PRECISE AND EFFICIENT THERAPEUTIC GENOME EDITING FOR THE CORRECTION OF GENETIC DISEASES IN ANIMALS

A Thesis

by

CARLOS ANDRES PINZÓN ARTEAGA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Charles R. Long
Co-Chair of Committee,	Michael C. Golding
Committee Member,	Michael C. Satterfield
Head of Department,	Larry Suva

December 2017

Major Subject: Biomedical Science

ABSTRACT

There are more than 6,052 identified genetic mutations linked to disease in humans and animals. Thanks to the advent of gene editing based on programmable nucleases and the advances in DNA sequencing and writing technologies, it is now possible to make precise changes in eukaryotic genomes with the potential to correct monogenic diseases, from affected cells, tissues, organisms and eventually whole populations. This is the concept behind therapeutic genome editing, which arises out of the idea that instead of pursuing palliative care, the ideal therapy for monogenic diseases would be to develop a method that can directly correct the disease-causing mutations.

Many of these disease alleles have been have been unknowingly co-selected when performing phenotypic genetic selection on plants and animals. Although selected breeding has been successful in the establishment and improvement of many different strains of plants and breeds of animals, we have been propagating these disease alleles in the populations. One of these deleterious alleles is the Glycogen Branching Enzyme Deficiency (GBED), which is caused by a nonsense mutation (C > A) in the first exon of the *GBE1* gene that severely disrupts glycogen metabolism. This mutation is lethal in homozygotes and an estimated 9% of Quarter Horse and Paint Horse lineages are heterozygote carriers. In this work, we corrected this mutation in a heterozygous cell line

derived from a high genetic merit American Quarter Horse stallion, by using CRISPR-Cas9. The long-term goal is to use the corrected cell lines for somatic cell nuclear transfer (SCNT) thereby generating a cloned animal that maintains the genetic merit of its predecessor, but is free of the GBED mutation.

Precise genome editing requires the introduction of a double stranded break (DSB) at an exact location in the genome and the correct DNA repair outcome. Although CRISPR-Cas9 has allowed for the introduction of precise DSBs in a very efficient manner, the lack of control over cell-autonomous repair mechanisms namely non-homologous end-joining (NHEJ) and homologous recombination (HR), is still the major bottle neck for seamless genome editing. The DNA-dependent protein kinase (DNA-PK), composed of the Ku 70 - Ku 80 heterodimer and the DNA-PK catalytic subunit ($DNA\text{-}PK_{cs}$), is best known as the NHEJ molecular sensor for DNA damage, but has been also identified as a pattern recognition receptor (PRR) that defends against the invasion of foreign nucleic acids. Here we devised a novel strategy that capitalizes on the natural ability of the *Vaccinia virus* (VACV) C16 protein that evolved as an elegant subversion mechanism to inhibit the detection of the VACV genome by the host cytoplasmic PRR defenses, specifically the Ku-mediated DNA sensing.

DEDICATION

To my family

ACKNOWLEDGEMENTS

I would like to thank my mentors Dr. Charles Long and Dr. Michael Golding for their trust, support, and guidance. I would also like to thank the help of Mathew Snyder, Ciccera Lazzarotto, Diego Fernando Carrillo, Nicolas Moreno, Dalton Horn, Camilo Hernandez, Megan Mikkelson, Haley White, James Oldeschulte, Carly Turner, Yudi Bedi, William Skiles, Charlotte Whittaker and Cassandra Skenadore for their help in different aspects of these projects. Finally, I would also like to thank Dr. Gus Wright for his help on the FACS sort and the gating analysis on the flow data, Dr. Andrew Hillhouse for his help with the ddPCR, Dr. Terje Raudsepp for her help with karyotyping and BAC sequencing, Dr. Eun-Gyu No for his help on Sanger Sequencing and Dr. Rytis Juras for his help on the independent mutation correction verification and typing of the GBED colonies.

CONTRIBUTORS AND FUNDING SOURCES

This work was supported by a dissertation committee consisting of Dr. Charles Long and Dr. Michael Golding of the Department of Veterinary Physiology and Pharmacology; and Dr. Carrie Satterfield of the Department of Animal Science. All work for the dissertation was completed independently by the student.

This work was financed by the Link Equine Research Program, the Legends Premier Stallion Season Auction, the CVM Graduate Student Trainee Grant, the Center for Organ Biotechnology & Texas Heart Institute Innovation Kitchen Grant and the CVM Advanced Developmental Training Award.

NOMENCLATURE

A-EJ	Alternative End Joining
BFP	Blue Fluorescent Protein
BLG	ß-Lactoglobulin
bESC	Bovine Embryonic Stem-Like Cells
Cas	CRISPR-Associated Genes
CDKs	Cyclin-Dependent Kinases
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
CSR	Class-Switch Recombination
CtIP	C-Terminal Dinding Protein Interacting Protein
DNA	Deoxyribonucleic Acid
DNA-PK	DNA-Dependent Protein Kinase
DNA-PKcs	DNA-PK Catalytic Subunit
DSB	Double Stranded Break
ESCs	Embryonic Stem Cells
GE	Genetic Engineering
GBE	Glycogen Branching Enzyme
GBE1	Glucan (1,4-Alpha-) Branching Enzyme 1 Gene

GBED	Glycogen Branching Enzyme Deficiency
GFP	Green Fluorescent Protein
HDR	Homology-Directed Repair
iPSC	Induced Pluripotent Stem Cell
LIG4	DNA Ligase 4
MRE11	Meiotic Recombination 11
MTBC	*Mycobacterium tuberculosis complex*
NBS1	Nijmegen Breakage Syndrome Protein 1
NHEJ	Non-Homologous End-Joining
PRR	Pattern Recognition Receptor
PNMI	Pronuclear Microinjection
PAM	Protospacer Adjacent Motif
RAD50	RAD50 Double Strand Break Repair Protein
RNA	Ribonucleic Acid
sgRNA	Single Guide RNA
SRSR	Short Regularly Spaced Repeats
SSA	Single-Strand Annealing
ssODN	Single-Stranded Oligodeoxynucleotide
SV-40	Simian Virus- 40
SCNT	Somatic Cell Nuclear Transfer

tracrRNA	Trans-Activating crRNA
TALE	Transcription Activator-Like Effectors
TALEN	TALE Nuclease
VACV	*Vaccinia virus*
WR	Western Reserve
RVDs	Variable Di-Residues
WT	Wild Type
ZF	Zinc Finger
ZFN	Zinc Finger Nuclease

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

LIST OF APPENDIX A

SUPPLEMENTARY FIGURES

Page

LIST OF APPENDIX B

SUPPLEMENTARY TABLES

Page

LIST OF APPENDIX C

SUPPLEMENTARY MATERIAL

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

"They thought their classified species were more fixed and unchangeable than anything in heaven or earth that we can now imagine. We have learned that they are as plastic in our hands as clay in the hands of the potter or color on the artist's canvas, and can readily be molded into more beautiful forms and colors than any painter or sculptor can ever hope to bring forth".

Luther Burbank, 1901

1.1 History of Genetic Engineering.

"Appreciation of the history of a discipline helps shape its future." (1)

The field of genetic engineering (GE) began in the early 1970s, thanks to the development of the recombinant DNA technology (2), which allowed for the establishment of the first recombinant viral vector by Paul Berg in 1971 and the first genetically modified organism, a bacteria transformed with a recombinant plasmid conferring antibiotic resistance in 1973 by Stanley Cohen and John Morrow et al., (3, 4). This group was also the first to report first direct transfer of DNA between different kingdoms by expressing Xenopus genes in bacteria(4).

In 1974, the first genetically modified animal was produced by Rudolf Jaenisch and Beatrice Mintz (5) by microinjecting the whole DNA of the Sivian Virus - 40 (SV-40) into an expanded blastocyst and showing the presence of the transgene in the cells of healthy adult mice. Followed by Gordon in 1980 (6), were a recombinant plasmid containing DNA fragments of the SV-40 virus was microinjected into the pronuclei of fertilized mouse embryos, with later demonstration of the transgene transmission to subsequent generations in 1981 (7-9).

But the process of how the DNA was finding its way into the genome was still not understood until July of 1982, when Mario Capecchi published the intriguing observation that when many copies of a gene were inserted into the genome, the pattern of integration was the opposite of the expected random integration. He found that the genes were always clustered in one or a few regions, with many copies overlapping one another. Cappecchi had discovered that mammalian cells could undergo homologous recombination. At the end of his 1982 article he stated, “It will be interesting to determine whether we can exploit this machinery to "target" a gene by homologous recombination to a specific chromosomal location” (10). Just three years later, Oliver Smithies published a paper where they exploited this possibility to precisely target the human ß-globin locus and introduce a neomycin resistance gene in a predictable fashion, albeit with the planned modification occurring in one per thousand transformed cells (0.001%) (11). These were milestone papers that began the exponential growth in the field of genetic engineering.

By December of 1982, Richard Palmiter and Ralph Brinster published the establishment of the famous growth enhanced “gigantic mice” (12). Then in 1985, Robert Hammer created the first genetically modified farm animals (13). With these innovations, the field of genetically engineered animals opened its doors to the agricultural production, providing the tools for the introduction of new traits that could not be produced by selective breeding, or traits that would take several decades of breeding to achieve.

Pronuclear microinjection (PNMI) gave rise to the initial development of the field and allowed the production of many different GE animals including rodents, rabbits, pigs, sheep, goats, cattle, salmon and other fishes (13-16). However, the technology was still very inefficient (≤1% of injected zygotes will produce a viable transgenic animal) and did not allow for any control in the copy number, stable expression, or the integration site of the transgene (16). This was important because if the integration occurred within a heterochromatic region, the transgene was most often silenced. Further, it was later revealed that the regulatory elements of nearby genes affected the level of expression, or conversely, the transgene could interfere with the transcriptional control of critical genes impacting growth and development (16, 17). This pushed the field of genetic engineering to improve integration and expression efficiency, and it rapidly increased our understanding of the functional importance of regulatory elements such as promoters, enhancers and insulator sequences for reduced gene silencing and increased transcription of a transgene.

By 1983 Jack Szostack, Terry Weaver and Rodney Rothstein at Harvard Medical School, published a pioneering model based on experimental data on meiotic recombination in yeast that explained the mechanism behind homologous recombination and gene targeting. They proposed that the event that precipitated homologous

recombination was a double stand break (DSB) in the DNA, in which the free ends of the DNA at the site of the break were prone to fusing, making the flanking sequences far likelier to engage in the exchange of genetic information with the homologous chromosome, or in the case of gene editing, with the repair template provided by the scientist (18).

By the end of the 1980s seminal work in Martin Evans laboratory at the University of Cambridge, demonstrated that by targeting genes in mouse embryonic stem cells (ESCs) with retroviral vectors and then injecting those modified stem cells into mouse embryos, chimeric mice with the designed edits in the germ line could be obtained and could subsequently be used them for the derivation of transgenic strains (19). This milestone, conferred the possibility for having much more control in the type and complexity of the genetic modification. The breakthroughs by Capecchi, Smithies and Evans earned them the 2007 Nobel prize "for their discoveries of principles for introducing specific gene modifications in mice by the use of embryonic stem cells" (20).

Although this ESCs technology has been a huge success in the establishment of GE mice and rats, the culture conditions tested so far are not supportive of livestock ESCs self-renewal and proliferation making this technology unavailable for the making of GE livestock (21). In contrast, numerous reports of derivation of livestock induced pluripotent

stem cells (iPSCs) have been made with limited success (22-26). iPSCs were first derived in 2006 using mouse fibroblasts by overexpression of four key transcription factors: Oct4, Sox2, Klf4, and c-Myc (OSKM) by the Shinya Yamanaka group in Japan (27). The livestock iPSCs have demonstrated capacity for long term proliferation and in vivo pluripotency, as indicated by teratoma formation assay (22). However, to what extent these iPSCs represent fully reprogrammed PSCs remains controversial as most livestock iPSCs depend on continuous expression of reprogramming factors and robust germline chimerism has not been demonstrated. Although transgenic bovine chimeric offspring from embryonic in bovine embryonic stem-like cells (bESC) (28, 29) have been produced no germline transmission was demonstrated (29). In summary, alternative technologies were necessary to produce transgenic livestock since stem cells were not an option.

In 1994, Maria Jasin's laboratory at Memorial Sloan Kettering center in New York published a ground-breaking paper (30) that would lay the foundation of genome engineering using engineered nucleases. Her lab generated for the first time a specific artificial DSB in the mouse genome by using a rare-cutting endonuclease (I-SceI) and showed that gene targeted clones were nearly undetectable without the nuclease expression, while in contrast, the presence of the nuclease inducing artificial DSBs displayed an approximately 10% increase in homologous recombination. The implementation of this discovery was not quick because the recognition specificity of

meganucleases proved too difficult to tailor to desired target sites. Thus, generation of a targeted DSB remained the rate-limiting step in the development of HDR technology for genome engineering of plant and mammalian cells, including human cells (31).

In 1996 at the Roslin Institute, Keith Campbell and Ian Wilmut adapted the nuclear transfer technique for the use of mammalian somatic cells (SCNT) (Wilmut et al., 1997), based on the previous work that Jon Gurdon had done in 1962 using *Xenopus laevis* at Oxford (32-34). This milestone was established by the birth of "Dolly", the first mammal ever cloned from adult somatic cells. SCNT made possible the controlled insertion of precise genetic modifications in livestock by facilitating the verification of genetically engineered cell lines prior to SCNT. Using this technology 100% of born animals were transgenic. However, due to epigenetic reprograming and developmental issues, the abortion rates of cloned fetuses were high, approximately 17%, and there is also a high perinatal mortality, making the overall efficiency very low (<5%) and with a high cost of implementation(17, 35, 36).

The first proof of concept, showing the potential of this technology was provided by a collaboration between the Roslin Institute and PPL - Therapeutics, upon the birth of two transgenic sheep, 'Polly' and 'Molly', the first genetically engineered (GE) animals produced by using SCNT. Here, fetal fibroblasts that were modified to carry the human

blood factor IX under the control of the sheep milk ß-lactoglobulin (BLG) promoter and a neomycin resistance gene, were used as donor nucleus to generate these sheep (37).

By 1996, the laboratory of Srinivasan Chandrasegaran in Johns Hopkins University developed a novel site-specific hybrid restriction enzyme that fused two different zinc finger (FZ) proteins to the non-specific type IIs enzyme FolkI cleavage domain, a breakthrough that set the stage for genome engineering using programmable nucleases (38). After a few years of technology refinement, the Chandrasegaran lab started a collaboration with Dana Caroll that resulted in the use of the designed zinc finger nucleases (ZFN) to stimulate homologous recombination in frog oocytes and to edit the *yellow* gene in Drosophila. This was a profound development that showed the promise this technology had for precise genome editing in living organisms (39-41).

In spite of their promise, ZFNs were never widely adopted. Each ZF usually recognizes 3-bp sequence and binds DNA by inserting the α-helix into the major groove of the double helix; however, when there is an aspartic acid residue present at +2 position of the α-helix, it can enforce an adenine or a cytosine base outside the 3-bp site at the next base on the non-contact strand of DNA via a cross-strand contact, changing the ZFs recognition to a 4-bp site. This ZF contact outside the 3-bp site further influences the specificity of neighboring ZFs. Once it became apparent the recognition of DNA by the

ZFs was not truly modular, where each ZF recognizes a triplet sequence as one had expected and that each ZF's recognition was greatly influenced by its neighbors, this complicated the design. This meant the selection of each ZF had to be performed in a context-dependent fashion to obtain highly sequence specific ZFPs, which is laborious, time consuming and requires extensive experience in protein engineering (31). The ZFNs had proven that designer nucleases were the right tool to pursue, but the field still needed a new kind of technology that would be more reliable and straightforward to use.

This technology arrived in 2009 due to an accidental discovery made in Xanthomonas bacteria: a novel DNA-binding protein known as transcription activator-like effectors (TALE) (42). The TALE central repeat domain consists of repeating units of 33–35 amino acids, where each repeat is largely identical except for two highly variable amino acids at positions 12 and 13, referred to as the repeat variable di-residues (RVDs). While each ZF recognizes 3–4 bases, each TALE motif recognizes a single nucleotide, and the recognition specificity is determined by the RVD. More importantly, unlike the ZFs, the recognition of DNA by individual TALE modules appears to be largely independent of neighboring modules. The DNA recognition code thus provides a one-to-one correspondence between the array of amino acid repeats and the nucleotide sequence of the DNA target, which made them ideal for constructing custom nucleases (43).

TALENs development built largely upon the experience gained from ZFN development, which made the implementation of this technology much faster than ZFNs. But even though TALENs are easier to generate than ZFNs, the genes encoding TALENs are about three times larger than ZFNs and the TALE consensus sequences are invariant and highly repetitive in nature. This makes it more difficult to assemble the genes encoding for TALENs in E. coli and its delivery into mammalian cells is also problematic; finally, the initial commercial pricing for TALENs of ~$5000 per target, restricted it use. But only 3 years after the elucidation of the TALE recognition code, the CRISPR-Cas platform arrived on the scene.

1.2. CRISPR-Cas9

This technology had unpretentious origins. It was first reported in 1987 by Atsuo Nakayta in Japan, as an incidental finding of unknown biological significance, where odd repetitive sequences were found separated by unique spacer sequences of similar length and were clustered next to the *iap* gene of *Escherichia coli* K12 (44). By the 1990s, these regions were named interspaced direct repeats (DR), and the DR-intervening sequences, known as spacers, were found to differ among isolates and were harnessed for strain typing in the *Mycobacterium tuberculosis complex* (MTBC) (45, 46). By the late 1990s, thanks to the improvement of sequencing technologies, similar elements were found in archaea and bacteria. Bioinformatics analysis on the then called short regularly spaced repeats (SRSR) in a large number of unrelated microorganisms suggested a biological relevant

function (47). By 2002, the groups of Ruud Jansen in the Netherlands credited the proposal of Fransisco Mojica in Spain to unify the diversity of names and labels for these DNA repeated elements as CRISPR (clustered regularly interspaced short palindromic repeats), and at the same time a set of four genes in vicinity to the clustered repeats was detected and named CRISPR-associated (Cas) genes (cas1-cas4) (48).

In 2005, those three bioinformatics groups (49-51) reported that the spacer sequences matched the sequences of phages, suggesting a possible role for CRISPR in bacterial immunity. This was an important clue that allowed for a hypothesis to be formulated: CRISPR was somehow a bacterial immune system involved in the targeted recognition by CRISPR-RNA molecules (52), reminiscent of the eukaryotic interference RNA that defended the bacteria against viruses (50). One of these studies (51) was also the first to observe a short conserved DNA motif next to the protospacers, which would later be termed the protospacer adjacent motif (PAM) (53).

Two years later in 2007, this hypothesis was confirmed experimentally when the group of Rodolphe Barrangou and Philippe Hovarth at Danisco, in their landmark publication showed that they could alter the resistance of S. Thermophilus to bacteriophage infection of a susceptible strain by adding spacer DNA that matched the phage's genome (54). In 2008 Brouns et al., (55) demostrated that the small RNAs

produced by the CRISPR arrays (crRNAs) had a crucial role driving antiviral defense and that the DNA was the target of CRISPR action. This was subsequently confirmed by Marraffini et al., (56) by showing that CRISPR could interfere with the horizontal transfer of plasmid sequences through DNA targeting. Together these findings established CRISPR-Cas as an efficient adaptive immune system in prokaryotes.

Over the following years a detail characterization of the biochemical mechanism of action was progressively made. PAM sequences seemed important for the interference (57) and the motifs emerged as common features in many systems (53) until it was defined that the CAS protein cleavage occured next to the PAM (58). By 2011 the trans-activating crRNA (tracrRNA) was discovered and its crucial role for the generation of mature crRNA molecules was shown (59). That same year, the first classification of the CRISPR-Cas systems was made into three different functional and structural types (Type I, II and III) and subtypes (60). This was later changed to two Classes I (Type I, III and IV) and II (Type II and V) with 16 subtypes (61, 62).

By 2012, two independent teams published the biochemical characterization of in vitro characterized CRISPR-Cas systems (63, 64), and for the first time, both groups proposed that CRISPR-Cas could be used as a genetic engineering tool. The group of Jennifer Doudna in UC Berkeley and Emmanuelle Charpentier in Umeå University of

Sweden, showed that the *Streptococcus pyogenes* tracrRNA, crRNA and Cas9 could be used to create a precise DSB in a DNA sequence that matched the crRNA spacer. They simplified the system by making a chimeric synthetic RNA molecule between the tracrRNA and the crRNA, named single guide RNAs (sgRNA) which also directed sequence specific DNA cleavage (**Figure 1 A, B**), In parallel Sinsnys, Barrarangou and Hovart working with the *Streptococcus thermophilus* CRISPR system demonstrated similar results. Together, these two publications resulted in three independent US teams assessing in less than 6 months, the genome editing capabilities of the CRISPR system in bacteria by Luciano Marraffini group at Rockefeller University (65) and in mammalian genomes by the groups of George Church at Harvard Medical School (66) and Feng Zhang at MIT (67). These landmark papers triggered a wave of excitement throughout the Scientific community and CRISPR-Cas was widely adopted.

CRISPR-Cas technology is under a constant evolution and refinement; for example, researchers inspired by ZFNs and TALENs, have developed Cas9 variants from the Streptococcus pyogenes Cas9 (SpyCas9) for use as nickases, dual nickases or FokI fusion variants (68-72). Deactivated variants of Cas9 (dCas9) lacking DNA cleavage function allow for the fusion of many different domains. For example, sequence-specific transcriptional regulation can be made by fusing a transcriptional repressor (e.g., KRAB) or activator domain (e.g., VP64) to reduce or increase gene expression, respectively (73-

75). Epigenetic modifications have also been possible by the fusion of acetyltransferases and demethylases (76-78). Fusion of dCas9 to fluorophores, has enabled sequence-specific visualization of DNA, and dynamic imaging of chromatin and allowing up to 6 different loci using CRISPRainbow (79, 80). Crystallographic and electron-microscopy based structures of Cas9 have revealed the nature of its interactions with guide RNA and target DNA, which have been exploited for the engineering of Cas9 variants with altered PAM specificities (81) and lower off target effects (82). Transcriptional and functional control of Cas9-mediated genome editing has been made with optogenetics (83) and chemically inducible (doxycycline-regulated) approaches (84). Other Cas9 orthologues, such as the smaller *Staphylococcus aureus* Cas9 that allows for adeno-associated virus (AAV) packaging (85), and other nucleases derived from class II CRISPR–Cas systems including Cpf1 (86)and C2c1(87), have also been added to the CRISPR repertoire by bioinformatics mining of the abundant bacterial and archaeal CRISPR–Cas systems.

1.3. Applications of Gene Editing Technologies

Gene editing technologies have been used in many different fields such as research, therapeutic manufacturing, xenotransplantation, livestock and crop production, food and industrial microbes. This section aims to highlight the most important applications in each of the aforementioned fields.

1.3.1 Animal Disease Models for Research

The ability to generate targeted mutations in the mammalian genome has formed the backbone of genetic research since the establishment of the first knockout mouse by Capecchi, et al (88, 89). Since then, the mouse as a model organism has been at the forefront of genetic engineering and has led to many of the discoveries for the currently available drugs and treatments (90). However, a high proportion of clinical trials and consequently drug development efforts have been met with some unexpected failures. This is attributable in part to the mouse data not being translatable to human medicine, and in some cases a failure to replicate the symptoms associated with human mutations (91). This has pushed the development of new animal models, like the rat (92), rabbit, dog, pig (91), sheep (93) , goat, cow and primates that better phenocopy human diseases. For example, porcine models of Parkinson's disease and Von Willebrand disease have been made by a triple knockout of *DJ-1*, *parkin*, and *PINK1* genes (94) and a knock out of the *vWF* gene (95) respectively.

1.3.2 Agricultural Applications

Genetically engineered (GE) animals have been developed for several purposes, that target almost all aspects of animal agricultural production. Farm animals and fish have been genetically modified with the aim to enhance economically important traits (96). For example, increased growth rate has been attained in pigs (13, 97-100), Atlantic salmon

(101) and tilapia (102, 103) by modifying growth hormone expression. Cattle (104) and goats (105) lacking the prion protein have been made by knocking out the PrPC protein. Mastitis resistant cattle were produced by expressing in the mammary gland either the human lysozyme, lactoferrin or lysostaphin (106-109). An increase in meat quality was also accomplished in pigs with higher level of polyunsaturated fatty acids (PUFAs) in their meat (110, 111). Animals that are resistant to diseases have also been made, for example chickens that are resistant to the transmission of the influenza virus (112) and PRSSV resistant pigs (113, 114). Although not applicable yet to animals, there are efforts to make an organism resistant to all viruses by repurposing the genetic codons (115).

1.3.3 Biological Control Applications

Insects act as vectors of disease of plants and humans. The replacement of wild insect populations with genetically modified individuals unable to transmit disease, or carrying a lethal gene drive to kill the offspring, provides a friendly and self-perpetuating method for disease prevention (116). This can be done with gene drives, which are genetic systems that circumvent the traditional rules of Mendelian inheritance in sexually reproducing organisms (50:50), by greatly increasing the odds that the drive will be passed on to offspring, until an entire population bears that gene. These "selfish" genetic elements occur naturally in mice, beetles and many other organisms, and named homing endonuclease genes (HEGs) (117). The discovery of CRISPR-Cas9 brought gene drive

systems (GDS) from a remote possibility back to reality, via the making of efficient self-perpetuating gene drives that can be deployed in potentially any genomic location in any eukaryotic species (118). For example, the malaria transmitting mosquito *Anopheles gambiae* has been targeted with CRISPR-Cas gene drive systems (116).

1.3.4 Organ Transplant

The highly disproportional shortage of transplantable organs and the increasing number of patients added to the waiting lists, is causing that on average more than 20 persons die each day due to the lack of a transplant. This has pushed the search for new alternatives such as organ bioreactors and the use of pigs as organ donors for the xenogenic transplantation into humans, a process known as xenotransplantation. One approach is by eliminating the molecular immune incompatibility between the donor and the recipient, by genetically engineering the pig (119, 120). The other approach is to engineer the pig to grown human organs by doing embryo chimeric complementation with pluripotent stem cells (121).

1.3.5 Industrial Biotechnology and Therapeutic Manufacturing

The deficit in the bio-manufacturing capacity of therapeutic proteins and the increasing perceived viral risks associated with plasma derived products have led to the development of the production of recombinant therapeutic proteins by the use of genetic

engineering (122). This technology platform represents a powerful tool to address the growing needs of the biopharmaceutical market for a safer, more efficient and less costly product. It also enables the production of any type of protein with a physiologically relevant function. These proteins can be produced in various hosts, with multiple possibilities to design the protein product and also the production process. Currently recombinant proteins are the main end product of the biotechnology industrial pipeline and include protein hormones, monoclonal antibodies, protein based (sub-unit) vaccines and even spider silk (123-132). Different host have been used such as bacteria, yeast, insect, mammalian cells and transgenic animals for the production of recombinant proteins (127, 133-135). When choosing a system that is most suitable for any given protein target many different factors come into play such as post-translational modifications (PTM), cost, scale up easiness, total annual production, and speed of production (136).

1.4. Repair of DNA Double-Strand Breaks.

Gene editing technologies, including CRISPR-Cas9, rely on the precise introduction of a DNA double-strand break (DSB) at a precise location in the genome. DSBs are the most dangerous type of DNA damage, which if left unrepaired, can result in chromosomal translocations, the loss of large chromosomal regions or can activate cell cycle check-point arrests and induce signals for programmed cell death. In dividing mammalian cells, there are an estimated ten DNA double-strand breaks (DSBs) per day

per cell (137). These pathological DSBs arise from DNA replication errors, reactive oxygen species, and genotoxic agents such as ionizing radiation, radiomimetic chemicals or topoisomerase inhibitors (138). There are also physiological DSBs that arise during normal cellular processes, such as meiosis during gametogenesis or V(D)J recombination and class-switch recombination (CSR) which facilitate the rearrangements of antigen receptor genes in lymphogenesis (137, 138) (**Supplementary Figure 14**).

DSBs that occur throughout the cell cycle are repaired predominantly by the non-homologous DNA end joining (NHEJ) pathway that is present during the G1, S and G2 phases and to a lesser extent by the homology-directed repair (HDR) pathway that is restricted during the late S and G2 phases(139). The classical NHEJ (C-NHEJ) can be divided into five main stages: (I) The rapid binding of the Ku70-Ku80 heterodimer to free ends of DNA, which induces the recruitment and activation of the DNA-dependent protein kinase catalytic subunit (DNA-PKcs). The KU-heterodimer and the DNA-PKcs together form the DNA-PK complex. (II) The synaptic end bridging, that is dependent on the configuration of the DNA ends (which can include blunt ends, 5′ overhangs and 3′ overhangs), is carried out by the DNA-PK, where the broken DNA ends are kept in close proximity. (III) The DNA end processing is carried out because most DSBs have two incompatible DNA ends that preclude direct ligation, has many subpathways were end-processing factors (e.g. PNKP, TDP1), NHEJ polymerase (e.g Pol μ, Pol λ) and nuclease

(e.g Artemis) complexes, assure microhomolgy and prepares the DNA ends for (IV) ligation of the break by the XRCC4 -XLF-LIG4 complex (139-141) and finally (V) the Ku heterodimer removal from the restored DNA (**Figure 1**).

Structural and biochemical studies support a model in which the degree of microhomology (≤4 nucleotides) between the ends, direct repair subpathways; that is, which of the different sets of NHEJ proteins serve to align the two DNA ends in an end-to-end configuration (137). In order to create new microhomology ends, multiple rounds of resection and addition are possible, and nuclease and polymerase activities at each of the two DNA ends can act independently. The process is very error-prone and can result in diverse DNA sequences at the repair junction (139-141).

