmosphere (Hashimoto & Kunieda, 2017). Investigations into the molecular basis for the extremotolerant nature of these animals, aided by the sequencing of the tardigrade *Ramazzottius varieornatus* genome, led to the identification of the tardigrade-unique damage suppressor (Dsup) protein (Hashimoto et al., 2016). Dsup is an overall unstructured and highly charged protein that localizes to the nucleus (Hashimoto et al., 2016; Mínguez-Toral et al., 2020). In a study conducted by Hashimoto and colleagues, Dsup was shown to protect tumorigenic HEK293 human cells from DNA damage by hydrogen peroxide and ionizing radiation (2016). It was later demonstrated that Dsup protects DNA against hydroxyl radicals by binding to the nucleosome (Chavez et al., 2019). The optimization of Dsup, a protein derived from a non-vertebrate organism, to protect against DNA damage in vertebrates was inspired by the discovery that Dsup shares a stretch of similar amino acids with the nucleosome binding domain (NBD) of vertebrate high mobility group nucleosomal binding domain (HMGN) proteins (Chavez et al., 2019). While HMGN proteins are present in all vertebrates, thus far, Dsup is the only known HMGN-like factor found outside of vertebrates (Chavez et al., 2019). HMGN proteins contain a unique nucleosome binding domain for which the nucleosome contains two high affinity binding sites (González-Romero et al., 2014; Nanduri et al., 2020). Dsup's NBD sequence homology with that of the HMGN protein could be manipulated to improve binding of Dsup to nucleosomes in human cells and has yet to be applied to enhance human DNA damage protection mechanisms.

The goal of my research was to improve Dsup-mediated DNA damage protection in human cells. Because Dsup mechanistically binds to the nucleosome to protect DNA, I hypothesized that an increase in this binding affinity within human cells would result in greater tolerance to damage. I generated human cell lines containing a modified Dsup protein, "humanized" through the genetic fusion of Dsup to either the entire nucleosome binding domain (NBD) of the human HMGN2 proteins or just the core consensus region, which is critical for nucleosomal binding of all HMGN proteins. HMGN2 is one of the five members of the HMGN protein family (HMGN1-5) and is considered the most genetically conserved (Kugler et al., 2012). Humanized Dsup was shown to bind nucleosomes with a higher affinity than wildtype Dsup in preliminary *in vitro* experiments, and therefore may improve protection against DNA damage in human cells. Before assessing whether DNA damage protection by Dsup could be optimized in human cells, however, I first had to determine if Dsup protects non-tumorigenic human cells from DNA damage as Dsup has only been shown to increase DNA damage tolerance in tumorigenic human cells so far (Hashimoto et al., 2016; Ricci et al., 2021). DNA damage assessment of non-tumorigenic human cells stably expressing Dsup and Dsup-HMGN2 hybrid transgenes by could navigate Dsup research toward therapeutic applications.

Cell lines expressing Dsup or Dsup-HMGN2 hybrid proteins were generated so that their capacity to protect DNA from hydroxyl radicals in human cells could be tested. It was then necessary to ensure Dsup protects against DNA damage in non-tumorigenic MCF10A cells before beginning comparative analysis with established cell lines stably expressing Dsup and Dsup-HMGN2 hybrid proteins. A cell line containing the Dsup protein was subjected to hydrogen peroxide treatment (which produces hydroxyl radicals) before DNA damage was assessed with the alkaline comet assay. Despite the surprising increase in DNA damage exhibited

by the Dsup-containing cells when compared to untransfected cells, further experimentation using the genetic tools generated in this study should be conducted to investigate Dsup's versatility for applications in DNA damage protection.

If future experiments confirm that cells expressing Dsup and Dsup-HMGN2 hybrid proteins improve DNA damage protection in non-tumorigenic human cells, the molecular instruments developed in this study could function as biomedically useful reagents for protecting human cells from ROS. Human therapeutics such as gene therapy could benefit from a robust cell line that has a higher threshold for detrimental DNA damage. Overall, this study builds on previous evidence suggesting that the Dsup protein's ability to provide DNA damage protection holds potential for the advancement of human health and knowledge.

CHAPTER 1: CELL LINE DESIGN

Previous findings demonstrated that Dsup provides DNA damage protection by binding to nucleosomes (Chavez et al., 2019). Therefore, an increased affinity of Dsup-HMGN2 hybrid proteins to nucleosomes, when compared to wildtype Dsup, may suggest that damage protection in human cells could be improved by stably expressing Dsup-HMGN2 hybrid proteins. *In vitro* analysis of nucleosome affinity was conducted in the form of a gel shift assay (EMSA) to determine if Dsup-HMGN2 hybrids bind to nucleosomes with higher affinity than Dsup (Figure 1.1). This 4.5% native gel shows the migration of nucleosomal DNA before and after the addition of the indicated Dsup or Dsup-HMGN2 hybrid proteins. An upward band shift (away from the free mononucleosome) demonstrates the association of Dsup and Dsup-HMGN2 hybrid proteins with mononucleosomes. Rather than the wildtype Dsup sequence, the Dsup-HMGN2 hybrid sequences Dsup-coreNBD, Dsup-0.5NBD, and Dsup-NBD contain the core HMGN2 nucleosome binding domain, 50% of the HMGN2 nucleosome binding domain, and the entire HMGN2 nucleosome binding domain, respectively. Upon the addition of wildtype Dsup to mononucleosomes, an upward band shift is observed, indicating that Dsup associates with the nucleosomes, as previously observed (Chavez et al., 2019). Upon addition of Dsup-HMGN2 hybrids (*i.e*., Dsup-coreNBD, Dsup-0.5NBD, and Dsup-NBD) to mononucleosomes, however, a band shift of greater intensity than that of wildtype Dsup is observed. This suggests that the Dsup-HMGN2 hybrids, specifically Dsup-coreNBD and Dsup-NBD, have a higher binding affinity to mononucleosomes than wildtype Dsup (Figure 1.1). Alongside previous research, this evidence suggests that the generation of human cells stabling expressing Dsup-HMGN2 hybrid proteins could be promising for improving protection against DNA damage (Chavez et al., 2019). Constructs Dsup-coreNBD and Dsup-NBD were used throughout this study while Dsup-

0.5NBD was not further tested. These *in vitro* results strengthened my hypothesis and motivated the design of a CRISPR/Cas9 targeted integration strategy to stably express Dsup and Dsup-HMGN2 hybrid proteins in human cells.

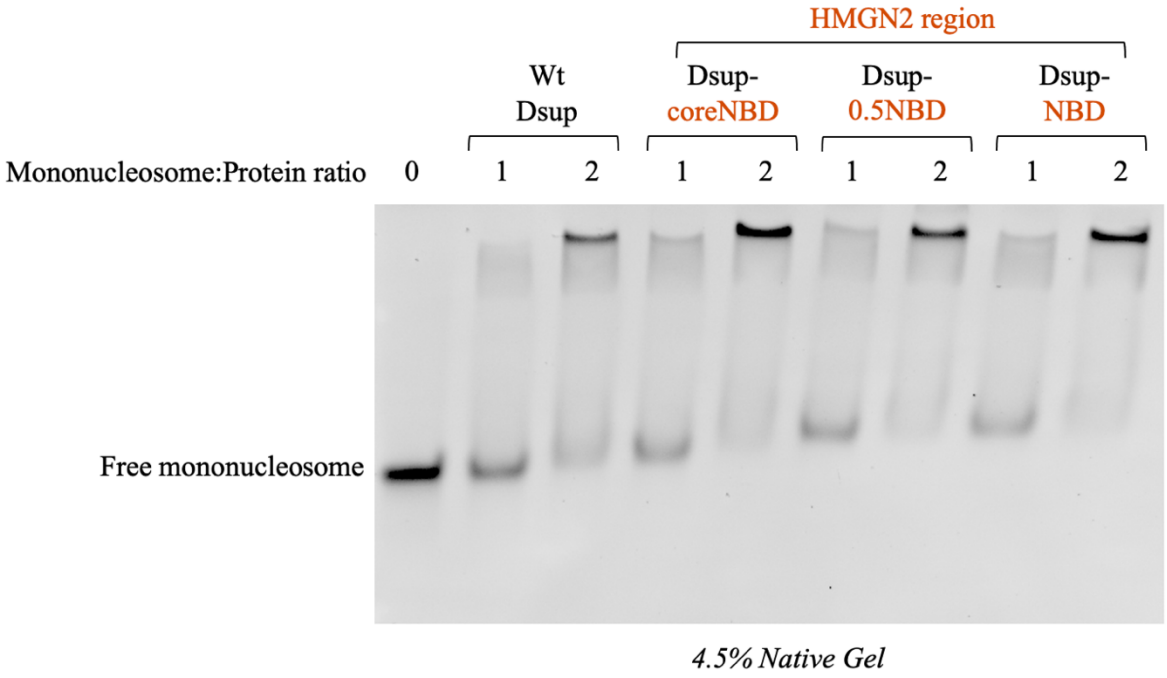


Figure 1.1: Dsup-HMGN2 variants bind with higher affinity to mononucleosomes than wildtype Dsup. A gel shift assay was performed to compare the binding affinity of wildtype Dsup and Dsup-HMGN2 hybrid proteins to the nucleosome *in vitro*. The nucleosome binding domain (NBD) corresponds to a 29 amino acid sequence of human HMGN2. The core consensus sequence (*i.e.*, RRSARLSA) of the NBD of human HMGN2 is denoted as *coreNBD*. *0.5NBD* corresponds to half of the NBD sequence of human HMGN2. *NBD* corresponds to the full-length NBD sequence of human HMGN2. An upward nucleosomal DNA band shift indicates the association of Dsup and Dsup-HMGN2 hybrid proteins with mononucleosomes. The greater the enrichment in shifted band from free mononucleosome toward the top of the gel, the greater the affinity the indicated protein has to the nucleosome. *Data is unpublished from an experiment conducted by A. A. Chavez, G. Cruz-Becerra, and J. T. Kadonaga.*

MCF10A cells, non-tumorigenic human epithelial cells, were selected for CRISPR-Cas9 mediated generation of human cell lines stabling expressing Dsup and Dsup-HMGN2 hybrid proteins (Debnath et al., 2003). Determining if Dsup and Dsup-HMGN2 variants can protect against DNA damage in non-tumorigenic cells rather than tumorigenic cells is critical for

potential biomedical applications, as discussed in *Conclusion*. It was determined that genomic editing of MCF10A cells would be conducted at the *AAVS1* locus as it is considered a safe harbor locus (Oceguera-Yanez et al., 2016; Hayashi et al., 2020). In order to direct Cas9 to cleave DNA at the *AAVS1* locus and induce homology directed repair (HDR), I generated two plasmids that would both express a Cas9 protein but would express different single-guide RNAs (sgRNAs) when transfected. The use of multiple sgRNAs for the intended locus is recommended as some fail to work, as previously reported (Ran et al., 2013). CRISPR RNA (crRNA) sequences 1 (GTCACCAATCCTGTCCCTAG) and 2 (GGGGCCACTAGGGACAGGAT) were identified on the *AAVS1* locus and the corresponding sequences were inserted into an expression vector for Cas9. See '*Generation of Cas9 plasmid containing sgRNA1 and sgRNA2'* in *Methods* for a detailed procedure. crRNA1 and crRNA2 target sequences were chosen due to their alignment between the *AAVS1* homology arms (HAs), which correspond to those included in the HDR donor repair templates. Both crRNAs are followed by 5'-TGG PAM sequences (Figure 1.2). The on- and off-target scores (high are better) of these crRNAs were determined using the IDT CRISPR designer tool: crRNA1 has an on-target score of 84 and an off-target score of 58 while crRNA2 has an on-target score of 46 and an off-target score of 22. Both demonstrated successful results in preliminary HDR experiments (see *Chapter 2*), therefore, only crRNA1 was used for the establishment of cell lines used in DNA damage analysis (see *Chapter 3*). crRNA1 was used over crRNA2 as it has better on- and off-target scores, therefore it is predicted to show higher specificity for the target locus. Following '*sequencing'* of the ligated region within Cas9 plasmid candidate clones, those that demonstrated correct alignment to predicted sequences were selected and amplified according to *'large-scale plasmid amplification'* in preparation for transfection reactions alongside donor repair templates (see *Methods*).

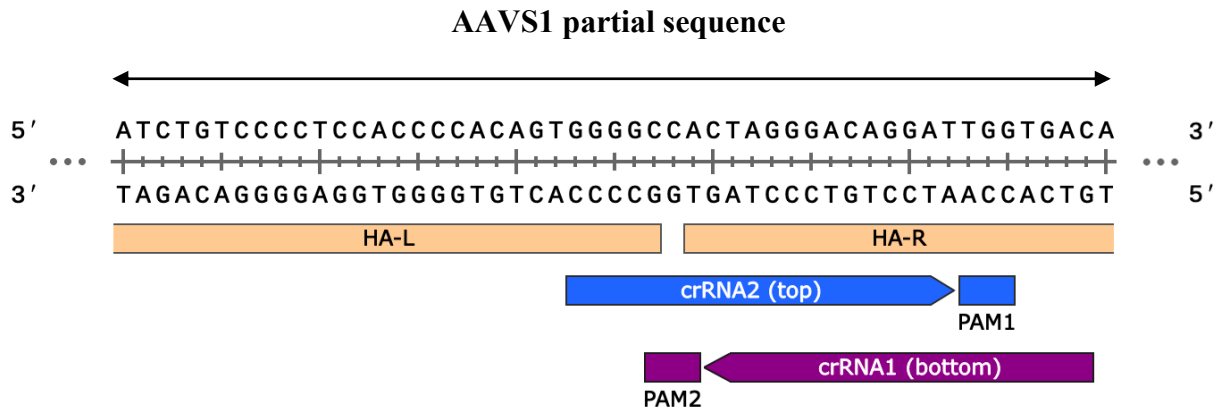


Figure 1.2: Two crRNA sequences were selected for targeting of the *AAVS1* locus. crRNA sequences were identified between the intersection of the right and left homology arms (HAs) that correspond to those found on the donor repair templates. Both crRNA1 (GTCACCAATCCTGTCCCTAG) and the crRNA2 (GGGGCCACTAGGGACAGGAT) precede 5'-TGG PAM sequences. The Cas9 cleavage sites are located 3 base pairs 5' of the PAM sequence. Image was produced with *SnapGene*.

