DEVELOPING A CLINCALLY RELEVANT ISOGENIC CELL LINE OF HUNTINGTON'S DISEASE

DEVELOPING A CLINICALLY RELEVANT ISOGENIC CELL LINE OF HUNTINGTON'S DISEASE

By NATASHA SAVIC, B.Sc.(Honours)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Master of Science

McMaster University MASTER OF SCIENCE (2022)

Hamilton, Ontario (Biochemistry and Biomedical Sciences)

TITLE: Developing a clinically relevant isogenic cell line of Huntington's Disease

AUTHOR: Natasha Savic, B.Sc. (McMaster University)

SUPERVISOR: Dr. Ray Truant

NUMBER OF PAGES: xiii, 82

LAY ABSTRACT

Huntington's Disease is a neurodegenerative disease caused by excessive CAG repeats in the HTT gene. Without a way of delaying or stopping symptoms, we need model systems to understand why some individuals develop disease later than others. We know that people develop symptoms earlier in life when they have more CAG repeats, but we don't fully understand the contribution of their other genes. We explore this by attempting to edit the DNA of a clinically relevant cell line to change the CAG repeats. Once we establish this system, we can compare the cells before and after editing to see exactly what changes are caused by the CAG repeat length. While we were not able to develop the cell line, this work should be used by future researchers as a troubleshooting guide as they develop a representative model system.

ABSTRACT

Huntington's Disease (HD) is a neurodegenerative disease caused by a CAG expansion in exon 1 of the HTT gene. The age of symptom onset is inversely correlated with the length of the CAG tract. CAA interruptions in the CAG tract have a protective effect, delaying the onset of symptoms despite still coding for a glutamine residue. Evidence suggests that genes outside of the CAG expansion also contribute to differences in age of symptom onset. To improve our knowledge of disease, a clinically relevant model system is necessary. Here, we identify that many model systems use overexpressed protein that fails to adhere to the stoichiometry of an endogenous system. Many systems also use CAG expansions beyond what would be seen in a clinical case of HD or do not appropriately account for the influence of other genes. TruHD fibroblast cells immortalized with human telomerase reverse transcriptase are derived from patient samples and have a stable genome. Here, we discuss attempts to develop isogenic cell lines by editing TruHD cells to have different CAG repeats. While we were unsuccessful in isolating clones carrying the desired edit, future researchers can use this work as a guide to navigate the successful creation of isogenic cells. We can compare phenotypes between pre and post edited cells to see how they change with a new CAG length in an otherwise stable genetic background. We also discuss the successful editing of TruHD cells with a CAG to CAA base edit to introduce a CAA interruption in diseased cells. We discuss how

to propagate these cells for future research to compare pre and post edited cells to better understand the protective action of CAA interruptions.

ACKNOWLEDGEMENTS

I want to thank the many, many people who were involved in this project. This work would not be possible without all your support.

First, I would like to extend my sincerest gratitude towards my supervisor, Dr. Ray Truant. Your support during my years in the Truant lab has been invaluable. I also want to thank my committee members, Dr. Lesley MacNeil and Dr. Matthew Miller for their continued guidance throughout this project.

I want to thank all the members of the Truant Lab: Dr. Tamara Maiuri, Dr. Claudia Hung, Celeste Suart, Nola Begeja, Carlos Barba-Bazan, Kaitlyn Neuman, and Yi Peng. I really appreciate all the time you spent teaching me, troubleshooting this project, and just being there to lean on for emotional support. You have all been amazing.

Finally, I want to thank all the friends and family that stuck by me throughout this project. I couldn't have done it without your unwavering support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTS OF ALL ABBREVIATIONS AND SYMBOLS

18Q donor plasmid	pJOP-HTT-HR18Q
45Q donor plasmid	pJOP-HTT-HR45Q
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
Cas9	CRISPR associated protein 9
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
Cryo-EM	Cryogenic electron microscopy
DSB	Double stranded break
ECL	Enhanced Chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
FACS	Fluorescence activated cell sorting
FBS	Fetal bovine serum
GFP	Green fluorescent protein
gRNA	Guide RNA
GWAS	genome wide association study
h	Hour(s)
HAP40	Huntingtin – associated protein
HD	Huntington's disease
HDR	Homology directed repair
HEAT	Huntingtin, Elongation Factor 3, protein phosphatase 2A, target of rapamycin 1
HEK	Human embryonic kidney
hiPSC	Human induced pluripotent stem cells
hTERT	Human telomerase reverse transcriptase
Htt	Huntingtin
ICE	Inference of CRISPR edits
IDR	Intrinsically disordered region
Indels	Insertions or deletions
KI	Knock-in
mCer	mCerulean
N17	Residues 1-17 of the N terminal region
nCas9-EGFP	NLS-nCas9-NLS-P2A-EGFP
NHEJ	Nonhomologous end joining
PAGE	Polyacrylamide Gel Electrophoresis
PAM	Protospacer adjacent motif
PID	PAM-interacting domain
PVDF	Polyvinylidene Difluoride
PY	Proline-tyrosine
R/2	R6/2
RIPA	Radio Immuno Precipitation Assay

RNA	Ribonucleic acid
RNP	Ribonucleoprotein complexes
ROS	Reactive oxygen species
RPE1	Retinal pigment epithelial
SDS	Sodium dodecyl-sulfate
sgRNA	Single guide RNA
TALEN	Transcription activator-like effector nucleases
TALEN 1	DSPQ-HTT-TALEN1 ZEO
TALEN 2	DSPQ-HTT-TALEN 2 BSD
tracrRNA	Trans-activating crRNA
WT	Wild type
YFP	Yellow Fluorescent Protein
ZFN	Zinc-Finger Nucleases

DECLARATION OF ACADEMIC ACHIEVEMENT

Hong Liang conducted Fluorescence Activated Cell Sorting of all sorted samples through the McMaster Core Flow Facility.

Dr. Pranam Chatterjee provided constructs of SpCas9-BE4Max, SpRYc-BE4Max and targeted gRNA for base editing experiment in Chapter 3.7 and Appendix. Natasha is an author on the work “PAM-Free Genome Editing with an Optimized Chimeric Cas9” originally submitted June 2021 to Nature.

All remaining work was completed by Natasha Savic.

CHAPTER 1.0 INTRODUCTION

1.1 Huntington's Disease

Huntington's Disease (HD) is a monogenic, dominant autosomal neurodegenerative disease that mainly targets the medium spiny neurons of the neostriatum for degeneration (1). HD develops in mid-life stages, with symptom onset establishing in individuals between 30-50 years of age (2). The prevalence of HD is 3-10 in 100 000 individuals (3). Juvenile cases of HD are very rare with one estimate of incidence at 0.7 per million in the United Kingdom (4). Juvenile cases present differently than the traditional development of HD with more severe degeneration observed (5). There are recognized sex-differences in HD, with women presenting with more severe disease than men (6). HD's progressive brain cell loss manifests as progressive dementia, chorea, and behavioral changes (7). Currently, there is no known method of slowing progression of the disease. Rather than curative, current treatments focus on attempting to improve symptoms through physiotherapy and psychotherapy. To develop a reliable therapeutic, considerable research has focused on learning more information about the pathogenic mechanism.

The monogenic nature of HD makes it a valuable research target. HD arises when an expansion of CAG is present in the 5' region of exon 1 of the *HTT* gene coding for the huntingtin protein (8). The *HTT* gene is comprised of 67 exons, and the corresponding protein is 3144 amino acids in length. A disease

phenotype arises when individuals possess above 40 CAG repeats (3). While records note a range of 40-121 repeats in patient cases, statistical outliers represent CAG expansions greater than 50 repeats, with 43 CAG repeats representing the clinical mean (3,9). Healthy individuals possess a limited 6-35 CAG repeat tract, with an intermediate phenotype presence associated with tract lengths of 36-39 repeats (3). HD and several other similar polyglutamine expansion diseases exhibit a correlation whereby CAG tract length is inversely correlated to age of symptom onset (10–12). In other words, a longer CAG repeat length tends to correspond to earlier symptom development. Studies show that the uninterrupted CAG length is the measure of disease onset rather than the length of polyglutamine repeats (13). CAA interruptions have been shown to contribute to variance in symptom onset age, with loss of interruptions hastening onset and gain of interruptions delaying symptom development (13–16). The loss and gain of interruptions seems to impact CAG lengths of 36-39 most prominently (14–16). CAG expansions are subject to somatic expansion, where an individuals' CAG lengths have the potential to expand over their lifetime and produce disease. Similarly, HD often arises or even progresses over generations during the process of genetic anticipation. This often leads to offspring developing more severe symptoms earlier than their ancestors (3,17). Though this is a useful tool for general predictive trends, this is not the only contributor to disease development. There is tremendous variation in age of symptom onset, with only 50-60% of symptomatic development attributable to CAG expansion

length (3,10). Findings have indicated that heritable components contribute 40-60% of the non-CAG component of disease onset (3). In 2015 a genome wide association study (GWAS) was used to unbiasedly study how genetic variations outside of the CAG expansion may impact symptom onset. By taking an unbiased approach and analyzing the entire genome, the GWAS identified DNA repair genes as potential modifiers of motor symptom onset (3,11).

1.2 Protein Structure and the Stress Response

Exploration into huntingtin's structure was a primary achievement used to uncover protein interactions that have enabled very promising potential therapeutics. Huntingtin has been shown to have critical involvement in autophagy, endocytosis, regulation of transcription, transport of vesicles, the stress response and recently DNA repair (18–21). Huntingtin is composed of a series of HEAT (Huntingtin, Elongation Factor 3, protein phosphatase 2A, target of rapamycin 1) repeats, producing a solenoid structure to facilitate protein interactions (22,23). This structural configuration supports a scaffolding role for huntingtin to mediate its functions. The proline-rich domain of huntingtin has also been shown to support its interaction with other proteins (24). Huntingtin has a proline-tyrosine nuclear localization signal and nuclear export signal that allow its entry to and from the nucleus (25,26). Residues 1-17 of the N terminal region (N17) of exon 1 mediate binding of huntingtin to peripheral surfaces and are evolutionarily conserved (27,28). Serine 13 and 16 of the N17 domain of

huntingtin are hypo-phosphorylated in HD, with restoration of phosphorylation shown to be protective (22,28,29). Also within the N17 domain, methionine 8 has been identified as a sensor for reactive oxygen species (ROS), implicating huntingtin's role in the stress response (18). Methionine 8 oxidation prompts huntingtin's release from the endoplasmic reticulum, promoting solubility and its availability to kinases for phosphorylation of N17 (18). Phosphorylation promotes nuclear retention at chromatin regions until oxidative stress is resolved. Upon resolution, huntingtin is released and returns to the endoplasmic reticulum (18). ROS is a by-product of mitochondrial respiration, but due to the high levels of oxygen and metabolic demands, it becomes increasingly problematic as it rapidly accumulates over time. Due to inefficiencies in metabolism with aging, ROS damage could explain the later onset typically seen in HD (19,30).

1.3 Evaluating Model Systems

To fully understand the progression of HD, it is necessary to have model systems that accurately reflect the disease as it presents in patients. Models of HD often involve parameters that fail to mimic endogenous conditions. Many models involve the expression of only exon 1 of huntingtin. This fails to contextualize the critical function of the rest of the huntingtin protein. Recent structural reconstructions of huntingtin show that exon 1 protein when investigated in the context of full length huntingtin, behaves differently than when in isolation (31). Current structural knowledge of huntingtin is largely focused on

its interaction with HTT – associated protein (HAP40). Isolated forms of huntingtin are more difficult to investigate due to higher aggregation levels, reduced stability, and less homogeneity than when in complex with HAP40 (31). It is not well understood if or how the behavior of huntingtin changes when in complex with HAP40. Furthermore, it is not clear if huntingtin functions endogenously in isolation or in complex with HAP40. Cryogenic electron microscopy (Cryo-EM) has not been able to uncover the full structure of huntingtin, in isolation or in complex. The structure of exon 1 and the intrinsically disordered region (IDR) (residues 400-660) of huntingtin are still unresolved by cryo-EM. Harding *et al.* were able to use complementary modeling software to reconstruct the structure of huntingtin when in complex with HAP40. Harding *et al.*, discovered that when exon 1 structure is investigated in the context of full-length huntingtin, expanded CAG lengths increase flexibility and movement of exon 1. This finding directly conflicts with past studies of isolated exon 1 that found expanded CAG lengths conferred a more compact exon 1 structure (32,33). Additionally, Harding *et al.* established that expanded exon 1 also changed the conformation of the IDR of huntingtin and seems to change its accessibility for binding interactions. The IDR has been raised as a target for many post-translation modifications and is thus a key area in protein function. The results by Harding *et al.* support the conclusion that expanded CAG lengths of exon 1 work synergistically across the entire protein to influence structure and function. This limits the applicability of model systems relying on only expression of mutant exon 1.

Mouse models have been developed to allow full length or exon 1 containing expanded CAG to be expressed in a living, full-body organism to study as the disease progresses. There have been several noted discrepancies between clinical cases and mouse models. The short lifespan of a mouse means this system cannot accurately model the slow progression of HD in humans (34,35). To combat the discrepancy in lifespan between mice and humans, CAG expansions that are far outside of the clinical window are used to quicken the disease progression (34). The R6/2 mouse line features 150 CAG repeats and the N171-82Q model features 82 CAG repeats (9,36,37). If these expansions were to be seen in clinical practice, it would represent a juvenile case of HD (4,9). Both the R6/2 and N171-82Q transgenic mouse models involve expression of segmented huntingtin, rather than expression of full-length huntingtin. Not only do these models neglect the function of the whole protein, but they also disregard the stoichiometry of a natural system. Yang *et al.* compared huntingtin protein levels of extracts from R6/2 mice, knock-in mice with varying huntingtin segment lengths and control mice (38). Knock-in mice were generated from an original HD140Q knock-in mouse strain of full-length huntingtin, and then gene editing tools were used to delete portions of the gene. Gene R6/2 mice show dramatically higher levels of exon 1 protein levels relative to knock-in models and much higher degree of soluble protein aggregation (38). The higher level of protein aggregation was not related to CAG length as the R6/2 mouse contains

150 CAG repeats, and the knock-in models contain 140 CAG repeats (38). Rather, it is hypothesized, that the degree of overexpression of mutant exon 1 in R6/2 mice may limit the ability for available chaperones to mediate misfolding events (38). This could further be used to explain the non-specific aggregation that is present in R6/2 mice compared to human HD (38). There is supporting evidence that overexpression of mutant exon 1 in these mouse models may be contributing to the rapid and extreme disease phenotype development (39,40). Additionally, mouse models such as the R6/2 model involve the development of disorders not correlated with human HD (epilepsy, diabetes, cardiac dysfunction) (40–44). While mouse models of full-length huntingtin tend to model disease more accurately, the progression of disease is slow (45,46). YAC128 and BACHD transgenic mouse models of full-length huntingtin show a weight gain phenotype, whereas weight loss is typically seen in human HD (47–49).

Extending beyond mouse models of disease, pig models have been generated to model the response of a large animal to CAG expansions (50). Similar to the mouse models, this had CAG expansions that were beyond clinical cases at around 140-150 CAG repeats (50). HD pig models displayed more severe phenotypes as compared to HD mouse models at the same CAG expansions (50–53). This could mean that mouse models show a greater tolerability to handle CAG expansions than larger animal models and humans (50). As such, this presents another confounding variable to the use of mice, where they again fail to represent the experience of a human patient.

Another widely used system is transformed cells. Cells are transformed from both human and animal models to have a high proliferative capacity and ease the practicality of culturing these cells. To extend the proliferative capacity, the genome integrity is compromised and the tumor suppressor p53 protein is inhibited (54–56). P53 is a transcription factor activated in response to cell stress events like DNA damage, ROS elevation, and hypoxia (57–60). Not only are these stress events implicated in HD, huntingtin directly interacts with p53 (57). In response to events of cell stress, p53 is activated and interacts with huntingtin to increase its transcription (57). The close regulation of huntingtin levels in response to cellular conditions present another limitation of many model systems. As described in mouse models, cellular models of HD also fail to recognize the importance of endogenous protein level and full-length huntingtin function. Cellular models of HD often employ the used of transformed human embryonic kidney (HEK) cells for their rapid growth rates (61), high transfection efficiency (62), and high level of protein production (63). For these practical reasons, researchers have transfected mutant exon 1 into HEK cells to answer questions related to HD. Researchers have studied the influence of various compounds on cell toxicity and aggregate formation (64,65) to identify a state of increased glycolysis (66), reduced levels of adenosine triphosphate (ATP) (67), and increased membrane fluidity with mutant exon 1 (68), A model system that is not reflective of clinical disease severely limits the conclusively of these results.

This dramatically limits the knowledge that can be gained from this model system as p53 is not only an important regulator of the stress response, DNA damage, and DNA repair, but has also been shown to directly influence huntingtin's transcription levels (57,69–71). In response to cellular stress Appropriate levels of huntingtin reflective of an endogenous system are a significant limitation to most model systems of HD. Stochastic levels of huntingtin expression represent an important caveat to most model systems of HD.

Cultured patient fibroblasts derived from skin biopsies overcome many of the difficulties of transformed and mouse models. Derived from patients, these cell lines represent a clinically accurate model system with CAG expansions that are seen in a clinical setting. These models are not widely used due to their limited reproducibility as they undergo normal cell division and are subject to telomere-controlled senescence (72,73).

The next breakthrough in model systems came with the development of TruHD immortalized patient fibroblasts (74). Derived from biopsies of 3 different patients, the primary fibroblasts were immortalized with human telomerase reverse transcriptase (hTERT) (74). Immortalization with hTERT extends their proliferative capacity while maintaining genome stability without impairment to p53 function (75,76). This enables strong reproducibility with an accurate method of observing huntingtin's response to cell stress and DNA damage. hTERT

immortalization was employed in the generation of the frequently used human retinal pigment epithelial (RPE1) cell line (ATCC). RPE1 cells were derived from a healthy female and are often used to for their practical ease in culture and their maintained genomic stability. The individuals used for derivation of TruHD cells were a control female (TruHD-Q21Q18) sampled at age 52, heterozygous male (TruHD-Q43Q17) sampled at age 54 with disease onset at age 50, and homozygous female (TruHD-Q50Q40) sampled at age 23 with disease onset at age 28 (74). Information beyond these attributes was very limited due to ethical concerns regarding confidentiality.

These cells were subject to several validation assays before their use as a representative model of HD (74). One method of validation was huntingtin's response to cellular stress. Mutant huntingtin in HD increases the cell's susceptibility to death during stress (22,28,74). This was demonstrated with a lower cell viability in response to oxidizing agent $KBrO_3$ in the diseased TruHD-Q43Q17 and TruHD-Q50Q40 cells compared to healthy TruHD-Q21Q18 cells (74). Past findings have also indicated an energy deficit in HD that was validated in TruHD cells with an adenosine diphosphate (ADP)/ATP ratio (54,74,77). Consistent with past results, the TruHD-Q43Q17 and TruHD-Q50Q40 show a high ADP/ATP ratio indicative of energy deficit as compared to healthy TruHD-Q21Q18 cells (54,74,77). TruHD cells have also been used to investigate the relationship between HD and DNA damage. An alkaline comet assay, a

technique used to measure DNA damage through the presence of a longer “tail” moment was used to measure the relative amount of DNA damage in the presence of $KBrO_3$ in TruHD-Q43Q17 and TruHD-Q21Q18 (19). The TruHD-Q43Q17 exhibited more DNA damage indicated by a longer tail moment (19). One critical limitation of TruHD cells is the different genetic background between cell lines. The GWAS in 2015 emphasizes the impact of genetic background aside from CAG length to onset of disease. This brings some uncertainty to how different genetic backgrounds are influencing results from these cells (11). Specifically, it cannot be concluded that differences between TruHD cells are only due to the length of the CAG expansion.

To isolate the contribution of the CAG expansion, it is necessary to derive cells from patients that are genetically identical except for the CAG expansion length. This is termed the development of isogenic cells. The cells would be directly comparable, with any changes in phenotype or results being directly due to a different CAG expansion. Various techniques in genetic editing would allow for the TruHD immortalized patient fibroblasts to be edited to a custom CAG expansion. We could attribute any changes in results to the length of the CAG expansion rather than confounding genetic differences between the patients’ genetics.

1.4 Genetic Editing Tools

Zinc-Finger Nucleases (ZFNs) are one method of making a precise change to the genome. They consist of a domain of DNA binding sequences that are adapted from zinc-finger transcription factors, and a FokI restriction enzyme endonuclease domain (78). They work in a paired system where the DNA targeted for cleavage is surrounded by a ZFN on the forward and reverse strand of the DNA (78). When both ZFNs flank the DNA, the FokI enzymes dimerize and create a double stranded break (DSB) in the DNA sequence (78). The DSB can be handled through nonhomologous end joining (NHEJ) or homology-directed repair (HDR) (78,79). While HDR is only available for the cell to use when there is a sister chromatid available during late S phase or G_2, NHEJ can be used at any point of the cell cycle (80). NHEJ is very error-prone and often leads to frameshift mutations that silence the target gene (78,79). When using ZFNs to adjust the CAG expansion to a custom length, a donor plasmid with a custom CAG expansion that is flanked by homologous sequences can be introduced into the cells. The cell can use the introduced plasmid to repair the DSB through HDR and introduce the desired CAG expansion. While ZFNs offer the potential for precise genome editing, there are several caveats that limit effective use. Off-target effects have to be considered due to potential binding to other sequences in the genome (81). To limit off-target effects, small modifications can be made to the technique such as using FokI nucleases that are obligate heterodimers (82,83). This improves targeting accuracy by requiring both nucleases to be

present and flanking the target DNA region. The other difficulty with implementation of ZFNs is their practical application. Expertise is required to assemble ZFNs that stretch to longer target DNA sequences (80,84). To combat this, a consortium has been developed with ZFN sequences to match various DNA sequences (85,86). However, even with this addition, researchers are still limited in targeting any DNA sequence, thus making specific site changes to the DNA difficult (78,80).

The discovery of transcription activator-like effector nucleases (TALENs) improved the foundation for precise genetic editing. TALENs function similarly to ZFNs in that they are composed of a DNA binding domain and a FokI nuclease that when dimerized create a DSB in the target DNA (80). The DNA binding domain of TALENs is composed of transcription activator-like effector repeats (80). These repeats contain a variable di-residue, or two amino acids in sequence, that match with a specific nucleotide base (80,87,88). Understanding the code that matches the adjacent amino acids to the DNA base has enabled precise gene editing. Unlike ZFNs, TALENs are far simpler to design, with repeats extending to whatever length is desired for targeting (89,90). TALEN specificity has been shown to be quite high in whole-genome sequencing, with an added layer of accuracy when FokI is modified to an obligate heterdimer (80,91). As with ZFNs, this means a DSB occurs only when both FokI are flanking the target DNA (80).