When NHEJ is compromised owing to the lack of one or more of its key protein components, the activity of the other end joining pathways becomes apparent. This typically involves amuch more extensive resection of the DNA ends to reveal sequence homology; the annealing of which stabilizes the two ends of a break to allow for more efficient joining and ligation (137, 142). This 5′ to 3′ endo- and exo-nucleolytic processing is performed by the carboxy-terminal binding protein interacting protein (CtIP) and the MRN (MRE11–RAD50–NBS1 (Nijmegen breakage syndrome protein 1)) complex, to generate stretches of single-strand DNA (ssDNA) 3’ overhangs. This is a very rapid (~30 minutes) (143) and highly regulated process by the cell cycle cyclin-dependent kinases

(CDKs) (137) by weakening the activity of two distinct blocks to resection: one mediated by the chromatin-binding protein 53BP1 (p53- binding protein 1), and a second embodied by HELB, an ssDNA translocase(144). These 3' overhangs are suitable substrate for two possible mechanisms: alternative end joining (A-EJ, also known as micro homology-mediated end joining (MMEJ) or Pol θ-mediated end joining), which uses small sequence homologies (~2-20 nt) and is 50-fold less frequent than NHEJ(145); and homologous recombination (HR), which relies on extensive homology (~20 - >100 nt). Neither of these pathways is reliant on Ku, and the binding of Ku to DNA ends may need to be attenuated for a-EJ and HR to proceed (137).

The HR pathway comprises of at least three sub-pathways: homology directed repair (HDR), single stranded annealing (SSA), and the recently recognized single stranded template repair (SSTR) (117, 146). Non-conservative homology-directed repair pathways (which involve the loss of nucleotides), such as SSA, requires >20 bp of homology (137, 147). The conservative HDR pathway, in which no nucleotides are lost, generally requires lengths of homology longer than 100 bp (148, 149). HR is a slow (~≥7h), but typically an error free process that uses the sister chromatid for repair during the late S and G2 phases of the cell cycle) (143). Mechanistically, replication protein A (RPA) binds to the ssDNA 3' ends and subsequently BRCA2 recruits the RAD51, which displaces RPA to form a RAD51–ssDNA nucleofilament recombinase, which is required

for homology search, strand invasion and ultimately the repair of the lesion (148). For heterochromatic regions that can engage in ectopic recombination during DSB repair, leading to chromosome rearrangements and widespread genome instability, the HDR repair relies on the relocalization of the DSBs to the nuclear periphery before Rad51 recruitment (150, 151).

The post replicative chromatin marking model is the simplest model for explaining HR activation during S phase and proposes that replicated chromatin is competent for HR whereas unreplicated chromatin remains refractory to this type of repair (144). The lack of methylation at the Lys20 residue of histone H4 (H4K20me0) represents a post-replicative chromatin mark as well as the presence of cohesin. H4K20me0 is recognized by MMS22L–TONSL, which promotes HR by promoting RAD51 loading in response to DNA replication stress (141, 144). In contrast, the presence of monomethylated and especially dimethylated, H4K20, marks of chromatin damage by DSBs that are induced by ataxia-telangiectasia mutated (ATM) monomers, are recognized by 53BP1 which rapidly accumulates on the chromatin surrounding the break site and promotes NHEJ and opposes HR (and A-EJ) at least in part by blocking DNA end resection (141). Single strand annealing, in contrast, mediates annealing between stretches of chromosome-internal homologies resulting in the loss of the intervening region, and is therefore considered an error prone repair pathway (117, 137, 147, 152, 153). Single stranded template repair has

not been fully characterized to date, but is known that requires multiple components of the Fanconi Anemia (FA) pathway, but is independent of BRCA2 or Rad51(146, 154) (**Figure 1 C-G**).

The high abundance of the Ku molecules in cells (KU70 1,290,000 and KU80 826,000 molecules per cell) (137) increases the likelihood that the Ku-heterodimer is the first protein to bind to a broken DNA end and, therefore, that repair is carried out through NHEJ. Although binding of DNA ends by Ku-heterodimer inhibits extensive resection by MRN (MRE11–RAD50–NBS1) complex and CtIP (C-terminal binding protein interacting protein) (155) favors repair by NHEJ, extensive resection is also dependent on the cell cycle owing to the action of cyclin-dependent kinases (CDKs). Factors that promote extensive end resection are more active during S and G2 phases and favor HR when a sister chromatid is present. This is another reason why repair by NHEJ is dominant throughout the cell cycle, whereas repair by HR and SSA is favored in S and G2 phases (156-158). Therefore, in G1 phase, NHEJ is favored by more than 50-fold for the repair of DSBs, owing to both the level of Ku and the suppression of extensive end resection by CtIP and MRN. Even in S and G2 phases, when extensive end resection can take place, the resection machinery must still overcome the presence of Ku at DNA ends either by outcompeting Ku for DNA-end binding or by processing the DNA ends to the point at

which Ku binding is less favored. The ratio of NHEJ to HR in wild-type mammalian somatic cells, even during S phase and G2 phase, is estimated to be 4:1 (137, 159)

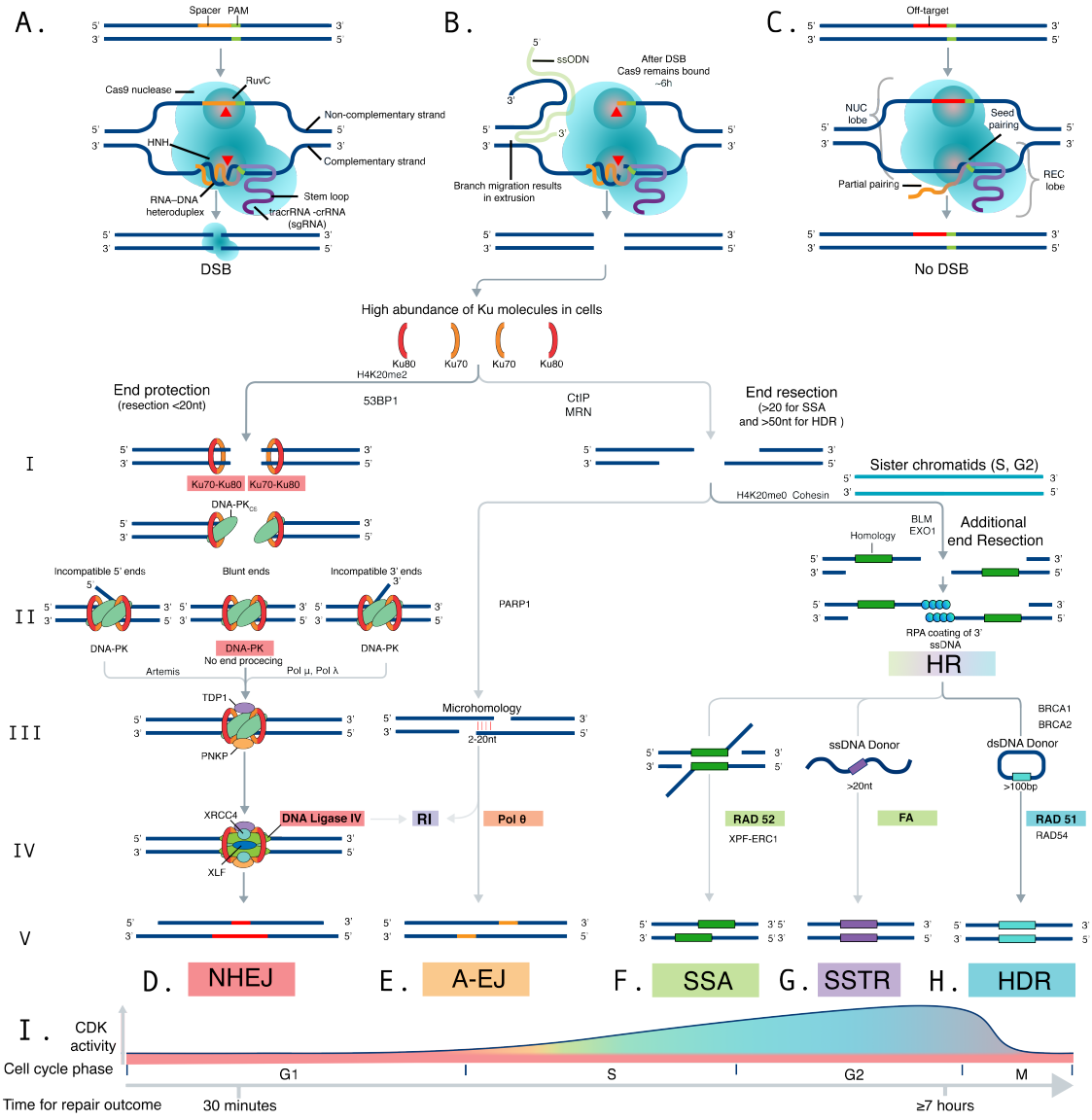
If Ku is absent (which is exceedingly rare in normal human tissues, as well as in neoplastic human tissues), a-EJ may be favored over SSA and HDR in G1 phase, owing to the limited amount of resection that a-EJ involves (137). It is still unknown what dictates the use of a-EJ versus SSA in S and G2 phases. However, time is likely to be a key determinant because the longer a DSB remains unrepaired, the more end processing can occur to generate longer 3′ ssDNA tails to favor SSA. Studies have shown that NHEJ can occur in approximately 30 minutes while HDR lakes 7h or longer, these efficiencies are strongly influenced by chromosomal location (143). Lastly, quantification of the relative ratio of various pathways is complicated because the absence of one pathway results in the accumulation of substrate for other pathway (137)

Finally, although unrelated with repair of DSBs, random integration (RI) of exogenous DNA hampers the precise engineering of genomes through recombination of exogenous DNA at unanticipated loci. RI does not rely on sequence homology, is more efficient than HDR, happens with no or little detectable sequence preference, results in insertional mutations, and is hence often referred to as “illegitimate recombination” (160). Although the mechanism of RI was unknown, recent publications describe that the dual

loss of Pol θ and C-NHEJ LIG4 eliminates off-target integration of exogenous DNA and demonstrates the role of A-EJ and C-NHEJ in RI. Here the term TMEJ (polymerase Theta-Mediated EJ) is proposed due to the notion that A-EJ may also encompass Pol θ-independent repair (160, 161).

Figure 1. CRISPR-Cas9 Introduction of DSB and DNA Repair Outcomes.

DNA double-strand breaks (DSBs) can arise in pathological (e.g ionizing radiation) and physiological conditions (e.g Meiosis), or by the genome editing technologies, like CRISPR-Cas. **A-C)** The *Streptococcus pyogenes* CRISPR-Cas9 system. Here Cas9 endonuclease (shown in blue) consists of a nuclease (NUC) lobe and a recognition (REC) lobe. Cas9 is targeted to specific DNA sequences by direct pairing of the chimeric single guide RNA (sgRNA), an engineered fusion between a crRNA (orange) and part of the tracrRNA sequence (purple), with the target DNA by forming a DNA-RNA heteroduplex. This targeting relies on the presence of a 5′protospacer-adjacent motif (PAM) in the DNA (green), which in S. pyogenes is usually NGG. After PAM recognition, Cas9 unwinds the DNA at the PAM-proximal region and creates a seed RNA-DNA heteroduplex pairing, next, the R-loop propagates via sequential unwinding to the PAM-distal regions (162). Only when extensive pairing has occurred (on-target), Cas9 suffers a conformational change in its HNH catalytic domain that triggers the RuvC domain catalytic activity ensuring a coordinated introduction of a DSB (163). While the HNH domain cleaves the gRNA complementary strand 3nt 5' of the PAM, the RuvC domain cleaves of the non-complementary strand in a variable location due to the ability for this strand to breathe in and out of the nuclease domain (164). The two domains can be mutated (i.e RuvC1 (D10A) and HNH (H841A)) to create a nuclease null dCas9. B) After the DSB Cas9-DNA complex remains bound to the cleaved products for ~6h partially releasing the PAM-distal non-target strand before complete dissociation (163, 165). **C)** If only partial pairing occurs there is no DNA cleavage and Cas9 remains unbound to the DNA. The DSB can be repaired by the D) classical non-homologous end joining (NHEJ) pathway, E) the alternative end joining (a-EJ) pathway, F) the single-strand annealing (SSA) pathway, G) single strand template reapair (SSTR) or H) by homologous recombination (HR). The major differences in pathway choice are the requirement for substantial DNA end resection. **D)** The high abundance of the Ku molecules in cells increases the likelihood that Ku-heterodimer is the first protein to bind to a broken DNA end and, therefore, that repair is carried out through NHEJ. p53-binding protein 1 (53BP1) is a chromatin remodeler and a positive regulator of NHEJ (137). Although the complex of Artemis and DNA-dependent protein kinase catalytic subunit (DNA-PKcs) can carry out some resection (typically <20 nucleotides), the NHEJ pathway does not require extensive end resection and the ends are mostly protected by the binding of Ku70–Ku80. The classical NHEJ (C-NHEJ) can be divided into five main stages: **I)** The rapid binding of the Ku70-Ku80 heterodimer to free ends of DNA, **II)** the synaptic end bridging, **III)** the DNA end processing, **IV)** ligation of the break by the XRCC4 -XLF-LIG4 complex (139-141) and finally **V)** the Ku heterodimer removal from the restored DNA. On the other side, the carboxy-terminal binding protein interacting protein (CtIP) and the MRN (MRE11–RAD50–NBS1 (Nijmegen breakage syndrome protein 1)) complex are involved in extensive 5′ to 3′ resection of regions of the duplex to generate stretches of single-strand DNA (ssDNA) at DNA ends for A-EJ, SSA and HR. **D)** SSA typically requires >20 bp of microhomology, whereas the requirement for **E)** a-EJ is <25 bp. Poly(ADP-ribose) polymerase 1 (PARP1) and DNA polymerase θ (Pol θ) are important for a-EJ**. F-G)** Bloom syndrome RecQ-like helicase (BLM) and exonuclease 1 (EXO1) provide additional resection, and replication protein A (RPA) binds to ssDNA to promote the SSA and the HR pathways. RAD52-mediated annealing of large regions of homology is key for the SSA pathway. The xeroderma pigmentosum group F (XPF)–ERCC1 complex cuts the remaining 3′ overhangs before ligation. **H)** By contrast, RAD51-mediated strand exchange and its association with BRCA1, BRCA2 and RAD54 are essential for promoting the HR pathway. XLF, XRCC4-like factor; XRCC4, X-ray repair cross-complementing 4. Figure adapted and modified with permission from Chang, et al., (137).


A.
Spacer
PAM
Cas9 nuclease
RuvC
HNH
Non-complementary strand
Complementary strand
RNA–DNA heteroduplex
Stem loop
tracrRNA -crRNA (sgRNA)
DSB
B.
ssODN
After DSB Cas9 remains bound ~6h
Branch migration results in extrusion
C.
Off-target
NUC lobe
Seed pairing
Partial pairing
REC lobe
No DSB
High abundance of Ku molecules in cells
Ku80
H4K20me2
Ku70
Ku70
Ku80
End protection (resection <20nt)
53BP1
CtIP
MRN
End resection (>20 for SSA and >50nt for HDR)
I
Ku70-Ku80
Ku70-Ku80
DNA-PKcs
Sister chromatids (S, G2)
H4K20me0
Cohesin
BLM
EXO1
Additional end Resection
Homology
II
Incompatible 5' ends
Blunt ends
Incompatible 3' ends
DNA-PK
DNA-PK
DNA-PK
No end procecing
PARP1
RPA coating of 3' ssDNA
HR
Artemis
Pol μ, Pol λ
III
TDP1
PNKP
Microhomology
2-20nt
BRCA1
BRCA2
ssDNA Donor
>20nt
dsDNA Donor
>100bp
IV
XRCC4
DNA Ligase IV
XLF
RI
Pol θ
RAD 52
XPF-ERC1
FA
RAD 51
RAD54
V
D. NHEJ
E. A-EJ
F. SSA
G. SSTR
H. HDR
I. CDK activity
Cell cycle phase
G1
S
G2
M
Time for repair outcome
30 minutes
≥7 hours

1.5 Manipulation of the DNA Repair Systems.

Precise genome editing not only requires the introduction of a DSB at an exact location in the genome, it requires the correct intended DNA repair outcome (18, 31). Inducing user-defined edits has become the major bottleneck in genetic engineering due to low rates of HR. Many efforts by several groups are taking place to increase HR rates, in general the strategies being taken can be classified into 4 groups: cell cycle synchronization strategies(158, 166), inhibition of the C-NHEJ repair pathway components(167-171), enhancement of the HDR pathway(172-174) and the rational design and use of ssODN donors, Cas9 and Cas9 variants(83, 84, 154, 164, 175, 176).

An example for the cell cycle synchronization strategies is the work by Lin et al,. (177) were they reported that Nocodazole synchronization of HEK293T cells enhanced the total editing frequencies more than twofold and HDR frequencies over six-fold when transfecting variable doses of Cas9 ribonuceoprotein complexes and donors.

Examples for the inhibition of the NHEJ is the use of the DNA-Ligase IV inhibitory compound SCR7. Chu et al., demonstrated that SCR7 increased HDR (up to 19-fold at a concentration of 1 μM) (167), other reports indicate a very high variability of this molecule between different experimental systems (158); or even that SCR7 is neither a selective nor a potent inhibitor of the human DNA ligase IV (178). The use of the

adenovirus 4 (Ad4) proteins 4E1B-E4orf6, which mediate the ubiquitination and proteosomal degradation of DNA ligase IV have also been reported (167). The co-expression of these proteins together with the CRISPR-Cas9 system using a 2A peptide was reported by Chu et al., (167) to improve HDR efficiency up to ~8-fold reaching HR frequencies of 50-66% and significantly decreased NHEJ activity in HEK293 and mouse Burkitt lymphoma-like cell lines. This same work reports the use of shRNA sequences to knock down KU70, KU80 or DNA ligase IV, which resulted in substantial suppression of the NHEJ and a ~5-fold improvement in HR. Following the same small molecule inhibition strategy, Robert el, al (170) reported the inhibition of the DNA-PK_{cs} using the small molecules NU7441 and KU-0060648, which caused a decrease of ~40 % in NHEJ events with a ~2-fold increase in HDR, although HDR rates were low ~4 % HR; they also show the additive effect of combining these two compounds with either RS-1, SCR7, siRNA suppression of Ku70 and Ku80 and adenovirus proteins E1B55K and E4orf6, these last being the most efficient inhibiting NHEJ ~8-fold and stimulating HR ~3.5 fold, in line Chu, et al., (167).

Examples for the enhancement of the HDR pathway is the use of RS-1 as a Rad51 stimulatory molecule identified by a small molecule screen where it was shown to stabilize association of RAD51 with DNA(179). Rad51 is a key molecule in the HDR pathway that displaces RPA to form a RAD51–ssDNA nucleofilament recombinase, which is required

for homology search, strand invasion and ultimately the repair of the DSB lesion (148). Song et al.,(172) reported RS-1 improves HR rates 2-5 fold at a concentration of 7.5μM in vitro and in vivo (from 4.4 to 26.1%). RS-1 was also shown to improve HR rates in HEK cells (from 3.5 to 21%) at and U2OS cells (from 1.9 to 2.4%) at a concentration of 10μM, but this varied depending on the locus and transfection method, with electroporation (~3-fold) having lower efficiency than with lipofectamine (~6-fold) (180).

Finally, an example of the rational design strategies is the report by the Corn group (164). After investigating the interaction of Cas9 with target DNA they discovered that Cas9 releases the PAM-distal non-target strand after cleavage, but before complete dissociation. They used this finding to rationally design a ssDNA that matches this strand, they increased HR rates with up to 60% HR in the absence of any chemical intervention (164). Similarly, the group of Liang et al., used asymmetric ssODN with PS modifications that had up to 56% HR in HEK293 cells (181). Another example is the work of Gutschner et al., (158) where a 1.87-fold increase was obtained when fusing Cas9 to the N-terminal region of human geminin. This was adapted by another group where they recently indicated that a Cas9- GFP-geminin enhanced the HDR/NHEJ ratio 2.7 fold in U2-OS and 1.8 fold in K562 cells and is now commercially available (182).

CHAPTER II

THERAPEUTIC GENOME EDITING IN THE HORSE

2.1 Abstract

The Glycogen Branching Enzyme Deficiency (GBED) is caused by a nonsense mutation (C > A) in the first exon of the *GBE1* gene that severely disrupts glycogen metabolism. This mutation is lethal in homozygotes and an estimated 9% of Quarter Horse and Paint Horse lineages are heterozygote carriers. Advances in the development of genome editing technologies have substantially improved our ability to make precise changes in the genomes of cells and gives us the opportunity to eliminate monogenic, highly penetrant diseases, such as the GBED mutation. To correct this mutation in a heterozygous cell line derived from a high genetic merit American Quarter Horse stallion, we used CRISPR-Cas9 to induce a double stranded break (DSB) to stimulate the DNA repair via homologous recombination (HR). The long-term goal is to use the corrected cell lines for somatic cell nuclear transfer (SCNT) by generating a cloned animal that maintains the genetic merit of its predecessor, but is free of the GBED mutation. To accomplish our objective, a series of sgRNAs flanking the mutation were cloned into the px458 plasmid (pSpCas9(BB)-2A-GFP), and then co-transfected with different single-stranded oligodeoxynucleotide (ssODN) repair templates. Forty-eight hours post transfection, cells were enriched for GFP+ by flow cytometry and plated at low density for clonal isolation and expansion. Distal sgRNAs (+44, +15, -13) co-transfected with and

without a symmetrical repair template showed variable insertions and deletions (INDEL) formation (8.3% - 81.8% n=124) but no HR was observed. Proximal sgRNAs (+1, -1) co-transfected with an asymmetric (67-30nt) ssODN with 5' and 3' phosphorothioate modifications showed 20.0% (n=20), and 23.1% (n=13) HR rates respectively. In order to decrease the possible number of off-target effects, truncated versions of these sgRNAs (+1T, -1T) were used, showing 4.3% (n=23) and 0.0% (n=13) HR, respectively. Interestingly, T-sgRNA -1 produced three HR positive colonies (20.0 %, n=15) when no repair template was present, this suggested that the homologous allele was used as the repair template. To verify this possibility, we targeted in an allele specific manner the non-mutated allele and we obtained one colony homozygous for the mutation (7.7%, n=13). These results demonstrate that the CRISPR-Cas9 system can be used to correct the GBED mutation in primary equine cell lines, even in some instances, without the need for a repair template.

2.2 Introduction

Humans have been performing phenotypic genetic selection on plants and animals for centuries in order to suit their needs. Although this form of breeding has been successful in the establishment and improvement of many different breeds of plants and animals, we have unknowingly co-selected for deleterious alleles that have been propagating in animal populations (183). One of these deleterious alleles is the Glycogen Branching Enzyme

Deficiency (GBED) which is caused by a nonsense mutation (C > A) in the first exon of the *GBE1* gene, that severely disrupts glycogen metabolism by decreasing the end points in the glycogen molecule for myophosphorylase to use as a substrate to liberate glucose (184). This mutation is lethal in homozygotes, due to the lack of GBE activity in liver, cardiac and skeletal muscle (184-187). An estimated 9% of Quarter Horse and Paint Horse lineages are heterozygote carriers, which have half of the normal tissue GBE activity (184-187). This disease is also a significant cause of second and third trimester abortion and foal mortality in the Quarter horse breed (185, 188, 189). The majority of the Quarter Horses are descendants from two stallions and many of the GBED carriers can be traced to at least one of them: the sire King P234 (185). Other horse breeds like Thoroughbreds have been screened for GBED without finding this genetic mutation (185-187, 190).

There are more than 6,052 identified genetic mutations linked to disease in humans and animals (191). Thanks to the advent of genome editing technologies based on programmable nucleases (31), it is now possible to make precise changes in eukaryotic genomes with the potential to correct monogenic diseases, such as the GBED mutation from affected cells(192), tissues(193) and eventually whole populations. This is the concept behind therapeutic genome editing, which arise out of the idea that instead of pursuing palliative care, the ideal therapy for monogenic diseases would be to develop a method that can directly correct the disease-causing mutations (194).

Since the early days of genome editing, it was realized that the precise targeting of a gene locus was hampered by the overall low frequencies, one per thousand transformed cells (0.001%) (11). However, this changed with the discovery of that a site specific double stranded break (DSB) in the DNA was the event that precipitated a particular locus to exchange genetic information with a homologous chromosome or provided repair template, a process known as homologous recombination (HR) (18). Since then, the field of genome editing has pursued the specific introduction of DSB at a particular locus with engineered nucleases (30): first with the zinc-finger nucleases (ZFNs) (31, 38), then transcription activator-like effector nucleases (TALENs) (42, 43), and more recently the CRISPR-Cas system (clustered, regularly interspaced, short palindromic repeats/CRISPR-associated protein) (64, 72, 195).

The CRISPR-Cas system was discovered as a natural adaptive immune system of prokaryotes, that protects them against the invasion of foreign viral nucleic acids (54). This biological discovery (44, 49-51) was subsequently engineered (64, 196, 197) as a precise and efficient genome engineering tool (72, 195, 198). The most commonly used *Streptococcus pyogenes* Cas9 (spCas9) is a type II CRISPR-Cas system (62) that uses a dual RNA molecule composed of the tracrRNA and the crRNA to create a double stranded break (DSB) at a precise location that matches the unique 20nt sequence in the crRNA,

requiring only the presence of a 5'-NGG protospacer adjacent motif (PAM). This dual RNA system, can be simplified for an engineered single chimeric small guide RNA (sgRNA) (64). After PAM recognition, Cas9 unwinds the DNA at the PAM-proximal region and creates a seed RNA-DNA heteroduplex pairing, then the R-loop propagates via sequential unwinding to the PAM-distal regions (162). Only when extensive pairing has occurred (on-target), Cas9 undergoes a conformational change in its HNH catalytic domain that triggers the RuvC catalytic domain activity ensuring a coordinated introduction of a DSB (163). While the HNH domain cleaves the gRNA complementary strand 3nt 5' of the PAM, the RuvC domain cleaves of the non-complementary strand in a variable location due to the ability for this strand transition in and out of the nuclease domain (164). Finally, Cas9-DNA complex remains bound to the cleaved products for ~6h partially releasing the PAM-distal non-target strand before complete dissociation (163, 165). Finally, DSBs are sensed by the cells endogenous DNA repair system (137, 139, 141, 153) which repairs the DSBs predominantly by the canonical non-homologous end joining (c-NHEJ) pathway and to a lesser extent by the HR pathway, that comprises at least three sub-pathways: homology directed repair (HDR), single stranded annealing (SSA) and the recently recognized single stranded template repair (SSTR) (117, 146) (**Figure 1**).

There are many approaches for therapeutic genome editing in humans and animals. For example, by using somatic therapy, it is now possible to cure the acquired immune deficiency syndrome (AIDS), by taking a patient own CD4 T cells and using ZFNs to target the CCR5 gene that encodes a co-receptor required by most strains of HIV-1 to infect T cells(199, 200) and recently, the first chimeric antigen receptor (CART) T-cell gene therapy was approved by the FDA in the United States (201, 202). In addition, research for the germline genome editing of human embryos is taking place (203). But in the field of veterinary medicine, we have a unique and powerful tool that is not available in human medicine: reproductive cloning (204, 205). Producing cloned gene edited animals from edited cell lines has been successfully employed in the past (37, 110, 123, 124, 205-215). More recently, hornless edited cattle were produced by the introgression of the polled gene into Holstein breed primary cells by using TALENs followed by reproductive cloning (216, 217).

In this study, we utilized the CRISPR-Cas9 system to successfully correct for the first time, the GBED mutation in a heterozygous cell line derived from an American Quarter Horse stallion of high genetic merit. The long-term goal of our work is to use these corrected cell lines for somatic cell nuclear transfer (SCNT), generating a cloned animal that maintains the genetic merit of its predecessor, but is free of the GBED mutation (**Figure 3**).

2.3 Materials and Methods

2.3.1 Primary Fibroblast Establishment and Culture

A skin biopsy was collected in an aseptic manner under local anesthetic from an American Quarter Horse Stallion and donated to our laboratory by the horse owner. The sample was placed in 1x Ca and Mg free DPBS (Life Technologies®) and transported on ice. The sample was washed in a 0.2% (v/v) chlorhexidine gluconate in DBPS for 5 seconds and then in two DPBS washes. The sample was placed in a 10 cm petri dish and it was cut into small pieces (<5 mm). The tissue sample was then washed through a series of DPBS washes in a 15 ml conical tube; for this, the sample was mixed by inverting the tube, the tissue was allow to briefly sediment and the supernatant media was aspirated. The tissue was re-suspended in Dulbecco's Modified Eagle Medium with nutrient mixture F-12 (DMEM/F12) (Life Technologies®) supplemented with 10% fetal bovine serum (FBS), 1x antibiotic-antimycotic (Anti-Anti™, Thermo®) and 0.05mg/ml gentamicin (Life Technologies). The tissue sample was finally placed in a T25 tissue culture flasks (Falcon) and cultured at 37°C in a 5% CO_2 and 5% O_2 humidified incubator (Nuaire) to reduce oxidative stress and prolong the Hayflick limit. Cells were passaged when 80% confluence was reached, in a split ratio of no more than 1 to 3.

2.3.2 Karyotyping

A normal karyotype integrity is essential for SCNT procedures (218). For this, primary and edited cell were sent for karyotypic evaluation to Dr. Terje Raudsepp at the Texas A&M Laboratory of Molecular Cytogenetics and Genomics. GTG-banding (G-banding) was performed by trypsin treatment followed by Giemsa staining (**Supplementary Figure 4**).

2.3.3 Short Guide RNA Design

sgRNAs were designed using different bioinformatics web based tools follSowed by a manual sequence verification. The different programs used were the Broad institute CRISPRko which uses the on-target scoring described by Doench, et al., (219), and the WU-CRISPR tool of the Washington University described by Wong, et al., (220). All DNA sequences were manipulated using Benchling® (221) and APE. RNA folding was assessed using the Mfold web server for nucleic acid folding and hybridization prediction. Each guide and its complementary sequence was ordered as synthetic 25nmole oligos from Thermo® with attached BbsI cloning sites: Sense: 5' – CACCGNNNNNNNNNNNNNNNNNN – 3' and antisense: 3' – CNNNNNNNNNNNNNNNNNNCAAA – 5' (**Figure 2, Supplementary Figure 1**).

