PRO136ALA CTLA4 Mutation

PRO136ALA *CTLA4* Mutation

CODY PALMER

A Thesis Submitted to the Faculty of

The Harvard Medical School

in Partial Fulfillment of the Requirements

for the Degree of Master of Medical Sciences in Immunology

Harvard University

Boston, Massachusetts.

May, 2017

Thesis Advisor: Dr. Shiv Pillai Cody Palmer

PRO136ALA *CTLA4* Mutation

Abstract

Cytotoxic T lymphocyte antigen–4 (CTLA-4) is an inhibitory receptor involved in the regulation of immune responses. Deficiency of Ctla4 in mice causes fatal multiorgan lymphocytic infiltration. We identified a heterozygous, missense c.406C>G, p.P136A mutation in CTLA4 in a CVID patient. This mutation is within the binding motif of CTLA-4 and is predicted to interfere with ligand binding. To test whether this mutation is causal in the phenotype of the patient, we will use CRISPR/Cas9-mediated genome editing to clone the mutation into a regulatory T cell-like cell line and use the edited cell line for functional studies.

Table of Contents

Figures

Acknowledgements

I would like to express my sincerest gratitude to the Pillai lab at the Ragon Institute of MGH, MIT, and Harvard for all of the assistance and guidance they have provided during this work.

This work was conducted with support from Students in the Master of Medical Sciences in Immunology program of Harvard Medical School. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard University and its affiliated academic health care centers.

1. Chapter 1: Background

1.1. Background

Systems of host defense have evolved since the first primitive immune system that was in existence by the time that plants and animals diverged. Unicellular organisms such as bacteria have simple immune systems that protect against bacteriophage infections. Other defense mechanisms evolved in eukaryotes. These mechanisms are the basis of the innate immune system and include phagocytosis, antimicrobial peptides, and the complement system. Jawed vertebrates and higher members of subphylum Vertebrata have evolved more complex defense mechanisms that includes the ability to form immunological memory after an initial response to a specific pathogen which allows for a faster and more efficient immune response upon re-exposure to that pathogen[1].

Humans and other vertebrates of the class Mammalia have an adaptive immune system that consists of lymphocytes that are able to recognize specificity among pathogens via antigen receptors on their surface that are generated through somatic recombination. The adaptive immune system allows for more efficient and targeted immune responses based on recognition of a large number of self- and non-self antigens. With this capacity to generate such a large repertoire of immune cells comes the generation of self-reactive T cells and, thus, increased opportunities for inappropriate responses against self-antigens[1]. Therefore, it is critical that mechanisms are in place that aim to prevent these autoimmune events. Tolerance is controlled by multiple mechanisms, including inhibitory receptors and regulatory T cells.

Immune system dysregulation can lead to autoimmune diseases, inflammatory diseases, and cancer. Inflammation is a normal component of the biological response of tissue to harmful stimuli. The function of inflammation is to eliminate the initial cause of injury and to initiate tissue repair. Inflammatory diseases are caused by an inappropriate inflammatory response to an antigen that should not elicit an immune response under normal circumstances. Cancer is also the result of a disordered immune response. The immune system normally is capable of recognizing and eliminating cells that have undergone malignant transformation. The loss of this immunosurveillance ability is pathogenic in the development of cancer. Immunodeficiency is the result of a hypoactive immune system which causes recurring infections. Immunodeficiency can be the result of either a genetic disease such as severe combined immunodeficiency, acquired conditions such as HIV/AIDS, or the use of immunosuppressive medication. In contrast, autoimmunity is caused by a hyperactive immune system which forms

an immune response to self-tissues. Common autoimmune diseases include diabetes mellitus type 1, systemic lupus erythematosus, and rheumatoid arthritis[2, 3].

Although it may seem unlikely due to their seemingly opposite phenotypes, autoimmunity and immunodeficiency can manifest concurrently in a single individual[2]. The mechanisms behind this association are not fully known but is commonly seen in disorders such as CVID. Common variable immunodeficiency (CVID) is the most common clinical primary immunodeficiency. The immune disorder is defined by high susceptibility to infection, hypogammaglobulinemia, and impaired specific antibody response. By definition, all CVID patients have low serum IgG with low IgA or low IgM, and impaired antibody response to vaccination. The symptoms of CVID are highly variable, as the name suggests. Most commonly, patients have recurrent bacterial infections, usually of the upper respiratory tract. Infections are a direct result of the low antibody levels in the circulation, which do not adequately protect against pathogens. The most frequent cause of infections in CVID are due to the bacteria *Haemophilus influenzae* and *Streptococcus pneumoniae*. Infections mostly affect the sinopulmonary tract, however, they can also occur at other sites, such as the eyes, skin and gastrointestinal tract. Pneumonia, bronchitis, and sinusitis are common in people with CVID. About 50% of patients have at least one occurrence of pneumonia. These infections can become recurrent, leading to chronic lung disease. This chronic lung diseases manifests as bronchiectasis or interstitial lung disease. Patients with interstitial lung diseases are most likely to also have autoimmunity. Recurrent and chronic diarrhea is seen in approximately 40% of patients, frequently due to *Giardia*, *Salmonella* and *Campylobacter jejuni*. Acute and chronic gastritis related to *H. pylori* infection is also seen.

Over 50% of patients also develop complications, including enteropathy, lymphocytic infiltration of tissues, malignancy, and autoimmunity. Immune thrombocytopenic purpura and autoimmune hemolytic anemia are the most common autoimmune features, but other manifestations include rheumatoid arthritis, inflammatory bowel disease (IBD), systemic lupus erythematosus, diabetes mellitus, multiple sclerosis, and psoriasis. Accordingly, CVID presents in many patients as an immune dysregulation syndrome with variable symptoms. CVID has a nonspecific clinical presentation and laboratory definition which makes diagnosis difficult. Diagnosis of CVID is based mostly on exclusion of other causes, such as secondary immunodeficiencies. The cause of CVID is unknown in approximately 90% of cases. An autosomal dominant inheritance pattern is observed in most cases of familiar CVID. Most mutations that are causal to CVID or increase susceptibility to diseases have not been identified[3, 4, 8].

Most cases of CVID are sporadic and occur in people with no apparent family history. Not all individuals who inherit a gene mutation will develop CVID. In many instances, affected children have an unaffected parent with the same mutation. Penetrance may also appear incomplete due to late onset of symptoms. It is likely that both environmental and genetic factors play causal roles. Although the specific environmental factors are unclear, mutations in at least 13 genes have been associated with CVID. However, these known mutations only account for approximately fifteen percent of cases. Since CVID is characterized by defective differentiation of B cells to plasma cells and memory B cells, it is believed that the genetic influences in CVID are mutations in genes that are involved in the development and function of B cells. Mutations in genes that encode BCR-associated complex or are involved in BCR signaling (*CD19*, *CD81*, *CD21*) or are involved in B cell development and plasma cell differentiation (*CD20*) have been associated with CVID. The most frequent mutations occur in the *TNFRSF13B* gene, also known as *TACI*, a receptor for BAFF and APRIL. The protein produced from this gene plays a role in the survival and maturation of B cells and in immunoglobulin production. *TNFRSF13C*, a receptor for BAFF also known as *BAFFR*, is also associated with CVID. Other mutations associated with CVID are in genes involved in the function and maturation of immune cells, such as T cells. It has been observed that, in addition to intrinsic B cell defects, CVID can be caused by deficient T cell co-stimulation. Accordingly, alterations in the function and frequency of T cells have also been documented in patients with CVID. Mutations in the T cell co-stimulatory molecule ICOS as well as the T cell inhibitory molecule CTLA-4 have been identified in CVID. These defects may cause changes in the interaction between B and T cells which may provide potential mechanisms into the various complications seen in CVID[3, 5, 6].

There are different types of CVID that are distinguished by differing genetic causes. Patients with the same type of CVID may have varying symptoms and clinical presentation. The heterogeneity of CVID has led to several attempts to classify CVID into subgroups based on clinical presentation and immunological features. When a molecular diagnosis is made, the disease becomes a specific genetic deficiency and is considered distinct entities separate from CVID. However, it is likely that CVID is a common disease that can result from many various genetic mutations and in the future, these subcategories will be re-categorized as CVID with specific molecular etiology that will not affect classification of the disease but rather give physicians a molecular basis for personalized treatment, much like the way cancer is treated.

CTLA4 mutations were first reported in subjects clinically diagnosed with CVID. Heterozygous *Ctla4* deficiency in mice shows no apparent phenotype. However, heterozygous human *CTLA4* deficiency causes immune dysregulation characterized by recurrent infections, hypogammaglobulinemia, and various autoimmune manifestations, much like symptoms characteristic of CVID. However, instead of falling under the classification of a complex disease with multifactorial causes like most cases of CVID, this phenotype has a monogenic cause and can be treated accordingly[7, 8].

The inhibitory receptor CTLA-4 is an excellent therapeutic target. As a key molecule in the regulation of the immune response, CTLA-4 expression can be therapeutically modulated in order to reduce or increase immune activity. CTLA-4 is a critical component of the effector function of regulatory T (T_{reg}) cells. T_{reg} cells are used by the mammalian immune system as a mechanism of tolerance to control self-reactive T cells. T_{reg} cells are a subpopulation of T cells which control immune homeostasis and tolerance to self-antigens. This population of cells are immunosuppressive and function to downregulate the proliferation of effector T cells. Naturally occurring $CD25^+CD4^+$ Treg cells specifically express Forkhead box P3-positive ($FOXP3^+$) which is a master regulator of T_{reg} cell development and function. Mutations in *FOXP3* resulting in deficiency of T_{reg} cells cause an autoimmune syndrome called IPEX (immune dysregulation polyendocrinopathy X-linked). In mice, homozygous deficiency of *Ctla4* results in a fatal immune dysregulation with clear similarities to FOXP3 deficiency[9-12]. CTLA-4 is constitutively expressed by T_{reg} cells and mediates their suppressive function through the inhibition of naive T cell activation. CTLA-4 is also expressed by T cells upon activation. Upon ligation to ligand, CTLA-4 inhibits proliferation of activated T cells[1]. The modulation of inhibitory immune receptors such as CTLA-4 is changing the way many conditions and diseases are being treated including inflammatory diseases, immunodeficiencies, and cancer.

CTLA-4 was only recently discovered in 1987 when scientists at the Centre d'Immunologie in France isolated cDNA clones from activated $CD8^+$ T cells that defined a sequence encoding a 223 amino acid protein which they called cytotoxic T cell antigen-4 (CTLA-4). Structural features of CTLA-4 identified the protein as a member of the immunoglobulin superfamily. Subsequent to the identification of CTLA-4, it was cloned and mapped to the same chromosomal region in both human and murine genomes as another member of the immunoglobulin superfamily, CD28. CTLA-4 is homologous to the T cell co-stimulatory protein CD28. CTLA-4 was identified as a second receptor for the T cell co-stimulation ligand B7 in 1991. The genetic and molecular similarities between

CTLA-4 and CD28 suggested that the two T cell surface molecules may also share similar functionality. However, the function of CTLA-4 remained largely unknown until 1995 when it was identified as an inhibitor of T cell activation[13-15].

In order to become properly activated, T cells requires two signals. The first signal is provided by binding of the T cell receptor (TCR) to its cognate peptide presented on major histocompatibility complex (MHC) on an antigen-presenting cell (APC). This signal provides specificity to the subsequent immune response based upon the identity of the present antigen. The second signal comes from co-stimulation, in which B7 co-stimulatory molecules on the APC bind to the surface receptor CD28 on the T cell. Naive T cells express the co-stimulatory receptor CD28, so co-stimulation for these cells comes from the CD80 and CD86 proteins, which together constitute the B7 protein, (B7-1 and B7-2, respectively) on the APC. The expression of B7 is induced by various stimuli, such as products of pathogens or cell damage, in order to restrict immune activation only when appropriate. Co-stimulation is requisite to appropriately activate T cell functionality. In the absence of co-stimulation, TCR signaling results in anergy. On re-stimulation, anergic T cells are unable to produce IL-2 or to proliferate, even in the presence of co-stimulatory signals. This is one of the mechanisms the immune system uses to ensure that T cells are only activated in appropriate circumstances. CD28 co-stimulation initiates intracellular signaling that increases cellular metabolism, proliferation, and differentiation[16, 17].