The current popular method for genetic editing is clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein (Cas)9. The CRISPR-Cas9 system overtook the use of ZFNs and TALENs with its practical advantages. This system was first discovered as an adaptive immune response to invading pathogens in bacteria and archaea (92). The immune response has three stages: adaptation, expression and interference (93). During the adaptation stage, bacteria incorporate invading DNA into their genome into spacer sequences at the CRISPR locus (93). At the expression stage, these spacer sequences are transcribed with CRISPR RNA (crRNA) and activated by complementary sequences between crRNA and a separate molecule of trans-activating crRNA (tracrRNA) (93). This complex combines with a Cas endonuclease expressed from an adjacent locus (94). At the interference stage, the spacer sequences in the complex act as a guide for the Cas enzyme to cleave and inactivate the target DNA (93). The CRISPR system is incredibly diverse, with a rapidly changing classification system. The current classification lists 2 classes, 6 subtypes and 33 subtypes (95). The Type II CRISPR system is the most widely used method for gene editing, with only Cas9 needed to recognize and cleave target DNA.

Researchers were able to manipulate this system of defense to create a very practical gene editing technique for creating DSBs (96–99). Plasmids can be

developed carrying a sequence for Cas9 and a guide RNA (gRNA) that is composed of crRNA, tracrRNA and a sequence complementary to the target DNA (80). This plasmid can be transfected into mammalian cells and expressed. As in the adaptive immune defense system, the gRNA and Cas9 combine to form a complex. The gRNA targets the Cas9 to its target DNA sequence, and the Cas9 enzyme targets it for cleavage and introduces a DSB. As with ZFN and TALEN gene editing techniques, the DSB can be repaired with NHEJ or HDR (78,80,97). To change the DNA sequence, researchers can transfect a donor plasmid with the gRNA/Cas9 plasmid containing the desired DNA sequence to enable HDR at the target DNA. Unlike ZFNs and TALENs which require changing protein sequences, CRISPR-Cas9 only requires a change to a segment of the gRNA for DNA targeting (80). The Cas9 enzyme is expressed in the cell as protein but does not require recoding. This enables simple, and almost limitless changes to be made to the gRNA for DNA targeting without the expertise, time and cost required for ZFNs and TALENs.

The most concerning limitation for CRISPR-Cas9 is the high rates of off-target mutagenesis that have been reported in human cells (100–102). One method used to overcome this limitation is the use of a Cas9-nickase enzyme with paired single gRNAs (sgRNAs) (103). This modifies the Cas enzyme to produce nicks in the DNA rather than DSBs. The targeted DNA is flanked by two sgRNA each with their own Cas9-nickase in the DNA, that only when both are at

their target DNA will produce a DSB. Off-target nicks that would occur when one gRNA/Cas complex cleaves DNA would be repaired with high fidelity, and the addition of a second sgRNA extends the DNA sequence required for binding to further promote accuracy (103,104). Editing efficiency largely depends on the parameters of the experiment (knock-out, knock-in, substitution, length of substitution, delivery method, cell type), but is generally low for traditional plasmid-based delivery of CRISPR constructs. One approach to improving efficiency to 50-66% is genetically silencing the NHEJ pathway to promote HDR (105). Alternatively, the use of viral vectors is a highly efficient delivery method of constructs (106). Lentiviruses are a common approach for their ability to transduce non-dividing cells (107). Transduction results show editing efficiency of greater than 90% in certain cases (108). The major limitation of lentiviral based approaches is the long-term stable expression of Cas9 to introduce off-target cleavage events (109). There are methods of circumventing this long-term exposure, by pre-packing a Cas9 protein and gRNA into the viral vector for delivery (110).

Similarly, the use of ribonucleoprotein complexes (RNPs) also involves the direct delivery of a packaged construct into cells. This method involves the direct transfection of a pre-packaged Cas9 protein and gRNA into cells (111). This direct delivery of protein promotes rapid clearance of Cas9 from cells and

subsequently limits the opportunity for off-target cleavage and cell toxicity (111,112).

Another consideration for CRISPR-Cas9 is the required protospacer-adjacent motif (PAM) downstream of the target DNA (97). A ternary complex is formed between the gRNA binding the target DNA, and the Cas9 binding the PAM sequence (97). The PAM sequence changes depending on the Cas variant, with Cas9 from *S.pyrogenes* requiring an NGG motif (99,113). This does introduce some limitations into the sequences available for targeting, but many variants of Cas have been developed with a focus on improving the flexibility of the PAM to increase the available DNA sequence for targeting (114). Chatterjee *et al.* were able to relax the PAM requirement by using ScCas9, an ortholog to SpCas9 (115). This Cas9 enzyme has a more flexible NNG PAM requirement. They were able to improve the specificity and efficiency of this Cas9 enzyme by replacing its critical positively charged loop structure with another positively charged loop carrying a flexible “SG” motif. They called this new Cas9 enzyme Sc++ (114,116). Similarly, Walton *et al.*, were able to expand the PAM requirement by mutating the PAM-interacting domain (PID) of SpCas9 (117). This Cas9 enzyme was named SpRY and had PAM requirements that were specific to NRN and weakly specific to NYN sequences. As improvements continue towards improving the efficiency and targetability of CRISPR-Cas9, it becomes the leading gene editing tool.

1.5 Creating Isogenic Lines in HD

CRISPR-Cas9 genetic editing has been used to generate isogenic lines in HD. Studies have used CRISPR-Cas9 to correct disease to a normal CAG expansion or cause disease from a healthy CAG repeat length. In 2014, An *et al.* used CRISPR-Cas9 to modify the CAG expansion in human induced pluripotent stem cells (hiPSCs) and transformed 293-F cells (118). They modified the hiPSCs and 293-F cells with a donor plasmid carrying a 97 CAG repeat expansion to develop an array of isogenic cells with CAG repeats of 21, 72 and 97. In 2017, Xu *et al.* used CRISPR-Cas9 to correct a repeat length of 180 CAGs in hiPSCs with a donor plasmid containing 18 CAG repeats (119). Malankhanova *et al.* in 2020 established a protocol that was modeled in HEK293 cells to create an isogenic line using CRISPR-Cas9 technology that enables editing of any CAG expansion length (120). The technology has been extensively studied, with minor modifications to the CRISPR-Cas9 system to improve the efficiency and accuracy of the technique. These studies feature mostly iPSCs as their target for isogenic lines. While iPSCs are enticing for their reprogram-ability into neurons, there are significant concerns about the use of these cells as a model system for HD. Mainly, when iPSCs are reprogrammed into neurons, these cells lose their age-specific profiles and are not representative of the mature cells that are lost in HD (121) Additionally, many studies use CAG expansions that are not representative of clinical cases that may give results that are not applicable for the vast majority

of HD cases (118–120). The creation of an isogenic line from immortalized TruHD cells would present a model of HD that is clinically representative and accurate without the limitations of most model systems.

1.6 Project Rationale

Aim 1: Generating an isogenic cell line

Isolating the effect of the CAG expansion in HD is a critical component to better understanding the disease. The model systems currently available are limited by a failure to represent a clinical case of HD. The immortalization of TruHD fibroblasts derived from patients offer an accurate model of HD, but as highlighted by the 2015 GWAS study, they are confounded by other genetic differences between the patients (11). To isolate the effect of the CAG expansion, gene editing will be used to modify the CAG expansion of the TruHD cells. CRISPR-Cas9 gene editing offers the most ease in practice for an accurate and efficient cleavage system, with a modified Cas9-nickase and paired gRNA system used to minimize off-target events. To account for the patient's genetics outside of the CAG expansion, we intended to use two systems of genetic editing. TruHD-Q21Q18 were to be modified from healthy to disease with the introduction of a 45 CAG donor plasmid, and TruHD-Q43Q17 were to be corrected from disease with a healthy 18 CAG repeat donor plasmid. The inclusion of separate modified CAG expansion lines adds an additional layer of consistency to ensure that phenotypes are not influenced by the patient's

background and allows us to investigate the impact of sex differences in HD. After facing technical challenges with transfection and growth in TruHD cells we turned to an alternative cell line, RPE1 cells, with the intention of editing in a diseased level of CAG repeats in otherwise healthy cells. This would retain the same purpose of isolating the effect of a changing CAG expansion by comparing pre and post edited cells in a genetically stable cell line. We hoped to use these generated isogenic cells to identify differences in phenotypes before and after editing to directly identify how the CAG expansion influences phenotypes.

Aim 2: Propagate TruHD-Q43Q17M cells with CAA interruption

CAA interruptions are shown to have a protective effect in HD (15). I will establish a cell line with a CAA interruption in the polyglutamine tract of TruHD-Q43Q17 cells using a novel, PAM-less Cas9 enzyme. SpRYc was combined with a cytosine base editor to form SpRYc-BE4Max and targeted to the CAG repeat tract with a corresponding sgRNA. SpRYc-BE4Max was targeted to modify the 4th CAG repeat to CAA, reducing the number of CAG repeats below 40, the threshold for development of disease. Propagating cells with this genetic modification would be useful as prior studies have already established that the length of the uninterrupted CAG expansion, rather than the length of glutamine residues, are associated with symptom onset (14–16). A stable cell line containing this CAA interruption would be useful as a comparison against TruHD-Q43Q17 cells, to further investigate how CAA interruptions influence HD. In this

case, TruHD-Q43Q17 cells and cells carrying the CAA interruption would again be genetically identical aside from the single mutation reducing the uninterrupted CAG length to 39 repeats.

Aim 3: Expression of transfected exon 1 relative to endogenous huntingtin

The levels of endogenous huntingtin are tightly controlled in a cell. Many models of HD rely on models of overexpressed wild type and mutant exon 1 to make conclusion about disease. It is important to recognize that experiments involve greatly overexpression exon 1 that is not reflective of the stoichiometric conditions of a human person with HD. We transfected wild type and mutant exon 1 in increasing concentrations in the frequently used, HEK293 transformed cell line. We lysed cells and conducted a western blot to compare the difference between transfected exon 1 and endogenous huntingtin. If there is a large degree of difference between exogenous and endogenous huntingtin levels, it casts doubt on the use of transfected exon 1 as a model to assess HD.

Chapter 2.0 Experimental Procedures

2.1 Tissue Culture

Human fibroblast cells immortalized with hTERT from heterozygous male TruHD-Q43Q17 and healthy female TruHD-Q21Q18 were cultured in MEM supplemented with 10% fetal bovine serum (FBS) and 1% Glutamax (74). Human

RPE1 cells immortalized with hTERT were acquired from American Type Culture Collection (122). RPE1 cells were cultured in DMEM/F12 1:1 supplemented with 10% FBS and 0.01mg/mL hygromycin B. HEK 293 cells were cultured in DMEM/F12 1:1 supplemented with 10% FBS. All cells were maintained at 37℃ in 5% CO_2 incubator.

2.2 Plasmid Constructs

Oligos for sgRNA-a/nCas9 (Addgene, no. 87201) and sgRNA-b/nCas9 (Addgene, no. 87200) in cloned Cas9-nickase expression vector pX335, and donor plasmids pJOP-HTT-HR45Q (45Q donor plasmid) (Addgene, no. 92248), and pJOP-HTT-HR18Q (18Q donor plasmid) (Addgene, no. 87228) were ordered. After seeing high toxicity levels, we developed new gRNA vectors: sgRNA-a and sgRNA-b oligos were designed according to Xu *et al.*, 2017. Both oligos were PCR amplified and cloned into separate BPK1520 vectors (Addgene, no. 65777) using BsmBI restriction enzyme digestion and subsequent ligation. Colonies were confirmed as successful ligation products with Sanger sequencing. NLS-nCas9-NLS-P2A-EGFP (nCas9-EGFP) (Addgene, no. pRZ70) was ordered. To replace green fluorescent protein (GFP) selection marker with zeocin resistance, pJOP-HTT-HR45Q (Addgene, no. 92248) was digested with AsiSI and XhoI. For zeocin resistance, pFUSE-hIgG1-Fc1 (InvivoGen) was ordered. Zeocin resistance marker in pFUSE-hIgG1-Fc1 was PCR amplified with forward primer carrying AsiSI recognition site:

TTAGCGATCGCGTACAAGAATTAATTGGACG and reverse primer carrying XhoI recognition site: TAACCTCGAGAATAGGAGTCAGGACG. The PCR amplified DNA fragment with zeocin resistance at ~1700 base pairs was digested with AsiSI and XhoI. We attempted to ligate the digested PCR fragment into the pJOP-HTT-HR45Q donor plasmid, though no colonies were successfully isolated. For experiments with TALENs, downstream (DSPQ) vector pairs of TALENs were ordered: DSPQ-HTT-TALEN1 ZEO (TALEN1) (Addgene, no. 92245) and DSPQ-HTT-TALEN 2 BSD (TALEN2) (Addgene, no. 92246). Plasmid constructs for HTT gRNA, SpRYc BE4Max, and SpCas9 BE4Max were provided by Koseki *et al.*

2.3 Transfection of TruHD-Q43Q17 cells for development of isogenic cells

TruHD-Q43Q17 cells at 95-100% confluency were transfected with varying concentration of sgRNA-a/nCas9, sgRNA-b/nCas9 and pJOP-hTT-HR18Q (concentration ranged from 0-12µg depending on experimental conditions) according to Amaxa 4D Lonza Nucleofection protocol with SG solution. Samples were pulsed with program EN-158 and remained undisturbed for 45 minutes and then plated with fresh growth media. Growth media was replaced after 24hours (h) and cells were imaged 24-48h after transfection to assess GFP expression.

2.4 Transfection of RPE1 cells with Amaxa 4D Lonza Nucleofector

RPE1 cells were transfected according to 3 separate methods. In method 1, RPE1 cells at 90-100% confluency were transfected with varying concentration

of Method 1 constructs : sgRNA-a/nCas9, sgRNA-b/nCas9 and pJOP-HTT-HR45Q (concentration ranged from 0-5µg depending on experimental conditions); Method 2 constructs : sgRNA-a/BPK1520, sgRNA-b/BPK1520, pJOP-hTT-HR18Q and NLS-nCas9-NLS-P2A-EGFP (concentrations ranged from 0-5µg); Method 3 constructs: DSPQ-HTT-TALEN1 ZEO, DSPQ-HTT-TALEN 2 BSD and pJOP-HTT-HR45Q (concentrations ranged from 0-5µg) . RPE1 cells were transfected with constructs according to Amaxa 4D Lonza Nucleofection protocol with SF solution. Samples were pulsed with either DN-100 or EA-104 programs depending on experiment. Following pulsing, cells remained undisturbed for 45 minutes and then plated with fresh growth media. Growth media was replaced after 24h and cells were imaged 24-48h after transfection to assess GFP expression. Transfections were also attempted using Lipofectamine 3000 Transfection Reagent (Invitrogen) protocol and *Trans*IT-X2 (Mirus). RPE1 cells at 90-100% confluency were transfected according to Lipofectamine 3000 Transfection Reagent (Invitrogen) protocol and *Trans*IT-X2 (Mirus) into serum free media. Serum free media was replaced with fresh growth media 2h after transfection and cells were imaged 24-48h after transfection to assess GFP expression.

2.5 Transfection of HEK 293 cells for development of isogenic cells

HEK293 cells at 70-90% confluency were transfected according to Lipofectamine 3000 Transfection Reagent (Invitrogen) protocol and *Trans*IT-X2

(Mirus) into serum free media. Serum free media was replaced with fresh growth media 2h after transfection and cells were imaged 24-48h after transfection to assess GFP expression.

2.6 Transfection of TruHD-Q43Q17 cells for SpRYc base editor

TruHD-Q43Q17 cells at 95-100% confluency were transfected with (1) 500ng sgRNA, (2) 250ng SpCas9-BE4Max and 250ng sgRNA or (3) 250ng SpRYc-BE4Max and 250ng sgRNA according to Amaxa 4D Lonza Nucleofection protocol with SG solution. Samples were pulsed with program EN-158 and remained undisturbed for 45 minutes and then plated with growth media. Growth media was replaced after 24h. Cells were allowed to grow for 5 days, and then DNA was extracted with PureLink Genomic DNA Mini Kit (Invitrogen).

2.7 Fluorescence Activated Cell Sorting

McMaster Core Flow Facility was used for sorting. Depending on experimental conditions, 1-3 days following transfection, cells were prepared for sorting. Cells were sorted during log phase. Ideally, cells are sorted during log phase to promote cell survival and reduce cell adhesion. Cells were prepped directly before sorting. Ethylenediaminetetraacetic acid (EDTA) will be added to cells prior to sorting to reduce cell adhesion. For sorting of cells with GFP expression of pJOP-HTT-HR45Q/pJOP-HTT-HR18Q in development of isogenic cells, preps included: (1) sorting sample, cells transfected with gene editing

constructs, stained with eBioscience 7-AAD Viability Staining Solution (Invitrogen); (2) GFP control, sorting sample of cells transfected with gene editing constructs, unstained for eBioscience 7-AAD Viability Staining Solution; (3) unstained control, cells transfected with sonicated salmon sperm DNA at equal concentration to gene editing construct, unstained for eBioscience 7-AAD Viability Staining Solution. For sorting of cells with SpRYc-BE4Max base editor, prepared samples included: (1) sorting sample, sample transfected with SpRYc-BE4Max and sgRNA, stained with eBioScience 7-AAD Viability Staining Solution (2) GFP control, cells transfected with 0.5µg pmaxGFP Vector (Lonza) and brought to DNA concentration levels of gene editing sample with sonicated salmon sperm DNA, one population stained for eBioscience 7-AAD Viability Staining Solution, and one population unstained; (3) unstained control, cells transfected with sonicated salmon sperm DNA at equal concentration to gene editing construct, unstained for eBioscience 7-AAD Viability Staining Solution. Prior to sorting, cells were centrifuged at 20℃ at ~700rpm for 5 minutes, washed in flow buffer (PBS with 2.5 mM EDTA and 0.5% bovine serum albumin) and resuspended in flow buffer. Cells were sorted into a 96-well plate with fresh growth media and grown from individual cell populations. When sorting TruHD-Q43Q17 cells, conditioned media from a fully confluent plate of TruHD-Q43Q17 cells was used to support growth.

2.8 Microscopy

Cells were stained with NucBlue Live ReadyProbes Reagent (Hoechst 33342) (Invitrogen) for live cell imaging. Images were acquired with the EVOS FL Auto 2 inverted widefield microscope (Thermofisher Scientific).

2.9 Transfection Efficiency Calculation

To estimate transfection efficiency, CellProfiler was used to compare 5µg pJOP-HTT-HR18Q with 1µg sgRNA-a/nCas9 and 1µg sgRNA-b/nCas9 to GFP Control (0.5µg pmaxGFP Vector with 6.5µg of sonicated salmon sperm DNA) in TruHD-Q43Q17 cells. Transfected efficiency was calculated as GFP-positive cells above threshold background fluorescence/nuclei identified with Hoechst stain across 10 images for each condition.

2.10 Antibodies

Antibodies against N-terminal domain of huntingtin (anti-N17) were previously characterized and validated (29). For secondary, goat anti-rabbit IgG, HRP (Invitrogen) was used.

2.11 Transfection of HEK293 cells for analyzed exon 1 expression

HEK293 cells at 70-90% confluency were transfected with mCerulean (mCer) – Q17 Exon1- Yellow Fluorescent Protein (YFP) vector at 250ng, 500ng, 1000ng, 2000ng, and Q138 Exon1 -YFP plasmid at 250ng, 500ng, 1000ng,

2000ng, and control YFP vector according to Lipofectamine 3000 Transfection Reagent (Invitrogen) protocol into serum free media. DNA concentrations at transfection were balanced with sonicated salmon sperm DNA to be 2000ng per condition. Serum free media was replaced with fresh growth media 2h after transfection.

2.12 Western Blot analysis

24-48h after transfection, cells were scaped and lysed with lysis buffer (200µl of RIPA (Radio Immuno Precipitation Assay) buffer (50mM Tris HCl pH 7.4, 50mM NaCl, 2mM EDTA, 0.1% Sodium dodecyl-sulfate (SDS)) with 10% phosphatase and protease inhibitors (Roche). Cells were centrifuged at 10 000 x g for 12 minutes. The supernatant was collected and prepared for loading in 4-20% polyacrylamide gradient gels (Biorad) for SDS-polyacrylamide gel electrophoresis (PAGE). Gel was blotted on polyvinylidene difluoride (PVDF) membrane (Millipore). Membrane was treated with blocking buffer (5% skim milk powder in 1X TBS-T (1ml of TWEEN 20 (Millipore) in 1L of 1X TBS)) overnight at 4℃ or for 1 hour at room temperature. Membrane was treated with anti-N17 primary diluted at 1:1000 in blocking buffer for 1h at room temperature. Membrane was washed 3 times in TBS-T, and 3 times in TBS-T/Blocking Buffer for 5 minutes. Membrane was incubated with secondary antibody diluted at 1:50 000 in blocking buffer for 30 minutes at room temperature. Membrane was washed 3 times in TBS-T, and 3 times in TBS-T/Blocking Buffer for 5 min each.

Finally, membrane was treated with enhanced chemiluminescence (ECL) (Millipore) and visualized using Microchemi (DNR Bio-Imaging Systems). mCer – Q17 Exon1- YFP appeared at approximately 70 kDa, Q138 Exon1- YFP appeared at 45 kDa, endogenous huntingtin appears at 348 kDa. Levels of exon 1 were quantified using ImageJ relative to endogenous huntingtin as loading control.

CHAPTER 3.0 RESULTS

3.1 Developing stable isogenic cells

Our goal was to develop an isogenic cell line only differing in the length of the CAG expansion. Due to practical concerns, we resorted to various strategies and cell types to attempt to achieve this goal (Fig. 1).