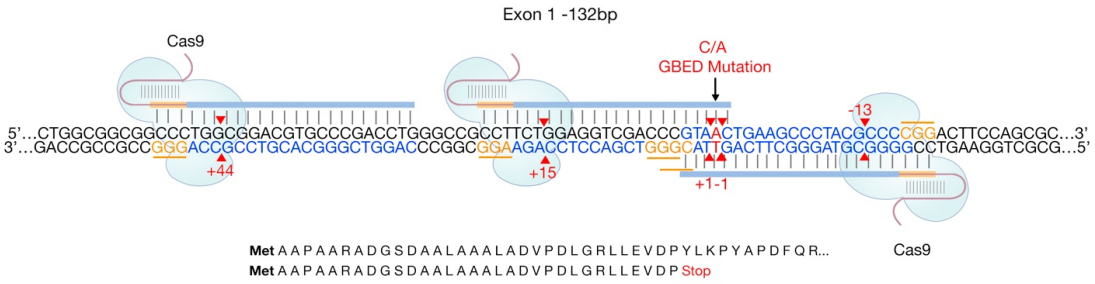


Figure 2. GBE Locus and CRISPR-Cas9 Targeting Strategy.
Schematic illustration of the GBE1 locus with the different sgRNA/Cas9 targeting complexes used. The C to A mutation at base 102 is indicated by the black arrow; this mutation causes a premature stop codon (bottom). The Cas9 enzyme shown in blue is directed by the sgRNA (composed of the crRNA and the tracrRNA linked by a loop) is shown in red and the 20nt guide sequence in blue. Upstream of the guide sequence, lies the protospacer adjacent motif (PAM), shown in orange. Cas9 mediates the DSB approximately 3 bp upstream of the PAM, denoted by the red arrow heads for each guide used (i,e. +44, +15, +1, -1, -13).

2.3.4 Short Guide RNA Cloning Into the PX458 Plasmid

The designed guides were cloned into the px458 (pSpCas9(BB)-2A-GFP) plasmid (Addgene, #48138). Synthetic oligonucleotides were suspended in water to 100µM and then 10µM of oligos was phosphorylated using T4 PNK (NEB) at 37°C for 1 hour. A 0.5µM mix of both oligos, was annealed by heating at 95°C for 5 minutes, oligos were allowed to cool slowly at room temperature. The px458 plasmid was digested with fast digest BbsI (Thermo®) for 1 hour at 37°C, dephosphorylation was carried out with rSAP (NEB) for 30 minutes and the plasmid was run on a 1% agarose gel, followed by gel purification (Qiagen®). Ligation was performed with 50ng of the linearized vector and 1µl of the annealed oligos (0.5µM) using Quick ligase™ (Roche®) for 1h at room temperature.

One Shot TOP10 competent *E. coli* cells (Thermo®) were transformed following manufacturer instructions and plated in 100µg/ml Ampicillin (Sigma) LB agar plates followed by an overnight incubation at 37°C. Insert was verified by colony PCR using the sense oligo as a forward primer and a common Cbh reverse primer: 5' GTCAATAGGGGGCGTACTTGG 3', at a 50°C annealing using the HiFi PCR premix (Clonetech®). A Maxi prep (Qiagen®) was performed followed of a final verification by Sanger sequencing using the LKO forward primer for the human U6 promoter: 5' GACTATCATATGCTTACCGT 3'.

2.3.5 Repair Templates Design

All repair templates were designed by hand using the web application Benching. All single stranded oligonucleotide (ssODN) repair templates were ordered PAGE Ultramer® DNA Oligos from IDT® and suspended to a 100µM concentration in DNAse free water (**Supplementary material 1**).

2.3.6 Transfection

In order to improve attachment and viability after transfection, plates were pre-coated with 0.1% gelatin in 1xDPBS and incubated at 37°C for 1 h. Early passage cells were passed as described above and plated at density of 300,000 cells per well in a 6 well plate. Cells were cultured for at least 36h until an 80% confluence was reached. Cells were

transfected using Lipofectamine 3000™ (Thermo®). For this, each well was transfected with 1µg of CRISPR-Cas9 plasmid (Px458 with SgRNA) and 3µl of the 10mM ssODN (~600ng/µl). Media was replaced 12h after transfection.

2.3.7 Flow Cytometry

After verifying GFP expression under an inverted fluoresce microscope (Nikon eclipse TE300) (**Figure 4 B, C**), cells were trypsinized and were suspended in transport media (10% FBS in DPBS, 1xAnti-Anti and 1mM EDTA). Cells were transported in ice to the Texas A&M CVM flow cytometry CORE facility. Before the flow sort, 2.5ng/ml of propidium iodide (PI) was added to the cells as a viability stain. Cells were sorted for single, PI - and GFP +. Cells were sorted into 45% FBS, 45% DMEMF12, 10% conditioned media from confluent healthy cells, 1x Anti- Anti and 100µM Y-27632, ROCK inhibitor (Stemcell®). Cells were plated at a density of 1000 cells per 150 mm dish for subsequent single cell colony isolation.

2.3.8 Single Cell Colony Isolation

After ~10 days of culture single cell colonies were recovered using agarose embedded cloning rings for single cell clone isolation as described by Mathupala, et al., (222). Briefly, single colonies were quickly marked with a permanent marker under a stereo microscope on the bottom surface of the plate. The plates were placed in the tissue

culture hood and cloning cylinders carefully placed around clones with sterile curved forceps; then one percent (w/v) LMP agarose (Sigma®) in 1xDBPS (37°C) was then slowly dispensed dropwise around the outside of the cloning cylinders. Cells were lifted with 40µl of 0.05% trypsin-EDTA (Invitrogen®) at 37°C for 5 minutes and then 100µl of normal culture media was added to inactivate the trypsin and pipetted several times to detach and suspend cells before moving each colony to a 48 well plate. Subsequent passages to larger wells were made as the cells became confluent. Cells from 6 well plates were passaged and half were frozen in 10% DMSO, 45% FBS, 45% DMEMF12 media using an isopropyl alcohol freezing container (Thermo®). The other half was used for DNA extraction.

2.3.9 Colony Genotyping

DNA was extracted using the DNeasy Blood & Tissue Kits (Qiagen®) with the addition of 1µl of tRNA (10µg/µl) for improved DNA yield. DNA was eluted in water and the concentration was measured using the nanoDrop™ spectrophotometer (Thermo®). PCR was performed with approximately 200ng of genomic DNA by using 2.5 units of the the HotStarTaq plus DNA polymerase (Qiagen®) with 1x Q solution, 100µM dNTpMix, and 0.1µM of each forward and reverse primers in a 50µl reaction. The PCR was run at an annealing temperature of 57^{o}C with an extension time of 30s for 40 cycles. PCR products were purified using the QIAquick™ PCR purification kit (Qiagen®) and eluted

with water. Samples were sent to the Texas A&M Institute for plant Genomics and Biotechnology for Sanger sequencing reaction. Forward primer: 5' CTCGCCGCTATAAAGGGCCCC 3', reverse primer: 5' TGCGCTGGAAGTCCGGGG 3'.

2.3.10 Sanger Sequencing Data Analysis.

All chromatogram files were aligned in Benchling® (**Supplementary Figure 2**). The web application CRISPR-ID was used to de-convolute the overlapping spectra from the Sanger sequencing of PCR products. These overlapping spectra arise from the random C-NHEJ DNA repair of the CRISPR-Cas9 induced DSB. Because of the single cell origin of the colonies and the allele specific nature of the sgRNA (+15, +1, -1), the exact sequence of the resulting alleles can be identified using this tool. The sequence of a colony of sgRNA +15 with INDELS is shown (**Figure 4, I**). After de-convoluting the reads fasta files were converted manually to fastq files, then fastq groomer was run on galaxy before submitting them as single end reads to the web application CRISPResso, that was built for the analysis of CRISPR-Cas9 genome editing outcomes from deep sequencing data (223). The following parameters were used. Amplicon sequence: GCCGCCTTCTGGAGGTCGACCCGTAACTGAAGCCCTACGCCCCGGACTTCCA Expected HDR amplicon sequence:

GCCGCCTTCTGGAGGTCGACCCGTACCTGAAGCCCTACGCCCCGGACTTCCA. Sequence homology for an HDR occurrence: 95%, sgRNA sequence/s: GTAaCTGAAGCCCTACGCCC. Window size (bp around each side of cleavage site) to quantify NHEJ edits: 30 Exclude bp from the left side of the amplicon sequence for the quantification of the mutations: 5 Exclude bp from the right side of the amplicon sequence for the quantification of the mutations: Disabled (**Supplementary Figure 3**).

2.3.11 Microsatellite Analysis and RT-PCR Taqman Assay.

Genomic DNA from gene edited, non-carrier single cell colonies identified by Sanger sequencing and control carrier samples were sent for independent TaqMan DNA typing and carrier status verification to the Texas A&M Genetics Laboratory. A total of 13 microsatellite markers (AHT4, AHT5, ASB17, ASB23, HMS6, HMS7, HTG4, VHL20, HMS3, ASB2, HTG10, LEX33 and HTG6) specific to *Equus caballus* were used (224, 225). All markers are included in the panel recommended by the International Society for Animal Genetics. GBED carrier status was determined by RT-PCR TaqMan Assay. Forward primer: CCTGGGCCGCCTTCT. Reverse primer: GCGCTGGAAGTCCGGG. VIC probe: CCCGTACCTGAAGCC. FAM probe: CCCGTAACTGAAGCC.

2.3.12 BAC DNA Isolation and BAC End Sequencing.

GBE1 containing BACs 93G22 and 112C9, verified by PCR, were set grown in 50 ml tubes with LB media and chrloramphenicol overnight at 37°C in a shaker incubator. Two different kits were used: the Qiagen® midiprep and the PacBio® High Pure Plasmid Isolation kit with modifications. Briefly the Qiagen® kit manufacture instructions were followed with an isopropanol precipitation and resuspension performed instead of the filter based purification. The PacBio® kit was performed in triplicate combining the triplicate extraction in one column. DNA quality assessment was performed in gel and quantified in a NanoDrop™ spectrophotometer (Thermo®).

Bac end sequencing was performed in a 10.5µl reaction mix of 2 µl of Big Dye™, 10 µM Primer, 1µl of BAC DNA and 10µl of MasterAMP™. The reaction was at 96°C for 2 minutes, followed by 8 cycles of 96°C for 30s, 50°C 30s, 65°C 4 min, 60 cycles of 96°C 30s, 50°C 30s, 60°C 4 minutes and 4 cycles of 96°C 1 min, 50°C 1min and 65°C for 15 min. BioMax™ spin-50 columns were used to purified sequencing reactions and sent to be resolved in a sequencer. The following set of BAC end primers were used:

TAMU and INRA M13 Reverse: 5'-CAGGAAACAGCTATGACC-3'

TAMU and INRA T7: 5'-TAATACGACTCACTATAGGG-3'

CHORI-241 T7-29: 5'-GCCGCTAATACGACTCACTATAGGGAGAG-3'

CHORI-241 SP6: 5'- CCGTCGACATTTAGGTGACACTATAG-3'

2.3.13 Fluorescent In-Situ Hybridization.

Fluorescent in-situ hybridization (FISH) was performed with the help of Dr. Terje Raudsepp as described (226). Briefly, two slides one with the 93G22 BAC + 112C9 BAC isolated with the Genopure™ Plasmid Maxi Kit (Roche®) and the second with the 112C9 BAC + 93G22 BAC isolated with the HiSpeed™ plasmid maxi kit (Qiagen®), were detected conjointly with the control probes ETSTY7 bio + ETSTY dig (Y chromosome) with 3 layers of avidin-FITC; biotinylated anti-avidin; avidin-FITC Digoxigenin – 1 layer of anti-dig Rhodamine. Both 93G22 and 112C9 map to the same (overlapping) region in chr26. Control probes mapped to the Y chromosome as expected. Both BACs corresponded to the region of interest.

2.3.14 Statistical Analysis.

All experiments were performed using three or more independent biological replicates. Unless otherwise indicated, statistical analyses of categorical variables were conducted using Pearson's chi-square with Fisher's exact test (227-229). For viability and fluorescence intensity analyses, after verifying for the assumptions of equal variance and normality, P values were calculated using One-Way ANOVA with Tukey's HSD (230). Unless otherwise indicated, error bars represent standard deviation. Analyses were

performed with Prism™ (Graphpad®). Values with different subscripts are different ($P<0.05$).

2.4 Results

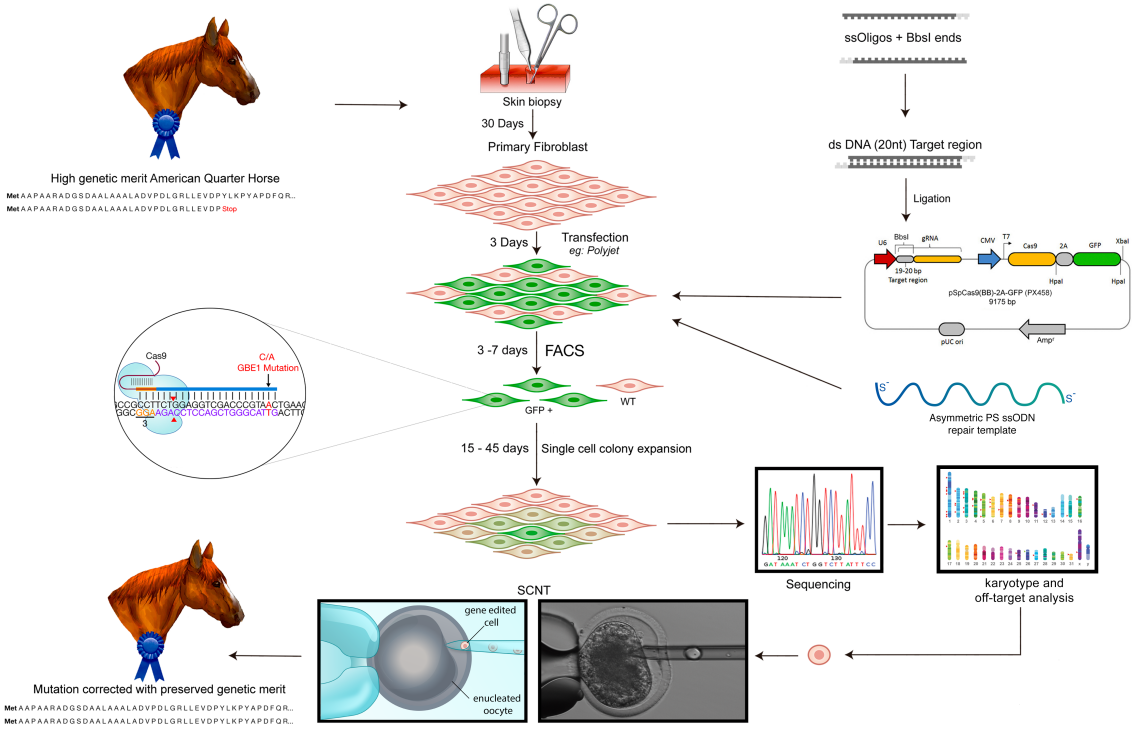


Figure 3. Schematic Illustration of the Experimental Design.
Schematic illustration of the experimental design. A skin biopsy was recovered from a high genetic merit American Quarter Horse stallion and a primary fibroblast cell line was established and cryopreserved. Different sgRNAs at a varying distance from the mutation were designed and cloned into the px458 plasmid (Addgene®) which allows for GFP selection. Different single-stranded oligo DNA (ssODN) repair templates (IDT®) were tested. Cells were co-transfected with the px458 plasmid and the repair template. Cells were incubated for 72H in four treatment groups. Cells were subsequently selected by flow cytometry for GFP positive and PI negative cells and plated at very low density in order to recovered single cell colonies. DNA was extracted for PCR amplification of the target region and sequenced for verification of the genetic correction.

We first established a karyotypically normal primary fibroblast cell line from an American Quarter Horse stallion (**Figure 4 A**) and a created a small bank of early passage cells. sgRNAs that flanked the mutation at a variable distance were designed (**Figure 2**) and cloned them into the CRISPR-Cas9 px458 plasmid (Addgene). Cells were transfected (**Figure 4 B, C**) and sorted by FACS for PI- and GFP+ cells (**Figure 4 D**). Cells were plated at low density and after ~10 days of culture, single cell colonies were recovered by using agarose embedded cloning rings, described in (222) (**Figure 4 E, F**).

Due to the primary nature of our cells, the initial viability of the cells after FACS, measured as the number of total colonies divided by the number of plated cells was low 6.96%±1.71 and was significantly lower 3.18%±0.89 ($p<0.001$) when a phosphorothioate (PS) modified ssODN was used but remained similar when a regular ssODN was used 4.95%±0.73 (**Figure 4 G**). This decreased the number of colonies that could be sampled per replicate. The Rho-associated protein kinase (ROCK) among other functions activates the caspase signaling cascade leading to cellular apoptosis (231). To improve viability, we added 100μM of the Rho-associated protein kinase (ROCK) inhibitor (Y27632) to our recovery media and observed an approximate 6-fold increase in the number of viable colonies, for a total average viability post FACS of 16.03%±0.89 (**Figure 4 H**).

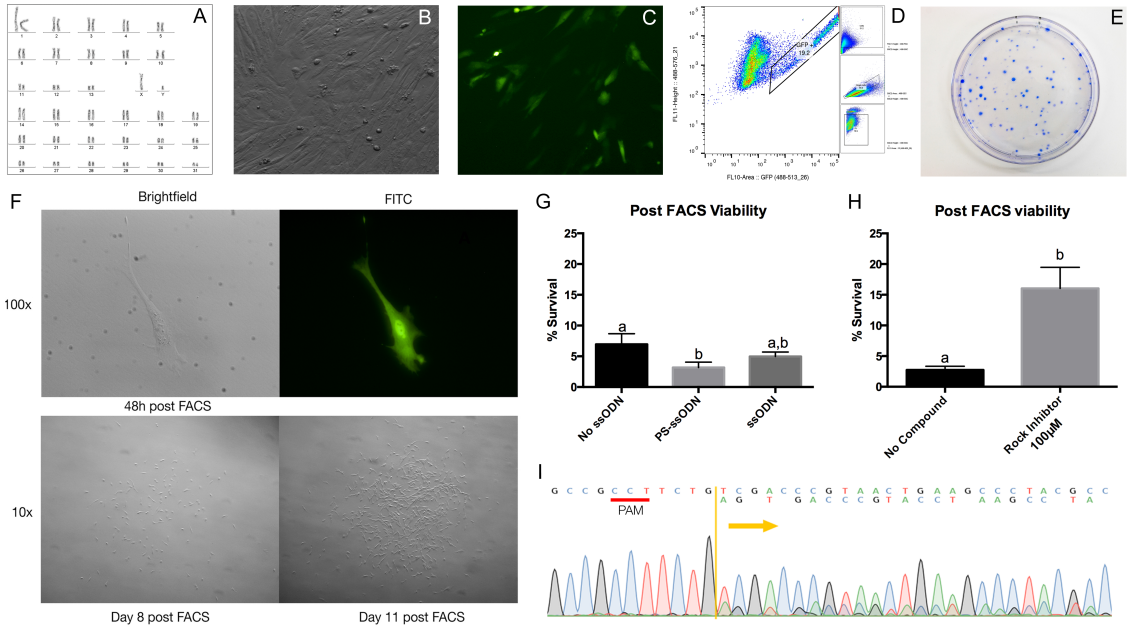


Figure 4. Primary Fibroblast Cell Line Establishment and Isolation of Single Cell Edited Colonies.

A) Normal karyotype of equine primary fibroblasts. B) bright-field and C) fluorescent image 12h after transfection CRISPR-Cas9. D) fluorescent activated cell sorting (FACS) enrichment for GFP. E) Single cell colonies 15 days after of FACS, stained with Coomassie blue. F) Isolation of single cell colony after FACS enrichment and low-density plating. G) Phosphorothioate (PS) modified single-stranded oligodeoxynucleotide (ssODN) repair template decreased cell survival, but H) an apoptosis inhibitor (ROCK inhibitor Y27632) improves viability of cells. I) Chromatograph showing a shift in reading frame 4 bp 5' of the PAM of sgRNA+15 demonstrating Cas9 mediated DNA allele specific cleavage. One way ANOVA, different letters signify statistical differences $p < 0.05$.

Figure 5. Gene Editing Outcomes from Designed sgRNAs.
A) Indel formation of sgRNAs flanking the GBED mutation without a repair template. B) Total events with modifications and C) Percent of INDELS (p=0.5047) and HR (p=0.1256) of proximal sgRNAs (+1,+1T,-1,-1T) and their truncated versions with an asymmetric ssODN. Values with different subscripts are different Chi-Square (P<0.05).

Table 1. Distal sgRNAs With and Without a Symmetrical ssODN.

	+ ssODN RT			- ssODN RT		
sgRNA	**n Sequenced**	**% INDEL**	**% HR**	**n Sequenced**	**% INDEL**	**% HR**
+44	16	18.8[a]	0.0	12	8.3 [a]	0.0
+15	33	81.8[b]	0.0	28	85.7 [b]	0.0
-13	19	15.8 [a]	0.0	16	18.8 [a]	0.0

We observed variable sgRNA efficiency, with the sgRNA +15 being the most efficient and sgRNA +44, -13, +1, -1 showing similar level of activity (**Figure 5A**). Distal sgRNAs (+44, +15, -13) co-transfected with and without a symmetrical repair template showed variable insertions and deletions (INDEL) formation (8.3% - 85.7% n=124) but no HR was observed (**Table 1**).

Due to the high level of activity of sgRNA+15, together with the unidirectional characteristic of the INDEL tracts, that extended over the GBED mutation site (**Supplementary Figure 2**), we hypothesized that the use of the RAD-51 stimulatory compound RS-1 (15μM), the DNA-Ligase IV inhibitory compound SCR7 (80μM) or their combination with either the previously used symmetric ssODN (**Table 2**) or a silenced PAM ssODN (**Table 3**). No HR events obtained and similar INDEL rates were observed between groups.

Table 2. RS-1 and SCR7 Effect on sgRNA +15 With ssODN.

Treatment	n Sequenced	% INDEL	% HR
No Compound	33	81.8	0.0
15μM RS-1	36	86.1	0.0
80μM SCR7	43	79.1	0.0
Combined	43	81.0	0.0

Table 3. RS-1 and SCR7 Efect on sgRNA +15 With Mutated PAM ssODN.

Treatment	n Sequenced	% INDEL	% HR
No Compound	17	70.6	0.0
15μM RS-1	6	83.3	0.0
80μM SCR7	10	90	0.0
Combined	9	88.9	0.0

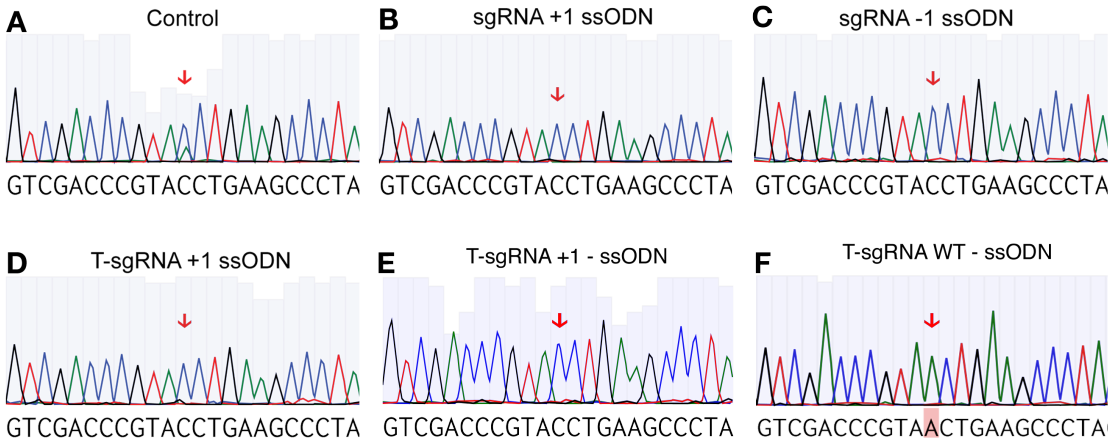


Figure 6. Sequence Verification of Genetic Correction of the GBED Mutation in Multiple Isolated Cell Lines.

Subset of Sanger sequencing chromatograms of PCR fragments from the isolated single cell colonies. A) Control unmodified colony: red arrow indicates the double peak showing the C (blue) to A (green) conversion in one of the alleles, representing the heterozygous state of the horse. B), C) and D) subset of colonies of sgRNA +1, -1 and Truncated +1, respectively, with phosphorothioate modified single stranded oligonucleotide repair template (ssODN). E) Colony of truncated sgRNA -1 without a repair template. F) Colony of the truncated sgRNA targeting the WT allele without a repair template. In the corrected colonies (B-E) there is an absence of the double peak observed in the control colony (red arrows) and the height of the both adjacent cytosines read peaks is approximately the same, indicating the correction of the mutation in the mutated allele.

Table 4. Proximal sgRNAs With and Without an Asymmetric PS-ssODN

	+ Asymmetric PS-ssODN			**- Asymmetric PS-ssODN**		
sgRNA	**n Sequenced**	**% INDEL**	**% HR**	**n Sequenced**	**% INDEL**	**% HR**
+1	20	30.0	20.0	21	33.3	0.0
+1T	23	30.4	4.3	7	57.1	0.0
-1	13	15.4	23.1	8	37.5	0.0
-1T	12	41.7	0.0	15	40.0	20.0

CO-transfection of the proximal sgRNAs (+1, -1) with an asymmetric (67-30nt) ssODN with 5' and 3' phosphorothioate modifications (181) resulted in, 20.0% (n=20), and 23.1% (n=13) HR rates respectively Figure 5**C, B**). To decrease the possible number of off-target effects (232), truncated versions of these sgRNAs (+1T, -1T) were used, showing 4,3% (n=23) and 0.0% (n=13) HR, respectively (**Table 4, Figure 4 C , B**). Interestingly, T-sgRNA -1 produced three HR positive colonies (20 %, n=15) when no repair template was present. This suggested that the homologous allele was used as the repair template. To verify this possibility, we targeted, in an allele specific manner the non-mutated allele and we obtained one colony homozygous for the mutation (7,7%, n=13) (**Table 5**). No significant differences were found. Representative chromatograms are shown in **Figure 6**.

Table 5. Targeting of the Wild Type Allele

		- **ssODN**		
sgRNA	**n Sequenced**	**% INDEL**	**% HR**	**% R**
WT	20	5.0	0.0	0.0
WT - T	13	7.7	7.7	50.0

Finally, analysis of sequencing chromatograms showed that INDEL tracts in the GBED locus were unidirectional, these tracts extended at least 50 bp from the DSB for sgRNA +15 and the INDELS introduced by Cas-9 were non-random with at least ~50%

of the events having a repeatable outcome (**Supplementary Figure 2).** Furthermore, the allele specific nature of the proximal sgRNAs allowed us to do a more in-depth analysis of the genome editing outcomes by using the web application CRISPResso (223). We found on a subset of the events of sgRNA+1 (n=7), that 7.7% corresponded to HDR events together with some other mutation, and that the majority of the INDEL events were substitutions, followed by deletions and insertions (**Supplementary Figure 3)**.

2.5 Discussion

Many genetic engineering strategies incorporate DSBs in close proximity to the target region to be repaired (233-235). This arose from the seminal study by Maria Jasin on DSB repair tracts (236), where it was reported that 80% of the repair tracts were ≤58bp. In primary equine fibroblasts, under the conditions of this study, only the sgRNAs targeting the mutation site in very close proximity (+1, -1) were capable of producing HR events (**Table 4**). This is in agreement with other published reports. However, at the GBE1 locus the allowable range for sgRNA recognition seems to be much narrower, as we did not obtain HR with sgRNA-guided Cas9-cutting greater than 13nt away from the mutation. For sgRNA -13, this could be due to the low DSB efficiency. We did not observe any HR events, even when we used a very active sgRNA (+15) with a symmetrical ssODN, PAM silenced ssODN.

The use of the RAD-51 stimulatory compound RS-1 (15μM) and the DNA-Ligase IV inhibitory compound SCR7 (80μM), at the concentrations tested, did not provide any benefit in improving HR rates (**Table 2, 3**). It is important to note that the concentration of SCR7 in the original publication of Chu, et al., had not been corrected at the time of the experiment. The concentration in the initial report of SCR7 was later corrected to 1 μM (167), however this molecule has been highly controversial and other reports indicate a very high variability of this molecule between different experimental systems (158) or even that SCR7 is neither a selective nor a potent inhibitor of the human DNA ligase IV (178). In contrast, RS-1 as a Rad51 stimulatory molecule identified by a small molecule screen where it was shown to stabilize association of RAD51 with DNA(179). Rad51 is a key molecule in the HDR pathway that displaces RPA to form a RAD51–ssDNA nucleofilament recombinase, which is required for homology search, strand invasion and ultimately the repair of the DSB lesion (148). Song et al, (172) reported RS-1 improves HR rates 2-5 fold at a concentration of 7.5μM in vitro and in vivo (from 4.4 to 26.1%) . RS-1 was also shown to improve HR rates in HEK cells (from 3.5 to 21%) and U2OS cells (from 1.9 to 2.4%) at a concentration of 10μM, but this varied depending on the locus and transfection method(180). It is important to note that the HR repair with a ssODN has been recently referred to as single stranded template repair (SSTR) (117, 237) and it has been identified that is a RAD51/BRCA2 independent pathway (154) so the type of donor being

used is one of the possible explanations for why we did not observe any benefit of the RS-1 molecule.

In contrast to the work by Maria Jasin, Hollywood et al.; (238) reported that after analyzing CRISPR-Cas9 repair tracts in the human CFTR gene, they found that 90% of the editing events in the CFTR locus were long continuous repair tracts in excess of 100bp from the DSB with no bias towards bidirectional or unidirectional correction. With this finding, they created sgRNAs that induced a Cas9 DSB 100bp away from the target and obtained 1.9 % HR events when providing a repair plasmid that harbored seven nucleotide differences with ~2 kb homology arms. They suggest that there is a~200 bp window in which to select gRNAs for template-dependent editing. We found that INDEL tracts in the GBED locus were unidirectional, for sgRNA +15 which had a very high INDEL efficiency ~85 % (**Figure 4 A, Table 1**), these tracts extended at least 50 bp from the DSB (**Supplementary Figure 2**). We were not able to employ long flanking homology arms because of the to the lack of the 3' intronic sequence. Despite numerous attempts to either amplify or sequence from verified BAC clones, we were not able to characterize this sequence, possibly due to the high GC content surrounding this region. This illustrates the importance of well annotated and sequenced genome (239) for the successful gene editing of agricultural and companion animals.