After the immune system is activated, the strength and duration of immune responses must be tightly regulated. The degree of T cell activation is determined by ligation of activating versus inhibitory receptors of the CD28 gene family. Upon activation of the T cell, other receptors are expressed, such as ICOS and PD-1, that provide additional co-stimulatory and inhibitory signaling. T cell activation through the TCR and CD28 also leads to increased expression of CTLA-4. CTLA-4 is constitutively expressed in T_{reg} cells but only upregulated in conventional T cells after activation. CTLA-4 is homologous to CD28 and both molecules bind to B7-1 and B7-2 on APCs. B7-1 and B7-2 are shared ligands of CTLA-4. The expression profiles of B7-1 and B7-2 differ. B7-2 is constitutively expressed in APCs at low levels and rapidly up-regulated upon activation, whereas B7-1 is expressed only upon activation. CTLA-4 binds mostly to B7-1 while CD28 is the preferential receptor for B7-2[17-22]. Upon ligation, CTLA-4 downregulates immune responses and prevents the over activation of T cells by inhibiting T cell proliferation, differentiation, cell cycle progression, and IL-2 production. It has recently been shown that CTLA-4 functions by binding to and physically removing B7 ligands from APCs via a process called transendocytosis.

Through this ligand downregulation, CTLA-4 reduces activation of conventional T cells by reducing the availability of co-stimulatory ligands on APCs, thus suppressing T cell activation[24].

CD28 and CTLA-4 both bind B7 via the same motif. A hexapeptide motif (^{99}MYPPPY^{104}) in the FG loop, conserved between CD28 and CTLA4, is essential for ligand binding. Mutagenesis experiments have shown that this sequence is essential for binding to B7 and alanine substitutions of these residues reduce or abolish binding. Although CD28 and CTLA-4 share B7 as ligands, CTLA-4 binds with greater avidity and affinity thus enabling it to outcompete CD28 for its ligands. CTLA-4 is able to outcompete CD28 for B7 due to the manner in which it binds. CTLA-4 is able to bind B7 bivalently, resulting in a very stable lattice structure of alternating CTLA-4 and B7 dimers. This oligomeric binding pattern displaces CD28 from the synapse due to steric crowding[18-23].

CD28 and CTLA-4 are closely linked on human chromosome 2q33. The structure of *CTLA4* contains four exons, two untranslated regions upstream of exon 1 (5′ UTR) and downstream of exon 4 (3′ UTR), and a promotor region up to −335 bp. Exon 1 encodes the signal peptide sequence; exon 2, an extracellular IgV-like domain that contains the B7 binding domain; exon 3, the transmembrane region, and exon 4, the cytoplasmic tail. Polymorphisms of the *CTLA4* gene region have been associated with increased susceptibility to several autoimmune diseases such as type 1 diabetes, celiac disease, rheumatoid arthritis, and multiple sclerosis[12, 17]. However, further research is required to determine how polymorphisms directly affect CTLA-4 function.

Gene editing technology offers a powerful tool for modeling disease-associated mutations. Bacterial CRISPR/Cas adaptive immune systems have been modified for use in editing of the eukaryotic genome. The CRISPR/Cas system is a microbial immune system used by bacteria and archaea that provides acquired immunity against viruses and plasmids by conferring resistance to foreign genetic elements (Figure 1). Type II CRISPR systems incorporate foreign DNA within the host genome at CRISPR loci and the corresponding CRISPR RNAs (crRNAs) are used to guide targeted degradation of homologous sequences. The recognition of exogenous DNA by a Cas complex prompts the capture of viral or plasmid DNA and insertion of the exogenous DNA into the CRISPR locus in the form of a novel repeat-spacer. Transcripts from the repeat-spacer array are processed into CRISPR RNAs (crRNAs) that contain a sequence transcribed from the exogenous DNA called a protospacer sequence. The part of the crRNA encoded by the protospacer directs Cas to cleave the complementary DNA target sequence if it is adjacent to a short sequence called a protospacer adjacent motif (PAM) that is rarely present in the host genome, the identity of which varies depending on the specific CRISPR system. The PAM sequence is used as a mechanism of

tolerance to prevent cleavage of host DNA and is requisite for binding of the Cas complex to the target sequence. Transactivating CRISPR RNA (tracrRNA) binds to the crRNA and forms an active complex with Cas nuclease. The active Cas complex uses the spacer target sequence to recognize and cut exogenous genetic elements[29, 30].

CRISPR/Cas9 genome editing most commonly uses a Type II CRISPR system from *Streptococcus pyogenes*. This system utilizes Cas9, crRNA, tracrRNA, and an optional repair template (Figure 2). Cas9, crRNA, tracrRNA, and a repair template must be introduced into the target cell. The crRNA and tracrRNA can be joined together to form a single-guide RNA (sgRNA). The sgRNA consists of the guide or target sequence and a scaffold. The protospacer portion of the crRNA contains the target sequence that Cas9 uses to identify and directly bind to the DNA of the cell. The target sequence is a 20-nt guide sequence that provides the specificity of the Cas9 nuclease. When the Cas9 protein is expressed, it will form a complex with the sgRNA by binding to the scaffold domain of the sgRNA. Cas9 is able to select the correct location on the host DNA by using the guide sequence to bond with base pairs on the host DNA. For the *S. pyogenes* system, the target sequence must immediately precede a 5′-NGG PAM sequence. The 20-nt guide sequence base pairs with the opposite DNA strand to mediate Cas9 cleavage 3 bp upstream of the PAM sequence. Cas9 has two endonuclease domains, RuvC and HNH, which allows cleavage of both strands of the target DNA resulting in a double-stranded break (DSB). DSBs are typically repaired by non-homologous end-joining (NHEJ), which results in nonspecific insertions, deletions or other mutations. This is a commonly used approach to engineer cell lines or animal models with specific gene knockout mutations. Specific sequences can be inserted or replaced by using a DNA repair template to direct repair of DSBs by homology-directed repair (HDR). A repair template introduced into the cell can be used as an alternative template in order to produce specific nucleotide changes. The cell can use the repair template containing the chosen edit to repair the break, thus incorporating the mutation of interest into the genome of the cell. Although CRISPR/Cas9 is widely used to engineer gene knock-outs through NHEJ, the inefficiency of HDR prevents its widespread use for modeling genetic disorders through the introduction of disease-associated mutations[30-35].

1.2. Schematic figures

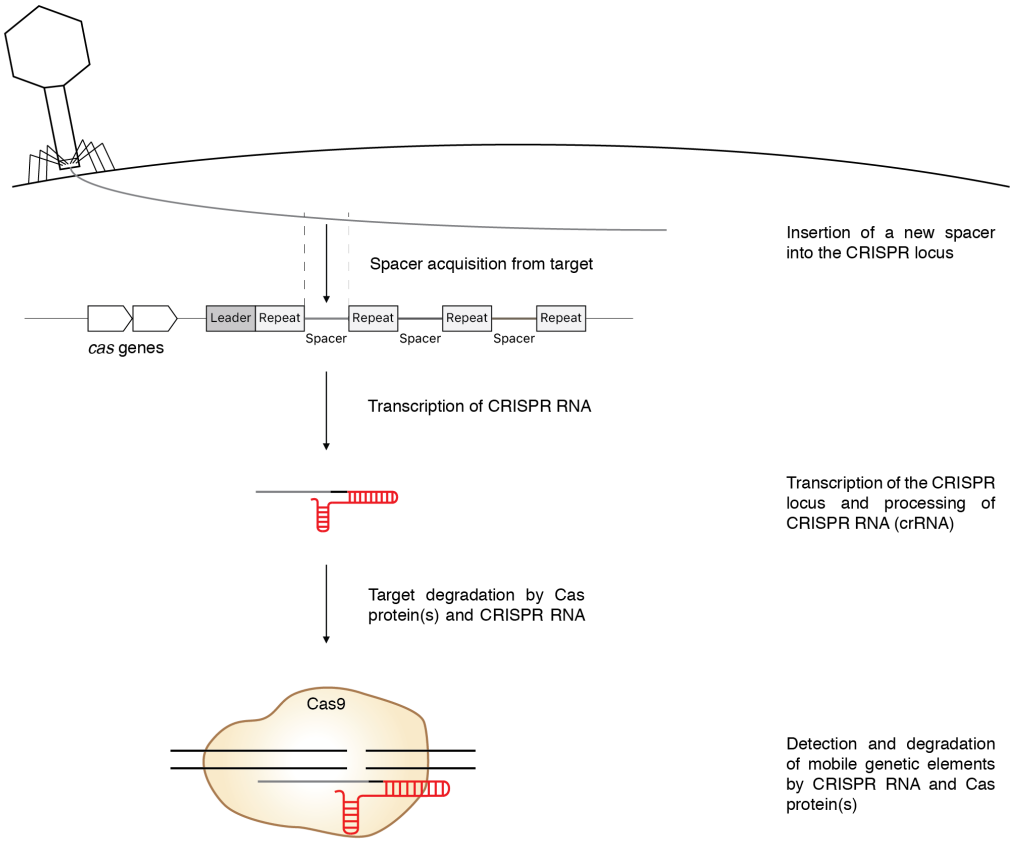

Figure 1 CRISPR in bacteria and archaea

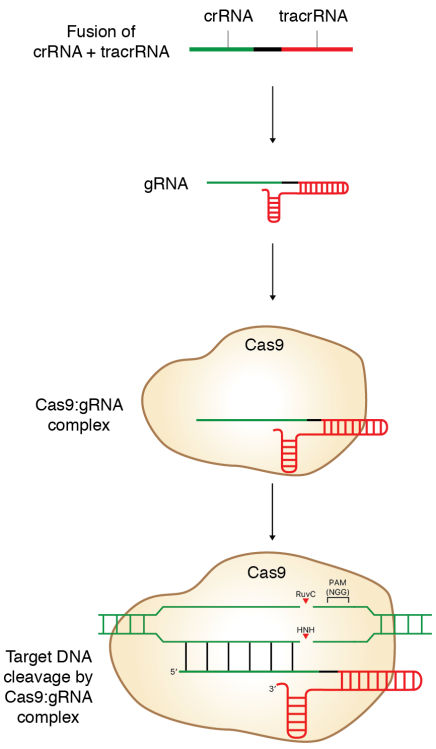

Figure 2 CRISPR in genome engineering

2. Chapter 2: Data and Methods

2.1. Short Introduction

Using a combined whole exome sequencing (WES) and targeted screening approach, we identified a novel heterozygous, *de novo*, missense c.406C>G, p.P136A mutation in *CTLA4* in a subject with CVID. Proline 136 is within the binding motif of CTLA-4. A missense mutation within this ligand-binding sequence is predicted to result in CTLA-4 deficiency due to an inability of the protein from the mutated allele to bind to B7. The protein product of the allele carrying the mutation likely acts antagonistically to the wild-type allele. Since CTLA-4 functions as a homodimer, the protein carrying the p.P136A mutation may dimerize with wild-type proteins forming a dimer that is unable to oligomerize due to the inability of the mutant polypeptide to bind ligand. It is predicted that this mutation acts in a dominant-negative manner, reducing the activity of the CTLA-4/B7 complex. It is unknown whether the mutant protein is expressed at a normal level. If overexpressed, the proportion of CTLA-4 dimers containing only normal subunits will decrease.

The clinical presentation of the patient was typical of CVID with autoimmune manifestations. Patient history showed low IgG and IgA, low switched memory B cells, recurrent infections of the respiratory tract, infiltrative granulomatous disease, ITP, and AIHA. To determine whether the *CTLA4*^{P136A} mutation is causal in the phenotype of the patient, we used CRISPR/Cas9-mediated genome editing to clone the mutation into a regulatory T cell-like cell line for use in functional studies.