Next, donor repair templates were designed to induce HDR-mediated gene editing when transfected alongside the Cas9 plasmid containing an sgRNA targeting the *AAVS1* locus. Donor repair templates including Dsup, Dsup-NBD, Dsup-coreNBD, and DsupM6 were produced according to '*Generation of donor repair templates'* in *Methods*. DsupM6 is a truncated version of the Dsup protein that only contains the first 357 amino acids of the sequence, therefore lacking the NBD region and nucleosome binding activity (Figure 1.3). The transgene sequences (*i.e.*, Dsup, Dsup-NBD, Dsup-coreNBD, and DsupM6) were inserted into an *AAVS1* targeting vector so that they would be located between left homology arm (HA-L) and right homology arm (HA-R) and under the control of a CMV early enhancer/chicken β actin (CAG) promoter. Both monomeric enhanced green fluorescent protein (mEGFP) tagged and untagged donor repair templates were generated in this study. Downstream of the CAG promoter, a mEGFP was attached to the N-terminus of Dsup, Dsup-NBD, Dsup-coreNBD, and DsupM6 sequences via a

glycine rich flexible glycine (G)-serine (S) rich linker sequence (3x-GGGS). Following transfection of a Cas9 plasmid and donor repair template pair, it was predicted that CRISPR/Cas9-mediated HDR would result in the genomic insertion of the transgenic sequences, either tagged with mEGFP or untagged, between the corresponding HAs at the *AAVS1* locus (Figure 1.4). *'Colony PCR'* was performed to select for donor repair template clones positive for the desired transgenes before confirmation by '*sequencing.'* Confirmed clones were amplified according to *'large-scale plasmid amplification'* in preparation for transfection reactions (see *Methods*). Donor repair templates lacking an mEGFP tag were not further tested in this study.

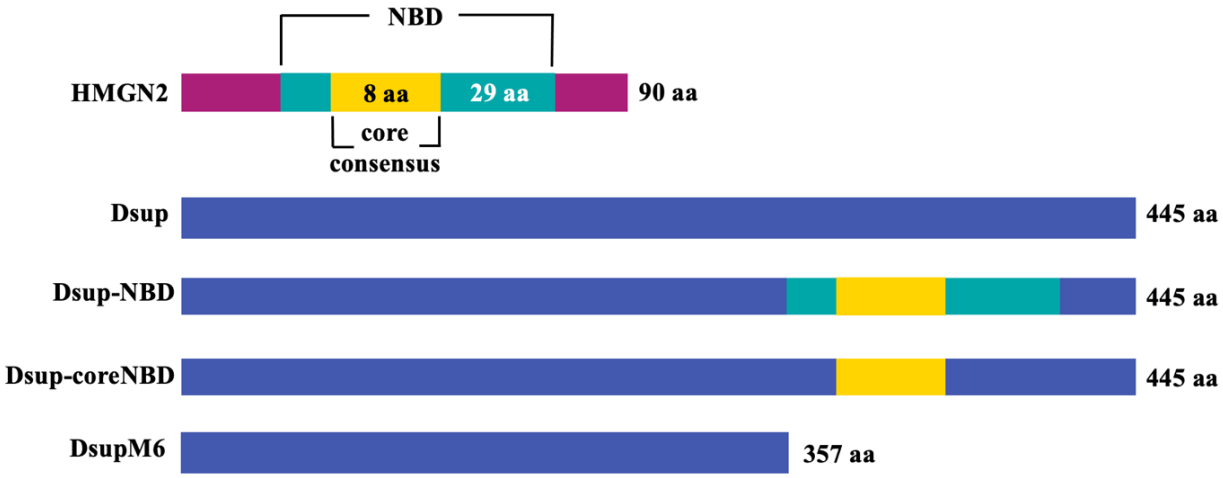


Figure 1.3: Dsup, Dsup-NBD, and Dsup-coreNBD protein sequences were designed for expression in human cells. The entire human HMGN2 nucleosome binding domain (NBD), which is composed of 29 amino acids, replaces that of Dsup in the Dsup-NBD sequence. The core consensus amino acid sequence within the NBD of HMGN2 region is RRSARLSA. This sequence replaces the homologous 8 amino acid Dsup sequence in the Dsup-coreNBD construct. Dsup, Dsup-NBD and Dsup-coreNBD all contain 445 amino acids while DsupM6 is a truncated version of Dsup that does not contain an NBD region and is only 357 amino acids long. Figure is not to scale.

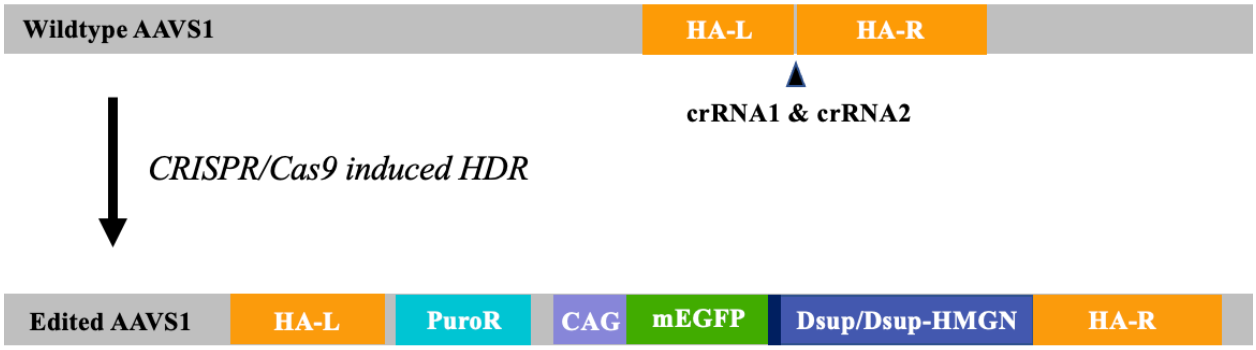
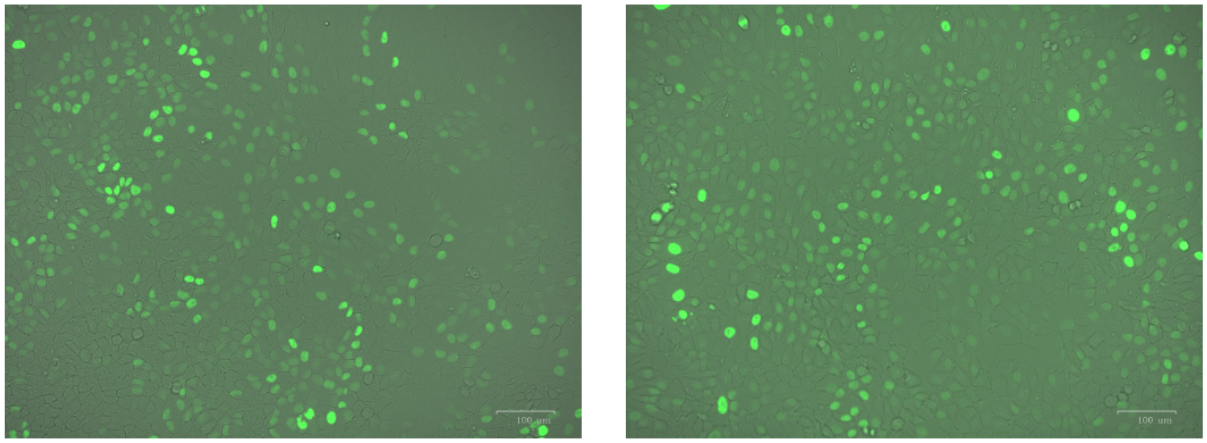


Figure 1.4: The *AAVS1* locus in MCF10A cells was targeted for CRISPR-Cas9 HDR directed integration of Dsup, Dsup-NBD, Dsup-coreNBD, and DsupM6. The left homology arm (HA-L) and right homology arm (HA-R) regions correspond to those included on Addgene Plasmid 64215 for targeting of the AAVS1 region. The crRNA1 and crRNA2 Cas9 directed cut sites are located between HA-L and HA-R. The transgenic sequence includes a puromycin resistance gene (PuroR), a monomeric enhanced green fluorescent protein (mEGFP) followed by the Dsup sequence (DsupM6, Dsup-NBD, and Dsup-coreNBD are not shown but replace the Dsup sequence for the corresponding transgenic cell lines). mEGFP is linked to Dsup by a 3x(GGGS) linker sequence (not detailed). Both mEGFP and Dsup are preceded by a CMV early enhancer/chicken β actin (CAG) promoter. Figure is not to scale.

CHAPTER 2: CELL LINE ESTABLISHMENT

The next goal of this study was to generate MCF10A cells stably expressing Dsup and Dsup-HMGN2 hybrid proteins in preparation for comparative DNA damage analysis. Once high-quality, endotoxin-free plasmid DNA was prepared for both Cas9/crRNA plasmids and donor repair templates, wildtype MCF10A cells obtained from Kadonaga Laboratory stocks (referred to as 'laboratory MCF10A cells' throughout this study) were transfected (see '*Transfection'* in *Methods).* Clonal cell lines were isolated through two rounds of fluorescence-activated cell sorting (FACS) (see *'FACS and clonal amplification'* in *Methods).* The first round of cell sorting selected cells positive for the transient expression of mCherry (associated with Cas9 expression) and mEGFP (associated with donor template expression) fluorescence about 24 hours after transfection. About two weeks later, a second round of sorting selected for cells stabling expressing mEGFP. As clonal populations were expanded, however, a mosaic expression pattern was observed in all cell populations and therefore epigenetic silencing was suspected. Cells within one clonal population showed varying levels of expression with at least 30% of cells showing no mEGFP expression across multiple clones for all donor repair templates (Dsup, Dsup-NBD, Dsup-coreNBD) and both Cas9 plasmids (crRNA1, crRNA1) (Figure 2.1). While mEGFP expression was localized to the nucleus in cells transfected with Dsup, Dsup-NBD, and Dsup-coreNBD donor repair templates, DsupM6 cells did not demonstrate nuclear localization and this construct was therefore not included in repeated experiments. In order to perform comparative DNA damage analysis, clonal populations would need to exhibit more homogeneous patterns of mEGFP expression in which most cells (>90%) show at least some level of expression.



mEGFP-Dsup/crRNA1 mEGFP-Dsup/crRNA2

Figure 2.1: Silencing was observed in all laboratory MCF10A cells transfected with a donor repair template and a Cas9 plasmid. mEGFP-Dsup (shown) and mEGFP-Dsup-HMGN2 hybrid (not shown) donor templates were transfected alongside both crRNA1 and crRNA2 Cas9 plasmids. Similar patterns of expression were observed for each transfection pair. Each image consists of a white light image merged with a green fluorescence image. Images were captured on ZOE Fluorescence Cell Imager (Bio Rad #1450031). Image scale: 100 µm.

Both FACS clonal isolation and antibiotic selection have been shown to improve homogeneity of expression during clonal cell line isolation (Kaufman et al., 2008). Therefore, treatment of FACS-isolated clonal populations with the antibiotic corresponding to the resistance gene located adjacent to the transgene was performed to assist in the recovery of a homogeneous phenotype. A puromycin resistance gene was included upstream of the CAG promoter within the donor repair template so that resistance to the antibiotic would be conferred to only successfully transfected cells. It was hypothesized that treatment with puromycin would eliminate cells in which epigenetic silencing had reduced or eliminated mEGFP expression and therefore that of the adjacent Dsup and Dsup variants (Kaufman et al., 2008). Wildtype laboratory MCF10A, laboratory MCF10A cells transfected with Cas9/sgRNA1 and Dsup donor vector, and laboratory MCF10A cells transfected with Cas9/crRNA2 and Dsup donor vector were treated with puromycin at varying concentrations to determine the ideal concentration to reduce silencing

without causing any visual toxicity to cells (See *'Puromycin kill curve'* in *Methods*). Although puromycin did appear to improve homogeneity of mEGFP expression, survival of wildtype laboratory MCF10A cells at surprisingly high concentrations of puromycin (up to 10μg/mL) led to the conjecture that the wildtype laboratory MCF10A cells may contain inherent puromycin resistance (Figure 2.2). It was concerning that the wildtype MCF10A cells used in transfection reactions contained puromycin resistance as this was the antibiotic resistance gene located within the donor repair templates designed to aid in isolation of positive Dsup and Dsup-HMGN2 clones. I amplified the puromycin resistance gene from wildtype laboratory MCF10A cells to determine if I needed to transfect an alternative stock of MCF10A cells before proceeding to DNA damage analysis (Figure 2.3).

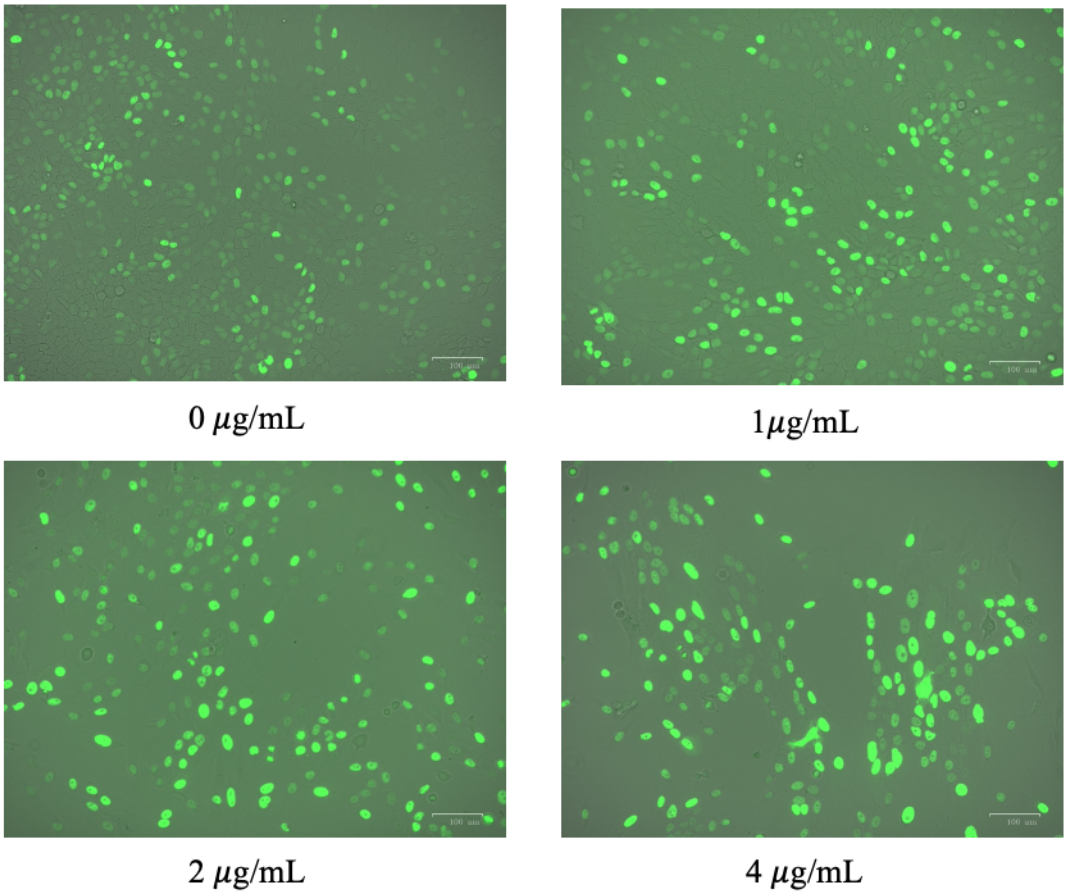


Figure 2.2: Puromycin treatment improves homogeneity of mEGFP-Dsup expression in laboratory MCF10A transfected with mEGFP-Dsup donor repair template and Cas9/sgRNA1 plasmid. A puromycin kill curve was performed with increasing amounts of puromycin, including 1, 2, and 4 μg/mL, and compared to untreated cells (0 μg/mL). Each image consists of a white light image merged with a green fluorescence image. Images were captured on ZOE Fluorescence Cell Imager. Image scale: 100 μm.