The INDELS introduced by Cas-9 were non-random with at least ~50% of the events having a repeatable outcome (**Supplementary Figure 2).** This is in line with the work of Overbeek, et al., (240) where after studying the DNA repair patterns at 223 sites in the human genome they demonstrated that the distribution of INDELS resulting from repair of Cas9-mediated DSBs is nonrandom and is composed of contributions from the C-NHEJ and A-EJ pathway with the protospacer sequence and not the genomic context, determining the outcomes. Other studies (241) have found that HDR/NHEJ ratios were highly dependent on gene locus, nuclease platform, and cell type.

A single stranded donor of an approximate length of 90nt was reported to produce high HR rates by Yang, et al., (234). This finding was later confirmed by the group of Liang, et al., (181), who reported that an asymmetric ssODN (67-30) with PS modifications had up to 56% in HEK293 cells. A mechanistic explanation was provided by Jacob Corn's group. Which suggested that after investigating the interaction of Cas9 with target DNA, Cas9 releases the PAM-distal non-target strand after cleavage but before complete dissociation. They used this finding to rationally design an asymmetric ssDNA that matches this strand, they increased HR rates with up to 60% HR in the absence of any chemical intervention(164)

In our work, to improve viability, we added 100μM of the Rho-associated protein kinase (ROCK) inhibitor (Y27632) to our recovery media and observed an approximate 6-fold increase in the number of viable colonies, for a total average viability post FACS of 16.03%±0.89. Other groups have reported that in primary porcine cells, without the combined use of a p53 inhibitor, PFTα, and a growth factor, bFGF, it was not possible to isolate single cell modified colonies(119).

Besides the proximity and type of repair template, another, and much more interesting possibility for the lack of HR with the distal sgRNAs at the GBE1 locus, is the genomic context and the chromatin state at this locus. It has been shown that Cas9 diffusion and chromatin binding is reduced but not eliminated at heterochromatic regions (242) and that nucleosomes, in fact, directly impede Cas9 binding and cleavage, while chromatin remodeling can restore Cas9 access (243). Miyaoka et al., showed that depending on the genomic context more HDR than NHEJ can be induced and viceversa (241). The repair of the DSB is also influenced by the chromatin context, although the repair of heterochromatic DSB is just starting to be understood, it is known that pericentrometric heterochromatin relies on the relocalization of repair sites to the nuclear periphery before Rad51 recruitment and repair progression (150, 151). Due to the tissue specific expression of the GBE1 enzyme, is likely that in primary fibroblast, this gene is silenced, which is

normally enriched with marks like H3K27me3 that are associated to the lamina-associated domains (LADs) for which DNA DSBs repair mechanisms are still unknown.

2.6 Conclusions

In conclusion, we were able to correct the GBED mutation with moderate efficiency (~20%) when using proximal sgRNAs cutting within 1 nucleotide of the intended correction site. We have also shown that sgRNA INDEL efficiency does not always correlate with the incorporation of HR events, even in the presence of small molecule compounds that have been reported to improve HDR. We showed that INDELS resulting from repair of Cas9-mediated DSBs are nonrandom and unidirectional in the GBE1 locus. Finally, we show that in rapidly dividing equine fibroblasts, when using allele specific sgRNAs, the use of the homologous chromatid for the repair of a DSB can be more common than we initially thought.

2.7 Acknowledgments

We would like to thank the help of Mathew Snyder, Ciccera Lazzarotto, Dalton Horn, Nicolas Moreno, Megan Mikkelson, Haley White, James Oldeschulte, Camilo Hernandez, Carly Turner and Cassandra Skenandore for their help in different aspects of this project.

CHAPTER III

IMPROVING THE EFFICIENCY OF HR USING C16

3.1 Abstract

Precise genome editing requires the introduction of a double stranded break (DSB) at an exact location in the genome and the correct DNA repair outcome. Although CRISPR-Cas9 has allowed for the introduction of precise DSBs in a very efficient manner, the lack of control over cell-autonomous repair mechanisms namely non-homologous end-joining (NHEJ) and homologous recombination (HR), is still the major bottle neck for seamless genome editing. The DNA-dependent protein kinase (DNA-PK), composed of the Ku 70 - Ku 80 heterodimer and the DNA-PK catalytic subunit (DNA-PK_{cs}), is best known as the NHEJ molecular sensor for DNA damage, but has been also identified as a pattern recognition receptor (PRR) that defends against the invasion of foreign nucleic acids (244). Here we devised a novel strategy that capitalizes on the natural ability of the *vaccinia virus* (VACV) C16 protein that evolved as an elegant subversion mechanism to inhibit the detection of the VACV genome by the host cytoplasmic PRR defenses, specifically the Ku-mediated DNA sensing (245). We evaluated the effects of this protein in a HEK293 BFP to GFP conversion assay. By localizing C16 into the nucleus we were able to obtain a ~ 2-fold increase (~10% HR) for HDR rates when they were low (~5% HR). After optimization of our assay for higher HR rates (~20% HR) we no longer saw a

benefit, and we believed were masked by the optimized conditions and the use of a ssODN that uses the SSTR and not the HDR pathway for repair. Future spatiotemporal refinements of this protein may enable highly efficient genome editing.

3.2 Introduction

The CRISPR-Cas system (clustered, regularly interspaced, short palindromic repeats/CRISPR-associated protein) was discovered as a natural adaptive immune system of prokaryotes that protects against the invasion of foreign viral nucleic acids (54). This biological discovery (44) was subsequently engineered (64, 196, 197) as a precise and efficient genome engineering tool (72, 195, 198). The most commonly used *Streptococcus pyogenes* Cas9 (spCas9) is a Type II CRISPR-Cas system (62) that uses a dual RNA molecule composed of the tracrRNA and the crRNA, to create a double stranded break (DSB) at a precise location. The target, matches the unique 20nt sequence in the crRNA, requiring only the presence of a 5'-NGG protospacer adjacent motif (PAM). This a dual RNA system, can be simplified for an engineered single chimeric small guide RNA (sgRNA) (64). After PAM recognition, Cas9 unwinds the DNA at the PAM-proximal region to create a seed RNA-DNA heteroduplex pairing. Next, the R-loop propagates via sequential unwinding to the PAM-distal regions (162). Only when extensive pairing has occurred (on-target), Cas9 undergoes a conformational change in its HNH catalytic domain that triggers the RuvC domain catalytic activity ensuring a coordinated

introduction of a DSB (163). While the HNH domain cleaves the gRNA complementary strand 3nt 5' of the PAM, the RuvC domain cleaves the non-complementary strand in a variable location due to the ability for this strand to breathe in and out of the nuclease domain (164). Finally, Cas9-DNA complex remains bound to the cleaved products for ~6h partially releasing the PAM-distal non-target strand before complete dissociation (163, 165). These DSBs are then identified by the cell's endogenous DNA repair system (137, 139, 141, 153) (**Figure 1 A-C**).

DSBs are the most dangerous type of DNA damage, which if left unrepaired, can result in chromosomal translocations, the loss of large chromosomal regions or can activate cell cycle check-point arrests that can induce signals for programmed cell death. In dividing mammalian cells, there are an estimated ten DNA double-strand breaks (DSBs) per day per cell (137). These pathological DSBs arise from DNA replication errors, reactive oxygen species, and genotoxic agents such as ionizing radiation, radiomimetic chemicals or topoisomerase inhibitors (138). There are also physiological DSBs that arise during normal cellular processes, such as meiosis, V(D)J recombination and class-switch recombination (CSR) which facilitate the rearrangements of antigen receptor genes in lymphogenesis (137, 138) (**Supplementary Figure 14**).

Organisms have evolved sophisticated DNA damage response pathways that sense the DNA lesion and activate repair mechanisms to correct the genomic insult (153, 154). DSBs are repaired predominantly by the canonical non-homologous end joining (c-NHEJ) pathway that is present throughout the cell cycle with peak activity during the G1, S and G2 phases (139). This is the main repair pathway due to the high abundance of the Ku70 and Ku80 molecules in cells (137), which increases the likelihood that the Ku-heterodimer is the first protein to bind the broken DNA ends, and therefore repair is carried out through NHEJ. This pathway can either result in a perfect repair by ligation of the two ends, or depending on the degree of microhomology (≤4 nucleotides) between the ends (112), different repair sub pathways are directed that can create multiple rounds of end resection and addition. This is a very error-prone process that can result in diverse DNA sequences at the repair junction, here on referred to as insertions and deletions (INDELS) (114-116).

A subset of the DSBs undergo 5′ to 3′ endo- and exo-nucleolytic processing by the carboxy-terminal binding protein interacting protein (CtIP) and the MRN (MRE11–RAD50–NBS1 (Nijmegen breakage syndrome protein 1)) complex, which lead to the generation of stretches of single-strand DNA (ssDNA) 3’ overhangs. This is a very rapid (~30 minutes) (143) and highly regulated process, controlled by the cell cycle cyclin-dependent kinases (CDKs) (137). These 3’ overhangs are suitable substrate for two possible mechanisms: alternative end joining (A-EJ, also known as micro homology-

mediated end joining (MMEJ), Pol θ-mediated end joining or alternative NHEJ), which uses small sequence homologies (~2-20 nt) and the homologous recombination (HR) pathway, which relies on extensive homology (~20 - >100 nt). Neither pathway is reliant on Ku, and the binding of Ku to DNA ends may need to be attenuated for a-EJ and HR to proceed (112).

The HR pathway comprises at least three sub-pathways: homology directed repair (HDR), single stranded annealing (SSA) and the recently recognized single stranded template repair (SSTR)(117, 146). HR is slow (~≥7h), but is normally an error free process that uses the sister chromatid for repair during the late S and G2 phases of the cell cycle)(143). Mechanistically, replication protein A (RPA) binds to the ssDNA 3' ends and subsequently, BRCA2 recruits the RAD51, which displaces RPA to form a RAD51–ssDNA nucleofilament recombinase. This nucleofilament is required for homology search, strand invasion and ultimately the repair of the lesion (148). For heterochromatic regions that can engage in ectopic recombination during DSB repair, leading to chromosome rearrangements and widespread genome instability, the HDR repair relies on the relocalization of the DSBs to the nuclear periphery before Rad51 recruitment(150, 151). Single-strand annealing, in contrast, mediates annealing between stretches of chromosome-internal homologies resulting in the loss of the intervening region, and is therefore considered an error prone repair pathway (117, 137, 147, 152, 153). Single

stranded template repair has not been fully characterized to date but is known to require multiple components of the Fanconi Anemia (FA) pathway but is independent of BRCA2 or Rad51 (146, 154) (**Figure 1 D-H**).

Precise genome editing not only requires the introduction of a DSB at an exact location in the genome but most importantly, requires the correct intended DNA repair outcome (18, 31). Inducing user-defined edits has become the mayor bottleneck in genetic engineering due to low rates of HR occurrence. Many efforts by several groups are taking place to increase HR rates. In general the strategies being taken can be classified into 4 groups: cell cycle synchronization strategies (158, 166), inhibition of the C-NHEJ repair pathway components(167-171), enhancement of the HDR pathway (172-174) and the rational design and use ssODN donors, Cas9 and Cas9 variants (83, 84, 154, 164, 175, 176, 246). There has been only moderate success and HR rates over 50% have just been recently been achieved (154, 167). It is likely that the best HR rate improvements will be achieved using combinatorial approaches, employing many of these strategies (247).

It is well documented that viruses have evolved proteins that counteract host detection mechanisms by binding and inhibiting signaling molecules. The DNA-dependent protein kinase (DNA-PK) , composed of the Ku 70 - Ku 80 heterodimer and the DNA-PK catalytic subunit (DNA-PK_{cs}), is best known as the molecular sensor for DNA damage(137) and

was recently identified as a pattern recognition receptor (PRR), which activates the innate immune response against DNA viruses(244, 245). As such, viruses have evolved strategies to inhibit host defense mechanisms and circumvent detection. The *Vaccinia virus* (VACV) is a dsDNA virus that is a member of the genus *Orthopoxvirus* and of the family *Poxviridae*, and is best known to be the live vaccine used to eradicate smallpox (248). The VACV encodes numerous proteins that inhibit the host innate immune system and reprogram cellular biochemistry to favor viral replication in the cytoplasm (245, 249). The C16 protein evolved as an elegant subversion mechanism to inhibit the detection of the VACV genome by the host cytoplasmic pattern recognition receptors (PRRs) defenses, specifically the KU-mediated DNA sensors (245). This results in the inhibition of the assembly of the DNA-PK complex and its stimulation of the innate immune pathway(245) (**Figure 1 D**).

In this study, we generated a novel genetic engineering strategy that capitalizes on the natural ability of the vaccina virus C16 protein to inhibit the binding of the KU heterodimer molecules. In these experiments, we hypothesize that we could re-localize C16 from the cytoplasm to the nucleus and use it to interfere with the C-NHEJ pathway by obstructing the KU-mediated DNA DSB sensing and subsequent DNA-PK_{CS} recruitment (245). We show the importance of the spatiotemporal expression of C16 with respect to CRISPR-Cas9 for its activity. Our findings show that C16 can be used to

enhance HDR rates up to 2-fold, and future spatiotemporal refinements may enable a more efficient genome editing.

3.3 Materials and Methods

3.3.1 Vector Design

In order to compartmentalize C16 in the nucleus and assess its efficiency, we generated a mCherry fusion construct that allowed for different plasmid versions to be generated by standar molecular cloning. For this, the 311aa protein coding sequence of the *Vaccinia virus* (VACV) strain Western Reserve (WR) C16 protein (GenBank: AGJ92502.1) and its cDNA sequence was determined by using the NCBI GenBank®. All the DNA sequences were worked using Benchling®. Codon optimization for homo sapiens was made by using the IDT codon optimization software. Correct translation of the constructs was made by using the ExPASy® translate tool software. In order to create a nuclear compartmentalization of C16, 3 repeats of the nuclear localization signal (DPKKKRKV) were attached to either the N or C terminus of the fusion protein with either Bglll and KpnI flanking sites respectively. In order to visualize the nuclear compartmentalization of the C16 protein, a fusion protein was made by attaching codon optimized mCherry protein to the C terminus of the C16 protein by a Gly_4Ser_2 linker flanked with EcoRI sites. The sequence was ordered as synthetic vector from Vector Builder® (**Figure 7**).

In order to create the different variants of the C16 vector (Vector Builder®) described above, (**Figure 7, A-G**) standard molecular cloning was used. For this a total of 1 µg of DNA of the C16 vector (986 ng/µl) was digested using either Kpn I or Eco RI in CutSmart™ buffer (NEB®) or Bgl II in buffer 3.1™ (NEB®) and the reaction was incubated at 37ºC during 3 hours. Digest products that were going to be double digested (Bgl II digest of KpnI digest product and Eco RI of Kpn I digest product) were purified with the QIAquick™ PCR purification kit (Qiagen®). The digested products were run on a 0.5% agarose (Invitrogen®) gel during 3h at 90V, appropriate bands were cut (**Supplementary Figure 5, 6**) and purified using the QIAquick™ gel extraction kit (Qiagen®). All DNA concentration in samples were measured using NanoDro™ spectrophotometer (Thermo Scientific®).

The ligation of ends was made as follows, in order to calculate the amount of vector and insert needed, the molarity of ends for both with a 3:1 ratio was calculated (Molarity = [(µg/µl) ÷ (base pairs x 650 daltons)] x 2 ends), the ligation was made with using the Rapid DNA Ligation kit (Roche®) in a 21µl total volume. One Shot™ TOP10 Chemically Competent *E. coli* (Invitrogen®) were transformed using 2µl of the ligation reaction. Briefly, the ligated DNA was chilled on ice for 5 minutes, 25µl previously ice-thawed One Shot™ TOP10 Chemically Competent E. coli (Invitrogen®) was gently added to the DNA,

the mixture was incubated for 30 minutes on ice, then the bacteria were subjected to a heat shock at 42 ºC for 30 seconds and returned to ice for 2 minutes, 300µl of S.O.C media (Invitrogen®) was gently added and the bacteria was incubated at 37ºC in a shaking incubator during 1 hour, and transformed cells were plated down in 100µg/ml ampicillin (Sigma®) LB - agar plates. After a 12h incubation at 37ºC colonies were selected and inoculated in 5 ml of LB-Broth with 100µg/ml of Ampicilin, incubated overnight in a 37ºC shaking incubator. The plasmid DNA was purified using a QIAprep™ miniprep Kit (Qiagen®) following manufacture instructions. DNA concentration was measured using NanoDrop™ spectrophotometer (Thermo Scientific®).

A restriction fragment analysis was made in order to verify the sequence of the corrected clones, once verified we proceeded to do a maxi prep isolation of the plasmid DNA for each of the variants of the C16 constructs using the DNA HiSpeed Plasmid Maxi kit (Qiagen®). DNA concentration was measured and concentrated up to 1µg/µl using a SpeedVac concentrator (Thermo Scientific®). All isolated vectors (**Figure 7**) were verified by Sanger sequencing using the primers listed in (**Supplementary table 1**); alignment is shown in **(Supplementary Figure 7)**.

3.3.2 Generation of HEK293 BFP Cell Line

In order to create a genetically modified cell line with a unique genomic integration of the blue fluorescent protein gene (BFP), lentiviral particles expressing BFP and puromycin under the EF1a and SV40 promoters respectively were used (Sigma®).For this, HEK 293T cells were seeded at 3 x 10^5 cells/well in 6-well plates and maintained for 18 h and the supernatant was then replaced with 1 ml of antibiotic free 20% FBS DMEM-12 with 100µM Y-27632, 8µg/ml hexadimethrine bromide and diluted lentiviral particles containing 3 x 10^5 TU/ml, 1.5 x 10^5 TU/ml, 3 x 10^4 TU/ml, for a MOI of 1, 0.5 and 0.1 respectively, followed by incubation overnight[1]. 48h post-transduction cells were analyzed and sorted by flow cytometry for BFP+ and PI- into 96 well plates and 5 ml tubes (**Supplementary Figure 8)**.

To identify the best concentration of puromycin to make a gradual selection to insure high levels of the transgene expression, HEK 293 were cells seeded in a 48 well plate at a concentration of 5 x 104 cells per well. Puromycin dihydrichloride (Thermo®) stocks were made in water to a concentration of 1mg/ml. Puromycin stocks were diluted 1.5ml of media for testing a range of concentrations (10,8,6,4,2,1,0.1,0 µg/ml) in triplicate wells. Cells were analyzed for cell dead by FACS 48h after the treatment. Transduced HEK 293 T cells were subject to a gradual puromycin selection starting at 0.5µg/ml 48h

after flow cytometry up to 2µg/ml changing the media every 72h in 0.5µg/ml increments (250).

A commonly encountered problem in cell lines produced by lentiviral transduction is heterogeneity expression of the vector-encoded transgene obtained in cell populations; this is due to copy number and position effects conferred by different integration sites (250, 251). Furthermore, retroviral vectors are often subject to transcriptional silencing shortly after transduction or extinction (progressive silencing of an initially expressing vector), through known mechanisms of gene repression, such as DNA methylation and histone modification, as well as through uncharacterized mechanisms. In order to have a homogenous expression cell line, the HEK 293T cells originally transduced with 1.5×10^5 TU/ml, sorted and subject to gradual puromycin selection, were subject to a second round of flow cytometry were cells were gated for a homogenous medium - to high expression of the BFP transgene (**Supplementary Figure 8, B**). Sorted cells were recovered, expanded and cryopreserved.

To have an accurate targeting of the BFP sequence, a PCR reaction was performed with the following primers: Forward 5' CCTGAAGTTCATCTGCACCACC 3', Reverse 5' CCATGCCCGAAGGCTACGTC 3'; using the CloneAmp™ HiFi PCR Premix (Clonetech®). Samples were sent to the Laboratory for Plant Genome Technologies

(LPGT) from the Institute for Plant Genomics and Biotechnology of Texas A&M University.

3.3.3 ddPCR for Copy Number Quantification of GFP Transgene

DNA samples were taken to the Texas A&M Institute for Genome Science and Scociety. Briefly, the ddPCR Master mix was performed following the manufacturer instructions using a reaction mixture in a final volume of 20 µL with 20 ng of the genomic DNA as the template. Restriction digest was performed in the ddPCR reaction using the enzyme HAEIII (NEB). Each reaction mixture was then loaded into a DG8 cartridge (Bio-Rad) with 70 µL of droplet generation oil to generate a droplet. The droplets from each well were then transferred into a 96-well PCR plate. The plates were heat-sealed and then thermally cycled under the following conditions: 37°C for 20 min (one cycle), 95°C for 10 min (one cycle); 40 PCR cycles of 94°C for 30 seconds and 60°C for 1 min; followed by one cycle of 95°C degrees for 10:00 minutes and a hold at 4°C. After PCR, the plates were placed on a QX200 droplet reader (Bio-Rad®) that analyzed the droplets of each well of the plate and quantified the target DNA. The PCR data were analyzed using QuantaSoft (Bio-Rad®) to determine the copy number variation (CNV). Copy numbers were calculated based on droplet numbers of the target, (BFP) compared to the single copy reference (RPP30) (e.g. BFP/RPP30: ~ 1 = Mono-allelic, ~ 2.0 = Bi-allelic, ~ 3.0 = Tri-allelic). Data represents two HEK 293 negative controls and four technical replicate

samples of genomic DNA extracted from FACS sorted GFP+ cells together with a no template control (NTC) as a negative control (**Supplementary Figure 9**).

3.3.4 Short Guide RNA Design and Cloning

sgRNAs Targeting the BFP protein chromophore region were designed using Benchling®. Each guide and its complementary sequence was ordered as synthetic 25nmole oligos from Thermo fisher® with attached BbsI cloning sites: Sense: 5' – CACCGNNNNNNNNNNNNNNNNN – 3' and antisense: 3' – CNNNNNNNNNNNNNNNNNNCAAA – 5' (**Supplementary Table 2**).

In order to clone the designed guides into the pU6-(BbsI)_CBh-Cas9-T2A-mCherry plasmid (Addgene®), synthetic oligos were suspended in water to 100µM and then 10µM of oligos was phosphorylated using T4 PNK (NEB) at 37°C for 1 hour. A 0.5µM mix of both oligos, was annealed by heating at 95°C for 5 minutes, oligos were allowed to cool down slowly at room temperature. The pU6-(BbsI)_CBh-Cas9-T2A-mCherry plasmid was digested with fast digest BbsI (Thermo®) for 1 hour at 37ºC, dephosphorylation was carried out with rSAP (NEB) for 30 minutes and the plasmid was run on a 1% agarose gel, followed by gel purification (Qiagen®). Ligation was performed with 50ng of the linearized vector and 1µl of the annealed oligos using Quick ligase (Roche) for 1h at room temperature. One Shot TOP10 competent E. coli cells (Thermo®)

were transformed following manufacturer instructions and plated in 100µg/ml Ampicillin (Sigma) plates followed by an overnight incubation at 37°C. Insert was verified by colony PCR using the sense oligo as a forward primer and a common Cbh reverse primer: 5' GTCAATAGGGGGCGTACTTGG 3', at a 50°C annealing using the HiFi PCR premix (Clonetech®). A Maxi prep (Qiagen®) was performed, followed of a final verification by Sanger sequencing using the LKO forward primer for the human U6 promoter: 5' GACTATCATATGCTTACCGT 3'.

3.3.5 Electroporation

In order to achieve high editing rates, we used the Neon™ electroporation system (Thermo®) manufacturer instructions for program optimization (**Supplementary Figure 10**). HEK293 cells were trypsinized and centrifuged at 250×g for 5min, the supernatant was carefully aspirated and the cell pellet was suspended once with 10 mL of DPBS without Ca^{2+} and Mg^{2+}, a 10µl sample was mixed 1:1 with 0.4% trypan blue and cell concentration was quantified in a Countess® automated cell counter (Thermo®). The sample was centrifuged again and the supernatant was carefully aspirated and the cell pellet was suspended in Buffer R to a final concentration of 5×10^5 cells/µl. DNA plasmids (i.e 750ng pU6-BFP-sgRNA#1_CBh-Cas9-T2A-mCherry, 750ng C16 vector) and repair template (500ng of ssODN) were diluted in R buffer up to 9µl. A 24µl cell suspension mix for two reactions was made with 15µl of the cell suspension mix and the 9µl of the

DNA mix. The parameters used for electroporation were: tree pulses of 1500 v with a 10ms width. The electroporated cells were transferred to a 25cm Nunc™ Cell Culture Treated EasYFlasks™ (Thermo®) with 4ml of antibiotic free 20% FBS DMEM-F12 with 10µM Y-27632 followed by incubation for 48h in a 5% CO_2 humidified incubator before replacing the media for 10% FBS DMEM-F12 1xAnti-Anti (Thermo®) with 2µg/ml of puromycin. For the compound treatments media was replaced 24h after electroporation with the compounds (i.e 15µM RS-1 and 80µM SCR7) and incubated for 72h before replacing the media. For the two electroporations, 100µl tips were used with the same propotions of DNA, but using an in-house made buffer (SBB: 250 mM sucrose and 1 mM MgCl2 in 1x DPBS) as reported by Brees et al., (252). All tips were recycled as reported by Brees et al., plus overnight UV incubation.

3.3.6 Nuclear Re-Localization of C16 From the Cytoplasm to the Nucleus

In order to evaluate the nuclear compartmentalization of C16, HEK 293 cells were electroplated as described above with the isolated vectors that contained a fused mCherry gene for visualization (**Figure 8, A-D**). Cells were imaged under an inverted fluoresce microscope (Nikon eclipse TE300) and mean grey value was quantified using Adobe Photoshop CC 2017 (**Figure 8**).

3.3.7 Flow Cytometry for BFP to GFP Analysis

In order to quantify editing outcomes, transfected cells were analyzed by flow cytometry after 8 days of culture. Cells were transported in 10% FBS 1x PBS and 1mM EDTA. Cell were gated for single cells and a total of 20,000 single cell positive events (GFP+) were collected. Analysis was performed in FlowJo. Statistical analysis was performed in Prism (**Figure 9**).

3.3.8 Statistical Analysis

All experiments were done with two or more independent biological replicates each with two or more technical replicates. Only the means of the biological replicates are used for the analysis (253). Unless otherwise indicated, after verifying for the assumptions of equal variance and normality, P values were calculated using One-Way ANOVA with Tukey's HSD (230). Unless otherwise indicated, error bars represent standard deviation. Analyses were performed with Prism (Graphpad®).

3.4 Results

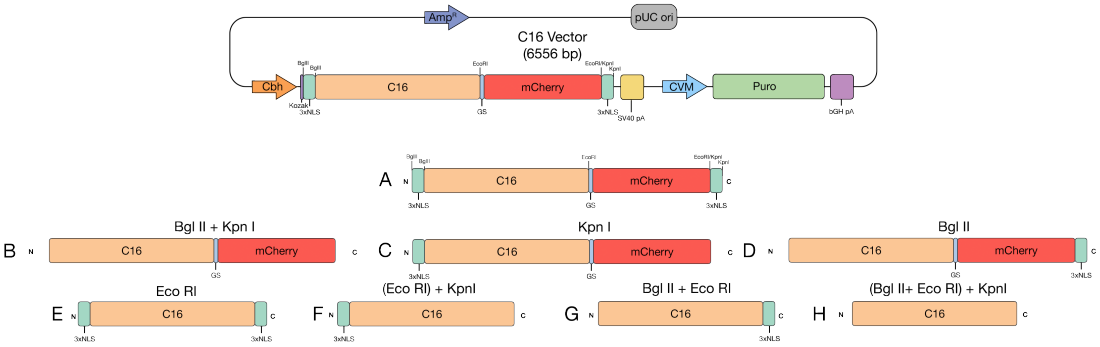


Figure 7. C16 Designed Vector and Variants.
Schematic illustration of the designed vector. A) Complete synthetic gene with two 3xNLS in the C and N terminus. B) Synthetic gene with no NLS produced from a Bgl II and KpnI digestion. C) Synthetic gene with N terminus 3xNLS produced from a KpnI digestion. D) Synthetic gene with C terminus 3xNLS produced from a Bgl II digestion. E) C16 protein with both N and C terminus 3xNLS. F) C16 protein with N terminus 3xNLS. G) C16 protein with C terminus 3xNLS. G) C16 protein with C terminus 3xNLS. G) C16 protein with no NLS.

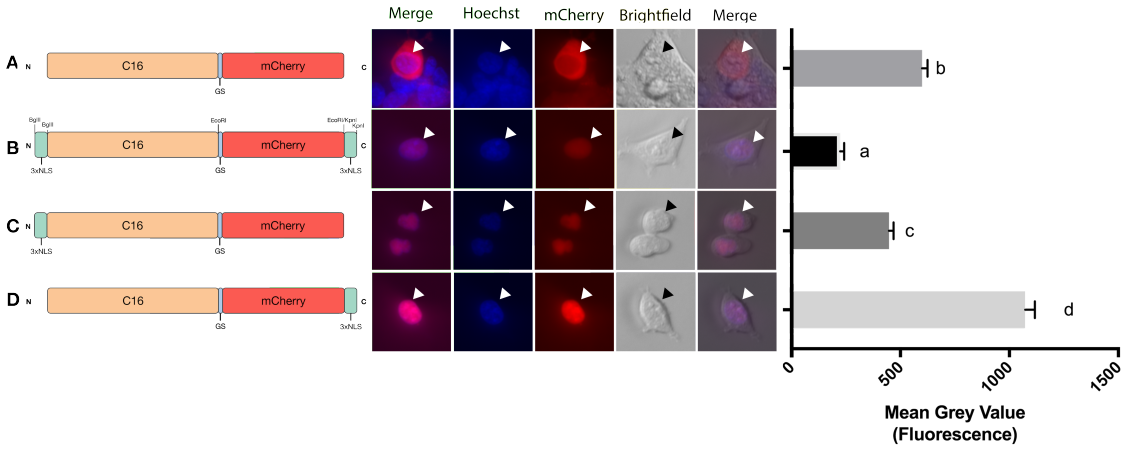


Figure 8. Nuclear Localization Assessment of C16.
Brightfield and fluorescence imaging for mCherry and Hoechst. A) C16-mCherry fusion gene with no NLS. B) Complete synthetic gene with two 3xNLS in the C and N terminus. C) Synthetic gene with N terminus 3xNLS. D) Synthetic gene with C terminus 3xNLS. Values with different subscripts are different, nonparametric One-Way ANOVA (Kruskal-Wallis) test, $p<0.05$, error bars represent standard error of mean (SEM).