2.2. Materials and Methods

MLM3636 gRNA plasmid construction. The target sequence for the gRNA was selected based upon proximity the intended Cas9 cleavage site with the lowest number of off-target sites. gRNA efficiency and off-target sites were analyzed using a CRISPR design tool (http://crispr.mit.edu). gRNA sequences targeting *CTLA4* were cloned into plasmid MLM3636. MLM3636 was a gift from K. Joung (Addgene plasmid number 43860). Two DNA oligonucleotides with vector-specific overhangs were ordered and then annealed. The top (FORWARD) and bottom (REVERSE) phosphorylated oligonucleotides were suspended to 100 μM in 0.1x TE buffer and diluted 1:10 to 10 μM in H2O. The phosphorylated oligonucleotides were annealed in the reaction buffer detailed below.

Top strand oligo FOR (10 μM)	1 μl
Bottom strand oligo REV (10 μM)	1 μl

10x T4 DNA Ligase buffer	2 µl
dH_2O (nuclease free)	16 µl
	20 µl total volume

The annealing reaction was heated to 95°C for 5 mins and then cooled to 10°C. The double-stranded DNA product was ligated into the BsmBI sites of the MLM3636 plasmid backbone. The ligation reaction is detailed below.

Plasmid backbone (cut MLM3636 vector)	1 µl
Annealed primers	1 µl
10x buffer (T4 DNA ligase buffer)	1 µl
T4 DNA ligase	0.5 µl
dH_2O (nuclease free)	6.5 µl
	10 µl total volume

The ligation reaction was incubated at 16°C overnight. The ligation product was transformed into *E. coli*. Bacterial cells were thawed on ice. 1 µl of DNA in H_20 and 25 µl of cells were combined and left on ice for 15 mins and then heat shocked at 42°C for 45 s. The reaction volume was then placed back on ice for 2 mins. 250 µl of LB medium was added to the reaction volume and was incubated at 37°C for 1 h. 100 µl of the reaction volume was plated on each of four LB plates with antibiotics and left to grow at 37°C overnight.

The plates were inspected for *E. coli* growth and eight colonies were selected. Each colony was inoculated into a loose-capped culture tube containing 5 ml LB medium plus antibiotic and incubated with 250 rpm shaking at 37°C overnight. After 12-16 h of growth, the bacterial cells were harvested by centrifugation at 6800 x g for 20 mins. Glycerol stocks of the eight colonies were prepared and stored at -80°C.

The plasmid DNA was isolated from the bacteria via miniprep. The pelleted bacterial cells were resuspended in 250 µl 2–8°C Buffer P1 with RNase A solution and transferred to a 1.5 ml tube. 250 µl Buffer P2 was added to the bacterial cell solution and the tube was inverted 5 times to homogenize the solution. 350 µl Buffer N3 was added to the solution and the tube was inverted 5 times to homogenize the solution. The solution was centrifuged in a microcentrifuge at 13,000 rpm for 10 mins. The supernatant was decanted into a spin column. The spin column was centrifuged at 13,000 rpm for 30 s and the flow-through was discarded. The spin column was washed with 0.5 ml Buffer PB. The spin column was centrifuged at 13,000 rpm for 30 s and the flow-through was discarded. The spin column was washed with 0.75 ml Buffer PE. The spin column was centrifuged at 13,000 rpm for 30 s and the flow-through was discarded. The spin column was centrifuged at 13,000 rpm for 1 min to remove wash buffer. Plasmid DNA was eluted from the spin column into a 1.5 ml tube by the addition of 50 µl Buffer EB (10 mM TrisCl, pH 8.5)

to the spin column. After 1 min, the spin column and 1.5 ml tube was centrifuged at 13,000 rpm for 1 min. DNA concentration was determined by UV spectrophotometry at 260 nm. The sequences of the resulting clones were verified via Sanger sequencing.

Bacteria from sample 1 and sample 2 were plated from glycerol stock on agar plates. The plates were incubated at 37°C overnight. Each colony was inoculated in 1–5 ml LB medium with antibiotic to form a starter culture and incubated 4–6 hours at 37°C. Each starter culture was diluted into 150 ml LB medium and incubated for 12–16 h at 37°C at 300 rpm until cell density reached approximately 3–4 x 10^9 cells per ml. The bacterial cells were pelleted by centrifugation at 6800 x g for 20 mins. Plasmid DNA was purified from the bacteria via maxiprep. The pelleted bacterial cells were resuspended in 10 ml 2–8°C Buffer P1 with RNase A solution. 10 ml Buffer P2 was added to the bacterial cell solution and the tube was inverted 5 times to homogenize the solution. The lysis reaction solution was incubated at 25°C for 5 mins. 10 ml of pre-chilled Buffer P3 was added to the lysate and the tube was inverted 5 times to homogenize the solution. The lysate was transferred into a filter cartridge and the solution was incubated at 25°C for 10 mins. The HiSpeed maxi tip was equilibrated with Buffer QBT. 10 ml Buffer QBT was added to the maxi tip and allowed to drain. After the precipitate of protein, genomic DNA, and detergent floated to the surface of the lysate solution, a plunger was used to push the cell lysate through the filter cartridge into the equilibrated tip. The tip was washed with 60 ml Buffer QC. The DNA was eluted from the tip filter with 15 ml Buffer QF. The DNA was precipitated by adding 10.5 ml 25°C isopropanol to the eluted DNA. The solution was incubated at 25°C for 5 mins. The eluate/isopropanol solution was transferred into a 30 ml syringe with attached QIAprecipitator. The eluate/isopropanol solution was filtered through the QIAprecipitator using a plunger. The QIAprecipitator was removed and the plunger was removed from the syringe. The The QIAprecipitator was reattached to the syringe. The DNA was washed by pressing 2 ml 70% ethanol through the QIAprecipitator using the plunger of the syringe. The QIAprecipitator was removed and the plunger was removed from the syringe. The QIAprecipitator was reattached to the syringe. The filter membrane of the QIAprecipitator was dried by using the plunger to press air through the QIAprecipitator. This step was repeated. The QIAprecipitator was then attached to a 5 ml syringe. The DNA was eluted from the filter membrane into a 1.5 ml tube using a plunger to press 1 ml Buffer TE through the syringe. The QIAprecipitator was removed from the syringe, the plunger was removed, and the QIAprecipitator was reattached to the syringe. The eluate was transferred from the 1.5 ml tube into the 5 ml syringe. The DNA was eluted again. DNA concentration was determined by UV spectrophotometry at 260 nm.

HDR template construction. The 101-nt ssODN repair template was designed with homologous genomic flanking sequence centered around the Cas9 cleavage site as described in Figure 6.

pSpCas9n(BB)-2A-GFP gRNA/Cas9 plasmid construction. The target sequence was the same guide sequence selected for MLM3636. gRNA sequences targeting *CTLA4* were cloned into plasmid pSpCas9n(BB)-2A-GFP. pSpCas9n(BB)-2A-GFP (PX461) was a gift from Feng Zhang (Addgene plasmid # 48140). Two DNA oligonucleotides with vector-specific overhangs were ordered and then annealed. The top (FORWARD) and bottom (REVERSE) oligonucleotides were suspended to 100 µM in 0.1x TE buffer and diluted 1:10 to 10 µM in H2O. The oligonucleotides were phosphorylated and annealed in the reaction buffer detailed below.

Top strand oligo FOR (10 µM)	1 µl
Bottom strand oligo REV (10 µM)	1 µl
10x T4 DNA Ligase buffer	1 µl
T4 Polynucleotide Kinase	1 µl
dH_2O (nuclease free)	6 µl
	10 µl total volume

The reaction was incubated at 37°C overnight to allow phosphorylation of the oligonucleotides to allow subsequent ligation. The phosphorylated reaction mixture was then heated to 95°C for 5 mins and then cooled to 10°C. The pSpCas9n(BB)-2A-GFP vector was digested in the reaction volume detailed below.

pSpCas9n(BB)-2A-GFP plasmid	3 µl
10x buffer (NEBuffer 2.1)	3 µl
BbsI (restriction enzyme)	2 µl
dH_2O (nuclease free)	22 µl
	30 µl total volume

The reaction was incubated at 37°C overnight to allow digestion of the vector backbone. The resulting concentration of the vector was 75 ng/µl. The vector backbone was dephosphorylated according to the reaction outlined below.

Plasmid backbone (cut pSpCas9n(BB)-2A-GFP vector)	30 µl
Antarctic Phosphatase (AnP)	3 µl
10x buffer (AnP buffer)	4 µl
dH_2O (nuclease free)	3 µl
	40 µl total volume

The reaction was incubated at 37°C for 2.5 h. The reaction mixture was then heated to 80°C for 30 mins to heat inactivate BbsI and AnP. The annealed oligo duplex was ligated into the BbsI sites of the pSpCas9n(BB)-2A-GFP plasmid backbone. The ligation reaction is detailed below.

Plasmid backbone (cut pSpCas9n(BB)-2A-GFP vector)	0.5 µl
Annealed primers	1 µl
10x buffer (T4 DNA ligase buffer)	1 µl
T4 DNA ligase	0.5 µl
dH_2O (nuclease free)	7 µl
	10 µl total volume

The ligation reaction was incubated at 16°C overnight. The ligation product was transformed into *E. coli*. Bacterial cells were thawed on ice. 10 µl of DNA in H_2O and 100 µl of cells were combined and left on ice for 30 mins and then heat shocked at 42°C for 45 s. The reaction volume was then placed back on ice for 2 mins. 250 µl of LB medium was added to the reaction volume and was incubated at 37°C for 1 h. 150 µl, 100 µl, 60 µl, and 50 µl of the reaction volume was plated on each of four LB plates with antibiotics and left to grow at 37°C overnight.

The plates were inspected for *E. coli* growth and 4 colonies were selected. Each colony was inoculated into a loose-capped culture tube containing 5 ml LB medium plus antibiotic and incubated with 250 rpm shaking at 37°C overnight. After 12–16 h of growth, the bacterial cells were pelleted by centrifugation at 6800 x g for 20 mins.

The plasmid DNA was isolated from the bacteria via miniprep. The pelleted bacterial cells were resuspended in 250 µl 2–8°C Buffer P1 with RNase A solution and transferred to a 1.5 ml tube. 250 µl Buffer P2 was added to the bacterial cell solution and the tube was inverted 5 times to homogenize the solution. 350 µl Buffer N3 was added to the solution and the tube was inverted 5 times to homogenize the solution. The solution was centrifuged in a microcentrifuge at 13,000 rpm for 10 mins. The supernatant was decanted into a spin column. The spin column was centrifuged at 13,000 rpm for 30 s and the flow-through was discarded. The spin column was washed with 0.5 ml Buffer PB. The spin column was centrifuged at 13,000 rpm for 30 s and the flow-through was discarded. The spin column was washed with 0.75 ml Buffer PE. The spin column was centrifuged at 13,000 rpm for 30 s and the flow-through was discarded. The spin column was centrifuged at 13,000 rpm for 1 min to remove wash buffer. Plasmid DNA was eluted from the spin column into a 1.5 ml tube by the addition of 50 µl Buffer EB (10 mM TrisCl, pH 8.5) to the spin column. After 1 min, the spin column and 1.5 ml tube was centrifuged at 13,000 rpm for 1 min. DNA concentration was determined by UV spectrophotometry at 260 nm. The sequences of the resulting clones were verified via Sanger sequencing.

Sanger sequencing. 10 µl of plasmid DNA at a concentration of 200 ng/µl and 10 µl of sequencing primer at a concentration of 1.5–3.0 µM were added to a 1.5 ml tube for submission. Human U6 promoter, forward primer was

used for sequencing.