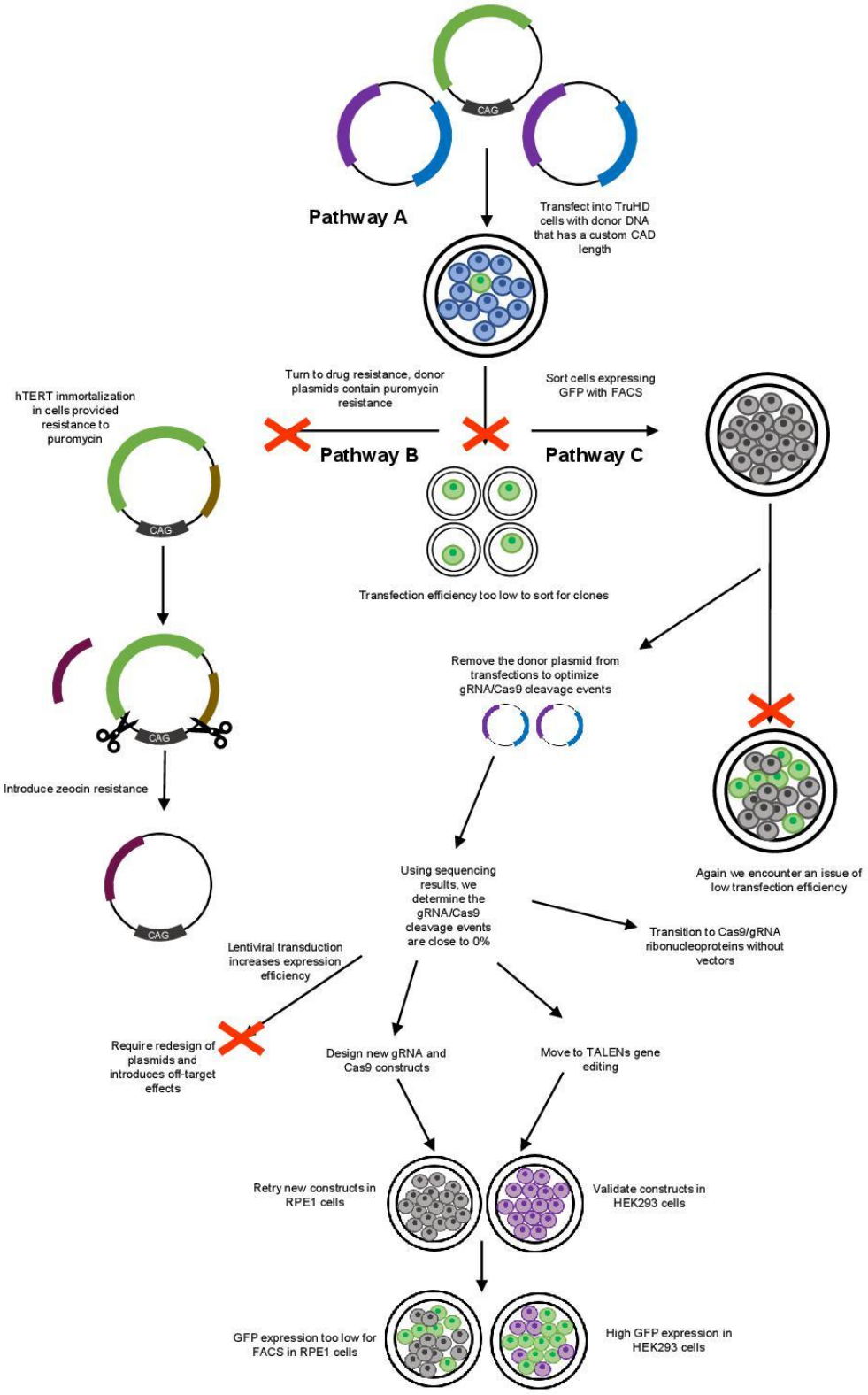

CAG
Pathway A
Transfect into TruHD cells with donor DNA that has a custom CAD length
Turn to drug resistance, donor plasmids contain puromycin resistance
Pathway B
Sort cells expressing GFP with FACS
Pathway C
hTERT immortalization in cells provided resistance to puromycin
Transfection efficiency too low to sort for clones
Remove the donor plasmid from transfections to optimize gRNA/Cas9 cleavage events
Introduce zeocin resistance
Again we encounter an issue of low transfection efficiency
Using sequencing results, we determine the gRNA/Cas9 cleavage events are close to 0%
Transition to Cas9/gRNA ribonucleoproteins without vectors
Lentiviral transduction increases expression efficiency
Require redesign of plasmids and introduces off-target effects
Design new gRNA and Cas9 constructs
Move to TALENs gene editing
Retry new constructs in RPE1 cells
Validate constructs in HEK293 cells
GFP expression too low for FACS in RPE1 cells
High GFP expression in HEK293 cells

Figure 1 – Workflow for the generation of isogenic cells using gene editing

Outline of 3 attempted methods in the development of isogenic cells. Pathway A describes the original workflow to generate isogenic cells using CRISPR gene editing in TruHD cells sorted by the presence of GFP expression through FACS. When transfection efficiency proved to be too low for efficient sorting, we turned to two alternative solutions. Pathway B involves using drug resistance for selection of donor plasmids. Pathway C overcomes issues of transfection efficiency/toxicity by using RPE1 cells. We first attempted to again use GFP from the donor plasmid as a selection marker, but transfection efficiency remained too low. We then analyzed the presence of any editing events by optimizing the transfection of gRNA/Cas9 plasmids and identifying the presence of insertions or deletions (indels). We then considered the use of lentiviruses, a redesign of our CRISPR constructs, a transition to TALEN constructs, or a transition to RNPs. We validated that GFP expression was functional by expressing system constructs in HEK293 with subsequent GFP expression that allowed us to perform FACS. We again, attempted experiments in RPE1s, though the GFP expression remained too low for FACS.

3.2 Developing Isogenic Cells, Pathway A: FACS selection of pJOP-HTT-HR18Q donor plasmid in TruHD-Q43Q17 cells

We aimed to reduce the CAG repeat length to 18 repeats in TruHD-Q43Q17 cells prior to gene editing TruHD-Q21Q18 cells. We followed the

experimental framework as presented in Xu *et al.* 2017. Several optimization experiments to determine the optimal DNA concentration of the 18Q donor plasmid were conducted in TruHD-Q43Q17 cells (Table 1). We established the brightest GFP signal 48h post transfection of 5µg of 18Q donor plasmid (Fig. 2). Following transfection, we attempted to use Fluorescence Activated Cell Sorting (FACS) to select for cells expressing the GFP signal from the 18Q donor plasmid. We attempted to sort GFP-positive cells individually into a single well in a 96-well plate and incubated for growth into more cells (Fig. 1, Pathway A). After cell growth, we planned to sequence cells to ensure they carried the intended 18 CAG repeat modification. To assess transfection efficiency prior to sorting, we imaged cells 48h following transfection of CRISPR constructs and control GFP vector (Fig. 3A). To estimate transfection efficiency, 10 images were taken of each condition and analyzed through CellProfiler to identify nuclei and GFP signal. Transfection efficiency was calculated as GFP signal beyond threshold background signal/number of nuclei stained with Hoechst. We found that cells transfected with CRISPR constructs had a transfection efficiency of 2% compared to the 16% transfection efficiency of cells transfected with control GFP vector (Fig. 3B). When we attempted to sort TruHD-Q43Q17 cells through FACS, we confirmed that transfection efficiency was too low for sorting (Fig. 4A). During sorting, cell events are progressively grouped into categories to sort for GFP-positive cells. We found a high degree of cell death with only 3.9% of cells assessed as viable through 7-AAD viability stain. Of these viable cells, cells were

compared to a GFP-threshold identified by a transfected GFP-control vector to identify GFP-positive cells. Very few GFP-positive events were found, indicated by 1 event appearing in GFP-A 10^3 to 10^4 (Fig. 4A). Since there were too few GFP-positive cells, we sorted for GFP-negative, viable cells to assess how TruHD-Q43Q17 cells would grow from a single cell. The population of GFP-negative, 7-AAD positive cells used for sorting is represented by group P5 (Fig. 4A). Without any intervention, cells were not able to grow from a single-cell, and no living cells were visualized 1 week following sorting. Transfection efficiency may be improved with higher DNA concentrations at transfection, but will also increase cell toxicity (123,124). After encountering low transfection efficiency and high toxicity, we modified the plan to include alternative strategies (Fig. 1).

Table 1 – Attempted parameters during optimization of GFP expression from donor plasmid using original constructs from Xu *et al.*, 2017

Cell type	Transfection Method	Concentrations of each gRNA/Cas9 vector (sgRNA-a/Cas9, sgRNA-b/Cas9)	Concentrations of Donor Plasmid (pJOP-HTT-HR18Q/pJOP-HTT-HR45Q)
TruHD-Q43Q17	Lonza 4D-Nucleofector	0.5-1µg	0.5-12µg

RPE1	*Trans*IT-X2	1µg	4µg
		0.5-1µg	0.5-5µg
		0.5-5µg	0
	Lonza 4D-Nucleofector	0.5-1µg	0.5-5µg
		0.5-3µg	0
		1.5-5.5µg	0
HEK293	*Trans*IT-X2	1-2µg	2-5µg
	Lipofectamine 3000	1-2µg	2-5µg

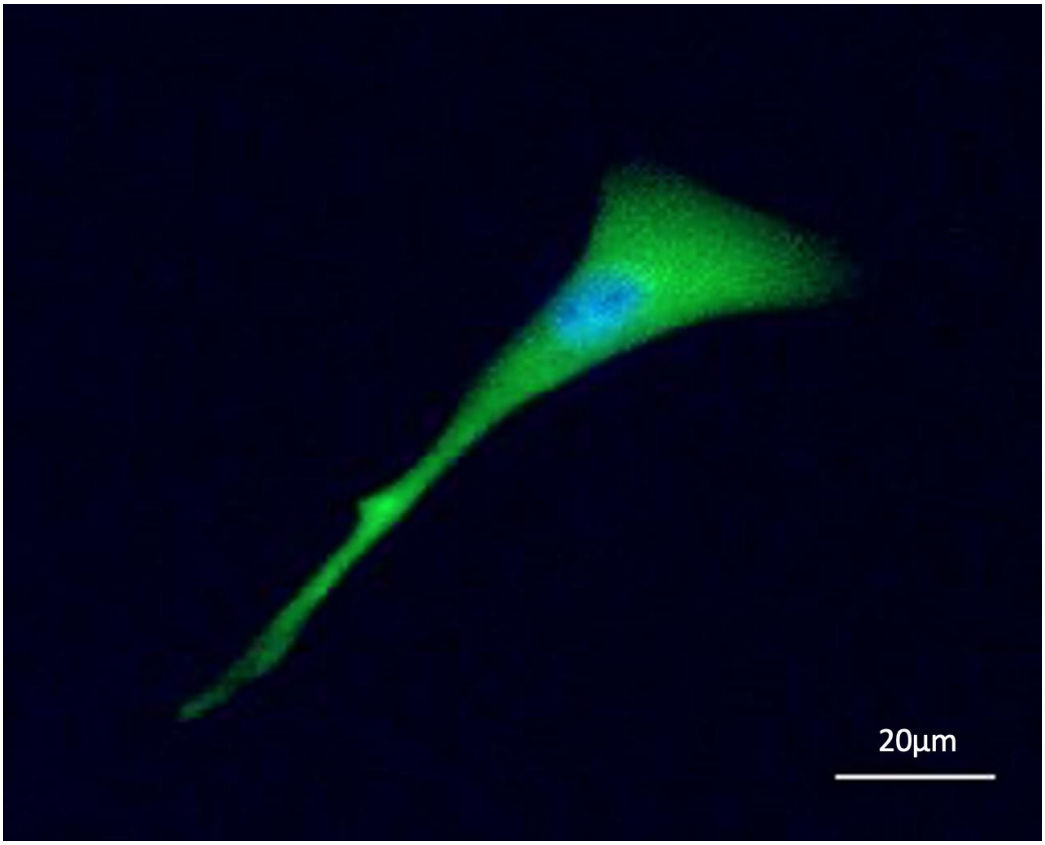


Figure 2 - Expression of pJOP-HTT-HR18Q donor plasmid in TruHD-Q43Q17 cells

TruHD-Q43Q17M cells were transfected with 5µg of pJOP-HTT-HR18Q alongside 1 µg sgRNA-a/nCas9 and sgRNA-b/nCas9 CRISPR constructs. GFP signal indicates successful expression of pJOP-HTT-HR18Q donor plasmid.

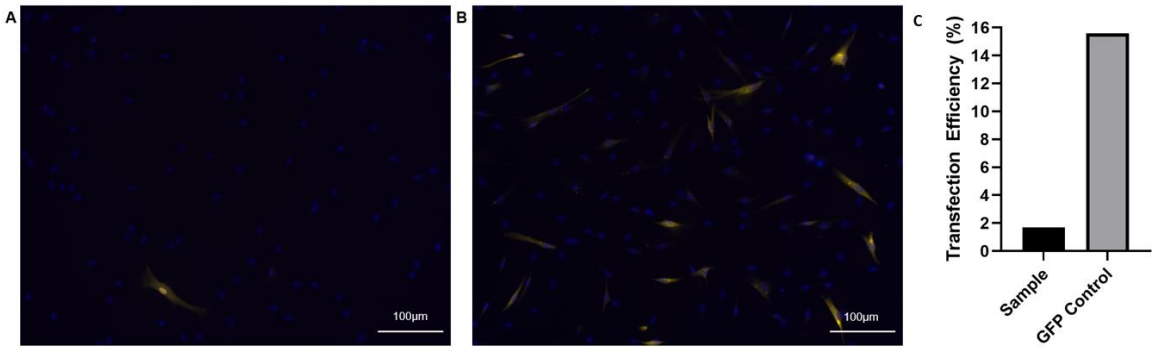
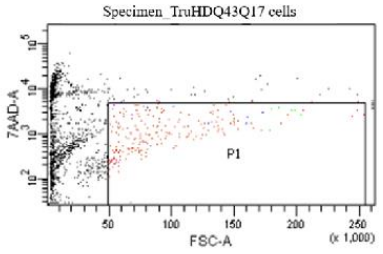
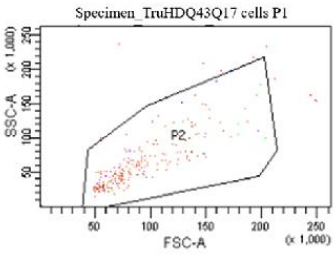


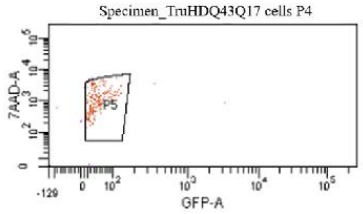
Figure 3 – Toxicity and low transfection efficiency of CRISPR constructs in TruHD-Q43Q17 cells

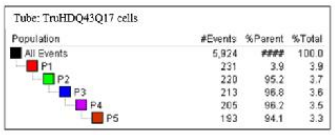
(A) Transfection of 5µg pJOP-HTT-HR18Q with 1µg sgRNA-a/nCas9 and 1µg sgRNA-b/nCas9 in TruHD-Q43Q17 cells. (B) Control transfection of 0.5µg pmaxGFP Vector with 6.5µg of sonicated salmon sperm DNA in TruHD-Q43Q17 cells. Images were acquired 48h after transfection. (C) Transfection efficiency of sample transfection (5µg pJOP-HTT-HR18Q with 1µg sgRNA-a/nCas9 and 1µg sgRNA-b/nCas9) compared to GFP Control (0.5µg pmaxGFP Vector with 6.5µg of sonicated salmon sperm DNA) in TruHD-Q43Q17 cells. Transfected efficiency was estimated as GFP-positive cells/nuclei identified with Hoechst stain across 10 images for each condition. N=1.

A

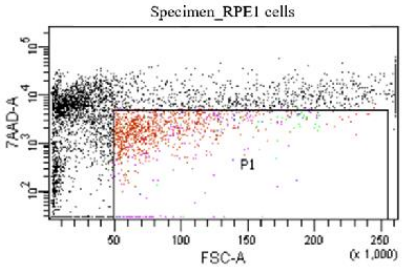

Specimen_TruHDQ43Q17 cells
7AAD-A
FSC-A
(x 1,000)
P1

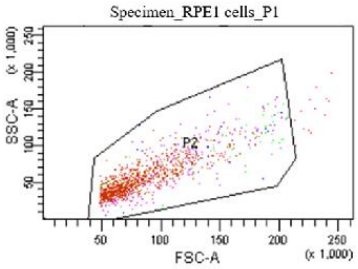

Specimen_TruHDQ43Q17 cells P1
SSC-A
FSC-A
(x 1,000)
P2

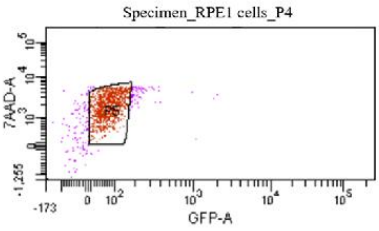

Specimen_TruHDQ43Q17 cells P4
7AAD-A
GFP-A
P5

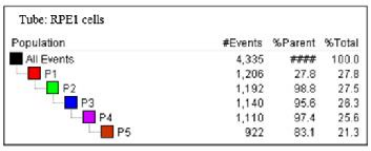

Tube: TruHDQ43Q17 cells
Population #Events %Parent %Total
All Events 5,924 #### 100.0
P1 231 3.9 3.9
P2 220 95.2 3.7
P3 213 96.8 3.6
P4 205 96.2 3.5
P5 193 94.1 3.3

B

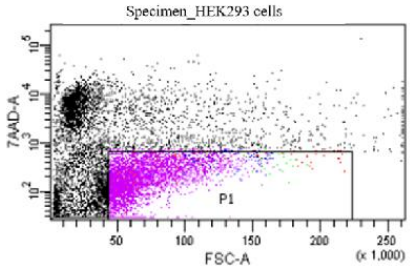

Specimen_RPE1 cells
7AAD-A
FSC-A
(x 1,000)
P1

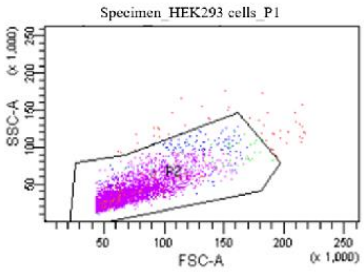

Specimen_RPE1 cells_P1
SSC-A
FSC-A
(x 1,000)
P2

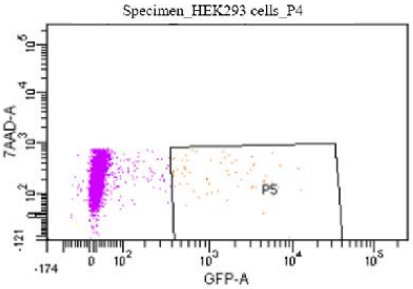

Specimen_RPE1 cells_P4
7AAD-A
GFP-A
P5


Tube: RPE1 cells
Population #Events %Parent %Total
All Events 4,335 #### 100.0
P1 1,206 27.8 27.8
P2 1,192 98.8 27.5
P3 1,140 95.6 26.3
P4 1,110 97.4 25.6
P5 922 83.1 21.3

C


Specimen_HEK293 cells
7AAD-A
FSC-A
(x 1,000)
P1


Specimen_HEK293 cells_P1
SSC-A
FSC-A
(x 1,000)
P2


Specimen_HEK293 cells_P4
7AAD-A
GFP-A
P5


Tube: HEK293 cells
Population #Events %Parent %Total
All Events 10,000 #### 100.0
P1 4,388 43.9 43.9
P2 4,341 98.9 43.4
P3 4,297 99.0 43.0
P4 4,139 96.3 41.4
P5 83 1.5 0.6

Figure 4 – FACS Data of TruHD-Q43Q17, HEK293 and RPE1 cells of transfected CRISPR constructs

All cell types were sorted 48-72h following transfection with 5µg pJOP-HTT-HR18Q (in TruHD-Q43Q17 cells (A)) or 5µg pJOP-HTT-HR45Q (in RPE1 (B) and HEK293 cells (C)) with 1µg sgRNA-a/nCas9 and 1µg sgRNA-b/nCas9. Cells were stained with 7-AAD viability dye as a marker for living cells. GFP positive cells indicated successfully transfected cells, expressing donor plasmid. Cells were grouped into P1 by size and viability. P1 was sorted into group P2 based on cell size and granularity. Specimen graph P4 indicates separation between GFP-positive cells and GFP-negative cells against cells stained with 7-AAD viability dye. (A) and (B) carried too few GFP-positive cells, and we selected to sort for non-GFP expressing, viable cells represented by P5 grouping (C) had adequate cell numbers for sorting GFP-positive cells, and thus P5 indicates GFP-positive, 7-AAD stained cells.

3.3 Developing Isogenic Cells, Pathway B: Drug Selection of pJOP-HTT-HR18Q donor plasmid in TruHD-Q43Q17 cells

As an alternative approach, we turned to drug resistance of the donor plasmid as a method of selection (Fig. 1 Pathway B). We used the same protocol and constructs as described in Method 1, where the 18Q donor plasmid is transfected with sgRNA-a/nCas9 and sgRNA-b/nCas9 plasmids in TruHD-Q43Q17 cells. Both the 18Q donor plasmid and the 45Q donor plasmid

constructs contain puromycin resistance, as this was the method of selection in Xu et al, 2017. TruHD cells also possess puromycin resistance, as this was used for the selection marker for successful transduction of hTERT during the development of the TruHD cells (74). We attempted to see if puromycin could still be a viable method of selection of CRISPR constructs, where cells expressing the 18Q donor plasmid carried additional puromycin resistance. We determined that puromycin resistance was not additive after finding comparable cell survival in TruHD-Q43Q17 cells transfected at up to 7µg of 18Q donor plasmid and control TruHD-Q43Q17 cells with puromycin concentration between 0.5-5 µg/ml (Table 2). To ensure our results were not influenced by the stress of transfection, we conducted a control transfection with pmaxGFP and single-stranded sonicated salmon sperm DNA balanced to 18Q donor plasmid transfected concentration. All trials included application of intended puromycin concentrations every 48h for up to 2 weeks. No difference was noted between any of the conditions, so we were not able to use puromycin resistance as a method of selection for CRISPR constructs.

Table 2 - TruHD-Q4317M transfected with pJOP-HTT-HR18Q do not gain additional puromycin resistance compared to resistance conferred during hTERT immortalization

Trial	DNA Concentration of pJOP-HTT-HR18Q	Control Condition	Puromycin Concentration (µg/ml)
1	5µg	TruHD-Q43Q17 cells untransfected	0, 0.5, 1, 2, 3, 4
2	5µg	TruHD-Q43Q17 cells untransfected	0, 1, 2, 3, 4, 5
3	5µg, 7µg	TruHD-Q43Q17 cells transfected with pmaxGFP (500ng) with sonicated salmon sperm DNA (6.5µg)	2,3,4

We attempted to replace puromycin with zeocin drug resistance from pFUSE-hIgG1-Fc plasmid. We successfully digested the 18Q donor plasmid with AsiSI and XhoI, removing puromycin resistance and GFP tag sequences. We designed primers targeting 1700 base pairs around the zeocin resistance sequence in pFUSE-hIgG1-Fc, adding restriction enzyme sites for AsiSI and XhoI. The zeocin resistance sequence was PCR amplified and digested with

AsiSI and XhoI. We attempted several ligation experiments, though these were unsuccessful, and no colonies were isolated afterwards.

3.4 Developing Isogenic Cells, Pathway C: Establishing an isogenic line in RPE1 cells

After experiencing very low transfection efficiency and high toxicity while troubleshooting FACS in TruHD-Q43Q17 cells, we questioned if the plasmids may be causing an unusually toxic effect in TruHD cells. We shifted our focus to using the established CRISPR constructs in RPE1 cells immortalized with hTERT (Fig. 1 Pathway C).

We attempted to optimize the transfection in RPE1 cells using various transfection methods, nucleofection protocols, and DNA concentrations (Table 1). Over 3 transfections into RPE1 cells at 0.5-5µg using *Trans*IT-X2, transfection efficiency was very poor. We attempted transfections using Lonza Nucleofector to improve transfection efficiency, which again yielded a poor transfection efficiency. We attempted FACS on transfected RPE1 cells, but transfection efficiency was still too low for sorting (Fig. 4B). RPE1 cells did have a higher population of viable cells compared to TruHD-Q43Q17 cells. During the first population of grouping by cell size and viability stain, 27.8% of cells were assessed as viable in RPE1s, compared to 3.9% in TruHD-Q43Q17 cells. To assess for GFP-positive cells, GFP-control vector was transfected with RPE1 cells to establish a threshold

fluorescence level. There is a sparse number of approximately 4 GFP-positive cells expressing donor construct in the range of 10^3 to 10^4 GFP (Fig. 4B). Like with TruHD-Q43Q17 cells, since we could not select for cells expressing GFP, we selected for GFP-negative cells and instead assessed growth from an individual cell population. The cells sorted for FACS are present in population P5 (Fig. 4B). Unlike TruHD-Q43Q17 cells, we were able to grow RPE1 cells from a single cell population. Transitioning to RPE1 cells alleviate many other practical concerns. The toxic effect in TruHD-Q43Q17 cells was resolved, and we did see a minor increase in GFP expression from the 45Q donor plasmid as seen by increase in events in FACS (Fig. 4A, Fig. 4B). Still, we grew concerned over the repeated low GFP-positive events across cell types, and again repeated quality control measures of plasmids. All CRISPR constructs were validated using gel electrophoresis and Sanger sequencing showing clean, high quality DNA samples.