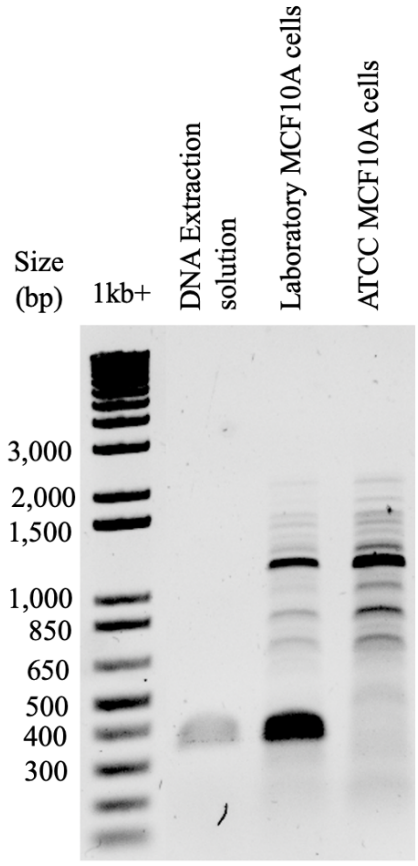


Figure 2.3: Puromycin resistance gene is present in wildtype laboratory MCF10A cell lines. The puromycin resistance gene was amplified by PCR in wildtype laboratory MCF10A cells and MCF10A cells obtained from ATCC. The expected product size of the puromycin gene is 330 bp. DNA extraction solution was used to isolate DNA from cells and included as a negative control. PCR products were run on 0.8% agarose gel in 1x TBE buffer and stained with ethidium bromide.

In order to harness the potential of antibiotic selection to improve homogeneity of expression and prevent transcriptional silencing, the wildtype MC10A cell line could not contain the resistance gene present within the donor repair template (Kaufman et al., 2008). The presence or absence of a puromycin resistance gene was analyzed by PCR in both the laboratory MCF10A cells and an MCF10A cell line obtained from American Type Culture Collection (ATCC) (referred to as 'ATCC MCF10A cells' throughout this study) (see *'PCR amplification of puromycin gene,' 'Cellular DNA extraction,' and* '*Cell line'* in *Methods*). The presence of a band

at size 330 bp indicates that presence of the puromycin gene, therefore the ATCC MCF10A cell line lacks puromycin resistance while the puromycin resistance gene is indeed present in the laboratory MCF10A cell line (Figure 2.3). DNA extraction solution was run as a negative control, however, the faint signal observed within this sample was determined to be inconsequential contamination by repeated experiments. The conclusion that the laboratory MCF10A cells contained puromycin resistance was surprising as this cell line was not known to contain resistance to puromycin. I anticipated that treating the ATCC MCF10A cell lines with puromycin following transfection with the Dsup and Dsup-HMGN2 hybrid donor repair templates alongside the Cas9 plasmid would allow for selection of successful HDR-edited cells. Ensuring that the transfected MCF10A cell line did not contain puromycin resistance was a critical step in the development of stable cell lines expressing Dsup and Dsup-HMGN2 hybrid proteins that could be then used in a DNA damage assay.

Before repeating transfection reactions in ATCC MCF10A cells, I assessed whether the designed CRISPR/Cas9 system was indeed effective for inserting the desired transgene at the intended locus. I separately amplified the right and left homology arm regions to confirm the insertion of the donor repair template at the *AAVS1* locus (see *PCR reactions* in *Methods*). Primers for amplification of HA-L and HA-R matched to the *AAVS1* locus (outside of the homologous region) and the transgene-containing sequence, which was expected to be located between the homology arms (Figure 2.4). For HA-L, a band with the expected size of 1,258 bp was amplified from laboratory MCF10A cells transfected with mEGFP-Dsup repair template alongside both sgRNA1 and sgRNA2 Cas9 plasmids (Figure 2.5a). For the HA-R, a band around expected size of 1,466 bp was amplified from laboratory MCF10A cells transfected with mEGFP-Dsup repair template alongside both sgRNA1 and sgRNA2 Cas9 plasmids (Figure

2.5b). DNA extraction solution was included as a negative control along with the untransfected MCF10A cell lines (Figure 2.5). This successful amplification of HA-L and HA-R demonstrates that HDR was successful with both sgRNA sequences. As discussed in *Chapter 1*, only the sgRNA1 Cas9 plasmid was used throughout repeated transfection experiments in order to reduce sample load.

Figure 2.4. PCR amplification of the homology arm regions was designed to confirm crRNA1 and crRNA2 efficacy. Homology arms were amplified from both a non-homologous location on the *AAVS1* sequence and a region within the transgene sequence to confirm intended transgenic insertion. HA-L oligos include: (a) CATTGTCACTTTGCGCTGC and (b) GACGCGCGTGAGGAAGA. HA-R oligos indicated are as follows: (c) GGAAAAGCCGCTGCCACAAA and (d) GGAGGAGAATCCACCCAAAAGGC. Figure is not to scale.

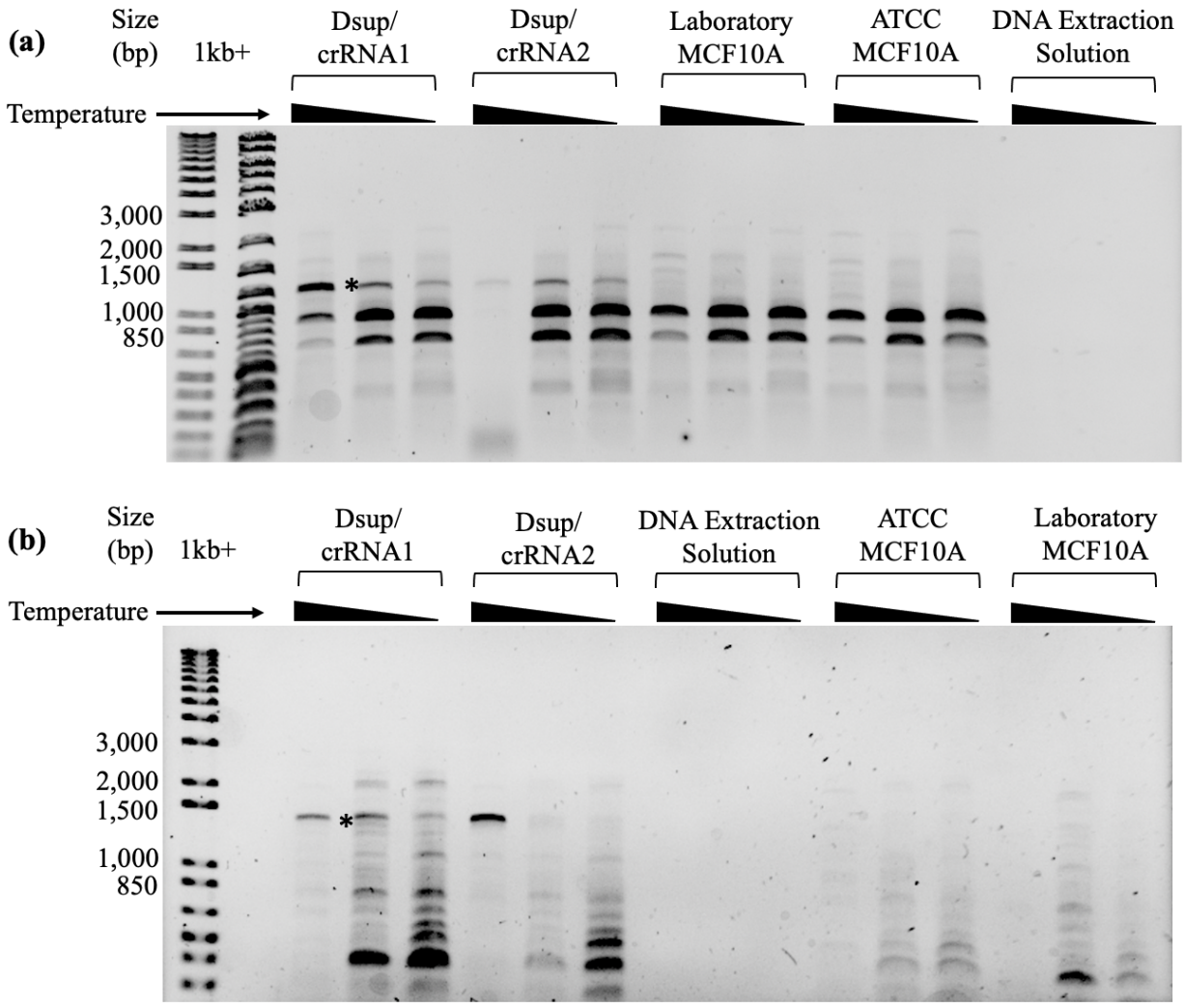


Figure 2.5: Insertion of transgene at *AAVS1* locus was confirmed for both crRNA sequences. The expected product size of the amplified region containing the left homology arm is 1,258 bp (a). The expected product size of the amplified region containing the right homology arm is 1,466 bp (b). Expected product sizes are denoted with asterisks. Three annealing temperatures were used for each experimental condition. PCR products were run on 0.8% agarose gel in 1x TBE buffer and stained with ethidium bromide.

ATCC MCF10A cells were transfected according to '*Transfection protocol'* detailed in *Methods*. Donor repair templates Dsup, Dsup-NBD, and Dsup-coreNBD were transfected in separate reactions alongside the sgRNA1/Cas9 plasmid. 24 hours after transfection, cells were pool sorted based on transient expression of mCherry (Cas9 plasmid) and mEGFP (donor repair

template) (See *'FACS and clonal amplification'* in *Methods*). Culture medium containing 0.5 μg/mL puromycin was added approximately 72 hours following time of sort to induce antibiotic selection of edited cells. This concentration was determined to be the lowest effective concentration that would kill wildtype MCF10A cells after 7 days (see *'Puromycin kill curve'* in *Methods*). About 2 weeks after the initial sort, a second round of FACS was performed using the pooled populations to isolate Dsup and Dsup-HMGN2 clonal populations stabling expressing mEGFP. In the week following this second sort, the plates were screened for single cell population: clones suspected to have started from more than one cell were eliminated. This was estimated based on the doubling time of MCF10A cells. Not only was puromycin able to enhance the population of successful HDR cells in the pooled cell populations collected by FACS following transfection of ATCC MCF10A cells, but puromycin treatment also successfully reduced observed silencing in clonal populations. Established MCF10A cells containing mEGFP-tagged Dsup, Dsup-NBD and Dsup-coreNBD proteins were preserved according to '*Freezing cells*" in *Methods*.

CHAPTER 3: CELL LINE ANALYSIS

In order to perform comparative DNA damage protection analysis, all mEGFP-tagged Dsup or Dsup-HMGN2 hybrid proteins had to not only demonstrated similar expression, but they had to contain the same zygosity: all cell lines had to be entirely homozygous (*i.e*., contain two copies of the transgene) or heterozygous (*i.e*., contain one copy of the transgene). To determine zygosity, the *AAVS1* region was PCR amplified from potential cell lines (see '*PCR amplification of the AAVS1 region'* in Methods). Potential clonal cell line sets were selected after discarding potential clones that did not demonstrate nuclear localization and after selection of the most homogeneous clones demonstrating the greatest overall level of expression. *'Cellular DNA extraction'* was performed to isolate DNA from cell pellets for two candidate clones for each of the following: Dsup, Dsup-NBD, and Dsup-coreNBD. The *AAVS1* region was amplified from outside both the left and right homology arms regions to determine if the candidate cell lines contained wildtype *AAVS1* alleles (Figure 3.1). The presence of the expected 2,026 bp band suggests that at the least one wildtype *AAVS1* allele is present and therefore the cell line may contain one wildtype allele and one edited allele or two wildtype alleles. Meanwhile, the absence of a band at 2,026 bp suggests that the cell line lacks wildtype *AAVS1* alleles. Untransfected ATCC MCF10A cells were included as a positive control and were expected to demonstrate a band at the expected size (corresponding to two wildtype alleles), while DNA extraction solution was included as a negative control. The Dsup cell line was selected from set B and the Dsup-NBD and Dsup-coreNBD cell lines were selected from set A as all showed a band at the expected 2,026 bp size (Figures 3.2). Precise insertion of the different transgenes at the *AAVS1* locus of the selected cell lines remains to be confirmed by PCR amplification of the *AAVS1* region containing the entire repair template, however, PCR reactions were unsuccessful and

therefore require optimization to account for the large size of the desired amplicon. The selected populations demonstrate nuclear localization characteristic of Dsup as well as relatively homogeneous expression (Figure 3.3). Within each population of the selected cell lines, > 90% of cells show some degree of expression, although there is still a large range in the expression levels observed. Differences in mEGFP expression may be attributed to the lack of explicit cell cycle synchronization. Variances in expression, however, were considered mitigated for the purposes of this study by selecting comparable high cell densities for the alkaline comet assay, as discussed in *Methods* (see *'Alkaline comet assay'*).