MLM3636/Cas9D10A-GFP nucleofection. 10^6 MT-2 cells were nucleofected with 0.4 µg gRNA (MLM3636) and 1.6 µg Cas9 (Cas9D10A-GFP) in a 1:1 molar ratio with 1 µl ssODN template per 500 ng of plasmid DNA. Cells used for nucleofection have >95% viability. 10^6 MT-2 cells were put into a 15 ml conical and washed once with PBS with the supernatant aspirated out. The cells were resuspended in 100 µl of supplemented SE Nucleofector solution (2.25 ml SE Nucleofector solution + 0.5 ml supplement). 2 µg of plasmid DNA and 1 µl of the ssODN template was added to the 100 µl of cells in SE solution. 100 µl of the cell and DNA mixture was transferred into a Nucleocuvette vessel and the vessel was placed in the 4-D Nucleofector X unit. The Jurkat CL-120 program was selected and run. After the run (~1 second), the Nucleocuvette vessel was removed and the contents were transferred in 2 ml of 37°C RPMI 1640 + 2 mM Glutamine + 10% Fetal Bovine Serum (FBS) + 5 ml P/S.

Live/dead viability staining. Cells were centrifuged at 1600 rpm for 5 mins and resuspended in 1 ml PBS. 1 µl of live/dead viability dye was added. Cells were incubated with dye for 30 mins and centrifuged at 1600 rpm for 5 mins. Cells were washed with 1 ml PBS and centrifuged at 1600 rpm for 5 mins. Cells were resuspended in 200 µl PBS.

Fluorescence-activated cell sorting. GFP^+ cells were isolated and collected by FACS 48 h post-transfection and resuspended in 50% conditioned RPMI 1640 + 2 mM Glutamine + 20% Fetal Bovine Serum (FBS) + 5 ml P/S.

Cell culture. MT-2 lymphocyte cell cultures were maintained at a cell concentration between 5 X 10^4 and 4 X 10^5 cells/ml in RPMI 1640 + 2 mM Glutamine + 10% Fetal Bovine Serum (FBS) + 5 ml Penicillin Streptomycin (P/S) at 37°C 5% CO_2.

Cell cloning by limiting dilution. Cells were homogenized by passing several times through a serological pipet. A dilute cell solution (<10^6 cells/ml) was obtained before counting the cells in order to increase accuracy of the number of cells present. The cell concentration in the homogenized cell solution was quantitated via cell counter. Cells from the homogenized cell solution were transferred into the conditioned medium to make a new cell solution at concentrations of 5 cells/ml, 10 cells/ml, and 50 cells/ml. 10 ml of the new cell solution at each concentration was prepared per 96-well plate. 100 µl of the new cell solution was pipetted into each well of a 96-well plate. The 96-well plates were placed in a 37°C 5% CO_2 incubator for 7 days. The growth was observed 7 days and 10 days after plating.

Restriction digest. Cas9D10A-GFP was digested using three enzymes as detailed below.

Buffer	Enzyme	Cuts	Temperature
NEB Cutsmart	XhoI	5289-5288 (9271 bp) (Linearize)	37°C
NEBuffer 3.1	BglII	1996-2932 (937 bp), 2933-1995 (8334 bp)	37°C
NEB Cutsmart	EcoRI-HF	1728-5567 (3840 bp), 5568-1727 (5431 bp)	37°C

For each enzyme digest reaction, the enzyme and its corresponding buffer were added to 1 µg of Cas9D10A-GFP and water was added to make a 50 µl reaction mixture. Below is the reaction component and corresponding volume for a generic digestion 50 µl total volume reaction.

DNA (1 µg)	2 µl
Buffer	5 µl
Enzyme	1 µl
dH_2O (nuclease free)	42 µl
	50 µl total volume

The digestion was run at 37°C for 1 h. The resultant digest samples were run on an agarose gel.

Gel electrophoresis. 1.25 g of agarose was measured and mixed with 100 ml 1x TAE in a flask. The mixture was microwaved for 1–3 minutes until the agarose was completely dissolved. The agarose solution was cooled to 20–25°C. SYBR Safe gel stain was added to the solution. The 1.25% agarose solution was poured into a gel tray and allowed to solidify. The agarose gel was placed into the electrophoresis unit and the unit was filled with 1x TAE. The molecular weight ladder was loaded into the first lane of the gel. The digest samples were loaded into subsequent lanes of the gel. The gel was run at 80 V for 1.5 h. The gel was visualized under UV light.

Sleeping Beauty transposon system. High fidelity Cas9 was cloned into the SfiI cloning sites in an empty SB transposon vector with constitutive bi-directional promoter (pSbbi-GP). pSBbi-GP was a gift from Eric Kowarz (Addgene plasmid # 60511). 10^6 MT-2 cells were nucleofected with 10.0 µg SB transposon plasmid (pSBbi-GP) and 5.00 µg SB transposase (SB100X). Cells used for nucleofection had >90% viability. 10^6 MT-2 cells were put into a 15 ml conical and washed once with PBS with the supernatant aspirated out. The cells were resuspended in 100 µl of supplemented SE Nucleofector solution (2.25 ml SE Nucleofector solution + 0.5 ml supplement). Transposon and transposase plasmid DNA was added to the 100 µl of cells in SE solution. 100 µl of the cell and DNA mixture was transferred into a Nucleocuvette vessel and the vessel was placed in the 4-D Nucleofector X unit. The Jurkat CL-120 program was selected and run. After the run (~1 second), the Nucleocuvette vessel was removed and the contents

were transferred in 2 ml of 37°C RPMI 1640 + 2 mM Glutamine + 10% Fetal Bovine Serum (FBS) + 5 ml P/S.

2.3. Results

The goal of this project was to use CRISPR/Cas9-mediated genome editing to clone the novel heterozygous *CTLA4* mutation identified in our CVID patient into a T_{reg}-like cell line to confirm this mutation as causal in the autoimmune phenotype of the patient. There are many different methods to deliver and express Cas9 and the gRNA in target cells that differ according to the goals of the experiment and ease of transfection of the target cells[31-33]. However, the most common method uses transfection of the target cells with plasmids containing Cas9 and the gRNA, as used by us in our workflow (Figure 3). Plasmids with Cas9 and the gRNA were constructed and nucleofected into MT-2 cells. MT-2 cells are T_{reg}-like cell line that has the phenotypic and functional characteristics of human T_{reg} cells[36]. This cell line was used as a model due to the fact that it constitutively expresses CTLA-4. Since our aim was to model a specific point mutation as seen in our patient, we used a repair template in addition to the gRNA and Cas9 in order to incorporate the mutation of interest into the cell through HDR. The patient mutation is heterozygous and therefore the edited cell must only incorporate the mutation from the repair template into one allele.

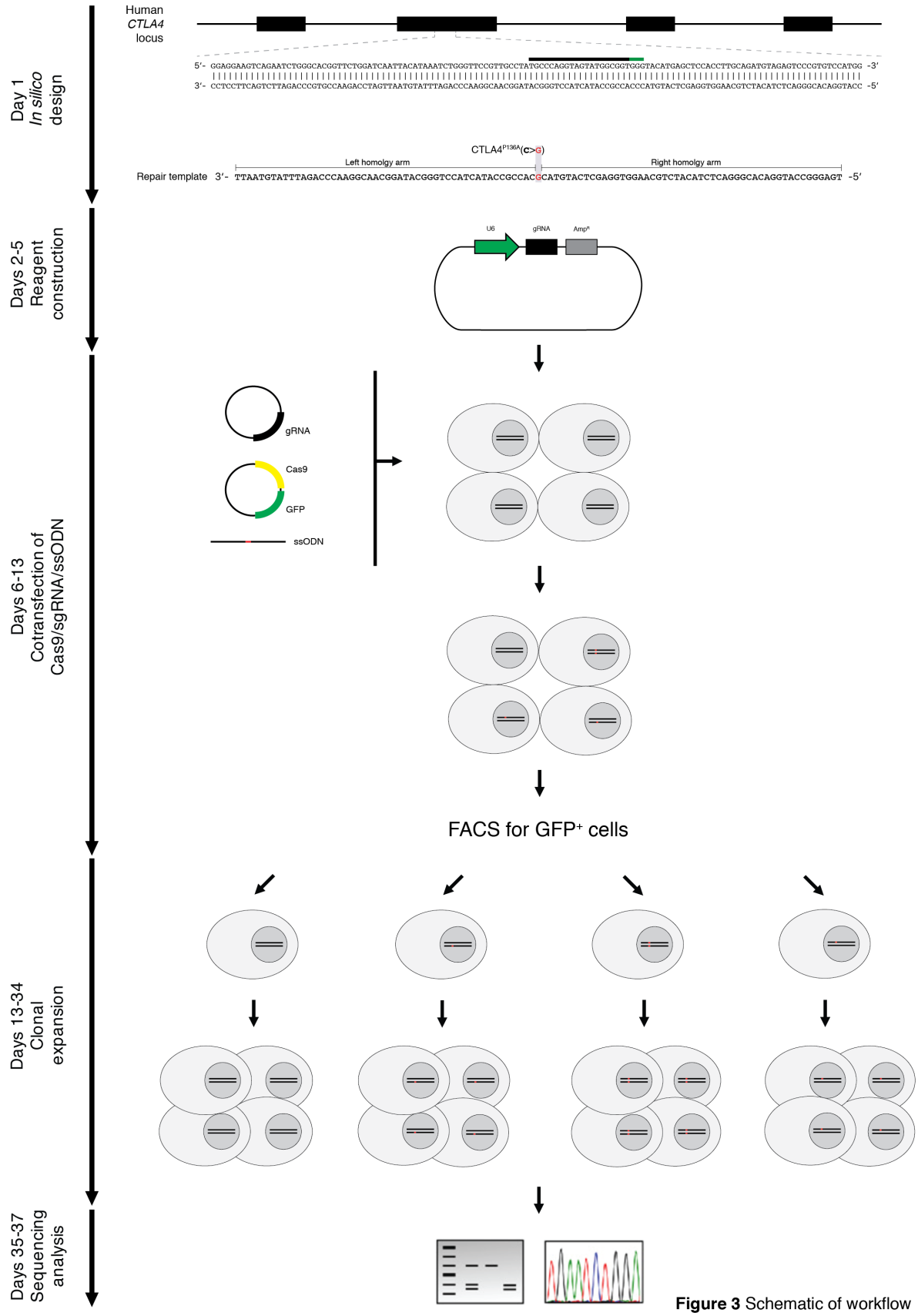

Figure 3 Schematic of workflow

To increase the frequency of HDR, we used a Cas9 with only one active catalytic domain (Cas9 nickase) that is only able to cut one strand, resulting in a nick rather than a DSB. Wild-type Cas9 has two nuclease domains, HNH and RuvC, which cleave DNA by nicking the gRNA-complementary and noncomplementary strands, respectively. We used a Cas9 with a D10A mutation in the RuvC domain (Figure 4). This Cas9 from *Streptococcus pyogenes* requires a 5′-NGG PAM for target cleavage. We used a plasmid for Cas9 expression that contained Cas9 (D10A) nickase and GFP. The Cas9D10A-GFP plasmid was prepared as a maxiprep using a QIAprep Miniprep kit.

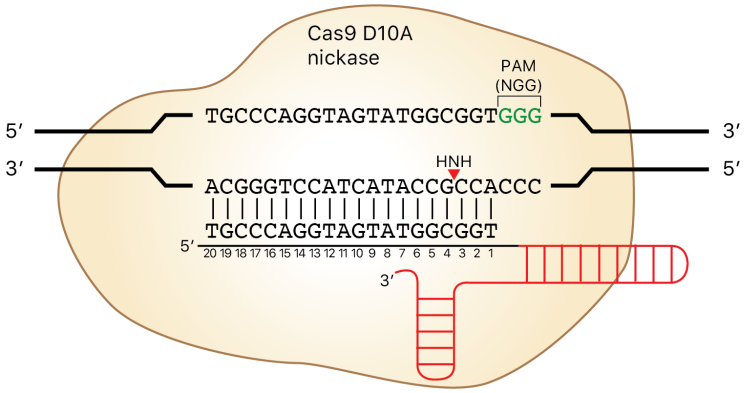

Figure 4 An aspartate-to-alanine (D10A) mutation in the RuvC catalytic domain of Cas9 results in a Cas9 with only one active catalytic domain (Cas9 nickases) that are only able to cut one DNA strands, resulting in a single-stranded break (nick) which are preferentially repaired by HDR. Cas9D10A nickases cleave only the DNA strand that is complementary to and thus recognized by the gRNA.