We shifted our focus to analyzing editing efficiency rather than optimizing the expression of GFP expression. To assess the editing efficiency in RPE1 cells, we focused on optimizing the transfection of sgRNA/nCas9 plasmids without transfection of the 45Q donor plasmid. If the sgRNA/nCas9 plasmids are working efficiently, they should be producing indels at the target DNA. We transfected each sgRNA/nCas9 plasmid at 0.5-5µg in RPE1s over multiple trials using *Trans*IT-X2 and Lonza Nucleofector protocols to determine if we can optimize the

presence of indels (Table 1). We extracted the genomic DNA of transfected cells 1-5 days following transfection and performed Sanger sequencing. We used the ICE (Inference of CRISPR edits) software that identifies the presence of indels by comparing Sanger sequencing data of transfected cells against untransfected cells (125). Using the sequence of gRNA, this software uses an algorithm to align edited vs unedited chromatograms looking for any base differences around the target DNA (125). Across various experiments of nucleofection involving altered method of transfection, DNA concentration and timing of extraction, there is consistently between 0-5% indels in the transfected RPE1 cells. In response to the consistently low level of transfection efficiency in combination with the low editing efficiency, we felt it was necessary to revisit the plasmid design. We considered 4 different alternative solutions to overcome the low transfection efficiency: lentiviral transduction, new gRNA/Cas9 constructs, TALENs constructs, and Cas9/gRNA RNPs (Fig. 1).

3.5 The redesign of smaller CRISPR constructs

To combat the consistently low transfection efficiency, we constructed new, smaller CRISPR constructs (123,124). BPK1520 gRNA expression vector was used for both newly designed sgRNAs. BPK1520 was digested with BsmBI restriction enzyme. Oligos for sgRNA-a and sgRNA-b sequences designed by Xu et al,, 2017 were ordered and digested with BsmBI restriction enzyme. Digest

sgRNAs were separately cloned into digested BPK1520 for a total of 2, 2kb sgRNA vectors. Separately, a Cas9-nickase, NLS-nCas9-NLS-P2A-EGFP (nCas9-EGFP) carrying a GFP signal was ordered that totalled 8kb in size. This system relies on the expression of 45Q donor plasmid for HDR of the CAG tract. Both gRNA plasmids did not contain a selection marker, and as such we were not able to assess for their expression in transfection. We were able to assess for GFP expression of nCas9-EGFP. To assess the expression of Cas9 expression vector, RPE1 and HEK293 cells were transfected according to TransIT-X2 and lipofectamine 3000 protocols. High GFP expression was consistently observed at DNA concentrations of Cas9 expression vector at 3-5µg in RPE1 cells and HEK293 cells.

3.6 TALENs gene editing

A transition to TALENs may address the low editing efficiency that we have experienced in our CRISPR constructs. We acquired TALENs: DSPQ-HTT-TALEN1 ZEO (TALEN1) and DSPQ-HTT-TALEN 2 (TALEN2) designed by Ooi *et al.* to develop isogenic hiPSC models of HD (126). The use of this system still relies on the 45Q donor plasmid and 18Q donor plasmid constructed by Xu *et al.*, 2017, and as such is still a barrier when we consider the low expression of GFP in our system. We conducted transfection of RPE1 cells and HEK293 cells according to *Trans*IT-X2 protocols with TALEN1 and TALEN2 at 1-4µg and 45Q

donor plasmid at 3-5µg. Expression of GFP from 45Q donor plasmid was very low in RPE1 cells, but high in HEK293 cells. Transfections in HEK293 cells showed high levels of GFP expression of 45Q donor plasmid, consistent across all construct systems: original constructs sgRNA-a/nCas9, sgRNA-b/nCas9; newly designed sgRNA-a/BPK1520, sgRNA-b/BPK1520, nCas9-EGFP; TALEN1 and TALEN2. Since transfected RPE1 cells were considered inadequate for FACS, we attempted FACS in HEK293 cells. Transfected HEK293 cells with 45Q donor plasmid, sgRNA-a/nCas9, sgRNA-b/nCas9 had adequate cell populations positive for GFP, and were successfully sorted for GFP-positive, viable cells (Fig. 4C). Transfection efficiency in HEK293 cells was still very low, with only 63 events identified across a population of 4139 parent cells. This was not pursued as transformed cell lines are unsuitable for a stable cell line investigating HD dynamics.

3.7 Propagate TruHD-Q43Q17 cells with CAA interruption

I used the novel, PAM-less, SpRYc Cas9 enzyme to introduce a CAA interruption in TruHD-Q43Q17 cells (127). SpRYc was modified with a cytosine base editor to generate a CAG to CAA base modification in the fourth CAG repeat in the CAG repeat tract in TruHD-Q43Q17 cells. Successfully editing this base reduces the uninterrupted CAG repeat length to fewer than 40 CAG repeats while maintaining a polyglutamine expansion beyond 40 repeats.

I used the Amaxa 4D Lonza Nucleofector to transfect a sgRNA construct targeting the fourth CAG repeat with SpRYc-BE4Max. I extracted the genomic DNA for PCR and Sanger sequencing. Sequencing results were analyzed using Base Editing Evaluation Program to reveal a 35% efficiency in base editing of CAG to CAA (Fig. 5).

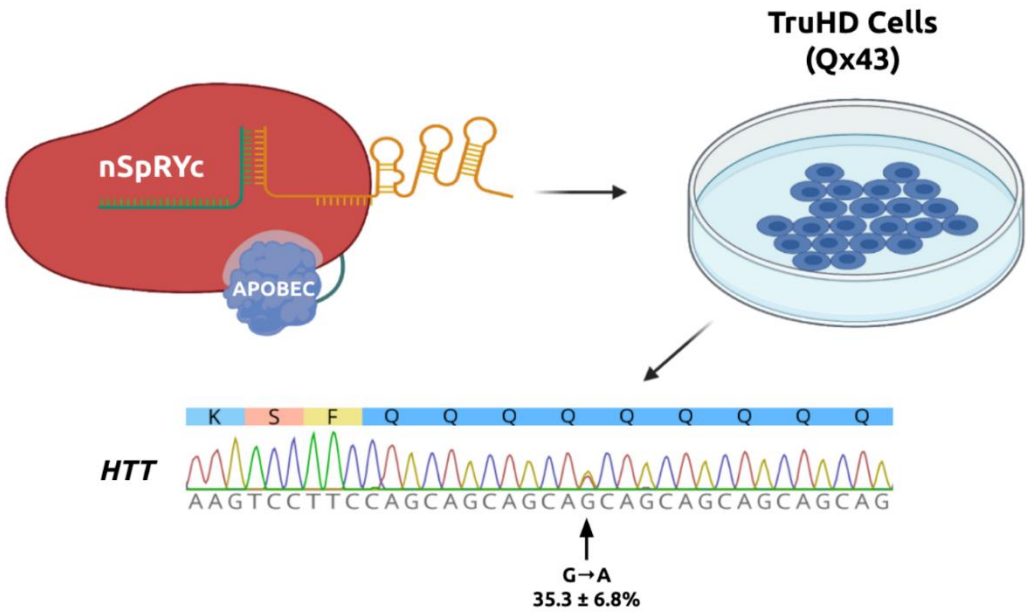


Figure 5 - SpRYc-BE4Max has 30% efficiency in conversion of fourth CAG repeat to CAA in TruHD-Q43Q17 cells

sgRNA and SpRYc enzyme with APOBEC base editor were transfected into TruHD-Q43Q17 cells in triplicate (n=3). Genomic DNA was extracted with expansion of the target region through PCR. Editing efficiency was calculated

using mean values from a Base Editing Evaluation Program that evaluates the level of intended modification in edited compared to unedited cells (115). Reproduced with permission from Koseki *et al.* unpublished.

To generate a stable cell line carrying this genetic mutation, we repeated transfection in TruHD-Q43Q17 cells and performed FACS using 7-AAD viability stain to individually sort for viable cells (Fig. 6). Cells were grouped according to cells stained with 7-AAD viability stain, cell size and complexity, and GFP-negative cells as this signal was not present in the CRISPR vector. The final grouping for to sorting to individual cells is represented by group P5 (Fig. 6). The 35% editing rate establishes a likely chance that one of the single cells will carry the CAA base edit. We intended to grow cells from the single cell population, extract genomic DNA and conduct Sanger sequencing to determine which cell populations carry the CAA edit. FACS was originally conducted with fresh growth media, and all isolated single cells died following sorting. To promote cell growth from a single colony we used conditioned media, the use of media from a fully confluent plate of TruHD-Q43Q17 cells. With the use of conditioned media, we were able to grow colonies from a single cell.

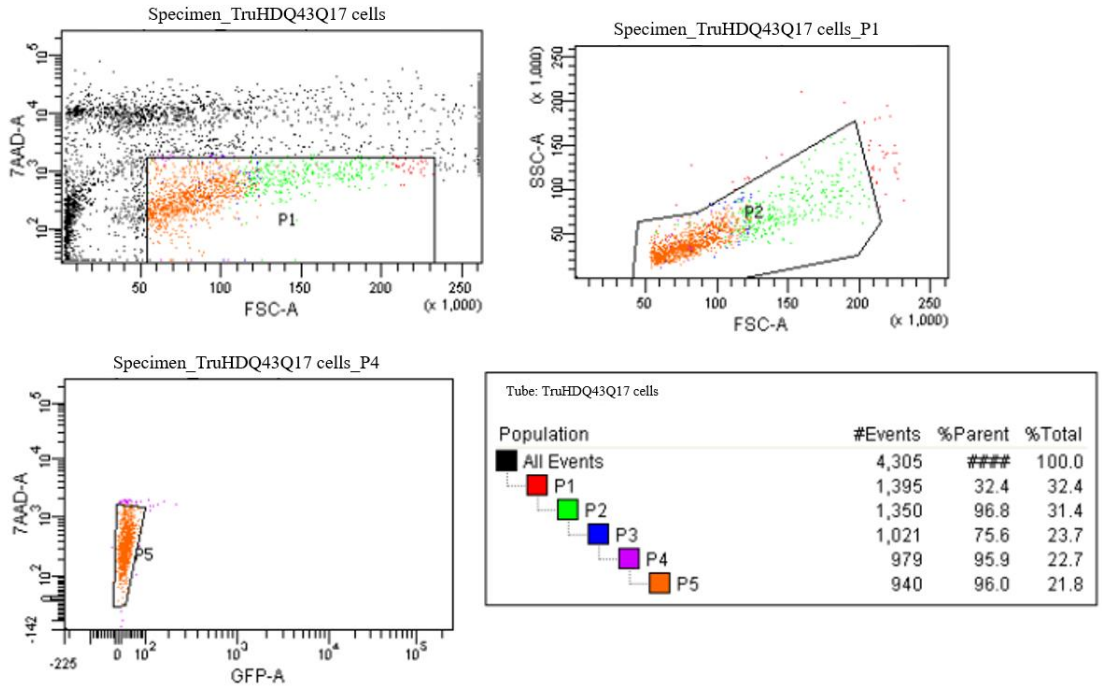


Figure 6- FACS sorting of living TruHD-Q43Q17 cells transfected with CRISPR base editor

TruHD-Q43Q17 cells were transfected with SpRYc-BE4Max enzyme and gRNA targeting a single base edit of CAG to CAA in the HTT gene. All cells expressing 7-AAD viability stain were sorted to P1, then grouped by size and complexity, and then further sorted to GFP-negative cells. P5 indicates GFP-negative, 7-AAD stained TruHD-Q43Q17 cells that were sorted to single cells.

3.8 Transfected exon 1 expression relative to endogenous huntingtin

To assess how increasing levels of transfected exon 1 compare to endogenous huntingtin levels, we performed a western blot analysis in HEK293 cells (Fig. 7A). We used increasing DNA concentrations of 250ng, 500ng,

1000ng, 2000ng of exon 1. Transfections were balanced to 2000ng with inactive DNA to ensure overall DNA concentration was not a contributing factor. To assess if there was a difference in non-expanded vs expanded exon 1, both healthy mCer-Q17Exon1-YFP, and diseased, Q138 Exon1-YFP were assessed in increasing concentrations. Additionally, an empty YFP backbone vector without exon 1 was transfected to act as a control. Protein extracts of cells were analyzed by western blot with anti-huntingtin specific to the N-terminal region of protein (Fig. 7A). mCer-Q17Exon1-YFP appears at a larger size than Q138 Exon1-YFP, despite the larger size of the expanded polyglutamine length due to having two fluorescent tags. Levels of exon 1 were quantified relative to endogenous huntingtin which acted as a loading control (Fig. 7B). Results indicate that in a typical transfection experiment of 1000ng, exon 1 has a density of 3-4 relative to endogenous huntingtin. This represents a problem in data that use transfected exon 1 to analyze the function of huntingtin. Results with this level of exon 1 do not represent the stochastic conditions of endogenous huntingtin. According to this data, to achieve stochastic conditions, less than 500ng of transfected exon 1 should be used.

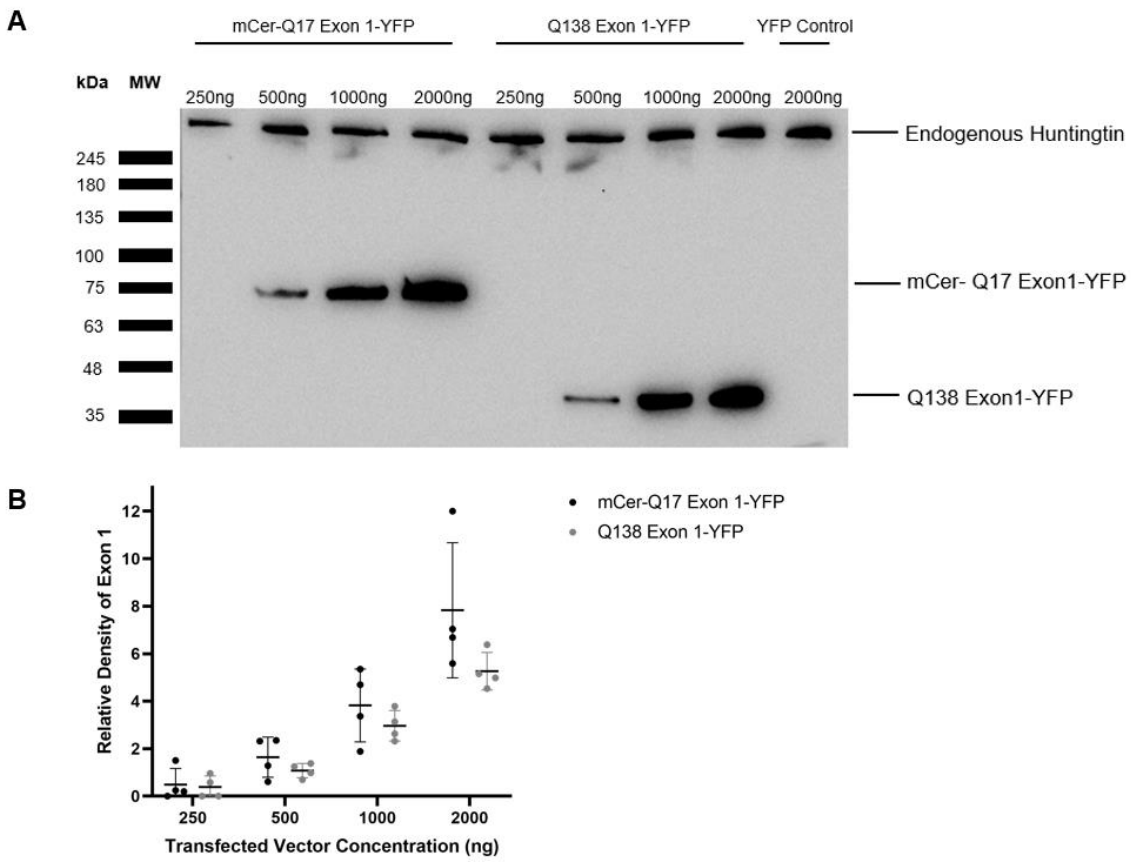


Figure 7 – Overexpression of transfected exon 1 relative to endogenous huntingtin

(A) Representative western blot with anti-huntingtin of transfected HEK293 cells with increasing concentration of indicated constructs. (B) Indicated exon 1 levels were quantified relative to endogenous huntingtin as loading control. The mean relative density of exon 1 was calculated with standard deviation depicted over 4 trials. N=4.

CHAPTER 4.0 DISCUSSION

4.1 Development of an isogenic cell line of HD

The development of a clinically relevant isogenic cell line would be an invaluable tool in understanding the role of the CAG expansion in HD. We attempted the development of an isogenic cell line in the clinically relevant, genome stable, TruHD cellular model of HD. We used a CRISPR system with paired sgRNAs and a Cas9-nickase to reduce off-target effects. The Cas9-nickase works in a paired system where both sgRNA/nCas9 pairs are necessary for a DSB to occur, where the presence of a single sgRNA/nCas9 complex only creates a nick. We planned to use FACS to sort cells for the expression of the 18Q donor and 45Q donor plasmids by GFP expression and conduct sequencing to ensure the intended edit occurred (Fig. 1). Our experiments originally focused on TruHD-Q43Q17 cells, as TruHD-Q21Q18 cells are more difficult to grow and may present a challenge after selection where cells are grown from sparse environments. We planned to complete the editing TruHD-Q43Q17 cells fully to overcome basic troubleshooting steps, before transitioning to the more sensitive TruHD-Q21Q18 cells. We felt it was important to conduct editing on both cell lines as a control experiment to ensure any observed effects are directly due to the CAG length, and not due to the specific genetics of the TruHD-Q43Q17 cells. Since these cell lines are derived from patients with different genetic backgrounds, we are not certain that changes to the CAG length of one cell type are uninfluenced by their other genetics. If we see similar results with changing CAG lengths in both cell types, we can be certain it is the direct consequence of

the CAG repeats. The 18Q donor plasmid serves as the donor in TruHD-Q43Q17 cells for its intended use to substitute the 43 CAG tract with 18 CAG repeats by HDR. In healthy TruHD-Q21Q18 cells, the 45Q donor plasmid serves as the substitution construct. Both donor plasmids are the largest CRISPR constructs in size, both approximately 12kb. They are followed closely, by sgRNA-a/nCas9 and sgRNA-b/nCas9 plasmids each at 8kb. The donor plasmids both carry a GFP selection marker; we used expression of fluorescence as an estimate for the transfection efficiency of the donor plasmid. We optimized the expression of GFP the 18Q donor plasmid in TruHD-Q43Q17 cells (Fig. 2) and compared relative levels of GFP expression to a control GFP vector (Fig.3). We found that compared to a control GFP vector, there were very few cells expressing GFP from the 18Q donor plasmid (Fig. 3). Our attempt at FACS for 18Q donor plasmid in TruHD-Q43Q17 cells show too few GFP-positive events for successful sorting (Fig. 4A). Our attempt at sorting also showed a high degree of unviable cells. To combat the cells lost during sample preparation prior to FACS, we added EDTA to flow buffer and aimed to sort cells during log phase. Both factors were used with the intention of reducing cell adhesion and promoting cell growth. Still, even with these considerations, there were not enough cells expressing GFP for successful sorting. We were unsure if this level of death was abnormal for TruHD-Q43Q17 cells. It may be that death was a result of general stress from transfection, specific toxicity from this set of CRISPR constructs, or over confluency due to the timing of sorting. Future research should further investigate

this phenotype by performing a sorting experiment with a control non-toxic construct and modifying the timing of sorting. While we did have a control GFP construct in this sorting experiment to establish a baseline threshold to identify GFP-positive cells, data was not available for cell viability. In this FACS attempt, we sorted for viable, GFP-negative cells to assess how TruHD-Q43Q17 cells would grow in a single cell environment. While cells were not able to grow from this environment, we were unsure if this was due to the toxicity of this construct, general transfection stress, or a result of the sensitive growth patterns of the TruHD-Q43Q17 cells.

Several alternative pathways were attempted to achieve expression of gene editing constructs (Fig.1). Drug resistance was pursued as an alternative selection marker to avoid the need for high cell density following transfection. After several experiments, we determined that puromycin resistance in the 18Q and 45Q donor plasmids was not additive to overcome the inherent resistance in cells immortalized with hTERT (Table 2). Puromycin was applied at 0.5-5 µg/ml to transfected TruHD-Q43Q17 cells with up to 7µg of donor plasmid every 48 for 2 weeks without any notable difference to control untransfected TruHD-Q43Q17 cells (Table 2). In response to puromycin as a non-viable selection method, we turned to substituting zeocin resistance as an alternative source of drug resistance. While we believed that this method would be more fruitful, I would

consider returning to puromycin resistance, as I believe we did not test for additive resistance extensively. Our main barrier has been a low transfection efficiency, this means even in a drug resistance selection method, there will be very few cells expressing resistance. It is possible transfected vs control cells appeared to have a similar cell death pattern, but perhaps if we had continued to apply puromycin, differences would have emerged. Puromycin resistance affects mammalian cells at concentrations between 0.5-10µg/ml, and thus I would also consider greatly increasing the levels of puromycin to see if there is eventually a toxicity difference between transfected and untransfected cells. While there may have been a difference in phenotype eventually, we still considered it beneficial to pursue another method of drug resistance. We conducted preliminary cloning of zeocin drug resistance into 18Q and 45Q donor plasmids. Cloning zeocin to substitute puromycin resistance allowed us to also remove the GFP tag that was no longer needed for selection. Cloning zeocin resistance has the added benefit of reducing the size of the plasmid to improve transfection efficiency and reduce toxicity (124). Future work should consider revisiting puromycin drug resistance and conclude preliminary cloning of zeocin drug resistance into the 18Q and 45Q donor plasmids.

The issue of low transfection efficiency in TruHD-Q43Q17 cells meant that we were unable to successfully isolate clones expressing CRISPR constructs (Fig. 4A). We transitioned to the use of hTERT immortalized RPE1 cells derived

from a healthy patient harboring a CAG length under 40 repeats (122). These cells were chosen because they retained the value of hTERT immortalization in TruHD cells, but without the relatively low transfection efficiency associated with fibroblast cells. Additionally, RPE1 cells are more resistant to growth from single cell populations, and thus are practically easier to use for selection of CRISPR constructs. We prioritized experiments conducted in RPE1 cells as we can troubleshoot and generate edited cells faster while still determining the isolated effect of the CAG expansion.

