Development of Novel CRISPR Technologies for the Precise Genome Editing of Human Pluripotent Stem Cells

By

Jared Matthew Carlson-Stevermer

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biomedical Engineering)

at the

UNIVERSITY OF WISCONSIN-MADISON

2018

Date of final oral examination: 08/13/2018

The dissertation is approved by the following members of the Final Oral Committee:

Krishanu Saha, PhD, Assistant Professor, Biomedical Engineering
Shaoqin Gong, PhD, Professor, Biomedical Engineering
Kate O'Conner-Giles, PhD, Associate Professor, Genetics
Sean Palecek, PhD, Professor, Biomedical Engineering
Masatoshi Suzuki, PhD, Associate Professor, Comparative Biosciences

TABLE OF CONTENTS

ABSTRACT

Development of Novel CRISPR Technologies for the Precise Genome Editing of Human Pluripotent Stem Cells

Jared Carlson-Stevermer

Under the supervision of Professor Krishanu Saha

at the University of Wisconsin-Madison

Precision genome surgery has the potential to revolutionize the field of personalized medicine through applications ranging from basic science, to creation of *in vitro* disease models, all the way to the correction of genetic disease *in vivo*. Although the field has rapidly advanced since the development of CRISPR-based technology, the ability to generate precise edits at multiple locations within the genome remains an outstanding challenge. As a result, edited cell populations as well as isolated clones must be rigorously screened for the presence of undesired mutations before any form of clinical application such as *ex vivo* cell therapy or *in vivo* somatic cell editing can be undertaken. Thus, there is a need to increase rates of precise editing and decrease rates of imprecise mutation to improve the clinical viability of gene editing.

This thesis aims to develop new technologies to address challenges in precision genome editing and subsequently demonstrate their utility in a human pluripotent stem cell disease model. First, I describe a method to track hundreds of cell populations in real time using high-content analysis, allowing the facile selection of those that have undergone genome editing. Second, I demonstrate a new technology that is capable of tethering together CRISPR ribonucleoproteins and donor DNA to increase the frequency of precise gene surgery. Finally, I combine these technologies to correct mutations in an induced pluripotent stem cell model of Pompe disease, an autosomal recessive disorder. These methods and their applications

demonstrate advances in the field of precision genome surgery and represent an important step toward the clinical use of genetic therapies.

ACKNOWLEDGEMENTS

This thesis is a product of many people who have pushed me and supported me along every step of my educational journey.

First of all, to my family, especially my mom Beth, who encouraged my curiosity from an early age and provided countless opportunities to learn and grow. I always knew I would have their support no matter where my path took me. They enabled me to take chances and go places that I never would have been able to on my own. Without them, I certainly would not be where I am today.

Second, to my advisor, Dr. Krishanu Saha, who took a chance on me and gave me an opportunity to grow as a scientist. Your expertise and guidance has been invaluable along this journey.

Third, to the friends who have supported me both personally and scientifically over the years. Thank you for serving as a sounding board for crazy ideas and motivating me to keep pushing.

Finally to my fiancée Jill. You have been the backbone of my graduate school career and a constant source of support regardless of the circumstances. Whether it was late night dinners brought to the office or moving me to bed from the couch after a long night of writing you are always there. You have pushed me to be better than what I am by myself. I am so excited that we will spend our lives together.

LIST OF ABBREVIATIONS

4-MU	4-methylumbelliferyl
4-MUG	4-methylumbelliferyl-α-D-glucopyranoside
AAV	adeno-associated virus
BFP	blue fluorescent protein
bp	base pair
CRISPR	clustered regularly interspaced short palindromic repeats
Cas9	CRISPR-associated protein 9
DLS	dynamic light scattering
DNA	deoxyribonucleic acid
DSB	double strand break
dsDNA	double stranded DNA
ECM	extracellular matrix
EMSA	electrophoretic mobility shift assay
ERT	enzyme replacement therapy
FACS	fluorescence-activated cell sorting
GAA	acid α-glucosidase
GFP	green fluorescent protein
HCA	high-content analysis
HDR	homology directed repair
HEK	human embryonic kidney cell
hESC	human embryonic stem cell
hiPSC	human induced pluripotent stem cell
hPSC	human pluripotent stem cell
indel	insertion/deletion mutation
IVT	*in vitro* transcription
LAMA5	laminin-alpha 5
MMEJ	microhomology mediated end joining
NHEJ	non-homologous end joining
NGS	next-generation sequencing
NLS	nuclear localization signal
nt	nucleotide
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
PDMS	polydimethylsiloxane
PEG	polyethylene glycol
qPCR	quantitative PCR
rhGAA	recombinant human acid α-glucosidase
RNA	ribonucleic acid
RNP	ribonucleoprotein
RT-PCR	reverse transcription PCR
sgRNA	short guide ribonucleic acid
ssDNA	single stranded DNA
ssODN	single-stranded oligo deoxynucleotide
TCPS	tissue-culture polystyrene
TALEN	tale-like effector nuclease
μCP	micro-contact printing
ZFN	zinc-finger nuclease

CHAPTER 1: Introduction

Portions of the work in this chapter were adapted from:

Increasing the precision of gene editing *in vitro, ex vivo,* and *in vivo*.

K. Mueller*, J. Carlson-Stevermer*, and K. Saha.

Current Opinion in Biomedical Engineering, 2018

Article in Review

* *these authors contributed equally*

Genome Surgery

Genome editing is the method by which one is able to change the DNA, or base code of an organism itself[1]. This capability has a wide variety of uses in the biomedical field from basic science and the understanding of cellular mechanisms to the potential to cure genetic diseases. Many complex health issues such as Amyotrophic Lateral Sclerosis, obesity, and Alzheimer's disease may arise from genetic abnormalities[2–4]. The ability to correct these abnormalities could greatly increase the quality of life for countless patients and provide a one-time treatment paradigm rather than a lifetime of drug prescriptions[5].

Genome editing is typically initiated through the induction of a break within the DNA strands using a pair of molecular scissors, typically endonuclease enzymes. Once the strand is broken, it must be repaired by cellular mechanisms or the cell will die. DNA repair typically proceeds through two main processes called non-homologous end joining (NHEJ)[6] and homology directed repair (HDR)[7]. In most cells, NHEJ is the predominant mechanism of repair. In NHEJ, the strands are reattached through addition and subtraction of nucleotides until a match is made. Being a trial-and-error mechanism, NHEJ may alter the original sequence and cause new mutations in the genome known as insertion/deletion mutations (indels). Undesirable mutations caused by NHEJ may lead to a wide range of challenges for the cellular environment, ultimately leading in some cases to disease formation (*e.g.* cancer)[8]. The second repair mechanism, HDR, uses a secondary piece of DNA to act as a molecular glue to tie the broken strands together. In this way, the DNA acts as a template to precisely correct (or modify in the case of gene editing) the genome with the desired sequence. To tether strands together, this piece of DNA contains bases on either end that match the genomic sequence around the cut site known as homology arms. In between the homology arms, single nucleotide variations or novel sequences of DNA can be incorporated and are subsequently inserted into the genome. In this manner, disease-causing variants can be corrected to create isogenic cell lines[9] or

fluorescent proteins can be incorporated to tag proteins of interest[10]. The process of crafting these precise manipulations within the genome is often referred to as genome surgery. The possibility of favoring HDR over NHEJ carries major implications for the adoption of genome editing therapeutics in the clinic. To reflect this, clinical trials for focusing on gene knockout such as those for HIV[11] or PD-1[12] are being explored. These gene knockout strategies are both NHEJ based, and are dependent on aberrant repair causing dysregulation of the disease causing gene. However, addressing more complicated disorders such as polygenic conditions or autosomal recessive disorders remains an outstanding challenge in the field.

Genome editing using CRISPR systems

The Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) and the CRISPR Associated Protein 9 (Cas9) system is a the latest pair of molecular scissors that overcomes many of the hurdles of previous systems[13–15]. The key advancement of the CRISPR/Cas9 system is the use of a single guide RNA (sgRNA) in combination with a universal endonuclease that creates double strand breaks (DSBs) at a defined point in the genome (**Figure 1.1**). Use of RNA rather than protein binding motifs such as those utilized by tale-like effector nucleases (TALENs)[16] and zinc-finger nucleases (ZFNs)[17] imparts a higher level of specificity to gene editing. Additionally, as RNA is easy to produce *in vitro,* CRISPR systems can be rapidly tested and highly multiplexed[15]. A sgRNA consists of two distinct regions; first, a sequence of 20 base pairs known as the protospacer at the 5' end of the RNA denotes the locus of targeting by Watson-Crick base pairing. This sequence is generally 20 nucleotides in length and only occurs every 10^{12} bp, imparting a high level of specificity. Early research in the CRISPR field has shown that this length is malleable[18] and that certain regions of the sequence may contribute more to the effectiveness of genomic targeting[19]. Second, following the target sequence is a scaffolding region that consists of three stem loops used for recognition by the Cas9 protein. This region has been shown previously to be tolerant of mismatches and

additions while preserving activity, allowing for regions that can be engineered to provide added functionality[20,21]. Cas9 binding of DNA also requires a protospacer adjacent motif (PAM). This sequence is recognized by Cas9 and varies between species[15,22–24]. Studies using labelled Cas9 molecules and dynamic tracking have shown that when a Cas9 ribonucleoprotein (a complex of Cas9 and sgRNA, RNP) enters the nucleus the locus is not actively targeted. Rather, RNPs slide through the genome looking for a PAM site and match the sequence of the protospacer[25]. At this point, a conformational change is undergone and the two cutting domains of the Cas9 protein break the DNA strand, forming a double strand break (DSB) to be repaired by cellular machinery. However, due to this mechanism of action, incorrect sequences are sometimes recognized as the editing site, leading to off-target effects.

Precise on-target nuclease activity

From the inception of genome editing, researchers have been concerned with the ability to edit the genome at the target site, while limiting edits elsewhere within the genome, commonly called “off-target effects” (**Figure 1.2**). Shortly after CRISPR systems were identified as genome editing tools[26,27], several groups raised concerns that Cas9 may create excessive undesirable mutations[28–30]. Varying rates of off-target events were reported ranging from >1000 per sgRNA sequence[31] to negligible effects[32], resulting in calls to develop better off-target screening methods. Recent discoveries that mutations within individual patients that introduce novel PAM sites have amplified these calls[9]. Popular techniques to quantify off-target sites include Digenome-seq, Circle-Seq, Guide-Seq, and integrase-deficient lentiviral vectors[33–37]. Use of these methods illuminated off-target sites that were not predicted by standard bioinformatic sequence analysis and necessitated introduction of new techniques that may avoid these issues.

Following these observations, many researchers adopted new methodologies to controllably introduce genome editing components, including tighter stoichiometric control that replaced plasmid-based systems with ribonucleoproteins[38] (RNPs), and strategies to regulate when and where Cas9 is expressed following viral integration[39–41]. These controlled methods showed a concurrent decrease in the number of off-target effects[40], and it is likely that RNP-based editing systems will remain popular in the clinic. The timing and location of editing events can also be modulated by light and small molecules to control nuclease activity[42]. Other modifications allow Cas9 nucleases can be selectively activated with small molecules to decrease the gene editing time window[39]. These ligand-dependent nucleases demonstrate 25-fold higher specificity with regards to on-target vs. off-target edits, and can also be used to induce Cas9 functionality *in vivo* at time-specific intervals during development. Combined, these technologies highlight the power of dynamic temporal control over genome engineering machinery.

Additional efforts to decrease off-target effects have emphasized further modifying the nuclease, including engineered “nickase” Cas9 proteins featuring only one active nuclease domain. When used alone, nickases are unable to create a full double strand break (DSB). However, when two nickases are paired, the resultant break is treated as a DSB and can be repaired via non-homologous end-joining (NHEJ)[43]. While this method lowers off-target effects, the efficiency of genome editing is greatly decreased, as two nickases and two sgRNAs need to be introduced into the cell simultaneously; thus, practical applications of nickase therapeutics may be limited by this delivery challenge. Others have engineered Cas9 RNP by mutating residues that interact with the protospacer-adjacent motif (PAM) (5`-NGG-3` in *S.Pyogenes*) region of the sgRNA[44]. These modifications expand the targeting capabilities of Cas9 to recognize PAM sites that occur less frequently throughout the genome, thereby decreasing off-target binding capability. Cas9 proteins from different species (*N. Menengitis*[45], *S. Aureus*[46]) have a similar potential.

Insights into Cas9 structural biology[47,48] yielded a rational design approach to create high-fidelity variants of Cas9: eSpCas9[49],Cas9-HF1[50], and xCas9[51]. Cas9 variants function by decreasing the binding time of the sgRNA to the target sites within the genome, resulting in a decrease in off-target binding and cutting. All variants also claim to only slightly decrease the frequency of on-target DSB formation. Recent work has shown that they actually decrease off-target cuts by modifying the kinetics of the change in the structural formation[52], but may also work poorly when introduced into human cells without modifications to the sgRNA[53]. At the moment, these high-fidelity Cas9 variants may represent a quick path to clinical relevance as they can greatly reduce off-target events. Further, due to the flexible nature of CRISPR engineering, many of the prior strategies can be combined to further drive precise on-target activity.

A promising new method to completely avoid the potential of off-target DNA editing that has been is to skip that part of the central dogma of biology entirely. Instead, a new member of the CRISPR family, Cas13 has been shown to target RNA post-transcription[54] and has RNase capability. In this manner, proteins of interest can be knocked out while maintaining genomic integrity. However, the specificity of this platform is still under investigation. Use of Cas13 has shown particular promise in the SHERLOCK system[55] for diagnostics and is currently a highly active area of research looking for new problems to solve. In the context of genome surgery, Cas13 could be used for short term knockdown of genes in developing systems but would need to be integrated in the genome to have a long-term therapeutic effect. This requirement highlights the need for development of low off-target methodologies described above.

Precise scarless incorporation of new sequences

While some desired outcomes can be accomplished via error-prone DSB repair[56] (e.g. NHEJ), there are still challenges that can only be solved through precise point mutations or transgene insertion via homology-directed repair (HDR), herein termed “scarless editing”

(**Figure 1.2**). Researchers have thus attempted to increase both the overall efficiency of HDR as well as the ratio of precise to imprecise mutations (**Figure 1.3**).

Early work to increase scarless editing levels focused on modulating cellular DSB repair pathways. Multiple groups showed that small molecule mediators (e.g. SCR7, L755507) can increase the relative proportion of scarless editing events[57,58]. More recently, co-introduction of an i53 protein[59] was able to increase scarless editing. These methodologies are most applicable for *in vitro* cell culture applications where potential toxicity is less limiting.

Building on this idea, other research demonstrated that timed delivery of gene-editing particles to certain points in the cell cycle corresponding to DNA synthesis could also increase HDR rates[60]. This method has been further applied by synchronizing cell cycles and subsequent timed delivery during S-phase[61]. Others have combined these ideas with Cas9 protein engineering methods. This technique, called Cas9-hGem, retains Cas9 protein within the cell only when HDR is favored during the cell cycle[62]. These methodologies hold great promise for *in vitro* cell culture applications, but may be untenable for *in vivo* editing where the majority of cells are in a post-mitotic state.

For point mutations and short knock-ins, single-stranded oligodeoxynucleotide (ssODN) templates hold significant promise for treating disease variants due to their ease of synthesis. However, sequence changes encoded by the ssODN are infrequently incorporated after editing (<10%), and desired edits are typically outnumbered by other sequence outcomes (presumably NHEJ). Recent reports have shown that ssODN design can significantly alter how the DSB is repaired. By making ssODN homology arms asymmetrical around the cut site, HDR can be promoted up to 5-fold higher over symmetrical ssODNs[63,64]. Other methods have made use of synthetic DNA synthesis to add in unnatural phospothioetherate bonds to prevent the ssODN from degredation[35,65] However, these methods still require free-floating foreign DNA that is not necessarily available to create the desired edit at the cut site. To solve this problem, several groups have tried strategies to link the ssODN to Cas9. For example, the sgRNA and ssODN

can be chemically tethered and assembled together in a nanoparticle[66]. Other techniques have leveraged avidin/biotin binding capabilities to link the ssODN directly to the CRISPR protein[67] or the sgRNA through the use of accessory proteins and RNA aptamers[68]. Each of these methods increase the ratio of precise edits:imprecise mutations and could potentially be delivered as a preassembled RNP for an all-in-one therapeutic. In spite of the widespread application, use of methodologies that promote HDR have recently been met with caution. Multiple reports have suggested that cells which incorporate foreign DNA into the genome, are more likely to have mutations in p53, a protein that is responsible for tumor suppression[69,70]. Future use of gene-edited cells in a cell therapy approach must account for this possibility and be rigorously screened to avoid potential complications.

Finally, other methods attempt to avoid the use of HDR altogether, and instead leverage other DNA repair pathways[71,72]. One method that has been explored, especially in post-mitotic cells, is microhomology-mediated end-joining (MMEJ). This pathway is thought to use small regions of homology between single strands that form via exonuclease activity during resection after the DSB. This pathway may contribute to commonly occurring indel mutations[73] or small insertions[74], although recent work has shown that integration of novel DNA may be possible. Several groups have developed protocols using small homology arms that undergo resection to form regions that overlap with genomic DNA[75]. The most well-known technology using this pathway is homology-independent targeted integration[76], which requires only 8 bp of homology around the cut site to insert full transgenes in post-mitotic cells *in vivo*. In other recent work, it has been shown that if the disease mutant is only presence on one allele within the cell, the other allele may function as a repair template[77]. This idea may be useful in the correction of recessive disorders. These approaches avoid some of the challenges in assembling long homology arms into donor constructs. As with precision on-target editing mentioned previously, strategies to increase scarless incorporation, such as the introduction of small molecule

inhibitors can be combined with protein engineering to further increase the precision of on-target integration.

In combination with donor DNA design, novel CRISPR systems have been discovered that may help precise gene correction. Cas12a (formerly Cpf1) functions similarly to Cas9 nickase but requires only one protein component, instead using the RNP to make a staggered DSB around the target site[78]. Similar to a Cas9 nickase, by creating a DSB in this fashion, the cell does not easily recognize break points as DSBs, and thus decreases the rate of NHEJ initiation. Further work has also shown that Cas12a has a lower rate of off-target effects, contributing to the precision of the nucleases[79]. However, following DNA cleavage, HDR can still be initiated. This increases the ratio of precise to imprecise mutations, and reduces the risk of undesired NHEJ products[80]. Most recently this has been shown in spinal muscular atrophy patient iPSCs[81], suggesting untapped potential for precision gene correction. While Cas12a has obvious benefits for precise gene editing, it has recently been suggested to possess ssDNA cleavage activity, even in the absence of a PAM site[82]; thus, it may be better suited for dsDNA templates used for large transgenes.

Base editors are particularly attractive for clinical translation, as they avoid DSBs entirely. They employ a catalytically dead version of Cas9 fused to a DNA deaminase to modify existing base pairs in the sgRNA protospacer region. Base editors deaminate cytidine bases to form uridine. These modified bases are then recognized by the cell as mismatched and corrected to thymidine[83]. Current work in this area mostly focuses on C>T (or the analogous G>A) conversions, although future versions will aim to allow modifications of any single base[84].While this technology should avoid unwanted genomic instability through the breaking of DNA strands, imprecise editing can also occur, as all C nucleotides within the protospacer region are capable of being modified. Further modifications to the base-editor system may address this shortcoming by shortening the available editing region[85].

The use of base editors to repair specific points within genome follows a larger trend within the field. As the HDR techniques become more refined, other recent work has suggested that insertion of transgenes or direct repair of mutant DNA at endogenous loci may be preferable to the insertion of foreign constructs[86,87]. Use of endogenous promoters within the cell may allow for correct expression of edited transcripts as well as controlled expression along cell differentiation pathways.

Precise editing within specific cells and tissues

While the transcriptional regulation of gene editing outcomes is a critical consideration, delivery to appropriate tissues is equally if not more important for any somatic editing approach. Precise delivery of editing components remains an extant challenge within the field, as many delivery agents suffer from low efficiency, high toxicity, and immunogenicity (**Figure 1.2**). Both viral and nonviral delivery agents have been engineered to achieve cell and tissue specificity.

Viral vectors are one of the most commonly used methods for delivering genetic payloads[88]. There is an increasing trend towards the use of adeno-associated viruses (AAV), which are capable of transducing non-mitotic cells while avoiding integration into the target genome. These vectors come in various serotypes with tropism specific to particular tissues, and have been used to edit the mammalian CNS[89,90] and retina, for which a first-in-kind AAV gene therapy, Luxterna™, has received FDA approval[91]. AAV can handle genetic payloads up to 5 kb, which limits their efficacy for some constructs; however, when used with smaller nucleases such as SaCas9, this issue is somewhat mitigated[46]. Viral constructs can also be engineered to harbor cell and tissue-specific promoters driving expression of the gene editing system[46,92], such that editing machinery is not expressed in non-desired cell types.

In spite of the relative efficiency of AAV delivery vectors, capsid immunogenicity remains a barrier[93]. Additionally, if used to deliver the nuclease sequence along with template DNA, there are significant concerns about the effects of long-term nuclease expression within the

target cell that severely dampen the potential for clinical use[94]. Thus, nonviral delivery methods, such as nanocarriers and other customized biomaterials, are being explored to circumvent these problems. In order for gene editing components to produce therapeutic effects, they must traffic to the desired tissue without producing an immune response, enter the target cell, escape the endosome, and enter the nucleus[95]. This is particularly challenging in the context of nuclease delivery, as Cas9 and other proteins are sizeable and sgRNAs carry a negative charge, two characteristics that limit cell penetration[96].

Several designs have demonstrated high gene-editing efficiencies when used with RNPs, ranging from 30-40% in cell lines, and up to 90% delivery efficiency[38,97–99]. These nanocarriers have demonstrated comparable efficiency to conventional electroporation or lipid-based reagents (*e.g.* Lipofectamine CRISPRMAX[100]), and have the potential to facilitate nuclease delivery *in vivo* at therapeutically active rates while remaining biocompatible in patients. To complement these nanoparticle designs, others have engineered gene-editing components themselves for improved tissue specificity. A recent paper described engineered Cas9 proteins featuring glycoprotein receptor ligands conferring specificity to liver cells[101]. These engineered nucleases were able to both penetrate liver cells *in vitro* and escape the endosome to confer organ-specific edits. While not yet validated *in vivo*, these findings raise the possibility for future precision editing designs featuring tissue-specific nucleases.