3.4.1 Nuclear Re-Localization of C16 From the Cytoplasm to the Nucleus

We designed a vector carrying the codon optimized C16 protein fused to a C terminus mCherry via a GS linker with two 3xNLS on both the N or C Terminus (**Figure 7, A**). We then constructed different versions of this vector with and without mCherry, carrying either a N or C terminus NLS or no NLS at all (**Figure 7, B-H**). The mCherry fusion constructs were transfected into HEK293 cells, and nuclear localization was quantified as the mean grey value in the nuclear area determined by Hoechst stain (**Figure 2**). The presence of the 3 repeats of the NLS were able to re-localize C16 form the cytoplasm (**Figure 8 B**) to the nuclear compartment (**Figure 8A, C and D**) independently of the location. Nevertheless, the mean grey value (fluoresce) was significantly lower ($p<0.05$) with the presence of the NLS in both terminus (**Figure 8, A**) or in the N termini (**Figure 8, C**) compared to the wild type protein (**Figure 8, B**). The highest nuclear fluorescence was obtained in the C16 protein carrying a C terminus NLS (**Figure 8 D**).

3.4.2 Establishment of BFP to GFP Conversion Assay in HEK293 Cells

To evaluate the CRISPR genome editing outcomes we first generated a modified HEK293 cell line with single genomic integrations of a lentiviral transgene crrying the blue fluorescent protein gene (BFP), BFP and puromycin resistance cassette gene under

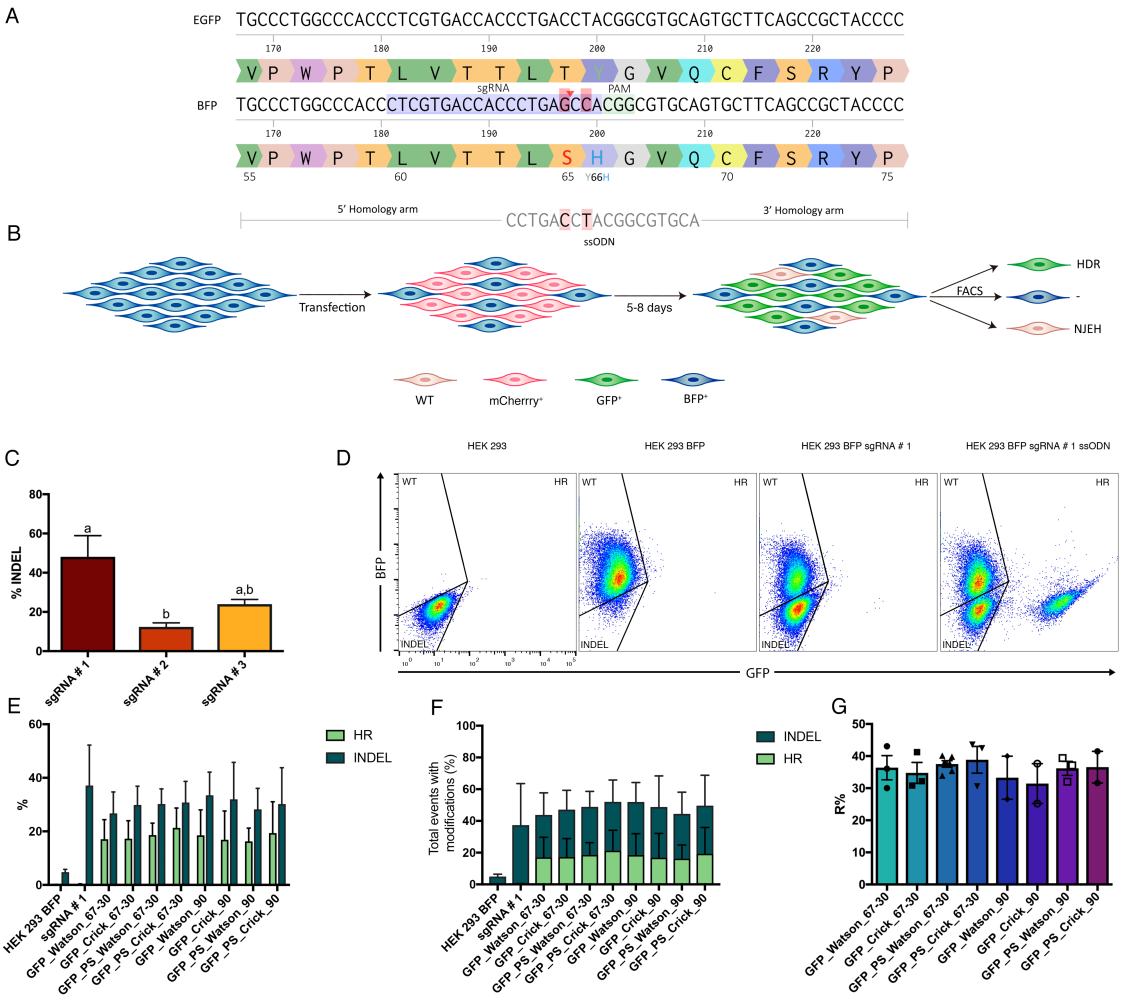
the EF1a and SV40 promoters respectively, using lentiviral particles. This cell line was sorted twice by flow cytometry (**Supplementary Figure 8**). Transgene copy number was determined using ddPCR, resulting in BFP/RPP30: ~ 3 (tri-allelic, 3 copies) (**Supplementary Figure 9**). We tested three different sgRNAs targeting the BFP protein chromophore region and using flow cytometry, assayed the loss of fluorescence as a direct indicator of insertion and deletions (INDELS) caused by CRISPR-Cas9 DSBs (representing the NHEJ pathway). The small guide RNA one (sgRNA #1) had the highest efficiency and was used for the rest of this study (**Figure 9 C, D**).

3.4.3 Effects of ssODN parameters on HDR

We tested different single-stranded oligodeoxynucleotide (ssODN) repair templates with either asymmetric 67-30nt or symmetric 45-45nt homology arms with and without 5' and 3' phosphorothioate (PS) modifications. In addition, we also evaluated the orientation in either a Watson (PAM strand for sgRNA#1) or Crick conformation (**Supplementary Figure 12**). We did not observe significant differences among the groups on either the percent of R ratios or HR. It is important to disclose that we tested same mass and not same molarity for each of the repair templates (**Figure 9 E-G**). Based on our results and on the work of Liang, et al., (181) we chose an asymmetric ssODN with PS modifications in a Watson orientation for the remainder of this study. However, these data could not support this choice over any other.

Figure 9. Gene Editing Outcomes Quantification by FLow Cytometry.

A. Comparison between the DNA and amino acid sequences of the enhanced green fluorescent protein (eGFP) and the blue fluorescent protein (BFP). We designed a sgRNA (purple) that guides Cas9 to introduce a specific double stranded break (DSB). When a single-stranded oligodeoxynucleotide (ssODN) repair template is provided, two single base pair changes in the chromophore region of the BFP gene, mediate the conversion from blue to green emission. B. Illustration of the experimental design, HEK293 BFP cells transfected with CRISPR-Cas9 and C16 plasmids that carry the mCherry marker for quantifying transfection efficiency. C. The CRISPR-Cas9 induction for insertions and deletions (INDELS) causes a loss of BFP fluorescence and can be used as an estimate of the non-homologous end joining pathway (NHEJ), sgRNA efficiency was quantify as the percent of BFP negative cells. D. Gene editing outcome quantification by flow cytometry, matrix composed of two gates for evaluating GFP and BFP expression. HEK293 cells show no fluorescence and locate in the bottom left third, HEK293 BFP cells locate upper left third. The CRISPR-Cas9 induction for insertions and deletions (INDELS) causes a loss of BFP fluorescence and can be used as an estimate of the non-homologous end joining pathway (NHEJ), when providing a ssODN the GFP positive cells locate in the right third. In order to stablish an efficient assay, different repair templates were tested, graph E shows HR and NHEJ rates, graph F shows HR rates and graph G shows editing ratios (R) calculated as (HDR/(NHEJ+HDR)). One way ANOVA, different letters signify statistical differences $p < 0.05$.


A
EGFP
TGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCC
V P W P T L V T T L T Y G V Q C F S R Y P
sgRNA
PAM
BFP
TGCCCTGGCCCACCCTCGTGACCACCCTGAGCCACGGCGTGCAGTGCTTCAGCCGCTACCCC
V P W P T L V T T L S H G V Q C F S R Y P
Y66H
B
5' Homology arm
CCTGACCTACGGCGTGCA
3' Homology arm
ssODN
Transfection
5-8 days
FACS
HDR
NJEH
WT
mCherrry+
GFP+
BFP+
C
% INDEL
a
b
a,b
sgRNA # 1
sgRNA # 2
sgRNA # 3
D
HEK 293
HEK 293 BFP
HEK 293 BFP sgRNA # 1
HEK 293 BFP sgRNA # 1 ssODN
WT
HR
INDEL
BFP
GFP
E
%
HR
INDEL
HEK 293 BFP
sgRNA # 1
GFP_Watson_67-30
GFP_Crick_67-30
GFP_PS_Watson_67-30
GFP_PS_Crick_67-30
GFP_Watson_90
GFP_Crick_90
GFP_PS_Watson_90
GFP_PS_Crick_90
F
Total events with modifications (%)
INDEL
HR
G
R%

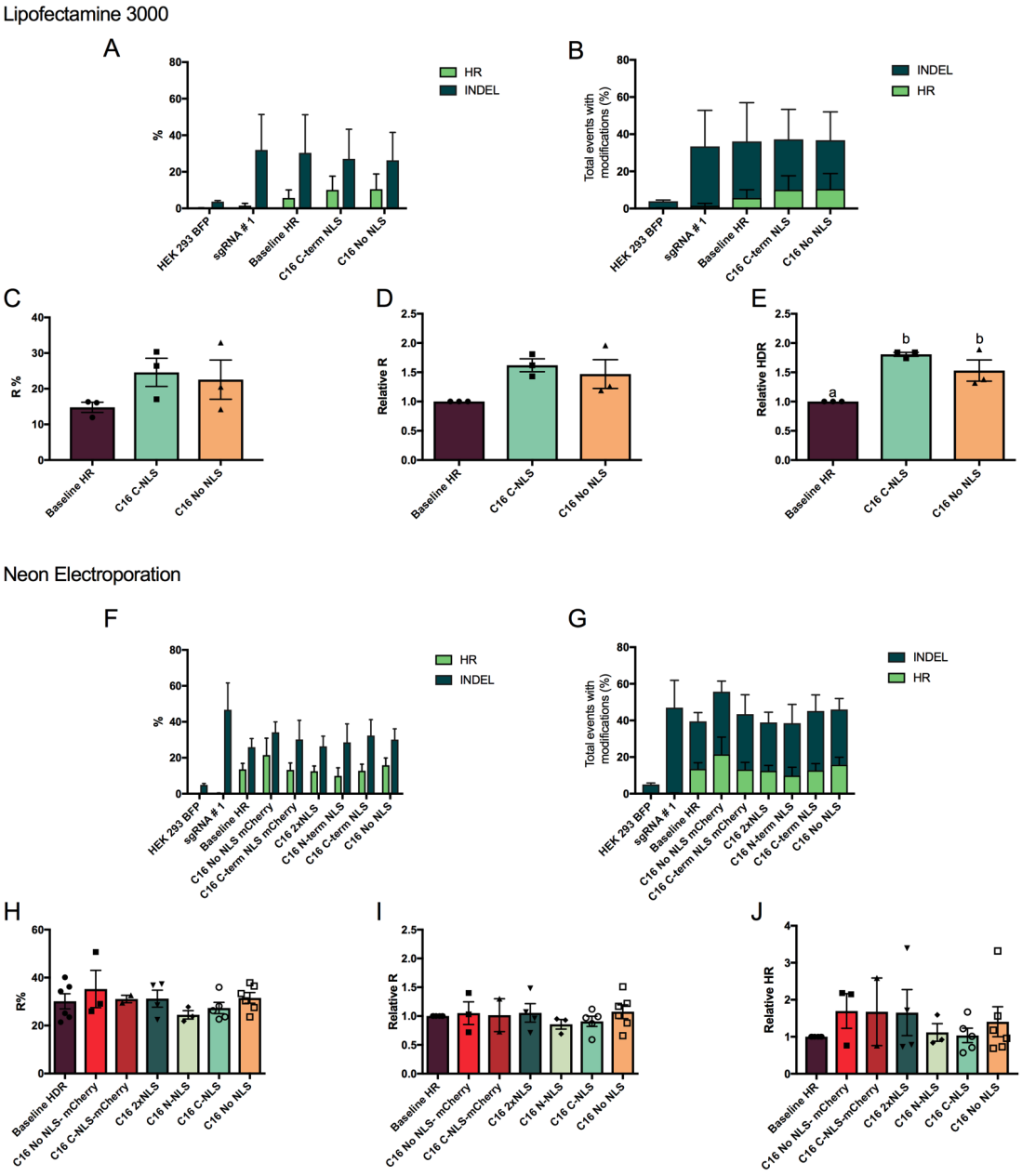


Figure 10. Effects of C16 on HDR Rates.

C16 effects on HDR rates when co-transfected CRISPR-Cas9 using either A-E) Lipofectamine 3000 or F-J) the Neon electroporation system. Each dot represents one biological replicate mean from two or more technical replicates. Values with different subscripts are different ($P<0.05$).

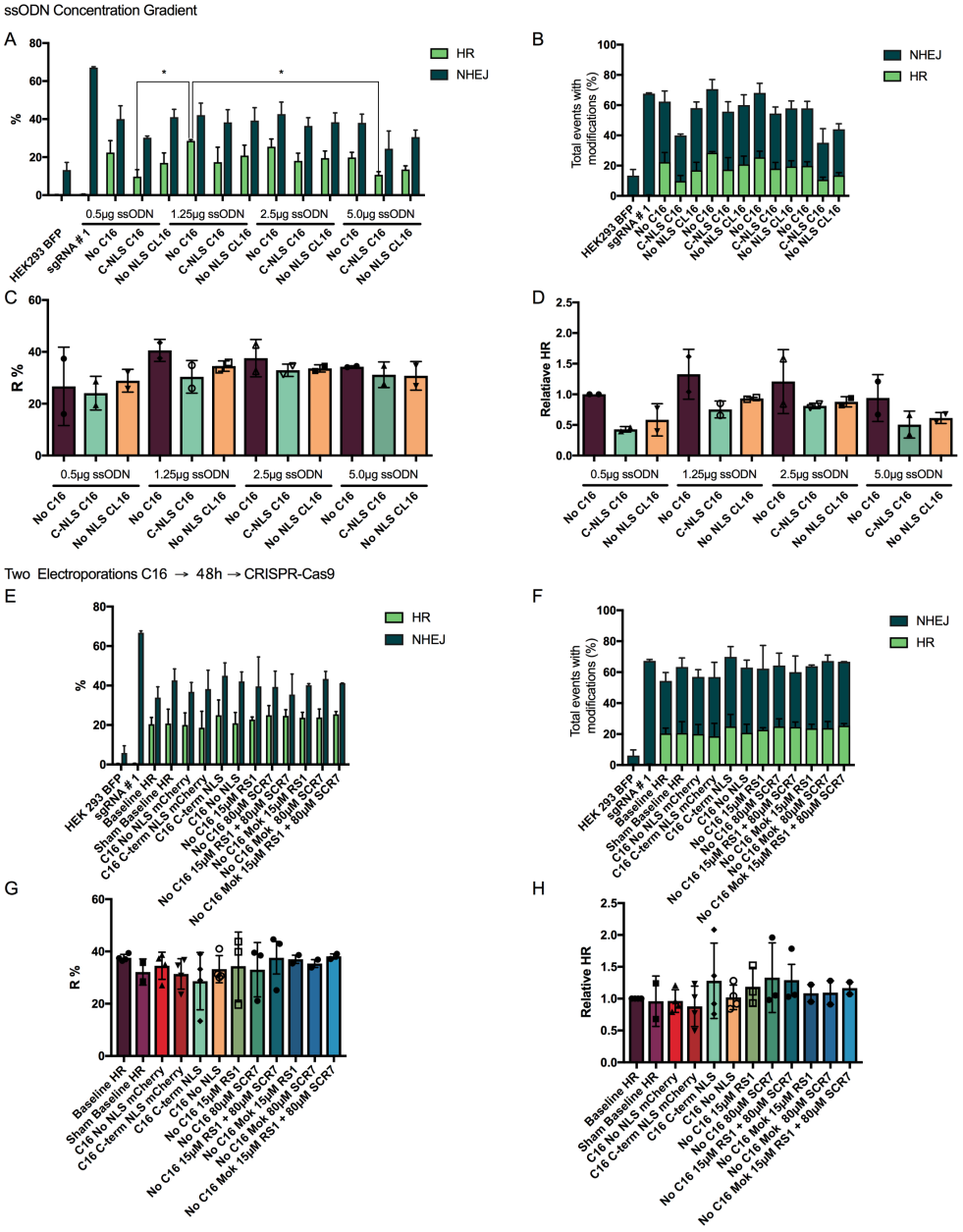


Figure 11. Effect of Increased Concentration of Donor Template and Sequential Delivery of C16.

A-D) C16 effects on HDR rates when co-transfected CRISPR-Cas9 under increasing concentrations of repair template (0.5, 1.25μg, 2.5μg and 5.0μg). E-F) C16 effects on HDR rates when transfected sequentially 48h later with CRISPR-Cas9.

3.4.4 Effects of C16 on HDR rate

In order to evaluate the effect of C16 on homologous recombination (HR) rates, we first employed liposomal transfection (Lipofectamine 3000, Thermo) to deliver C16 with either no NLS or a C-Terminus NLS (**Figure 7 G, H**), together with the CRISPR-Cas9 px458 plasmid and the asymmetric repair template (Watson-PS-ssODN) into our HEK293 BFP reporter cell line (**Figure 10 A-E**). Transfection efficiency was highly variable due to the poor adherence of our cells, that only mildly improved in either Type-I collagen or gelatin coated plates. However, the presence of the C-terminal 3xNLS C16 or C16 without NLS, increased the number of green cells, improving baseline HR (CRISPR-Cas9 and ssODN, without C16) from 5.73% (0.4% - 14.36%) to 10.09% (0.8% - 25.00%) and 10.47% (0.87% -27.18%) (**Figure 10 A**) representing a 1.80 and 1.53-fold improvement on normalized HR rates (**Figure 10 E**) respectively ($n=3$, $P<0.05$). Although, no significant differences were observed on R ratios (HR/NHEJ+HR) and normalized R ratios, there was a tendency ($p=0.0541$) for the C-terminus NLS C16 to improve the normalized R ratios.

We next evaluated C16 HR improvements using electroporation with the Neon Transfection System (Thermo®). Although initial baseline HDR improved with electroporation to 16.7% ± 7.2 compared to lipofectamine (5.7±7.5) ($P=0.0521$) (**Supplementary Figure 13**), no significant differences were found between the different

treatment groups on either R ratios, relative R ratios or relative HR (n= 3-6) (**Figure 10 F-J**). We hypothesized that this could be due because there was not enough repair template available to allow for HR rates to increase or because of the higher transfection efficiency was leading to different molecular stoichiometries between the different vectors that might be influencing gene-editing efficiencies, which may limit the amount of C16 protein available before the CRISPR-Cas9 DSB was introduced.

In order to test this hypothesis, we optimized our electroporation conditions (**Supplementary Figure 10**) and tested increasing concentrations of ssODN (0.5, 1.25µg, 2.5µg and 5.0µg) comparing the baseline HR (No C16) with a C-terminus NLS C16 or C16 without NLS. It is important to disclose that the baseline HR group in each of the treatments was not compensated for the extra 750ng of DNA from the C16 plasmid, so the total amount of DNA transfected in the control is lower (e.g 1.25µg total DNA in the 0.5µg of ssODN group) than in the C-NLS C16 and No NLS C16 groups (e.g 2µg total DNA in the 0.5µg of ssODN group). Significant differences were found in the HR% of the 1.25µg ssODN no C16 (28.54%) versus the 0.5µg ssODN C-NLS c16 (9.76%) and the 5.0µg ssODN C-NLS c16 (10.68%) (n=2, $p<0.05$) (**Figure 11 A**). Editing ratio (R) and HR rates peaked at an average of 40.59% and of 28.54 % for 1.25µg of ssODN. No significant differences were found in the HR rates or relative HR ratios normalized to the 0.5µg no C16 group (22.41 % HR) or (**Figure 11 A-D**).

To address our second hypothesis, we the tested the sequential delivery of C16 followed 48h later by the CRISPR-Cas9 plasmid and the ssODN repair template. During this experiment, we also evaluated under our conditions the effect of the RAD-51 stimulatory compound RS-1 (15μM) and the DNA-Ligase IV inhibitory compound SCR7 (80μM). No significant differences were found in either NHEJ, HR, R or relative R rates (n=3) (**Figure 11 E-H**)

3.5 Discussion

To test our initial hypothesis, we established of the BFP to GFP conversion assay first reported by Glasser et al., (254), which has been used by many other laboratories to study the improvement of genome editing outcomes and the biochemical optimization of the CRISPR-Cas9 system(146, 164, 255, 256). In our hands, we were able to obtain on average ~ 20% total HR (up to 30% HR, **Supplementary Figure 11**) rates representing ~ 40% of the total editing events (R), while using a plasmid based CRISPR-Cas9 system and an ssDNA donor template. We created a heterogeneous population of BFP cells with different single transgene integration sites at different locus in order to avoid locus bias in our HDR experiments. Although we maintained the cells under constant puromycin selection (2μg/ml) for 15days before beginning our experiments, we noticed that not continuing this selection after transfections lead to silencing of the transgene, as seen by

an increase of BFP negative cells in our control populations 72h after transfection cells were maintained under puromycin selection.

In this study, we did not find any significant differences in the gene editing outcomes between the different parameters being tested with respect to the ssODN repair template (**Figure 9 E-G**). The evaluated criteria were the symmetry and of the homology arms, the presence of PS modifications and the strand orientation. Although this is likely due to the large variability of the experiment, it is also worth noting that in this study comparisons were done with mass instead of molarity between the different repair templates. Other groups have reported that an asymmetric ssODN with PS modifications had up to 56% HR in HEK293 cells (181). Similarly, the Corn group reported that after investigating the interaction of Cas9 with target DNA and discovering that Cas9 releases the PAM-distal non-target strand after cleavage but before dissociation they used this finding to rationally design a ssDNA that matches this strand, they increased HR rates with up to 60% HR in the absence of any chemical intervention(164) (**Figure 1 B**).

We then examined the re-localization of C16 from the cytoplasmic to the nuclear compartment. As the VACV replication cycle occurs in the cytoplasm, C16 is a cytoplasmic protein (249). This was achieved by adding a nuclear localization signal (3xNLS) at either the N, C terminus or both. The lower fluoresce intensity of the constructs

carrying a N terminus NLS, suggests that the presence of the NLS might affect the folding or the protein production dynamics of the protein (**Figure 8**).

A 3xNLS C terminus tagged C16 was able to increase low baseline HR rates of an average of 5.73% to 10.09%, representing a moderate ~80% increase in the BFP to GFP conversion assay in HEK293. These experiments were performed using a liposomal based transfection system (Lipofectamine 3000) that produced large variability due to the inability of our cells to remain attached to the plates. These results have similar improvement rates to the work of of Gutschner et al.,(158) where a 1.87-fold increase was obtained when fusing Cas9 to the N-terminal region of human Geminin, although with very low overall rates of a maximum 0.59% HR. This same strategy was later taken and improved by another group where they recently indicated that a Cas9- GFP-geminin enhanced the HDR/NHEJ ratio 2.7 fold in U2-OS and 1.8 fold K562 cells and is now commercially available (182).

We decided to further characterize for how this protein would behave at the higher range of HDR rates. For this we decided to change methodology and use the Neon electroporation system (Thermo®) that would allow for higher and more consistent transfection rates. Although baseline HDR improved to 13.56 % ± 8.25 (**Supplementary Figure 13**), no significant differences were found between the different treatment groups

(**Figure 10 F-J**). We hypothesized that this could be either due because there was not enough repair template available to allow for HR rates to increase or because of the higher transfection efficiency was leading to different molecular stoichiometries between the different vectors that might be influencing gene-editing efficiencies by causing for not enough C16 protein to be present before the CRISPR-Cas9 DSB was introduced. This hypothesis was tested by evaluating the effect of a gradual increase in the amount of repair template and by evaluating the transfection of the C16 plasmid followed 48h later of the CRISPR-Cas9 plasmid and ssODN. However, no significant differences were found (**Figure 11**). During this second experiment, we also evaluated under our conditions the effect of RS-1 (15μM) and SCR7 (80μM) but did not observe any significant differences in HDR (**Figure 11 E-H**). It is important to note that we used a lower dose of SCR7 due to reagent availability in our laboratory and the time frame to perform these experiments. Although the original publication by Chu et al., demonstrated that SCR7 increased HDR (up to 19-fold at a concentration of 1 μM) (167), other reports indicate a very high variability of this molecule between different experimental systems (158); or even that SCR7 is neither a selective nor a potent inhibitor of the human DNA ligase IV (178).

RS-1 is a Rad51 stimulatory molecule identified by a small molecule screen where it was shown to stabilize association of RAD51 with DNA(179). Rad51 is a key molecule in the HDR pathway that displaces RPA to form a RAD51–ssDNA nucleofilament

recombinase, which is required for homology search, strand invasion and ultimately the repair of the DSB lesion (148) (**Figure 1**). Song et al.,(172) reported RS-1 improves HR rates 2-5 fold at a concentration of 7.5μM in vitro and in vivo (from 4.4 to 26.1%). RS-1 was also shown to improve HR rates in HEK cells (from 3.5 to 21%) at and U2OS cells (from 1.9 to 2.4%) at a concentration of 10μM, but this varied depending on the locus and transfection method, with electroporation (~3-fold) having lower efficiency than with lipofectamine (~6-fold) (180). It is important to note that the HR repair with a ssODN has been recently referred to as single stranded template repair (SSTR) (117, 237) and it has been identified as a RAD51/BRCA2 independent pathway (154) so the type of donor being used is one of the possible explanations of the variability of the results obtained with RS-1. Zhang et al.,(257) reported that RS-1 did not improved HR rates in hiPSCs at the same concentration of 10μM. Interestingly, they also report that the overexpression of RAD51 or of the adenovirus 4 (Ad4) proteins 4E1B-E4orf6, which mediate the ubiquitination and proteosomal degradation of DNA ligase IV; were detrimental to their HR rates. One explanation they give is that after optimization of their donor vector, they obtained high-level of HR and these high rates masked the subtle changes mediated by many of this inhibitors and proteins. The co-expression of these same adenovirus proteins together with the CRISPR-Cas9 system using a 2A peptide was reported by Chu et al., (167) to improve HDR efficiency up to ~8-fold reaching HR frequencies of 50-66% and significantly decreased NHEJ activity in HEK293 and mouse Burkitt lymphoma-like cell

lines. This same work reports the use of shRNA sequences to knock down KU70, KU80 or DNA ligase IV, which resulted in substantial suppression of the NHEJ and a ~5-fold improvement in HR.

Following the same small molecule inhibition strategy, Robert el, al (170) reported the inhibition of the DNA-PK_{cs} using the small molecules NU7441 and KU-0060648, which caused a decrease of ~40 % in NHEJ events with a ~2-fold increase in HDR, although HDR rates were low ~4 % HR; they also show the additive effect of combining these two compounds with either RS-1, SCR7, siRNA suppression of Ku70 and Ku80 and adenovirus proteins E1B55K and E4orf6, these last being the most efficient inhibiting NHEJ ~8-fold and stimulating HR ~3.5 fold, in line Chu, et al., (167) .

In our study, we observed a transfection methodology variability effect similar to the one described by Song et al.,(172), We also found no effect of the C16 protein after optimizing for higher HDR rates just as described Zhang et al.,(257). Zhang et, observed that the use of the Ad4 proteins 4E1B-E4orf6d were even detrimental to their HR rates, but Chu et al., (167) obtained ~8-fold HR improvement when co-expressed with the CRISPR-Cas9 system using a 2A sequence. This would lead us to believe that expressing C16 in a bi-cistronic construct with CRISPR-Cas9 would lead to a better performance of the system. Further, in this study, we used an optimized asymmetric sODN that is using

the SSTR pathway and if we considered that HDR is required for repairing lesions using double-stranded, but not single-stranded DNA as a template(154), in order to evaluate the effects of C16 on the HDR pathway, we would need to evaluate using a dsDNA donor with homology arms longer than 100bp (137).

Smith et al., demonstrated that C16 could inhibit the DNA-mediated activation of the innate immune system, resulting in approximately a 50% reduction in the production of the pro-inflammatory molecules Cxcl10 and Il-6 (245). This was in line with a report where MEFs lacking the Ku heterodimer induce lower levels of cytokines and chemokines induced upon stimulation with DNA (244). This residual signaling, was explained by the Smith group as by DNA-PKcs having DNA-binding capability independent of Ku (258), although this interaction is greatly enhanced by the presence of the Ku heterodimer, and by the existence of other DNA sensing mechanism (249, 259, 260) such as IFI16 (261) and cGAS (262). Another reason is that C16 does not bind Ku70/80 in direct competition with DNA-PKcs, although it reduces the amount of DNA-PKcs bound to DNA (245, 249). Additionally, there is a multi-functional role of C16, as this protein was also show to reprogram cellular energy metabolism towards increased synthesis of the metabolic precursors utilized during viral replication, via the stabilization of the hypoxia-inducible transcription factor (HIF)-1 alfa. Finally an incomplete penetrance of the cells with the C16-encoding plasmid and the high abundance of the Ku molecules in cells (KU70

1,290,000 and KU80 826,000 molecules per cell) (137) can make the inhibition of these molecules under our conditions challenging.