To introduce the monoallelic *CTLA4*^{P136A} mutation into target cells, we selected a gRNA at the human *CTLA4* locus whose cleavage site is the least distance to the intended mutation with the lowest amount of off-target sites (Figure 5). Although Cas9 generally cleaves its target reliably, off-target cleavage events can occur. Cas9 nuclease activity is guided by the gRNA and requires a PAM motif for cleavage of the target. Therefore, as long as the Cas9 is able to recognize a guide sequence that is immediately upstream of a 5′-NGG PAM sequence, it will cleave the DNA. Thus, if the guide sequence of the gRNA is nonspecific and is frequently present at sites outside of the intended target site, off-site targeting and cleavage will occur[31]. To minimize off-target activity, an online guide sequence design tool that computationally determines off-target cleavage activity was used to select a guide sequence. The program estimates off-target activity of Cas9 at potential off-target sites by evaluating the frequency of sequences within the genome with homology to the guide sequence or off-target sequences that vary by only a

few nucleotides as compared to the intended target.

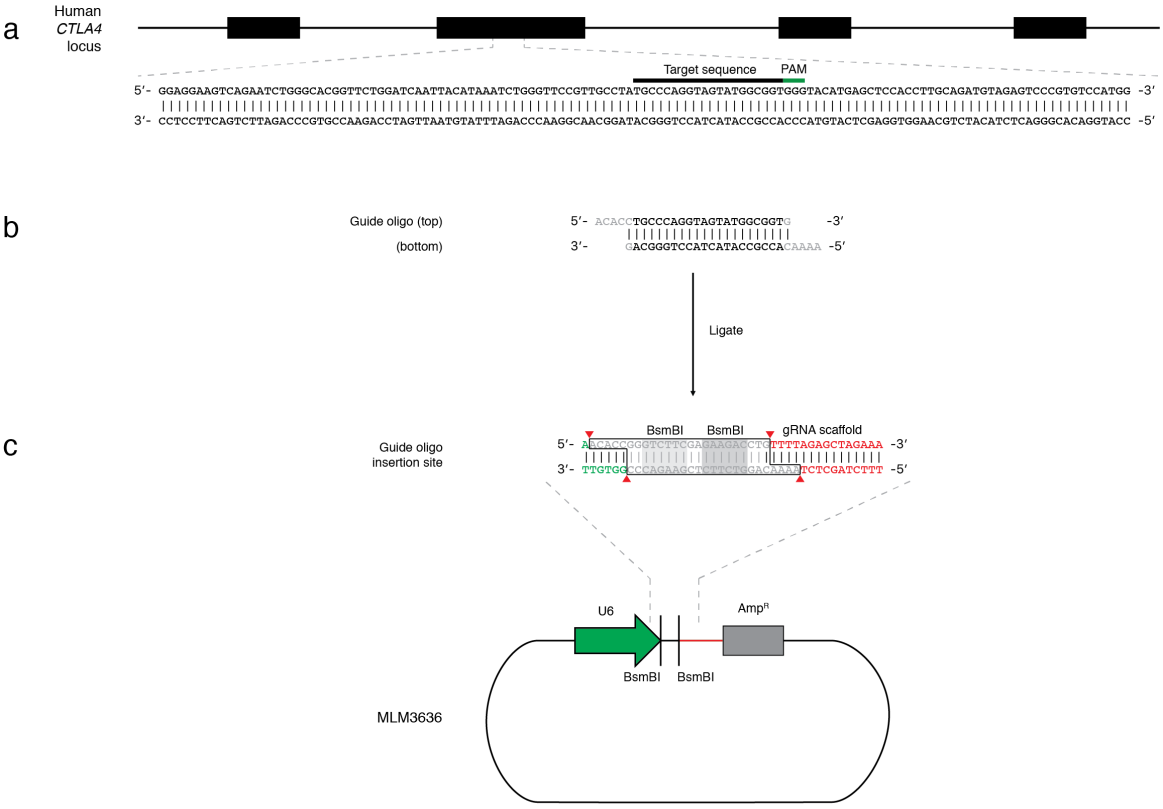

Figure 5 Targeting strategy and gRNA plasmid construction (**a**) The 20 bp target sequence must be immediately 5' of a 5'-NGG PAM sequence. *CTLA4* sequencing alignment showing target sequence selection. (**b**) The top strand oligo is the target sequence and the bottom strand oligo is the reverse complement of the target sequence. The top and bottom strand oligos have the same orientation as the sequences in the genomic target. The guide sequence oligos have overhangs specific for ligation into the pair of BsmBI sites in MLM3636 (gray). (**c**) The MLM3636 plasmid was digested with BsmBI, which enables the replacement of the Type II restriction sites (black outline) with direct insertion of annealed oligos into the guide oligo insertion site of the vector.

A single-stranded oligonucleotide (ssODN) construct containing the intended mutation was used as a repair template. For high HDR efficiency, ssODNs should contain sequences that flank the intended edit of at least 40 bp that are homologous to the sequence of the target site, called homology arms. The ssODN sequence can be homologous to the sense or antisense sequence of the target locus. We designed a 101-nt ssODN with 50 bp left and right homology arms on either side of the intended C>G mutation (Figure 6a). The intended nucleotide in the ssODN template is 4 bp away from the Cas9 cleavage site. The intended mutation is within the PAM of the gRNA sequence and is expected to prevent re-cutting of Cas9 upon successful repair via the ssODN template (Figure 6b).

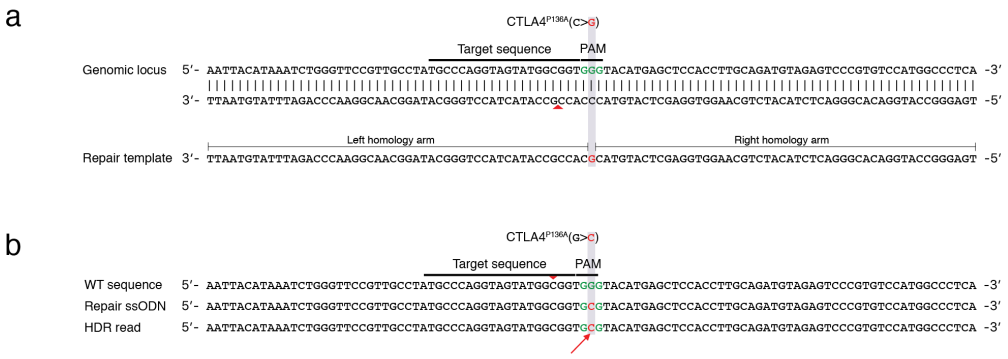

Figure 6 ssODN HDR template design using CRISPR/Cas9-blocking mutations increase HDR accuracy by preventing re-editing by Cas9 (**a**) *CTLA4* sequencing alignment showing ssODN sequence with c.406C>G mutation (red). (**b**) *CTLA4* sequence modification using repair ssODN resulting in a CRISPR/Cas-blocking mutation of the PAM sequence. *CTLA4* sequencing alignment showing the introduction of a CRISPR/-Cas-blocking mutation by including a blocking mutation in the PAM sequence (red) using the antisense sequence for illustrative purposes. Although the pathogenic c.406C>G mutation is in the sense strand and the sequence used in the repair ssODN is sense, a sense or antisense repair ssODN can be used for HDR. The introduced mutation in the antisense strand would be a G>C base substitution, resulting in a blocking mutation (red arrow) in the PAM. This blocking mutation prevents re-editing by Cas9 by mutating the NGG sequence required by Cas9 for targeting.

We first aimed to determine the efficiency of nucleofection by nucleofecting the Cas9D10A-GFP plasmid into MT-2 cells and Jurkat cells. MT-2 cells are a human T-cell leukemia virus type 1 (HTLV-1) infected cell line that has the phenotypic and functional characteristics of human T_{reg} cells. A T_{reg}-like cell line was used to study the effect of the identified *CTLA4* mutation given the role of CTLA-4 in T_{reg} cell function. This cell line was used as a model due to the difficulty of identifying and isolating $CD4^{+}FoxP3^{+}$ Treg cells. Jurkat cells were used as a control in comparison of nucleofection efficiency for MT-2s based on their well-documented ability to undergo successful nucleofection and their similarity in size and composition to MT-2s. 10^6 MT-2s and 10^6 Jurkat cells were nucleofected with pmaxGFP to determine nucleofection efficiency. 10^6 MT-2s and 10^6 Jurkat cells were nucleofected with the Cas9D10A-GFP plasmid. Control pmaxGFP plasmid-transfected MT-2s were used as a reference for GFP expression. The efficiency of the Cas9D10A-GFP plasmid was determined by comparing live versus dead cells and GFP expression of the positive control pmaxGFP plasmid versus the Cas9D10A-GFP plasmid.

48 hours post-nucleofection, flow cytometry was used to evaluate GFP expression and percentage of live versus dead cells reported by staining of the cells with live/dead viability dye. Results showed very low levels of live cells and GFP expression after nucleofection with the Cas9D10A-GFP plasmid. Previous Cas9 plasmid preparation via maxiprep yielded low concentrations of plasmid DNA. The low quality Cas9 plasmid may have lowered the efficiency of nucleofection since the quality and the concentration of DNA used for nucleofection is a

critical factor in the efficiency of gene transfer. The Cas9D10A-GFP plasmid was purified via maxiprep and the DNA concentration was determined to be 495.5 ng/µl by UV spectrophotometry at 260 nm. The identity of the plasmid was confirmed by gel electrophoresis.

Jurkat cells and MT-2 cells were nucleofected with the Cas9D10A-GFP plasmid after maxiprep preparation. 30 hours post-nucleofection, the percentage of live versus dead cells and GFP expression was measured by flow cytometry. Nucleofected cells showed 70% viability (Figure 7). Jurkat cells had a higher level of GFP expression at 17% positive cells, whereas only 2% of MT-2s were positive for GFP expression. MT-2s showed a nucleofection efficiency of 65% with the pmaxGFP control plasmid. A nucleofection efficiency rate of 65% is consistent with expected efficiency rate as determined by optimization experiments. This suggests that MT-2s are more resistant to nucleofection with larger plasmids than Jurkat cells. These results also demonstrate that the low efficiency and high toxicity of the previous nucleofection was due to low quality plasmid DNA purification of the Cas9D10A-GFP plasmid which was rectified upon re-purification of the plasmid.

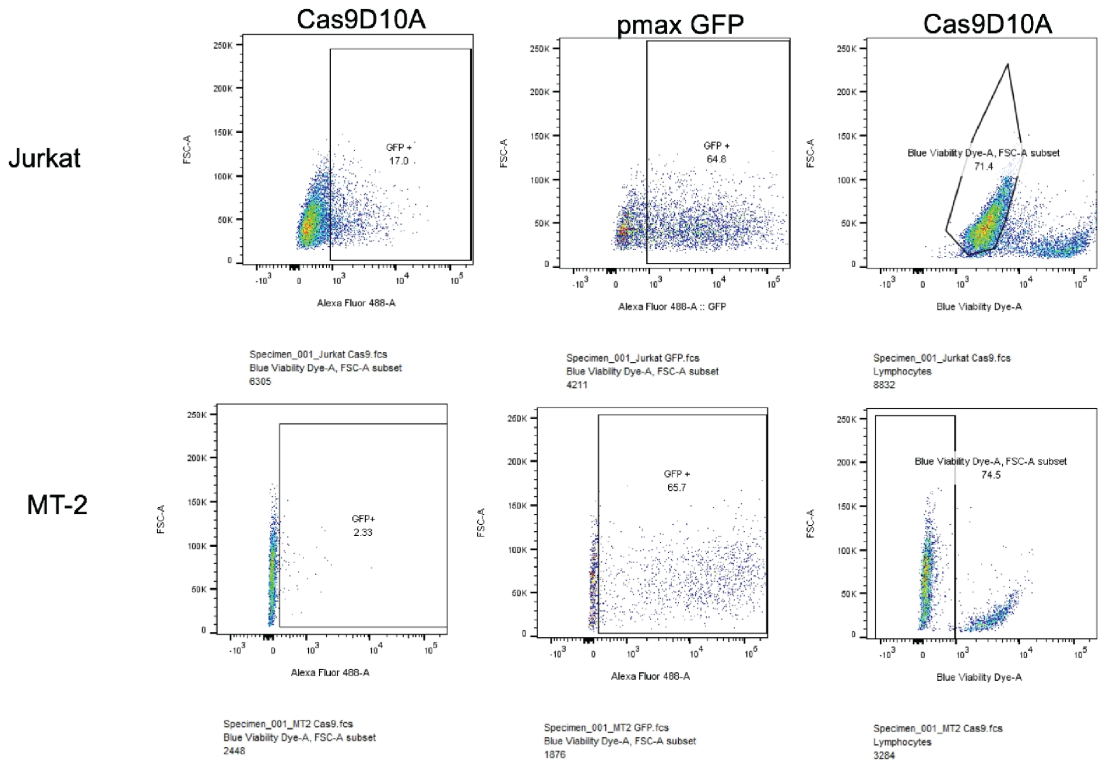

Figure 7 Nucleofection efficiency controls

The gRNA (MLM3636) and Cas9 (Cas9D10A-GFP) plasmids and ssODN were nucleofected into MT-2s (Figure 8). The gRNA and Cas9 plasmids were nucleofected at a 1:1 molar ratio at a total of 2 µg of total DNA. The

ssODN template was nucleofected at an amount of 1 µl/500 ng. In order to obtain single-cell colonies, GFP^+ cells were single-cell sorted via flow cytometry into three 96-well plates. Growth was monitored every seven days for 21 days. A total of 288 cells were plated among which 0 colonies were recovered. These results suggest that MT-2s may have not grown due to difficulty of the cell line to grow in single cell conditions.