Application of Precise Genome Editing to Correct an Autosomal Recessive Disorder, Pompe Disease

Pompe disease is a progressive and often fatal muscular disease caused by deficiency in acid-α-glucosidase (GAA) (**Figure 1.4**). There are two described forms of the disease that have been broadly characterized as infantile and late onset[102]. While late onset disease has been shown to respond to changes in diet and exercise, infantile onset often results in rapid disease progression and death within one year without intervention[103,104]. Recent advances in

enzyme replacement therapy (ERT) and viral mediated delivery of recombinant GAA have helped to increase the prognosis of patients although both approaches retain significant hurdles that must be cleared.

ERT has been successfully used in clinical trials and two related products (Myozyme® and Lumizyme®) have been approved by the FDA for use in infantile onset patients[105], although the two products have been combined and sold under a single label. In clinical studies, patients who have been dosed with rhGAA show a reduced risk of death and achieve new motor milestones over 52 weeks of treatment[106]. However, patients require dosing with high levels (20-40 mg/kg) of rhGAA[107] every two weeks and must remain on this regiment throughout their life[108] leading to high costs, up to $300,000 per year. Additionally, repeated dosing can also lead to attenuation of therapeutic effect, particularly in skeletal muscle[109,110] and when delivered systemically the majority of rhGAA remains trapped in the liver and is ineffective in muscle tissue[111]. Additional caution must be exercised as delivery of rhGAA to the blood stream has resulted in development of IgG antibodies in the majority of patients[105,106]. Recent studies have attempted to ameliorate this response through gene therapy[112–116].

Gene therapy has been proposed as an option to treat Pompe disease as it may allow a long-term correction of GAA deficiency. Both lentiviral and adenoviral vectors have been shown to induce efficient expression of hGAA cDNA in human cells and mice models[117,118]. Viral loads are injected either directly into muscle or administered systemically and transported to the liver. Direct injection to cardiac or skeletal muscle provided long term recovery of phenotype (10-fold reduction in glycogen content) to transduced cells but did not affect non-transduced cells and required high viral loads ($>10^{10}$ viral genomes/kg)[116,119–121]. Conversely, transduction of hepatocytes resulted in a 100-fold increase in secreted GAA found in the bloodstream and 10-fold increase in tissues of treated mice when under control of a constitutive or liver-specific promoter[116–118]. Systemic delivery of recombinant AAV has also resulted in development of IgG antibodies against hGAA[122]. To avoid this response, cell therapy of transduced HLA-matched

hematopoietic stem cells have also show glycogen clearance in the heart, diaphragm, and liver[123,124] suggesting that *ex vivo* cell therapy may be a viable treatment option. New therapies have also recently been proposed that suggest genetic engineering of patient cells, either *ex vivo* or *in vivo* may be a viable treatment option that removes the hurdles of other methods. These strategies rely on inclusion of exons into the *GAA* mRNA and thus promote correct expression of endogenous protein[125]. Another study by the same group has shown that anti-sense oligonucleotides can achieve the same effect[126]. While these strategies can be used for correction of splice variants, a significant proportion of patients suffering from Pompe disease possess mutations within the coding sequence itself and would require alternate therapies.

Among these many treatment modalities, none has attempted to correct the underlying cause of Pompe disease, mutations in the *GAA* gene. This may be a promising strategy to generate corrected cells either *ex vivo* for cell therapy or *in vivo* using somatic cell editing. In this thesis, I tackle the challenge of correcting two different mutant alleles within compound heterozygous cases of infantile-onset Pompe disease. First, due to the difficult nature in isolating cells that have been edited at multiple loci, I developed a high content screening platform that utilizes phenotypic changes to identify populations of interest. This platform, ArrayEdit, is able quickly and efficiently identify edited pluripotent stem cells. Second, to investigate the role of DNA localization in the induction of HDR, I developed new nanoassemblies of CRISPR machinery. This assembly, the S1mplex, is able to promote precise editing in multiple cell types. Finally, I combine ArrayEdit with S1mplex technology to explore the effects of performing genome surgery on iPSCs from Pompe patients.

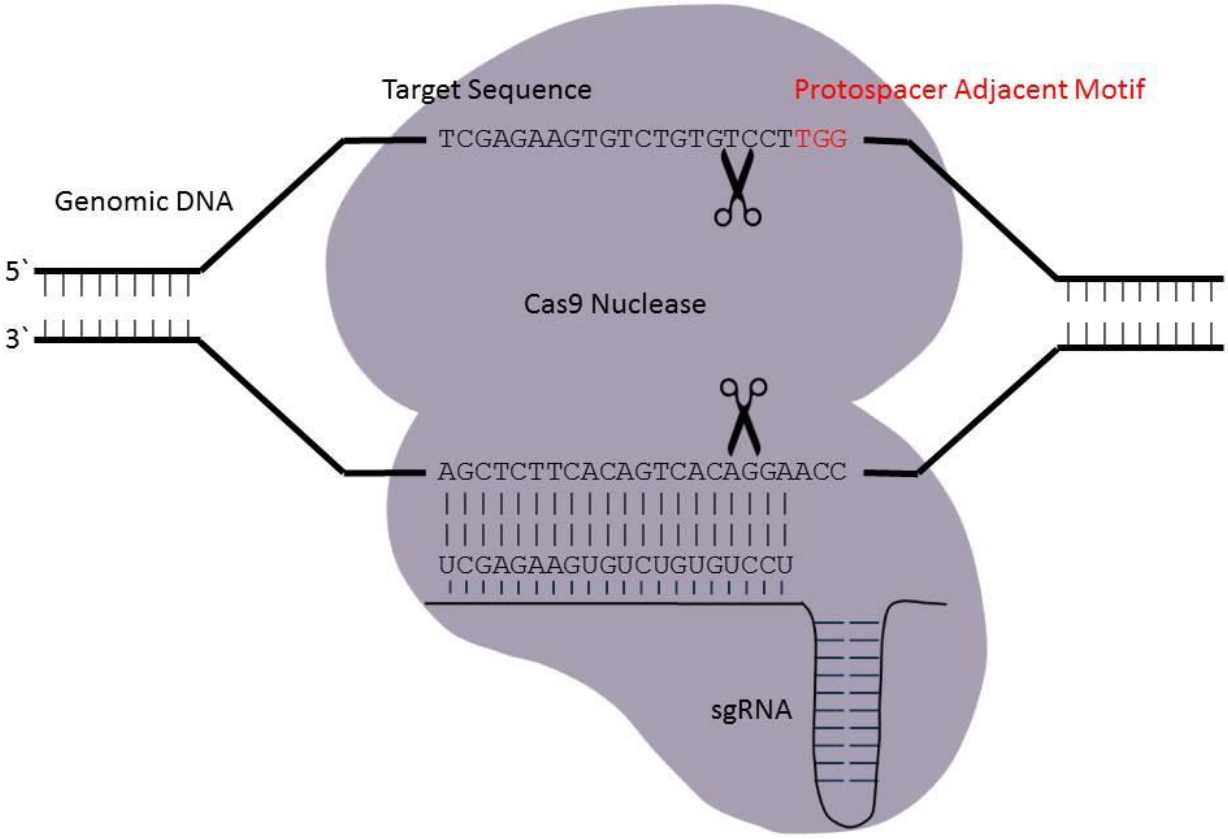


Figure 1.1 | **Schematic of Cas9 RNP complexed with genomic DNA**. Cas9 RNPs contain two components; a sgRNA and a Cas9 nuclease. sgRNAs consist of a 20nt sequence that invades double stranded genomic DNA and searches for a matching sequence to complex with. In order for a match to be found by a sgRNA, a protospacer adjacent motif (red) must be present. After this sequence is found, Cas9 undergoes a conformational change to cut both strands of the DNA, forming a double strand break that is the starting point of genome editing.

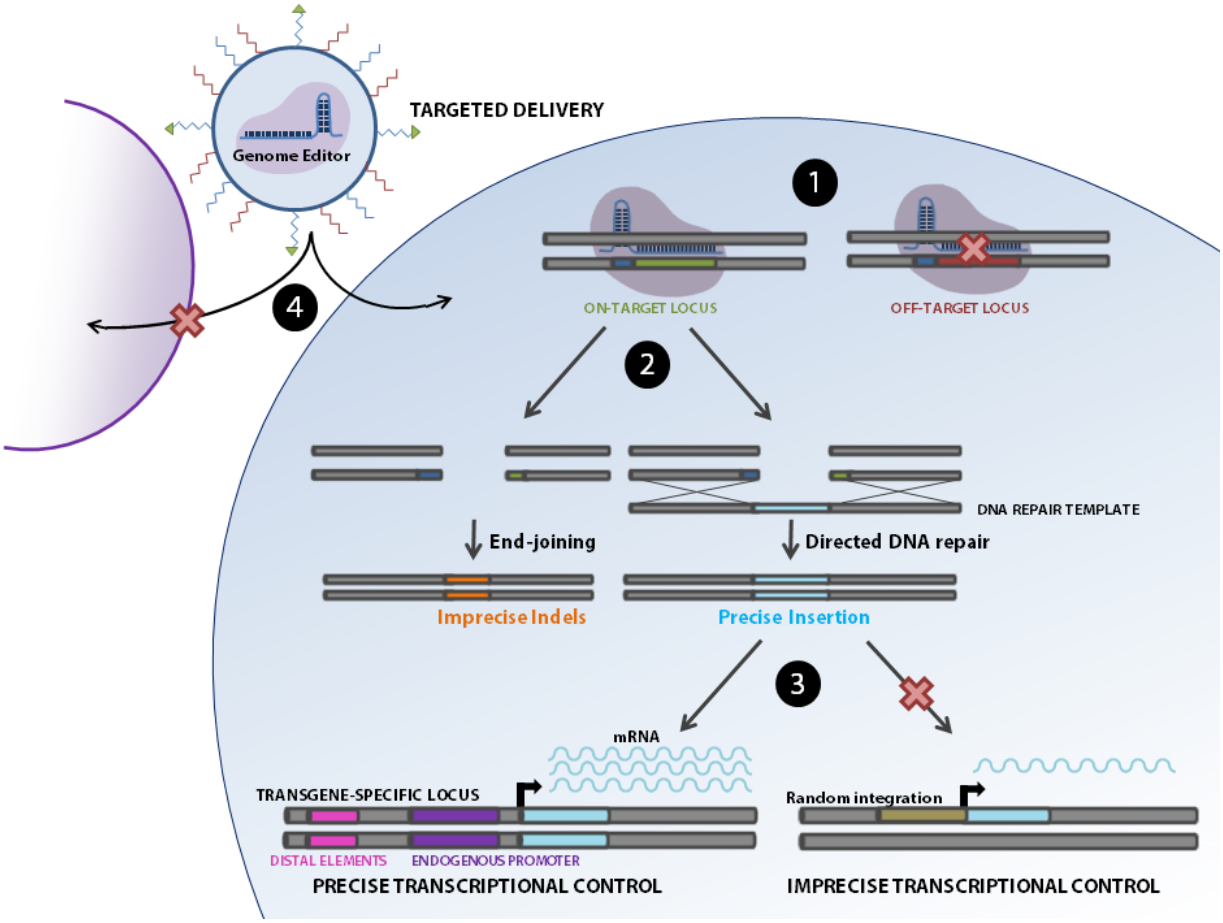


Figure 1.2 | **Precision genome editing.** Schematic illustrates four ways in which precision genome editing can be achieved: (1) the binding of genome editing machinery to the desired target genomic locus, (2) the incorporation of the correct sequence into the edited locus following DSB formation or after base editing (not shown), (3) precise regulation of integrated transgenes by endogenous promoters and distal elements in comparison to random integration, and (4) delivery to specific cell types by engineered nanomaterials or viral capsids.

Figure 1.3 | **Examples of precise and imprecise gene editing at a single locus.** Double strand breaks caused by Cas9 may be repaired in a variety of ways. Of particular interest is homology directed repair (HDR) where DNA matches exactly the provided template. However, imprecise mutations such as incorrect HDR and NHEJ events are often more prevalent in edited cell populations.

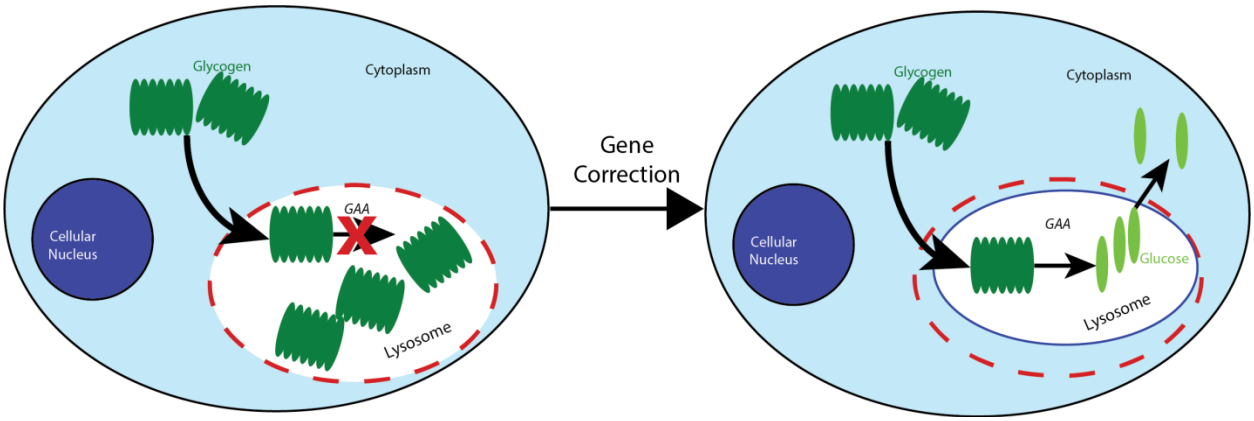


Figure 1.4 | Mechanism of Pompe Disease. Pompe disease is caused by a defect in acid-α-glucosidase (GAA). This enzyme is responsible for breakdown of glycogen within the lysosome. Loss of GAA through mutations causes the buildup of glycogen and an increase in organelle size (red dashes). Correction of these mutations through genome editing may restore enzyme function and ameliorate disease phenotype.

CHAPTER 2: High-Content Analysis of CRISPR-Cas9 Gene-Edited Human Embryonic Stem Cells

Work in this chapter was adapted from:

High-Content Analysis of CRISPR-Cas9 Gene-Edited Human Embryonic Stem Cells

Jared Carlson-Stevermer, Madelyn Goedland, Ben Steyer, Arezoo Muvaghar, Meng Lou, Lucille Kohlenberg, Ryan Prestil, and Krishanu Saha.

Stem Cell Reports 6 (1) 109-120, 2016

Abstract

CRISPR-Cas9 gene-editing of human cells and tissues holds much promise to advance medicine and biology, but standard editing methods require weeks to months of reagent preparation and selection where much or all of the initial edited samples are destroyed during analysis. ArrayEdit, a simple approach utilizing surface-modified multiwell plates containing one-pot transcribed sgRNAs, separates thousands of edited cell populations for automated, live high-content imaging and analysis. The approach lowers the time and cost of gene-editing and produces edited human embryonic stem cells at high efficiencies. Edited genes can be expressed in both pluripotent stem cells and differentiated cells. This preclinical platform adds important capabilities to observe editing and selection *in situ* within complex structures generated by human cells, ultimately enabling optical and other molecular perturbations to the editing workflow that could refine the specificity and versatility of gene-editing.

Introduction

CRISPR-Cas9, an emerging genome surgery tool, exploits an engineered ribonucleoprotein complex consisting of two essential components: *1)* a protein, Cas9; and, *2)* a single-guide RNA (sgRNA). Together, the Cas9-sgRNA complex cuts a specific target sequence in the genome. Human cells and tissues edited by CRISPR-Cas9 are important resources for drug target identification[127–129], regulatory science[130], medicine[131] and basic biology[130,132]. However, human gene editing experiments frequently require laborious cloning of expression plasmids for each sgRNA, and there are limited opportunities in these culture systems to watch and perturb genome surgery in action, as it is difficult to isolate and image living mutant cells during and shortly after the DNA cleavage event. Overall, there is a need to expand the throughput and capabilities of current *in vitro* human culture systems where novel genome surgery approaches can be evaluated with human cells and tissues[133]. Advanced capabilities

with human pluripotent stem cells in particular could eventually expand the suite of human preclinical model systems, ranging from patient-specific cell lines to complex human embryonic tissues established from stem cells.

Current gene-editing techniques generate heterogeneous human cell populations that require significant subsequent characterization. It is crucial to analyze the sequence of genome of the edited cells by sequencing before continuing with other studies, and several protocols require destruction of mutant cell populations during sequencing analysis[127–129,134–137]. For example, targeted gene disruption followed by selection and next-generation sequencing can identify drug targets, but a separate, subsequent gene-editing experiment is required to obtain living mutant cells for downstream analysis[127–129,138–140], a process that is often infeasible for slowly dividing or primary cells. This slows epigenomic and functional characterization of properly-edited cells, and it is currently unknown whether there are persistent epigenomic and functional problems within the edited cells[141]. Further sequence-level characterization is also required at the single clone level, as there is frequent and variable disruption of, or insertion of donor DNA into, the non-targeted allele in edited cell lines[142]. Finally, efficiencies of isolating precisely edited cells remains a challenge with current methods, typically with 20% or lower efficiencies to make near precise deletions in the human genome[143].

Here, we describe a platform, termed ArrayEdit, which combines two capabilities: one-pot transcription, and the combination of microcontact printed plates and high-content analysis (HCA). First, we detail a method that can generate many sgRNAs in parallel, within hours, using chemically-synthesized oligonucleotides ordered in a multi-well format. One-pot transcribed sgRNAs can be delivered without purification and can efficiently generate desired gene-edits within human embryonic stem cells (hESCs) when co-delivered with Cas9. Second, we describe a versatile combination of culture and imaging to select for edited cells and tissues using non-destructive analysis of thousands of spatially-defined features that localize edited cell

colonies/aggregates. We were able to isolate gene-edited hESC lines within two weeks, 82% of which were mutant for our desired edit at a proof-of-concept locus (*LAMA5)* without any detectable off-target mutations. This platform adds important capabilities to easily observe editing and selection *in situ* within complex structures generated by human cells.

Results

Simplified "one-pot" transcription of sgRNAs in multiwell plates

One key feature of ArrayEdit is the generation of "one-pot" transcribed sgRNAs with chemically-synthesized oligonucleotides within a multiwell format. One-pot transcription is similar to one-pot synthesis in chemistry, because products of the reaction are created at high yields without any intermediate purification steps. As outlined in **Figure 2.1A**, our method consists of three components: *1)* a forward primer containing a minimal T7 primer, sgRNA target sequence, and a region for PCR amplification; *2)* a double stranded sequence of DNA encoding the sgRNA conserved region; and *3)* a universal reverse primer for PCR amplification (see **Figure A-1A** and **Tables A-1 and A-2** for sequences). This method is versatile and can generate any desired sgRNA within hours, regardless of sequence complexity. In contrast to other methods [144–146], this process is modular, such that advances in the sgRNA backbone that refine specificity or increase editing efficiency [20,147] do not necessitate recreating entire sgRNA libraries. All primers can be chemically synthesized and delivered from commercial vendors overnight, decreasing the time between design and experiments.

To demonstrate the multiplexed synthesis of sgRNAs, we designed six sgRNAs targeting two genes (*MCHERRY*, *GFP;* **Table A-2**) and ordered four replicates of each primer set in a 96-well format for subsequent one-pot transcription. Next, PCR amplification was performed yielding a DNA product that was of consistent size and concentration across all targets (**Figure A-1B**). These DNA products are reusable and can be stored for months, or used

immediately for *in vitro* transcription (IVT) with T7 RNA polymerase. IVT was allowed to proceed for as little as two hours up, to overnight. Longer incubation times resulted in increased concentration without any undesired products or degradation of the sgRNA. When allowed to progress overnight, IVT consistently produced greater than 90 µg sgRNA in a 20 µL reaction (**Figure 2.1B**), suggesting that the concentration of rNTPs is the limiting reagent. With consistent yields, this strategy renders post-production purification or quantification unnecessary and allows us to use approximate molar quantities of sgRNA during transfection directly following IVT. This method of production also reliably produces only the desired product of the correct size as seen in a lone narrow peak when profiled by a RNA Bioanalyzer 2100 (**Figure 2.1C**).

We designed and created one-pot sgRNAs to target 74 additional loci (**Table A-2**) in both hESCs and human embryonic kidney (HEK) cells. After transfecting these sgRNAs into cells, we determined the percent gene modification in an *EGFP* transgene via flow cytometry (**Figure A-1C**) and in endogenous *FGFR2* via restriction fragment length polymorphism on genomic DNA (**Figure A-1D**). Further, we detected large-scale genome deletions in *MUC16* and *DPH7* using gel electrophoresis on genomic DNA (**Figure A-1E, A-1F**). Large-scale genome deletions were created by transfecting cells with two sgRNAs that are located 200 to 4000 bp apart from each other in the genome. A standard agarose gel was then able to resolve if the sequence between the two sgRNAs was deleted based off of amplicon size. In all cases, each one-pot sgRNA was capable of creating a targeted DNA double strand break that was likely resolved using non-homologous end joining (NHEJ).

Next, we compared the gene-editing efficiency of one-pot transcribed sgRNAs against the established methods of sgRNA production. To quantify single cell editing efficiency, we used flow cytometry in conjunction with a transgenic HEK-H2B-mCherry line which was engineered to constitutively express a fusion protein of histone 2B (H2B) and mCherry from a single *AAVS1*

'safe-harbor' locus (**Figure A-1G**). sgRNAs were produced via one-pot transcription, plasmid transfection[135], commercially (termed a U6-gBlock), and via previously described IVT methods[144]. One-pot transcribed sgRNAs resulted in the highest percentage of fluorescence expression loss, successfully knocking out expression in 60% of cells **(Figure 2.1D-E)**. Additionally one-pot sgRNA production was the quickest method, requiring only two days from design to experiment whereas commercially-produced sgRNAs required 4-5 business days and previously-described methods required a cloning step and plasmid production scale-up (~4 days). We then used one-pot transcribed sgRNAs with a transgenic hESC line: WA09-H2B-mCherry[148]. When one-pot transcribed sgRNAs targeting mCherry were introduced via electroporation into this line, some hESCs lost mCherry expression after four days of culture. Additionally, one-pot transcribed sgRNAs generated 5x more mCherry-negative cells when analyzed by flow cytometry than standard techniques that express the same sgRNAs off of plasmids[135] (**Figure 2.1E**). One-pot transcribed sgRNAs also performed nearly as well as methods that enrich for transfected cells[134] via fluorescence-activated cell sorting (FACS) for GFP in a co-transfection with a GFP-expressing plasmid.