Figure 3.1: The AAVS1 region was PCR amplified to determine cell line zygosity. Left and right homology arms contained within the donor repair template correspond to complementary sequences within the human *AAVS1* locus. PCR amplification of the wildtype *AAVS1* locus was performed to determine cell line zygosity. Oligos used for PCR amplification, which align to the *AAVS1* sequence outside of the homologous regions, are as follows: (a) GGAACTCTGCCCTCTAACGC (b) GCTTCTTGGCCACGTAACCT.

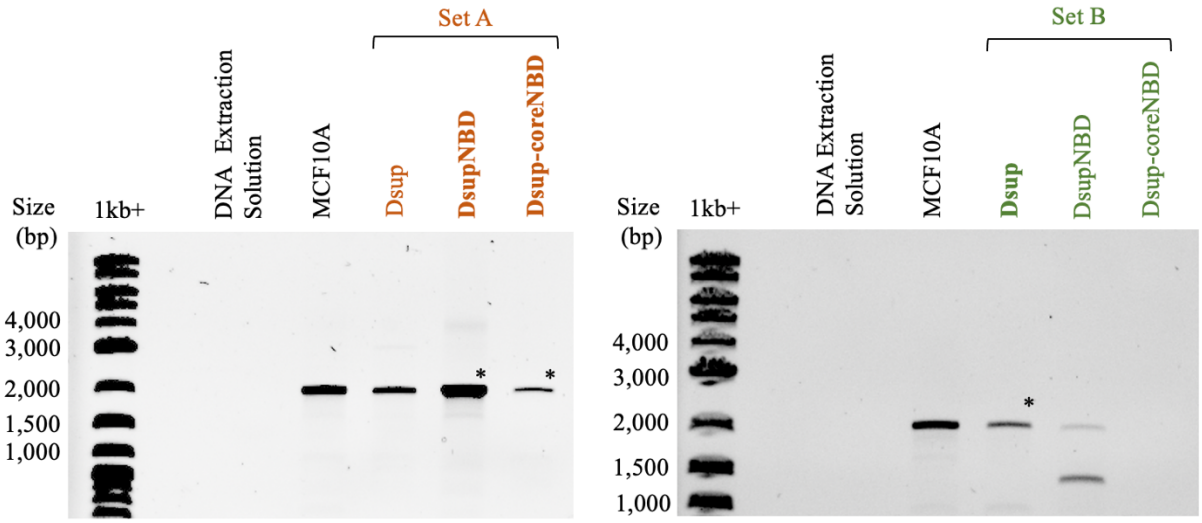
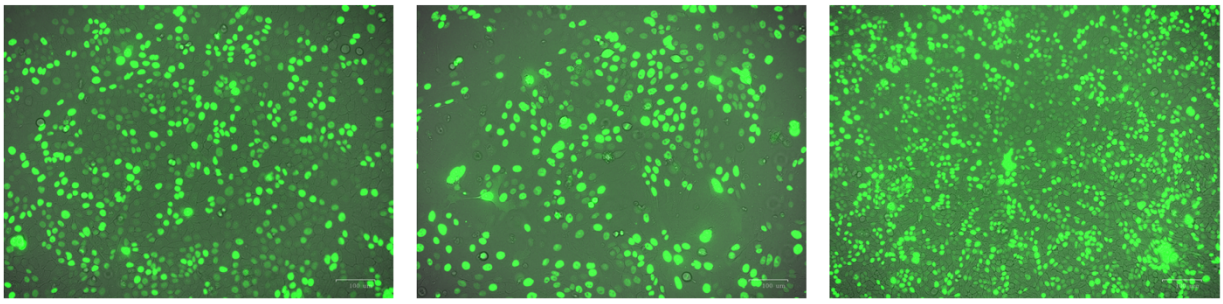


Figure 3.2 mEGFP-Dsup and mEGFP-Dsup-HMGN2 hybrid cell lines containing at least one wildtype *AAVS1* allele were selected for future comparative DNA damage analysis. PCR amplification of the *AAVS1* region was performed to confirm the presence of at least one wildtype allele and therefore the potential of heterozygosity. Oligos for amplification were designed to amplify the region containing the entire homology arms used for integration of transgenic region. The expected product size is 2,026 bp. From set A, clonal populations for mEGFP-Dsup-NBD* and mEGFP-Dsup-coreNBD** were selected for use in future experiments. From cell line set B, the clonal population for Dsup*** was selected. Laboratory stock reference: *2h, **3h, ***1c.



mEGFP-Dsup (a) mEGFP-DsupNBD (b) mEGFP-Dsup-coreNBD (c)

Figure 3.3: Homogeneous expression across selected cell lines was required for comparative DNA damage analysis. The donor repair templates indicted were transfected alongside the crRNA1 Cas9 plasmid in ATCC MCF10A cells and treated with 0.5 μg/mL puromycin throughout culture. Cell populations (a), (b), and (c) were selected as mEGFP expression is localized to the nucleus and >90% of cells demonstrate some level of mEGFP expression. Displayed cells lines were selected for use in the alkaline comet assay as they are all suspected to be heterozygous for transgene expression. Each image consists of a white light image merged with a green fluorescence image. Images captured on ZOE Fluorescence Cell Imager. Image scale: 100 µm.

Selection of clonal Dsup and Dsup-HMGN2 hybrid populations was based on observed mEGFP expression, however, expression of mEGFP did not guarantee expression of the downstream Dsup or Dsup-HMGN2 hybrid proteins. Because the mEGFP protein sequence was linked to Dsup and Dsup-HMGN2 hybrid proteins via a 3x(GGGS) linker sequence at the N-terminus, successful production of full-length proteins had to be confirmed by western blot as premature termination could have resulted in mEGFP expression without Dsup or Dsup-HMGN2 expression. In order to confirm the expression of Dsup and Dsup-HMGN2 hybrid proteins by western blot analysis, an anti-Dsup antibody from rabbit serum had to be generated and validated.

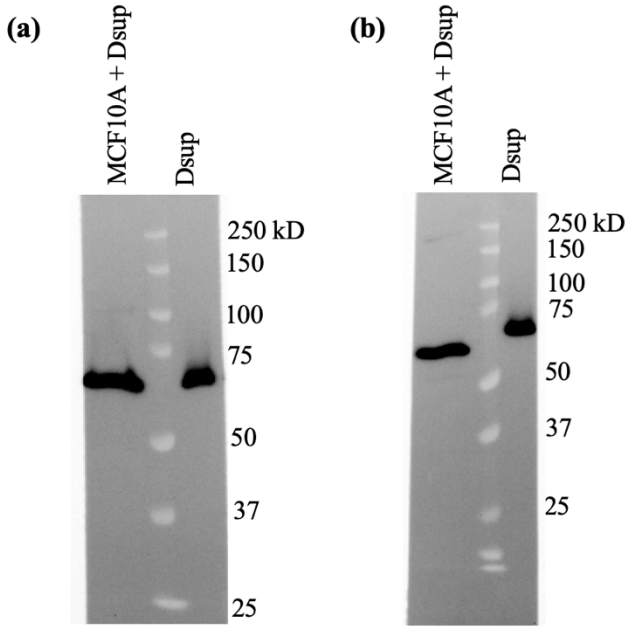


Figure 3.4: Dsup antibodies were successfully produced in rabbits 37838 and 37839. Rabbit (a) 37838 and (b) 37839 sera were collected by Pocono Rabbit Farm. Anti-Dsup antibodies were diluted in 5% milk at a concentration of 1:10,000. The expected band size for Dsup is about 65 kD. Anti-Dsup antibodies were tested against Dsup extract in the presence or absence of MCF10A cellular extract.

Rabbits 37838 and 37839 were selected for antibody generation after western blot results confirmed that existing sera did not exhibit antibody background signal at the approximate 65 kD molecular weight of Dsup. Immunization with the Dsup antigen was performed by the Pocono Rabbit Farm using a recombinant Dsup protein containing only the first 357 amino acids of Dsup (Dsup M6). Because the Dsup and Dsup-HMGN2 hybrids are homologous for the first 357 amino acids of the protein sequences, this anti-Dsup antibody would be able to detect both Dsup and Dsup-HMGN2 hybrid protein expression. Western blots were performed with sera collected after immunization of the two rabbits to determine if anti-Dsup antibodies were successfully generated. Dsup antibody candidates from both rabbits (37838 and 37839) were diluted in 5% milk in TBST buffer at a dilution of 1:10,000 (Figure 3.4). Samples included both 50 ng of

recombinant Dsup and 20 ng of recombinant Dsup added to MCF10A protein extract to simulate detection in established transgenic cell lines. The anti-Dsup antibodies were successfully generated in both rabbits as a signal at the expected Dsup size was observed (Figure 3.4 and 3.5). These anti-Dsup antibodies could therefore be used to determine if Dsup and Dsup-HMGN2 hybrid proteins are being expressed in the cell lines generated in this study (See the *'Western blot analysis'* section in *Methods*).

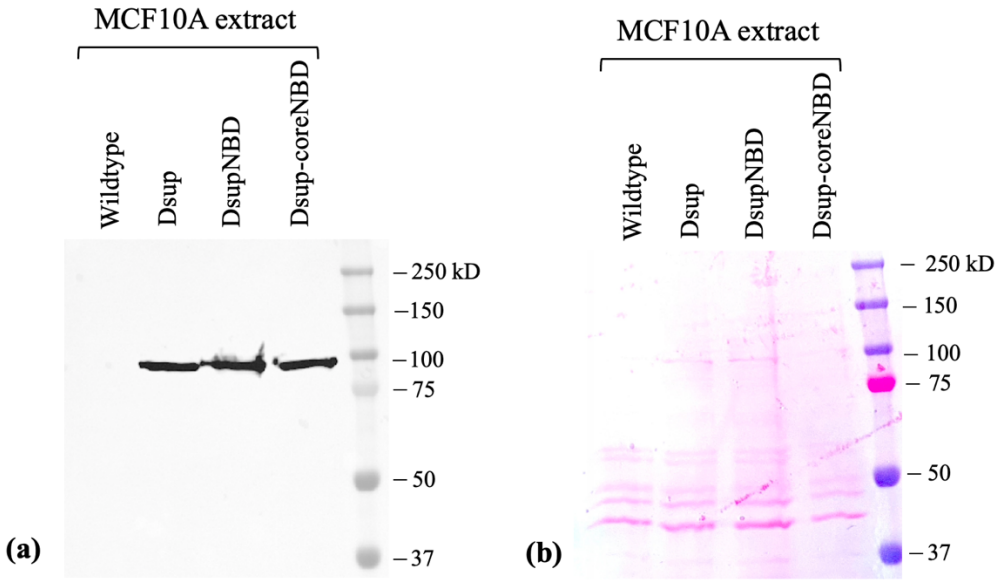


Figure 3.6: Dsup and Dsup-HMGN2 hybrid protein expression in MCF10A cells confirmed by western blot analysis. (a) Membrane was incubated with the primary anti-Dsup antibody followed by secondary anti-rabbit antibody. mEGFP-3x(GGGS)-Dsup and mEGFP-3x(GGGS)-Dsup-HMGN2 hybrid proteins were expected to show a band at around 100 kD, or the combined molecular weight of mEGFP (~ 27 kD) and Dsup/Dsup-NBD/Dsup-coreNBD (~ 65 kD). (b) Ponceau stained membrane* shows background protein content.
*Error: Ponceau stain performed after western blot experiment. Image edited to reduce background stain.

Once anti-Dsup antibodies demonstrated successful functionality, protein expression in the selected cell lines was confirmed by western blot analysis (Figure 3.6). A band appearing

slightly below 100 kD for all lanes containing Dsup or Dsup-HMGN2 cellular extract reflects the expression of a Dsup or Dsup-HMGN2 hybrid protein linked to mEGFP. Western blot results demonstrate that selected MCF10A cell lines, which were transfected with mEGFP-Dsup, mEGFP-Dsup-NBD, or mEGFP-Dsup-coreNBD donor repair templates alongside the crRNA1/Cas9 plasmid, stably express the transgenes (Figure 3.6a). After protein expression was confirmed, DNA damage protection capacity of the established transgenic cell lines could be evaluated with the alkaline comet assay.

Prior to determining if the modification of Dsup with the NBD of human HMGN2 could optimize DNA damage protection in human cells, wildtype Dsup protection against DNA damage in non-tumorigenic MCF10A cells had to be confirmed. An alkaline comet assay was conducted to assess the tail DNA damage percentage of wildtype ATCC MCF10A cells (Control) and MCF10A cells expressing Dsup (Dsup) after exposure to 25 μM or 50 μM of hydrogen peroxide (See *Methods* for complete *'Alkaline comet assay'* protocol). During the electrophoresis step of the alkaline comet assay, fragmented DNA migrates away from the nucleus forming a tail which can then be measured to indicate the fraction of DNA breaks, which include both single and double stranded DNA breaks (Lu et al., 2017). Control and Dsup cells not treated with hydrogen peroxide showed a similar background mean tail DNA percentage of 4.73% and 4.76%, respectively. Control cells treated with 25 μM hydrogen peroxide surprisingly showed a significantly smaller mean tail DNA percentage than Dsup cells (30.50% and 56.09% for Control and Dsup cells, respectively; $p < 0.001$). Control cells treated with 50 μM hydrogen peroxide also showed a significantly smaller mean tail DNA percentage than Dsup cells (68.53% and 69.88% for Control and Dsup cells, respectively; $p < 0.05$) (Figure 3.7, Table 3.1). Under the conditions of my experiment, MCF10A cells containing Dsup do not demonstrate a reduction in

the amount of DNA damage inflicted by hydrogen peroxide when compared to untransfected MCF10A cells. Further experimentation, therefore, was not conducted to assess the DNA damage protection capacity of MCF10A cells containing Dsup-NBD or Dsup-coreNBD. Future experiments that refute my finding would be necessary before proceeding to investigate whether Dsup-HMGN2 hybrid proteins could improve the DNA damage protection against hydrogen peroxide.