The gRNA (MLM3636) and Cas9 (Cas9D10A-GFP) plasmids and ssODN were nucleofected into MT-2 cells. 9000 GFP^+ live cells were bulk sorted via flow cytometry into 2 ml of conditioned RPMI medium (Figure 8). 24 hours post-sort, GFP^+ MT-2s were viewed by microscopy. The cells were not visible under the microscope. Of the 9000 cells that were positive for GFP, approximately 100 cells were visible. This result suggests an error during the sorting process or a low tolerance of the cells to sorting process resulting in the loss of viable cells.

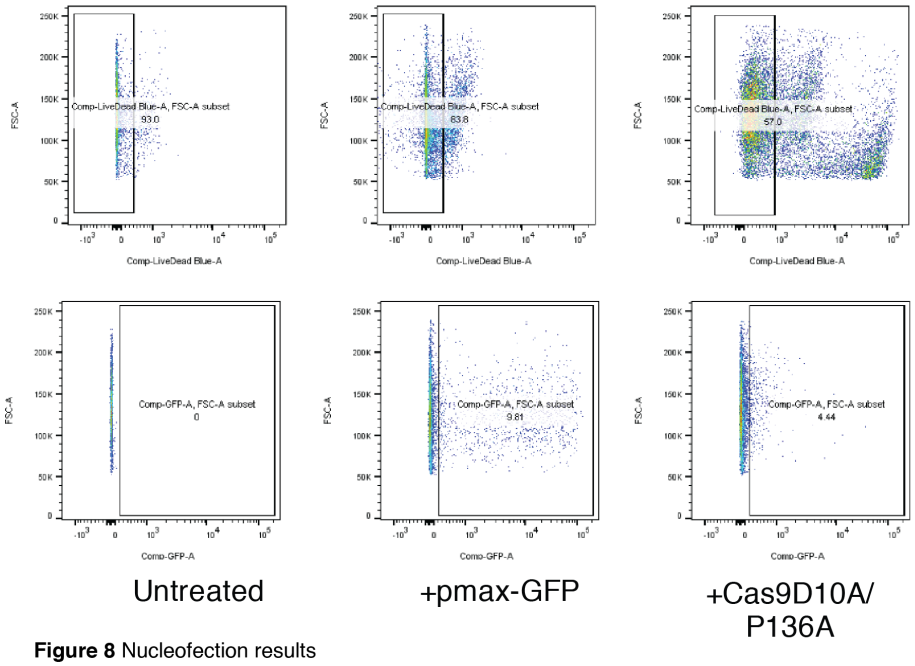

Figure 8 Nucleofection results

The gRNA (MLM3636) and Cas9 (Cas9D10A-GFP) plasmids and ssODN were again nucleofected into MT-2 cells. 72 hours post-nucleofection, the cells were stained with live/dead viability dye and sorted for GFP expression. 0 cells were positive for GFP, suggesting an error during the nucleofection procedure.

The nucleofection was repeated. The GFP^+ cells were both bulk sorted and single-cell sorted. 181 GFP^+ cells were bulk sorted into 1 ml of RPMI and 8 GFP^+ cells were single-cell sorted into a 96-well plate. RPMI was added every 3 days to the bulk sorted cells. The cells were expanded at 37°C with 5% CO_2. These cells are still currently growing out for sequencing.

We also constructed a plasmid that contained both the gRNA expression vector and Cas9 on the same vector. Use of a plasmid that contains both the gRNA and Cas9 has the potential to allow higher throughput introduction of pathogenic mutations into target cells by requiring construction and preparation of a single plasmid. The 20-nt target sequence used in the MLM3636 plasmid was cloned into a vector backbone encoding a human U6 promoter-driven gRNA expression cassette and a CBh-driven Cas9-D10A (pSpCas9n(BB)-2A-GFP) (Figure 9). The top strand oligo includes a guanine added to the 5′ end of the guide sequence not present in the target site in order to increase transcription from the U6 promoter (Figure 9b, c).

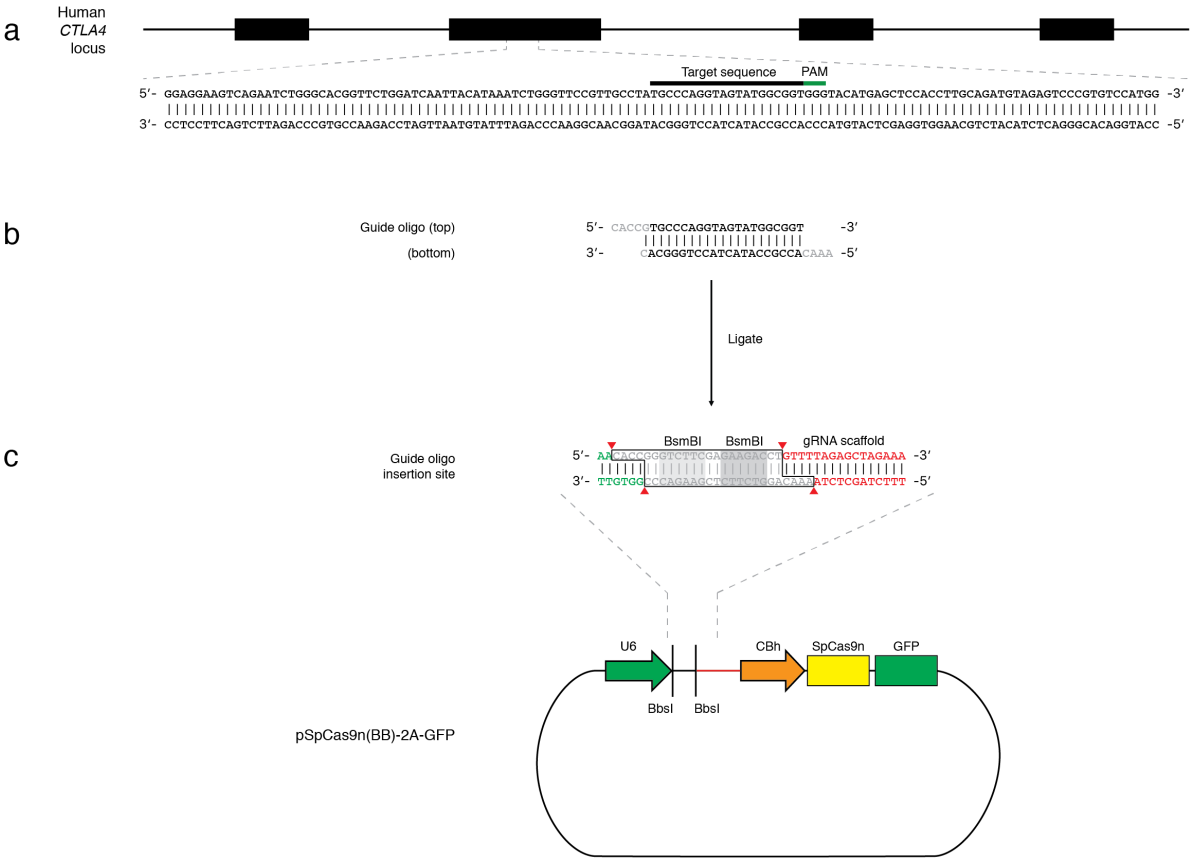

Figure 9 Targeting strategy and gRNA plasmid construction (**a**) The 20 bp target sequence must be immediately 5' of a 5'-NGG PAM sequence. *CTLA4* sequencing alignment showing target sequence selection. (**b**) The top strand oligo is the target sequence and the bottom strand oligo is the reverse complement of the target sequence. The top and bottom strand oligos have the same orientation as the sequences in the genomic target. The guide sequence oligos have overhangs specific for ligation into the pair of BbsI sites in pSpCas9n(BB)-2A-GFP (gray). (**c**) The pSpCas9n(BB)-2A-GFP plasmid was digested with BbsI, which enables the replacement of the Type II restriction sites (black outline) with direct insertion of annealed oligos into the guide oligo insertion site of the vector.

The low efficiency of nucleofection observed in MT-2 cells is, at least partially, due to the large size of the Cas9 plasmid. Stable expression of Cas9 by the target cell would negate the need for a Cas9 expression plasmid. The

use of transposable elements allows transgene integration of a gene of interest into a target cell. The Sleeping Beauty (SB) transposable system was used to generate a MT-2 cell line that stably expresses Cas9[41]. Plasmids that contain the Sleeping Beauty transposon and transposase were nucleofected into MT-2 cells. The transposon plasmid has Cas9, GFP, and puromycin, which is all stably integrated into genome of the cell. MT-2 cells were nucleofected with pSBbi-GB and SB100X plasmids (Figure 10). 2 days after nucleofection, GFP^+ cells were bulk sorted into RPMI. 7 days post-sort, single cells were plated via limiting dilution to expand clonal populations.