Deep sequencing of edited stem cell derivatives

To gain a more detailed analysis of genome-editing events in hESCs and their matured cell derivatives, we performed deep sequencing of cells edited by one-pot transcribed sgRNAs. We first generated one-pot transcribed sgRNAs targeting seven genes that mark pluripotent, ectodermal, mesodermal and endodermal cells. These one-pot sgRNAs were electroporated into a HUES8 hESC line with an inducible Cas9 transgene[144]. The electroporated hESCs were cultured for several passages and then matured into embryoid bodies (EBs) for five days to allow cells to differentiate into all three germ layers. mRNA was extracted from the EBs and reversed transcribed into cDNA. PCR using primers flanking each of sgRNA target sites was

performed and prepared for sequencing via Illumina Hi-Seq. We found 20% to 92% of reads overlapping the target sites contained at least one insertion or deletion (indel) within a ~100-nucleotide window around the expected cut site (**Figure 2.2A**). Consistent with NHEJ repair at the expected cut site, both frameshift and in-frame indels were observed for all of these loci (**Figure 2.2A,** grey, orange). We also observed that many more reads with an indel contained a deletion event (85%) than contained an insertion event (15%) (**Figure 2.2B**). Similar observations were reported in deep sequencing analysis of human cells edited by *S. pyogenes* Cas9 [135,149].

The sequencing results allowed quantitative analysis of observed indel mutations and their spatial distribution in the target region. The results in Figure 2C show the frequency of indels in the endodermal markers, *CDH20* and *FOXP2*, and an ectodermal marker *PAX6*. In EBs derived from edited cells, we found the highest frequency of indels three to four nucleotides upstream from the PAM sequence **(Figure 2.2C)**, consistent with reports of type II CRISPR systems. Taken together, one-pot sgRNAs, when combined with Cas9, can generate targeted genomic edits in hESCs that can be expressed in differentiated cells.

Patterning adhesive microfeatures to separate gene-edited hESCs

In the prior experiments, all gene-edited cells within standard cell culture were interspersed with wild type cells, so colonies would need to be selected, dissociated and subcloned to isolate gene-edited cells for subsequent culture and analysis. Such a mixture can be easily visualized in the *MCHERRY*-edited cells **(Figure A-1H)**. To overcome laborious downstream clonal selection steps in the editing workflow, we designed our ArrayEdit platform to separate edited cells by exploiting microcontact printing (μCP) on the surface of multiwell plates**.** μCP was performed on gold-coated glass[148] to create surfaces within standard culture multiwell plates that contained greater than 400 circular μFeatures of 300 μm diameter per well,

allowing for the spatially controlled growth of up to 2,400 separate gene-edited cells per standard 6-well tissue culture plate (**Figure A-2A**). The poly(ethylene glycol) (PEG) brush surface layer after μCP does not contain defects common to other stamped PEG surfaces, and μFeatures are stable upon extended culture, for over 30 days[150].

ArrayEdit enabled the facile isolation of living, gene-edited hESCs. We electroporated one-pot transcribed sgRNAs against mCherry into our WA09-H2B-mCherry labeled line along with a plasmid encoding Cas9. Cells were then seeded at clonal density on ArrayEdit according to a Poisson distribution, such that there would be high probability for 0 or 1 cell to be within each μFeature. Four days post transfection, it was trivial to identify WA09-H2B-mCherry clones of interest via fluorescent microscopy (**Figure 2.3A**). We then randomly selected four clones that lost fluorescence and transferred them to separate wells of a 24-well plate. After five days of subsequent culture, genomic DNA was harvested and Sanger-sequenced across the expected sgRNA target site. The sequences revealed indel mutations at the desired target causing loss of mCherry fluorescence (**Figure 2.3B**). Interestingly, two of the clones isolated possessed the same modification suggesting the presence of local DNA microhomology influencing DNA repair pathways[151].

HCA to identify properly edited hESCs

To enable marker-less identification of gene-edited cells, we developed an automated high-throughput, HCA within ArrayEdit (**Figure 2.3C**). Twenty-four hours post-seeding and each subsequent 24 hours, fluorescence microscopy was used to individually image each μFeature. Due to the array-based format and spatial control of the μFeatures, daily images can be used to create a time-lapse image of cell number within each μFeature (**Figure 2.3D**). An automated analysis pipeline to assess cell number was created in CellProfiler[152], a software package for image analysis. From the image, the number of nuclei per image was identified (**Figure A-2B**)

and written to a database for further analysis (**Code A-1**). Using high-throughput computing, ~15,000 images can be analyzed in the span of a few hours. Data from individual µFeatures over multiple days were joined together to provide the growth rate for cells within each µFeature (see example in **Figures 2.3E, A-2C and A-2D**).

We sought to test the speed and efficiency of ArrayEdit against prior work on *LAMA5*, because specific domains of *LAMA5* gene-edits are predicted to render a selectable phenotype in several culture conditions[153]. α-5 laminin, encoded by *LAMA5*, is an extracellular matrix protein recently identified as an autocrine/paracrine factor regulating self-renewal of hESCs[153]. Isolating growth-deficient *LAMA5* mutants using standard gene editing to study the function of this factor is difficult in standard cultures, as they are quickly outcompeted by unwanted, wild type cells. In prior work[153], through FACS followed by laborious subcloning of 50 gene-edited colonies, we isolated only a few lines edited at *LAMA5*.

On ArrayEdit, three sgRNAs were designed that targeted consecutive exons in the globular domains of the integrin binding 3' region of *LAMA5* (**Figure 2.4A**). One-pot transcribed sgRNAs were electroporated as a pool such that any cell that receives separate two sgRNAs can experience a deletion on the order of 100s of base pairs via NHEJ repair, which can disrupt *LAMA5* function. Exploiting HCA over a period of six days, we tracked 480 potential gene-edited *LAMA5* clones and observed clones that expanded rapidly as well as clones that lagged behind (**Table A-3, Figure 2.3E**). Many clones followed an exponential growth model of proliferation, suggesting that they had not been edited and maintained a wild type phenotype. However, there was another population of clones exhibiting a non-standard growth phenotype, suggesting that increased rates of apoptosis or decreased rates of self-renewal may be due to *LAMA5* edits. We manually separated the clones into three different categories: low, intermediate, and high growth (**Figures 2.3E, A-3A, and A-3B**). Comparison of cell number per µFeature as well as doubling times between the high and low growth population revealed a significant difference

between the populations (Student's two-tailed t-test, $p<5x10^{-5}$) despite large variations in the slow growth population (**Figure A-3B and A-3C**). The exact µFeature on ArrayEdit corresponding to the growth profile was identified using HCA, and we were able to easily pick clones of interest for expansion and further analysis. Twelve clones were isolated and expanded from each of the high and intermediate growth populations, and subsequently subjected to Sanger sequencing. As expected, all of the high growth clones maintained a wild type genotype at all three sgRNA cut sites (**Figure A-3D and A-3E**, data not shown). All of the intermediate growth clones also sequenced correctly at all three loci, and expanded in a manner similar to high growth populations after isolation, suggesting that there is some modest post-transfection transient variability in growth rate of clones (**Figure A-3A and A-3E**). Therefore, we proceeded to focus our analysis on clones from the low growth population.

We isolated, expanded and genotyped low-growth clones identified by HCA on ArrayEdit for further characterization. The deletions generated by the sgRNA pool within hESC lines can easily be resolved on a standard agarose gel (**Figure 2.4B**). Agarose gels are intended as a semi-quantitative measure that can quickly screen clones to identify alleles that have undergone large-scale deletions. Due to differential repair pathways in edited cells such as microhomology mediated end-joining [151], this quick assay may contain variable band sizes, and products require follow-on sequencing to confirm that the exact modification has occurred. Regardless, this gel assay was used to reveal cells that have been cut by two sgRNAs, producing either single or biallelic deletions. Agarose gels of isolated clones revealed 82% of clones had at least one large-scale deletion in the *LAMA5* allele (**Figs. 2.4C, 2.4D, and A-3E**). Of all selected clones, 46% contained a single allele edit, and 36% contained a biallelic edit, although not necessarily homozygous. The remaining 18% harbored wild type genotypes, as determined by agarose gel screening assay. For the three sgRNAs transfected, three different large-scale deletions are predicted, all of which were observed (**Figs. A-4A and A-4B**) suggesting that all sgRNAs within

ArrayEdit were effective in cleaving genomic targets at similar rates. These results demonstrate significant improvement over prior methods used to obtain *LAMA5* gene-edited hESCs, where sequencing over 50 clones isolated via standard FACS methods yielded only one single allelic mutant and no biallelic modifications[153] (**Figure 2.4D**).

On ArrayEdit, we selected five clones that exhibited a biallelic mutation pattern on the agarose screening gel for further analysis via sequencing. Of these five lines, three clones were found to contain near-precise, homozygous deletions between two sgRNAs (**Figure 2.4E**). The remaining two clones contained complex deletions around the sgRNA targets (**Figure 2.4F**). These lines were further subjected to off-target analysis as they are the most likely to have contained functional Cas9-sgRNA complexes and are therefore most likely to have experienced off-target activity. Off-target analysis on all five selected cell lines was performed by Sanger sequencing at three loci predicted by bioinformatics analysis[154] to be the most likely off-target sites. This methodology does not preclude the possibility that there are other, unpredicted, gene edits. However, whole-genome sequencing would be required to find these modifications and was not pursued in this work. Sequencing at the most likely off-target loci revealed perfect alignment with the reference genotype in all five isolated biallelic clones, indicating that ArrayEdit can be used to generate targeted edits with minimal off-target effects (**Figures 2.4G and A-4C**).

Phenotypic characterization of *LAMA5*-edited hESCs

The five selected biallelic *LAMA5* gene-edited hESC lines were subsequently cultured on commonly used culture substrates: matrigel, and laminin-111, both of which supply a low-level of exogenous α-5 laminin but are insufficient to rescue the complete growth phenotype[153]. Importantly, all of the isolated clones expressed high levels of pluripotency markers, indicating that use of ArrayEdit does not lead to differentiation (**Figure A-5**).

After three days of culture, all five biallelic-edited clones had significantly less cell number than the wild type cells (n=4, Student's two-tailed t-test, $p<0.05$) (**Figures 2.5A and 2.5C**) on both matrigel and laminin-111 substrates. All five clones had increased levels of apoptosis on laminin-111 when compared to wild type cells, while four out of the five clones had higher levels of apoptosis on matrigel when compared to wild type (**Figures 2.5B and 2.5D**). Comparison of specific mutation to cell growth rate revealed that the clone with the earliest predicted stop codon in α-5 laminin, "G10," with complex frame-shift edits from exon 66 onwards, had the least number of cells (**Figure 2.5C**) and among the highest levels of apoptosis (**Figure 2.5D**), while the clone with the latest predicted stop codon in α-5 laminin, "C8," with only an in-frame deletion between exons 67 and 68, had levels of apoptosis similar to wild type levels on matrigel (**Figure 2.5D**). The longer α-5 laminin produced by clone C8 may come together with an important component of matrigel to rescue apoptotic cells.

We then cultured all five *LAMA5* gene-edited lines on laminin-521, which has been proven to rescue growth phenotype in *LAMA5* knockouts[153], to show that growth phenotype differences are a direct cause of the biallelic knockout. After three days in culture, four of the five lines had similar cell numbers to the wild type, while one, "C8," had a larger cell number (**Figure 2.5C**). Four of five clones also had similar levels of apoptosis as measured by cleaved caspase-3 percentage as the wildtype cells. Clone "G10," with its complex frame shift and early stop codon, had a slightly elevated percentage of apoptotic cells (**Figure 2.5D**).

Overall, these results indicate that *LAMA5* gene-edited cells on ArrayEdit exhibit the expected phenotype[153] - involving decreased rates of proliferation, high rates of apoptosis (**Figure 2.5C, 2.5D**), or a combination of both - on both matrigel and laminin-111. This phenotype was fully rescued by supplying a source of exogenous laminin-521, supporting the conclusion that precise edits were the only cause of the phenotype. ArrayEdit for *LAMA5* took approximately fifteen days from conception and design to clonal expansion, which is

substantially faster than the state-of-the-art methods that can take upwards of a month **(Figure 2.5E)**.

Discussion

Here we tested whether a combination of improvements to the gene-editing workflow for hESCs would increase the efficiency and throughput of the overall process. One-pot transcription combined with µCP plates and HCA increased the multiplexing of edits within a single well or multiwell plate (3 sgRNAs with 3 sgRNA combinations), increased the efficiency of isolating edited clones to 82%, and reduced the time necessary for complete gene-editing workflow (15 days versus >30 days). To our knowledge, no other platform has combined these capabilities to gene-edit cell lines. This combination of capabilities is most powerful when edited cells have phenotypes that can be distinguished with HCA. Lists of genes and sgRNA targets are now being assembled across pooled growth selection screens[155], and customized cells edited at many of these targets could be readily generated on ArrayEdit.

ArrayEdit is capable of identifying potentially edited cells by HCA in a non-destructive manner without single-cell dissociation or added transgenes, in contrast to FACS and sequencing based methods to identify edited cells. This capability permits longitudinal, temporal tracking of phenotypes within cells, with subcellular resolution, and avoids the need for subcloning which typically adds at least 3-7 days to the workflow. ArrayEdit thus significantly reduces the culture and passaging required to isolate edited cells, enabling gene editing in human cell types that undergo substantial phenotypic changes with prolonged culture or primary cells that can only be cultured for few passages prior to senescence. Additionally, ArrayEdit allows for the continuous culture of mutants that experience a deficit in self-renewal due to gene-edits and would be out competed in standard culture. We found that 36% of our edited clones harbored biallelic editing of *LAMA5*, which is comparable to other CRISPR-Cas9 studies

in human pluripotent stem cells[134,135,144,156]. Efficiencies of editing on ArrayEdit would likely increase with advances in delivery modalities of DNA, RNA and protein[156–158] and application of additional selection pressure, such as drug selection[142,156]. The 82% efficiency that we demonstrate on ArrayEdit is below the 95-100% purity of isolating edited cell lines by FACS frequently followed by drug selection, but our platform facilitates the identification of phenotypes by HCA– phenotypes that are very difficult or impossible to track with FACS. Further, sorting can cause cellular stress, increasing cell death, which is especially challenging when isolating edited cells with a growth defect.

While we only implemented ArrayEdit for fluorescence loss and growth rate differences, the method employing HCA is extendable in principle to any image-based phenotype. Phenotypes could be defined by changes in uptake of cytoxicity dyes, live immunocytochemistry for cell surface or ECM markers, calcium flux dyes or mitochondrial functional dyes[159]. A distinguishable phenotype in edited cells may not be readily apparent by HCA in the pluripotent stem cell state. Hence, differentiation on ArrayEdit may be required to distinguish edited phenotypes. These capabilities on ArrayEdit seem possible, as arrayed neural organoid culture has already been achieved on μCP plates[160]. Further, more sophisticated computational methods could be easily implemented in our analysis pipeline in CellProfiler to prospectively identify imaging phenotypes that connect to proper or abnormal biological and epigenomic characteristics of edited cells[161]. Because all edited clones share the same culture media, the intra-well benchmarking of phenotypes on ArrayEdit could enable identification of phenotypes that may be lost due to noise or fluctuations in media composition among culture wells or plates. Variations in signaling factors in the media could also lead to hESC clones that become more or less lineage committed [162], opening a window to study the biological variability within clones that are difficult to ascertain during standard genotyping of clones. The stringent definition of

phenotypes enabled by HCA with ArrayEdit will likely permit a more thorough characterization of the biological and functional consequences of various gene-editing protocols.

Current limitations of ArrayEdit arise from setting up the platform and screening for phenotypes in pluripotent cells. Although μCP is a straightforward technique, ArrayEdit is not a turn-key ready platform for many traditional biology or industrial labs who may need access to laser cutting or automated microscopy units. However, many commercial HCA instruments are available on the market, and our HCA pipeline can be readily employed using standard cloud-based computing or even a personal computer. The simple, versatile, and well-characterized μCP chemistry requires only standard laboratory equipment. The chemistry could also be flexibly modified to create various hydrophilic and hydrophobic areas on a single surface[150,163], even "well-of-the-well," water-in-oil culture platforms that are routinely used in pre-implantation embryo culture. Such chemically-defined surfaces may be particularly attractive for clinical application in future work.

One-pot transcription of sgRNAs from PCR amplicons generated by oligonucleotide DNA primers produced clean and functional sgRNAs. One-pot transcribed sgRNAs can be rapidly designed and made from commercial vendors overnight with costs scaling with 20-60 base pair synthesis, which are anticipated to decrease over time (currently <$1 USD per sgRNA per experiment, see **Table A-4**). In contrast to other methods that require purchasing multiple oligonucleotides, our method requires only one unique oligonucleotide that can be synthesized in a multiwell plate format by commercial vendors, decreasing the setup time and the possibility of pipetting error. Errors in long (>60 nt), chemically synthesized oligonucleotides (up to 10%) have been observed[145], and our method notably avoids the use of long oligonucleotides by using a sequence-verified, synthesized double stranded DNA for the long universal region of the sgRNA. Our modular design also permits facile incorporation of additional RNA elements and devices[20,164]. Further, our method performed better than several previously described

methods[135,144] (**Figure 2.1D-E**) and generated sgRNAs in less than two days. When analyzed via deep sequencing, one-pot sgRNAs had an efficiency of editing between 20%-92% of mRNA transcripts after EB differentiation. Interestingly, many of the transcripts analyzed (5 of 7) had a higher percentage of in-frame mutations than would be caused by random chance [33%,[128]]. This may suggest that there were selection pressures in the EB cultures that modify the mutation spectrum observed.

We observed no off-target mutations in our edited cell lines at eight bioinformatically predicted sites by using one-pot sgRNAs, although we cannot rule out the presence of mutations at other loci. Because optimized Cas9 mRNA and protein delivery can reduce off-target mutagenesis[165], one-pot transcribed sgRNAs could be combined with these methods of delivering Cas9 to reduce the risk of off-target mutagenesis.

Overall, ArrayEdit provides a window into the editing process that could be useful in refining the specificity and versatility of CRISPR-Cas9 gene-editing techniques. *In situ* HCA could incorporate new optical and imaging techniques that monitor and perturb CRISPR-Cas9 editing. Inducible[166,167] and optical control[168,169] of Cas9 activity has been demonstrated, and various spatial and temporal perturbations could be tested in future work on ArrayEdit. New nucleic acid probes could also visualize, or "paint," specific loci of cells[147,170] within complex cellular structures on ArrayEdit, permitting the specific editing and isolation of edited clones with enhanced precision. ArrayEdit is fully compatible with existing screening platforms, so that small molecule or other biologic screens could be tested to enhance the efficacy of any desired gene-editing protocol. Also, ArrayEdit could be easily adapted to isolate cells with appropriate phenotypes after application of engineered, "nuclease-dead" Cas9 protein fusions, such as those designed to activate or repress gene transcription. Finally, ArrayEdit will be likely applicable to many other human cell types, providing an attractive route to generating gene-edited human cells for a variety of industrial and preclinical purposes.

Materials and Methods

One-pot transcription of sgRNAs: *In vitro* transcribed (IVT) sgRNAs were synthesized in parallel in a 96-well plate within one day. The first step employed PCR with two chemically synthesized primers and a 125 bp double stranded DNA template (Integrated DNA Technologies; see Figure S1A). Forward primers were ordered in 96-well plate format to enable high-throughput synthesis. PCR was performed using Phusion High-Fidelity Polymerase (New England Biolabs) according to manufacturer protocols and was placed in the thermocycler at 98^{o}C for 30 s followed by 35 cycles of: 98^{o}C for 5 s, 52^{o}C for 10 s and 72^{o}C for 15 s before a final extension period of 72^{o}C for 10 min. A truncated T7 promoter was included in the forward primer, which allowed the transcription of sgRNAs via a 37^{o}C overnight reaction with a MEGAshortscript kit (Life Technologies).

High-content image acquisition and analysis: Automated microscopy was performed using a Nikon Eclipse TI epifluorescent microscope and NIS Elements Advanced Research (V4.30) software. The ND acquisition 6D module was used to establish a 20 x 20 grid pattern such that one 10x image was taken at each µFeature and combined in a single file. Nikon Perfect Focus was used to ensure that all images were in the same Z-plane and in focus. Each image was then corrected for illumination defects using CellProfiler [152] and the number of nuclei was determined (Figure S2). Analysis was performed in a massively parallel manner using the Center for Throughput Computing (UW-Madison) and results were written to a local MySQL database. MySQL Workbench 6.1 CE was used to retrieve data and join tables from time points on the basis of well and position.

RNA quantification and characterization: For quantification and characterization of sgRNAs, purification was performed with the MEGAclear transcription kit (Life Technology) and quantified on a Nanodrop 2000. An Aligent 2100 Bioanalyzer was used according to manufacturer protocols to determine the uniformity of the produced sgRNA. Due to consistent yields post-purification, approximate molar quantities were used for transfections based off of sgRNA volume used.

Cell culture: WA09 hESCs (WiCell, Madison, WI) and HUES9-iCas9 (generously provided by the lab of Dr. Danwei Huangfu, Sloan-Kettering Institute, New York, NY) were maintained in E8 medium on Matrigel (WiCell) coated tissue culture polystyrene (TCPS) plates (BD Falcon) [171]. Cells were passaged every 3-4 days at a 1:6 ratio using Versene solution (Life Technologies). For experimental procedures on µCP plates, cells were singularized using Accutase (StemPro) and seeded on to plates in mTESR1 medium (WiCell) containing 10 µM ROCK inhibitor (Y27632, Selleckchem).

Human embryonic kidney cells (293T) were maintained between passage 15-40 in Growth medium containing DMEM (Life Technologies), 10% v/v FBS (Life Technologies), 2mM L-Glutamine (Life Technologies), and 50 U/mL Penicillin-Streptomycin (Life Technologies). Cells were passaged 1:40 with Trypsin-EDTA (Life Technologies) onto Gelatin-A (Sigma) coated plates.