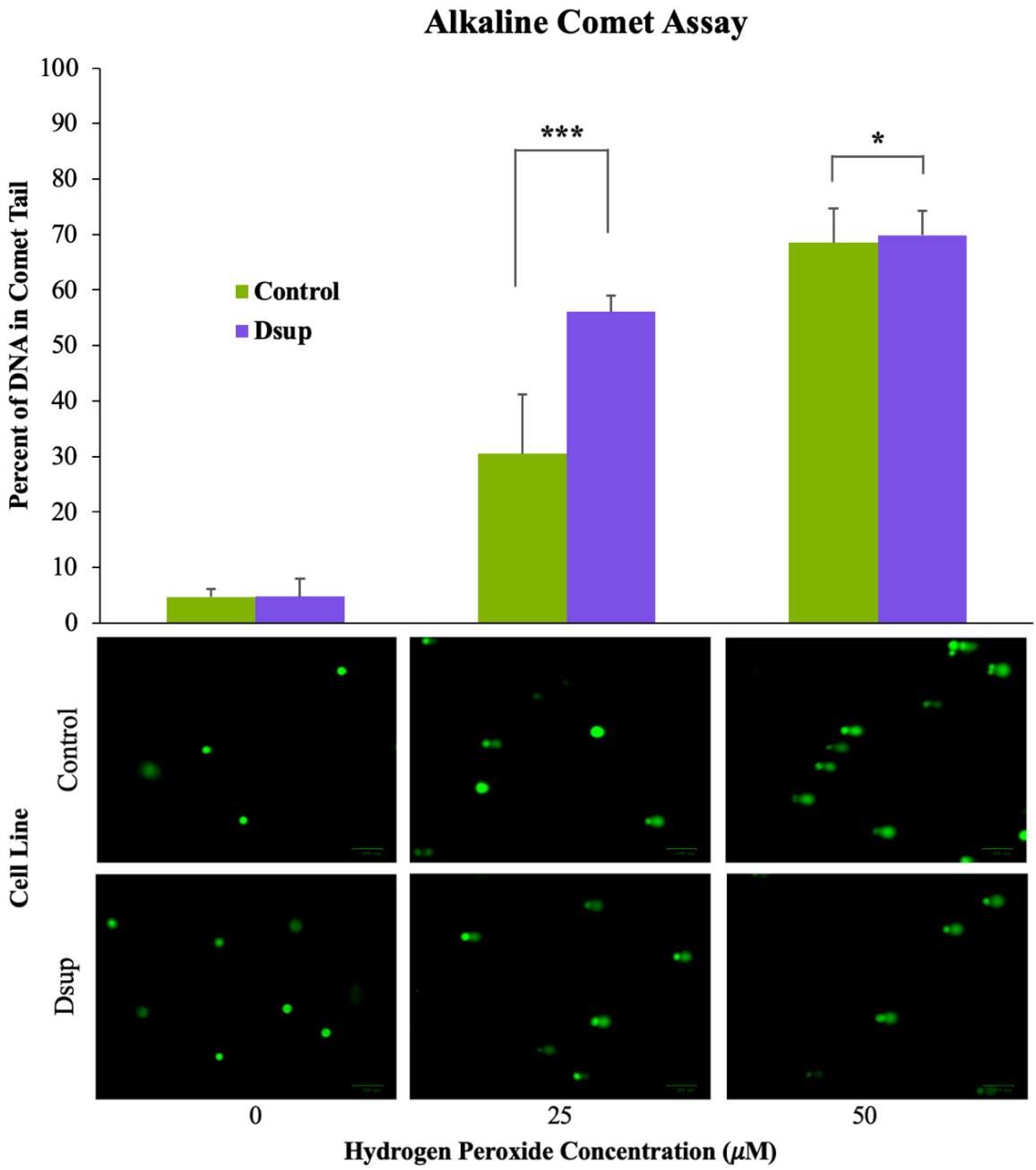


Figure 3.7: Dsup does not appear to protect against DNA damage by hydrogen peroxide in MCF10A cells. An alkaline comet assay was performed, and the percentage of DNA found in the comet tail was quantified after exposure to 25 μM and 50 μM hydrogen peroxide for 3 minutes at 4 °C. Significance values were calculated for each exposure: *** $p < 0.001$, *$p < 0.05$ (Welch's t-test). At least 150 nuclei were quantified for each condition and three separate experiments were averaged (top). Representative images for each cell line under each experimental condition are shown with the direction of electrophoresis running from left to right. Image scale: 100 μm (bottom).

Table 3.1: Dsup does not appear to protect against DNA damage by hydrogen peroxide in MCF10A cells.

Hydrogen Peroxide	Cell Line	Average Tail DNA Percentage	Standard Deviation Tail DNA Percentage	Number of Nuclei
0 μM	Control	4.726	1.464	187
	Dsup	4.758	3.177	170
25 μM	Control	30.499	10.763	263
	Dsup	56.094	2.880	185
50 μM	Control	68.526	6.122	191
	Dsup	69.884	4.380	162

CHAPTER 4: CONCLUSION

The goal of my experimental inquiry was to determine if a damage suppressor (Dsup) protein containing the nucleosome binding domain (NBD) of the human high mobility group nucleosomal binding domain 2 (HMGN2) protein could optimize DNA damage protection in human cells. Dsup was shown to protect against DNA damage in tumorigenic HEK293 cells by Hashimoto and colleagues in 2016 and by Ricci and colleagues in 2021, however, Dsup's protection capacity had not yet been assessed in a non-tumorigenic cell line. I established stable MCF10A cells (non-tumorigenic epithelial cells) expressing human codon-optimized Dsup and Dsup-HMGN2 hybrid proteins under the regulation of a CAG promoter at the *AAVS1* safe-harbor locus using CRISPR/Cas9 technology. In order to perform comparative DNA damage analysis across Dsup and Dsup-HMGN2 cell lines, similar expression patterns and zygosity had to be confirmed for the transgenes. Following mitigation of transcriptional silencing with the addition of puromycin as a selective pressure throughout establishment of stable cell lines and zygosity analysis by PCR, expression of Dsup was confirmed by western blot analysis. Before performing comparative DNA damage analysis across established Dsup and Dsup-HMGN2 cell lines, Dsup damage protection in non-tumorigenic cells had to be confirmed. Following treatment with 25 μM and 50 μM of hydrogen peroxide, an alkaline comet assay was conducted, and Dsup-containing cells showed a significantly greater mean tail DNA damage percentage when compared to wildtype MCF10A cells. This result was surprising as previous studies have demonstrated a reduced mean tail DNA damage percentage in cells (albeit tumorigenic) containing Dsup following exposure to hydrogen peroxide (Hashimoto at al., 2016). Before continuing studies that investigate whether the established Dsup-HMGN2 cell lines contain

optimized DNA damage protection, Dsup must demonstrate the capacity to protect against cellular damage in non-tumorigenic cells.

Under the conditions of the comet assay performed in this study, Dsup does not appear to protect MCF10A cells from DNA damage by hydrogen peroxide. This unexpected finding contrasts with previous conclusions in which HEK293 cells expressing Dsup improved protection against hydrogen peroxide-induced DNA damage relative to untransfected HEK293 cells (Hashimoto et al., 2016). In contrast to previous studies, the Dsup sequence used in this study was codon-optimized to be expressed in human cells. This could have resulted in the over-abundant expression of Dsup, resulting in a toxic level of protein present in the established cell lines. The significant increase in DNA damage I observed in non-tumorigenic MCF10A cells expressing Dsup may be attributed to Dsup interference with transcription and replication processes. Previous research demonstrated the inability to establish stable HEK293 cell lines containing only the C-terminal region of Dsup and the formation of abnormal DNA aggregates within transfected cells (Hashimoto et al., 2016). Dsup may form aggregates when over-expressed that sterically interfere with cellular processes in non-tumorigenic cells.

Despite my observation that MCF10A cells stabling expressing Dsup do not appear to increase tolerance to DNA damage, additional studies should be conducted to confirm or refute this finding. One limitation of my experimental design was that the Dsup-containing cell line was treated with puromycin throughout culture while the control MCF10A cell line was not under puromycin treatment prior to use in the alkaline comet assay. Puromycin resistance was contained within the homology directed repair (HDR) template adjacent to Dsup, and preliminary experiments confirmed that treatment with puromycin as a selective pressure was necessary to increase homogeneity and persistence of apparent mEGFP expression and therefore

Dsup (Kaufman et al., 2008). Puromycin resistance, however, was not present in the untransfected, wildtype MCF10A cell line used as the control in the DNA damage analysis, therefore, this cell line was not treated with puromycin. While DsupM6 was designed as a control cell line containing puromycin resistance but lacking an NBD, it was determined to be unsuitable due to the lack of nuclear localization exhibited in preliminary experiments. To improve the experimental design, an MCF10A cell line containing a Dsup protein with an inactive NBD but an active nuclear localization signal (NLS) should be included as an additional control in repeated DNA damage assessment experiments. An inactive NBD would prevent Dsup from binding to nucleosomes and therefore inhibit its functional mechanism, while an NLS would allow for nuclear localization exhibited by the mEGFP-Dsup and mEGFP-Dsup-HMGN2 hybrid cell lines. This control cell line could be under puromycin treatment and could therefore serve a superior experimental control when compared to untransfected MCF10A cells. An alternative solution would be to cease the addition of an antibiotic selective pressure once homogeneous expression has been observed. For example, Hashimoto and colleagues treated HEK293 cells with the antibiotic compound G418 for only three weeks following transfection (2016). Future experiments continuing the work performed in this study would need to confirm that homogeneity of expression is not lost following the removal of puromycin as a selective pressure. The establishment of an improved control cell line in future studies is critical in verifying if Dsup can protect against DNA damage in non-tumorigenic MCF10A cells. It could then be determined if hybridized Dsup-HMGN2 proteins could provide greater DNA protection within human cells. A robust non-tumorigenic cell line containing Dsup-HMGN2 hybrid proteins could hold potential for advancements of biotechnologies such as gene therapy.

CHAPTER 5: METHODS

Cloning procedures. Laboratory bench was sterilized using dH_2O and 75% EtOH before starting procedures.

Bacterial Transformation. A 10 μL aliquot of chemically competent 5-alpha cells (NEB C2987I) was incubated on ice for about 10 minutes (or until thawed). 20 ng of cold plasmid DNA (incubated on ice) was added to cells and the tube was mixed by flicking 3 times. Cells and DNA mixture were incubated on ice for 30 minutes. Heat shock was then performed at 42 ℃ for 30 seconds and followed by a 2-minute incubation on ice. 140 μL to 160 μL SOC media was then added, and cells were incubated at 37 ℃ for 1 hour with shaking. Cells were plated using autoclaved glass beads on pre-warmed Lennox broth (LB) agar plates (prepared by the Kadonaga Laboratory) containing antibiotics and incubated for 12 to 16 hours at 37 ℃. Plates were wrapped in parafilm and stored at 4 ℃.

Liquid culture. An aliquot of LB liquid medium was retrieved from laboratory stocks. Antibiotics were vortexed and added to LB to achieve the desired final concentrations: ampicillin reaching a final concentration of 80 μg/mL, kanamycin reaching a final concentration of 40 μg/mL (antibiotic dependent on plasmid resistance). LB containing antibiotic was then added to a conical tube, ensuring that the volume of the container for liquid culture was greater or equal to 5x the volume of the liquid medium. Sterile plastic inoculating loops or sterile pipet tips were used to add a bacteria colony to the medium. Lids were taped so that tubes were open to air but still covered. Samples were incubated for about 16 hours at 37 ℃ with shaking.

Gel electrophoresis. Agarose gels prepared in this study (not including agarose slides prepared in comet assay) were either 0.8% or 1% agarose in 1x Tris-Borate-EDTA (TBE) buffer. 6x Gel Loading Dye (NEB #B7024S) was added to samples to reach a final concentration of 1x

and vortexed to mix. Gel electrophoresis was run until red/pink dye reached the middle to the bottom of the gel (dependent on anticipated product size). Gel was stained in 1 μg/mL ethidium bromide solution for 10 minutes with gentle shaking/rocking. Gel was then rinsed with dH_2O before a 10-minute dH_2O wash with gentle shaking/rocking. Gel was drained of excess liquid before DNA was visualized using AlphaImager software. A 1 kb Plus DNA Ladder (NEB #N3200S) was run alongside samples to determine product sizes throughout the study.

Digestion. Digestion protocols associated with specified NEB restriction enzymes were used throughout the study. The following components were combined in 50 μL final volume reactions: 1x buffer (final concentration), 2 μg of DNA, 10 units of enzyme, ddH_2O was added to reach 50 μL final volume. Reaction was mixed by pipetting before initiation of a 3-hour incubation at 37 ℃.

Ligation. 100 ng of vector was ligated to desired insert using the following calculation: ng insert = [(ng vector * insert size) / (vector size)] * molar ratio. A molar ratio of 5 mols of insert:1 mol of plasmid was used throughout this study. T4 ligase (NEB #M0202S) was used with T4 DNA ligase buffer (NEB #B0202) and the reaction was prepared according to the associated protocol. The ligation was incubated overnight at 16 ℃.

Generation of Cas9 plasmid containing crRNA1 and crRNA2. Addgene plasmid #64324 is a vector for simultaneous expression of Cas9 and a single guide RNA (sgRNA) in mammalian cells. A T2A peptide links the Cas9 to an mCherry reporter gene. Two different crRNA sequences were cloned into Addgene plasmid #64324 for targeting of the *AAVS1* locus. To generate each of the two crRNA (crRNA1 and crRNA2)-containing plasmids, two pairs of oligos were used in two separate cloning reactions. Because the Cas9 plasmid contains a U6 promoter and requires RNA polymerase III mediated transcription of the sgRNA, the crRNA

sequences begin with the letter "G" (Ran et al. 2013). The crRNA sequences, their corresponding PAM sequences, and the oligo sequences used for cloning are as follows.

crRNA1: GTCACCAATCCTGTCCCTAG
PAM: TGG
Oligo sequences for crRNA1:
AAVS1-crRNA1T – CACC GTCACCAATCCTGTCCCTAG
AAVS1-crRNA1B – AAAC CTAGGGACAGGATTGGTGAC

crRNA2: GGGGCCACTAGGGACAGGAT
PAM: TGG
Oligo sequences for crRNA2:
AAVS1-crRNA2T – CACC GGGGCCACTAGGGACAGGAT
AAVS1-crRNA2B – AAAC ATCCTGTCCCTAGTGGCCCC

To integrate crRNA1 and crRNA2 into Addgene plasmid #64324, the plasmid DNA was first digested with BbsI-HF (NEB #R3539S) in CutSmart buffer (NEB #B6004) (see *'Digestion'*). *Gel electrophoresis* was performed using a 0.8% agarose gel. Digested DNA was recovered using Monarch DNA Gel Extraction Kit (NEB #T1020) and associated protocol. The modifications that were made to Monarch DNA Gel Etraction Kit protocol include: the first wash step was performed using 500μL DNA Wash Buffer and incubated for 5 minutes before centrifugation, the repeated wash step was performed using 200 μL DNA Wash Buffer and incubated for 2 minutes incubation, and DNA elution was performed using 20 μL ddH_2O rather than DNA Elution Buffer.

crRNA oligos were annealed and phosphorylated in preparation for ligation reaction with BbsI-digested Addgene #64324. Oligos were first resuspended to 100 μM in ddH_2O. The oligo annealing-phosphorylation reaction was then set up to reach the indicated final concentrations: 10 μM of both top and bottom oligos, 1x T4 ligase buffer (NEB #B0202), 10 units T4 PNK (NEB #M0201S), in ddH_2O. The reaction mixture was incubated for 30 minutes at 37 ℃

followed by a 5-minute incubation at 95 ℃, and finally a cool down to 25 ℃ at a rate of 0.5 ℃/min.