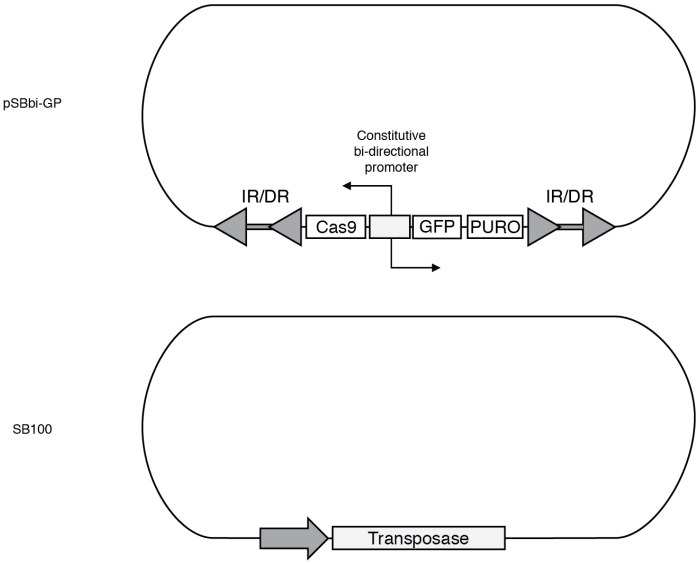

Figure 10 Sleeping Beauty transposon and transposase plasmids

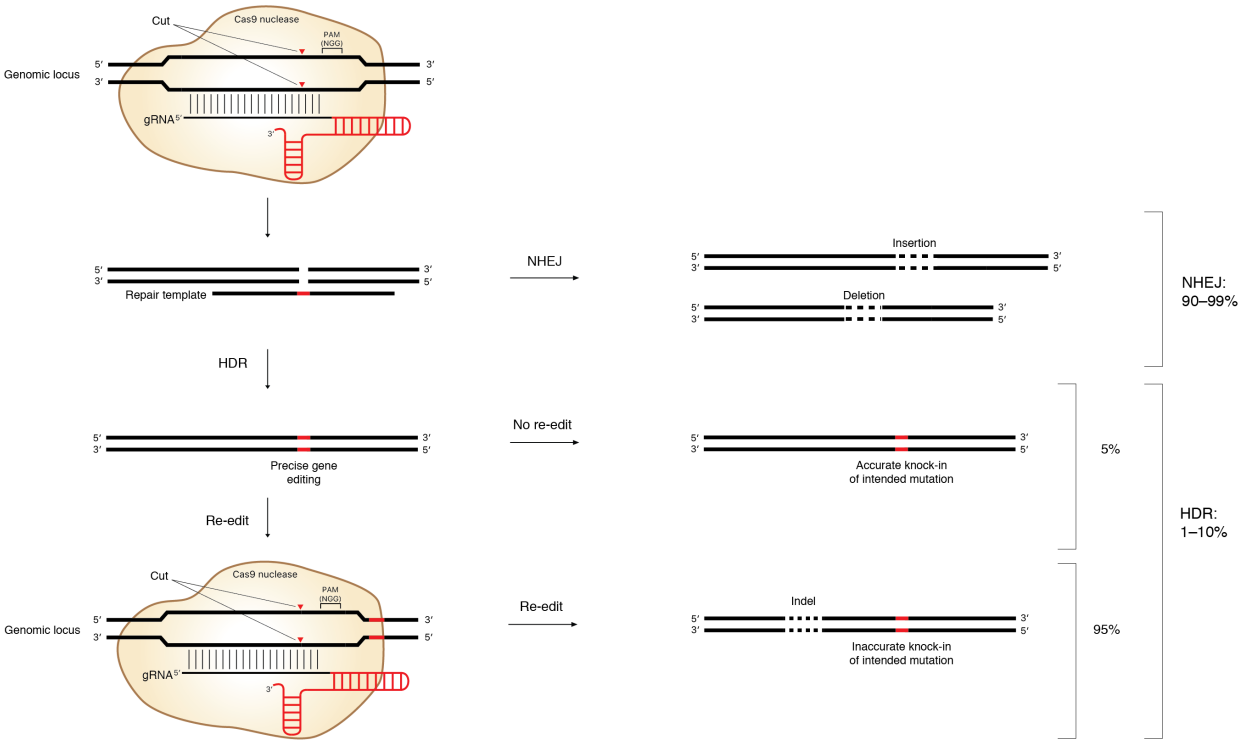

Figure 11 ssODN HDR template design using CRISPR/Cas9-blocking mutations increase HDR accuracy by preventing re-editing by Cas9 (**a**) *CTLA4* sequencing alignment showing ssODN sequence with c.406C>G mutation (red). (**b**) *CTLA4* sequence modification using repair ssODN resulting in a CRISPR/Cas-blocking mutation of the PAM sequence. *CTLA4* sequencing alignment showing the introduction of a CRISPR/-Cas-blocking mutation by including a blocking mutation in the PAM sequence (red) using the antisense sequence for illustrative purposes. Although the pathogenic c.406C>G mutation is in the sense strand and the sequence used in the repair ssODN is sense, a sense or antisense repair ssODN can be used for HDR. The introduced mutation in the antisense strand would be a G>C base substitution, resulting in a blocking mutation (red arrow) in the PAM. This blocking mutation prevents re-editing by Cas9 by mutating the NGG sequence required by Cas9 for targeting.

3. Chapter 3: Discussion and Perspectives

3.1. Limitations

Widespread use of CRISPR-mediated gene editing for modeling disease-associated mutations is currently limited by low HDR efficiency. This has prompted the development of various new strategies that all aim to increase the efficiency of precise genome editing by use of CRISPR/Cas9 systems[42-44]. Although this work did not result in an edited cell line with our patient mutation, it has furthered our search for an optimized protocol for modeling mutations using CRISPR/Cas9.

Genome editing in cell lines via homologous-recombination-mediated CRISPR/Cas9 is inefficient due to multiple variables. Delivery of DNA repair templates is difficult and methods for efficient transfection vary widely between cell type. Difficult to transfect cell lines, like MT-2 cells, require the use of viral transduction or nucleofection as transfection methods. Viral transduction has very high transfection efficiencies but requires that DNA be introduced as viral vectors. Introduction of a repair template is not feasible using this approach. Nucleofection can reach high transfection efficiencies and also allows us to deliver all necessary CRISPR/Cas9 system components.

Nucleofection of MT-2 cells has been successfully reported but is not well characterized. Nucleofection programs are cell-type specific and apply specific high-voltage pulses dependent on cell type to facilitate transfer of exogenous DNA into the nucleus. Thus, cytotoxicity and efficiency are dependent on optimization of the program being used for the cell-type being nucleofected. MT-2 cells do not have a preset nucleofection program. Due to similar structure and composition, we used a nucleofection program for Jurkat cells. Optimization data for a closely related human T cell leukemia cell line called MT-4 reports a nucleofection efficiency of 60% for MT-4 cells. This suggests that our program is optimal for MT-2 cells since we obtained 65.7% efficiency of nucleofection with the control pmaxGFP plasmid. Thus, it is hypothesized that low GFP expression is likely due to reasons other than nucleofection conditions.

The pmaxGFP plasmid is 3486 bp whereas the Cas9D10A-GFP plasmid is 9271 bp. Size is a factor for uptake efficiency in nucleofection. Nuclear delivery of large plasmids is more difficult than small plasmids. This is seen when using equivalent mass or molar concentrations of plasmid constructs of varying sizes. Alternative methods for Cas9 expression in the target cell would negate the need for a Cas9 expression plasmid. Given the high

toxicity of nucleofection that leaves us with very few viable cells, using an approach to increase the percentage of these viable cells that have successfully incorporated the foreign DNA is critical. The number of cells that we have been able to recover after the nucleofection has been consistently low, despite the method we have used post-nucleofection. This suggests that MT-2 cells are highly dependent on support from other cells to grow. Control experiments we performed subjecting MT-2 cells to nucleofection without a plasmid present resulted in similar results. This shows that the cell death observed is due to the nucleofection process itself and not to toxicity associated with the plasmids or Cas9. Although sorting the cells is a stressor that the cells do not tolerate well, selection of positive cells is necessary due to the low frequency of viable cells positive for plasmid incorporation. Otherwise, the nucleofected cells would have to be single-cell sorted by limiting dilution. This would require a very large number of cell populations to be expanded and sequenced, the vast majority of which would be negative for plasmid incorporation. This process would be both expensive and time consuming with low throughput.

Transposable systems can be used for integration of a gene of interest into a target cell line. Stable gene expression can be achieved using transposable elements (TEs) as a gene transfer system, negating the need for plasmid-based gene expression. The Tc1/mariner-type Sleeping Beauty (SB) transposable system consists of a transposase and a transposon that contains a gene-expression cassette for the gene that is to be transferred. The transposase is an enzyme that catalyzes the movement of the transposon. The transposon plasmid contains a promotor that directs transcription of the gene of interest. A second plasmid contains the gene for expression of the transposase enzyme. The SB transposase is expressed and the enzyme binds to the inverted terminal repeats (ITR) that flank the transposon and cuts the transposon out of the plasmid via an endonuclease reaction. The released transposon can then bind a different DNA molecule with a TA sequence. The transposase creates a DSB in the DNA which allows the transposon to integrate into the DNA. Since transgene integration occurs only at TA dinucleotides, there is no insertion into promoters or first introns of actively transcribed genes as seen when using retro- and lentiviral systems. The SB transposon vector can be designed for constitutive or inducible expression of the gene expression[41]. The transposon plasmid used to generate our Cas9 stable MT-2 cell line has Cas9, GFP, and puromycin, which is all stably integrated into genome of the cell. Although a large Cas9 expression plasmid is no longer required for Cas9 expression in these cells, our use for this stable transgenic cell line is currently limited. The gRNA plasmid must contain a selection marker of cells positive for the plasmid post-nucleofection. The MLM3636 gRNA plasmid used with the Cas9D10A-GFP plasmid did not contain a selection marker. GFP expression in our

nucleofected cells indicates successful uptake of the larger Cas9 plasmid. We thus used GFP expression for a marker for uptake of both plasmids due to the smaller size of MLM3636 and the tendency that cells have to incorporate all plasmids present together. Since the SB transposon vector contained a GFP expression cassette under a constitutive promoter as part of the transgene, the MT-2 Cas9 stable cell line constitutively expresses GFP. Therefore, GFP cannot be used to select for positive gRNA plasmid uptake in the stable cell line. Due to very low percentage of MT-2 cells that are GFP^+ after nucleofection, using a gRNA plasmid without a selection marker is not an option. Use of a gRNA vector backbone with RFP was considered but was not an optimal option due to a difficult and expensive cloning protocol. Although use of a gRNA-RFP vector may have allowed us to increase the efficiency of nucleofection closer to the rates we observed with the control pmaxGFP plasmid, use of this method for cloning at a higher-throughput is not favorable due to cost and difficulty.

3.2. Future Research

The Alt-R CRISPR/Cas9 system was recently developed by IDT as an alternative to traditional plasmid-based delivery of CRISPR/Cas9 components. The Alt-R CRISPR/Cas9 system uses a Cas9 protein that is combined with a crRNA:tracrRNA complex to form a ribonucleoprotein (RNP) complex, which is then delivered to the target cells. The Alt-R CRISPR/Cas9 system uses a 2-oligo system of a 67-nt tracrRNA and a 36-nt crRNA. The lengths of these RNAs were determined by systematic variation to provide the highest gene editing efficiency. The tracrRNA can be fluorescently labeled for use in imaging or FACS sorting. The crRNA and tracrRNA are combined to form the gRNA complex, which is then combined with Cas9 to form the RNP complex. The RNP and HDR template can be delivered by lipofection or nucleofection. Although the components would still need to be delivered into MT-2s by nucleofection, this method offers several potential advantages over the approach we have been using. This system has shown higher on-target gene editing efficiency with less toxicity to the cells. By delivering our CRISPR/Cas9 components as a RNP complex, this may reduce the toxicity observed in the cells post-nucleofection. IDT also reports higher editing efficiency using the Alt-R system, which is still an issue for precise gene editing even after optimizing delivery methods of the CRISPR/Cas9 system components.

The use of CRISPR/Cas9 for the insertion of precise genetic modifications is limited by the low efficiency of HDR as compared to the high efficiency of NHEJ[32-35]. HDR is less frequent than NHEJ and only occurs during S and G2 phase, whereas NHEJ occurs throughout the cell cycle. HDR does not occur sequentially but instead concurrently with NHEJ. Inhibition of NHEJ improves the frequency of HDR and therefore methods to inhibit

components of the NHEJ pathway are being used to promote homologous recombination. Molecules involved in NHEJ can be suppressed by gene silencing or inhibited by use of small molecule inhibitors. DNA ligase IV is a key enzyme for NHEJ. Scr7 is a DNA ligase IV inhibitor and functions by targeting the DNA binding domain of DNA ligase IV, reducing its affinity for DSBs and inhibiting its function. The use of Scr7 has been shown to increase the efficiency of HDR-mediated genome editing in both mammalian cell lines and in mice up to 19-fold[42, 43]. NHEJ repair can also suppressed by targeting DNA ligase IV using adenovirus 4 (Ad4) E1B55K and E4orf6 proteins, which mediate the ubiquitination and proteasomal degradation of DNA ligase IV. Gene silencing of KU70, KU80, or DNA ligase IV by short hairpin (sh)RNA have also been used for NHEJ suppression[43]. Inhibition of NHEJ can induce apoptosis and therefore cell lines have varying sensitivity to treatment with NHEJ inhibitors. Manipulation of cellular repair pathways can also affect the ability of the cell to respond to and repair damage at other non-target sites, which may lead to tumor formation and, as a result, may not be feasible in therapeutic gene editing. However, manipulation of cell cycle and cellular repair pathways can be used to increase the frequency of CRISPR/Cas9-mediated precise gene modifications to optimize disease modeling in both cell lines and mouse models[42-44].