Cell culture substrate coatings: 24-well tissue cultue polystyrene plates (TCPS) were coated with matrigel, laminin-111, or laminin-521. Matrigel (WiCell) diluted in DMEM-F12 (Life Technologies) and coated onto TCPS at a density of 8.7 µg/cm^2 overnight at 37^{o}C. Laminin-111 (Cultrex) was diluted in PBS and coated onto poly-ornithine-coated TCPS well at a density of 1

μg/cm^2 then incubated overnight at 37^{o}C. Laminin-521 (BioLamina) was diluted in 1x DBPS with Ca^{2+} and Mg^{2+} and coated onto poly-ornitihine-coated TCPS well at a density of 1.5 μg/cm^2 then incubated at 37^{o}C for 2 hours.

Nucleic acid delivery to human cell lines: Nucleic acids were electroporated into two hESC lines: one harboring a mCherry transgene, WA09-H2B-mCherry (Figure S1) and another harboring an inducible Cas9 transgene, HUES9-iCas9. Doxycycline was added to inducible Cas9 line at 5 μM 48 hours prior to experiments and both lines were treated with 10 μM ROCK inhibitor (Y27632, Selleckchem) 24 hours prior to experiments. Electroporation was performed using a GenePulser Xcell, with attached CE module (BioRad). Cells were first singularized using Accutase (StemPro), then 10^6 cells were centrifuged and resuspended in 400 μL mTESR1 medium containing 5 μg of each sgRNA, and 2.5 μg hCas9 plasmid (Addgene, 41815) for cell lines that did not harbor inducible Cas9 transgenes. Electroporation was performed with an exponential decay waveform, 250V, and mTESR1 medium, counted, and subsequently plated at a density of 2000 cells/cm^2 on ArrayEdit in mTESR1 medium (WiCell) containing 10μM ROCK inhibitor (Y27632, Selleckchem). Cells remained undisturbed and allowed to attach for 48 hours followed by daily media changes.

Human embryonic kidney cells were plated at 50,000 cells/well in 24 well plate (~25,000 cells/cm^2) at day 0. On day 1, 50 μL of Transfection media was added to cells in 450 μL of freshly exchanged Growth media. Transfection media contained 50 μL Opti-MEM (Life Technologies), 1.25 μg Cas9 plasmid, and 0.75 μL of Lipofectamine 2000 (Life Technologies). On day 2 50 μL of Transfection media was added to cells in 450 μL of freshly exchanged Growth media. Transfection media contained 100 μL Opti-MEM (Life Technologies), 0.75 μL of Lipofectamine 2000 (Life Technologies), and either 1.25 μg one-pot sgRNA. 1.25 μg plasmid-based IVT, 0.5 μg sgRNA plasmid, or 3 nM U6-gBlock [Integrated DNA Technologies[172]]. Cells

were cultured in Growth media plus Transfection media for 4 additional days before downstream analysis.

Embryoid body differentiation: Cells were cultured created using the Aggrewell (Stem Cell Technologies) system per the manufacturer's protocol. EBs were cultured in non-adherent cell culture plates (Corning Inc.) suspended in 4 mL per well of "EB Medium" consisting of 80% Iscove's Modified Dulbecco's Medium (IMDM, Life Technologies) with 20% Fetal Bovine Serum (Life Technologies). After four days in suspension, RNA was harvested from aggregates using SV Total RNA Isolation System (Promega) according to manufacturer instructions.

Deep sequencing library preparation: RNA was reverse transcribed using SuperScript III (Life Technologies). 30 ng of each RNA extraction were combined with random hexamer primers and water before incubating for 5 minutes at 65 ^{o}C. SuperScript III enzyme mix was then added and the reaction was thermocycled for 5 minutes at 25 ^{o}C , 50 minutes at 50 ^{o}C followed by termination of the reaction at 85 ^{o}C for 5 minutes. DNA was then amplified using gene specific primers with Illumina adapters using Taq AccuPrime HiFi polymerase as described in Genomic Analysis. Genomic PCRs were then purified using Wizard SV Gel and PCR Clean-Up System (Promega) and quantified using a Nanodrop2000. Samples were then pooled and ran on an Illumina HiSeq 2000 at a run length of 1x100bp.

Deep sequencing data analysis: A custom python script was developed to perform sequence analysis. The pipeline starts with preprocessing, which consists of filtering out low quality sequences and finding the defined ends of the reads. For each sample, sequences with frequency of less than 100 were filtered from the data. Sequences in which the reads matched

with primer and reverse complement subsequences classified as “target sequences”. Target sequences were aligned with corresponding wildtype sequence using global pairwise sequence alignment from “Biopython” package (http://biopython.org). Each aligned sequence was classified as follows. If there is no insertion or deletion in the aligned sequence, it classified as the wild type otherwise it considered a mutant. If the mutant sequence preserves the reading frame, the sequence is called in-frame mutant; otherwise, it is a frame-shift mutant. This sequencing assay would not lose information if a premature STOP codon is produced by the editing, but rather can detect these edited transcripts. The frequency, length and position of matches, insertions, deletions and mismatches were identified in the resulting aligned sequences. The frequency and position of in-frame shifts were also extracted and compared in different samples. The custom python script is available upon request.

Microcontact-printed (μCP) well plate fabrication: μCP was performed using previously described methods[148]. The surface modification involved printing of an alkanethiol initiator to nucleate the polymerization of hydrophilic poly(ethylene glycol) (PEG) chains (Figure S2A). Briefly, double sided-adhesive was attached to the bottom of a standard tissue culture plate, after which a laser cutter was used to cut out the well bottoms. Glass sheets were purchased at a size slightly smaller than a well plate. A metal evaporator was then used to deposit a thin layer of titanium, followed by a layer of gold onto one side of the glass sheet. Using previously described chemistry[150], patterns were transferred to gold-coated glass via a polydimethylsiloxane stamp after which the glass was submerged in a poly(ethylene glycol) (PEG) solution overnight to build hydrophillic PEG chains surrounding μFeatures. After submersion, sheets were washed with deionized water to remove residual copper deposited by the reaction and 70% ethanol to sterilize. Standard tissue culture plates with well bottoms cut out were then fastened to processed sheets using a custom-made alignment device.

Genomic analysis: DNA was isolated from cells using QuickExtract (Epicentre) following treatment by 0.05% trypsin-EDTA and centrifugation. QuickExtract solution was incubated at 65^{o}C for 15 minutes, 68^{o}C for 15 minutes, and finally 98^{o}C for 10 minutes. Genomic PCR was performed following manufacturer's instructions using AccuPrime HiFi Taq (Life Technologies) and 500 ng of genomic DNA. Products were then prepared for sequencing by purifying the PCR product using the QIAquick PCR purification kit (Qiagen) and quantified using a Nanodrop2000. PCR amplicons were submitted to the UW Biotechnology Center for Sanger Sequencing, and the sequencing chromatograms were analyzed in Benchling. For the *LAMA5* sgRNAs used in this study, all off-target sites scoring 1.0 and greater on a 100 point scale were assayed. The remaining sites dropped off in score significantly and are predicted to be modified at even lower probabilities.

Flow cytometry: Flow cytometry was performed using a C6 Accuri (BD) and analyzed using BD CSampler Software. mCherry fluorescence was detected using the 610/20 filter in the FL3 position. Gates were established by running wild type WA09 hESCs. Data in Figure 1 are for two independent experimental replicates (n=2). For the Plasmid+Sort experiments, hESCs were co-electroporated with a GFP plasmid (Addgene #6085-1) in addition to plasmids encoding sgRNAs and Cas9, and then sorted for GFP+ cells using an Aria III sorter (BD Biosciences) 1 day after electroporation. The level of apoptosis was detected by trypsinization followed by fixation in 4% paraformaldehyde (IBI Scientific) at room temperature for 15 min. Cells were resuspended with 0.5% Triton-X 100 (Electron Microscopy Sciences) and incubated at room temperature for 30 min. Samples were then resuspended in 5% goat serum and a primary antibody against cleaved caspase-3 (1:400 Cell Signaling Technologies 9661S) and placed at 4 ^{o}C overnight. The next day, cells were centrifuged and resuspended in 5% goat serum plus a

goat anti-rabbit secondary antibody conjugated to AlexaFluor488 (1:400 Santa Cruz Biotechnology sc-362262). Gates were established by running a secondary antibody only control.

Immunocytochemistry: Biallelic-edited hESC clones were cultured on matrigel for >10 passages and then fixed in 4% paraformaldehyde (IBI Scientific) at room temperature for 15 min. Cells were washed with PBS then permeablized with 0.5% Triton-X 100 (Electron Microscopy Sciences) by incubation at room temperature for 30 min. Primary antibodies against Oct4 conjugated to AlexaFluor594 (1:10 BD Pharmingen 560186), Nanog (1:200 R&D Systems AF1997), and TRA-1-60 (1:100, EMD Millipore MAB4360) were prepared in 5% donkey serum and incubated overnight at 4^{o}C. For primary antibodies without conjugated antibodies, samples were washed with 5% donkey serum then incubated in PerCP/Cy5.5 (406511) or AlexaFluor-647 (ThermoFisher A21238) anti-mouse or goat secondary antibodies diluted 1:400 in 5% goat serum at 4^{o}C for 1 hour. To detect apoptosis, a primary antibody against cleaved caspase-3 (1:250 Cell Signaling Technologies 9661S) was used as described above and detected using anti-rabbit, AlexaFluor-488 conjugated secondary antibody (ThermoFisher A11034). Cells were analyzed using fluorescent microscopy with a Nikon TiE microscope as described in the high-content imaging section.

Statistical analysis: All statistical tests were performed using Excel standard packages (Microsoft). Student's t-tests were performed using two-tailed, two-sample unpaired data with unequal variance and were deemed significant at $p<0.05$.

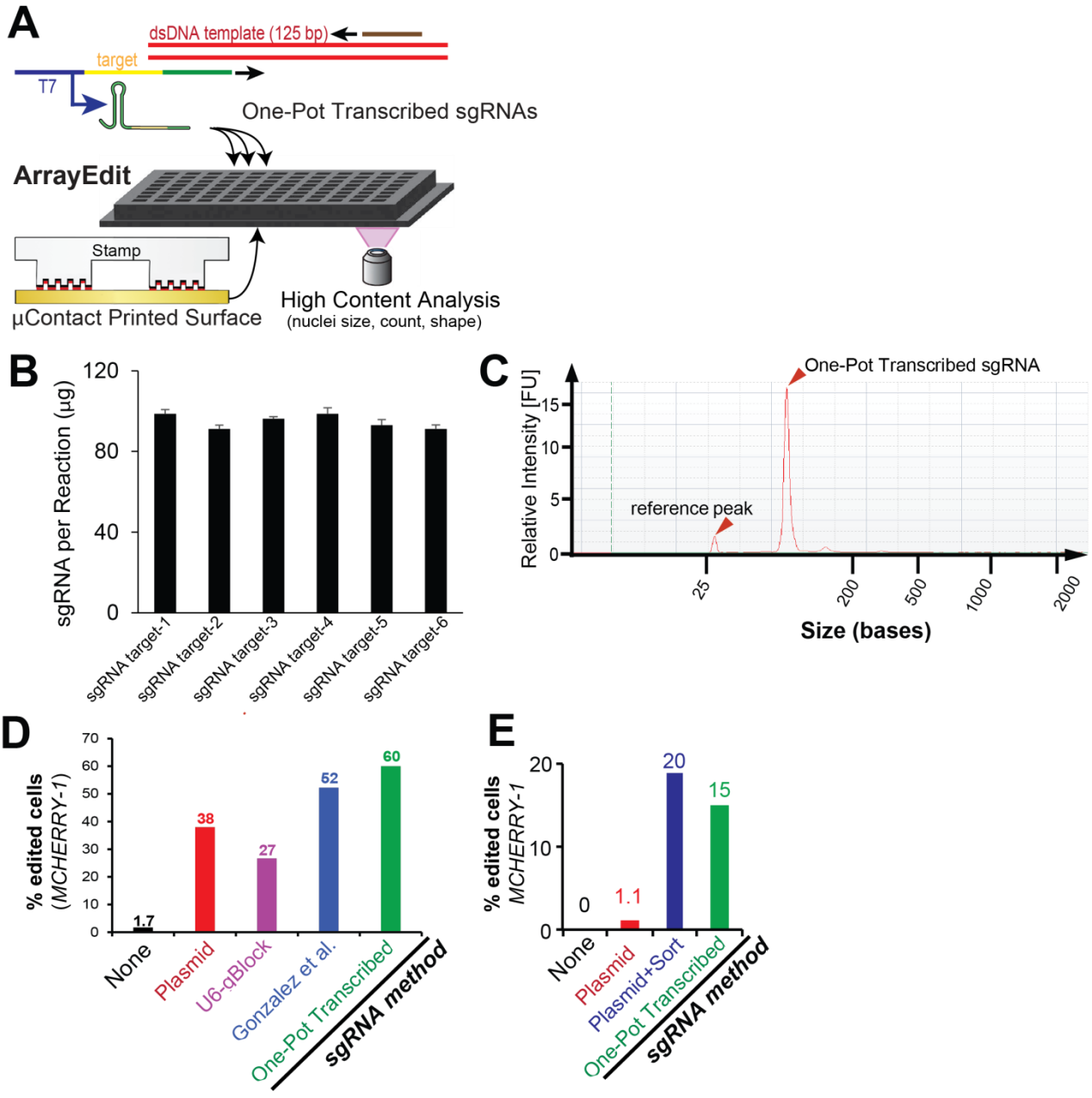

A
dsDNA template (125 bp)
target
T7
One-Pot Transcribed sgRNAs
ArrayEdit
Stamp
µContact Printed Surface
High Content Analysis
(nuclei size, count, shape)
B
sgRNA per Reaction (µg)
120
80
40
0
sgRNA target-1
sgRNA target-2
sgRNA target-3
sgRNA target-4
sgRNA target-5
sgRNA target-6
C
One-Pot Transcribed sgRNA
Relative Intensity [FU]
15
10
5
0
reference peak
25
200
500
1000
2000
Size (bases)
D
% edited cells
(MCHERRY-1)
70
60
50
40
30
20
10
0
1.7
38
27
52
60
None
Plasmid
U6-qBlock
Gonzalez et al.
One-Pot Transcribed
sgRNA method
E
% edited cells
MCHERRY-1
20
10
0
0
1.1
20
15
None
Plasmid
Plasmid+Sort
One-Pot Transcribed
sgRNA method

Figure 2.1 | ArrayEdit: an arrayed, high-content platform to monitor and isolate gene-edited cells via one-pot transcribed single-guide RNAs.

(A) Overview of ArrayEdit assembly and key components. Top: Schematic of “one-pot” PCR and T7 transcription. All components can be mixed and reacted within a single well without any intermediate purification steps. Primers are synthesized as custom oligonucleotides. The forward primer defines the genomic target of editing by Cas9. Bottom: Surface modification to the bottom of multiwell plates generates cell-adhesive μFeatures on a glass bottom. Each μFeature can be tracked over time via high-content imaging and stitched together to form a time-lapse visualization of edited cell phenotypes. **(B)** Amount of sgRNA produced within each well via one-pot transcription. Data are represented as mean ± 95% C.I. from four independent one-pot transcriptions on each sgRNA target (targets 1-3: MCHERRY-1-3; 4-6: GFP-1-3) and are not significantly different (Student’s t-test, $p>0.05$). **(C)** RNA bioanalyzer spectra of one-pot transcribed sgRNA. Narrow peak (arrowhead) is consistent with only the desired product being produced. Reference peak is used by the bioanalyzer to standardize size measurements. **(D-E)** Flow cytometry histograms of HEK-H2B-mCherry cells (D) and WA09-H2B-mCherry hESCs (E) four days after delivery of MCHERRY-1 sgRNAs. sgRNA was either expressed of a plasmid, ordered commercially (U6-gBlock), created using previously described methods (Gonzalez), or one-pot transcribed (n=2; independent experiments). Sorted: a GFP plasmid is co-electroporated and then sorted for GFP+ cells, leading to enrichment of cells that contained exogenous nucleic acids.

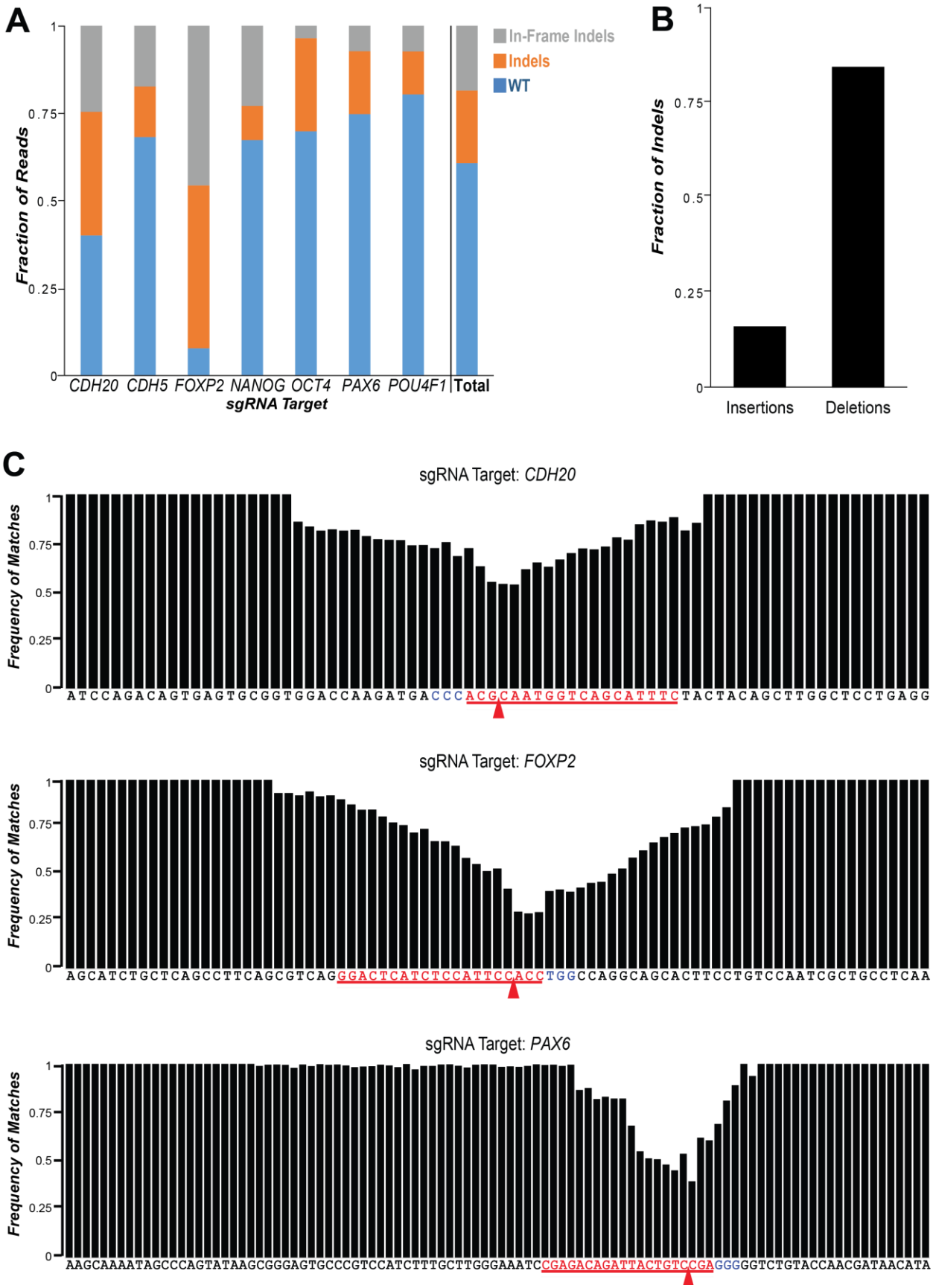

A
Fraction of Reads
1
0.75
0.5
0.25
0
In-Frame Indels
Indels
WT
CDH20
CDH5
FOXP2
NANOG
OCT4
PAX6
POU4F1
Total
sgRNA Target
B
Fraction of Indels
1
0.75
0.5
0.25
0
Insertions
Deletions
C
sgRNA Target: CDH20
Frequency of Matches
ATCCAGACAGTGAGTGCGGTGGACCAAGATGACCCACGCAATGGTCAGCATTTCTACTACAGCTTGGCTCCTGAGG
sgRNA Target: FOXP2
AGCATCTGCTCAGCCTTCAGCGTCAGGGACTCATCTCCATTCCACCTGGCCAGGCAGCACTTCCTGTCCAATCGCTGCCTCAA
sgRNA Target: PAX6
AAGCAAAATAGCCCAGTATAAGCGGGAGTGCCCGTCCATCTTTGCTTGGGAAATCCGAGACAGATTACTGTCCGAGGGGGTCTGTACCAACGATAACATA

Figure 2.2 | Deep sequencing reveals high-efficiency modification by one-pot transcribed sgRNAs.

(A) Fraction of sequencing reads from gene-edited, hESC-derived embryoid bodies that matched the wildtype sequence (blue), contained at least one insertion or deletion (indel, orange), or contained an in-frame indel (grey). Edits occurred in genes and transcripts marking all three germs layers [ectoderm (PAX6, POU4F1), mesoderm (CDH5), endoderm (CDH20, FOXP2)] and pluripotent stem cells (NANOG, OCT4). Total: all reads across the seven loci. **(B)** Frequency of insertion events to deletion events in the sequencing reads, with deletion being ~5x more common than insertion. **(C)** Per base frequency of matches to the wildtype sequence for three genes (CDH20, FOXP2, PAX6). Red bases denote sgRNA target sequence while blue bases denote the protospacer adjacent motif (PAM). Red arrows indicate the predicted site of double strand break formation by Cas9. The beginning and ending sequences (17-22 bp) for each gene are uniform, because they contain reads that were amplified using primers (17-22 bp) during PCR.