3.6 Conclusions

In summary, our study introduces a new approach to the efforts of the manipulation of the gene editing outcomes, by the adaptation of the naturally evolved VACV-C16 protein to target the Ku heterodimer in the mammalian nucleus. We recognize that the observed enhancement in our study is moderate and that under the conditions tested is only beneficial at low basal HR rates. We show the need for a refinement of the spatiotemporal expression of this protein. More research will be needed in order to reveal the full potential of this genome engineering tool in the future.

3.7 Acknowledgments

We would like to thank Gus Wright for his help on the FACS sort and the gating analysis on the flow data. We would also wile to thank Andrew Hillhouse for his help with the ddPCR. We would also like to thank the support of the Innovation Kitchen Laboratory and its members, for providing funding, equipment and ordering of reagents. Finally, we would like to thank Diego Fernando Carrillo, Nicolas Moreno, Yudi Bedi, Camilo Hernandez, Charlotte Whittaker and Cassandra Skenadore, for their help in different aspects of this project.

CHAPTER IV

SUMMARY AND CONCLUSIONS

In this work, we have shown that genome editing technologies can be used for the therapeutic genome editing of genetic diseases in animals by successfully correcting the GBED mutation with moderate efficiency (~20%) in a heterozygous cell line derived from a high genetic merit American Quarter Horse stallion using CRISPR-Cas9 that will be used to generate a cloned animal. In order to make the correction of genetic diseases it is necessary to increase the efficiency of precise genome editing, this was the main objective for the second part of our work where we created a new approach to the efforts of the manipulation of the gene editing outcomes, by the adaptation of the naturally evolved VACV-C16 protein to target the Ku heterodimer in the mammalian nucleus. We recognize that the observed enhancement in our study is moderate and that under the conditions tested is only beneficial at low basal HR rates. We show the need for a refinement of the spatiotemporal expression of this protein. More research will be needed in order to reveal the full potential of this genome engineering tool in the future.

REFERENCES

1. Bettegowda A, Patel OV, Lee KB, Park KE, Salem M, Yao J, et al. Identification of novel bovine cumulus cell molecular markers predictive of oocyte competence: functional and diagnostic implications. Biology of reproduction. 2008;79(2):301-9.

2. Wivel NA. Historical perspectives pertaining to the NIH Recombinant DNA Advisory Committee. Hum Gene Ther. 2014;25(1):19-24.

3. Cohen SN, Chang AC, Boyer HW, Helling RB. Construction of biologically functional bacterial plasmids in vitro. Proc Natl Acad Sci U S A. 1973;70(11):3240-4.

4. Morrow JF, Cohen SN, Chang AC, Boyer HW, Goodman HM, Helling RB. Replication and transcription of eukaryotic DNA in Escherichia coli. Proc Natl Acad Sci U S A. 1974;71(5):1743-7.

5. Jaenisch R, Mintz B. Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA. Proc Natl Acad Sci U S A. 1974;71(4):1250-4.

6. Gordon JW, Scangos GA, Plotkin DJ, Barbosa JA, Ruddle FH. Genetic transformation of mouse embryos by microinjection of purified DNA. Proc Natl Acad Sci U S A. 1980;77(12):7380-4.

7. Brinster RL, hen HY, Trumbauer M, Senear AW, Warren R, Palmiter RD. Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs. Cell. 1981;27(1 Pt 2):223-31.

8. Costantini F, Lacy E. Introduction of a rabbit beta-globin gene into the mouse germ line. Nature. 1981;294(5836):92-4.

9. Gordon JW, Ruddle FH. Integration and stable germ line transmission of genes injected into mouse pronuclei. Science. 1981;214(4526):1244-6.

10. Folger KR, Wong EA, Wahl G, Capecchi MR. Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. Mol Cell Biol. 1982;2(11):1372-87.

11. Smithies O, Gregg RG, Boggs SS, Koralewski MA, Kucherlapati RS. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature. 1985;317(6034):230-4.

12. Palmiter RD, Brinster RL, Hammer RE, Trumbauer ME, Rosenfeld MG, Birnberg NC, et al. Dramatic growth of mice that develop from eggs microinjected with metallothionein-growth hormone fusion genes. Nature. 1982;300(5893):611-5.

13. Hammer RE, Pursel VG, Rexroad CE, Jr., Wall RJ, Bolt DJ, Ebert KM, et al. Production of transgenic rabbits, sheep and pigs by microinjection. Nature. 1985;315(6021):680-3.

14. Ebert KM, Schindler JES. Transgenic farm animals: Progress report. Theriogenology. 1993;39(1):121-35.

15. Pursel VG, Rexroad CE, Jr. Status of research with transgenic farm animals. Journal of animal science. 1993;71 Suppl 3:10-9.

16. Wall RJ. Pronuclear microinjection. Cloning and stem cells. 2001;3(4):209-20.

17. Panarace M, Aguero JI, Garrote M, Jauregui G, Segovia A, Cane L, et al. How healthy are clones and their progeny: 5 years of field experience. Theriogenology. 2007;67(1):142-51.

18. Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. The double-strand-break repair model for recombination. Cell. 1983;33(1):25-35.

19. Robertson E, Bradley A, Kuehn M, Evans M. Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. Nature. 1986;323(6087):445-8.

20. Nobelprize.org. The Nobel Prize in Physiology or Medicine 2007: Nobel Media AB; 2007 [Available from: <http://www.nobelprize.org/nobel_prizes/medicine/laureates/2007/>.

21. Malaver-Ortega LF, Sumer H, Liu J, Verma PJ. The state of the art for pluripotent stem cells derivation in domestic ungulates. Theriogenology. 2012;78(8):1749-62.

22. Soto DA, Ross PJ. Pluripotent stem cells and livestock genetic engineering. Transgenic Res. 2016;25(3):289-306.

23. Kumar D, Talluri TR, Anand T, Kues WA. Induced pluripotent stem cells: Mechanisms, achievements and perspectives in farm animals. World journal of stem cells. 2015;7(2):315-28.

24. Goncalves NN, Ambrosio CE, Piedrahita JA. Stem cells and regenerative medicine in domestic and companion animals: a multispecies perspective. Reprod Domest Anim. 2014;49 Suppl 4:2-10.

25. Gandolfi F, Pennarossa G, Maffei S, Brevini T. Why is it so difficult to derive pluripotent stem cells in domestic ungulates? Reprod Domest Anim. 2012;47 Suppl 5:11-7.

26. Blomberg LA, Telugu BP. Twenty years of embryonic stem cell research in farm animals. Reprod Domest Anim. 2012;47 Suppl 4:80-5.

27. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126(4):663-76.

28. Cibelli JB, Stice SL, Golueke PJ, Kane JJ, Jerry J, Blackwell C, et al. Transgenic bovine chimeric offspring produced from somatic cell-derived stem-like cells. Nature biotechnology. 1998;16(7):642-6.

29. Y. S. Bogliotti, J. Wu, M. Vilariño, K. Suzuki, J. C. Belmonte, Ross PJ. BOVINE EMBRYONIC STEM-LIKE CELLS DERIVED FROM IN VITRO-PRODUCED BLASTOCYSTS. Reproduction, Fertility and Development. 2017;29(1):108.

30. Rouet P, Smih F, Jasin M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol Cell Biol. 1994;14(12):8096-106.

31. Chandrasegaran S, Carroll D. Origins of Programmable Nucleases for Genome Engineering. J Mol Biol. 2016;428(5 Pt B):963-89.

32. Elsdale TR, Gurdon JB, Fischberg M. A description of the technique for nuclear transplantation in Xenopus laevis. Journal of embryology and experimental morphology. 1960;8:437-44.

33. Gurdon JB. The developmental capacity of nuclei taken from differentiating endoderm cells of Xenopus laevis. Journal of embryology and experimental morphology. 1960;8:505-26.

34. Gurdon JB. Factors responsible for the abnormal development of embryos obtained by nuclear transplantation in Xenopus laevis. Journal of embryology and experimental morphology. 1960;8:327-40.

35. Sutovsky P, Prather RS. Nuclear remodeling after SCNT: a contractor's nightmare. Trends in biotechnology. 2004;22(5):205-8.

36. Chavatte-Palmer P, Lee R, Bertolini M, Jammes H, Schmidt M, Callesen H. Chapter 9 - Pregnancy and Neonatal Care of SCNT Animals. In: Cibelli J, Gurdon J, Wilmut I, Jaenisch R, Lanza R, West MD, et al., editors. Principles of Cloning (Second Edition). San Diego: Academic Press; 2014. p. 107-26.

37. Schnieke AE, Kind AJ, Ritchie WA, Mycock K, Scott AR, Ritchie M, et al. Human factor IX transgenic sheep produced by transfer of nuclei from transfected fetal fibroblasts. Science. 1997;278(5346):2130-3.

38. Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A. 1996;93(3):1156-60.

39. Smith J, Bibikova M, Whitby FG, Reddy AR, Chandrasegaran S, Carroll D. Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. Nucleic Acids Res. 2000;28(17):3361-9.

40. Bibikova M, Carroll D, Segal DJ, Trautman JK, Smith J, Kim YG, et al. Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. Mol Cell Biol. 2001;21(1):289-97.

41. Bibikova M, Golic M, Golic KG, Carroll D. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics. 2002;161(3):1169-75.

42. 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;326(5959):1509-12.

43. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, et al. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics. 2010;186(2):757-61.

44. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. Journal of bacteriology. 1987;169(12):5429-33.

45. Jeffreys AJ, MacLeod A, Tamaki K, Neil DL, Monckton DG. Minisatellite repeat coding as a digital approach to DNA typing. Nature. 1991;354(6350):204-9.

46. Groenen PM, Bunschoten AE, van Soolingen D, van Embden JD. Nature of DNA polymorphism in the direct repeat cluster of Mycobacterium tuberculosis; application for strain differentiation by a novel typing method. Molecular microbiology. 1993;10(5):1057-65.

47. Mojica FJ, Diez-Villasenor C, Soria E, Juez G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Molecular microbiology. 2000;36(1):244-6.

48. Jansen R, Embden JD, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. Molecular microbiology. 2002;43(6):1565-75.

49. Pourcel C, Salvignol G, Vergnaud G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology. 2005;151(Pt 3):653-63.

50. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. Journal of molecular evolution. 2005;60(2):174-82.

51. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology. 2005;151(Pt 8):2551-61.

52. Tang TH, Bachellerie JP, Rozhdestvensky T, Bortolin ML, Huber H, Drungowski M, et al. Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. Proc Natl Acad Sci U S A. 2002;99(11):7536-41.

53. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology. 2009;155(Pt 3):733-40.

54. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007;315(5819):1709-12.

55. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes

Virus population dynamics and acquired virus resistance in natural microbial communities. Science. 2008;321(5891):960-4.

56. Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science. 2008;322(5909):1843-5.

57. Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, Boyaval P, et al. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. Journal of bacteriology. 2008;190(4):1390-400.

58. Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature. 2010;468(7320):67-71.

59. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature. 2011;471(7340):602-7.

60. Horvath P, Barrangou R. CRISPR/Cas, the Immune System of Bacteria and Archaea. Science. 2010;327(5962):167.

61. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, et al. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol. 2011;9(6):467-77.

62. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, et al. An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Micro. 2015;13(11):722-36.

63. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proceedings of the National Academy of Sciences. 2012;109(39):E2579–E86.

64. 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;337(6096):816-21.

65. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature biotechnology. 2013.

66. Mali P, Esvelt KM, Church GM. Cas9 as a versatile tool for engineering biology. Nat Methods. 2013;10(10):957-63.

67. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339(6121):819-23.

68. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013;8(11):2281-308.

69. Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nature biotechnology. 2014.

70. Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. 2014;24(1):132-41.

71. 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;154(6):1380-9.

72. Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science. 2014;346(6213):1258096.

73. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell. 2014;159(3):647-61.

74. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat Protoc. 2013;8(11):2180-96.

75. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013;154(2):442-51.

76. Kearns NA, Pham H, Tabak B, Genga RM, Silverstein NJ, Garber M, et al. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. Nat Methods. 2015;12(5):401-3.

77. Hilton IB, Gersbach CA. Enabling functional genomics with genome engineering. Genome Res. 2015;25(10):1442-55.

78. Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nature biotechnology. 2015;33(5):510-7.

79. Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell. 2013;155(7):1479-91.

80. Ma H, Tu LC, Naseri A, Huisman M, Zhang S, Grunwald D, et al. Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. Nature biotechnology. 2016;34(5):528-30.

81. Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015;523(7561):481-5.

82. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature. 2016;529(7587):490-5.

83. Nihongaki Y, Yamamoto S, Kawano F, Suzuki H, Sato M. CRISPR-Cas9-based photoactivatable transcription system. Chemistry & biology. 2015;22(2):169-74.

84. Dow LE, Fisher J, O'Rourke KP, Muley A, Kastenhuber ER, Livshits G, et al. Inducible in vivo genome editing with CRISPR-Cas9. Nature biotechnology. 2015;33(4):390-4.

85. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, et al. In vivo genome editing using Staphylococcus aureus Cas9. Nature. 2015;520(7546):186-91.

86. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell. 2015;163(3):759-71.

87. Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI, Gootenberg JS, Semenova E, et al. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Molecular cell. 2015;60(3):385-97.

88. Thomas KR, Folger KR, Capecchi MR. High frequency targeting of genes to specific sites in the mammalian genome. Cell. 1986;44(3):419-28.

89. Thomas KR, Capecchi MR. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell. 1987;51(3):503-12.

90. Dow LE. Modeling Disease In Vivo With CRISPR/Cas9. Trends in molecular medicine. 2015;21(10):609-21.

91. Park KE, Powell A, Sandmaier SE, Kim CM, Mileham A, Donovan DM, et al. Targeted gene knock-in by CRISPR/Cas ribonucleoproteins in porcine zygotes. Sci Rep. 2017;7:42458.

92. van der Worp HB, Sena ES, Donnan GA, Howells DW, Macleod MR. Hypothermia in animal models of acute ischaemic stroke: a systematic review and meta-analysis. Brain : a journal of neurology. 2007;130(Pt 12):3063-74.

93. Williams D, Pinzon C, Oldeschulte J, Huggins S, Gaddy D, Suva L. Modeling Hypophosphatasia in Ovies aries (sheep) using CRISPR/Cas9: Development of large animal models that phenocopy rare human bone diseases. The FASEB Journal. 2017;31(1 Supplement):902.1.

94. Wang X, Cao C, Huang J, Yao J, Hai T, Zheng Q, et al. One-step generation of triple gene-targeted pigs using CRISPR/Cas9 system. Sci Rep. 2016;6:20620.

95. Hai T, Teng F, Guo R, Li W, Zhou Q. One-step generation of knockout pigs by zygote injection of CRISPR/Cas system. Cell Res. 2014;24(3):372-5.

96. Forabosco F, Löhmus M, Rydhmer L, Sundström LF. Genetically modified farm animals and fish in agriculture: A review. Livestock Science. 2013(0).

97. Hammer RE, Brinster RL, Rosenfeld MG, Evans RM, Mayo KE. Expression of human growth hormone-releasing factor in transgenic mice results in increased somatic growth. Nature. 1985;315(6018):413-6.

98. Pursel VG, Hammer RE, Bolt DJ, Palmiter RD, Brinster RL. Integration, expression and germ-line transmission of growth-related genes in pigs. Journal of reproduction and fertility Supplement. 1990;41:77-87.

99. Pursel VG, Bolt DJ, Miller KF, Pinkert CA, Hammer RE, Palmiter RD, et al. Expression and performance in transgenic pigs. Journal of reproduction and fertility Supplement. 1990;40:235-45.

100. Vize PD, Michalska AE, Ashman R, Lloyd B, Stone BA, Quinn P, et al. Introduction of a porcine growth hormone fusion gene into transgenic pigs promotes growth. Journal of cell science. 1988;90 (Pt 2):295-300.

101. Cook JT, McNiven MA, Richardson GF, Sutterlin AM. Growth rate, body composition and feed digestibility/conversion of growth-enhanced transgenic Atlantic salmon (Salmo salar). Aquaculture. 2000;188(1):15-32.

102. Martínez R, Arenal A, Estrada MP, Herrera F, Huerta V, Vázquez J, et al. Mendelian transmission, transgene dosage and growth phenotype in transgenic tilapia (Oreochromis hornorum) showing ectopic expression of homologous growth hormone. Aquaculture. 1999;173(1):271-83.

103. Rahman MA, Mak R, Ayad H, Smith A, Maclean N. Expression of a novel piscine growth hormone gene results in growth enhancement in transgenic tilapia (Oreochromis niloticus). Transgenic Res. 1998;7(5):357-69.

104. Richt JA, Kasinathan P, Hamir AN, Castilla J, Sathiyaseelan T, Vargas F, et al. Production of cattle lacking prion protein. Nature biotechnology. 2007;25(1):132-8.

105. Yu G, Chen J, Xu Y, Zhu C, Yu H, Liu S, et al. Generation of goats lacking prion protein. Mol Reprod Dev. 2009;76(1):3.

106. Maga EA, Cullor JS, Smith W, Anderson GB, Murray JD. Human lysozyme expressed in the mammary gland of transgenic dairy goats can inhibit the growth of

bacteria that cause mastitis and the cold-spoilage of milk. Foodborne pathogens and disease. 2006;3(4):384-92.

107. van Berkel PH, Welling MM, Geerts M, van Veen HA, Ravensbergen B, Salaheddine M, et al. Large scale production of recombinant human lactoferrin in the milk of transgenic cows. Nature biotechnology. 2002;20(5):484-7.

108. Wall RJ, Powell AM, Paape MJ, Kerr DE, Bannerman DD, Pursel VG, et al. Genetically enhanced cows resist intramammary Staphylococcus aureus infection. Nature biotechnology. 2005;23(4):445-51.

109. Kerr DE, Plaut K, Bramley AJ, Williamson CM, Lax AJ, Moore K, et al. Lysostaphin expression in mammary glands confers protection against staphylococcal infection in transgenic mice. Nature biotechnology. 2001;19(1):66-70.

110. Lai L, Kang JX, Li R, Wang J, Witt WT, Yong HY, et al. Generation of cloned transgenic pigs rich in omega-3 fatty acids. Nature biotechnology. 2006;24(4):435-6.

111. Saeki K, Matsumoto K, Kinoshita M, Suzuki I, Tasaka Y, Kano K, et al. Functional expression of a Delta12 fatty acid desaturase gene from spinach in transgenic pigs. Proc Natl Acad Sci U S A. 2004;101(17):6361-6.

112. Lyall J, Irvine RM, Sherman A, McKinley TJ, N√∫√±ez A, Purdie A, et al. Suppression of Avian Influenza Transmission in Genetically Modified Chickens. Science. 2011;331(6014):223-6.

113. Burkard CA-Ohoo, Lillico SGA-Ohoo, Reid EA-Ohoo, Jackson B, Mileham AJ, Ait-Ali T, et al. Precision engineering for PRRSV resistance in pigs: Macrophages from genome edited pigs lacking CD163 SRCR5 domain are fully resistant to both PRRSV genotypes while maintaining biological function. (1553-7374 (Electronic)).

114. Whitworth KM, Rowland RRR, Ewen CL, Trible BR, Kerrigan MA, Cino-Ozuna AG, et al. Gene-edited pigs are protected from porcine reproductive and respiratory syndrome virus. Nat Biotech. 2016;34(1):20-2.

115. Ostrov N, Landon M, Guell M, Kuznetsov G, Teramoto J, Cervantes N, et al. Design, synthesis, and testing toward a 57-codon genome. Science. 2016;353(6301):819-22.

116. Hammond A, Galizi R, Kyrou K, Simoni A, Siniscalchi C, Katsanos D, et al. A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. Nat Biotech. 2016;34(1):78-83.

117. Jasin M, Haber JE. The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair. DNA Repair (Amst). 2016;44:6-16.

118. Bull JJ, Malik HS. The gene drive bubble: New realities. PLoS Genetics. 2017;13(7):e1006850.

119. Niu D, Wei HJ, Lin L, George H, Wang T, Lee IH, et al. Inactivation of porcine endogenous retrovirus in pigs using CRISPR-Cas9. Science. 2017;357(6357):1303-7.

120. Niemann H, Petersen B. The production of multi-transgenic pigs: update and perspectives for xenotransplantation. Transgenic Res. 2016;25(3):361-74.

121. Wu J, Platero-Luengo A, Sakurai M, Sugawara A, Gil MA, Yamauchi T, et al. Interspecies Chimerism with Mammalian Pluripotent Stem Cells. Cell. 2017;168(3):473-86.e15.

122. Dove A. Uncorking the biomanufacturing bottleneck. Nature biotechnology. 2002;20(8):777-9.

123. Gavin W, Chen L, Schofield M, Masiello N, Meade H, Echelard Y. Chapter 26 - Transgenic Cloned Goats and the Production of Recombinant Therapeutic Proteins. In: Cibelli J, Gurdon J, Wilmut I, Jaenisch R, Lanza R, West MD, et al., editors. Principles of Cloning (Second Edition). San Diego: Academic Press; 2014. p. 329-42.

124. Echelard Y, Williams JL, Destrempes MM, Koster JA, Overton SA, Pollock DP, et al. Production of recombinant albumin by a herd of cloned transgenic cattle. Transgenic Res. 2009;18(3):361-76.

125. Gavin W, Chen L, Melican D, Ziomek C, Echelard Y, Meade H. 10 - Transgenic Cloned Goats and Cows for the Production of Therapeutic Proteins. In: Nerem AALATM, editor. Principles of Regenerative Medicine. San Diego: Academic Press; 2008. p. 168-88.

126. Echelard Y, Meade HM, Ziomek CA. The First Biopharmaceutical from Transgenic Animals: ATryn®. Modern Biopharmaceuticals: Wiley-VCH Verlag GmbH; 2008. p. 995-1020.

127. Behboodi E, Ayres SL, Memili E, O'Coin M, Chen LH, Reggio BC, et al. Health and reproductive profiles of malaria antigen-producing transgenic goats derived by somatic cell nuclear transfer. Cloning and stem cells. 2005;7(2):107-18.

128. Pollock DP, Kutzko JP, Birck-Wilson E, Williams JL, Echelard Y, Meade HM. Transgenic milk as a method for the production of recombinant antibodies. J Immunol Methods. 1999;231(1-2):147-57.

129. Echelard Y. Recombinant protein production in transgenic animals. Current opinion in biotechnology. 1996;7(5):536-40.

130. Wang Y, Zhao S, Bai L, Fan J, Liu E. Expression systems and species used for transgenic animal bioreactors. Biomed Res Int. 2013;2013:580463.

131. Chung H, Kim TY, Lee SY. Recent advances in production of recombinant spider silk proteins. Current opinion in biotechnology. 2012.

132. Rising A, Widhe M, Johansson J, Hedhammar M. Spider silk proteins: recent advances in recombinant production, structure-function relationships and biomedical applications. Cell Mol Life Sci. 2011;68(2):169-84.

133. Liljeqvist S, Stahl S. Production of recombinant subunit vaccines: protein immunogens, live delivery systems and nucleic acid vaccines. J Biotechnol. 1999;73(1):1-33.

134. Houdebine LM. Production of pharmaceutical proteins by transgenic animals. Comparative immunology, microbiology and infectious diseases. 2009;32(2):107-21.

135. Houdebine LM. Antibody manufacture in transgenic animals and comparisons with other systems. Current opinion in biotechnology. 2002;13(6):625-9.

136. Dyck MK, Lacroix D, Pothier F, Sirard MA. Making recombinant proteins in animals--different systems, different applications. Trends in biotechnology. 2003;21(9):394-9.

137. Chang HHY, Pannunzio NR, Adachi N, Lieber MR. Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat Rev Mol Cell Biol. 2017;18(8):495-506.

138. Frit P, Barboule N, Yuan Y, Gomez D, Calsou P. Alternative end-joining pathway(s): bricolage at DNA breaks. DNA Repair (Amst). 2014;17:81-97.

139. Heyer WD, Ehmsen KT, Liu J. Regulation of homologous recombination in eukaryotes. Annual review of genetics. 2010;44:113-39.

140. Collis SJ, DeWeese TL, Jeggo PA, Parker AR. The life and death of DNA-PK. Oncogene. 2004;24(6):949-61.

141. Panier S, Boulton SJ. Double-strand break repair: 53BP1 comes into focus. Nat Rev Mol Cell Biol. 2014;15(1):7-18.

142. Lieber MR. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annual review of biochemistry. 2010;79:181-211.

143. Mao Z, Bozzella M, Seluanov A, Gorbunova V. Comparison of nonhomologous end joining and homologous recombination in human cells. DNA Repair (Amst). 2008;7(10):1765-71.

144. Hustedt N, Durocher D. The control of DNA repair by the cell cycle. Nature cell biology. 2016;19(1):1-9.

145. Wyatt DW, Feng W, Conlin MP, Yousefzadeh MJ, Roberts SA, Mieczkowski P, et al. Essential Roles for Polymerase theta-Mediated End Joining in the Repair of Chromosome Breaks. Molecular cell. 2016;63(4):662-73.

146. Richardson CD, Kazane KR, Feng SJ, Bray NL, Schaefer AJ, Floor S, et al. CRISPR-Cas9 Genome Editing In Human Cells Works Via The Fanconi Anemia Pathway. bioRxiv. 2017.

147. Bhargava R, Onyango DO, Stark JM. Regulation of Single-Strand Annealing and its Role in Genome Maintenance. Trends in genetics : TIG. 2016;32(9):566-75.

148. Lisby M, Rothstein R. Cell biology of mitotic recombination. Cold Spring Harbor perspectives in biology. 2015;7(3):a016535.

149. Moynahan ME, Jasin M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. Nat Rev Mol Cell Biol. 2010;11(3):196-207.

150. Amaral N, Ryu T, Li X, Chiolo I. Nuclear Dynamics of Heterochromatin Repair. Trends in genetics : TIG. 2017;33(2):86-100.

151. Caridi PC, Delabaere L, Zapotoczny G, Chiolo I. And yet, it moves: nuclear and chromatin dynamics of a heterochromatic double-strand break. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2017;372(1731).

152. Ceccaldi R, Rondinelli B, D'Andrea AD. Repair Pathway Choices and Consequences at the Double-Strand Break. Trends in cell biology. 2016;26(1):52-64.

153. Ciccia A, Elledge SJ. The DNA damage response: making it safe to play with knives. Molecular cell. 2010;40(2):179-204.

154. Bothmer A, Phadke T, Barrera LA, Margulies CM, Lee CS, Buquicchio F, et al. Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. Nature communications. 2017;8:13905.

155. Mimitou EP, Symington LS. Ku prevents Exo1 and Sgs1-dependent resection of DNA ends in the absence of a functional MRX complex or Sae2. Embo j. 2010;29(19):3358-69.

156. Jazayeri A, Falck J, Lukas C, Bartek J, Smith GC, Lukas J, et al. ATM- and cell cycle-dependent regulation of ATR in response to DNA double-strand breaks. Nature cell biology. 2006;8(1):37-45.

157. Ira G, Pellicioli A, Balijja A, Wang X, Fiorani S, Carotenuto W, et al. DNA end resection, homologous recombination and DNA damage checkpoint activation require CDK1. Nature. 2004;431(7011):1011-7.

158. Gutschner T, Haemmerle M, Genovese G, Draetta GF, Chin L. Post-translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair. Cell reports. 2016;14(6):1555-66.

159. Beucher A, Birraux J, Tchouandong L, Barton O, Shibata A, Conrad S, et al. ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. Embo j. 2009;28(21):3413-27.

160. Zelensky AN, Schimmel J, Kool H, Kanaar R, Tijsterman M. Inactivation of Pol theta and C-NHEJ eliminates off-target integration of exogenous DNA. Nature communications. 2017;8(1):66.

161. Saito S, Maeda R, Adachi N. Dual loss of human POLQ and LIG4 abolishes random integration. Nature communications. 2017;8:16112.

162. Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nature biotechnology. 2014.

163. Sternberg SH, LaFrance B, Kaplan M, Doudna JA. Conformational control of DNA target cleavage by CRISPR–Cas9. Nature. 2015;527(7576):110-3.

164. Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nature biotechnology. 2016;34(3):339-44.

165. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature. 2014;507(7490):62-7.

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

167. 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. Nature biotechnology. 2015;33(5):543-8.

168. Maruyama T, Dougan SK, Truttmann MC, Bilate AM, Ingram JR, Ploegh HL. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nature biotechnology. 2015;33(5):538-42.

169. Vartak SV, Raghavan SC. Inhibition of nonhomologous end joining to increase the specificity of CRISPR/Cas9 genome editing. The FEBS journal. 2015;282(22):4289-94.

170. Robert F, Barbeau M, Ethier S, Dostie J, Pelletier J. Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. Genome medicine. 2015;7:93.

171. Bertolini LR, Bertolini M, Maga EA, Madden KR, Murray JD. Increased gene targeting in Ku70 and Xrcc4 transiently deficient human somatic cells. Mol Biotechnol. 2009;41(2):106-14.

172. Song J, Yang D, Xu J, Zhu T, Chen YE, Zhang J. RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. Nature communications. 2016;7.

173. Maga EA, Sargent RG, Zeng H, Pati S, Zarling DA, Oppenheim SM, et al. Increased efficiency of transgenic livestock production. Transgenic Res. 2003;12(4):485-96.

174. Yu C, Liu Y, Ma T, Liu K, Xu S, Zhang Y, et al. Small molecules enhance CRISPR genome editing in pluripotent stem cells. Cell Stem Cell. 2015;16(2):142-7.

175. Wright AV, Sternberg SH, Taylor DW, Staahl BT, Bardales JA, Kornfeld JE, et al. Rational design of a split-Cas9 enzyme complex. Proc Natl Acad Sci U S A. 2015;112(10):2984-9.

176. Nihongaki Y, Kawano F, Nakajima T, Sato M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. Nature biotechnology. 2015;33(7):755-60.

177. Lin S, Staahl BT, Alla RK, Doudna JA. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife. 2014;4.

178. Greco GE, Matsumoto Y, Brooks RC, Lu Z, Lieber MR, Tomkinson AE. SCR7 is neither a selective nor a potent inhibitor of human DNA ligase IV. DNA Repair (Amst). 2016;43:18-23.