Several transfection experiments were conducted in RPE1 cells with limited success (Table 1). When FACS was attempted in RPE1 cells, we saw small improvements compared to TruHD-Q43Q17 cells (Fig. 4). We saw an increase in the first grouping of viability, with 27.8% viable in RPE1 cells compared to 3.9% in TruHD-Q43Q17 cells. We also saw a mild increase in the number of GFP-positive events with an estimate of 4 GFP-positive cells compared to 1 event seen in TruHD-Q43Q17 cells. While these were mild improvements, they were not enough for successful sorting of GFP-positive cells. As with TruHD-Q43Q17 cells, we sorted for living, GFP-negative cells to determine how RPE1 cells grow from individual populations. Unlike the TruHD-Q43Q17 cells, RPE1 cells were able to successfully grow from individual cells when left on their own. This gave us confidence, that if we were able to improve

the transfection efficiency, we would be able to grow cells from individual populations to produce an identical colony. The transition to RPE1 cells also brought several limitations. The development of isogenic cells in TruHD cells would allow us to study both the expansion and reduction of the CAG tract in different genetic backgrounds to confirm that any changes are directly related to CAG length. Since TruHD cells are derived from both male and female individuals, identifying common changes related to the length of the CAG expansion would eliminate sex-related differences and other background genes that may influence phenotypes (6,11). RPE1 cells were derived from a single healthy 1-year old female meaning isogenic RPE1 cells would not allow for the same level of analysis as in TruHD cells. Despite these limitations, an isogenic RPE1 cell line differing only in CAG lengths would still allow us to compare pre and post edited cells for differences related to CAG expansion length. After encountering consistently low GFP expression, we considered the possibility that this may not be a valid method of assessing the occurrence of editing events. This system only assesses the presence of the 45Q donor plasmid and does not consider the expression of the sgRNA/nCas9 vectors. The assessment of sgRNA/nCas9 vectors is limited by the lack of any selectable marker in the vectors. Additionally, GFP expression may be quickly diminishing, and thus while we may not be seeing GFP expression, editing events may still be occurring. We calculated editing efficiency using the presence of indels to be 0-5% using the Synthego ICE tool. After troubleshooting this process, we determined that

presence of indels may not necessarily reflect the editing efficiency of our system. Our system is composed of a Cas9-nickase meaning, any generated nicks at the targeted site may be repaired more efficiently than with a conventional Cas9 enzyme. Without this as a guide for editing efficiency, we returned to GFP expression as a substitute for assessing editing efficiency.

While we attempted various strategies to achieve editing, the most consistent hurdle we were not able to overcome was the low transfection efficiency of the 18Q and 45Q donor plasmids. We were not able to employ protocol differences to improve the transfection efficiency of the constructs we had prepared in the desired cell types. After confirming by Sanger sequencing that there was not an issue with the constructs themselves, we felt it was necessary to consider new constructs.

We pursued a re-design of our current CRISPR constructs. We successfully cloned sgRNA-a and sgRNA-b into separate BPK1520 empty gRNA vector and ordered a separate nCas9-EGFP. In doing so, we replaced the 2 vectors: sgRNA-a/nCas9 and sgRNA-b/nCas9 with 3 vectors: sgRNA-a/BPK1520, sgRNA-b/BPK1520 and nCas9-EGFP. While this increases the number of plasmids, we are reducing the redundancy of each gRNA plasmid carrying a copy of Cas9 as in the current method. In the original system, we were only able to assess GFP expression from the 18Q or 45Q donor plasmids. We

hoped that the large size of the donor plasmids meant they would be the only problematic construct for expression. This new system allows us to assess GFP expression from the nCas9-EGFP plasmid. This allows slightly more certainty that each construct is being expressed. In the original constructs, the donor plasmid was 12kb, and each gRNA/Cas9 vector was 8kb. In this system, each sgRNA/BPK1520 vector is 2kb, and the nCas9-EGFP plasmid is 8kb. While this new design still requires the expression of 45Q donor plasmid, we can optimize the transfection conditions based on GFP expression from the nCas9-EGFP and 45Q donor plasmids separately. We successfully transfected nCas9-EGFP into RPE1 cells. There was no notable toxicity associated with any newly designed plasmids. Still, without completion of zeocin drug resistance cloning or sufficient GFP expression from the 45Q donor plasmid in RPE1 cells, we were not able to move towards selecting for clones with the desired edit. Future experiments should conclude zeocin resistance cloning into the 45Q donor plasmid. This system would enable FACS of the Cas9 expression vector and zeocin drug resistance selection for 45Q donor plasmid.

In moving towards a new construct system, we also pursued TALEN gene editing with 45Q donor plasmid in RPE1 cells. One advantage of the TALEN constructs over both systems of CRISPR constructs, is each TALEN vector carries selectable drug resistance. TALEN1 contains zeocin resistance, and TALEN2 contains blasticidin resistance. This system ensures that each

component can be optimized to achieve highest editing efficiency. If 45Q donor plasmid is successfully modified to contain zeocin resistance to replace GFP selection, it would be necessary to optimize these transfections separately before performing a combined transfection. TALEN constructs were transfected at 1-4µg each and 45Q donor plasmid at 3-5µg. Increasing DNA concentrations of TALEN vectors did not confer a notable toxic effect.

In these alternative approaches, are greatest barrier remained the low transfection efficiency of the 18Q and 45Q donor plasmids which were still necessary to achieve CAG substitution. To increase transfection efficiency, we considered the use of lentiviral transduction as a delivery method (106,107). This system was not used for implementation as most systems of lentiviral transduction involve stable expression of Cas9, which would increase off-target effects and would not be suitable for a stable cell line (109). Additionally, our constructs were not compatible with lentiviral vectors and would require a new design. Lentiviral transduction could be conducted on gRNA and Cas9 vectors with separate delivery of 18Q or 45Q donor plasmids, but this would still not overcome the issue of low expression of donor plasmid.

Another alternative solution to overcoming the consistently low transfection efficiency was an RNP delivery method (128). An RNP method would involve direct nucleofection of a prepared complex of Cas9 protein with gRNA directly

into RPE1 or TruHD-Q43Q17 cells. This method would improve transfection efficiency of the gRNA/Cas9 complex but would still require expression of the 45Q donor plasmid or 18Q donor plasmid to substitute CAG length. This introduced a significant limitation, as transfection efficiency has mainly been limited by donor plasmid expression. To increase the efficiency of 18Q donor plasmid and 45Q donor plasmid in an RNP system, future work should consider the design of single stranded donor DNA. The use of single stranded DNA would involve redesigning the current homology arms in the 18Q and 45Q donor plasmids, but is shown to have increased editing efficiency (129).

To ensure the lack of GFP expression was not due to an issue with the constructs, control experiments were conducted in HEK293 cells to compare to transfection efficiency RPE1 cells. RPE1 cells retained consistently low GFP expression, with HEK 293 cells showing GFP expression from 45Q donor plasmid across all system constructs: original constructs sgRNA-a/nCas9, sgRNA-b/nCas9; newly designed sgRNA-a/BPK1520, sgRNA-b/BPK1520, nCas9-EGFP; TALEN1 and TALEN 2. HEK293 cells were a suitable cell type for a control experiment because of their high transfection efficiency (62). HEK293 cells were assessed for FACS to determine if there was an adequate number of GFP-positive cells (Fig. 4C). While HEK293 cells expressed GFP at levels suitable for FACS, HEK293 cells are not a valid system to study HD dynamics (Fig. 4C). As a transformed cell line, their p53 function is interrupted and their

genome is unstable (54–56). Huntingtin's involvement with p53 in the stress response and DNA damage would limit the usefulness of any resulting findings from HEK293 cells (57,58). The FACS data of HEK293 cells shows a high degree of cell death, like FACS results in RPE1 and TruHD-Q43Q17 cells. In HEK293 cells, 43.9% of cells were identified as 7-AAD viability stained in the first grouping, compared to 27.8% in RPE1 and 3.9% in TruHD-Q43Q17 cells. In HEK293 cells, only 63 events were considered GFP positive, out of a parent cell population of 4139 events, with a transfection efficiency of 1.5%. The results of this data confirm the integrity of the 45Q donor plasmid, but still present a very challenging transfection efficiency in a highly expressive cell type. Comparing the results of FACS across the three cell types provided some perspective for next steps. One differing factor between the unsuccessful sorting events, and successful sorting samples is the total population of events. 10 000 total events were investigated in HEK293 cells, 4335 events were investigated in RPE1 cells, and 5924 events in TruHD-Q43Q17 cells. While this population of cells includes unviable cells, one method of overcoming the high cell death is simply using a larger starting number. It may be that we are not investigating as many events in TruHD-Q43Q17 and RPE1 cells, and by increasing the population of cells by a large factor we will see enough GFP-positive cells for sorting. Future work should consider increasing the pool of cells analyzed for FACS to identify if this provides a large enough population of GFP-positive cells. Another method of increasing the population of GFP-positive cells, is identifying the source of cell death. Future

work should address if this is an issue with the constructs, timing of sorting or general stress from transfection. The amount of cell death seen in the TruHD-Q43Q17 cells is much higher than that of the RPE1 and HEK293 cell data. This may indicate a specific toxicity of this construct to this cell type, but the experiment would need to be repeated to provide more confidence. Future researchers can transfect a control non-toxic GFP vector and perform sorting to obtain viability data. It is also important to consider that our results use a highly expressive control GFP vector in establishing a baseline level for events to be considered GFP-positive. Future researchers may consider using a control construct with low transfection efficiency to represent a more comparable GFP level to the CRISPR constructs. It may be that our GFP vector is expressing such high levels of GFP compared to any GFP-events from the CRISPR constructs, that some GFP-events from the CRISPR constructs are being missed. Once the issue of low transfection efficiency of the 18Q donor plasmid and 45Q donor plasmid are overcome, the development of an isogenic cell line of HD would be an incredible resource to investigating the direct impact of CAG length.

4.2 CAA base editing interruption in TruHD-Q43Q17 cells

Currently, CRISPR is limited to targeting DNA regions that satisfy the specific PAM of the Cas enzyme. Koseki *et al.* have developed a novel Cas9 enzyme termed, SpRYc that is not limited by a PAM requirement (127). By modifying SpRYc to form a base editor, SpRYc-BE4Max, we were able to test the

efficiency of this enzyme to introduce base modifications. SpRYc-BE4Max was able to edit TruHD-Q43Q17 cells at approximately 35% efficiency to introduce a CAG to CAA modification (Fig. 4). The ability to target any DNA region effectively greatly expands the value of CRISPR in biological systems.

The CAG to CAA base edit is a key modification that would be beneficial for future investigations. The generation of a stable cell line that carries a CAA interruption in the polyglutamine expansion is a valuable phenotype for investigating variation in the age of symptom onset. Past studies have shown it is the length of the CAG expansion, rather than the length of the polyglutamine tract that contributes to disease development (14). Additionally, CAA interruptions have shown to be protective in delaying disease development (14,15). This base modification represents a shift from a disease cell line to a healthy cell line. Stable propagation of these cells would allow future work into differences between cell lines to better uncover the mechanism of protection.

We were able to conduct FACS to sort for TruHD-Q43Q17 cells stained with 7-AAD viability dye for future work to establish a stable cell line. Following the first attempt of sorting TruHD-Q43Q17 cells, all single cells in a 96-well plate were monitored for growth for 3 weeks. All TruHD-Q43Q17 cells appeared to die at the single-cell stage. We were not surprised by this result, as cells sorting during the attempt in TruHD-Q43Q17 cells with 18Q donor plasmid, sg-RNA-

a/nCas9, sg-RNA-b/nCas9 also did not grow. With this result, we knew that the resistance to grow from a single cell would be a consistent issue that was likely not specific to the transfected construct. We conducted a second attempt of sorting TruHD-Q43Q17 cells and replaced fresh growth media with conditioned media (Fig. 6). Conditioned media, or media from a confluent plate, provides additional growth factors from healthy cells to promote proliferation from a single cell (130). Sorting TruHD-Q43Q17 cells into conditioned media was successful in promoting growth of cells. The use of conditioned media should be used in future experiments for sorting of TruHD-Q43Q17 cells to establish isogenic cells. When this FACS result is compared to the attempts of CRISPR constructs in TruHD-Q43Q17, RPE1 and HEK293 cells there is a continued issue of high cell death. In this set of FACS data, the first grouping of viable cells represents 32.4% of all events. This is a much higher percentage of viable cells compared to the TruHD-Q43Q17 cell data of the previous CRISPR constructs that had the first grouping at 3.9% of all events. This supports the idea of a specific toxicity of these constructs in TruHD-Q43Q17 cells but would require further investigation to be conclusive. As previously discussed, it is important to use a control construct that was previously validated to not possess abnormal toxicity. Using this method of sorting a non-toxic construct we can establish a baseline level of cell death to interpret if these constructs are toxic.

4.3 Exon 1 overexpression relative to endogenous huntingtin

The results of the western blot assessing transfected exon 1 levels reveal that experiments using over 500ng of transfected exon 1 do not reflect stochastic conditions of the cell (Fig. 7). We assessed how increasing levels of exon 1 change relative to endogenous huntingtin in both unexpanded (mCer-Q17 Exon 1-YFP) and expanded (Q138 Exon 1 YFP) exon 1 DNA. We found a similar phenotype where mCer-Q17 Exon 1-YFP at 1000ng represents a 3.8-fold mean increase, and Q138 Exon 1 YFP represents a 3.0-fold increase. These results show that there is a similar level of overexpressed exon 1 regardless of the length of the polyQ expansion. Since Q17 exon 1 is flanked by two fluorescent tags, and Q138 contains only one fluorescent tag at its C-terminal domain, we can also control for a changing vector. Since the patterns of exon 1 overexpression are similar between both sets of exon 1, and are not apparent in the empty vector control, we can feel confident that our results are not confounded by the vector backbone.

As previously discussed, we cannot trust results that do not represent the stochastic conditions of a cell. While results from transfected exon 1 studies are already problematic for their overly segmented approach that neglects the full function of huntingtin, the results of this analysis show they also do not reflect appropriate protein levels in the cell. We chose this range of DNA concentration to represent a typical concentration for transfection of 1-5μg depending on

manufacturer's instructions (131–133). Here we show that careful consideration is necessary when evaluating findings based on transfected exon 1 concentration at these levels. Future work should consider a very conservative approach to DNA concentration at transfection to ensure findings are reflective of the endogenous environment.

CHAPTER 5.0 CONCLUSIONS AND FUTURE DIRECTIONS

In this study we explore the importance of developing a clinically relevant model system of disease. Successful creation of isogenic cells from TruHD or RPE1 cells would improve the knowledge gained about HD by improving the accuracy of our model system and isolating the role of the CAG length in future experiments. These isogenic cells would be derived from stable human cells, with clinically representative alleles, differing only by CAG expansion length. Once these cells have been established, future research can be conducted on pre and post edited cells to determine the effect of a changing CAG expansion.

Here we describe the various troubleshooting attempts to establish a clinically relevant isogenic cell line of HD. This should be used by future researchers as a baseline framework to accompany their own optimization experiments in developing isogenic cells. We discuss the significant barrier of low transfection efficiency in our current system and propose the development of a

new set of constructs according to an RNP delivery method. RNPs largely overcome the low editing efficiency and transfection efficiency of plasmid-based approaches (128). RNPs involve a direct delivery of Cas9 protein and gRNA as a complex. To further increase the low transfection efficiency of the donor plasmid, a single stranded DNA donor with shorter homology arms should be designed (129,134). While we largely focused on the use of FACS to select for expression of the donor plasmid, it may have been more beneficial to focus our efforts on drug resistance. With concerns of low confluency and low transfection efficiency, it is easier to select for positive cells from a drug selection method, rather than through FACS.

Once an isogenic cell line is established in TruHD or RPE1 cells, key experiments will be analyzing differences in DNA damage and repair in pre and post edited cells. Since this isogenic model uses a genomically stable, clinically relevant cell line, an isogenic line will allow us to study huntingtin's role in the stress response while identifying the effect of the CAG expansion.

Here, we use the SpRYc-BE4Max base editing enzyme to introduce a CAG to CAA base interruption in TruHD-Q43Q17 cells. We discuss how to improve growth conditions of TruHD-Q43Q17 cells to conduct FACS in future work. The development of a stable cell line carrying this interruption will allow us

to investigate the mechanism behind the protective actions of CAA in delaying disease development in HD.

Our work with transfected exon 1 levels in HEK293 cells identifies a problem in many models using this system to represent HD. We show that at typical concentrations of DNA at transfection, exon 1 levels are greatly overexpressed relative to what is seen in an endogenous system. These models do not reflect the tightly controlled levels of huntingtin present in clinical HD. By departing from a stoichiometric model system of huntingtin, we cannot trust results from systems relying on exon 1 overexpression in transformed cells.

Works Cited

1. Vonsattel JP, Myers RH, Stevens TJ, Ferrante RJ, Bird ED, Richardson EP. Neuropathological Classification of Huntington's Disease. J Neuropathol Exp Neurol. 1985 Nov 1;44(6):559–77.

2. Roos RA. Huntington's disease: a clinical review. Orphanet J Rare Dis. 2010 Dec 20;5:40.

3. Bettencourt C, Hensman-Moss D, Flower M, Wiethoff S, Brice A, Goizet C, *et al.* DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases: DNA Repair Pathways Modify polyQ Disease Onset. Ann Neurol. 2016 Jun;79(6):983–90.

4. Douglas I, Evans S, Rawlins MD, Smeeth L, Tabrizi SJ, Wexler NS. Juvenile Huntington's disease: a population-based study using the General Practice Research Database. BMJ Open. 2013 Jan 1;3(4):e002085.

5. Bates G, Harper PS, Jones L. Huntington's Disease. Oxford University Press; 2002. 574 p.

6. Hentosh S, Zhu L, Patino J, Furr JW, Rocha NP, Furr Stimming E. Sex Differences in Huntington's Disease: Evaluating the Enroll-HD Database. Mov Disord Clin Pract. 2021;8(3):420–6.

7. Bruyn G. Huntington's chorea: historical, clinical and laboratory synopsis. Elsevier Amsterdam; 1968. 298–378 p. (Handbook of Clinical Neurology; vol. 6).

8. MacDonald ME, Ambrose CM, Duyao MP, Myers RH, Lin C, Srinidhi L, *et al.* A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell. 1993 Mar 26;72(6):971–83.

9. Myers RH. Huntington's disease genetics. NeuroRX. 2004 Apr 1;1(2):255–62.

10. Lee JM, Ramos EM, Lee JJ, Gillis T, Mysore JS. CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashionAuthor Response: Neurology. 2012 Feb 8;78(10):952–3.

11. Genetic Modifiers of Huntington's Disease Consortium. Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. Cell. 2015 Jul 30;162(3):516–26.

12. Rosenblatt A, Liang KY, Zhou H, Abbott MH, Gourley LM, Margolis RL, *et al.* The association of CAG repeat length with clinical progression in Huntington disease. Neurology. 2006 Apr 11;66(7):1016–20.

13. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium, Lee JM, Correia K, Loupe J, Kim KH, Barker D, *et al.* Huntington's disease onset is determined by length of uninterrupted CAG, not encoded polyglutamine, and is modified by DNA maintenance mechanisms [Internet]. Genetics; 2019 Jan [cited 2021 Aug 26]. Available from: http://biorxiv.org/lookup/doi/10.1101/529768

14. Wright GEB, Collins JA, Kay C, McDonald C, Dolzhenko E, Xia Q, *et al.* Length of uninterrupted CAG repeats, independent of polyglutamine size, results in increased somatic instability and hastened age of onset in Huntington disease [Internet]. Genetics; 2019 Jan [cited 2021 Aug 26]. Available from: http://biorxiv.org/lookup/doi/10.1101/533414

15. Wright GEB, Black HF, Collins JA, Gall-Duncan T, Caron NS, Pearson CE, *et al.* Interrupting sequence variants and age of onset in Huntington's disease: clinical implications and emerging therapies. Lancet Neurol. 2020 Nov 1;19(11):930–9.

16. Ciosi M, Maxwell A, Cumming SA, Hensman Moss DJ, Alshammari AM, Flower MD, *et al.* A genetic association study of glutamine-encoding DNA sequence structures, somatic CAG expansion, and DNA repair gene variants, with Huntington disease clinical outcomes. EBioMedicine. 2019 Oct 1;48:568–80.

17. Dragileva E, Hendricks A, Teed A, Gillis T, Lopez ET, Friedberg EC, *et al.* Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. Neurobiol Dis. 2009 Jan 1;33(1):37–47.

18. DiGiovanni LF, Mocle AJ, Xia J, Truant R. Huntingtin N17 domain is a reactive oxygen species sensor regulating huntingtin phosphorylation and localization. Hum Mol Genet. 2016 Sep 15;25(18):3937–45.

19. Maiuri T, Mocle AJ, Hung CL, Xia J, van Roon-Mom WMC, Truant R. Huntingtin is a scaffolding protein in the ATM oxidative DNA damage response complex. Hum Mol Genet. 2017;ddw395.

20. Li S, Li XJ. Multiple pathways contribute to the pathogenesis of Huntington disease. Mol Neurodegener. 2006 Dec 16;1(1):19.

21. MacDonald ME. Huntingtin: Alive and Well and Working in Middle Management. Sci STKE. 2003 Nov 4;2003(207):pe48–pe48.

22. Nath S, Munsie LN, Truant R. A huntingtin-mediated fast stress response halting endosomal trafficking is defective in Huntington's disease. Hum Mol Genet. 2015 Jan 15;24(2):450–62.

23. Andrade MA, Petosa C, O'Donoghue SI, Müller CW, Bork P. Comparison of ARM and HEAT protein repeats11Edited by P. E. Wright. J Mol Biol. 2001 May 25;309(1):1–18.

24. Qin ZH, Wang Y, Sapp E, Cuiffo B, Wanker E, Hayden MR, *et al.* Huntingtin Bodies Sequester Vesicle-Associated Proteins by a Polyproline-Dependent Interaction. J Neurosci. 2004 Jan 7;24(1):269–81.

25. Xia J, Lee DH, Taylor J, Vandelft M, Truant R. Huntingtin contains a highly conserved nuclear export signal. Hum Mol Genet. 2003 Jun 15;12(12):1393–403.

26. Desmond CR, Atwal RS, Xia J, Truant R. Identification of a Karyopherin β1/β2 Proline-Tyrosine Nuclear Localization Signal in Huntingtin Protein. J Biol Chem. 2012 Nov 16;287(47):39626–33.

27. Rockabrand E, Slepko N, Pantalone A, Nukala VN, Kazantsev A, Marsh JL, *et al.* The first 17 amino acids of Huntingtin modulate its sub-cellular localization, aggregation and effects on calcium homeostasis. Hum Mol Genet. 2007 Jan 1;16(1):61–77.

28. Atwal RS, Xia J, Pinchev D, Taylor J, Epand RM, Truant R. Huntingtin has a membrane association signal that can modulate huntingtin aggregation, nuclear entry and toxicity. Hum Mol Genet. 2007 Nov 1;16(21):2600–15.