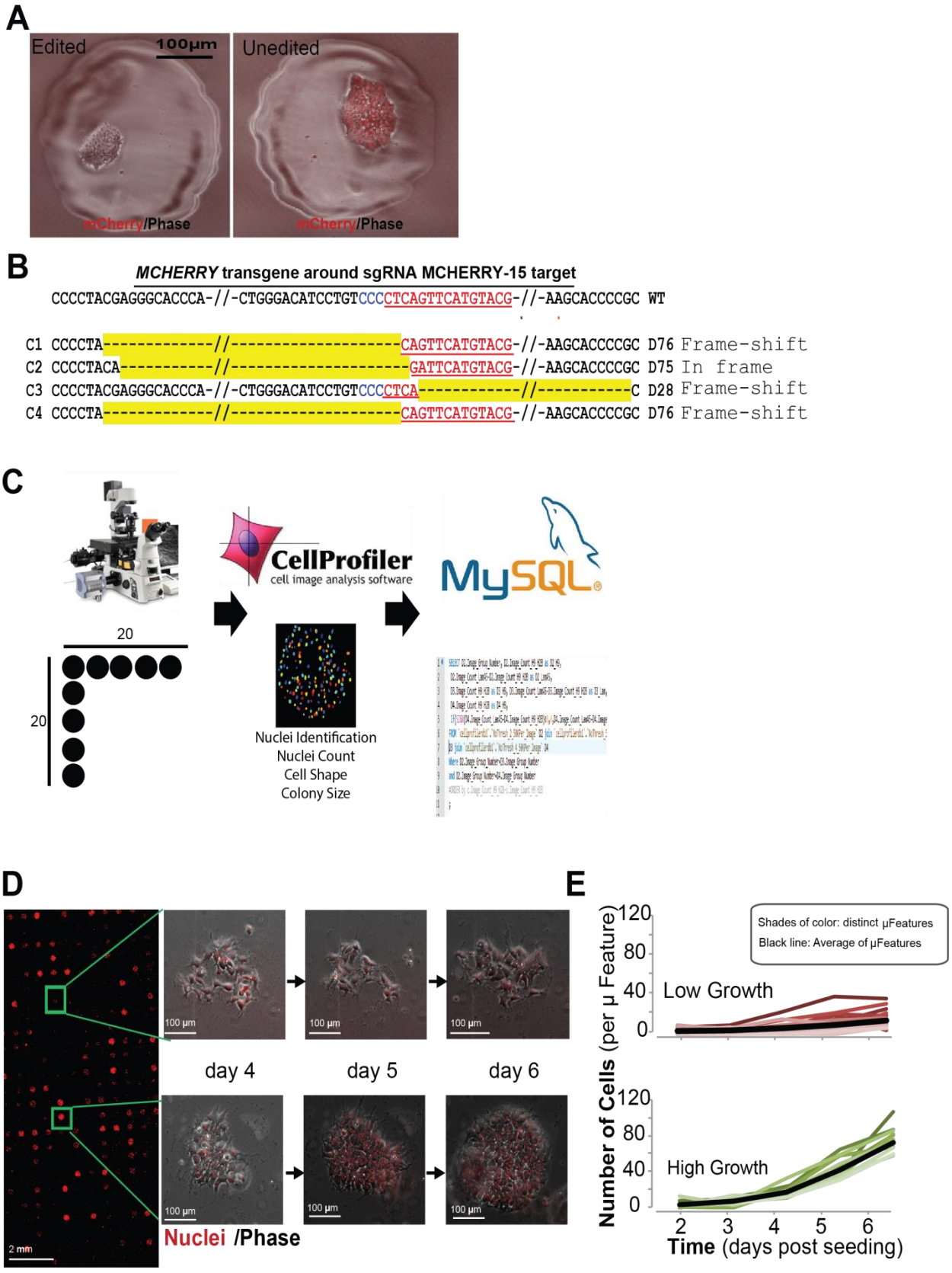
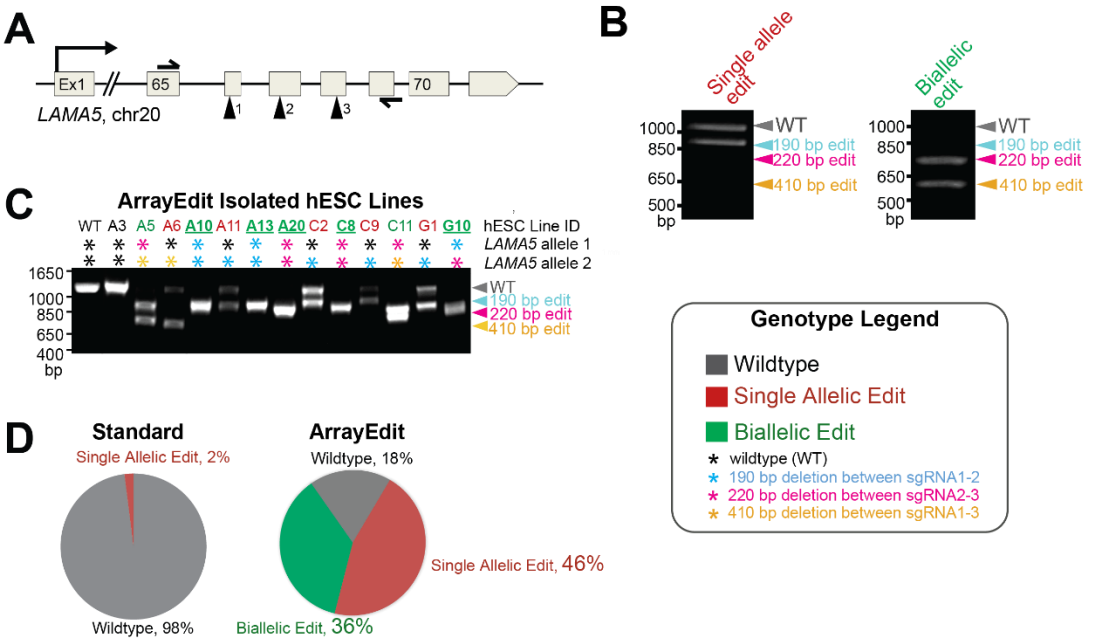

A
Edited
100µm
Unedited
mCherry/Phase
mCherry/Phase
B
MCHERRY transgene around sgRNA MCHERRY-15 target
CCCCTACGAGGGCACCCA-//-CTGGGACATCCTGTCCCCTCAGTTCATGTACG-//-AAGCACCCCGC WT
C1 CCCCTA----------//----------CAGTTCATGTACG-//-AAGCACCCCGC D76 Frame-shift
C2 CCCCTACA----------//----------GATTCATGTACG-//-AAGCACCCCGC D75 In frame
C3 CCCCTACGAGGGCACCCA-//-CTGGGACATCCTGTCCCCTCA----------//----------C D28 Frame-shift
C4 CCCCTA----------//----------CAGTTCATGTACG-//-AAGCACCCCGC D76 Frame-shift
C
20
20
CellProfiler
cell image analysis software
MySQL
Nuclei Identification
Nuclei Count
Cell Shape
Colony Size
D
100 µm
100 µm
100 µm
day 4
day 5
day 6
100 µm
100 µm
100 µm
2 mm
Nuclei /Phase
E
Shades of color: distinct µFeatures
Black line: Average of µFeatures
Number of Cells (per µ Feature)
120
80
40
0
Low Growth
120
80
40
0
High Growth
2
3
4
5
6
Time (days post seeding)

Figure 2.3 | **Substrate micropatterning enables live, HCA of gene editing.**

(A) Isolation of homogeneous mCherry gene-edited hESC lines on micropatterned plates. Clonal knockouts can be reliably identified and expanded in spatial isolation on each µFeature. **(B)** Sanger sequencing of mCherry edited clones isolated via ArrayEdit. sgRNA target is denoted in red and protospacer adjacent motif (PAM) in blue. Deletions are represented in yellow and the total length of deletion is to the right of the sequence (e.g., D76 indicates a deletion of 76 base pairs). **(C)** High-content image acquisition and analysis workflow. Images are taken in a 20x20 grid and are passed to CellProfiler. Different colors indicate distinct identified objects (nuclei). CellProfiler results are sent to a MySQL database. **(D)** Image of ArrayEdit within one standard culture well. Each µFeature can be tracked over time and stitched together to form a time-lapse visualization of edited cell phenotypes. Clones in two separate features are shown on days 4, 5, and 6. **(E)** Growth curves for cells within 24 µFeatures on ArrayEdit over 5 consecutive days after editing with LAMA5 sgRNAs. Curves were separated into high and low growth rate groups.



E

LAMA5 around sgRNA2 & sgRNA3 targets

```
GGCTGCAGTGGGACGGCCTTTTTCAGGCC-//-GTCCAGGTAGGAGCCGTCCGTGAGCCAC WT

C8  GGGATAACNNTTNGAAGACCTT--------//--------TAGGANCCGTTCTTGAACCAC D270 In frame
```

LAMA5 around sgRNA1 & sgRNA2 targets

```
GATGCGGGCGAAGCCGGTGCCGTCCAGGTAG-//-GAAGCAGGGGAGGGGGCTGTGGGCA WT

A13 GATTGTGGTAATTTANNTGTGGAAAGNNATTN//-GAAGCAGGGGAGGGGGCTGTGGGCA D232 Frame-shift
A20 GATGCGGGCGAAGCCGGTGCAGTACNGG----//---AGCAGGGGAGGGGGCTGTGGGCA D241 Frame-shift
```

F

LAMA5 around sgRNA2 target

```
ATGGCAGGACGGGTGTGCC-//-CCAGGTAGGAGCCGTCCGTGAGC-//-GCCACACGCACCGGGCA WT

A10 ATGGCAGGACGGGTGTGCC-//------------------------//--------GCTTCANGNA D237 Frame-shift
```

LAMA5 around sgRNA1 target

```
CCCTGTGCCCACAGCCCCCTCCCCTGCTTCGCTTCCCCGGCTACCGGGGCTGCATCG//ACCCCAGAATCTCCGACCTGCAGCTCC WT

G10 CCCTGTGCCCACAGCCCCCTCCCCTGCTTCGCTNATNAGGATANNGNGNCTNNATNC//NTGNTC--CCCNNNATGACCCTTCCGN Complex
```

G

Predicted top off-target site

```
GCACCGGGAGTGAGGCACTGCCTTTCCTTCGCTTCCCCGCCTTCCATCTCTTCGGAAAAG WT

C8  GCACCGGGAGTGAGGCACTGCCTTTCCTTCGCTTCCCCGCCTTCCATCTCTTCGGAAAAG
A10 GCACCGGGAGTGAGGCACTGCCTTTCCTTCGCTTCCCCGCCTTCCATCTCTTCGGAAAAG
A13 GCACCGGGAGTGAGGCACTGCCTTTCCTTCGCTTCCCCGCCTTCCATCTCTTCGGAAAAG
A20 GCACCGGGAGTGAGGCACTGCCTTTCCTTCGCTTCCCCGCCTTCCATCTCTTCGGAAAAG
G10 GCACCGGGAGTGAGGCACTGCCTTTCCTTCGCTTCCCCGCCTTCCATCTCTTCGGAAAAG
```

Figure 2.4 | Genetic characterization of hESC clones isolated from ArrayEdit.

(A) Schematic of CRISPR-targeted regions in *LAMA5* including primers used for genomic amplification. **(B)** Agarose gel of PCR products generated from amplification of genomic DNA isolated from an ArrayEdit gene-deleted clone. Gel indicates the expected deletions spanning hundreds of base pairs. Single allele mutants can be identified by the presence of a wildtype length band, while biallelic mutants are identified by the absence of a wildtype band. **(C)** Agarose gel of PCR products generated from amplification of genomic DNA isolated from ArrayEdit gene-edited hESC lines. ArrayEdit was implemented on hESCs with three sgRNAs targeting *LAMA5* as shown in part (A). Gel indicates the expected deletions spanning hundreds of base pairs. Red clone names denote single allele edited lines, while green denote biallelic modifications. Each allele is denoted by a colored asterisk corresponding to which sgRNA combination is presumed to have made the modification. Underlined clones are representative clones presented in the main text. **(D)** Summary of genotypes obtained from hESC lines isolated after gene-editing using ArrayEdit or standard procedures (see text). Efficiencies of generating edited cells are significantly higher on ArrayEdit. Genotypes are detailed in part (C). **(E)** Sanger sequencing analysis of representative biallelic-edited hESC lines chosen in part (C) that displayed gene editing between two sgRNAs. Wild type is denoted on top, and the biallelic-edited hESCs are below. Color codes are the same as in Figure 3. Non-faithful nucleotides in sequence alignment are in grey and are believed to be caused by 1 or 2 bp differences in alleles, causing misreads during sequencing. **(F)** Sanger sequencing analysis of representative biallelic-edited hESC lines chosen in part (C) that displayed unexpected deletions around 1 sgRNA site (A10) or potential differing modifications to both alleles (G10). **(G)** The sequence around the top potential off-target site for Cas9 with sgRNA1 is shown at the bottom. No modifications were observed in sequencing results from any of the edited hESC clones.

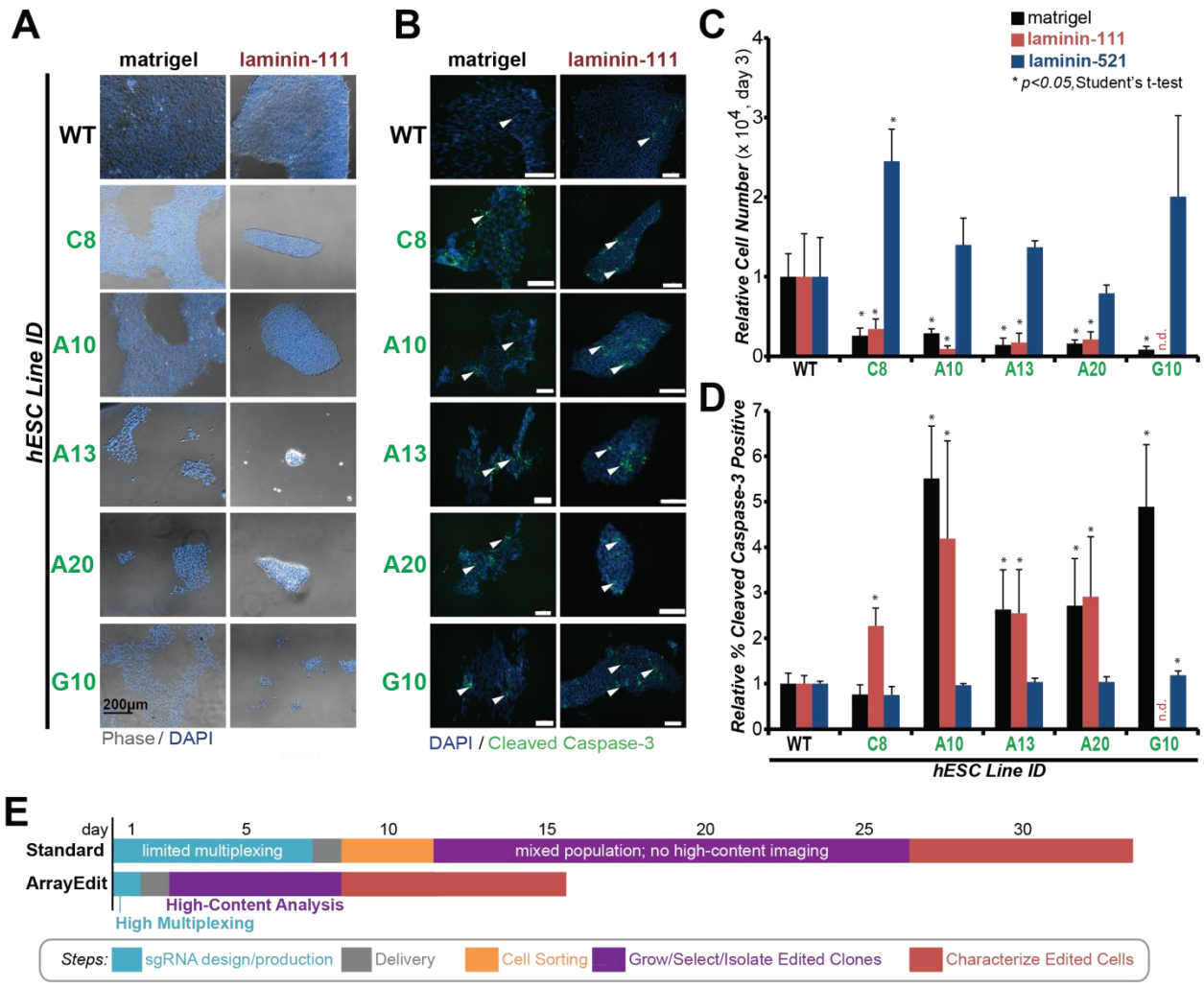

A
matrigel
laminin-111
hESC Line ID
WT
C8
A10
A13
A20
G10
200µm
Phase / DAPI
B
matrigel
laminin-111
DAPI / Cleaved Caspase-3
C
matrigel
laminin-111
laminin-521
* p<0.05, Student's t-test
Relative Cell Number (x 10^4, day 3)
n.d.
D
Relative % Cleaved Caspase-3 Positive
hESC Line ID
E
day
Standard
limited multiplexing
mixed population; no high-content imaging
ArrayEdit
High-Content Analysis
High Multiplexing
Steps:
sgRNA design/production
Delivery
Cell Sorting
Grow/Select/Isolate Edited Clones
Characterize Edited Cells

Figure 2.5 | ArrayEdit rapidly produces living, well-characterized and functional *LAMA5* gene-edited hESC lines.

(A) Representative images of hESC colonies formation of each biallelic-edited line on both matrigel- and laminin111-coated substrates. Laminin-111 contains less matrix proteins thereby accentuating self-renewal defects generated from loss of α-5 laminin protein domains in the biallelic-edited cells. Clone G10 notably displayed a drastic decrease in cell number on laminin-111. **(B)** Detection of apoptosis levels by immunocytochemistry of cleaved caspase-3. Each biallelic clone had detectable apoptotic cells (some are denoted by arrowheads). All scale bars are 100 μm. **(C-D)** Mean number of cells (C) and cleaved caspase-3 positive cells (D) for each hESC clone on matrigel-, laminin-111-, laminin-521-coated substrates after 4 days in culture on various substrates as determined by flow cytometry (n=4; independent experiments). On matrigel and laminin-111 substrates, each biallelic-edited line was less dense than wild-type cells (Student's two-tailed t-test, $p<0.05$), indicating a functional defect in self-renewal. On laminin-521 substrates each clone was at least as dense as wild-type cells, and had a similar number of apoptotic cells as wild-type except for G10. This may be due to complex deletions in the G10 clone. Clone G10 could not be detected ("n.d.") due to a low cell number on laminin-111. Clone C8 had an in-frame mutation that may be partially rescued on the matrigel substrate and had a significantly higher number of cells on laminin-521. **(E)** Comparison of ArrayEdit against standard methods to produce gene-edited cells. ArrayEdit is approximately 2x faster and produces significantly more gene-edited cell lines (see Figure 4D). ArrayEdit reduces or eliminates several steps in the gene-editing workflow and additionally has important multiplexing and HCA capabilities.

CHAPTER 3: Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing

Work in this chapter was adapted from:

Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing

Jared Carlson-Stevermer, Amr A. Abdeen, Lucille Kohlenberg, Madelyn Goedland, Kaivalya Molugu, Meng Lou, and Krishanu Saha

Nature Communications 8 (1) 1711, 2017

Abstract

Writing specific DNA sequences into the human genome is challenging with non-viral gene-editing reagents, since most of the edited sequences contain various imprecise insertions or deletions. We developed a modular RNA aptamer-streptavidin strategy, termed S1mplex, to complex CRISPR-Cas9 ribonucleoproteins with a nucleic acid donor template, as well as other biotinylated molecules such as quantum dots. In human cells, tailored S1mplexes increased the ratio of precisely-edited to imprecisely-edited alleles up to 18-fold higher than standard gene-editing methods, and enriched cell populations containing multiplexed precise edits up to 42-fold. These advances with versatile, preassembled reagents could greatly reduce the time and cost of *in vitro/ex vivo* gene editing applications in precision medicine and drug discovery and aid in the development of increased and serial dosing regimens for somatic gene editing *in vivo*.

Introduction

Precise editing of DNA sequences in the human genome correct mutations[157,173,174] or introduce novel genetic functionality[86] for many biomedical purposes. Specifically, non-viral delivery of preassembled CRISPR ribonucleoproteins (RNPs) is currently being developed for somatic gene editing applications[157,173–175]. RNPs combining *Streptococcus pyogenes* Cas9 nuclease (*Sp*Cas9, a high-affinity nuclease isolated from a type II CRISPR–associated system) and a single-guide RNA (sgRNA) generate on-target DNA double strand breaks (DSBs) with little to no off-target DNA cleavage[175,176]. This break is predominately repaired through one of two major DNA repair pathways: error prone non-homologous end joining (NHEJ) or precise homology directed repair (HDR), in which a template is used for precise gene editing (Figure 1a). Co-delivery of a nucleic acid donor template with the *Sp*Cas9 RNP (*Sp*Cas9 + sgRNA) is capable of producing precise edits at target loci through HDR of the DSB. However, variable delivery of the CRISPR system along with the donor templates generates a spectrum of edits, where a majority of cells include imprecise insertions and deletions (indels) of DNA bases from

NHEJ or microhomology mediated end-joining (MMEJ) of the DSB[177,178]. Even when precise HDR of the DSB occurs on one allele, there is a chance that both alleles are not identically edited, resulting in imprecise edits on the other allele[179,180]. Faithful writing of DNA, or scarless gene editing[137], within human cells at high efficiency remains an outstanding challenge.

Strategies to promote HDR include addition of small molecules to block NHEJ and restrict *Sp*Cas9 activity to particular phases of the cell cycle[146,181], but variability and toxicity has been observed across human cell lines when applying small molecules to promote HDR[182,183]. Also, selection strategies through viral integration and excision of drug[184] or cell-surface[173] selection cassettes, flow cytometry for co-expressed fluorescent proteins[180,185,186], or through transient drug selection[187] can assist in the isolation of cells with one or two precisely-edited alleles[177]. For all of these strategies, imprecise editing through NHEJ typically outnumbers precise HDR outcomes[174,176]. None of the current strategies precisely control the delivery of the RNP with the donor template, and many resort to 'flooding' the cell with high Cas9 expression and/or the donor template. We reasoned that some of the noise in gene editing outcomes could be reduced by preassembling RNPs with donor template or other moieties that enable the isolation of precisely-edited cells (**Figure 3.1a**). The S1mplex tool described here exploits high affinity interactions between a short RNA aptamer and streptavidin to promote more faithful writing of the human genome.