The vector-insert ligation reaction was set up following oligo annealing-phosphorylation. The oligos were first diluted 1:200 in ddH_2O. A 20 μL ligation reaction was then prepared with the following components: 100 ng of BbsI-digested Addgene plasmid #64324, 2 μL diluted annealed oligos, 1x T4 ligase buffer (final concentration), 10 units T4 DNA ligase (NEB #M0202S), and ddH_2O. The reaction was incubated at 16 ℃ overnight. 5 μL of the 20 μL ligation reaction was then used to transform a 10 μL aliquot of chemically competent 5-alpha (NEB #C2987I) cells according to *bacterial transformation* protocol. Liquid cultures for two clones per DNA construct were prepared according to *liquid culture* protocol. Nuclease-free plasmids were isolated from liquid cultures using ZymoPURE Plasmid Miniprep Kit (#D4209) and associated protocol. Clones were confirmed *sequencing*. Confirmed clones were selected for *large scale plasmid amplification.*

Generation of donor repair templates. Donor repair templates were constructed using sequential cloning in order to generate both monomeric enhanced green fluorescent protein (mEGFP) tagged and untagged versions of the Dsup and Dsup-HMGN2 transgenes. The Dsup sequence used to design the transgene sequences was derived from the tardigrade species *Ramazzottius varieornatus* and codon-optimized for expression in human cells.

Addgene plasmid #54759 contains both the gene that encodes for mEGFP and downstream restriction sites BspEI and BamHI. Addgene plasmid #54759 was first digested with BspEI (NEB #R0540) in NEB buffer r3.1 (NEB #B6003S). Following digestion, the entire reaction was run on a 0.8% agarose gel according to the *gel electrophoresis* protocol and the desired product was extracted using the Monarch DNA Gel Extraction Kit and associated

protocol (including specified modifications). A second *digestion* reaction was prepared using 2 μg of DNA isolated following BspEI digestion of Addgene #54759: 10 units BamHI-HF (NEB #R3136) in 1x final concentration of rCutSmart Buffer (NEB #B6004S). The double-digested plasmid was then isolated using the Monarch PCR&DNA Cleanup Kit (NEB # T1030S) and quantified using the Thermo Scientific One Microvolume UV-Vis Spectrophotometer (Thermo Scientific #8402742000) in preparation for use in ligation reactions.

A DNA fragment (gBlock) containing the Dsup-NBD sequence (see below) was ordered from IDT. This DNA contains restriction enzyme sites BspEI and BamHI (bolded and underlined text in the sequence below in the respective order) at the 5' and 3' ends, respectively. Randomized sequences (lowercase text) of 16 and 18 base pairs were added at the 5' and 3' ends of restriction enzyme sites BspEI and BamHI, respectively, in order to facilitate binding of restriction enzymes. A 3x(GGGS) flexible linker sequence (underlined and non-bolded text) was designed upstream of the Dsup-NBD coding sequence (uppercase text, not underlined or bolded).

BspEI-Dsup-NBD-BamHI gBlock sequence:

atcaagttagagctgtccactcgtt**TCCGGA**GGAGGGTCCGGTGGAGGGAGCGGAGGCGGGTCCAT
GGCCAGCACACACCAGTCTAGCACAGAGCCTAGCAGCACCGGCAAGTCCGAGGAAA
CAAAGAAGGACGCCAGCCAAGGCAGCGGCCAGGACAGCAAAAATGTGACCGTGAC
CAAAGGCACCGGCAGCTCTGCTACATCTGCCGCCATTGTGAAAACCGGCGGCAGCC
AGGGCAAAGACAGCTCTACAACAGCCGGCAGCAGCTCTACCCAGGGCCAGAAGTTT
AGCACCACACCTACAGACCCCAAGACCTTCAGCAGCGACCAGAAAGAGAAGTCTAA
GAGCCCCGCCAAAGAGGTGCCATCTGGCGGCGATTCTAAGTCCCAGGGCGATACCA
AGAGCCAGAGCGACGCCAAAAGCAGCGGACAGTCTCAGGGCCAGTCTAAGGATAG
CGGCAAGAGCAGCTCCGACAGCAGCAAGAGCCACTCTGTGATCGGCGCCGTGAAGG
ATGTTGTGGCTGGCGCCAAAGATGTGGCCGGAAAGGCTGTGGAAGATGCCCCTAGC
ATCATGCACACAGCCGTGGATGCCGTGAAGAATGCCGCCACCACCGTGAAAGACGT
GGCCTCTTCTGCCGCCTCTACAGTGGCCGAGAAAGTGGTGGATGCCTACCACTCTGT
CGTGGGCGACAAGACCGACGACAAGAAAGAGGGCGAGCACAGCGGCGATAAGAAG
GATGATAGCAAGGCCGGCAGCGGATCTGGCCAAGGCGGAGATAACAAGAAGTCTG
AGGGCGAGACATCCGGCCAGGCCGAATCTAGCTCTGGAAATGAAGGCGCCGCTCCT
GCCAAAGGCAGAGGAAGAGGTAGACCTCCTGCCGCCGCTAAAGGCGTTGCAAAAG
GTGCTGCAAAGGGCGCTGCCGCTTCTAAGGGTGCTAAATCTGGCGCCGAGAGCAGC
AAAGGCGGCGAACAATCTTCCGGCGACATCGAGATGGCCGACGCCTCTTCTAAAGG
GGGCAGCGATCAGAGAGATAGCGCCGCCACAGTTGGAGAAGGCGGAGCCTCTGGAT

CTGAAGGCGGCGCAAAAAAAGGCCGCGGAAGAGGCGCCGGAAAGAAAGCTGATGC
CGGCGACACAAAGGACGAGCCCCAGAGAAGAAGCGCCAGACTGTCTGCCAAACCT
GCTCCTCCTAAGCCTGAGCCTAAGCCAAAGAAGGCCCCTGCCAAAAGAGCCGCCAG
CTCCTCTAGCACACCCAGCAATGCCAAGAAGCAGGCTACAGGCGGAGCTGGAAAAG
CCGCTGCCACAAAAGCCACAGCTGCCAAGTCTGCTGCCTCTAAGGCTCCTCAAAATG
GCGCAGGCGCCAAGAAGAAAGGCGGAAAGGCAGGCGGACGGAAGCGGAAATGAtaa
ggatccacgattgccgtcacacgttactagg

The Dsup-NBD gBlock was sequentially digestedi with BspEI in NEB buffer r3 and BspEI in rCutSmart Buffer as before. Following these *digestion* reactions, isolated nuclease-free DNA was quantified using the Thermo Scientific One Microvolume UV-Vis Spectrophotometer in preparation for ligation. *Ligation* reactions were then prepared so that the digested Dsup-NBD DNA would be inserted into a digested Addgene plasmid #54759 between the BspEI and BamHI sites to yield the mEGFP-Dsup-NBD sequence. Ligation was followed by *bacterial transformation* and *liquid culture* protocols before plasmid DNA was prepared using ZymoPURE Plasmid Miniprep Kit. Products were quantified as before.

Next, a plasmid containing wildtype Dsup downstream of the mEGFP gene was generated by modifying the #54759 plasmid containing mEGFP-Dsup-NBD. A gBlock containing a partial Dsup sequence (sequence below) was ordered from IDT. The DNA sequence (uppercase type, not underlined or bolded) replaced the DNA fragment between the SacII and BamHI sites (bolded and underlined type in the sequence below in the respective order) in the #54759 plasmid containing mEGFP-Dsup-NBD to give a #54759 plasmid containing mEGFP-Dsup instead.

SacII-partial-Dsup-BamHI gBlock sequence:
agaggataacatcagatt**CCGCGG**AAGAGGCGCCGGAAAGAAAGCTGATGCTGGCGATACAA
GCGCCGAGCCTCCTAGAAGAAGCAGCAGACTGACAAGCTCTGGCACAGGCGCTGGA
TCTGCTCCTGCTGCTGCAAAAGGCGGAGCTAAGAGAGCCGCCAGCTCTAGCTCCAC
ACCAAGCAATGCCAAGAAGCAGGCTACAGGCGGCGCTGGAAAAGCCGCTGCCACA
AAAGCCACAGCCGCCAAATCTGCCGCTAGCAAGGCTCCTCAAAATGGCGCAGGCGC
CAAGAAGAAAGGCGGAAAGGCAGGCGGACGGAAGCGGAAATGAtaa**ggatcc**cactcatgag
atggaaga

The #54759 plasmid containing mEGFP-Dsup-NBD and the SacII-partial-Dsup-BamHI gBlock DNA were digested with both SacII (NEB #R0157S) and BamHI-HF in rCutSmart Buffer. Following these *digestion* reactions, the digested products were extracted using the Monarch DNA Gel Extraction and associated protocol including the specified modifications. Both the digested insert and digested plasmid were added to a *ligation* reaction so that an Addgene #54759 plasmid would contain the wildtype Dsup sequence downstream of mEGFP. Ligation was followed by *bacterial transformation* and *liquid culture* protocols before plasmid DNA was prepared using ZymoPURE Plasmid Miniprep Kit. This procedure was repeated using the following gBlock DNA (sequence below) ordered from IDT in order to generate a #54759 plasmid containing the Dsup-coreNBD sequence linked to mEGFP. The DNA sequence (uppercase type, not underlined or bolded) replaced the DNA fragment between the SacII and BamHI sites (bolded and underlined type) in the mEGFP-Dsup-NBD plasmid to give a mEGFP-Dsup-coreNBD-containing plasmid.

SacII-partial_Dsup-coreNBD-BamHI sequence:

agaggataacatcagatt**CCGCGG**AAGAGGCGCTGGAAAGAAAGCCGATGCTGGCGATACAA
GCGCCGAGCCTCCTAGAAGAAGCGCCAGACTGTCTGCCTCTGGAACAGGCGCAGGA
TCTGCTCCTGCTGCTGCAAAAGGCGGAGCTAAGAGAGCCGCCAGCTCTAGCTCCAC
ACCAAGCAATGCCAAGAAGCAGGCTACAGGCGGCGCTGGCAAAGCCGCTGCTACAA
AAGCCACAGCCGCCAAATCTGCCGCAAGCAAGGCTCCTCAAAATGGCGCAGGCGCC
AAGAAGAAAGGCGGAAAAGCCGGCGGACGGAAGAGAAAGTGAtaa**ggatcc**cactcatgagat
ggaaga

The mEGFP-Dsup, mEGFP-Dsup-NBD, and mEGFP-Dsup-coreNBD sequences were then subcloned into Addgene plasmid #64215, which is a vector used for targeted integration of DNA fragments at the *AAVS1* locus by homologous recombination. The desired DNA sequences were cloned between the two homology arms (HA) whose sequence homology corresponds to the established integration site. Addgene #64215 was digested in one single reaction with both

PacI (NEB #R0547S) and AsiSI (NEB #R0630S) in rCutSmart Buffer, and then treated with CIP (NEB #M0525S) to dephosphorylate the ends of complementary restriction sites after *digestion* to prevent complementary binding of these sites. 10 units CIP were added to the reaction mixture and incubated for 1 hour at 37 ℃ . The insert that was to be ligated into digested #64215 plasmid was first PCR amplified from the constructs containing mEGFP-Dsup, mEGFP-Dsup-NBD, and mEGFP-Dsup-coreNBD in Addgene #54759 backbone (see *PCR reactions)*. Oligos used to amplify both mEGFP tagged and untagged constructs added restriction enzyme sites PacI and AsiSI to the upstream and downstream ends of the PCR products, respectively (overhang regions were included for enzyme landing site). PCR amplification of mEGFP tagged constructs mEGFP-Dsup, mEGFP-Dsup-NBD, and mEGFP-Dsup-coreNBD included oligos that amplified this entire region whereas untagged constructs Dsup, Dsup-NBD, and Dsup-coreNBD were amplified by oligos that generated a product without the mEGFP sequence. DsupM6 and mEGFP-DsupM6 were amplified from plasmid 54759-Dsup-NBD using a downstream oligo containing AsiSI and annealing upstream of the NBD. The resulting DsupM6 therefore only contains the first 357 amino acids of the Dsup protein sequence.

DNA precipitation was performed by adding the following reagents to each PCR reaction: 0.5 μg/μL glycogen (final concentration), 0.27 M sodium acetate pH 5.5, 2.5x volume final of 100% ethanol. The samples were incubated at -20 ℃ overnight. Following the incubation, tubes were centrifuged at 16000xg for 30 min. at 4 ℃. The supernatant was then removed, and the pellet was allowed to air dry for 5 min. at room temperature. Gel Loading Dye was added to DNA pellets, vortexed, and then briefly spun down. A five-minute incubation took place at room temperature before samples were loaded onto a 0.8% agarose gel according to the *gel electrophoresis* protocol and the expected products were extracted using the Monarch DNA

Gel Extraction Kit and associated protocol, including the indicated modifications. Products were quantified using the Thermo Scientific One Microvolume UV-Vis Spectrophotometer. PCR products were then added to *digestion* reactions with PacI and AsiSI in rCutSmart Buffer and incubated for 5 hours at 37 ℃. Purified DNA was isolated using Monarch PCR&DNA Cleanup Kit and then quantified.

mEGFP-Dsup, Dsup, mEGFP-Dsup-NBD, Dsup-NBD, mEGFP-Dsup-coreNBD, Dsup-coreNBD, mEGFP-DsupM6, and DsupM6 sequences were then ligated into a digested Addgene Plasmid #64215 between the PacI and AsiSI sites. *Ligation* reactions were followed by *bacterial transformation,* preparation of *liquid culture,* and *colony PCR* to identify positive clones. ZymoPURE Plasmid Miniprep Kit was used to isolate nuclease-free DNA from liquid cultures of positive clones, which was then quantified. Clones were confirmed by *sequencing* before *large-scale plasmid amplification* was conducted in preparation for transfection reactions.