29. Atwal RS, Desmond CR, Caron N, Maiuri T, Xia J, Sipione S, *et al.* Kinase inhibitors modulate huntingtin cell localization and toxicity. Nat Chem Biol. 2011 Jul;7(7):453–60.

30. Maynard S, Schurman SH, Harboe C, de Souza-Pinto NC, Bohr VA. Base excision repair of oxidative DNA damage and association with cancer and aging. Carcinogenesis. 2009 Jan;30(1):2–10.

31. Harding RJ, Deme JC, Hevler JF, Tamara S, Lemak A, Cantle JP, *et al.* Huntingtin structure is orchestrated by HAP40 and shows a polyglutamine expansion-specific interaction with exon 1. Commun Biol. 2021 Dec 8;4(1):1–16.

32. Bravo-Arredondo JM, Kegulian NC, Schmidt T, Pandey NK, Situ AJ, Ulmer TS, *et al.* The folding equilibrium of huntingtin exon 1 monomer depends on its polyglutamine tract. J Biol Chem. 2018 Dec 21;293(51):19613–23.

33. Newcombe EA, Ruff KM, Sethi A, Ormsby AR, Ramdzan YM, Fox A, *et al.* Tadpole-like Conformations of Huntingtin Exon 1 Are Characterized by Conformational Heterogeneity that Persists regardless of Polyglutamine Length. J Mol Biol. 2018 May 11;430(10):1442–58.

34. Wheeler VC, White JK, Gutekunst CA, Vrbanac V, Weaver M, Li XJ, *et al.* Long glutamine tracts cause nuclear localization of a novel form of huntingtin in medium spiny striatal neurons in HdhQ92 and HdhQ111 knock-in mice. Hum Mol Genet. 2000 Mar 1;9(4):503–13.

35. White JK, Auerbach W, Duyao MP, Vonsattel JP, Gusella JF, Joyner AL, *et al.* Huntingtin is required for neurogenesis and is not impaired by the Huntington's disease CAG expansion. Nat Genet. 1997 Dec;17(4):404–10.

36. Mangiarini L, Sathasivam K, Seller M, Cozens B, Harper A, Hetherington C, *et al.* Exon 1 of the HD Gene with an Expanded CAG Repeat Is Sufficient to Cause a Progressive Neurological Phenotype in Transgenic Mice. Cell. 1996 Nov 1;87(3):493–506.

37. Schilling G, Becher MW, Sharp AH, Jinnah HA, Duan K, Kotzuk JA, *et al.* Intranuclear Inclusions and Neuritic Aggregates in Transgenic Mice Expressing a Mutant N-Terminal Fragment of Huntingtin. Hum Mol Genet. 1999 Mar 1;8(3):397–407.

38. Yang H, Yang S, Jing L, Huang L, Chen L, Zhao X, *et al.* Truncation of mutant huntingtin in knock-in mice demonstrates exon1 huntingtin is a key pathogenic form. Nat Commun. 2020 May 22;11(1):2582.

39. Crook ZR, Housman D. Huntington's Disease: Can Mice Lead the Way to Treatment? Neuron. 2011 Feb 10;69(3):423–35.

40. Carter RJ, Lione LA, Humby T, Mangiarini L, Mahal A, Bates GP, *et al.* Characterization of Progressive Motor Deficits in Mice Transgenic for the Human Huntington's Disease Mutation. J Neurosci. 1999 Apr 15;19(8):3248–57.

41. Hurlbert MS, Zhou W, Wasmeier C, Kaddis FG, Hutton JC, Freed CR. Mice transgenic for an expanded CAG repeat in the Huntington's disease gene develop diabetes. Diabetes. 1999 Mar 1;48(3):649–51.

42. Meade CA, Deng YP, Fusco FR, Del Mar N, Hersch S, Goldowitz D, *et al.* Cellular localization and development of neuronal intranuclear inclusions in striatal and cortical neurons in R6/2 transgenic mice. J Comp Neurol. 2002;449(3):241–69.

43. Ribchester RR, Thomson D, Wood NI, Hinks T, Gillingwater TH, Wishart TM, *et al.* Progressive abnormalities in skeletal muscle and neuromuscular junctions of transgenic mice expressing the Huntington's disease mutation. Eur J Neurosci. 2004;20(11):3092–114.

44. Mihm MJ, Amann DM, Schanbacher BL, Altschuld RA, Bauer JA, Hoyt KR. Cardiac dysfunction in the R6/2 mouse model of Huntington's disease. Neurobiol Dis. 2007 Feb 1;25(2):297–308.

45. Shenoy SA, Zheng S, Liu W, Dai Y, Liu Y, Hou Z, *et al.* A novel and accurate full-length HTT mouse model for Huntington's disease. eLife. 2022 Jan 13;11:e70217.

46. Pouladi MA, Morton AJ, Hayden MR. Choosing an animal model for the study of Huntington's disease. Nat Rev Neurosci. 2013 Oct;14(10):708–21.

47. Djoussé L, Knowlton B, Cupples LA, Marder K, Shoulson I, Myers RH. Weight loss in early stage of Huntington's disease. Neurology. 2002 Nov 12;59(9):1325–30.

48. Gray M, Shirasaki DI, Cepeda C, André VM, Wilburn B, Lu XH, *et al.* Full-Length Human Mutant Huntingtin with a Stable Polyglutamine Repeat Can Elicit Progressive and Selective Neuropathogenesis in BACHD Mice. J Neurosci. 2008 Jun 11;28(24):6182–95.

49. Van Raamsdonk JM, Gibson WT, Pearson J, Murphy Z, Lu G, Leavitt BR, *et al.* Body weight is modulated by levels of full-length Huntingtin. Hum Mol Genet. 2006 May 1;15(9):1513–23.

50. Yan S, Tu Z, Liu Z, Fan N, Yang H, Yang S, *et al.* A Huntingtin Knockin Pig Model Recapitulates Features of Selective Neurodegeneration in Huntington's Disease. Cell. 2018 May 3;173(4):989-1002.e13.

51. Loh DH, Kudo T, Truong D, Wu Y, Colwell CS. The Q175 Mouse Model of Huntington's Disease Shows Gene Dosage- and Age-Related Decline in Circadian Rhythms of Activity and Sleep. PLOS ONE. 2013 Jul 30;8(7):e69993.

52. Levine MS, Klapstein GJ, Koppel A, Gruen E, Cepeda C, Vargas ME, *et al.* Enhanced sensitivity to N-methyl-D-aspartate receptor activation in

transgenic and knockin mouse models of Huntington's disease. J Neurosci Res. 1999;58(4):515–32.

53. Menalled LB, Sison JD, Dragatsis I, Zeitlin S, Chesselet MF. Time course of early motor and neuropathological anomalies in a knock-in mouse model of Huntington's disease with 140 CAG repeats. J Comp Neurol. 2003;465(1):11–26.

54. Trettel F, Rigamonti D, Hilditch-Maguire P, Wheeler VC, Sharp AH, Persichetti F, *et al.* Dominant phenotypes produced by the HD mutation in STHdhQ111 striatal cells. Hum Mol Genet. 2000 Nov 22;9(19):2799–809.

55. Mittelman D, Wilson JH. The fractured genome of HeLa cells. Genome Biol. 2013 Apr 17;14(4):111.

56. Landry JJM, Pyl PT, Rausch T, Zichner T, Tekkedil MM, Stütz AM, *et al.* The Genomic and Transcriptomic Landscape of a HeLa Cell Line. G3 GenesGenomesGenetics. 2013 Aug 1;3(8):1213–24.

57. Feng Z, Jin S, Zupnick A, Hoh J, de Stanchina E, Lowe S, *et al.* p53 tumor suppressor protein regulates the levels of huntingtin gene expression. Oncogene. 2006 Jan;25(1):1–7.

58. Levine AJ. p53, the Cellular Gatekeeper for Growth and Division. Cell. 1997 Feb 7;88(3):323–31.

59. Jin S, Levine AJ. The p53 functional circuit. J Cell Sci. 2001 Dec 1;114(23):4139–40.

60. Vogelstein B, Lane D, Levine AJ. Surfing the p53 network. Nature. 2000 Nov;408(6810):307–10.

61. Thomas P, Smart TG. HEK293 cell line: A vehicle for the expression of recombinant proteins. J Pharmacol Toxicol Methods. 2005 May 1;51(3):187–200.

62. Wurm FM. Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol. 2004 Nov;22(11):1393–8.

63. Backliwal G, Hildinger M, Chenuet S, Wulhfard S, De Jesus M, Wurm FM. Rational vector design and multi-pathway modulation of HEK 293E cells yield recombinant antibody titers exceeding 1 g/l by transient transfection under serum-free conditions. Nucleic Acids Res. 2008 Sep 1;36(15):e96.

64. Pierzynowska K, Gaffke L, Hać A, Mantej J, Niedziałek N, Brokowska J, *et al.* Correction of Huntington's Disease Phenotype by Genistein-Induced Autophagy in the Cellular Model. NeuroMolecular Med. 2018 Mar 1;20(1):112–23.

65. TBK1 phosphorylates mutant Huntingtin and suppresses its aggregation and toxicity in Huntington's disease models. EMBO J. 2020 Sep;39(17):e104671.

66. Sameni S, Syed A, Marsh JL, Digman MA. The phasor-FLIM fingerprints reveal shifts from OXPHOS to enhanced glycolysis in Huntington Disease. Sci Rep. 2016 Oct 7;6(1):34755.

67. Toczek M, Pierzynowska K, Kutryb-Zajac B, Gaffke L, Slominska EM, Wegrzyn G, *et al.* Characterization of adenine nucleotide metabolism in the cellular model of Huntington's disease. Nucleosides Nucleotides Nucleic Acids. 2018 Nov 2;37(11):630–8.

68. Sameni S, Malacrida L, Tan Z, Digman MA. Alteration in Fluidity of Cell Plasma Membrane in Huntington Disease Revealed by Spectral Phasor Analysis. Sci Rep. 2018 Jan 15;8(1):734.

69. Fritsche M, Haessler C, Brandner G. Induction of nuclear accumulation of the tumor-suppressor protein p53 by DNA-damaging agents. Oncogene. 1993 Feb 1;8(2):307–18.

70. Smith M, Chen I, Zhan Q, O'Connor P, Fornace AJ. Involvement of the p53 tumor suppressor in repair of u.v.-type DNA damage. - Abstract - Europe PMC [Internet]. [cited 2021 Feb 22]. Available from: https://europepmc.org/article/med/7700629

71. Bae BI, Xu H, Igarashi S, Fujimuro M, Agrawal N, Taya Y, *et al.* p53 Mediates Cellular Dysfunction and Behavioral Abnormalities in Huntington's Disease. Neuron. 2005 Jul 7;47(1):29–41.

72. Marcotte R, Lacelle C, Wang E. Senescent fibroblasts resist apoptosis by downregulating caspase-3. Mech Ageing Dev. 2004 Oct 1;125(10):777–83.

73. Hayflick L, Moorhead PS. The serial cultivation of human diploid cell strains. Exp Cell Res. 1961 Dec 1;25(3):585–621.

74. Hung CLK, Maiuri T, Bowie LE, Gotesman R, Son S, Falcone M, *et al.* A patient-derived cellular model for Huntington's disease reveals phenotypes at clinically relevant CAG lengths. Drubin DG, editor. Mol Biol Cell. 2018 Nov 15;29(23):2809–20.

75. Ouellette MM, McDaniel LD, Wright WE, Shay JW, Schultz RA. The establishment of telomerase-immortalized cell lines representing human chromosome instability syndromes. Hum Mol Genet. 2000 Feb 12;9(3):403–11.

76. Bodnar AG, Ouellette M, Frolkis M, Holt SE, Chiu CP, Morin GB, *et al.* Extension of life-span by introduction of telomerase into normal human cells. Science. 1998 Jan 16;279(5349):349–52.

77. Milakovic T, Johnson GVW. Mitochondrial Respiration and ATP Production Are Significantly Impaired in Striatal Cells Expressing Mutant Huntingtin*. J Biol Chem. 2005 Sep 2;280(35):30773–82.

78. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. Nat Rev Genet. 2010 Sep;11(9):636–46.

79. Lieber MR. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. Annu Rev Biochem. 2010;79(1):181–211.

80. Gupta RM, Musunuru K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. J Clin Invest. 2014 Oct 1;124(10):4154–61.

81. Hockemeyer D, Soldner F, Beard C, Gao Q, Mitalipova M, DeKelver RC, *et al.* Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat Biotechnol. 2009 Sep;27(9):851–7.

82. Doyon Y, Vo TD, Mendel MC, Greenberg SG, Wang J, Xia DF, *et al.* Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat Methods. 2011 Jan;8(1):74–9.

83. Miller JC, Holmes MC, Wang J, Guschin DY, Lee YL, Rupniewski I, *et al.* An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol. 2007 Jul;25(7):778–85.

84. Ramirez CL, Foley JE, Wright DA, Müller-Lerch F, Rahman SH, Cornu TI, *et al.* Unexpected failure rates for modular assembly of engineered zinc fingers. Nat Methods. 2008 May;5(5):374–5.

85. Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichtinger M, *et al.* Rapid “Open-Source” Engineering of Customized Zinc-Finger Nucleases for Highly Efficient Gene Modification. Mol Cell. 2008 Jul 25;31(2):294–301.

86. Maeder ML, Thibodeau-Beganny S, Sander JD, Voytas DF, Joung JK. Oligomerized pool engineering (OPEN): an “open-source” protocol for making customized zinc-finger arrays. Nat Protoc. 2009 Oct;4(10):1471–501.

87. Streubel J, Blücher C, Landgraf A, Boch J. TAL effector RVD specificities and efficiencies. Nat Biotechnol. 2012 Jul;30(7):593–5.

88. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. Science. 2009 Dec 11;326(5959):1509–12.

89. Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. 2011 Jul 1;39(12):e82–e82.

90. Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, *et al.* A TALE nuclease architecture for efficient genome editing. Nat Biotechnol. 2011 Feb;29(2):143–8.

91. Smith C, Gore A, Yan W, Abalde-Atristain L, Li Z, He C, *et al.* Whole-Genome Sequencing Analysis Reveals High Specificity of CRISPR/Cas9 and TALEN-Based Genome Editing in Human iPSCs. Cell Stem Cell. 2014 Jul 3;15(1):12–3.

92. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science. 2007 Mar 23;315(5819):1709–12.

93. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. Nature. 2012 Feb;482(7385):331–8.

94. Gürel F, Zhang Y, Sretenovic S, Qi Y. CRISPR-Cas nucleases and base editors for plant genome editing. aBIOTECH. 2020 Jan 1;1(1):74–87.

95. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol. 2020 Feb;18(2):67–83.

96. Cho SW, Kim S, Kim JM, Kim JS. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol. 2013 Mar;31(3):230–2.

97. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science. 2012 Aug 17;337(6096):816–21.

98. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, *et al.* RNA-Guided Human Genome Engineering via Cas9. Science. 2013 Feb 15;339(6121):823–6.

99. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. Science. 2013 Feb 15;339(6121):819–23.

100. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol. 2013 Sep;31(9):822–6.

101. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol. 2013 Sep;31(9):827–32.

102. Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat Biotechnol. 2013 Sep;31(9):839–43.

103. Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Cell. 2013 Sep 12;154(6):1380–9.

104. Dianov GL, Hübscher U. Mammalian Base Excision Repair: the Forgotten Archangel. Nucleic Acids Res. 2013 Apr 1;41(6):3483–90.

105. Chu VT, Weber T, Wefers B, Wurst W, Sander S, Rajewsky K, *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol. 2015 May;33(5):543–8.

106. Maggio I, Holkers M, Liu J, Janssen JM, Chen X, Gonçalves MAFV. Adenoviral vector delivery of RNA-guided CRISPR/Cas9 nuclease complexes induces targeted mutagenesis in a diverse array of human cells. Sci Rep. 2014 May 29;4(1):5105.

107. Naldini L. Lentiviruses as gene transfer agents for delivery to non-dividing cells. Curr Opin Biotechnol. 1998 Oct 1;9(5):457–63.

108. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, *et al.* Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Science. 2014 Jan 3;343(6166):84–7.

109. Petris G, Casini A, Montagna C, Lorenzin F, Prandi D, Romanel A, *et al.* Hit and go CAS9 delivered through a lentiviral based self-limiting circuit. Nat Commun. 2017 May 22;8(1):15334.

110. Choi JG, Dang Y, Abraham S, Ma H, Zhang J, Guo H, *et al.* Lentivirus pre-packed with Cas9 protein for safer gene editing. Gene Ther. 2016 Jul;23(7):627–33.

111. Kim S, Kim D, Cho SW, Kim J, Kim JS. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res. 2014 Jun 1;24(6):1012–9.

112. DeWitt MA, Corn JE, Carroll D. Genome editing via delivery of Cas9 ribonucleoprotein. Methods. 2017 May 15;121–122:9–15.

113. Hou Z, Zhang Y, Propson NE, Howden SE, Chu LF, Sontheimer EJ, *et al.* Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proc Natl Acad Sci. 2013 Sep 24;110(39):15644–9.

114. Chatterjee P, Jakimo N, Lee J, Amrani N, Rodríguez T, Koseki SRT, *et al.* An engineered ScCas9 with broad PAM range and high specificity and activity. Nat Biotechnol. 2020 Oct;38(10):1154–8.

115. Chatterjee P, Jakimo N, Jacobson JM. Minimal PAM specificity of a highly similar SpCas9 ortholog. Sci Adv. 2018 Oct 1;4(10):eaau0766.

116. Chen X, Zaro JL, Shen WC. Fusion protein linkers: Property, design and functionality. Adv Drug Deliv Rev. 2013 Oct 15;65(10):1357–69.

117. Walton RT, Christie KA, Whittaker MN, Kleinstiver BP. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science. 2020 Apr 17;368(6488):290–6.

118. An MC, O'Brien RN, Zhang N, Patra BN, De La Cruz M, Ray A, *et al.* Polyglutamine Disease Modeling: Epitope Based Screen for Homologous Recombination using CRISPR/Cas9 System. PLoS Curr [Internet]. 2014 Apr 15 [cited 2020 Apr 23];6. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3994193/

119. Xu X, Tay Y, Sim B, Yoon SI, Huang Y, Ooi J, *et al.* Reversal of Phenotypic Abnormalities by CRISPR/Cas9-Mediated Gene Correction in Huntington Disease Patient-Derived Induced Pluripotent Stem Cells. Stem Cell Rep. 2017 Mar 14;8(3):619–33.

120. Malankhanova T, Sorokin M, Medvedev S, Zakian S, Malakhova A. Introducing an Expanded Trinucleotide Repeat Tract into the Human Genome for Huntington's Disease Modeling In Vitro. Curr Protoc Hum Genet. :21.

121. Mertens J, Paquola ACM, Ku M, Hatch E, Böhnke L, Ladjevardi S, *et al.* Directly Reprogrammed Human Neurons Retain Aging-Associated Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects. Cell Stem Cell. 2015 Dec 3;17(6):705–18.

122. hTERT RPE-1 | ATCC [Internet]. [cited 2022 Mar 1]. Available from: https://www.atcc.org/products/crl-4000

123. Lesueur LL, Mir LM, André FM. Overcoming the Specific Toxicity of Large Plasmids Electrotransfer in Primary Cells In Vitro. Mol Ther Nucleic Acids. 2016 Mar 8;5:e291.

124. Szostková M, Horáková D. The effect of plasmid DNA sizes and other factors on electrotransformation of Escherichia coli JM109. Bioelectrochem Bioenerg. 1998 Dec 1;47(2):319–23.

125. Conant D, Hsiau T, Rossi N, Oki J, Maures T, Waite K, *et al.* Inference of CRISPR Edits from Sanger Trace Data. CRISPR J. 2022 Feb;5(1):123–30.

126. Ooi J, Langley SR, Xu X, Utami KH, Sim B, Huang Y, *et al.* Unbiased Profiling of Isogenic Huntington Disease hPSC-Derived CNS and Peripheral Cells Reveals Strong Cell-Type Specificity of CAG Length Effects. Cell Rep. 2019 Feb;26(9):2494-2508.e7.

127. Koseki S, Tysinger E, Stan T, Amrani N, Savic N, Pacesa M, *et al.* PAM-Free Genome Editing with an Optimized Chimeric Cas9. Unpublished.

128. Seki A, Rutz S. Optimized RNP transfection for highly efficient CRISPR/Cas9-mediated gene knockout in primary T cells. J Exp Med. 2018 Feb 7;215(3):985–97.

129. Ranawakage DC, Okada K, Sugio K, Kawaguchi Y, Kuninobu-Bonkohara Y, Takada T, *et al.* Efficient CRISPR-Cas9-Mediated Knock-In of Composite Tags in Zebrafish Using Long ssDNA as a Donor. Front Cell

Dev Biol [Internet]. 2021 [cited 2022 Mar 1];8. Available from: https://www.frontiersin.org/article/10.3389/fcell.2020.598634

130. Llames S, García-Pérez E, Meana Á, Larcher F, del Río M. Feeder Layer Cell Actions and Applications. Tissue Eng Part B Rev. 2015 Aug;21(4):345–53.

131. Sugars KL, Brown R, Cook LJ, Swartz J, Rubinsztein DC. Decreased cAMP Response Element-mediated Transcription: AN EARLY EVENT IN EXON 1 AND FULL-LENGTH CELL MODELS OF HUNTINGTON'S DISEASE THAT CONTRIBUTES TO POLYGLUTAMINE PATHOGENESIS *. J Biol Chem. 2004 Feb 6;279(6):4988–99.

132. Chen ZJ, Kren BT, Wong PYP, Low WC, Steer CJ. Sleeping Beauty-mediated down-regulation of huntingtin expression by RNA interference. Biochem Biophys Res Commun. 2005 Apr 8;329(2):646–52.

133. Ramdzan YM, Polling S, Chia CPZ, Ng IHW, Ormsby AR, Croft NP, *et al.* Tracking protein aggregation and mislocalization in cells with flow cytometry. Nat Methods. 2012 May;9(5):467–70.

134. Dowling P, Clynes M. Conditioned media from cell lines: a complementary model to clinical specimens for the discovery of disease-specific biomarkers. Proteomics. 2011 Feb;11(4):794–804.

Appendix:

Koseki S, Tysinger E, Stan T, Amrani N, Savic N, Pacesa M, *et al.* PAM-Free Genome Editing with an Optimized Chimeric Cas9. Unpublished.

Submitted June 2021 to Nature. Presently, revisions are underway with the intention to resubmit to Nature. Natasha Savic conducted the transfections and genomic extractions. Genomic extracts were sent to Dr. Pranam Chatterjee at MIT Media Lab where extracts were amplified with PCR and analyzed for editing efficiency.