179. Jayathilaka K, Sheridan SD, Bold TD, Bochenska K, Logan HL, Weichselbaum RR, et al. A chemical compound that stimulates the human homologous recombination protein RAD51. Proc Natl Acad Sci U S A. 2008;105(41):15848-53.

180. Pinder J, Salsman J, Dellaire G. Nuclear domain 'knock-in' screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. Nucleic Acids Res. 2015;43(19):9379-92.

181. Liang X, Potter J, Kumar S, Ravinder N, Chesnut JD. Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. J Biotechnol. 2017;241:136-46.

182. Huang Y, McCann C, Samsonov A, Malkov D, Davis GD, Ji Q. Modulation of Genome Editing Outcomes by Cell Cycle Control of Cas9 Expression. bioRxiv. 2017.

183. Brosnahan MM, Brooks SA, Antczak DF. Equine clinical genomics: A clinician's primer. Equine Vet J. 2010;42(7):658-70.

184. Ward TL, Valberg SJ, Adelson DL, Abbey CA, Binns MM, Mickelson JR. Glycogen branching enzyme (GBE1) mutation causing equine glycogen storage disease IV. Mamm Genome. 2004;15(7):570-7.

185. Valberg SJ, Mickelson J. Glycogen branching enzyme deficiency. Proceedings of the 52nd Annual American Association of Equine Practitioners Convention, San Antonio, TX. 2006:351–3.

186. Valberg SJ, Dyson SJ. Chapter 83 - Skeletal Muscle and Lameness. In: Ross MW, Dyson SJ, editors. Diagnosis and Management of Lameness in the Horse (Second Edition). Saint Louis: W.B. Saunders; 2011. p. 818-39.

187. Valberg SJ. CHAPTER 12 - Muscle anatomy, physiology, and adaptations to exercise and training. In: Hodgson DR, McKeever KH, McGowan CM, editors. The Athletic Horse (SECOND EDITION): W.B. Saunders; 2014. p. 174-201.

188. Tryon RC, Penedo MC, McCue ME, Valberg SJ, Mickelson JR, Famula TR, et al. Evaluation of allele frequencies of inherited disease genes in subgroups of American Quarter Horses. J Am Vet Med Assoc. 2009;234(1):120-5.

189. Wagner ML, Valberg SJ, Ames EG, Bauer MM, Wiseman JA, Penedo MC, et al. Allele frequency and likely impact of the glycogen branching enzyme deficiency gene in Quarter Horse and Paint Horse populations. J Vet Intern Med. 2006;20(5):1207-11.

190. Tan WS, Carlson DF, Walton MW, Fahrenkrug SC, Hackett PB. Precision editing of large animal genomes. Adv Genet. 2012;80:37-97.

191. OMIM. OMIM Gene Map Statistics: Online Mendelian Inheritance in Man, Johns Hopkins University; 2017 [updated August 30th, 2017. Available from: www.omim.org/statistics/geneMap).

192. Bengtsson NE, Hall JK, Odom GL, Phelps MP, Andrus CR, Hawkins RD, et al. Muscle-specific CRISPR/Cas9 dystrophin gene editing ameliorates pathophysiology in a mouse model for Duchenne muscular dystrophy. Nature communications. 2017;8:14454.

193. Kemaladewi DU, Maino E, Hyatt E, Hou H, Ding M, Place KM, et al. Correction of a splicing defect in a mouse model of congenital muscular dystrophy type 1A using a homology-directed-repair-independent mechanism. Nat Med. 2017;23(8):984-9.

194. Porteus MH, Baltimore D. Chimeric nucleases stimulate gene targeting in human cells. Science. 2003;300(5620):763.

195. Barrangou R, Doudna JA. Applications of CRISPR technologies in research and beyond. Nature biotechnology. 2016;34(9):933-41.

196. Cong L, Zhang F. Genome engineering using CRISPR-Cas9 system. Methods Mol Biol. 2015;1239:197-217.

197. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013;339(6121):823-6.

198. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell. 2014;157(6):1262-78.

199. Tebas P, Stein D, Tang WW, Frank I, Wang SQ, Lee G, et al. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. The New England journal of medicine. 2014;370(10):901-10.

200. King A. Gene therapy: A new chapter. Nature. 2016;537(7621):S158-S9.

201. Couzin-Frankel J. Second chapter. Science. 2016;353(6303):983.

202. Mullard A, Dolgin E. FDA approves first CAR T therapy

Epic $12 billion deal and FDA's approval raise CAR-T to new heights. (1474-1784 (Electronic)).

203. Ma H, Marti-Gutierrez N, Park S-W, Wu J, Lee Y, Suzuki K, et al. Correction of a pathogenic gene mutation in human embryos. Nature. 2017;548(7668):413-9.

204. Wilmut I, Schnieke AE, McWhir J, Kind AJ, Campbell KH. Viable offspring derived from fetal and adult mammalian cells. Nature. 1997;385(6619):810-3.

205. Campbell KH, McWhir J, Ritchie WA, Wilmut I. Sheep cloned by nuclear transfer from a cultured cell line. Nature. 1996;380(6569):64-6.

206. Li H, Wang G, Hao Z, Zhang G, Qing Y, Liu S, et al. Generation of biallelic knock-out sheep via gene-editing and somatic cell nuclear transfer. Sci Rep. 2016;6:33675.

207. Liu J, Luo Y, Liu Q, Zheng L, Yang Z, Wang Y, et al. Production of cloned embryos from caprine mammary epithelial cells expressing recombinant human beta-defensin-3. Theriogenology. 2013;79(4):660-6.

208. Zhang P, Liu P, Dou H, Chen L, Chen L, Lin L, et al. Handmade cloned transgenic sheep rich in omega-3 Fatty acids. PLoS One. 2013;8(2):e55941.

209. Wu X, Ouyang H, Duan B, Pang D, Zhang L, Yuan T, et al. Production of cloned transgenic cow expressing omega-3 fatty acids. Transgenic Res. 2012;21(3):537-43.

210. Zhang P, Zhang Y, Dou H, Yin J, Chen Y, Pang X, et al. Handmade cloned transgenic piglets expressing the nematode fat-1 gene. Cell Reprogram. 2012;14(3):258-66.

211. Vajta G, Gjerris M. Science and technology of farm animal cloning: state of the art. Animal reproduction science. 2006;92(3-4):211-30.

212. Kurome M, Ueda H, Tomii R, Naruse K, Nagashima H. Production of transgenic-clone pigs by the combination of ICSI-mediated gene transfer with somatic cell nuclear transfer. Transgenic Res. 2006;15(2):229-40.

213. Galli C, Lagutina I, Crotti G, Colleoni S, Turini P, Ponderato N, et al. Pregnancy: a cloned horse born to its dam twin. Nature. 2003;424(6949):635.

214. Wells D. Nuclear Transfer from Established Cell Lines. eLS: John Wiley & Sons, Ltd; 2001.

215. Polejaeva IA, Chen SH, Vaught TD, Page RL, Mullins J, Ball S, et al. Cloned pigs produced by nuclear transfer from adult somatic cells. Nature. 2000;407(6800):86-90.

216. Tan W, Carlson DF, Lancto CA, Garbe JR, Webster DA, Hackett PB, et al. Efficient nonmeiotic allele introgression in livestock using custom endonucleases. Proc Natl Acad Sci U S A. 2013;110(41):16526-31.

217. Carlson DF, Lancto CA, Zang B, Kim ES, Walton M, Oldeschulte D, et al. Production of hornless dairy cattle from genome-edited cell lines. Nature biotechnology. 2016;34(5):479-81.

218. Galli C, Lagutina I, Duchi R, Colleoni S, Lazzari G. Somatic cell nuclear transfer in horses. Reprod Domest Anim. 2008;43 Suppl 2:331-7.

219. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nature biotechnology. 2016;34(2):184-91.

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

221. Sia SK, Owens MP. Share and share alike. Nat Biotech. 2015;33(12):1224-8.

222. Mathupala S, Sloan AA. An agarose-based cloning-ring anchoring method for isolation of viable cell clones. BioTechniques. 2009;46(4):305-7.

223. Pinello L, Canver MC, Hoban MD, Orkin SH, Kohn DB, Bauer DE, et al. Analyzing CRISPR genome-editing experiments with CRISPResso. Nat Biotech. 2016;34(7):695-7.

224. Khanshour A, Conant E, Juras R, Cothran EG. Microsatellite analysis of genetic diversity and population structure of Arabian horse populations. The Journal of heredity. 2013;104(3):386-98.

225. Khanshour AM, Juras R, Cothran EG. Microsatellite analysis of genetic variability in Waler horses from Australia. Australian Journal of Zoology. 2013;61(5):357-65.

226. Raudsepp T, Chowdhary BP. FISH for Mapping Single Copy Genes. In: Murphy WJ, editor. Phylogenomics. Totowa, NJ: Humana Press; 2008. p. 31-49.

227. Ugoni A, Walker BF. THE CHI SQUARE TEST: An Introduction. COMSIG review. 1995;4(3):61-4.

228. De Muth JE. Overview of biostatistics used in clinical research. American journal of health-system pharmacy : AJHP : official journal of the American Society of Health-System Pharmacists. 2009;66(1):70-81.

229. Xu B, Feng X, Burdine RD. Categorical data analysis in experimental biology. Developmental Biology. 2010;348(1):3-11.

230. Kao LS, Green CE. Analysis of Variance: Is There a Difference in Means and What Does It Mean? The Journal of surgical research. 2008;144(1):158-70.

231. Dakic A, DiVito K, Fang S, Suprynowicz F, Gaur A, Li X, et al. ROCK inhibitor reduces Myc-induced apoptosis and mediates immortalization of human keratinocytes. Oncotarget. 2016;7(41):66740-53.

232. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nature biotechnology. 2014;32(3):279-84.

233. Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, et al. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature. 2005;435(7042):646-51.

234. Yang L, Guell M, Byrne S, Yang JL, De Los Angeles A, Mali P, et al. Optimization of scarless human stem cell genome editing. Nucleic Acids Res. 2013;41(19):9049-61.

235. Byrne SM, Ortiz L, Mali P, Aach J, Church GM. Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. Nucleic Acids Res. 2015;43(3):e21.

236. Elliott B, Richardson C, Winderbaum J, Nickoloff JA, Jasin M. Gene conversion tracts from double-strand break repair in mammalian cells. Mol Cell Biol. 1998;18(1):93-101.

237. Vriend LE, Prakash R, Chen CC, Vanoli F, Cavallo F, Zhang Y, et al. Distinct genetic control of homologous recombination repair of Cas9-induced double-strand breaks, nicks and paired nicks. Nucleic Acids Res. 2016;44(11):5204-17.

238. Hollywood JA, Lee CM, Scallan MF, Harrison PT. Analysis of gene repair tracts from Cas9/gRNA double-stranded breaks in the human CFTR gene. Sci Rep. 2016;6:32230.

239. Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, et al. DNA sequencing at 40: past, present and future. Nature. 2017.

240. van Overbeek M, Capurso D, Carter MM, Thompson MS, Frias E, Russ C, et al. DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. Molecular cell. 2016;63(4):633-46.

241. Miyaoka Y, Berman JR, Cooper SB, Mayerl SJ, Chan AH, Zhang B, et al. Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. Sci Rep. 2016;6:23549.

242. Knight SC, Xie L, Deng W, Guglielmi B, Witkowsky LB, Bosanac L, et al. Dynamics of CRISPR-Cas9 genome interrogation in living cells. Science. 2015;350(6262):823-6.

243. Horlbeck MA, Witkowsky LB, Guglielmi B, Replogle JM, Gilbert LA, Villalta JE, et al. Nucleosomes impede Cas9 access to DNA in vivo and in vitro. Elife. 2016;5.

244. Ferguson BJ, Mansur DS, Peters NE, Ren H, Smith GL. DNA-PK is a DNA sensor for IRF-3-dependent innate immunity. Elife. 2012;1:e00047.

245. Peters NE, Ferguson BJ, Mazzon M, Fahy AS, Krysztofinska E, Arribas-Bosacoma R, et al. A Mechanism for the Inhibition of DNA-PK-Mediated DNA Sensing by a Virus. PLoS Pathog. 2013;9(10):e1003649.

246. Oakes BL, Nadler DC, Flamholz A, Fellmann C, Staahl BT, Doudna JA, et al. Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. Nature biotechnology. 2016;34(6):646-51.

247. Riesenberg S, Maricic T. Targeting Repair Pathways With Small Molecules Increases Precise Genome Editing In Pluripotent Stem Cells. bioRxiv. 2017.

248. Fenner F. The eradication of smallpox. Progress in medical virology Fortschritte der medizinischen Virusforschung Progres en virologie medicale. 1977;23:1-21.

249. Mazzon M, Castro C, Roberts LD, Griffin JL, Smith GL. A role for vaccinia virus protein C16 in reprogramming cellular energy metabolism. J Gen Virol. 2015;96(Pt 2):395-407.

250. Prieto C, Fontana D, Etcheverrigaray M, Kratje R. A strategy to obtain recombinant cell lines with high expression levels. Lentiviral vector-mediated transgenesis. BMC proceedings. 2011;5 Suppl 8:P7.

251. Lillico S, Vasey D, King T, Whitelaw B. Lentiviral transgenesis in livestock. Transgenic Res. 2011;20(3):441-2.

252. Brees C, Fransen M. A cost-effective approach to microporate mammalian cells with the Neon Transfection System. Anal Biochem. 2014;466:49-50.

253. Vaux DL, Fidler F, Cumming G. Replicates and repeats—what is the difference and is it significant?: A brief discussion of statistics and experimental design. EMBO reports. 2012;13(4):291-6.

254. Glaser A, McColl B, Vadolas J. GFP to BFP Conversion: A Versatile Assay for the Quantification of CRISPR/Cas9-mediated Genome Editing. Molecular therapy Nucleic acids. 2016;5(7):e334.

255. DeWitt MA, Corn JE, Carroll D. Genome editing via delivery of Cas9 ribonucleoprotein. Methods (San Diego, Calif). 2017.

256. Lee K, Mackley VA, Rao A, Chong AT, Dewitt MA, Corn JE, et al. Synthetically modified guide RNA and donor DNA are a versatile platform for CRISPR-Cas9 engineering. Elife. 2017;6.

257. Zhang JP, Li XL, Li GH, Chen W, Arakaki C, Botimer GD, et al. Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. Genome biology. 2017;18(1):35.

258. Boskovic J, Rivera-Calzada A, Maman JD, Chacón P, Willison KR, Pearl LH, et al. Visualization of DNA-induced conformational changes in the DNA repair kinase DNA-PKcs. The EMBO Journal. 2003;22(21):5875-82.

259. Paludan SR. Activation and regulation of DNA-driven immune responses. Microbiol Mol Biol Rev. 2015;79(2):225-41.

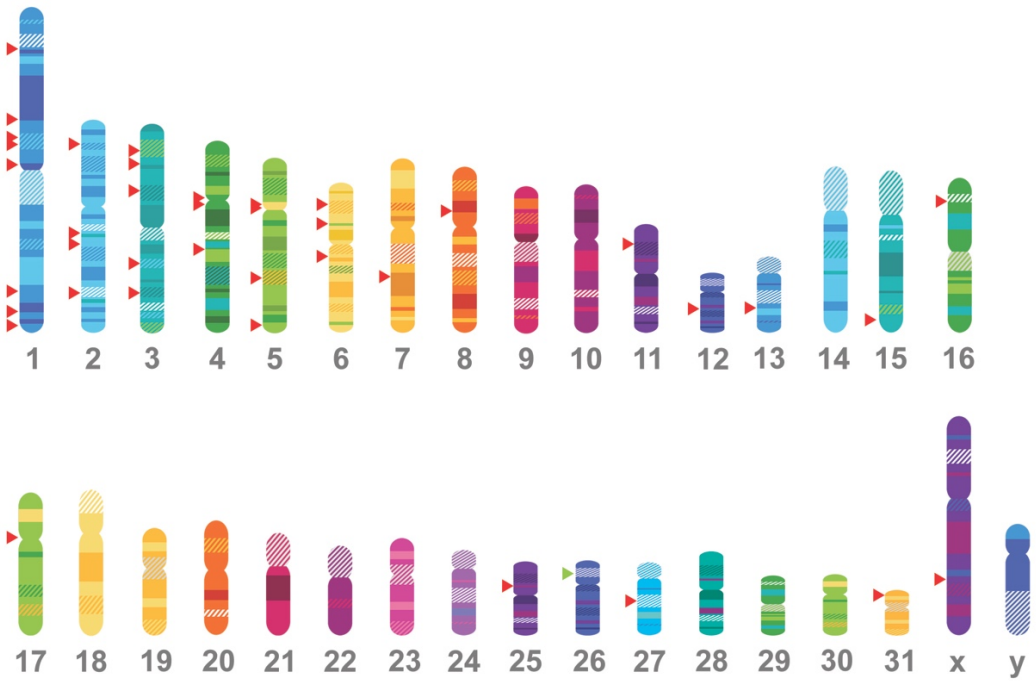
260. Orzalli MH, Knipe DM. Cellular sensing of viral DNA and viral evasion mechanisms. Annual review of microbiology. 2014;68:477-92.

261. Unterholzner L, Keating SE, Baran M, Horan KA, Jensen SB, Sharma S, et al. IFI16 is an innate immune sensor for intracellular DNA. Nature immunology. 2010;11(11):997-1004.

262. Sun L, Wu J, Du F, Chen X, Chen ZJ. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. Science. 2013;339(6121):786-91.

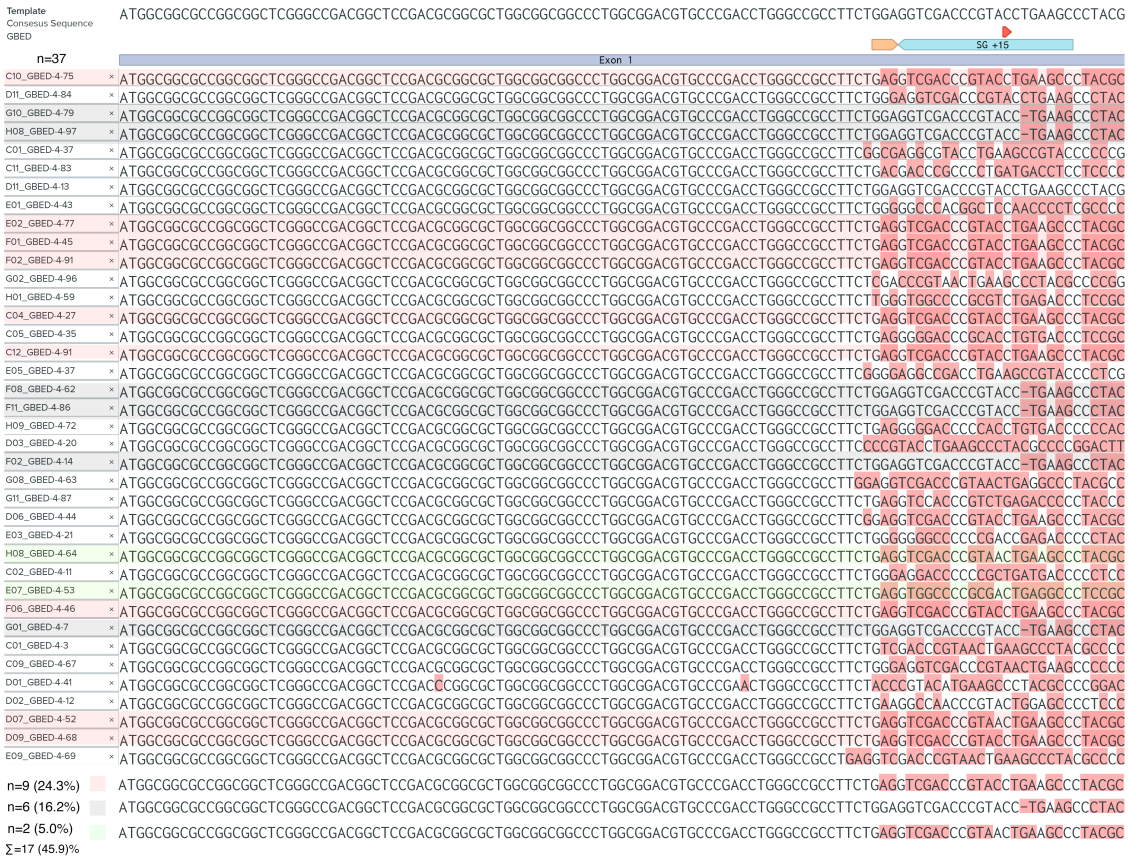
APPENDIX A

SUPPLEMENTARY FIGURES



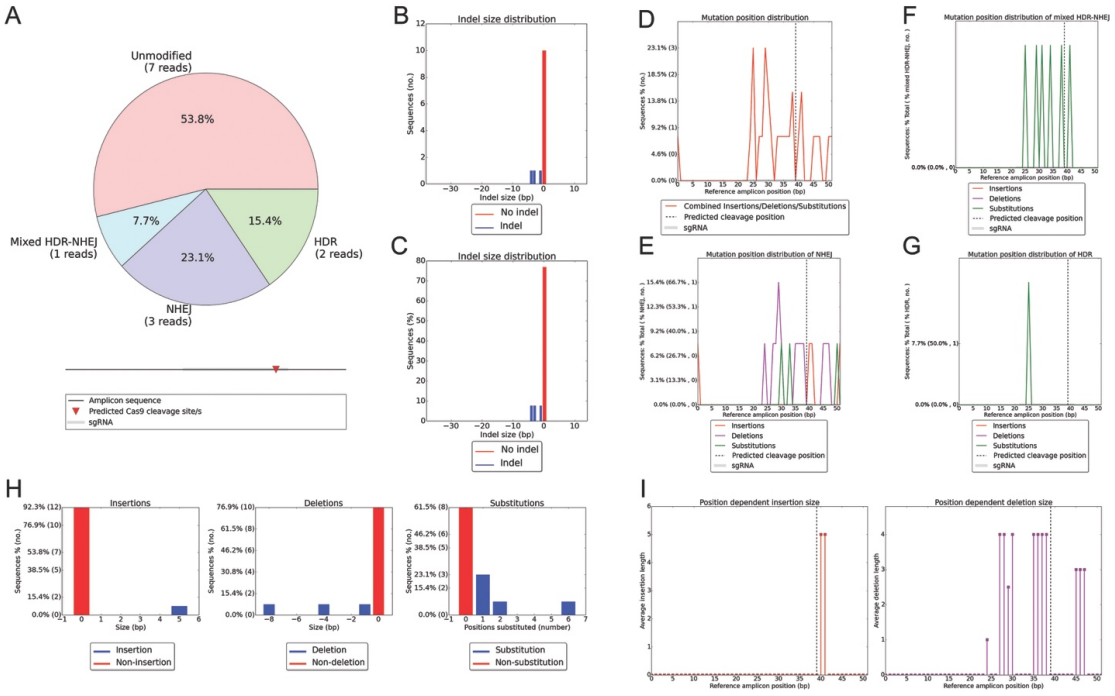
Supplementary Figure 1. Equine Chromosome Ideogram of CIRSPR-Cas9 On Target and Predicted Off Target Sites for the sgRNA +1.

Equine chromosome ideogram, illustrating only one chromosome of each of the 32 pairs with the exception of the X and Y chromosomes. Each predicted off-target chromosome and approximate location within the chromosome is shown (red arrow head). On-Target location in chromosome 26 is shown (green arrow head).



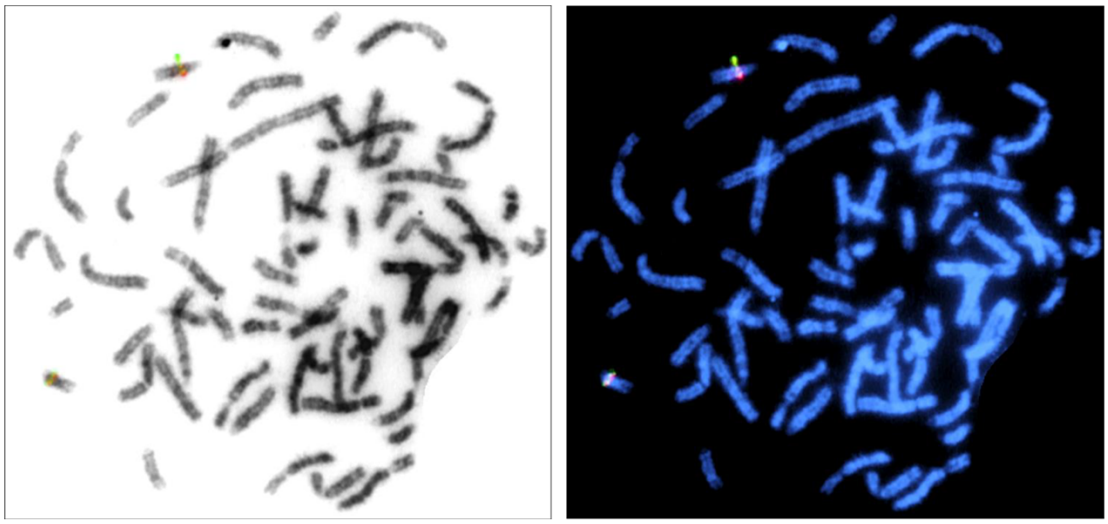
Supplementary Figure 2. CRISPR-Cas Induced DSB at the GBE1 Locus Have Unidirectional Repair Tracts and Non-Random INDEL Outcomes.

Fasta sequences of the aligned chromatographs (Benchling®) of a subset of colonies of sgRNA +15. Note that the majority of the INDEL events are located 3' of the sgRNA location, this represent the unidirectional repair tracts. A closer examination shows that several of this INDELS are non-random with approximately 45% of the events being repetitive, some events are more prevalent with 24% of the sequence reads being identical (Red square).

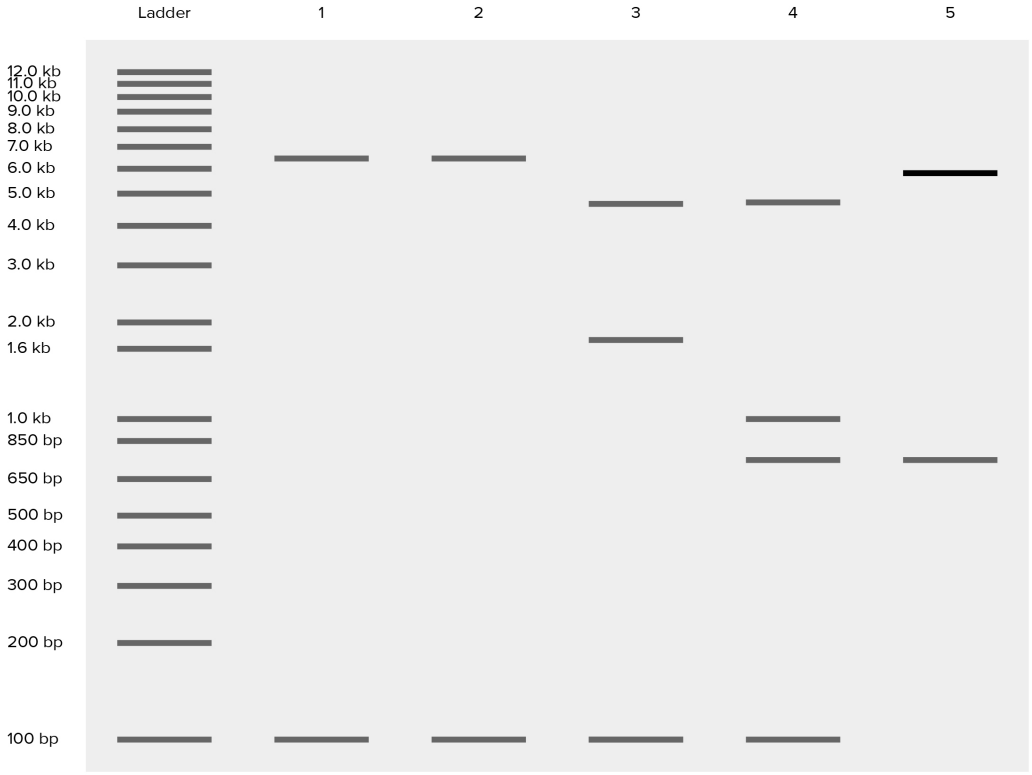


Supplementary Figure 3. CRISPResso Analysis of Genome Editing Outcomes on sgRNA +1.

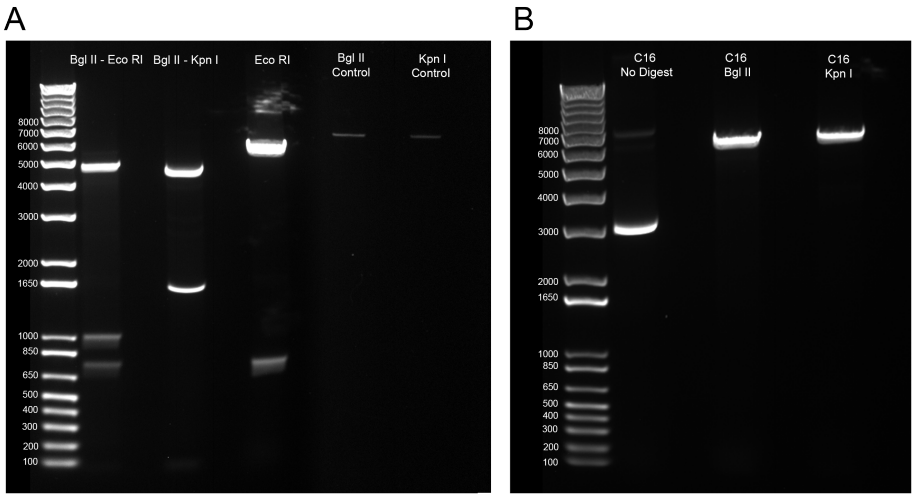
A) Quantification of editing frequency as determined by the percentage and number of sequence reads showing unmodified and modified alleles. Modified alleles are subdivided into C-NHEJ, HDR and mixed HDR-C-NHEJ alleles. Frequency distribution of **B)** size and **C**) frequency of alleles with indels (shown in blue) and without indels (in red). **D)**, **E)** C-NHEJ reads, **F)** mixed HDR-C-NHEJ reads and **G)** HDR reads with insertions (red), deletions (purpule) and substitutions (green) mapped to reference amplicon position. The predicted Cas9 clevage site is indicated by a vertical dash line. Only sequence positions directly adjacent to insertions or directly affected by deletions or substitutions are plotted. **H)** Left panel, frequency distribution of sequence modifications that increase read length with respect to the reference amplicon, classified as insertions (positive indel size). Middle panel, frequency distribution of sequence modifications that reduce read length with respect to the reference amplicon, classified as deletions (negative indel size). Right panel, frequency distribution of sequence modifications that do not alter read length with respect to the reference amplicon, which are classified as substitutions (number of substituted positions shown). **I)** Position dependent insertion size (left) and deletion size (right).