Sequencing. Sanger sequencing was performed by Eton Bioscience or Retrogen. crRNA1 (crRNA1.1), crRNA2 (crRNA2.1), mEGFP-Dsup (clone 1), mEGFP-DsupM6 (clone O2). mEGFP-NBD (clone GN7), mEGFP-Dsup-coreNBD (clone 8), Dsup (clone WR5), DsupM6 (clone 25), NBD (clone N1), and Dsup-coreNBD (clone 17) were all confirmed by sequencing alignment to designed DNA sequences in SnapGene.

Large-scale plasmid amplification. QIAGEN HiSpeed Plasmid Midi Kit (#12643) was used to isolate large-scale endotoxin-free plasmid DNA to be used in transfection reactions.

MCF10A cell maintenance. Gloves were changed upon entry of culture room to promote sterility of workspace. Biomedical hood was cleaned with 75% ethanol prior to each use. Plates and solutions were sprayed with 75% ethanol before entering hood.

Cell line. MCF10A cells (CRL-10317) were purchased from ATCC (American Type Culture Collection). The cells were maintained in complete culture medium (See *Complete culture medium preparation*). Cells were cultured in an incubator with 5% CO_2 and humidification at 37 ℃.

Complete culture medium preparation. The following reagents were combined to reach the final concentrations indicated in DMEM F-12 (Gibco): 5% horse serum (Gibco #16050122), 20 ng/mL epithelial growth factor (EGF) (Peprotech AF-100-15), 500 ng/mL hydrocortisone (Sigma #H-0888), 100 ng/mL cholera toxin (Sigma #C8052-1mg), 10 μg/mL human insulin (Sigma #I9278), and 1x penicillin/streptomycin (Gibco #15140-122). Medium was then filtered with the Millipore Sigma Steriflip Sterile Disposable Vacuum Filter (MilliporeSigma #SCGP00525). Complete culture medium was stored at 4 ℃.

Cell passage. Culture medium, PBS, and 0.05% trypsin were warmed in a 37 ℃ water bath for approximately 20 to 30 minutes (dependent on volume). Plates were gently mixed by swirling and culture medium was removed with an aspiration pipet. PBS was added to plate and gently mixed by swirling plate. PBS was aspirated and 0.05% trypsin was added to plate. Cells were then incubated for 15 to 25 minutes or until the majority of cells had detached. Once cells had detached, complete culture medium was added to plate to deactivate trypsin. Plate contents were collected in a sterile conical tube and centrifuged for 5 minutes at 750 rpm. Supernatant was then aspirated, and cell pellet was resuspended in 0.5 to 1 mL of complete culture medium. Cells were counted using EVE automated cell counter (NanoEnTek #EVE-MC) and associated

protocol. The desired number of cells were seeded in a new plate containing complete culture medium. Volumes used through cell passage were adjusted dependent on plate size (Table 5.1).

Table 5.1: Reagent volumes for cell passage dependent on plate size. Volumes indicated are per well.

Solution		Wells per plate		
	10-cm plate	**6**	**24**	**96**
PBS	3 mL	0.5 mL	0.5 mL	100 μL
0.05% Trypsin	1.5 mL	0.5 mL	100 μL	35 μL
Compete culture medium for trypsin deactivation	3.5 mL	1.5 mL	200 μL	100 μL
Complete culture medium volume in plate seeded	8 – 9 mL	2 mL	1 mL	200 μL

Thawing cells. Before commencing to thaw cells, I ensured that the 37 ℃ water bath was at an adequate level. Freeze medium was constituted of 5 – 10% DMSO to perforate cells and prevent explosion (7.5% DMSO for MCF10A cell line (ATCC #CRL-10317)). A conical tube containing 9 mL of complete culture medium was prepared prior to starting thaw protocol. Once prepared, cells were retrieved from liquid nitrogen storage and placed immediately on a 37 ℃ water bath. Tubes were moved continuously in 37 ℃ water bath for exactly two minutes. Tubes were dried and sprayed with 75% ethanol before transferring to hood. Cells were then transferred to prepared conical tube containing complete culture medium to dilute DMSO and the tube was inverted 3 – 4 times. Centrifugation took place for 5 minutes at 750 rpm. Supernatant was aspirated using caution to avoid cell pellets. 9 mL of medium was then added to a 10-cm plate. 1 mL from plate used to resuspend pellets and then entire resuspension added to the 10-cm plate. Plates were lightly shaken to improve uniformity of distribution. Plates were agitated again after 1 hour to ensure cells had not congregated in the center of the plate.

Freezing cells. Cells pellets were recovered from a 10 cm plate following cell detachment and trypsin aspiration detailed in *cell passage* protocol. Cells were resuspended in 3 mL freezing medium (7.5% DMSO and 92.5% horse serum) and mixed by pipetting 5 times. 1 mL of resuspension was added to 2 mL cryotubes. Cells were incubated at – 20 ℃ for at least 1 hour (less than 12 hours), then at – 80 ℃ for 72 hours before being transferred to a liquid nitrogen freezer for long term storage (Laboratory reference: Box B8).

Cellular DNA extraction. 100,000 – 200,000 cell pellets were collected from cell suspension. Pellets centrifuged at 400xg for 5 minutes at 4 ℃ and supernatant was removed. 100 μL of PBS was then added, and centrifugation was repeated. Following the removal of the supernatant, I snap froze the cell pellets in liquid nitrogen and then stored them at – 80 ℃. Genomic DNA was extracted using the QuickExtract DNA Extraction Solution (Lucigen #QE0905T) and associated protocol. Extracted DNA was quantified and diluted to approximately 50 ng/μL.

Establishment of cell lines containing transgenes

Puromycin kill curve. The quick reference protocol from Mirus (MIR 5940) was used to perform puromycin dihydrochloride solution kill curve. Wildtype MCF10A cells in a 6-well plate were grown in complete culture medium until about 80% confluency was reached. MCF10A cells obtained from a laboratory stock were treated with a concentrations range: 0, 0.5, 1, 2, 4, 6, 8, 10, 15, 20 μg/mL. Complete culture medium containing puromycin was replaced every 48 to 72 hours. The puromycin kill curve was repeated in MCF10A cells obtained from ATCC. Cells were treated with an initial concentration range of 0, 0.25, 0.5, 1, 2, 4 μg/mL and observed after one week. Prolific cell death in wells treated with 1, 2, 4 μg/mL prompted the

establishment of a narrowed kill curve including concentrations of 0, 0.1, 0.2, 0.3, 0.4, 0.5 μg/mL. Complete culture medium containing puromycin was replaced every 48 to 72 hours. I observed no difference in cell proliferation, confluency, or viability between cell treated with 0.2, 0.3, 0.4 μg/mL and untreated cells. Cells treated with 0.5 μg/mL puromycin reached confluency slightly slower and showed increased amounts of cell death. 0.5 μg/mL puromycin was therefore selected as an effective concentration for selection with puromycin.

Transfection. Wildtype MCF10A cells were seeded about 16 hours before transfection at a density of $0.7x10^6$ cells per well in 6 well plates. Cells transfections were performed using polyethyleneimine (PEI), pH 7.4, reagent. Cells were first rinsed with 0.5 mL of complete culture medium and then incubated with 2 mL fresh complete culture medium 1 hour prior to transfection. PEI was used at a PEI:DNA ratio of 3:1. 1.5 μg of Cas9 plasmid was added to 1.5 μg of donor template plasmid per reaction. The PEI reagent was added to DNA by gentle pipetting up and down three times and the reaction was then incubated for 10 minutes at room temperature. Optimem media was added to the DNA-PEI complex and mixed by flicking of the tube. The mixture was then incubated for 20 minutes at room temperature. The transfection mixture was then added dropwise to culture medium. Plates were shaken gently. Cells were cultured in incubator with 5% CO_2 and humidification at 37 ℃.

FACS and clonal amplification. Cell sorting was performed at the UCSD Human Embryonic Stem Cells core facility. Cells were resuspended in complete culture medium containing 200 ng/mL Dapi (Sigma #D9542) to exclude dead cells from being collected. Cells were sorted for mCherry and mEGFP positive signal 24 hours after transfection using ARIA FUSION cell sorting machine. Cells were pooled into complete culture medium containing 10% horse serum (rather than 5%) to aid in cellular recovery. 10% complete culture medium was

changed 3 to 5 days following sort. Clonal populations were amplified from 96-well plate to a 10-cm plate at about 80-90% confluency prior to each seed. About two weeks following the second sort, complete culture medium containing 10% horse serum was replaced with 5% using a 7.5% complete culture medium to transition over 3 to 4 days, all of which contained 0.5 μg/mL puromycin.

PCR reactions. *PCR reactions were performed using Taq DNA polymerase (NEB #M0273) in Thermopol buffer and associated protocol. **PCR reactions were performed using Phusion Polymerase (NEB #M0530) and HF Buffer (NEB #B0518S)

PCR amplification from edited 54759 for generation of 64215 plasmids containing transgenes.** Oligos 147F-2021 (tagact**ttaattaa**GCCACCATGGTGAGCAAG) and 148R (aagtc**gcgatcgc**TTATCATTTCCGCTTCCGT) were used to amplify mEGFP-Dsup from 54759 plasmid containing the Dsup sequence. The expected product size of this reaction was 2,119 bp and the anneal temperature was 61 ℃. The restriction sites PacI (bolded and underlined text) and AsiSI (bolded text) were attached to either end of the amplified product. Extra base pairs were included upstream and downstream of the restriction sites to aid in restriction enzyme binding. Oligos 149F-2021 (tagact**ttaattaa**gccaccATGGCCAGCACACACCAGT) and 157R-2021 (taagtc**gcgatcgc**GCTTATCATTTCCGCTTCCGTCCG) were used to amplify Dsup from the same Dsup-containing 54759 plasmid. The expected product size was 1,363 bp and the anneal temperature was 66 ℃. Oligos 149F-2021 and 150R-2021 (taagtc**gcgatcgc**TTATCATTTCCGCTTCCGTCCG) were used to amplify Dsup-NBD from the 54759 plasmid that contained Dsup-NBD. The expected product size was 1,363 bp and the anneal temperature used was 71 ℃. Oligos 147F-2021 and 150R were used to amplify mEGFP-

Dsup-NBD from the same plasmid. The expected product size was 2,119 bp and the anneal temperature used was 61 ℃. Oligos 147F-2021 and 158R-2021 (taagtc**gcgatcgc**TTATCACTTTCTCTTCCGTCC) were used to amplify mEGFP-Dsup-coreNBD from the Addgene plasmid 54759 modified to contain Dsup-coreNBD. The expected product size was 2,119 bp and the anneal temperature used was 61 ℃. Oligos 149F-2021 and 159R-2021 (taagtc**gcgatcgc**TTATCACTTTCTCTTCCGTC) were used to amplify Dsup-coreNBD. The expected product size was 1,363 bp and the anneal temperature was 66 ℃. Oligos 147F-2021 and 153R-2021 (taagtc**gcgatcgc**TTATCATGTGTCGCCGGCATCAGC) were used to amplify mEGFP-DsupM6. The expected product size was 1,855 bp with an annealing temperature of 64 ℃. Oligos 149F and 153R were used to amplify DsupM6. The expected product size was 1,099 bp and the annealing temperature was 68 ℃.

Table 5.2: Thermocycling conditions for amplification from edited 54759 for generation of 64215 plasmids containing transgenes

Step	Temperature	Time
Initial Denaturation	98 ℃	0:30
35 cycles	98 ℃	0:10
	–	0:20
	72 ℃	1:00
Final Extension	72 ℃	7:00
Hold	4 ℃	∞

Colony PCR*. Positive clones were screened using colony PCR. Oligos 142-2021 (GCAGCGATCAGAGAGATAGC) and 160-2021 (TCATTTACGCGTGCGATCGC) were used to perform colony PCR. The expected product size of 64215 plasmids containing Dsup, Dsup-NBD, and Dsup-coreNBD with and without mEGFP was 409 bp while the expected product size of 64215 plasmids containing DsupM6 and mEGFP-DsupM6 was 145 bp.

Table 5.3: Colony PCR thermocycling conditions

Step	Temperature	Time
Initial Denaturation	95 ℃	5:00
35 cycles	95 ℃	0:30
	61 ℃	0:30
	68 ℃	1:00 (1min/kb)
Final Extension	68 ℃	5:00
Hold	4 ℃	∞

The expected product size of 64215 plasmids containing Dsup, Dsup-NBD, and Dsup-coreNBD with and without mEGFP was 409 bp while the expected product size of 64215 plasmids containing DsupM6 and mEGFP-DsupM6 was 145 bp.

PCR amplification of the puromycin gene*. Oligos PuroR_FW_gec (ACCGAGTACAAGCCCACGG) and PuroR_RV_gec: (GCCTTCCATCTGTTGCTGCG) were used to amplify the puromycin gene. The expected product size was 330 bp.

Table 5.4: Thermocycling conditions for amplification of the puromycin gene

Step	Temperature	Time
Initial Denaturation	95 ℃	5:00
35 cycles	95 ℃	0:30
	56 ℃	0:30
	68 ℃	X (1min/kb)
Final Extension	68 ℃	5:00
Hold	4 ℃	∞

PCR amplification of HA-R*. Oligos Dsup_fw_HA-R* (GGAAAAGCCGCTGCCACAAA**)** and AAVS1_5'_HA-L **(**GGAGGAGAATCCACCCAAAAGGC) were used to amplify the right homology arm. The expected product size was 1,466 bp.

Table 5.5: Thermocycling conditions for amplification of HA-R

Step	Temperature	Time
Initial Denaturation	95 ℃	5:00
35 cycles	95 ℃	0:30
	60 ℃	0:30
	68 ℃	1:30 (1min/kb)
Final Extension	68 ℃	5:00
Hold	4 ℃	∞

PCR amplification of HA-L*. Oligos AAVS1-Puro_fw_HA-L_3 (CATTGTCACTTTGCGCTGC) and AAVS1-Puro_rv_HA-L_3 (GACGCGCGTGAGGAAGA) were used to amplify the left homology arm. The expected products size was 1,258 bp.