PAM-Free Genome Editing with an Optimized Chimeric Cas9

Sabrina R.T. Koseki,[1,2,*] Emma Tysinger,[1,2,*] Teodora Stan,[1,2,*]
Nadia Amrani,[3] Natasha Savic,[4] Martin Pacesa,[5]
Tomás Rodríguez,[3] Christian Kramme,[6,7] Manvitha R. Ponnapati,[1,2]
George M. Church,[6,7] Ray Truant,[4] Martin Jinek,[5]
Erik J. Sontheimer,[3] Joseph M. Jacobson,[1,2] Pranam Chatterjee,[1,2,6,7,†]

[1]Center for Bits and Atoms, [2]Media Lab, Massachusetts Institute of Technology
[3]RNA Therapeutics Institute, University of Massachusetts Medical School
[4]Department of Biochemistry and Biomedical Sciences, McMaster University
[5]Department of Biochemistry, University of Zurich
[6]Department of Genetics, Harvard Medical School
[7]Wyss Institute for Biologically Inspired Engineering, Harvard University
[*]These authors contributed equally.
[†]Corresponding author: pranam@mit.edu

CRISPR enzymes require a specified protospacer adjacent motif (PAM) flanking a guide RNA-programmed target site, limiting their sequence accessibility for robust genome editing applications. In this study, we recombine the PAM-interacting domain of SpRY, a broad-targeting Cas9 possessing a 5'-NRN-3' PAM, with the N-terminus of Sc++, a Cas9 with simultaneously broad, efficient, and accurate editing capabilities, to generate an enzyme with no distinct PAM preference: SpRYc. We demonstrate that SpRYc leverages structural properties of both enzymes to be highly active and accurate

on diverse 5'-NNN-3' PAM sequences and disease-related loci for potential therapeutic applications.

Introduction

To conduct programmable genome editing, CRISPR-associated (Cas) endonucleases require a protospacer adjacent motif (PAM) to immediately follow the target DNA sequence specified by the single guide RNA (sgRNA) (*1–3*). PAM binding triggers DNA strand separation, enabling base pairing between the sgRNA and the target DNA strand for subsequent nucleolytic cleavage and editing events (*4, 5*). The widely-utilized Cas9 from *Streptococcus pyogenes* bacteria (SpCas9), for example, requires a 5'-NGG-3' PAM (*1, 6, 7*), imposing severe accessibility constraints for therapeutically-relevant editing applications requiring precise genomic positioning, such as base editing (*8, 9*) and homology-directed repair (*10–12*).

To expand the targetable sequence space of CRISPR, we previously engineered Sc++, a variant of ScCas9 which employs a positive-charged loop that relaxes the base requirement at the second PAM position, thus enabling a 5'-NNG-3' preference, rather than the canonical 5'-NGG-3' (*13, 14*). Concurrent with the development of Sc++, Walton, et al. engineered a near-PAMless Cas9, termed SpRY, which contains mutations in the PAM-interacting domain (PID) of SpCas9 that enable strong 5'-NRN-3' specificity, alongside weaker 5'-NYN-3' targeting (*15*). Both Sc++ and SpRY thus represent exciting advances in CRISPR-based genome editing due to their robust editing characteristics and unprecedented genomic accessibility, respectively (*16*).

In this study, we combine Sc++ and SpRY to engineer an fully-optimized Cas9 enzyme

that can induce edits with no discernible PAM requirement. To do this, we employ computational modeling and experimental enzyme engineering to graft the PID of SpRY to the N-terminus of Sc++, generating the chimeric SpRYcCas9 (herein referred to as SpRYc). We demonstrate that SpRYc integrates the loop structure of Sc++ and the PID mutations of SpRY to obtain 5'-NNN-3' PAM preference, efficient editing capability in human cells, and reduced off-targeting propensity for diverse genome editing applications, including the targeting and editing of disease-related loci.

Results

Engineering of SpRYc

SpRY harbors ten substitutions in the PID of SpCas9 (L1111R, D1135L, S1136W, G1218K, E1219Q, A1322R, R1333P, R1335Q, and T1337R) which help reduce its specificity from the canonical 5'-NGG-3' to the more minimal 5'-NRN-3' PAM (*15*). Alternatively, ScCas9 and Sc++ both employ positive-charged, flexible loop-like structures in their N-terminus (residues 367 to 376) that do not exist in SpCas9 or SpRY, and relax the need for the second PAM base, enabling 5'-NNG-3' PAM preference rather than 5'-NGG-3' (*13, 14*).

Previously, we grafted the GC-independent PID of *Streptococcus macacae* Cas9 to the N-terminus of its ortholog, SpCas9, to generate iSpyMac, an efficient 5'-NAA-3' editor (*17*). Motivated by our previous domain grafting results, we engineered a single variant possessing the critical properties of SpRY and Sc++ by rationally exchanging the PID of Sc++ with that of SpRY to generate a chimeric hybrid Cas9: SpRYc. SpRYc consists of the N-terminus (residues 1-1119) of Sc++, including the flexible loop, followed by the region of SpRY (residues 1111-1368) spanning its PID mutations (Figure 1A).

In silico Modeling of SpRYc

We hypothesized that the optimized loop of Sc++ may improve the targeting breadth and efficiency of SpRY by generating sequence-nonspecific interactions with the PAM to relax the need for an A or G at position 2. Owing to the nearly 90% sequence similarity between ScCas9 and SpCas9, we conducted homology modeling of SpRYc in the DNA substrate bound-state using the SWISS-MODEL server (Figure 1A-B) (*18*). Our simulations indicate that the engineered positive-charged loop inserted into the REC1 domain points towards the PAM region of the target DNA strands and establishes new interactions with the phosphate backbone of the target strand (Figure 1Bi). In addition, the combination of ScCas9 and SpRY mutations results in several new non-specific backbone interactions with the non-target strand, thereby suggesting a relaxed PAM selectivity of SpRYc (Figure 1Bii-iii). Of note is a putative van der Waals interaction of the aromatic side chain of W1145 with the ribose moieties of the proximal non-target strand residues (Supplemental Figure 1Biv) (*19*). The new interactions resulting from the engineered mutations may thus energetically compensate for lack of PAM-specific recognition and facilitate local unwinding of double stranded DNA necessary for efficient R-loop formation in the absence of canonical PAM interactions.

PAM Characterization of SpRYc

To experimentally validate the relaxed PAM specificity of SpRYc in comparison to SpCas9, Sc++, and SpRY, we adapted a positive selection bacterial screen based on green fluorescent protein (GFP) expression conditioned on PAM binding, termed PAM-SCANR (*20*). Following transformation of the PAM-SCANR plasmid, harboring a fully randomized 5'-NNNNNNNN-3' (8N) PAM library, an sgRNA plasmid targeting the fixed PAM-SCANR protospacer, and a corresponding dCas9 plasmid, we conducted FACS analysis to isolate

GFP-positive cells in each population for subsequent library amplification and sequencing (Supplementary Figure 1). Our results demonstrate that while SpRY preferentially binds to an A or G at position 2 in the PAM, as expected, SpRYc does not bias against any specific base at any PAM position, further supporting our structural analyses (Figure 1C).

Human Genome Editing Capabilities of SpRYc

We compared the PAM specificities and DNA cleavage capabilities of SpRYc to SpCas9 and SpRY by transfecting HEK293T cells with plasmids expressing each Cas9 alongside one of sixteen sgRNAs which were directed to various genomic loci representing every two-base PAM combination (5'-N**NN**-3') (Supplementary Table 1). After five days post-transfection, indel formation was quantified following PCR amplification of the target genomic regions and subsequent sequencing analysis. Our results demonstrate that SpRYc generates indels at all tested genomic loci, compared to SpRY, which possesses minimal activity on select 5'-NYN-3' PAM sequences (Figure 2A). We similarly tested the performance of SpRYc in comparison to SpCas9 and SpRY for base editing applications by fusing each variant to ABE8e, a rapid, high-activity adenine base editor (*21, 22*). We quantified editing efficiency of the base with the highest conversion percentage in the editing window by utilizing the previously described Base Editing Evaluation Program (BEEP) following PCR amplification of the target genomic regions (*13*). Our results reveal that SpRYc-ABE8e can efficiently base edit at all tested genomic sequences, as compared to SpCas9-ABE8e and the slightly more restrictive SpRY-ABE8e (Figure 2B). Taken together, these results suggest that SpRYc is able to target, cleave, and base edit at genomic sites with minimal dependence on a specific PAM sequence.

Reduced Off-Target Propensity of SpRYc

Previously, we demonstrated that Sc++ is an intrinsically high-fidelity enzyme, with far reduced off-targeting as compared to the standard SpCas9 (*14*). We thus hypothesized that SpRYc may possess lower off-target propensity than its SpCas9-like counterpart, SpRY. To investigate this hypothesis, we employed the genome-wide, unbiased GUIDE-Seq method (*23*), by utilizing sgRNA sequences targeting two previously analyzed genomic loci (*VEGFA* and *EMX1*). From this analysis, we observe that SpRYc has nearly four-fold fewer off-targets on the more "difficult" *VEGFA* site, and two-fold reduction of off-targets on the *EMX1* site (Figure 2C, Supplementary Figures 2-3). We corroborated this data via a mismatch tolerance assay (*24*), in which we employed sgRNAs harboring double or single mismatches to a fixed protospacer for an endogenous *DNMT1* locus. SpRYc exhibited decreased activity on mismatched sequences, as compared to SpRY, with no detectable loss of on-target activity (Figure 2D).

SpRYc Base Editors Mediate Therapeutically-Relevant Edits

Having established SpRYc's broad, efficient, and accurate editing capabilities in human cells, we sought to investigate its utility as a potential therapeutic modality for the treatment of genetic diseases. Rett syndrome (RTT) is a progressive neurological disorder that predominantly affects young females. A majority of patients carry one of eight mutations in the *MECP2* gene (C316T, C397T, C473T, C502T, C763T, C808T, C880T, C916T), all of which are caused by C-to-T substitution mutations and can thus be potentially ameliorated by CRISPR adenine base editors, such as ABE8e (*9, 21, 25*). Notably, one of the eight mutations, C502T, can only be accessed at target sites consisting of a 5'-NCN-3' or 5'-NTN-3' PAM, preventing its correction by previous adenine base editors. To test whether SpRYc-ABE8e can effectively edit at all eight sites, we generated a universal

RTT HEK293T cell line via *piggyBac* transposase-mediated integration of a synthetic gene fragment encoding *MECP2* installed with all eight of the aforementioned RTT mutations. After puromycin-based clonal selection, we transfected the SpRYc-ABE8e plasmid alongside the appropriate sgRNAs for each site (Supplementary Table 1). After subsequent DNA extraction, loci amplification, and sequencing, we demonstrate that SpRYc-ABE8e can effectively edit all eight targets, including over 20% editing efficiency at the C502T mutation (Figure 3A).

Huntington's Disease (HD) is a fully-penetrant neurological disorder affecting over 1 in 10000 adults (*26*). It is caused by an expanded CAG repeat on chromosome 4 of the *HTT* gene, which encodes an extended polyglutamine (polyQ) tract in the resulting *huntingtin* protein (*26*). Recent studies have shown that there is an inverse relationship between the age of disease onset and the number of CAG repeats (rather than the length of the resulting polyQ tract) (*27*). We therefore assessed SpRYc's ability to introduce silent CAA interruptions in the CAG repeat region of *HTT*. To do this, we transfected patient-derived TruHD cells, possessing a clinically-relevant CAG repeat length of 43 (*28*), with a cytosine base editor SpRYc-BE4Max alongside an sgRNA targeting the antisense strand of the HTT repeat region (*29*) (Supplementary Table 1). Our sequencing results show that SpRYc can install a CAA interruption at the fourth CAG repeat, with an editing efficiency of over 30%, thus reducing the uninterrupted repeat length by 4 (Figure 3B). Taken together, these results illustrate SpRYc's potential utility for clinically-relevant applications and motivate its development as a therapeutic platform.

Discussion

While PAMs play a critical role in self-nonself discrimination by prokaryotic CRISPR-Cas9 immune systems, they limit the accessible sequence space for genome editing applications. Recent engineering and discovery efforts have yielded a host of Cas9 variants with altered or relaxed PAM sequences (*13–15, 17, 30–36*). In this study, we engineer an optimized Cas9 by harnessing the structural properties of SpRY and Sc++ to generate SpRYc, a Cas9 with undetectable PAM preference. SpRYc may thus increase the targetable sequence space for expanded base editing capabilites, more efficient homology-directed repair, and multiplexed screening platforms. We further show that SpRYc has reduced off-target effects as compared to SpRY, and due to high sequence homology of ScCas9 and SpCas9, we anticipate that high-fidelity mutations (*24, 37, 38*) can easily be ported into SpRYc for improved specificity, as has been shown previously for both Sc++ and SpRY (*14, 15*). Finally, we demonstrate that SpRYc can be integrated within base editing architectures to edit disease-related loci for potential therapeutic purposes.

Recently, Collias and Beisel highlighted the implications of a PAM-free nuclease, indicating that though powerful, such a tool would have severe drawbacks (*39*). For example, a PAM-free enzyme may edit its own sgRNA-expressing DNA construct and/or force interrogation of all target sequences in the genome, yielding hampered editing rates on-target while increasing accessibility to off-target sequences (*39*). During our genome editing experiments, we regularly amplified and sequenced the sgRNA plasmid cassette from various cell lysate samples and observed no detectable editing (Supplementary Figure 4), suggesting that SpRYc may have impaired targeting at the 5'-GTTTAGAG-3' PAM within the canonical SpCas9 and ScCas9 sgRNA scaffold. Furthermore, we observed editing rates

similar to SpCas9 and SpRY on overlapping PAM sequences, indicating no discernable trade-off between editing and accessibility. Finally, we showed that SpRYc has reduced off-targeting and mismatch tolerance than SpRY. Thus, SpRYc may overcome many of the limitations associated with a PAM-free CRISPR tool, though more investigation is needed.

While SpRYc serves as a step forward towards unrestricted, fully programmable genome editing, its development, more importantly, represents a culmination of a variety of state-of-the-art *in silico* and *in vitro* PAM engineering methods. ScCas9 was first identified via a high-throughput bioinformatics algorithm for ortholog discovery, dubbed SPAMALOT (*13*). Its derivative, Sc++, was engineered by computationally identifying and extracting motifs from *Streptococcus* orthologs, and splicing them into ScCas9 for improved functionality (*14*). Concurrently, SpRY was the result of a multi-year effort of SpCas9-based directed evolution and rational mutagenesis (*15, 30*). Finally, a combination of structure-based homology modeling and domain grafting methods, those that were instrumental in engineering other PAM variants such as iSpyMac (*17*) and cCas9 (*36*), enabled the fusion of SpRY and Sc++ into our final SpRYc variant. Together, these studies emphasize the power of integrating diverse engineering modalities to generate new and useful proteins and open the door for future integrative protein design.

Materials and Methods

Homology Modeling

Structural models of SpRYc were generated using the SWISS-MODEL server (*18*), using the PDB 4UN3 DNA substrate bound Cas9 model as template (*6*). Modelled sidechains and loop were curated and adjusted manually using COOT software (*40*).

Generation of Plasmids

To generate SpRYc, the N-terminal ORF of Sc++ (Addgene Plasmid #155011), corresponding to residues (1-1119) was PCR amplified and assembled using Gibson Assembly into the pCMV-T7-SpRY-P2A-EGFP backbone (Addgene Plasmid #139989), preserving residues 1111-1368 of SpRY's ORF. pCMV-T7-SpCas9-P2A-EGFP (Addgene Plasmid #139987) was used for SpCas9, and Sc++ was similarly integrated within the backbone. Analogously, the ORFs of SpCas9, SpRY, and SpRYc were integrated within the ABE8e (Addgene Plasmid #138489) and AncBE4Max (Addgene Plasmid #112094) backbones. sgRNA plasmids were constructed by annealing oligonucleotides coding for crRNA sequences (Table S1) as well as 4 bp overhangs, and subsequently performing a T4 DNA Ligase-mediated ligation reaction into a plasmid backbone immediately downstream of the human U6 promoter sequence. Assembled constructs were transformed into 50 μL NEB Turbo Competent *E. coli* cells, and plated onto LB agar supplemented with the appropriate antibiotic for subsequent sequence verification of colonies and plasmid purification.

PAM-SCANR Assay

Plasmids for the SpCas9 sgRNA and PAM-SCANR genetic circuit, as well as BW25113 ΔlacI cells, were generously provided by the Beisel Lab (North Carolina State University). Plasmid libraries containing the target sequence followed by either a fully-randomized 8-bp 5'-NNNNNNNN-3' library or fixed PAM sequences were constructed by conducting site-directed mutagenesis, utilizing the KLD enzyme mix (NEB) after plasmid amplification, on the PAM-SCANR plasmid flanking the protospacer sequence (5'-CGAAAGGTTTTGCACTCGAC-3'). Nuclease-deficient mutations (D10A and H850A) were introduced to the ScCas9 variants using Gibson Assembly as previously described.

The provided BW25113 cells were made electrocompetent using standard glycerol wash and resuspension protocols. The PAM library and sgRNA plasmids, with resistance to kanamycin (Kan) and carbenicillin (Crb) respectively, were co-electroporated into the electrocompetent cells at 2.4 kV, outgrown, and recovered in Kan+Crb Luria Broth (LB) media overnight. The outgrowth was diluted 1:100, grown to ABS600 of 0.6 in Kan+Crb LB liquid media, and made electrocompetent. Indicated dCas9 plasmids, with resistance to chloramphenicol (Chl), were electroporated in duplicates into the electrocompetent cells harboring both the PAM library and sgRNA plasmids, outgrown, and collected in 5 mL Kan+Crb+Chl LB media. Overnight cultures were diluted to an ABS600 of 0.01 and cultured to an OD600 of 0.2. Cultures were analyzed and sorted on a FACSAria machine (Becton Dickinson). Events were gated based on forward scatter and side scatter and fluorescence was measured in the FITC channel (488 nm laser for excitation, 530/30 filter for detection), with at least 10,000 gated events for data analysis. Sorted GFP-positive cells were grown to sufficient density, plasmids from the pre-sorted and sorted populations were isolated, and the region flanking the nucleotide library was then PCR amplified and submitted for Sanger sequencing or Amplicon-EZ NGS analysis (Genewiz). FCS files were analyzed using FCSalyzer (`https://sourceforge.net/projects/fcsalyzer/`)), and gating strategy is described in Supplementary Figure 1.

Cell Culture and DNA Modification Analysis

HEK293T cells were maintained in DMEM supplemented with 100 units/ml penicillin, 100 mg/ml streptomycin, and 10% fetal bovine serum (FBS). sgRNA plasmids (100 ng) and nuclease plasmids (100 ng) were transfected into cells as duplicates (2 x 10^4/well in a 96-well plate) with Lipofectamine 3000 (Invitrogen) in Opti-MEM (Gibco). After 5 days post-transfection, genomic DNA was extracted using QuickExtract Solution (Lucigen), and

genomic loci were amplified by PCR utilizing the Phusion Hot Start Flex DNA Polymerase (NEB). Amplicons were enzymatically purified and submitted for Sanger sequencing or Amplicon-EZ NGS sequencing (Genewiz). Sanger sequencing ab1 files were analyzed using the ICE web tool for batch analysis (`ice.synthego.com`) (*41*) in comparison to an unedited control to calculate indel frequencies via the ICE-D score. Select samples were further verified using the TIDE algorithm (`tide.deskgen.com`) to ascertain consistency of editing rates between replicates (*42*). NGS FASTQ files were analysed using a batch version of the software CRISPResso2 (`https://github.com/pinellolab/CRISPResso2`) (*43*). Base editing files were analyzed via the Based Editing Evaluation Program (BEEP) (`https://github.com/mitmedialab/BEEP`) in comparison to an unedited control. All samples were performed in independent duplicates or triplicates, as indicated.

GUIDE-Seq

We performed GUIDE-Seq as described previously (*23*). Briefly, HEK293T cells were electroporated in a 24-well plate with 500 ng of Cas9, 500 ng of sgRNA, 10 ng of mCherry plasmids, and 7.5 pmol of annealed GUIDE-Seq oligonucleotide using the Neon nucleofection system (Thermo Fisher Scientific). After 72 hours post-nucleofection, genomic DNA was extracted with a DNeasy Blood and Tissue kit (Qiagen 69504) according to the manufacturer's protocol. DNA libraries were prepared using custom oligonucleotides described in Tsai, et al. (*23*). Library preparations were done with original adaptors with each library barcoded for pooled sequencing. The barcoded, purified libraries were sequenced on a MiniSeq platform in a paired-end (150/150) run.

Raw sequencer output (BCL) was demultiplexed and aligned to hg38 using GS-Preprocess (`github.com/umasstr/GS-Preprocess`) (*44*). This software also constructed a reference

of UMIs unique to each read and merged technical replicate BAM files. Off-target analysis of this input was performed using the GUIDEseq Bioconductor package (*45*). Only sites that harbored a sequence with ≤6 mismatches relative to the gRNA were considered potential off-target sites. GUIDE-Seq read count data is indicated in Supplementary Figures 2-3.

Rett Syndrome Cell Line Generation

The *MECP2* editing locus containing all common Rett syndrome mutations was synthesized as a gBlock from IDT and inserted via Gateway cloning to a promoter-less *piggyBac* pMVP destination vector (Addgene 121874) harboring puromycin resistance. The RTT vector was then integrated into the HEK293T cell line via lipofection. Briefly, 600,000 cells were seeded in D10 media (DMEM + 10% FBS) to a six well plate 24 hours prior to lipofection. 2.5 ug of the RTT plasmid and 0.5 ug of a CMV-super PiggyBac transposase (System Biosciences) were then lipofected using Lipofectamine 3000 according to the manufacturer's protocol. Media was changed six hours post-transfection and cells were subjected to 1 ug/ml puromycin selection 48 hours post-transfection for 3 days. Cells were then expanded under no drug selection for three days to allow non-integrated plasmid loss, then again selected for 3 additional days to isolate a pure population.

TruHD Cell Culture

TruHD-Q43Q17M cells (*28*) were cultured in MEM supplemented with 15% FBS and 1% Glutamax and grown under 3% O2 and 5% CO2 at 37 degrees C in a 10 cm plate. At 95% confluence, cells were transfected through Lonza nucleofection using the SG Cell Line 4D-Nucleofector Kit. Growth media was replaced 24 hours post-nucleofection. 5 days post-nucleofection genomic DNA was extracted with PureLink Genomic DNA Mini

Kit (Invitrogen).

Statistical Analysis

Data are shown as mean of duplicate values. Data was plotted using Matplotlib and the Prism GraphPad software.

Supplementary Materials

Supplementary Table 1. Curated dataset of indel and base editing efficiencies at each site (annotated by target sequence, genomic locus and PAM), including relevant DNA and protein sequences related to this study.

Supplementary Figure 1. Gating Strategy for PAM-SCANR FACS analysis.

Supplementary Figure 2. GUIDE-Seq data including counts at each detected off-target with $\leq$6 mismatches for each nuclease tested.

Supplementary Figure 3. GUIDE-Seq data represented graphically of detected off-target read counts for each nuclease tested.

Supplementary Figure 4. Sequencing of select SpRYc sgRNA DNA constructs from genomic DNA extracts.

Acknowledgments

We thank Dr. Neil Gershenfeld and Dr. Shuguang Zhang for shared lab equipment. We further thank Dr. Noah Jakimo for critical insights.