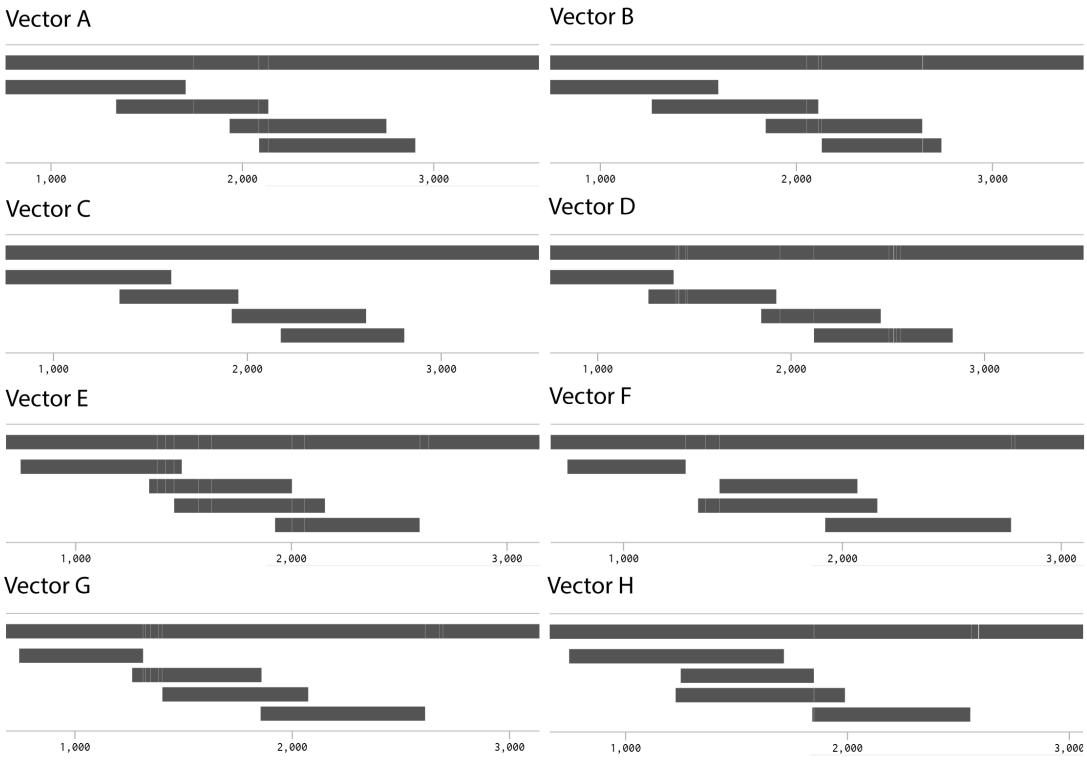
Supplementary Figure 4. FISH Stain of Isolated BAC Clones.
Fluorescent in-situ Hybridization of BACs 93G22 and 112C9 containing the GBE1 sequence of interest, that were detected conjointly with the control probes ETSTY7 bio + ETSTY dig targeting the Y chromosome.



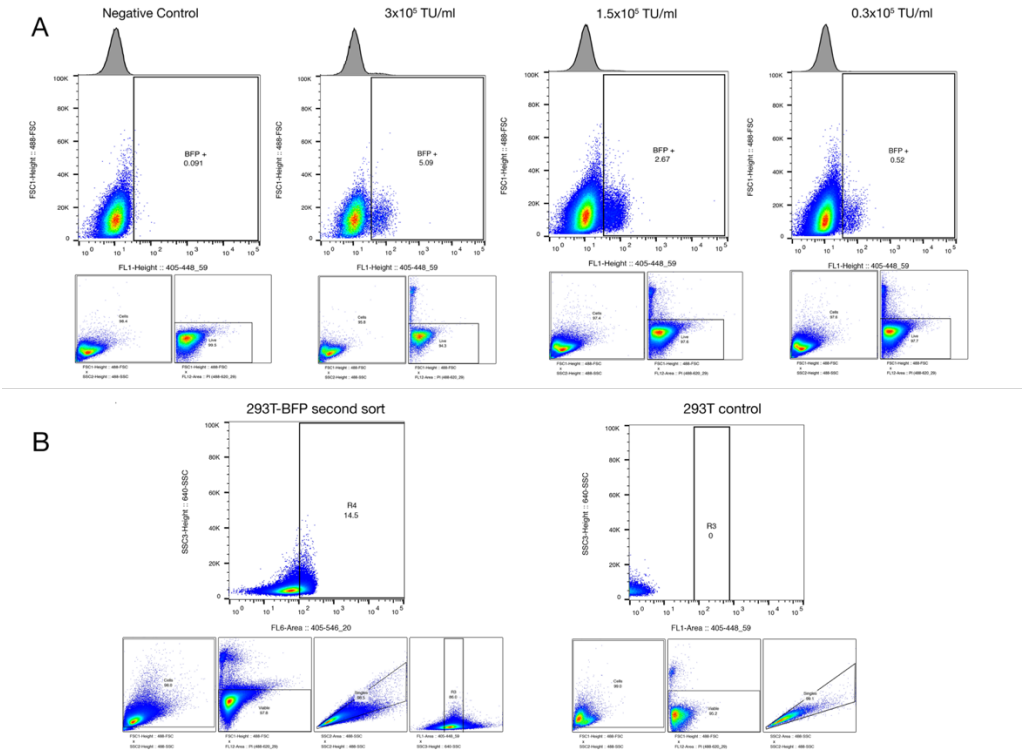
Supplementary Figure 5. Expected Digestion Fragments of C16 Vector.
This is an illustration of the ideal expected sizes of the C16 enzyme restriction fragments with the following enzymes: 1) Bgl Il; 2) Kpn I; 3) Bgl II and Kpn I; 4) Bgl II and Eco RI; 5) Eco RI.



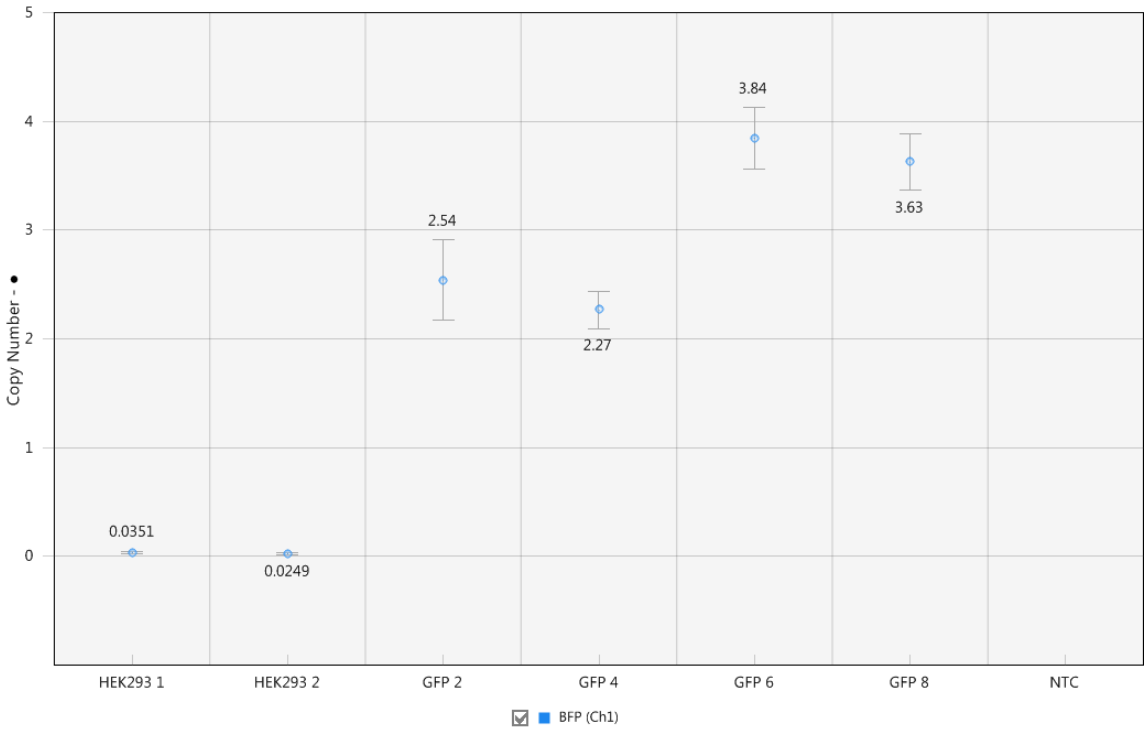
Supplementary Figure 6. Digestion Fragments of C16 Vector.



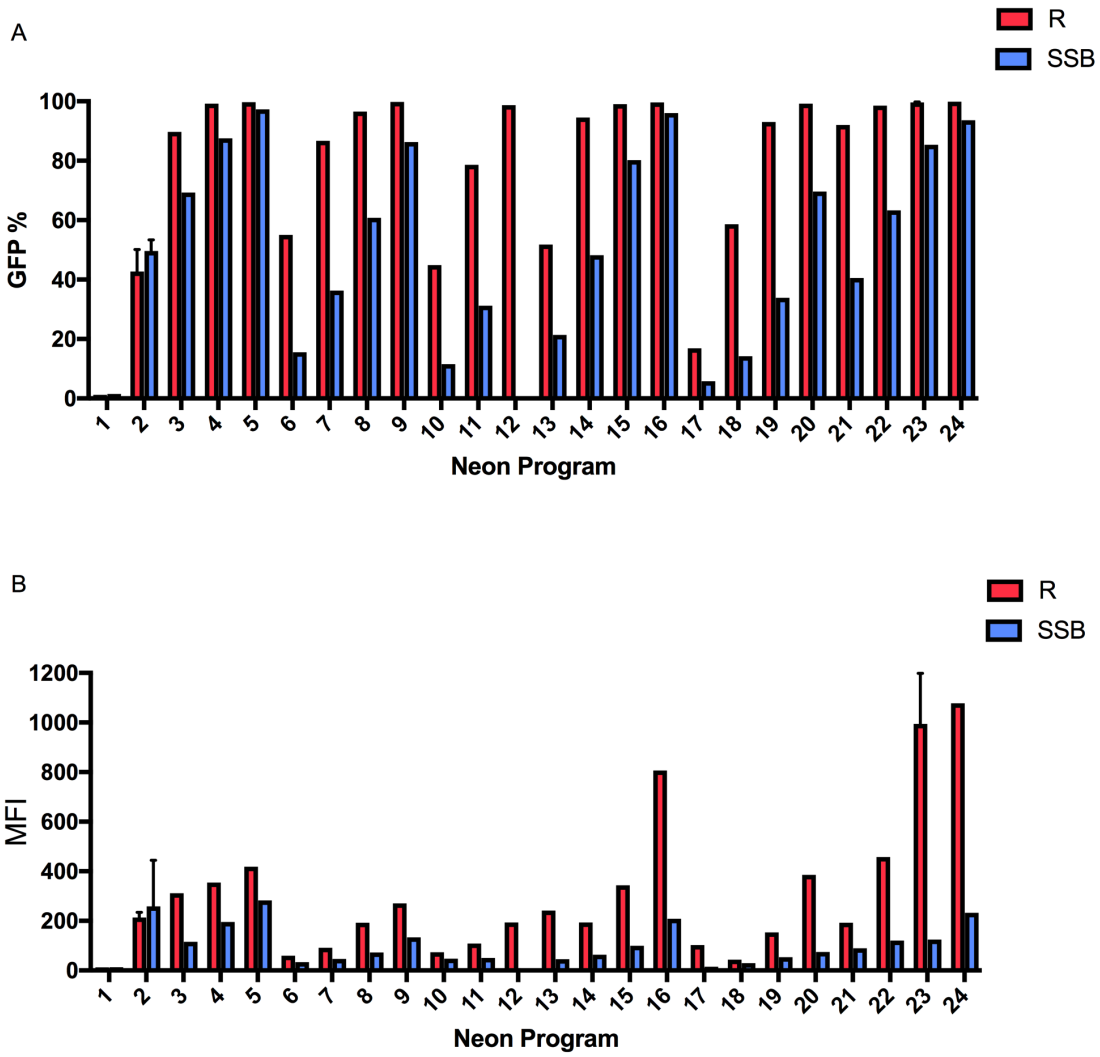
Supplementary Figure 7. Sequencing Alignment of C16 Vector Variants.



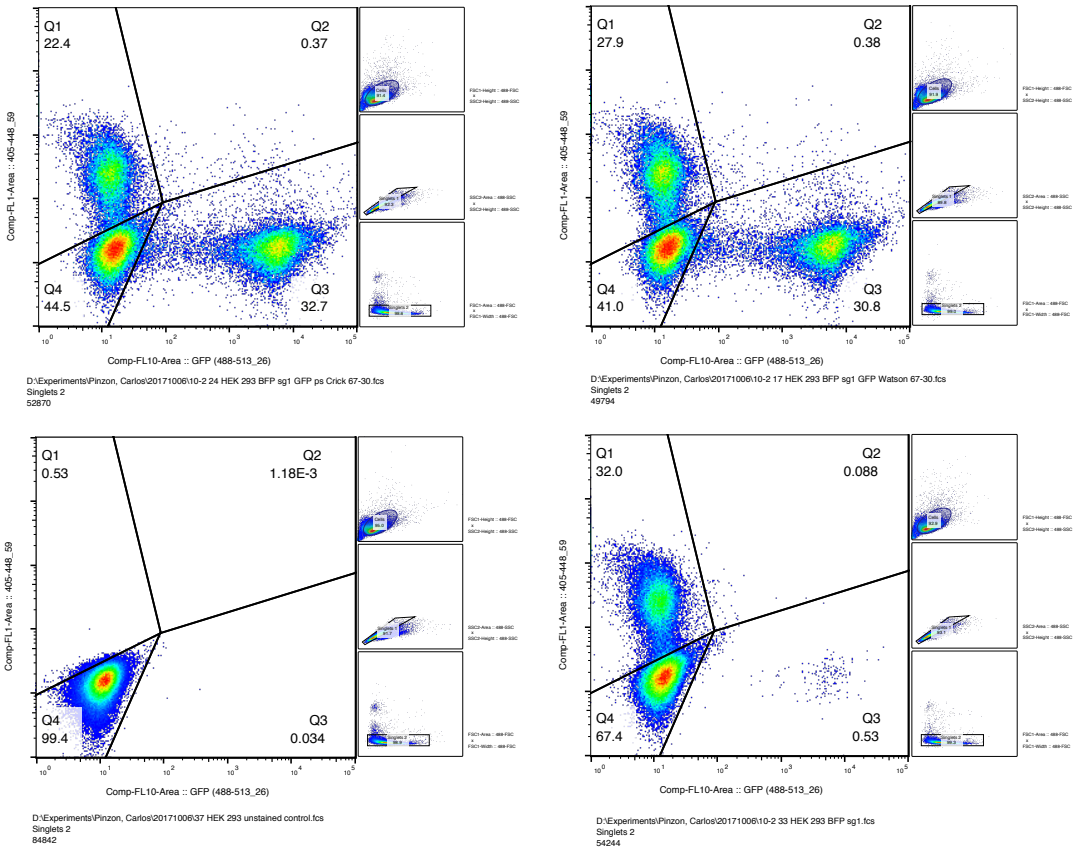
Supplementary Figure 8. FACS for the Establishment of a BFP-HEK293 Cell Line.
Fluorescent activated flow cytometry (FACS) of A) HEK293 cells transduced with lentivirus carrying puromycin and blue fluorescent protein (BFP) genes, and B) Second sort for homogeneous expression of BFP Transgene.



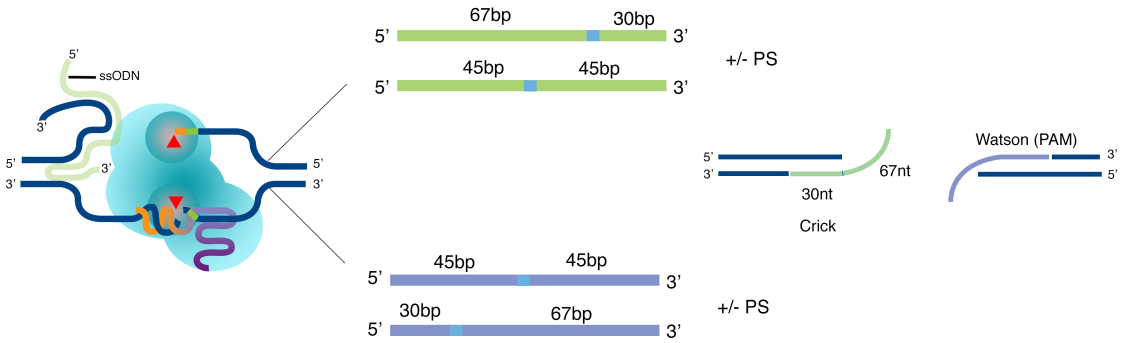
Supplementary Figure 9. ddPCR BFP Copy Number Quantification.
Copy numbers were calculated based on droplet numbers of the target, (BFP) compared to the single copy reference (RPP30) (e.g. BFP/RPP30:~ 1 = Mono-allelic, ~ 2.0 = Bi-allelic, ~ 3.0 = Tri-allelic). Data represents two HEK 293 negative controls and four technical replicate samples of genomic DNA extracted from FACS sorted GFP+ cells together with a no template control (NTC) as a negative control.



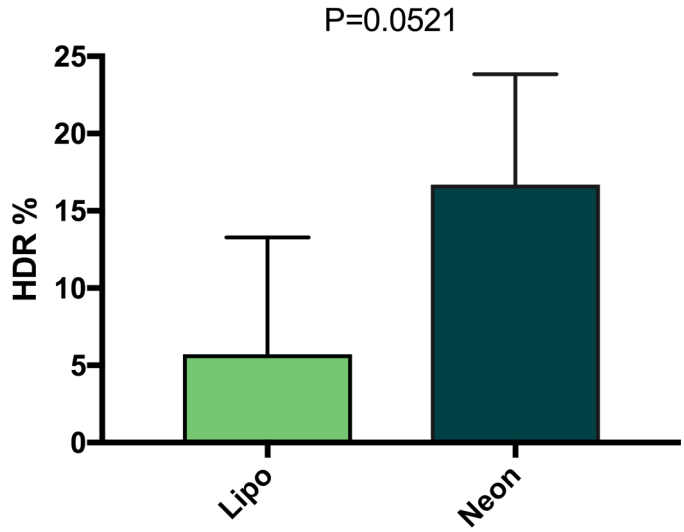
Supplementary Figure 10. Neon Transfection System Optimization.



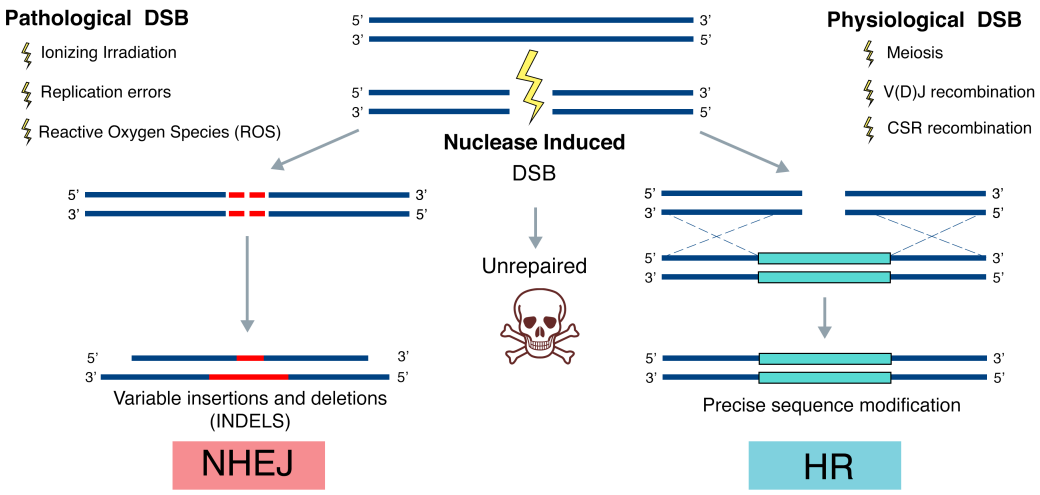
Supplementary Figure 11. Maximum HR Rates Obtained Under Our Conditions.



Supplementary Figure 12. Designed Repair Templates.
Illustration of the different single-stranded oligodeoxynucleotide (ssODN) repair templates with either asymmetric 67-30nt or symmetric 45-45nt homology arms with and without 5' and 3' phosphorothioate (PS) modifications. In in either a Watson (PAM strand for sgRNA#1) or Crick conformation.



Supplementary Figure 13. HDR Rates Comparison Between Lipofectamine and Electroporation.



Supplementary Figure 14. DNA Double Strand Breaks.

Gene editing technologies, including CRISPR-Cas9, rely on the precise introduction of a DNA double-strand breaks (DSBs) are the most dangerous type of DNA damage, which if left unrepaired, can result in chromosomal translocations, the loss of large chromosomal regions or can activate cell cycle check-point arrests and induce signals for programmed cell death. In dividing mammalian cells, there are an estimated ten DNA double-strand breaks (DSBs) per day per cell (137). These pathological DSBs arise from DNA replication errors, reactive oxygen species, and genotoxic agents such as ionizing radiation, radiomimetic chemicals or topoisomerase inhibitors (138). There are also physiological DSBs that arise during normal cellular processes, such as meiosis during gametogenesis or V(D)J recombination and class-switch recombination (CSR) which facilitate the rearrangements of antigen receptor genes in lymphogenesis (137, 138). DSBs that occur throughout the cell cycle are repaired predominantly by the non-homologous DNA end joining (NHEJ) pathway that is present during the G1, S and G2 phases and to a lesser extent by the homology-directed repair (HDR) pathway that is restricted during the late S and G2 phases(139).

APPENDIX B

SUPPLEMENTARY TABLES

Supplementary Table 1. Sequencing Primers Used.

Name	Sequence
C16 Seq FW # 1	CTT CTC CTC CGG GCT GTA AT T AG
C16 Seq FW # 2	GAT AAA ACT TTA CCA CCC TA C TG
C16 Seq FW # 3	CAC CGA ACC TCC GAC AGT CT
C16 Seq RV # 1	CTT GTT TAT TGC AGC TTA TAA TGG TTA C

Supplementary Table 2. Sequence of sgRNA Targeting the BFP Chromophore.

Name	Sequence	PAM	Strand
sgRNA BFP # 1	CTCGTGACCACCCTGAGCCA	CGG	+
sgRNA BFP # 2	GGCTGAAGCACTGGACGCCG	TGG	-
sgRNA BFP # 3	AGCACTGGACGCCGTGGCTC	AGG	-

APPENDIX C

SUPPLEMENTARY MATERIAL

Supplementary Material 1**.** GBED ssODN Repair Templates Used.

GBED Symmetric repair template

GCTCCGACGCGGCGCTGGCGGCGGCCCTGGCGGACGTGCCCGACCTGGGCC

GCCTTCTGGAGGTCGACCCGTACCTGAAGCCCTACGCCCCGGACTTCCAGCG

CA

GBED Symmetric repair template with PAM silence mutation

GCTCCGACGCGGCGCTGGCGGCGGCCCTGGCGGACGTGCCCGACCTGGGCC

GGCTTCTGGAGGTCGACCCGTACCTGAAGCCCTACGCCCCGGACTTCCAGCG

CA

GBED PS asymmetric repair template

A*C*G CGG CGC TGG CGG CGG CCC TGG CGG ACG TGC CCG ACC TGG GCC

GCC TTC TGG AGG TCG ACC CGT ACC TGA AGC CCT ACG CCC CGG ACT TCC

AGC* G*C

GBED PS asymmetric repair template

ACG CGG CGC TGG CGG CGG CCC TGG CGG ACG TGC CCG ACC TGG GCC

GCC TTC TGG AGG TCG ACC CGT ACC TGA AGC CCT ACG CCC CGG ACT TCC

AGCGC

Supplementary Material 2. GBE1 5’ UTR Exon 1 CDS.

CDS: 135…264, Feature: GBE1 exon1

CTCGCCGCTATAAAGGGCCCCGGGCCGCAGCCGCTCGCCTCGGCGTCCTCGG
CTCCGCCCTCGCGCCGGCCACTCCGCGGAGCTCGTTCCCGCTCGAGCGGCTC
GGGCCTCGGCTACTCGGGCTGCGGCCGAAGATGGCGGCGCCGGCGGCTCGG
GCCGACGGCTCCGACGCGGCGCTGGCGGCGGCCCTGGCGGACGTGCCCGAC
CTGGGCCGCCTTCTGGAGGTCGACCCGTACCTGAAGCCCTACGCCCCGGACT
TCCAGCGCA

Supplementary Material 3. C16 – mCherry Codon Optimized Construct Sequence.

CDS: 1…81, Feature: 3xNLS. CDS: 82...87, Feature: BglII. CDS: 88…1080, Feature: C16. CDS: 1081…1086, Feature: EcoRI. CDS: 1087…1116, Feature: $(Gly\ 4\ Ser\ 4)_2$ GS linker. CDS: 1117…1824, Feature: mCherry. CDS: 1825…1830, Feature: EcoRI. CDS: 1831…1836, Feature: KpnI. CDS: 1837…1911, Feature: 3xNLS. CDS: 1912…1917, Feature: KpnI. CDS: 1918…1920, Feature: Stop.

ATGAGATCTGATCCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAG
AAAGGTAGATCCAAAAAAGAAGAGAAAGGTAAGATCTATGGACATATATGA
TGATAAGGGCCTTCAGACGATTAAGCTGTTCAATAACGAATTTGATTGCATA
AGGAACGACATAAGAGAACTCTTTAAGCATGTAACTGACAGTGACTCAATTC
AACTGCCAATGGAGGACAACTCCGACATCATTGAAAATATACGAAAGATAC
TTTACAGACGACTTAAGAATGTGGAGTGCGTTGACATTGACTCAACAATTAC

CTTTATGAAGTACGATCCGAACGATGATAATAAGCGCACGTGTTCAAATTGG

GTTCCACTTACTAATAATTACATGGAATATTGCCTCGTGATATACCTCGAAA

CACCCATCTGTGGGGGCAAGATAAAACTTTACCACCCTACTGGGAATATCAA

ATCAGATAAGGATATCATGTTTGCGAAAACATTGGACTTCAAATCCAAGAAA

GTCTTGACAGGCCGGAAGACTATCGCGGTGCTTGATATTTCCGTTAGCTATA

ATCGATCCATGACGACGATCCATTACAACGACGACGTTGACATCGACATACA

CACAGATAAAAACGGAAAAGAACTCTGCTATTGCTATATCACGATCGACGA

TCACTATCTGGTCGATGTCGAAACAATCGGAGTAATCGTTAATCGGTCCGGC

AAGTGTCTGCTCGTTAATAATCATCTTGGCATCGGGATAGTCAAGGACAAAC

GCATCAGTGATTCATTCGGCGATGTGTGCATGGACACTATATTTGACTTCTCT

GAAGCTCGGGAACTCTTCAGTCTTACAAACGATGACAATCGCAATATCGCCT

GGGATACCGATAAGCTTGACGATGACACCGACATCTGGACACCTGTAACGG

AGGATGACTATAAATTTCTTAGTCGATTGGTGTTGTATGCCAAGTCTCAATC

AGATACGGTATTCGATTACTACGTCCTGACTGGGGACACCGAACCTCCGACA

GTCTTCATTTTCAAAGTCACGCGCTTCTACTTTAACATGCCGAAAGAATTCGG

TGGTGGTGGTTCTGGTGGTGGTGGTTCTATGGTGAGCAAGGGCGAGGAGGAT

AACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCT

CCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCT

ACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGC

CCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTAC

GTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGG

GCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCG

TGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCT

GCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCAT

GGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAA

GGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGC

TGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGC

CTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACC

ATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATG

GACGAGCTGTACAAGGAATTCGGTACCATGGATCCAAAAAAGAAGAGAAAG

GTAGATCCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAAAGGT

AGGTACCTAG.

Supplementary Material 4. BFP to GFP Repair templates.

GFP_WATSON_67-30:

CCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCT

CGTGACCACCCTGACCTACGGCGTCCAGTGCTTCAGCCGCTACCC

GFP_CRICK_67-30

GGCGGACTTGAAGAAGTCGTGCTGCTTCATGTGGTCGGGGTAGCGGCTGAA

GCACTGGACGCCGTAGGTCAGGGTGGTCACGAGGGTGGGCCAGGGC

GFP_PS_WATSON_67-30

C*C*CTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCC

TCGTGACCACCCTGACCTACGGCGTCCAGTGCTTCAGCCGCTAC*C*C

GFP_PS_CRICK_67-30

G*G*CGGACTTGAAGAAGTCGTGCTGCTTCATGTGGTCGGGGTAGCGGCTGA

AGCACTGGACGCCGTAGGTCAGGGTGGTCACGAGGGTGGGCCAGG*G*C

GFP_WATSON_90

CGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGC

GTCCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCA

GFP_CRICK_90

TGCTTCATGTGGTCGGGGTAGCGGCTGAAGCACTGGACGCCGTAGGTCAGG

GTGGTCACGAGGGTGGGCCAGGGCACGGGCAGCTTGCCG

GFP_PS_WATSON_90

C*G*GCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACG

GCGTCCAGTGCTTCAGCCGCTACCCCGACCACATGAAG*C*A

GFP_PS_CRICK_90

T*G*CTTCATGTGGTCGGGGTAGCGGCTGAAGCACTGGACGCCGTAGGTCAG

GGTGGTCACGAGGGTGGGCCAGGGCACGGGCAGCTTGC*C*G