Results

Design of chimeric sgRNA to bind streptavidin.

We devised a strategy inspired by CRISPR display[20] that leverages structural studies of the RNP to identify locations in the sgRNA sequence where RNA aptamers could be tolerated. Three novel sgRNAs with a modification either in a stem loop of the sgRNA or at the 3' end were designed (**Figure 3.1b**), as these locations have previously been shown to tolerate additions with a minimal loss in Cas9 binding activity[188]. Separately, at each location, we added

perfectly complementary 10 nucleotide block previously shown to aid aptamer addition to sgRNAs[20] and a 60 nucleotide S1m aptamer[189], which has a strong non-covalent interaction with streptavidin. The added sequence extends the sgRNA stem loop and contains two distinct bulges used for binding. We termed these new sgRNAs S1m-sgRNA-1, S1m-sgRNA-2, and S1m-sgRNA-3 in reference to their position in the sgRNA from 5´ to 3´ (**Figure 3.1b**).

We confirmed that S1m-sgRNAs can be made rapidly *in vitro* via one-pot transcription[190] and are larger than standard sgRNAs when analyzed by agarose gel electrophoresis (**Figure B-1a**). Next, we verified the ability of S1m-sgRNAs to complex with streptavidin *in vitro* by combining a constant amount of S1m-sgRNA with increasing amounts of streptavidin (0.1, 1, and 10 molar equivalents) via an electrophoretic mobility shift assay (EMSA). The electrophoretic front of each S1m-sgRNA slowed as streptavidin levels increased (**Figure B-1b**). However, S1m-sgRNA-3 demonstrated the least shift suggesting that the secondary structure necessary for binding may be partially disrupted at this location. At all levels of streptavidin, the electrophoretic front slowed, demonstrating that S1m-sgRNA-1 and S1m-sgRNA-2 interact with streptavidin. In contrast, when the same amount of standard (non-S1m) sgRNA was run with 10 molar equivalents of streptavidin, the electrophoretic front remained constant (**Figure B-1b**).

To demonstrate the ability of S1m-sgRNA-1 to complex with streptavidin and Cas9 protein simultaneously, we performed dynamic light scattering (DLS). When streptavidin and Cas9 were mixed in solution, two peaks were distinct at 3.0 nm and 7.8 nm (**Figure 3.2a**), both of which match closely the radii previously reported for each protein[191,192]. We next formed Cas9 RNPs with excess standard sgRNAs and observed that the species formed were larger than Cas9 alone and did not increase in radius with the addition of streptavidin. Excess sgRNA was not detected by DLS and was included in the DLS studies to ensure all key components were able to assemble together (**Figure B-1c**). Additionally, these samples had a discernable peak

corresponding to the presence of streptavidin alone. RNPs containing S1m-sgRNA-1s and Cas9 protein increased in radius by a larger amount than RNPs containing standard sgRNAs and Cas9 protein, likely due to the increased length of S1m-sgRNA-1. When streptavidin was added to S1m-sgRNA-1 RNPs, the average radius of the complex was increased by ~3 nm (green arrowhead), the radius of streptavidin protein. These tertiary complexes of Cas9, S1m-sgRNA-1, and streptavidin are termed “S1mplexes”. The second, larger peak in the S1mplex DLS trace (green asterisk) is attributed to the tetrameric nature of streptavidin that can harbor up to four RNPs.

While assembly of S1mplexes *in vitro* is important, the maintenance of complexes post-delivery is imperative to gene editing function. To demonstrate this capability, we delivered Cas9 protein and streptavidin in combination with either sgRNAs or S1m-sgRNAs into human pluripotent stem cells (hPSCs) via nucleofection and conducted immunohistochemistry for the two protein components. Multispectral imaging flow cytometric analysis of single fixed cells confirmed the co-localization of the two protein components within hPSCs (**Figures 3.2b, B-2**). Significantly higher correlation in the fluorescent signals from the two protein components were seen when S1m-sgRNA-1 was included ($p<10^{-5}$, **Figure 3.2c**). To gain further subcellular resolution of these components after S1mplex delivery, images obtained using confocal microscopy on fixed, intact hPSC cultures were analyzed using CellProfiler[152] for overlap between the two components within the nuclei. At 24 hours after delivery, the correlation between the fluorescent signals arising from Cas9 and streptavidin within the nucleus was significantly higher when using S1m-sgRNAs than sgRNAs ($p<0.05$, **Figure 3.2d, e**). Together, these results indicate that complexes between Cas9 and streptavidin are preserved specifically through the S1m aptamer during transfection and subsequent subcellular trafficking such as nuclear transport.

Next, we examined the ability of S1m-sgRNAs to edit genes within human cells. We created a human embryonic kidney (HEK) cell line that constitutively expressed blue fluorescent protein (BFP) from an integrated transgene[193]. DSBs produced by sgRNAs that target the fluorophore in combination with Cas9 expressed from a transfected plasmid are repaired predominantly through NHEJ, with indel formation at the DSB. NHEJ-mediated gene edits are expected to result in a loss of BFP fluorescence within this HEK line. After delivery of S1m-sgRNAs and a plasmid encoding Cas9 to this HEK line, BFP expression was analyzed via flow cytometry. All S1m-sgRNAs (1, 2, and 3) created indels at approximately half the frequency of standard sgRNAs (**Figure B-3a**). While the ~2-fold decrease in generating indel edits is significant, such decreases in indel formation have been linked to a concomitant decrease in off-target effects[194].

Assembly of DNA repair template to RNP.

We subsequently searched for a method to combine a donor DNA template with S1mplexes and form a quaternary complex in order to promote precise editing through HDR. Given the strong interaction between streptavidin and biotin ($K_D=10^{-15}$ M), we selected biotinylated single-stranded oligodeoxynucleotide (ssODNs) donor templates. All components (S1m-sgRNA, streptavidin, biotin-ssODN) were run individually on a gel and compared side-by-side with standard reagents (sgRNA, ssODN) to establish baseline migration characteristics via an EMSA. The biotin-ssODN ran slightly larger than the standard ssODN, presumably due to the biotin modification (**Figure B-3b**, lanes 4-5). Tertiary complexes of S1m-sgRNA-1, streptavidin, and ssODN were formed using 0.5, 1, or 2 molar equivalents of biotin-ssODNs. The primary band displayed a distinct electrophoretic shift from either the sgRNA or ssODN alone, indicating complex formation (**Figure B-3b**, lanes 6-9). To demonstrate that all components combined successfully, unmodified ssODNs were run in the place of biotin-

ssODNs. The unmodified ssODN displayed the expected electrophoretic shift despite the presence of the S1m-streptavidin complex (**Figure B-3b**, lanes 10-12). Finally, standard sgRNA was run with streptavidin and biotin-ssODN. In this condition, the band attributed to S1m-sgRNA-1-streptavidin or S1m-sgRNA-1-streptavidin-ssODN binding was not observed and instead solid bands representing sgRNA and ssODN-streptavidin were present (**Figure B-3b**, lane 13-14). Due to the strong interaction of biotin and streptavidin, we needed to ensure that biotin did not displace S1m-sgRNA-1 already bound to streptavidin when added in solution. To do so, we combined S1m-sgRNA-1s with streptavidin at a 1:1 molar ratio. We then added 4-fold molar excess of biotin to occupy every binding site on each streptavidin molecule and incubated the complex for 0, 5, 10, 20, or 30 minutes. After incubation, gel shift following electrophoresis was not different from bound S1m-sgRNA:streptavidin combinations suggesting that biotin did not interfere with the S1m-streptavidin interaction at four times the concentrations used in this study (**Figure B-3c**).

Increased HDR:indel ratios in human cells.

We tested the ability of all three ssODN-S1mplexes to induce HDR in a hPSC line containing a BFP-expressing transgene that can be switched to express GFP through a 3 nucleotide switch (**Figure B-4a**)[193]. S1mplexes with biotin-ssODNs (ssODN-S1mplexes) were assembled using one of the three S1m-sgRNAs and compared to standard sgRNAs and ssODN combinations. After delivery of ssODN-S1mplexes and subsequent deep sequencing of genomic DNA, we found that all three ssODN-S1mplexes had a higher ratio of HDR:indel editing than standard RNPs. ssODN-S1mplexes with S1m-sgRNA-1 and S1m-sgRNA-2 induced similar ratios of HDR:indel editing while ssODN-S1mplexes with S1m-sgRNA-3 had a slightly depressed HDR:indel ratio (**Figure 3.3a**). The decreased HDR:indel ratio found using S1m-sgRNA-3 may have been due to the lower binding affinity of this sgRNA with streptavidin, as

seen in the EMSA (**Figure B-1b**). In order to minimize the frequency of indel mutations while maximizing HDR, we decided to use S1m-sgRNA-1 for all remaining experiments and will refer to it henceforth simply as S1m-sgRNA.

With this knowledge, we then evaluated S1mplexes in multiple human cell lines for their ability to generate a variety of precise nucleotide changes. We assembled ssODN-S1mplexes to again switch BFP to GFP. After delivery to HEK cells, deep sequencing revealed that the ssODN-S1mplexes enriched the ratio of precise insertions to imprecise editing 18.4-fold over standard RNPs and approached a ratio of four precise edits to every one indel (**Figure 3.3b, Figure B-4b and Table B-9**). When the same experiments were conducted in hPSCs, results from flow cytometry assays were consistent with these conclusions from deep sequencing (**Text B-1, Figure B-4c-f and Table B-9**). Additionally, when introducing a 12 nucleotide insertion into the *EMX1* locus[195] of HEKs with ssODN-S1mplexes, the ratio of precise insertions to imprecise editing increased 2.7-fold over standard sgRNA RNPs (**Figure 3.3c, Figure B-4g and Table B-9**). Taken together, this shows that ssODN-S1mplexes are able to shift the balance of editing to enrich for small, precise edits within the genome.

We tested the ability of this strategy to create even larger sequence changes in hPSCs by designing an ssODN that carried a variable 18 nucleotide insertion. We deep sequenced the cell population after delivery of ssODN-S1mplexes, again targeting the *BFP* and *EMX1* loci. When standard sgRNA RNPs were transfected with streptavidin-ssODN complexes, minimal insertion was seen with a subsequently low ratio of precise HDR to imprecise indel alleles (**Figure 3.3d and Table B-9**). Equivalent precise:imprecise ratios were seen when standard sgRNA RNPs and ssODNs were transfected as when S1m-sgRNA RNPs were transfected with biotin-ssODN (without streptavidin) (**Figure 3.3d and Table B-9**). However, levels of indels were increased in the sgRNA RNP-free ssODN condition (**Figure B-4h and Table B-9**). When the full ssODN-S1mplexes were transfected into hPSCs, HDR insertion levels greatly increased

(**Figure B-4i**) as did the ratio of precisely-edited to imprecisely-edited alleles to 9.7 fold over standard RNP methods (**Figure 3.3d**). Again, we observed four precise edits to every one indel with ssODN-S1mplexes at this locus. At the endogenous *EMX1* locus, we delivered the S1m-sgRNA RNPs with biotin-ssODNs either with or without streptavidin. When streptavidin was added to generate the full ssODN-S1mplex, rates of insertion increased 51-fold (**Figure B-4j**), and the ratio of precise to imprecise gene-editing increased 15-fold (**Figure 3.3e**). Taken together, each component of the ssODN-S1mplex is necessary to drive higher HDR:indel ratios within human cells.

Design constraints on the ssODN-S1mplex.

Recent studies have reported that the design of the ssODN has a significant effect on the rate of HDR[183,193]. Accordingly, we explored various ssODN designs with ssODN-S1mplexes. Designs were limited to a 100 nucleotide length for ease of synthetic synthesis, but varied as follows: asymmetrical around the cut site, extending 30 upstream and 67 bp downstream or vice-versa, either identical to the sequence containing the PAM or the reverse complement (non-PAM), and biotinylated on either the 5´ or 3´ end of the ssODN (**Figure 3.4, left**). S1mplexes containing each unique ssODN were assembled and transfected separately into BFP-expressing hPSCs. Four days after delivery, genomic DNA from each condition was collected and analyzed using deep sequencing. Under these conditions, 2.8±2.2% of alleles in all samples were edited via HDR and NHEJ (**Figure 3.4a, top and Table B-9**). We observed that neither the asymmetry, sidedness, biotin, nor location on the ssODN had a significant effect on the HDR or indel outcomes using ssODN-S1mplexes (**Figure 3.4a, top and Table B-9**). Precise editing ranged from 2-10 times greater than imprecise editing (**Figure 3.4b, top and Table B-9**).

We next sought to test these ssODN designs at an endogenous *GAA* locus using a patient-derived hPSC line[196] that contains a pathogenic 1 bp deletion in exon 10 on one allele. We designed sgRNAs that target only the mutant allele as well as ssODNs to correct the mutation to wildtype and modify the PAM site. These ssODNs were again asymmetrical, 34 bp upstream and 66 bp downstream from the cut site, complementary to the PAM or non-PAM strand, and biotinylated at either the 5′ or 3′ end of the ssODN (**Figure 3.4, bottom**). At this locus ssODN-S1mplexes again had higher levels of precise to imprecise editing than RNPs consisting of sgRNAs, with 3-8 precise edits occurring for every imprecise edit (**Figure 3.4b, bottom and Table B-9**). Consistent with the sequencing results at the *BFP* locus, absolute levels of HDR and NHEJ editing were 2.0±1.1% (**Figure 3.4a-b, bottom and Table B-9**). There was still no significant difference between any of the ssODNs tested when complexed to the S1mplex.

Adding fluorescent cargoes to the RNP.

To facilitate isolation of the precisely-edited cells, we pursued a strategy to label the cells that received the S1mplexes by including additional biotinylated fluorescent cargoes. We preassembled standard streptavidin-conjugated quantum dots (QdotSA, 20 nm diameter) with S1mplexes (QdotSA-S1mplexes, **Figure 3.5a, top**). After transfection of QdotSA-S1mplexes into HEKs, a subpopulation of cells contained Qdots within the cytoplasm. High-intensity green fluorescence dots were distributed variably across the transfected cell population, indicating that standard transfection methods likely generate significant heterogeneity in the number of RNPs delivered to each cell. Despite the presence of Qdots in the cytoplasm, very low gene editing was observed upon further culture and analysis within a HEK H2B-mCherry reporter cell line (**Figure 3.5b, B-5**), suggesting linkage to Qdots was inhibiting RNP function within inside the cell, perhaps due to the Qdots being too large to allow RNP function within the nucleus. Even

transfecting QdotSA five hours post transfection of S1m-sgRNA RNPs had a significant effect on gene editing, suggesting QdotSA could complex with and inhibit S1m-sgRNA RNPs within cells (**Figure B-5**).

In the ensuing experiments, we chemically-modified Qdots so that the biotin linkage of the S1mplex to the Qdot was mediated through a cell-cleavable disulfide linker (Qdot-SS-S1mplex, **Figure 3.5a, bottom**). With this cleavable linker, we observed a gain in gene editing activity (**Figure 3.5b**), while the Qdots remained largely within the cytoplasm (**Figure 3.5c**), suggesting separation and nuclear transport of the RNP. The fluorescence from the Qdot at 24 hours post transfection was utilized for fluorescence activated cell sorting (FACS). There was a shift in fluorescence for the whole cell population, indicating uptake of Qdot-S1mplexes in most cells, although to differing extents (**Figure 3.5d**). Sorted cells based on Qdot positive fluorescent signal resulted in gene editing at 3.7-fold higher rates versus cells transfected using standard methods (**Figure 3.5e**).

Multiplexed gene editing with S1mplexes.

To obtain further control and refine the mutagenic spectrum of S1mplexes, we attached a fluorescent label directly to streptavidin that could be used for identification during flow cytometry. We preassembled an S1m-sgRNA and biotin-ssODN targeting *BFP* with a streptavidin labeled with a red fluorophore (AlexaFluor-594) (**Figure 3.6a**) and then performed a single cell FACS for the isolation of clones that had high fluorescence after delivery. Upon further cell culture, clones were analyzed by Sanger sequencing for editing at the *BFP* locus. Of the 34 isolated clones in the S1mplex-positive population, eight underwent HDR; eight harbored indels; and, the rest remained unedited (**Figure 3.6b**). In comparison, when using sgRNAs, seven of the 41 isolated clones harbored indels and none were positive for HDR. Cell populations did not contain mosaic gene editing, indicating that defined gene editing outcomes

could be enriched by FACS on the S1mplex fluorescence. Using this capability we tested whether if it was possible to multiplex edits using differently colored S1mplexes. We thus assembled the same ssODN-S1mplex targeting BFP, termed red-ssODN-S1mplex, and separately complexed an S1m-sgRNA and biotin-ssODN targeting *EMX1* with a streptavidin labeled with a green fluorophore (AlexaFluor-488), termed green-ssODN-S1mplex (**Figure 3.6a**). The two ssODN-S1mplexes were mixed and transfected simultaneously into HEKs (**Figure 3.6c**).

Twenty-four hours post transfection, we sorted cells using FACS into one of four populations: positive for either fluorophore, both, or neither (**Figure 3.6d**). Only the top 2% of each population was taken, as we observed some association of the fluorescent S1mplex with the cell membrane in addition to robust fluorescent signal within the nucleus of some of the cells (**Figure 3.6c**). One-week post sort, each of the four populations was analyzed for editing via deep sequencing as well as by flow cytometry for *BFP* editing or insert-based PCR for *EMX1*. Deep sequencing revealed that editing at the *EMX1* locus was increased in the presence of green-ssODN-S1mplexes (Green+ and double positive fractions) (**Figure 3.6d, B-6b and Table B-9**). In these populations the ratio of precise to imprecise edits increased and approached one and was 2-fold greater than that of the double negative fraction (**Figure B-6b and Table B-9**). Similarly, editing at the *BFP* locus was increased in the Red+ and double positive fractions. As was seen in previous deep sequencing experiments, the ratio of precise to imprecise edits was elevated in the presence of S1mplexes. With the addition and sorting of fluorescent S1mplexes, the ratio was greater than 10 insertions per indel (**Figure 3.6d and Table B-9**). Interestingly, the level of indels was highest in the double negative fraction (**Figure B-6b and Table B-9**); this may be due to the presence of unlabeled RNPs that did not complex with streptavidin. Results with conventional flow cytometry and PCR assays followed the same trends, consistent with these conclusions from deep sequencing (**Text B-1, Figure B-6c, d and Table B-9**). We

analyzed the top 5 off–target sites for both the BFP and EMX1 sgRNAs using TIDE[197] in the sorted fractions as well as previous samples used for deep sequencing. None of the sorted populations using ssODN-S1mplexes had modifications above the TIDE limit of detection (**Figure 3.6e, Figure B-7**). However, using standard sgRNA RNPs, notable off-target mutagenesis occurred at *EMX1* off-target site 2 (**Figure B-7**). Taken together, the assembly of S1mplex particles with a fluorescent tag can be used to create multiple, precise edits with increased efficiency without needing multiple transfections or extended culture.

Discussion

Together, our results indicate that addition of an S1m aptamer to the sgRNA can generate nanoparticles that shift the balance of precise gene editing outcomes to outnumber imprecise editing. This shift toward precise gene editing was seen across two different cell lines and three different loci with a variety of insertions and base pair changes. The S1m-streptavidin linkage was strong enough that excess biotin was unable to displace it *in vitro*. Further, the S1mplex was not observed to dissociate after nucleofection suggesting the complex is stable under a wide variety of conditions. We saw robust S1mplex localization both immediately following and one day post transfection. Editing with RNPs has been observed to peak within 24 hours of delivery[198]. This supports the idea that the S1mplex is intact within the cell and nucleus. Therefore, S1mplexes may enable HDR by creating a more defined stoichiometry of the RNP to the ssODN within each cell and presumably at the DSB. Each DSB would have at least one local HDR template, which could either suppress NHEJ repair or promote HDR. One group recently reported an approach to localize donor ssODN to the DSB by chemically conjugating the sgRNA with the ssODN and showed increased levels of HDR but no concurrent decrease in indel formation[199]. Further studies characterizing the location of ssODN relative to the DSB

within the nucleus may help to clarify the different gene editing outcomes among differing assembly strategies.

Interestingly, and contrary to recently published studies[183,193], the asymmetry and sidedness of donor DNA complexed to S1mplexes did not have a significant effect on the ratio of precise editing to imprecise editing. This could be due to the template being readily available for access by HDR mechanisms as opposed to variation based on factors such as cell cycle[146] or cell type. Increased HDR outcomes after Cas9 DNA cleavage have recently been linked to the ataxia-telangiectasia mutated (ATM) signaling pathway within human zygotes[176,200] and Fanconi anemia pathways within human hematopoietic progenitor cells[71], leading to the hypothesis that distinct cellular host factors can control the extent of HDR. Given the robust shift towards HDR in both HEK and hPSC lines with S1mplex, S1mplexes likely interact with common DNA repair machinery found in both HEKs and hPSCs. Further, because the polarity of ssODN biotinylation did not affect the shift towards HDR, differential modifications of the ssODN could be performed selectively at either end to protect from degradation and increase HDR efficiency[65].

Using ssODN-S1mplexes, absolute levels of editing via HDR was 1.6% on average across all cell types and loci, and ranged from 0.052% to 6.6% (**Tables B-8-9**). Only in one experiment did we see lower levels of HDR using ssODN-S1mplexes when compared to a standard strategy (12 bp insertion into *EMX1* within HEKs, **Figure B-3.4g and Table B-9**). While less than two previous reports in human cells[193,201], our absolute editing efficiency is similar to other reports in pluripotent stem cells *in vitro* (**Table B-8**). Producing HDR at the levels observed in this study while reducing the prevalence of indels has been a key challenge with hPSCs[202]: nevertheless, some studies have been able to target loci that may have important physiological effects[203,204] and model point mutations within inherited disorder[180] using high-throughput or selection strategies. Also, these levels of HDR with high levels of NHEJ can be

useful for hotspot mutation modeling[205] and for drug target validation[129] within human cancer cell lines. We note that several reports have suggested that 0.1-15% levels of HDR may have a significant effect *in vivo* on clinical phenotypes such as Duchenne muscular dystrophy[206], haemophilia B[207], and tyrosinemia type 1[208,209].