Cas9-mediated HDR frequencies can also be increased by specific design of the repair template. Linearized or double-stranded plasmid donors are often used as donor templates to incorporate large modifications or tags. For small modifications or insertions, ssODN templates are more effective than plasmid donors. HDR events are more frequent with repair templates with longer homology regions because the rate of recombination increases as the length of homology arms increases. Homology arms of a ssODN can be as low as 40-nt but longer homology regions have been reported to increase HDR efficiency. The length of ssODN donors can be optimized to increase the incorporation rate of intended mutations. Longer ssODNs have more homology between the repair template and the target site. However, longer ssODNs also have disadvantages in that they are more likely to be incorrectly synthesized or have secondary structure[45].

Recent research suggests that HDR frequency can be further improved by using donor ssDNA complementary to the nontarget strand (—i.e., the sequence of the target strand). Donor–nontarget strand complementarity increases the frequency of HDR events when using wild-type Cas9 or Cas9 variants (Cas9D10A or Cas9H840A). It has also been recently demonstrated that homology-directed genome editing can be improved by using asymmetric donor DNA as opposed to donor DNA symmetric around the break. Cas9 asymmetrically releases DNA at the 3′ end of the cleaved DNA strand not complementary to the gRNA (nontarget strand) before complete

dissociation. Asymmetric donor DNA with 36 bp on the PAM-distal side of the Cas9 cut site and 91 bp on the PAM-proximal side of the break complementary to the strand that is released first (nontarget strand) was found to have the optimal design for annealing and is reported to increase the rate of HDR up to 60%[45, 46].

The frequency of HDR can be improved using these methods, however, this does not address the accuracy of HDR. Most HDR events do not result in the intended sequence change due to re-editing (Figure 11). Cas9 can re-cut the edited locus, causing additional editing. The target locus can be re-cleaved until the guide sequence or PAM is sufficiently modified by NHEJ that Cas9 is no longer able to recognize the target. HDR-mediated genome-editing accuracy can be improved by blocking re-editing. Introducing CRISPR/Cas-blocking mutations in the guide sequence or PAM blocks re-cutting by Cas9 and, as a result, improves HDR accuracy. CRISPR/Cas-blocking PAM mutations are more efficient than gRNA sequence mutations. This is due to off-target activity by Cas9. Cas9 is able to tolerate up to five mismatches in the gRNA sequence, which means that Cas9 may still cleave a DNA sequence without complete guide sequence recognition[45-47].

CRISPR/Cas9 editing is mostly biallelic and selective homozygous or heterozygous HDR-mediated gene editing is not well characterized. Zygosity of the intended mutation can be controlled by designing the gRNA sequence and donor template with specific cut-to-mutation distances. The full ssODN template is not always incorporated during HDR. The probability that the mutation will be introduced decreases as the cut-to-mutation distance increases. This relationship does not vary between loci. Thus, there is an inverse relationship between the mutation incorporation rate and the distance from the cut site. Cut-to-mutation distance should be minimized for homozygous mutation introduction. Biallelic editing can be achieved at higher rates by selecting a gRNA that mediates a DSB at minimal distance (<10 bp) from the intended mutation site. In contrast, by selecting gRNA and donor sequences with longer cut-to-mutation distances (2–26 bp), there is a lower likelihood that the mutation will be incorporated and thus a higher frequency of heterozygous mutation incorporation. However, it is not always feasible to select a gRNA sequence that will mediate a cut-to-mutation distance of <10 bp. Another method to introduce monoallelic mutations in cases such as these is by using equimolar quantities of the donor template containing the intended mutation and a donor template that is wild-type at the intended mutation site[45, 47].

Homologous-recombination-mediated CRISPR/Cas9 gene editing offers great potential for modeling human disease in cell lines and mouse models[48-50]. CRISPR-mediated gene editing can be used to model the repression or activation of genes, disease-associated gene sequences, and epigenetic modifications involved in the

onset and progression of human disease. The use of the bacterial CRISPR/Cas9 system in editing and regulating genomes has the capacity to change the way human disease is researched and treated. However, limitations prevent its widespread use for precise gene editing. The future of CRISPR-mediated HDR depends on the development of novel strategies to optimize gene editing. Given the variability that exists between cell lines and models, it is important that different strategies be employed to develop an optimal protocol for the introduction of gene edits in the model being used. Using several different approaches in parallel allows testing of these different approaches to develop an efficient workflow in a timely manner. The combined use of various strategies to address the inefficiency of HDR-mediated genome editing has high potential to lead to the development of a protocol that allows a high throughput system for generating specific mutations.

4. Bibliography

1. Flajnik, M.F. & Kasahara, M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat. Rev. Genet.* **11**, 47–59 (2010).

2. Bacchetta, R. & Notarangelo, L.D. Immunodeficiency with autoimmunity: beyond the paradox. *Front. Immunol*. **4**, 77 (2013).

3. Maródi, L. & Notarangelo, L.D. Immunological and genetic bases of new primary immunodeficiencies. *Nat. Rev. Immunol.* **7**, 851–861 (2007).

4. Resnick, E.S. & Cunningham-Rundles, C. The many faces of the clinical picture of common variable immune deficiency. *Curr. Opin. Allergy Clin. Immunol.* **12**, 595–601 (2012).

5. Bogaert, D.J.A., Dullaers M., Lambrecht, B.N., *et al*. Genes associated with common variable immunodeficiency: one diagnosis to rule them all? *J. Med. Gen.* **53**, 575–590 (2016).

6. Maffucci, P., Filion, C.A., *et al*. Genetic diagnosis using whole exome sequencing in common variable immunodeficiency. *Front. Immunol*. **7**, 220 (2016).

7. Kuehn, H.S. *et al.* Immune dysregulation in human subjects with heterozygous germline mutations in CTLA4. *Science* **345**, 1623–1627 (2014).

8. Schubert, D. *et al*. Autosomal dominant immune dysregulation syndrome in humans with *CTLA4* mutations. *Nat. Med.* **20**, 1410–1418 (2014).

9. Sakaguchi S. *et al*. Regulatory T cells and immune tolerance. *Cell* **133**, 775–787 (2008).

10. Wing, K. *et al*. CTLA-4 control over Foxp3+ regulatory T cell function. *Science* **322**, 271–275 (2008).

11. Wang, C.J. *et al*. CTLA-4 controls follicular helper T-cell differentiation by regulating the strength of CD28 engagement. *PNAS* **112**, 524–529 (2015).

12. Sage, P.T., Paterson, A.M., Lovitch S.B. & Sharpe A.H. The coinhibitory receptor CTLA-4 controls B cell responses by modulating T follicular helper, T follicular regulatory, and T regulatory cells. *Immunity* **41**, 1026–

1039 (2014).

13. Brunet, J.F. *et al*. A new member of the immunoglobulin superfamily—CTLA-4. *Nature*. **328**, 267–270 (1987).

14. Linsley, P.S. *et al*. CTLA-4 is a second receptor for the B cell activation antigen B7. *J. Exp. Med.* **174**, 561–569 (1991).

15. Krummel, M.F. & Allison, J.P. CD28 and CTLA-4 have opposing effects on the response of T cells to stimulation. *J. Exp. Med.* **182**, 459–465 (1995).

16. Malissen, B. *et al*. Integrative biology of T cell activation. *Nat. Immunol.* **15**, 790–797 (2014).

17. Schildberg, F.A. *et al*. Coinhibitory Pathways in the B7-CD28 Ligand-Receptor Family. *Immunity* **44**, 955–972 (2016).

18. Zhang, X. *et al*. Crystal structure of the receptor-binding domain of human B7-2: Insights into organization and signaling. *PNAS* **100**, 2586–2591 (2003).

19. Pentcheva-Hoang, T. *et al*. B7-1 and B7-2 selectively recruit CTLA-4 and CD28 to the immunological synapse. *Immunity* **21**, 401–413 (2004).

20. Stamper, C.C *et al*. Crystal structure of the B7-1/CTLA-4 complex that inhibits human immune responses. *Nature* **410**, 608–611 (2001).

21. Yokosuka, T. *et al*. Spatiotemporal basis of CTLA-4 costimulatory molecule-mediated negative regulation of T cell activation. *Immunity* **33**, 326–339 (2010).

22. Schwartz, J.C. *et al*. Structural basis for co-stimulation by the human CTLA-4/B7-2 complex. *Nature* **410**, 604–608 (2001).

23. Darlington P.J., *et al*. Hierarchical regulation of CTLA-4 dimer-based lattice formation and its biological relevance for T cell inactivation. *J. Immunol*. **175**, 996–1004 (2005).

24. Qureshi, O.S. *et al.* Trans-endocytosis of CD80 and CD86: A molecular basis for the cell-extrinsic function of CTLA-4. *Science* **332**, 600–603 (2011).

25. Walker, L.S. & Sansom, D.M. The emerging role of CTLA4 as a cell-extrinsic regulator of T cell responses. *Nat. Rev. Immunol.* **11**, 852–863 (2011).

26. Kong, K.F. *et al*. Protein kinase C- η controls CTLA-4-mediated regulatory T cell function. *Nat. Immunol*. **15**, 465–472 (2014).

27. Valk, Elke *et al*. CTLA-4 trafficking and surface expression. *Trends Immunol*. **29**, 272–279 (2008).

28. Walker, L.S. & Sansom, D.M. Confusing signals: Recent progress in CTLA-4 biology. *Trends Immunol*. **36**, 63–70 (2015).

29. Marraffini, L.A. CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55–61 (2015).

30. Jackson, S.A. *et al*. CRISPR-Cas: Adapting to change. *Science* **356**, eaal5056 (2017).

31. Ran, F.A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).

32. Sander, J.D. & Joung, J.K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32**, 347–355 (2014).

33. Mali, P., Esvelt K.M. & Church, G.M. Cas9 as a versatile tool for engineering biology. *Nat. Met.* **10**, 957–963 (2013).

34. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339,** 823–826 (2013).

35. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol.* **31,** 233–239 (2013).

36. Hamano, R. *et al*. Characterization of MT-2 cells as a human regulatory T cell-like cell line. *Cellular & Molecular Immunol*. **12**, 780–782 (2015).

37. Lieber, M.R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* **79**, 181–211 (2010).

38. San Filippo, J., Sung, P. & Klein, H. Mechanism of eukaryotic homologous recombination. *Annu. Rev.*

Biochem. **77**, 229–257 (2008).

39. Davis, L. & Maizels, N. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *PNAS* **111**, E924–E932 (2014).

40. Symington, L. S. & Gautier, J. Double-strand break end resection and repair pathway choice. *Annu. Rev. Genet.* **45**, 247–271 (2011).

41. Kowarz, E., Löscher, D., & Marschalek, R. Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines. *Biotechnol. J.* **10**, 647–653 (2015).

42. Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* **33**, 538–542 (2015).

43. Chu, V.T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* **33**, 543–548 (2015).

44. Lin, S., Staahl, B.T., Alla, R.K. & Doudna, J.A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *elife* **3**, e04766 (2015).

45. Kwart, D., Paquet, D., Teo, S. & Tessier-Lavigne, M. Precise and efficient scarless genome editing in stem cells using CORRECT. *Nat. Protoc.* **12**, 329–354 (2017).

46. Richardson, C.D., Ray, G.J., DeWitt, M.A., Curie, G.L. & Corn, J.E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).

47. Paquet, D. *et al.* Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125–129 (2016).

48. Li, K. *et al*. Optimization of genome engineering approaches with the CRISPR/Cas9 system. *PLoS ONE* **9**, e105779 (2014).

49. Dow, L.E. Modeling disease in vivo with CRISPR/Cas9. *Trends Mol. Med*. **21**, 609–621 (2015).

50. Hsu, P.D., Lander, E.S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).