Author Contributions

P.C. conceived, designed, directed, and supervised the study. P.C. designed constructs, carried out genome editing experiments, and conducted data analyses. S.R.T.K., E.T., and T.S. built constructs, conducted transfections, carried out genome editing experi-

ments, and performed data analyses. N.A. and T.R. carried out GUIDE-Seq experiments and analysis. N.S. conducted genome editing assays in TruHD cells. M.P. and M.J. performed structural modelling of SpRYc. M.R.P. conducted *in silico* docking experiments. C.K. developed Rett cell lines. P.C. and S.R.T.K. wrote the paper. All authors reviewed and edited the paper. G.M.C. supervised cell line construction work. R.T. supervised Huntington Disease work. M.J. supervised homology modeling and structural analyses. E.J.S. supervised GUIDE-Seq experiments. J.M.J. reviewed the paper and provided critical insights.

Declarations

Work done by P.C., S.R.T.K., E.T., T.S., M.R.P. and J.M.J. was supported by the consortia of sponsors of the MIT Media Lab and the MIT Center for Bits and Atoms, as well as the CHDI Foundation and the Rett Syndrome Research Trust. GUIDE-Seq work done by N.A. and T.R. was supported by a grant (GM115911) to E.J.S from the U.S. National Institutes of Health (NIH). Homology modeling work done by M.P. and M.J. was supported by Swiss National Science Foundation Grant 31003A-182567 (to M.J.). R.T. and N.S. are supported by a grant from the Krembil Foundation. M.J. is an International Research Scholar of the Howard Hughes Medical Institute (HHMI) and Vallee Scholar of the Bert L. N. Kuggie Vallee Foundation.

Data and Materials Availability

All data needed to evaluate the conclusions in the paper are present in the paper and supplementary tables. All source data and sequencing files can be found at `https://tinyurl.com/yxm2wpfx`. SpRYc plasmids will be made available on Addgene.

References

1. Jinek, M. *et al.* A programmable dual-rna-guided dna endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)* **337**, 816–821 (2012).

2. Mojica, F. J. M., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic crispr defence system. *Microbiology (Reading, England)* **155**, 733–740 (2009).

3. Shah, S. A., Erdmann, S., Mojica, F. J. & Garrett, R. A. Protospacer recognition motifs. *RNA Biology* **10**, 891–899 (2013).

4. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. Dna interrogation by the crispr rna-guided endonuclease cas9. *Nature* **507**, 62–67 (2014).

5. Gong, S., Yu, H. H., Johnson, K. A. & Taylor, D. W. DNA unwinding is the primary determinant of CRISPR-cas9 activity. *Cell Reports* **22**, 359–371 (2018).

6. Anders, C. & Jinek, M. In vitro enzymology of cas9. *Methods in enzymology* **546**, 1–20 (2014).

7. Anders, C., Bargsten, K. & Jinek, M. Structural plasticity of PAM recognition by engineered variants of the RNA-guided endonuclease cas9. *Molecular Cell* **61**, 895–902 (2016).

8. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic dna without double-stranded dna cleavage. *Nature* **533**, 420–424 (2016).

9. Gaudelli, N. M. *et al.* Programmable base editing of a-t to g-c in genomic dna without dna cleavage. *Nature* **551**, 464–471 (2017).

10. Cong, L. *et al.* Multiplex genome engineering using CRISPR/cas systems. *Science* **339**, 819–823 (2013).

11. Mali, P. *et al.* RNA-guided human genome engineering via cas9. *Science* **339**, 823–826 (2013).

12. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2** (2013).

13. Chatterjee, P., Jakimo, N. & Jacobson, J. M. Minimal pam specificity of a highly similar spcas9 ortholog. *Science advances* **4**, eaau0766 (2018).

14. Chatterjee, P. *et al.* An engineered ScCas9 with broad PAM range and high specificity and activity. *Nature Biotechnology* **38**, 1154–1158 (2020).

15. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-cas9 variants. *Science* (2020).

16. Tang, L. PAM-less is more. *Nature Methods* **17**, 559–559 (2020).

17. Chatterjee, P. *et al.* A cas9 with PAM recognition for adenine dinucleotides. *Nature Communications* **11** (2020).

18. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **46**, W296–W303 (2018).

19. Wilson, K. A., Kellie, J. L. & Wetmore, S. D. DNA–protein -interactions in nature: abundance, structure, composition and strength of contacts between aromatic amino acids and DNA nucleobases or deoxyribose sugar. *Nucleic Acids Research* **42**, 6726–6741 (2014).

20. Leenay, R. T. *et al.* Identifying and visualizing functional PAM diversity across CRISPR-cas systems. *Molecular Cell* **62**, 137–147 (2016).

21. Richter, M. F. *et al.* Phage-assisted evolution of an adenine base editor with improved cas domain compatibility and activity. *Nature Biotechnology* **38**, 883–891 (2020).

22. Lapinaite, A. *et al.* DNA capture by a CRISPR-cas9–guided adenine base editor. *Science* **369**, 566–571 (2020).

23. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-cas nucleases. *Nature Biotechnology* **33**, 187–197 (2014).

24. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR–cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).

25. Liyanage, V. R. B. & Rastegar, M. Rett syndrome and MeCP2. *NeuroMolecular Medicine* **16**, 231–264 (2014).

26. McColgan, P. & Tabrizi, S. J. Huntington's disease: a clinical review. *European Journal of Neurology* **25**, 24–34 (2017).

27. Lee, J.-M. *et al.* CAG repeat not polyglutamine length determines timing of huntington's disease onset. *Cell* **178**, 887–900.e14 (2019).

28. Hung, C. L.-K. *et al.* A patient-derived cellular model for huntington's disease reveals phenotypes at clinically relevant CAG lengths. *Molecular Biology of the Cell* **29**, 2809–2820 (2018).

29. Koblan, L. W. *et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nature biotechnology* **36**, 843–846 (2018).

30. Kleinstiver, B. P. *et al.* Engineered CRISPR-cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).

31. Kleinstiver, B. P. *et al.* Broadening the targeting range of staphylococcus aureus CRISPR-cas9 by modifying PAM recognition. *Nature Biotechnology* **33**, 1293–1298 (2015).

32. Edraki, A. *et al.* A compact, high-accuracy cas9 with a dinucleotide pam for in vivo genome editing. *Molecular cell* **73**, 714–726.e4 (2019).

33. Gasiunas, G. *et al.* A catalogue of biochemically diverse CRISPR-cas9 orthologs. *Nature Communications* **11** (2020).

34. Nishimasu, H. *et al.* Engineered crispr-cas9 nuclease with expanded targeting space. *Science (New York, N.Y.)* **361** (2018).

35. Hu, J. H. *et al.* Evolved cas9 variants with broad pam compatibility and high dna specificity. *Nature* **556**, 57–63 (2018).

36. Ma, D. *et al.* Engineer chimeric cas9 to expand pam recognition based on evolutionary information. *Nature communications* **10**, 560 (2019).

37. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).

38. Vakulskas, C. A. *et al.* A high-fidelity cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nature Medicine* **24**, 1216–1224 (2018).

39. Collias, D. & Beisel, C. L. CRISPR technologies and the search for the PAM-free nuclease. *Nature Communications* **12** (2021).

40. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development ofCoot. *Acta Crystallogr D Biol Cryst* **66**, 486–501 (2010).

41. Hsiau, T. *et al.* Inference of CRISPR edits from sanger trace data (2018).

42. Brinkman, E. K. & van Steensel, B. Rapid quantitative evaluation of CRISPR genome editing by TIDE and TIDER. In *Methods in Molecular Biology*, 29–44 (Springer New York, 2019). URL `https://doi.org/10.1007/978-1-4939-9170-9_3`.

43. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nature Biotechnology* **37**, 224–226 (2019).

44. Rodríguez, T. C., Pratt, H. E., Liu, P., Amrani, N. & Zhu, L. J. GS-preprocess: Containerized GUIDE-seq data analysis tools with diverse sequencer compatibility (2020).

45. Zhu, L. J. *et al.* GUIDEseq: a bioconductor package to analyze GUIDE-seq datasets for CRISPR-cas nucleases. *BMC Genomics* **18** (2017).

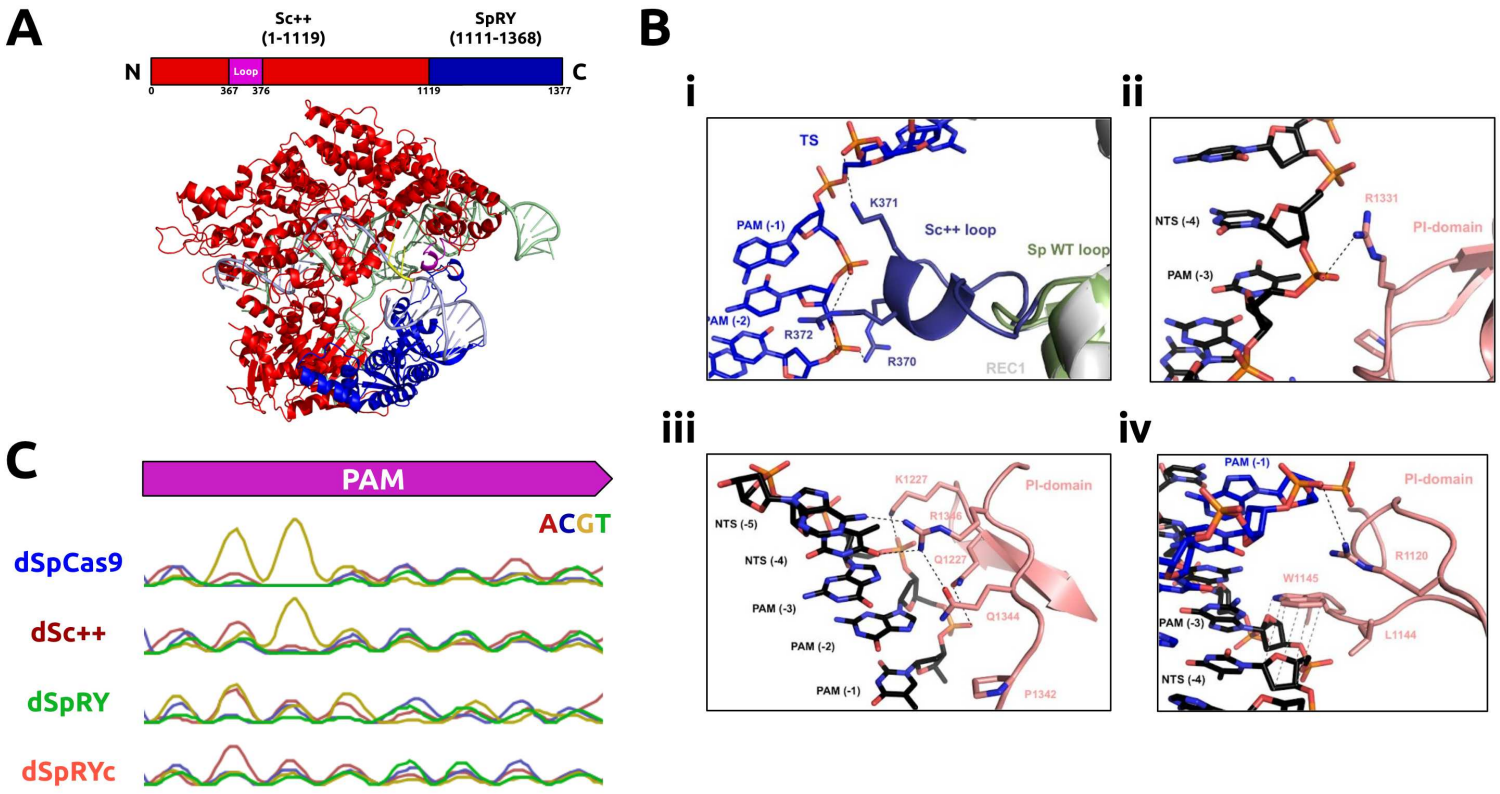


Figure 1: Engineering, Modeling, and PAM Characterization of SpRYc A) Homology model of SpRYc generated in SWISS-MODEL from PDB 4UN3 and visualized in PyMol. Original domain coordinates are indicated in parentheses above while SpRYc coordinates are indicated below. PAM is indicated in yellow, loop in purple, Sc++ N-terminus in red, and SpRY PID in blue. B) i. Interaction of the engineered Sc++ loop (purple) with the backbone of the target strand (TS) PAM region. The REC1 loop from wild type SpCas9 is indicated in green. ii. Potential interaction of residue R1331 with the non-target strand (NTS) backbone. iii. Multiple mutations within the PAM interaction loop allow for a more flexible PAM readout. iv. The potential van der Waals interaction of W1145 with the ribose moieties of non-target strand residues could further stabilize the PAM interaction. C) PAM enrichment for indicated dCas9 enzymes. Each dCas9 plasmid was electroporated in duplicates, subjected to FACS analysis, and gated for GFP expression based on a negative "No Cas9" control and a positive dSpCas9 control. All samples were performed in independent transformation duplicates (n=2), and the PAMs of the GFP-positive cells were sequenced via Sanger sequencing.

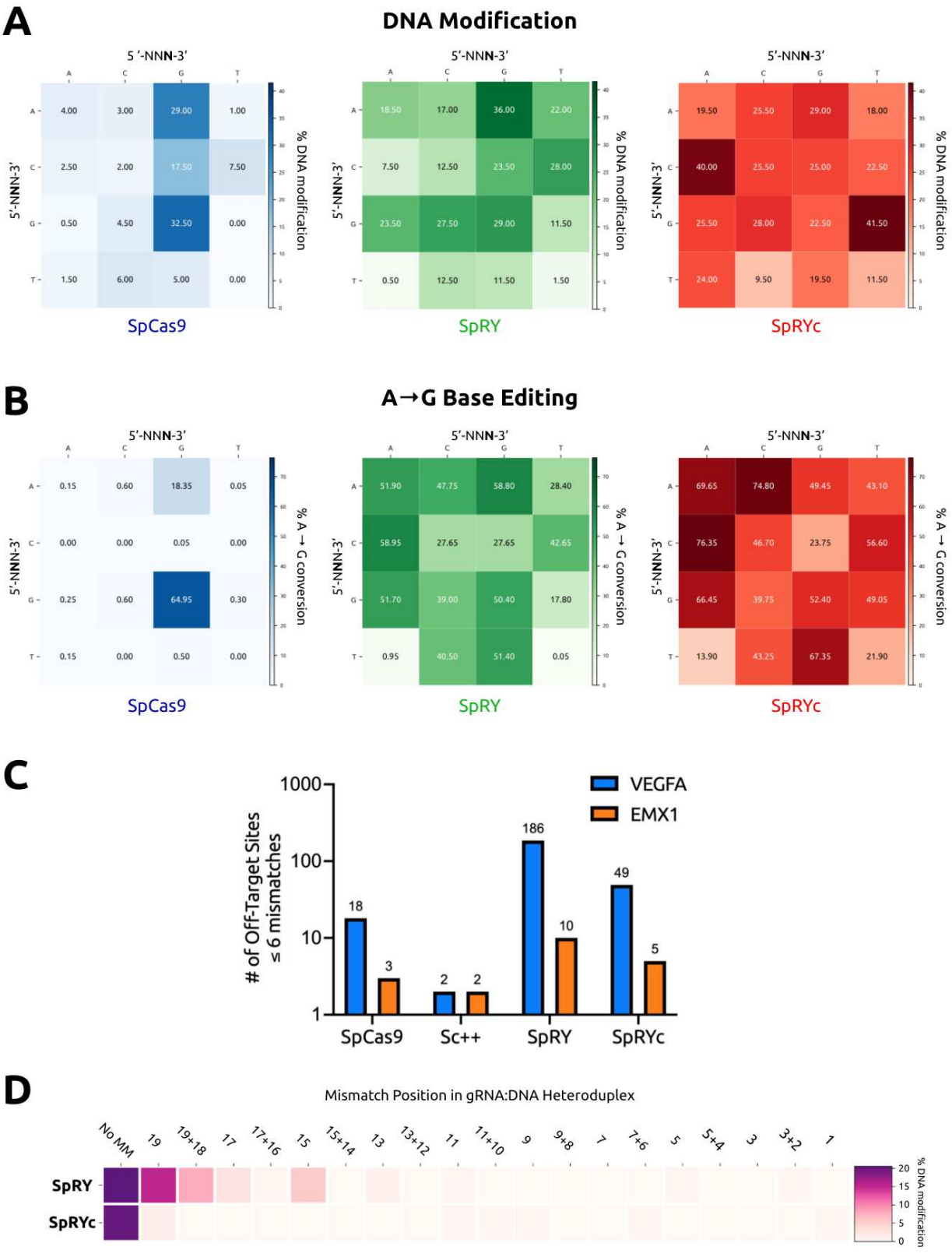

A
DNA Modification
5 '-NNN-3'
5'-NNN-3'
5'-NNN-3'
5'-NNN-3'
% DNA modification
SpCas9
SpRY
SpRYc
B
A→G Base Editing
% A → G conversion
C
# of Off-Target Sites ≤ 6 mismatches
VEGFA
EMX1
SpCas9
Sc++
SpRY
SpRYc
D
Mismatch Position in gRNA:DNA Heteroduplex
No MM
SpRY
SpRYc

Figure 2: Broad, Efficient, and Specific Genome Editing Capabilities of SpRYc. A) Quantitative analysis of indel formation with indicated Cas9 variants. Indel frequencies were determined via batch analysis following PCR amplification of indicated genomic loci, in comparison to unedited controls for each gene target. All samples were performed in independent transfection duplicates (n=2) and the mean of the quantified indel formation values was calculated. B) Quantitative analysis of A-to-G with indicated ABE8e variants. Base editing conversion rates were determined via BEEP following PCR amplification of indicated genomic loci, in comparison to unedited controls for each gene target. All samples were performed in independent transfection duplicates (n=2) and the mean of the quantified base editing formation values was calculated. C) Off-targets as identified by GUIDE-seq genome-wide for SpCas9, Sc++, SpRY, and SpRYc each paired with two sgRNAs targeting either *EMX1* or *VEGFA*. Only sites that harbored a sequence with $\leq$6 mismatches relative to the gRNA were considered potential off-target sites. D) Efficiency heatmap of mismatch tolerance assay on genomic targets. Quantified indel frequencies are exhibited for each labeled single or double mismatch (number of bases 5' upstream of the PAM) in the sgRNA sequence for the indicated Cas9 variant and indicated PAM sequence. All samples were performed in independent transfection duplicates (n=2) and the mean of the quantified indel formation values was calculated.

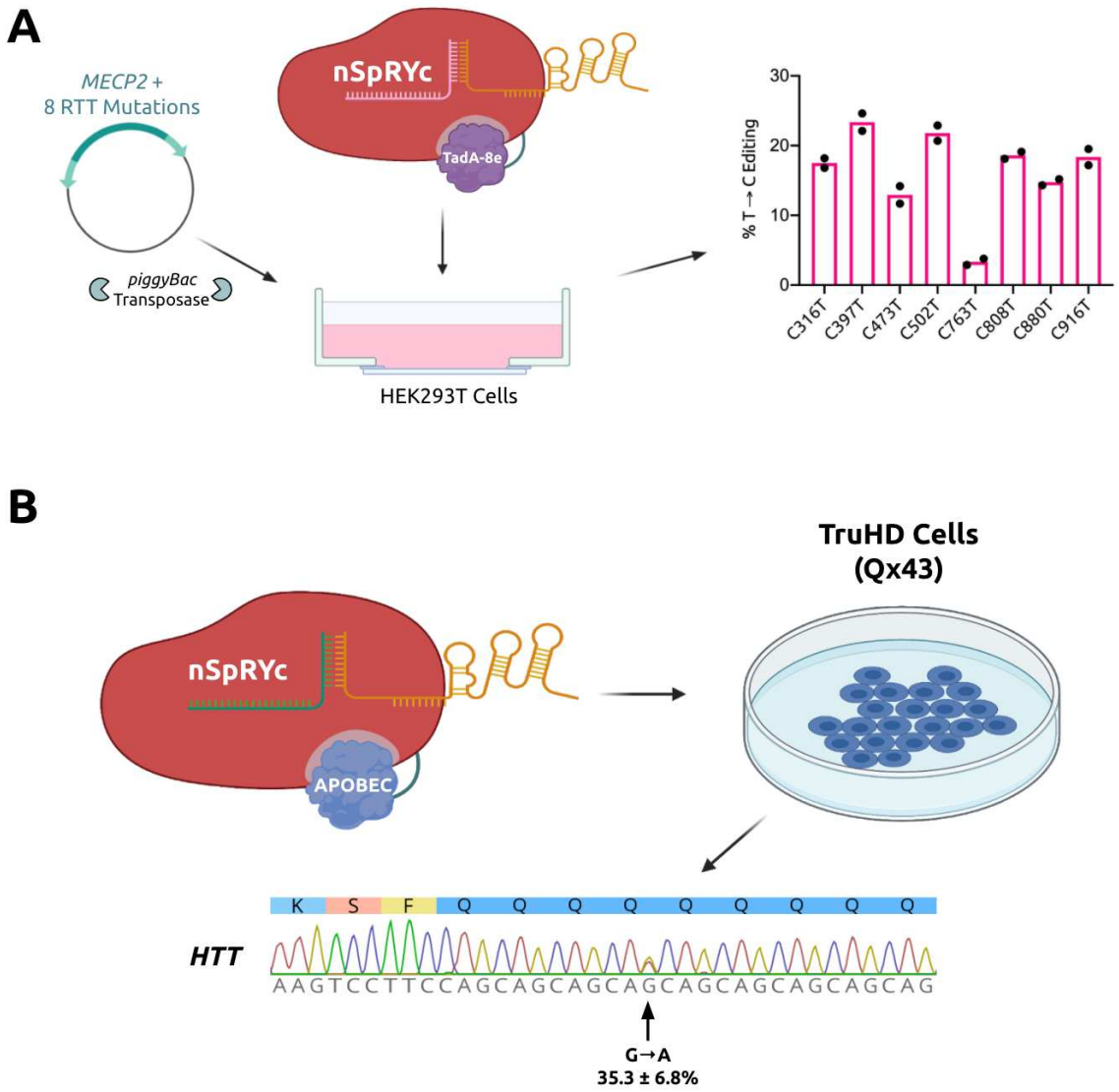


Figure 3: Targeting Disease-Associated Loci with SpRYc. A) Schematic of SpRYc RTT Experiment. Briefly, a synthetic *MECP2* gene fragment encoding eight specified RTT mutations was integrated into HEK293T cells, followed by transfection of SpRYc-ABE8e and sgRNAs targeting each mutation. Base editing conversion rates were determined via BEEP following PCR amplification of indicated genomic loci, in comparison to unedited controls for each mutation. All samples were performed in independent transfection duplicates (n=2) and the mean of the quantified base editing formation values was calculated. B) Schematic of SpRYc HTT Experiment. SpRYc-BE4Max was nucleofected into TruHD cells alongside an sgRNA targeting the *HTT* repeat. Base editing conversion rate was determined via BEEP following PCR amplification of indicated genomic loci, in comparison to an unedited control. Samples were performed in independent nucleofection triplicates (n=3) and the mean of the quantified base editing formation values was calculated.