We noticed that HDR levels decreased as the length of insertion increased and the length of homology arms decreased, which has been shown previously[137,183,210]. We achieved the highest ratios of precise editing to imprecise editing when generating small changes in an integrated transgene at one allele. This may account for decreased ratios observed at the *EMX1* locus where insertions were 20% the length of donor ssODN. Other reports have noted that repair at this allele poses a greater challenge than at the *BFP* locus, even when attempting small nucleotide changes,[193] and that larger homology arms play a significant role in the rate of HDR at this locus[146]. Utilizing larger homology arms for insertion of longer DNA stretches or even transgenes within the genome using S1mplexes has not been explored in this study but warrants further study. There is also the potential at this locus that that repair could be accomplished by a second allele within the genome. HDR using a second endogenous allele for our experiments would be classified as a wildtype deep sequencing read, so these editing events were not captured in the precise editing HDR frequencies. Importantly, ratios of precise editing to imprecise mutations at the pathogenic *GAA* allele, which underlies Pompe disease[196], were similar to those found at *BFP*. This consistency suggests that the variation in HDR:indel ratio observed when compared to large insertions may be due more to the number of nucleotide changes rather than locus-based variation. There may also have been further repair using the healthy *GAA* allele: however, we were unable to unravel this possibility with certainty in a bulk population. We anticipate that higher ratios of precise to imprecise editing could be generated for single nucleotide changes. 44,750 disease-associated single nucleotide or indel mutations

up to 50 nucleotides in length in the ClinVar database[211] can be corrected, in principle, by HDR via donor ssODN templates.

Gene editing in human cells could be controlled by the linkages within the S1mplex. For the Qdot-S1mplexes, a gain of RNP activity occurred only after switching to a labile disulfide bond, presumably because large cargoes such as Qdots (20 nm diameter), assembled with the RNP inhibit Cas9 nuclease activity. The smaller ssODN-S1mplexes without Qdots with mean diameters of 16 nm could generate edits at target loci. The Qdot-S1mplex results demonstrate that the biotin-streptavidin linkage is strong enough to associate biotinylated cargoes with the RNP within human cells, while disulfide bonds, which are labile at low pH[212], likely dissociate the S1mplex in acidic endocytotic trafficking compartments and release the RNP from the cargo to fully recover activity. Regulating CRISPR gene editing tightly through the release of large cargoes could be explored with other chemistries that generate labile cargoes upon excitation by light or heat[213]. Such strategies could advance targeted therapy to specific areas and cell types within the body and also be applied through a multiple dosing regimen, taking advantage of the high precise-to-imprecise editing ratio of S1mplexes.

One study recently reported that biotin-ssODNs could be recruited to RNPs within the cell produced by translation of injected Cas9-avidin mRNA[214]. mRNA strategies still require robust host translational machinery and must avoid an immune response to the delivered foreign mRNA. The Cas9-avidin fusion protein experiments with human cells did not demonstrate lower NHEJ and had a 2 fold change in the HDR frequency. These changes are lower than the 3-10 fold changes that we observed at different loci in hPSCs. While the use of fewer components, such as in a Cas9-avidin system, decreases the complexity of editing experiments, our ability to preassemble necessary moieties outside the cell has numerous advantages. The S1mplex strategy offers multiplexing capabilities, such as with the assembly of Qdots to the S1mplex, as well as the potential to purify and isolate only successfully assembled

particles. In contrast, in the use of Cas9-avidin mRNA system, ssODN must still be localized to gene-editing machinery within the cell and after translation, which causes significant lag time while still having significant free ssODN within the cells.

The S1mplex strategy provides a straightforward, robust and modular toolkit to regulate the gene editing activity of *Sp*Cas9 RNPs. RNA modification of the sgRNA with S1m aptamer can be performed readily through short nucleic acid synthesis methods, whereas other strategies that engineer the Cas9 protein can add challenges in protein expression, purification and stability[214,215]. Movement of the S1m aptamer around the sgRNA also showed S1mplex formation is viable in each stem loop, suggesting that multiple S1m sequences could be added simultaneously to increase avidity[189] or that other functional aptamers could be added in addition to S1m[20,216]. Use of RNA modifications could complement and add functionality to already engineered variants of *Sp*Cas9 (e.g., high fidelity[217], switchable[167], and optogenetic[168] nucleases). Preassembled, non-viral S1mplexes could also be readily manufactured as off-the-shelf reagents with well-defined critical quality attributes appropriate for clinical use: for example, avidin has previously been tolerated in clinical trials[218], and clinical grade *Sp*Cas9 is available from several vendors. We anticipate that the S1mplex strategy will be used as a versatile toolkit to further refine our abilities for precise, scarless gene editing *in vitro* and potentially *in vivo*.

Methods

Cell culture. All hPSCs were maintained in E8 medium on Matrigel (WiCell) coated tissue culture polystyrene plate (BD Falcon). Cells were passaged every 3-4 days at a 1:6 ratio using Versene solution (Life Technologies). WA09-BFP hESCs were generated through lentiviral transduction of BFP dest clone (Addgene #71825) and sorted to ensure clonal populations.

After expansion, lines were sorted monthly on a BD FACS Aria to maintain expression levels. hiPSCs were gifts from the T. Kamp and M. Suzuki (UW-Madison) labs.

Human embryonic kidney cells (293T) were maintained between passage 15-60 in Growth medium containing DMEM (Life Technologies), 10% v/v FBS (WiCell), 2mM L-Glutamine (Life Technologies), and 50 U/mL Penicillin-Streptomycin (Life Technologies). Cells were passaged 1:40 with Trypsin-EDTA (Life Technologies) onto Gelatin-A (Sigma) coated plates. HEK-H2B-mCherry lines were generated through CRISPR-mediated insertion of a modified AAV-CAGGS-EGFP plasmid (Addgene #22212) at the AAVS1 safe harbor locus using gRNA AAVS1-T2 (Addgene #41818). HEK-BFP lines were generated and maintained as mentioned above. All cells were maintained at 37ºC and 5%CO_2

One-pot transcription of S1m-sgRNA variants. S1m-sgRNAs were synthesized in three steps. Double-stranded DNA blocks encoding the sgRNA scaffold and the S1m aptamer were first created by overlap PCR using Phusion HighFidelity Polymerase (New England Biolabs) according to the manufacturer's protocols. The S1m-sgRNA-1 scaffold was generated under the following thermocycling conditions: 30 cycles of 98 ºC for 10 s and 72 ºC for 15 s with a final extension period of 72 ºC for 10 min. The S1m-sgRNA-2 and S1m-sgRNA-3 scaffolds were generated under the following thermocycling conditions: 10 cycles of 98 ºC for 10 s, 61.2 ºC for 10 s, and 72 ºC for 15 s, followed by 20 cycles of 98 ºC for 10 s and 72 ºC for 15 s with a final extension period of 72 ºC for 10 min. These scaffolds were combined with a second primer consisting of a truncated T7 promoter, the sgRNA target, and homology to the S1m scaffold and PCR was performed again under the following thermocycling conditions: 98 ºC for 30 s followed by 35 cycles of 98 ºC for 5 s, 60 ºC for 10 s, and 72 ºC for 15 s, with a final extension period of 72 ºC for 10 min. S1m PCR products were then incubated overnight at 37ºC in a HiScribe T7 IVT reaction (New England Biolabs) according to manufacturer's protocol. The resulting RNA

was purified using MEGAclear Transcription Clean-Up Kit (Thermo Fisher) and quantified on a Nanodrop2000.

S1m-sgRNA RNP formation. NLS-Cas9-NLS protein (Aldevron, Madison, WI) was combined with S1m-sgRNAs at a 1:1.2 molar ratio and allowed to complex for 5 minutes with gentle mixing. To this complex, streptavidin (Life Technologies) was added at an equimolar ratio to Cas9 and the mixture was allowed to complex for an additional 5 minutes. Finally, a 1.2 molar equivalent of biotin-ssODNs (Integrated DNA Technologies) to Cas9 was added to the tertiary complex and subsequently vortexed at low speed. This final mixture was then allowed to combine for 10 minutes to ensure complex formation.

S1m-sgRNA and streptavidin binding gel shift assays. S1m-sgRNAs were heated at 75 ºC for 5 min and cooled to room temperature for 15 min. 20 pmol S1m-sgRNA was combined with streptavidin at 10:1, 1:1, and 1:10 molar ratios in a final volume of 5 µl and the mixture was allowed to complex for 10 min. The S1m-sgRNA-streptavidin complexes were run on a 1% agarose gel. Tertiary complexes were assembled by first mixing 15 pmol each of S1m-sgRNA and streptavidin. To this mixture, 6, 15, or 30 pmol of ssODN was added prior to running the complexes through a 1% agarose gel.

Biotin competition assay. S1m-sgRNA was heated to 75ºC for 5 min and cooled to room temperature. 20 pmol each S1m-sgRNA and streptavidin were complexed for 10 min. 80 pmol biotin was added at 30, 20, 10, 5, and 0 min intervals prior to running the complexes through a 1% agarose gel.

RNP delivery. HEK transfections were performed using TransIT-X2 delivery system (Mirus Bio, Madison, WI) according to manufacturer's protocol. $2.5x10^5$ cells/cm^2 were seeded in a 24-well

plate 24 hours prior to transfection. RNP complexes were formed as described in 25 µL of Opti-MEM (Life Technologies). 5 pmol of Cas9 protein, 6 pmol sgRNA, 5 pmol streptavidin, and 6 pmol ssODN were used. In a separate tube, 25 µL of Opti-MEM was combined with 0.75 µL of TransIT-X2 reagent and allowed to combine for 5 minutes. TransIT-X2 and RNP solutions were then mixed by gentle pipetting and placed aside for 15 minutes. After this incubation, 50 µL of solution were added dropwise into the well. Media was changed 24 hours post transfection.

For HEK transfections involving quantum dots, Lipofectamine 2000 (Life Technologies) was used for delivery. Qdot-RNP complexes were formed using the following amounts of reagents (for 24 wells: 5 pmol of Cas9 protein, 6 pmol sgRNA, 5 pmol streptavidin, 3.125 pmol of quantum dots and 3ul lipofectamine per well; a quarter of these amounts were used when transfecting 5000 cells in 96 well plates).

All hPSC transfections were performed using the 4D-Nucleofector System (Lonza) in P3 solution using protocol CB-150. Cells were pretreated with Rho-kinase (ROCK) inhibitor (Y-27632 Selleck Chemicals) 24 hours prior to transfection. 50 pmol Cas9, 60 pmol sgRNA, 50 pmol streptavidin, and 60 pmol ssODN were used to form particles per ssODN-S1mplex as described above. Cells were then harvested using TrypLE (Life Technologies) and counted. $2x10^5$ cells per transfection were then centrifuged at 100x*g* for 3 minutes. Excess media was aspirated and cells were resuspended using 20 µL of RNP solution per condition. After nucleofection, samples were incubated in nucleocuvettes at room temperature for 15 minutes prior to plating into one well of a 6-well plate containing E8 media+10 µM ROCK inhibitor. Media was changed 24 hours post transfection and replaced with E8 medium.

Dynamic light scattering. DLS was performed using a DynaPro NanoStar (Wyatt Technology) using small volume (4 µL) disposable cuvettes. 10 µg of each component was added into the

cuvette and diluted as necessary with dH_2O to reach 4 µL of solution volume. Excess sgRNA was unable to be detected by DLS and was included to ensure all important species were able to complex together. In mixed component conditions, components were allowed to mix for 5 minutes while taking readings. Acquisitions were performed for 20 seconds with a minimum of 4 acquisitions per measurement. 5 measurements were performed per sample and were conducted at room temperature. Data was graphed as a function of percent intensity.

Immunocytochemistry. To measure correlation hPSCs were transfected with Cas9 protein and streptavidin-AF-647. 24 hours post transfection, cells were fixed using 4% PFA and incubated at room temperature for 10 minutes. Cells were then permeabilized using 0.05% Triton X-100 and incubated for 10 minutes. Following two washes with 5% goat serum, Cas9 antibody (Clontech #632607, 1:150) was added to cells and incubated overnight at 4 ºC. The next day, cells were rinsed twice with 5% goat serum and then incubated with a goat anti-rabbit secondary antibody (Santa Cruz Biotech #sc-362262, 1:500) for one hour at room temperature. Cells were then washed twice with PBS and mounted for imaging.

To visualize S1mplexes in the nucleus human embryonic kidney cells (HEK293T) were plated at 16,000 cells/well in an 8-well chamber slide at day 0. On day 1, 20 µL of transfection media was added to cells in 200 µL of maintenance media. Transfection media contained 20 µL Opti-MEM (Life Technologies), 10 pmol Streptavidin Alexa Fluor 488 conjugate (Thermo Fisher), and 0.6 µL *Trans*IT transfection reagent (Mirus). On day 3, cells were incubated with 1x CellMask Plasma Membrane Stain (ThermoFisher) and 1x Hoechst for 10 min. Following incubation at 37℃, cells were immediately washed with PBS and fixed in 4% paraformaldehyde (IBI Scientific) at room temperature for 15 min. Cells were analyzed using a Nikon Eclipse TI epifluorescent microscope and a Nikon *AR1* confocal microscope.

Multispectral imaging flow cytometry. hPSCs were transfected and stained as described above. After staining, cells were centrifuged and resuspended in 50 µL PBS. Fluorescence was detected on ImageStream X Mark II (EMD Millipore) according to manufacture instructions. Cellular colocalization was measured by IDEAS software package (Amnis) using predefined colocalization wizard.

Quantum dot biotin conjugation. To make Qdot-SS-S1mplexes, amine-PEG green fluorescent quantum dots (Qdot ITK 525 – ThermoFisher) were reacted with a degradable dithiol biotin linker (EZ-Link Sulfo-NHS-Biotin – ThermoFisher) as follows: First, 25 µL of an 8 µM Quantum dot solution in 50 mM Borate buffer was desalted into PBS using Zeba desalting columns (40K MWCO – ThermoFisher) and then reacted with excess sulfoNHS-dithiol-biotin linker for 2 hours at 4 ºC with shaking. The conjugate was purified from excess linker through buffer exchange in the desalting columns. Quantum dots retained their fluorescence and were stored at 4 ºC until use.

Flow cytometry. Flow cytometry of BFP expression and conversion to GFP was measured using a BD FACS Aria using the DAPI and FITC filters and analyzed using FlowJo. Voltages were established by running wild type WA09 hPSCs as well as WA09-BFP hPSCs. Sorting was performed on a BD FACSAria II with a nozzle size of 100 µm at room temperature and sorted into culture media.

Genomic analysis. DNA was isolated from cells using DNA QuickExtract (Epicentre, Madison, WI) following treatment by 0.05% trypsin-EDTA and centrifugation. The DNA extract solution was incubated at 65ºC for 15 minutes, 68ºC for 15 minutes, and finally 98ºC for 10 minutes. Genomic PCR was performed following manufacturer's instructions using AccuPrime HiFi Taq (Life Technologies) and 500 ng of genomic DNA. Products were then purified using AMPure XP

magnetic bead purification kit (Beckman Coulter) and quantified using a Nanodrop2000. For deep sequencing, samples were pooled and run on an Illumina HiSeq2500 High Throughput at a run length of 2x125bp or an Illumina Miseq 2x150bp.

Deep sequencing data analysis. A custom python script was developed to perform sequence analysis. The pipeline starts with preprocessing, which consists of filtering out low quality sequences and finding the defined ends of the reads. For each sample, sequences with frequency of less than 10 were filtered from the data. Sequences in which the reads matched with primer and reverse complement subsequences classified as “target sequences”. Target sequences were aligned with corresponding wildtype sequence using global pairwise sequence alignment. Sequences that were misaligned around the expected cut site were classified as NHEJ events while sequences that had insertions larger than 15 bp were classified as HDR events. The frequency, length, and position of matches, insertions, deletions, and mismatches were all tracked in the resulting aligned sequences.

Statistics. All error bars are shown as ±1 standard deviation. *p* values were computed using a Student’s two-tailed t-test and deemed significant at $\alpha<0.05$.

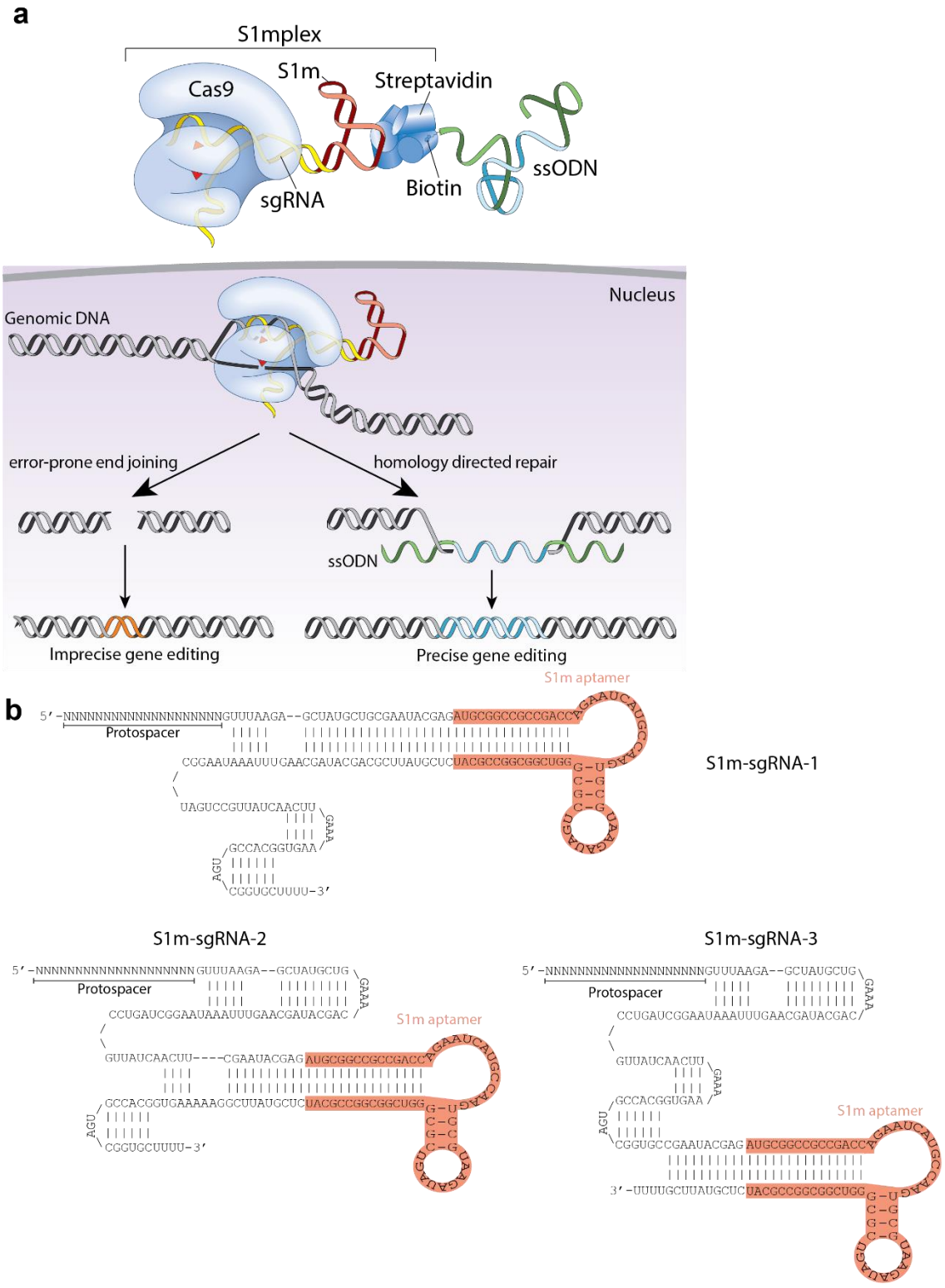

a
S1mplex
S1m
Cas9
Streptavidin
ssODN
Biotin
sgRNA
Nucleus
Genomic DNA
error-prone end joining
homology directed repair
ssODN
Imprecise gene editing
Precise gene editing
b
S1m aptamer
Protospacer
S1m-sgRNA-1
S1m-sgRNA-2
S1m-sgRNA-3
S1m aptamer
S1m aptamer

Figure 3.1 | **Design of S1mplexes with SpCas9.**

(a) Schematic showing preassembled ssODN-S1mplexes which are complexes of *Sp*Cas9 protein, sgRNA with a S1m aptamer, streptavidin, and a single-stranded oligodeoxynucleotide (ssODN) donor template. S1m-sgRNAs add an RNA aptamer at a stem loop of the sgRNA that is capable of binding streptavidin protein. A biotin-ssODN is then added to this tertiary complex. ssODN-S1mplex particles are designed to promote homology directed repair (HDR). **(b)** Sequence and secondary structure of each S1m-sgRNA. Protospacer designates the region of the sgRNA that defines the sequence to target in the human genome. S1m stem loop (red) binds streptavidin.