Table 5.6: Thermocycling conditions for amplification of HA-L

Step	Temperature	Time
Initial Denaturation	95 ℃	5:00
35 cycles	95 ℃	0:30
	56 ℃	0:30
	68 ℃	1:15 (1min/kb)
Final Extension	68 ℃	5:00
Hold	4 ℃	∞

PCR amplification of the AAVS1 region*. Oligos AAVS13'_3_rv (GGAACTCTGCCCTCTAACGC) and AAVS15'_3_fw (GCTTCTTGGCCACGTAACCT) were used to amplify the AAVS1 region. The expected product size of the AAVS1 region without the inserted region was 2,026 bp.

Amplification of *AAVS1* locus containing the transgenes attempted using LongAmp Hot Start Taq DNA Polymerase (NEB M0534S) with 5x LongAmp reaction buffer but was unsuccessful. The expected product size of the AAVS1 locus containing the mEGFP-Dsup, mEGFP-Dsup-NBD, or mEGFP-Dsup-coreNBD would have been 7,213 bp.

Table 5.7: Thermocycling conditions for amplification of the AAVS1 region

Step	Temperature	Time
Initial Denaturation	95 ℃	5:00
35 cycles	95 ℃	0:30
	57 ℃	0:30
	68 ℃	2:00 (1min/kb)
Final Extension	68 ℃	5:00
Hold	4 ℃	∞

Amplification of *AAVS1* locus containing the transgenes attempted using LongAmp Hot Start Taq DNA Polymerase (NEB M0534S) with 5x LongAmp reaction buffer but was unsuccessful. The expected product size of the AAVS1 locus containing the mEGFP-Dsup, mEGFP-Dsup-NBD, or mEGFP-Dsup-coreNBD would have been 7,213 bp.

Alkaline comet assay. The comet assay protocol was executed by combining two protocols: Lu et al. in 2017 and Clementi et al. in 2021. Light exposure was minimized throughout cell handling to prevent background DNA damage.

Preparation of agarose slides. 1% normal melting point agarose slides were prepared the afternoon prior to the assay by dipping "PTFE" printed slides (Electron Microscopy Sciences #63422-11) into the normal melting point agarose (Invitrogen #16500500) so that a thin layer covered each well. Slides were allowed to air-dry on bench overnight and then placed in 37 ℃ incubator on the day of the assay.

Preparation of cells for use in comet assay. Cells were collected for the comet assay at high confluency (~ 90%) to increase the likelihood of cell size and cycle similarity. Both wildtype MCF10A cells and MCF10A cells containing Dsup were seeded at differing confluences at least 24 hours in advance to ensure that comparable confluences would be available on the day of the assay. Wildtype MCF10A cells and MCF10A cells expressing Dsup

were cultured in complete culture medium and complete culture medium containing 0.5 μg/mL puromycin, respectively. Cells were collected from 6-well plates according to *'cell passage protocol'* and resuspended in 1 mL of complete culture medium. Cells were then counted using EVE automated cell counter and associated protocol and diluted to reach concentrations of 300,000 cells per 1 mL PBS per cell line. Cell dilutions were kept on ice.

Hydrogen peroxide treatment and comet assay. Cells in suspension were treated with 25 μM and 50 μM hydrogen peroxide for 3 minutes on ice. Treated cells were combined with 1% molten low melting point agarose (Invitrogen #6520100) to achieve a 1:10 ratio of cells:agarose. This mixture was then directly spread onto prepared agarose slides using wide orifice pipette tips. Slides were incubated in the dark at 4 ℃ for at least 30 min. and then immersed in chilled lysis buffer (2.5 M NaCl, 100 mM disodium EDTA, 10 mM Tris base, 200 mM NaOH in ddH2O, pH 10, 1% Triton X-100, 1% sodium lauroyl sarcosinate) for exactly 1 hour in the dark at 4 ℃. Next, slides were incubated for exactly 1 hour in the dark at 4 ℃ in cold alkaline comet electrophoresis buffer (200 mM NaOH and 1 mM disodium EDTA in ddH_2O, pH >13). Electrophoresis for the assay was performed at 22 V (0.8 V/cm) at 4 ℃ (in cold room) for 25 minutes. Following electrophoresis, slides were submerged in ddH_2O for 5 minutes two times, before being placed in 70% EtOH for 5 minutes at room temperature in the dark. Slides were dried in the dark at 37 ℃ for 15 minutes and then incubated for 15 minutes at room temperature in green fluorescent nucleic acid staining solution (green fluorescent nucleic acid stain in Tris-EDTA buffer (10 mM Tris-HCl, 1 mM disodium EDTA, pH 7.4)). Slides were rinsed and dried before images were captured using Zoe fluorescent cell imager. Images were analyzed using Open Comet Plugin in Fiji according to Clementi et al., 2021.

Western blot analysis

SDS-PAGE. 5 – 10 % SDS-Polyacrylamide 1 mm gel was created by first mixing the resolving gel (10% acrylamide final concentration) components which consisted of ddH_2O, 4x lower buffer (Kadonaga laboratory stock), 30:0.8% acrylamide:bis, 10% ammonium persulfate (APS), and tetramethylethylenediamine (TEMED). All components were vortexed and immediately added to a prepared gel mold. The solution was allowed to sit for at least 20 minutes before the addition of the stacking gel (5% acrylamide final concentration) mix which consisted of 4x upper buffer (Kadonaga laboratory stock), 30:0.8% acrylamide:bis, 10% APS, and TEMED. Immediately following addition of the upper buffer, a well comb was added. The mixture was left to harden for at least 20 minutes.

Protein extraction. Protein extraction was performed once cell pellets (100,000 - 200,000 cells) had thawed on ice for about 10 minutes. Pellet was resuspended gently in PBS containing 0.1% NP40. Centrifugation was performed for 30 seconds at 3500xg at 4 ℃. Supernatant containing the cytoplasmic fraction was then discarded. Next, the pellet was gently resuspended in PBS containing 0.05% NP40. Centrifugation was repeated as before, and the supernatant discarded. The pellet was finally resuspended in NLB buffer (50 mM $NaPO_4$ pH 8.0, 300mM NaCl, 5mM $MgCl_2$, 10% glycerol, 0.01% NP-40, 10mM Imidazole) in order to break the nuclei.

Electrophoresis. Room temperature 5x sample buffer (SB) (315 mM Tris pH 6.8, 25% BME, 15% SDS, 30% glycerol, and 0.05% BPB) was added to protein extracts to give a 1x final concentration and briefly mixed by pipetting before brief vortexing. The SB-extract mixture was then incubated at 95 to 100 ℃ for 5 min. The mixture was cooled on ice and directly loaded into the prepared gel using extra-long tips to ensure the sample reached the bottom of the well. A

protein ladder (BioRad #161037) was included alongside samples. The gel was run at 90 V until proteins left the stacking gel, then run at 150 to 200 V until the loading dye reached the bottom of the resolving gel.

Protein gel transfer. Transfer of gel content to a nitrocellulose membrane (0.20 μm; Bio-Rad cat. 1620112) was initiated by rehydration of the membrane in water. Membranes were then equilibrated in cold Towbin transfer buffer (Kadonaga Laboratory stock) for at least 5 minutes. The transfer cassette was assembled in Towbin buffer. The cassette was inserted into the transfer chamber and surrounded by ice. A current of 0.25 A (per chamber) was set for the gel transfer and run for 90 minutes at 4 °C. After transfer, the chamber was disassembled and the membrane was stained with Ponceau S (Sigma #P7170) for 5 min and the gel was stained with Coomassie Blue for 10 minutes. After imaging, the membrane was rinsed with water three times, allowed to air dry, and stored room temperature.

Western blot. Membranes were cut into desired sizes. Membranes were rehydrated in water before being rinsed with TBST buffer (10 mM tris-HCl pH 7.9, 150 mM NaCl, 0.05% Tween 20 (US Biochemicals #20605)). Membranes were then blocked with 5% milk in TBST buffer for 2 hours at room temperature with gentle shaking. The primary antibody was thawed on ice and added to membrane with specified dilution in 5% milk in TBST buffer. The membrane was incubated at 4 °C with gentle shaking overnight. The following day, membranes were incubated for 1 hour on gentle rocker at room temperature, before being rinsed with 1% milk in TBST for 10 minutes with vigorous shaking. Rinse step was repeated an additional two times, changing TBST buffer each time. The membrane solution was then replaced with 5% milk in TBST buffer solution containing secondary anti-rabbit antibody (Invitrogen #65-6120) at a dilution of 1:5,000 and membranes were incubated for 1 hour at room temperature with gentle

shaking. Following secondary antibody incubation, membranes were rinsed with TBST buffer three times as before: 10 min each with vigorous shaking. Membranes were stored in the dark at 4 ℃ until visualization with chemiluminescence, SuperSignal West Pico Plus Kit (ThermoScientific #34577), which was performed as soon as possible.

REFERENCES

Balasubramanian, B., Pogozelski, W. K., & Tullius, T. D. (1998). DNA strand breaking by the hydroxyl radical is governed by the accessible surface areas of the hydrogen atoms of the DNA backbone. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(17), 9738–9743. https://doi.org/10.1073/pnas.95.17.9738

Chavez, C., Cruz-Becerra, G., Fei, J., Kassavetis, G. A., & Kadonaga, J. T. (2019). The tardigrade damage suppressor protein binds to nucleosomes and protects DNA from hydroxyl radicals. *eLife*, *8*, e47682. https://doi.org/10.7554/eLife.47682

Clementi, E., Garajova, Z., & Markkanen, E. (2021). Measuring DNA Damage Using the Alkaline Comet Assay in Cultured Cells. *Bio-protocol*, *11*(16), e4119. https://doi.org/10.21769/BioProtoc.4119

Cooke, M. S., Evans, M. D., Dizdaroglu, M., & Lunec, J. (2003). Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, *17*(10), 1195–1214. https://doi.org/10.1096/fj.02-0752rev

Debnath, J., Muthuswamy, S. K., & Brugge, J. S. (2003). Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods (San Diego, Calif.)*, *30*(3), 256–268. https://doi.org/10.1016/s1046-2023(03)00032-x

González-Romero, R., Eirín-López, J. M., & Ausió, J. (2015). Evolution of high mobility group nucleosome-binding proteins and its implications for vertebrate chromatin specialization. *Molecular biology and evolution*, *32*(1), 121–131. https://doi.org/10.1093/molbev/msu280

Hashimoto, T., Horikawa, D. D., Saito, Y., Kuwahara, H., Kozuka-Hata, H., Shin-I, T., Minakuchi, Y., Ohishi, K., Motoyama, A., Aizu, T., Enomoto, A., Kondo, K., Tanaka, S., Hara, Y., Koshikawa, S., Sagara, H., Miura, T., Yokobori, S. I., Miyagawa, K., Suzuki, Y., … Kunieda, T. (2016). Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. *Nature communications*, *7*, 12808. https://doi.org/10.1038/ncomms12808

Hashimoto, T., & Kunieda, T. (2017). DNA Protection Protein, a Novel Mechanism of Radiation Tolerance: Lessons from Tardigrades. *Life (Basel, Switzerland)*, *7*(2), 26. https://doi.org/10.3390/life7020026

Hayashi, H., Kubo, Y., Izumida, M., & Matsuyama, T. (2020). Efficient viral delivery of Cas9 into human safe harbor. *Scientific reports*, *10*(1), 21474. https://doi.org/10.1038/s41598-020-78450-8

Jackson, S. P., & Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature*, *461*(7267), 1071–1078. https://doi.org/10.1038/nature08467

Kaufman, W. L., Kocman, I., Agrawal, V., Rahn, H. P., Besser, D., & Gossen, M. (2008). Homogeneity and persistence of transgene expression by omitting antibiotic selection in cell line isolation. *Nucleic acids research*, *36*(17), e111. https://doi.org/10.1093/nar/gkn508

Kugler, J. E., Deng, T., & Bustin, M. (2012). The HMGN family of chromatin-binding proteins: dynamic modulators of epigenetic processes. *Biochimica et biophysica acta*, *1819*(7), 652–656. https://doi.org/10.1016/j.bbagrm.2012.01.013

Lu, Y., Liu, Y., & Yang, C. (2017). Evaluating In Vitro DNA Damage Using Comet Assay. *Journal of visualized experiments: JoVE*, (128), 56450. https://doi.org/10.3791/56450

Mínguez-Toral, M., Cuevas-Zuviría, B., Garrido-Arandia, M., & Pacios, L. F. (2020). A computational structural study on the DNA-protecting role of the tardigrade-unique Dsup protein. *Scientific reports*, *10*(1), 13424. https://doi.org/10.1038/s41598-020-70431-1

Nanduri, R., Furusawa, T., & Bustin, M. (2020). Biological Functions of HMGN Chromosomal Proteins. *International journal of molecular sciences*, *21*(2), 449. https://doi.org/10.3390/ijms21020449

Oceguera-Yanez, F., Kim, S. I., Matsumoto, T., Tan, G. W., Xiang, L., Hatani, T., Kondo, T., Ikeya, M., Yoshida, Y., Inoue, H., & Woltjen, K. (2016). Engineering the AAVS1 locus for consistent and scalable transgene expression in human iPSCs and their differentiated derivatives. *Methods (San Diego, Calif.)*, *101*, 43–55. https://doi.org/10.1016/j.ymeth.2015.12.012

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, *8*(11), 2281–2308. https://doi.org/10.1038/nprot.2013.143

Ricci, C., Riolo, G., Marzocchi, C., Brunetti, J., Pini, A., & Cantara, S. (2021). The Tardigrade Damage Suppressor Protein Modulates Transcription Factor and DNA Repair Genes in Human Cells Treated with Hydroxyl Radicals and UV-C. *Biology*, *10*(10), 970. https://doi.org/10.3390/biology10100970

Veling, M. T., Nguyen, D. T., Thadani, N. N., Oster, M. E., Rollins, N. J., Brock, K. P., Bethel, N. P., Lim, S., Baker, D., Way, J. C., Marks, D. S., Chang, R. L., & Silver, P. A. (2022). Natural and Designed Proteins Inspired by Extremotolerant Organisms Can Form Condensates and Attenuate Apoptosis in Human Cells. *ACS synthetic biology*, *11*(3), 1292–1302. https://doi.org/10.1021/acssynbio.1c00572