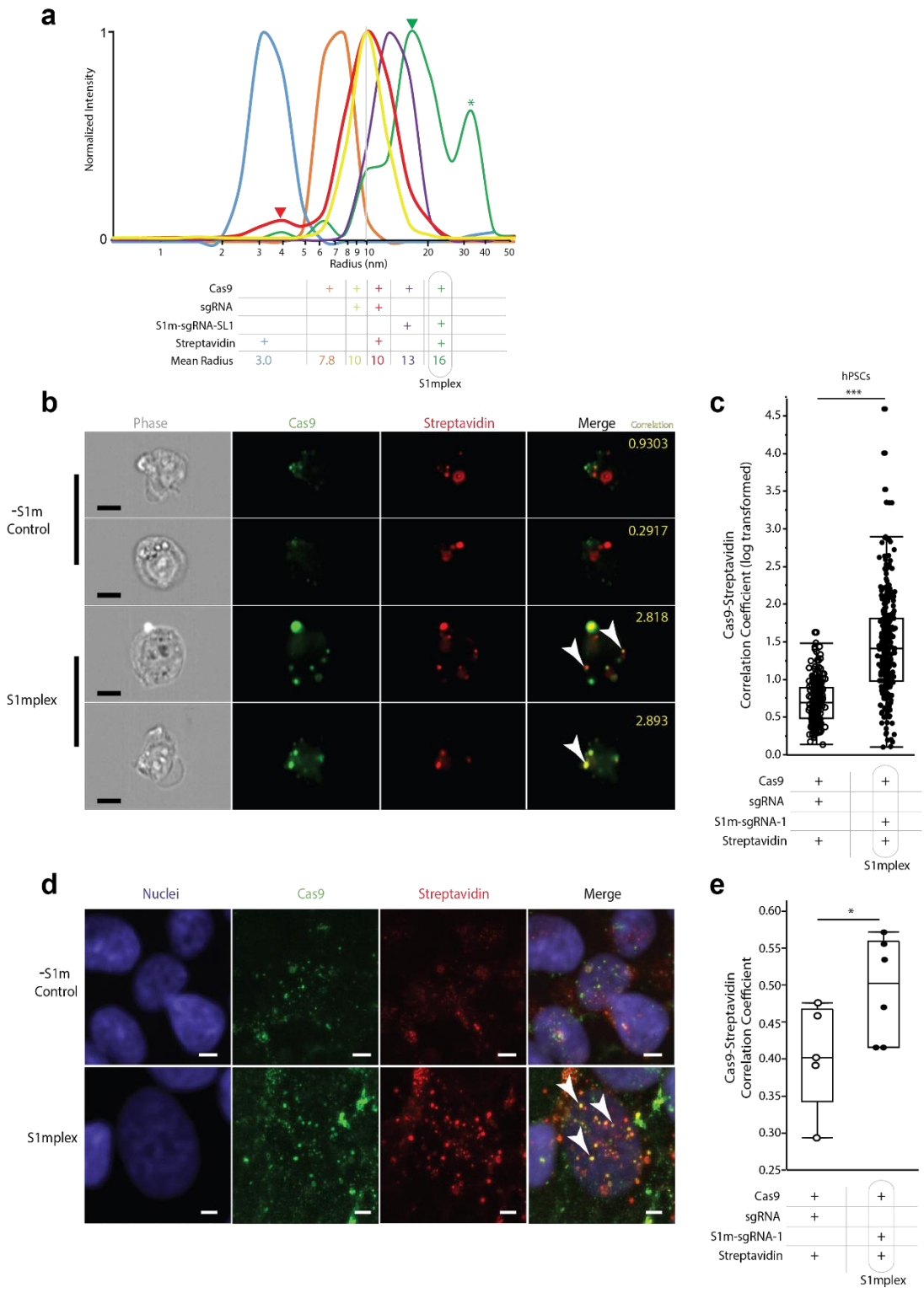

a
Normalized Intensity
Radius (nm)
Cas9
sgRNA
S1m-sgRNA-SL1
Streptavidin
Mean Radius
3.0
7.8
10
10
13
16
S1mplex
b
Phase
Cas9
Streptavidin
Merge
Correlation
0.9303
0.2917
2.818
2.893
-S1m Control
S1mplex
c
hPSCs
***
Cas9-Streptavidin Correlation Coefficient (log transformed)
Cas9
sgRNA
S1m-sgRNA-1
Streptavidin
S1mplex
d
Nuclei
Cas9
Streptavidin
Merge
-S1m Control
S1mplex
e
*
Cas9-Streptavidin Correlation Coefficient
Cas9
sgRNA
S1m-sgRNA-1
Streptavidin
S1mplex

Figure 3.2 | **Measuring and visualizing S1mplexes *in vitro* and within cells.**

(a) Dynamic light scattering of ssODN-S1mplex particle assembly. Cas9 (orange) and streptavidin (blue) proteins fail to assemble when in solution together and have a hydrodynamic radius consistent with published data. The addition of sgRNA to Cas9 protein increases the radius of the RNP particle to 10 nm (yellow). This radius does not change with the addition of streptavidin (red) and free streptavidin can be detected (red arrowhead). When S1m-sgRNAs are added to Cas9 (purple), the radius is increased by a larger amount than sgRNAs, potentially due to the larger size of the S1m-sgRNA. When streptavidin is added to this complex (green), a shift in size of about 3 nm occurs (green arrowhead), the size of streptavidin. A second peak at 35 nm (green asterisk) may be associated with multiple Cas9-S1m-sgRNA complexes connected to a single streptavidin. See also Figure B-1c. **(b)** Two representative single cell multispectral flow cytometric images of S1m-sgRNA and sgRNA transfected cells with Cas9 immunohistochemistry and fluorescent streptavidin (scale bar: 10 μm). Arrowheads indicate presence of overlapping colors. Numbers in yellow are measured log Pearson correlation coefficient as determined by IDEAS software (See also Figure B-2). **(c)** Correlation coefficient of Cas9 immunocytochemistry fluorescent signal and streptavidin fluorescence, as measured by multispectral image cytometry within hPSCs. Use of S1m-sgRNA significantly increased the correlation between the two signals (***$p<10^{-5}$, Student's two-tailed t-test). **(d)** Representative confocal images of S1m-sgRNA and sgRNA transfected cells with Cas9 immunohistochemistry and fluorescent streptavidin (scale bar: 5μm). Arrowheads indicate presence of overlapping colors. **(e)** Correlation coefficient of Cas9 immunocytochemistry and streptavidin fluorescence inside the nuclei of transfected cells. Introduction of S1m-sgRNAs significantly increased the correlation between the two molecules (*$p<0.05$, Student's two-tailed t-test).

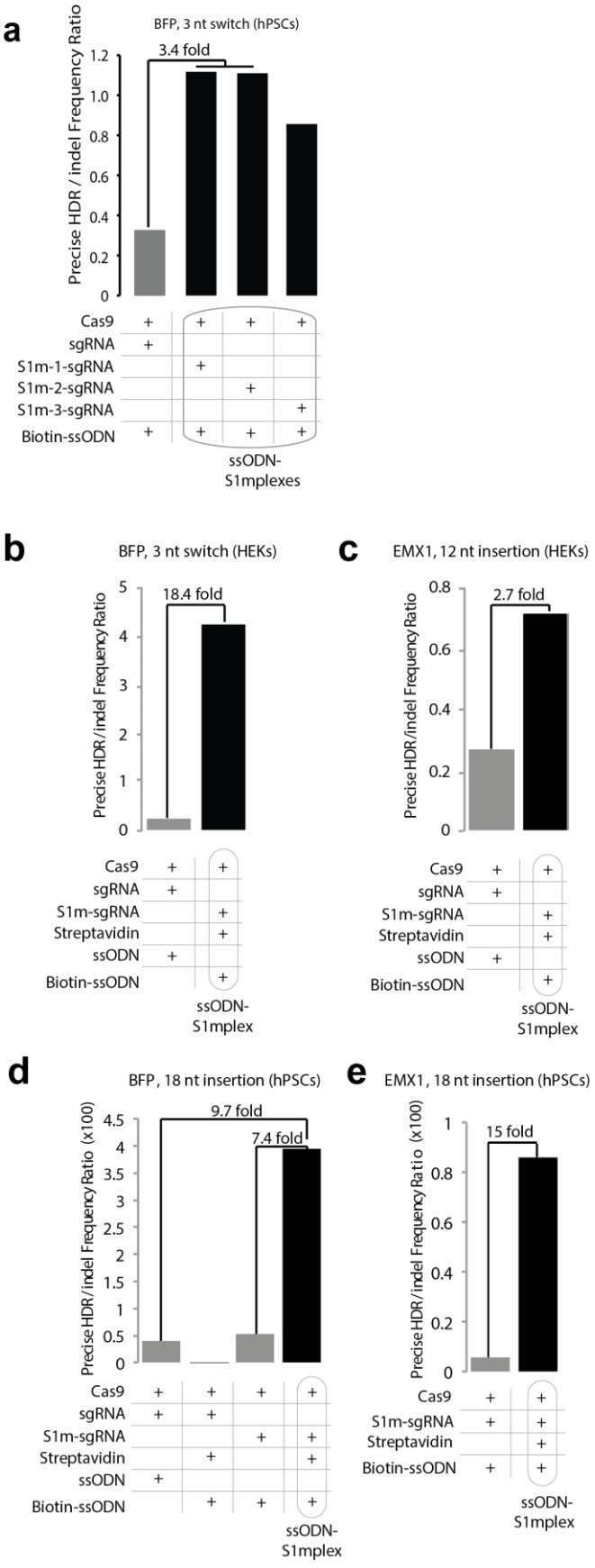

a
BFP, 3 nt switch (hPSCs)
3.4 fold
Precise HDR / indel Frequency Ratio
1.2
1.0
0.8
0.6
0.4
0.2
0
Cas9
sgRNA
S1m-1-sgRNA
S1m-2-sgRNA
S1m-3-sgRNA
Biotin-ssODN
ssODN-S1mplexes
b
BFP, 3 nt switch (HEKs)
18.4 fold
Precise HDR/indel Frequency Ratio
5
4
3
2
1
0
Cas9
sgRNA
S1m-sgRNA
Streptavidin
ssODN
Biotin-ssODN
ssODN-S1mplex
c
EMX1, 12 nt insertion (HEKs)
2.7 fold
Precise HDR/indel Frequency Ratio
0.8
0.6
0.4
0.2
0
Cas9
sgRNA
S1m-sgRNA
Streptavidin
ssODN
Biotin-ssODN
ssODN-S1mplex
d
BFP, 18 nt insertion (hPSCs)
9.7 fold
7.4 fold
Precise HDR/indel Frequency Ratio (x100)
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Cas9
sgRNA
S1m-sgRNA
Streptavidin
ssODN
Biotin-ssODN
ssODN-S1mplex
e
EMX1, 18 nt insertion (hPSCs)
15 fold
Precise HDR/indel Frequency Ratio (x100)
1
0.8
0.6
0.4
0.2
0
Cas9
S1m-sgRNA
Streptavidin
Biotin-ssODN
ssODN-S1mplex

Figure 3.3 | ssODN-S1mplexes: some assembly required to increase ratio of HDR to indels in human cells.

(a) Ratio of precise to imprecise editing using S1mplexes formed with different S1m-sgRNA variants in hPSCs. Each S1m-sgRNA increased the ratio of precise to imprecise editing when compared to sgRNAs. S1mplexes with S1m-sgRNA-1, and S1m-sgRNA-2 had the highest ratios of precise editing. **(b)** Ratio of precise to imprecise editing at *BFP* locus. ssODN-S1mplexes had an 18.4-fold higher ratio than sgRNAs and contained four precise edits to every one indel as analyzed by deep sequencing 8 days post lipofection of HEKs. **(c)** Ratio of precise to imprecise editing at *EMX1* locus. ssODN-S1mplexes had a 2.7-fold higher ratio than sgRNAs. **(d)** Ratio of precise insertions to imprecise indels at *BFP* locus in hPSCs as analyzed by deep sequencing. ssODN-S1mplexes had a 9.7-fold increase in comparison to standard sgRNAs and a 7.4-fold increase when compared with untethered ssODNs. **(e)** Ratio of precise insertions to imprecise indels at *EMX1* locus. Addition of streptavidin to S1mplex resulted in a 15-fold increase in the ratio of precise insertions to imprecise indels. See also Table B-9.

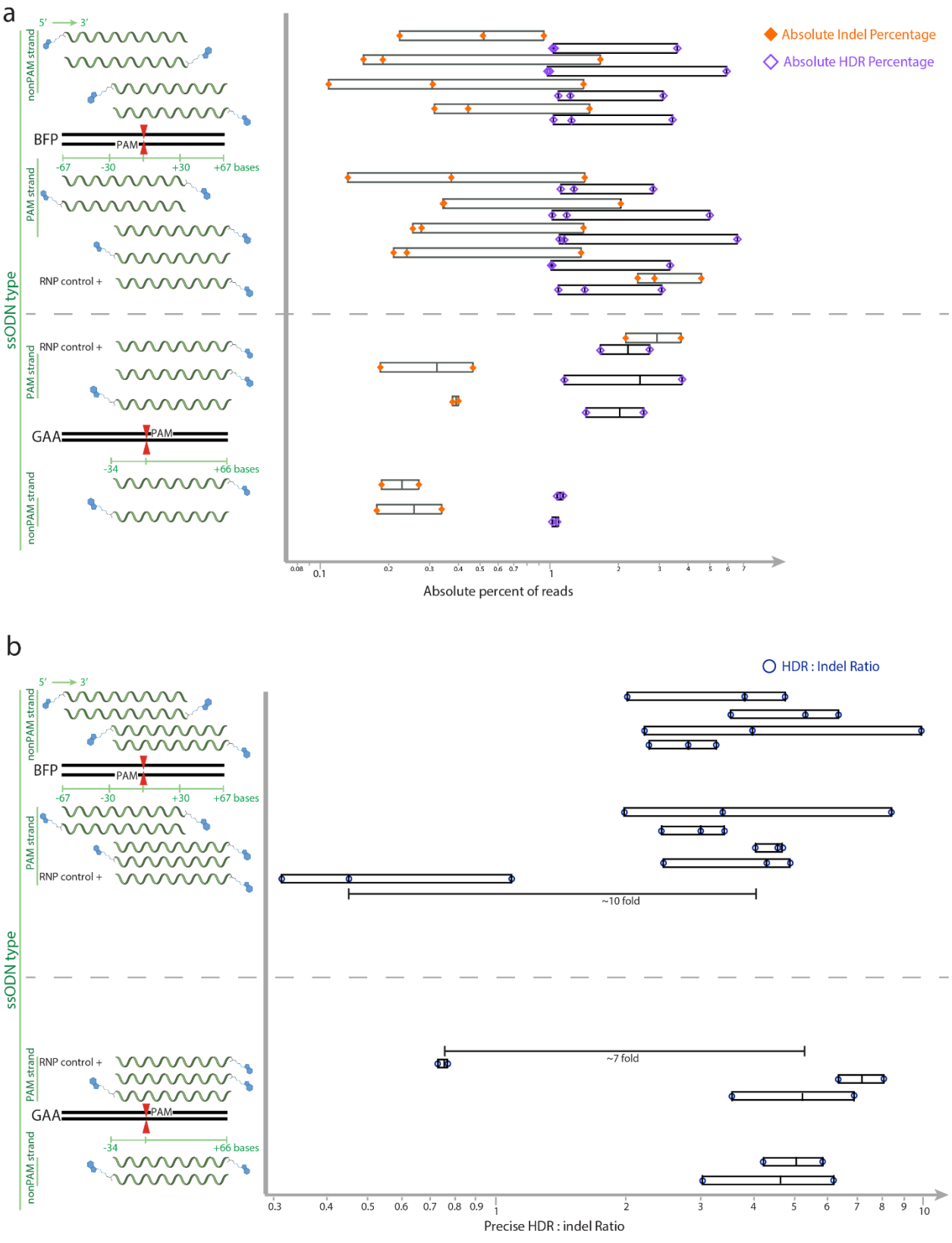

a
5′ → 3′
nonPAM strand
BFP
PAM
-67
-30
+30
+67 bases
PAM strand
RNP control +
ssODN type
RNP control +
PAM strand
GAA
PAM
-34
+66 bases
nonPAM strand
Absolute Indel Percentage
Absolute HDR Percentage
0.08
0.1
0.2
0.3
0.4
0.5
0.6
0.7
1
2
3
4
5
6
7
Absolute percent of reads
b
HDR : Indel Ratio
~10 fold
~7 fold
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
2
3
4
5
6
7
8
9
10
Precise HDR : indel Ratio

Figure 3.4 | ssODN asymmetry, strandedness and tether location minimally affects S1mplexes gene editing.

Left: ssODN design. Genomic sequence is denoted with black bars. sgRNA targeting site and PAM is denoted by 'PAM' inside genomic locus, while red triangles are the sgRNA cut site. ssODN length is measured around cut site either upstream (-) or downstream (+) as read by the reading frame. Biotin (blue hexagon) was attached to either the 5′ or 3′ end of the ssODN. ssODNs were identical in sequence to either the PAM or Non-PAM sequence as read in a 5′-3′ direction. RNP controls were standard sgRNAs plus corresponding ssODN. (**a)** Absolute NHEJ (orange diamonds) and HDR percentages (purple diamonds) as a function of total reads at two different loci in hPSCs using different ssODN designs. Each symbol represents a single replicate analyzed by deep sequencing 4 days after nucleofection into hPSCs. HDR levels were generally higher in each replicate than NHEJ levels. See also Table B-9. (**b)** Ratio of HDR:indel reads in deep sequencing using each ssODN combined with S1mplexes. Blue circles represent individual biological replicates. With each ssODN, S1mplexes increased the ratio of HDR:indel when compared to sgRNA controls but no significant trends as to symmetry, sidedness, or biotin location were observed.

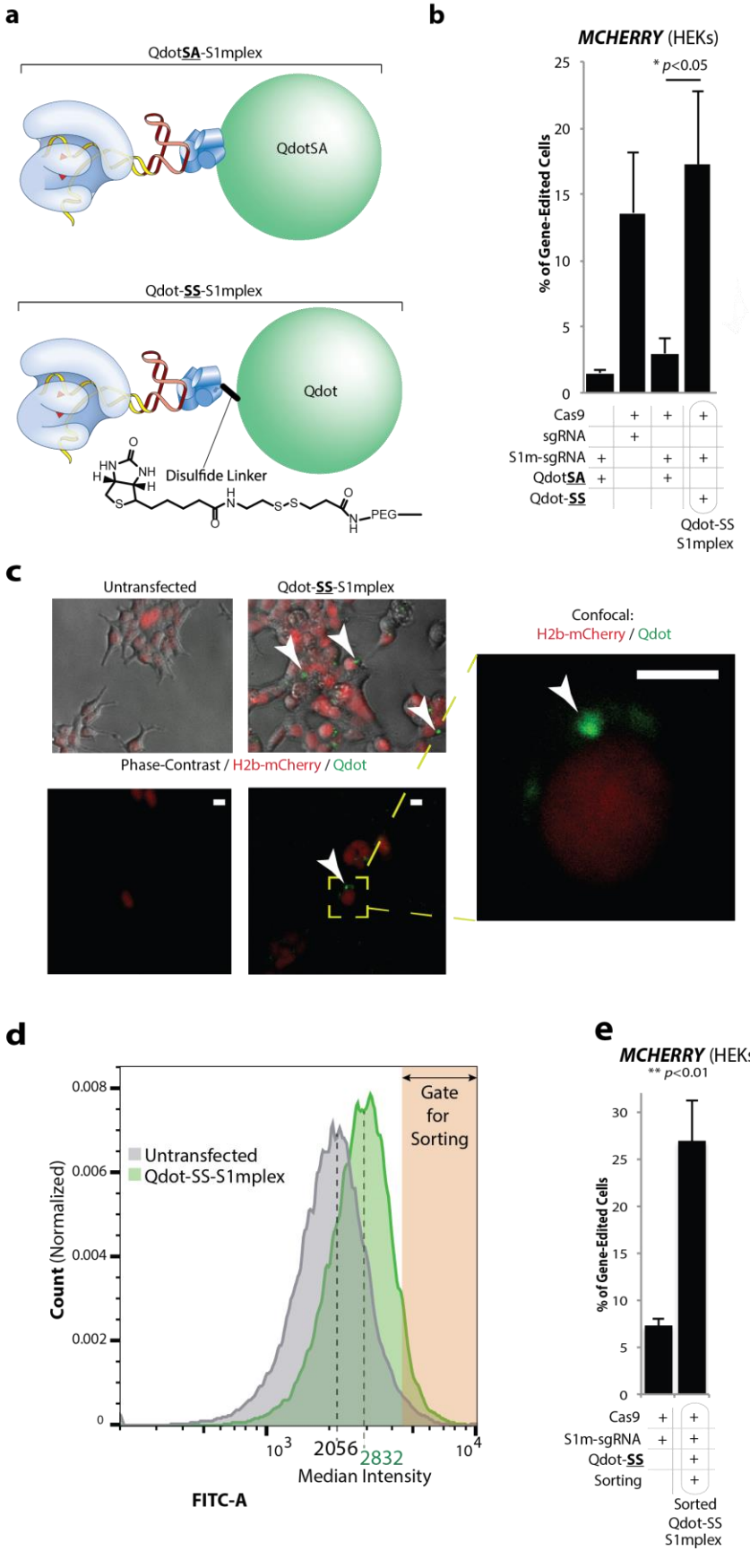

a
QdotSA-S1mplex
QdotSA
Qdot-SS-S1mplex
Qdot
Disulfide Linker
PEG
b
MCHERRY (HEKs)
* p<0.05
% of Gene-Edited Cells
25
20
15
10
5
0
Cas9 + + +
sgRNA +
S1m-sgRNA + + +
QdotSA + +
Qdot-SS +
Qdot-SS S1mplex
c
Untransfected
Qdot-SS-S1mplex
Phase-Contrast / H2b-mCherry / Qdot
Confocal:
H2b-mCherry / Qdot
d
Gate for Sorting
Untransfected
Qdot-SS-S1mplex
Count (Normalized)
0.008
0.006
0.004
0.002
0
10^3
2056
2832
10^4
Median Intensity
FITC-A
e
MCHERRY (HEKs)
** p<0.01
% of Gene-Edited Cells
30
25
20
15
10
5
0
Cas9 + +
S1m-sgRNA + +
Qdot-SS +
Sorting +
Sorted Qdot-SS S1mplex

Figure 3.5 | Addition of Qdots with degradable linkers to identify S1mplex-positive cells. **(a)** Schematic of S1mplexes with quantum dot cargoes. Qdots can be complexed with the S1mplex by streptavidin covalently attached directly to the quantum dot (QdotSA-S1mplex, top), or a cell-labile disulfide linker (Qdot-SS-S1mplex, top). The quantum dot has a mean diameter of 20 nm. **(b)** Gene editing comparison of different Qdot S1mplexes. Gene editing of HEK H2B-mCherry reporter cells five days post sorting as assayed by flow cytometry. DSBs within *MCHERRY* transgene repaired by NHEJ will generate indels near the fluorophore resulting in a loss of fluorescence. Percent gene editing refers to percent of cells without mCherry fluorescence. QdotSA interferes with RNP activity, while Qdot-SS has equivalent gene editing activity as the free RNP (n=3 technical replicates). **(c)** Representative epifluorescence images of untransfected and Qdot-SS-S1mplex transfected cells 24 hours post transfection (Scale bar: 10 μm). Arrowheads indicate Qdot fluorescence in the cytoplasm. **(d)** Increased fluorescence of Qdot-S1mplex allows sorting out of quantum dot positive fractions compared to untransfected cells 24 hours post transfection. **(e)** Quantum dot conjugation to S1mplex via a cleavable disulfide linker allows fluorescent enrichment of gene-edited human cells. (n=3 biological replicates).

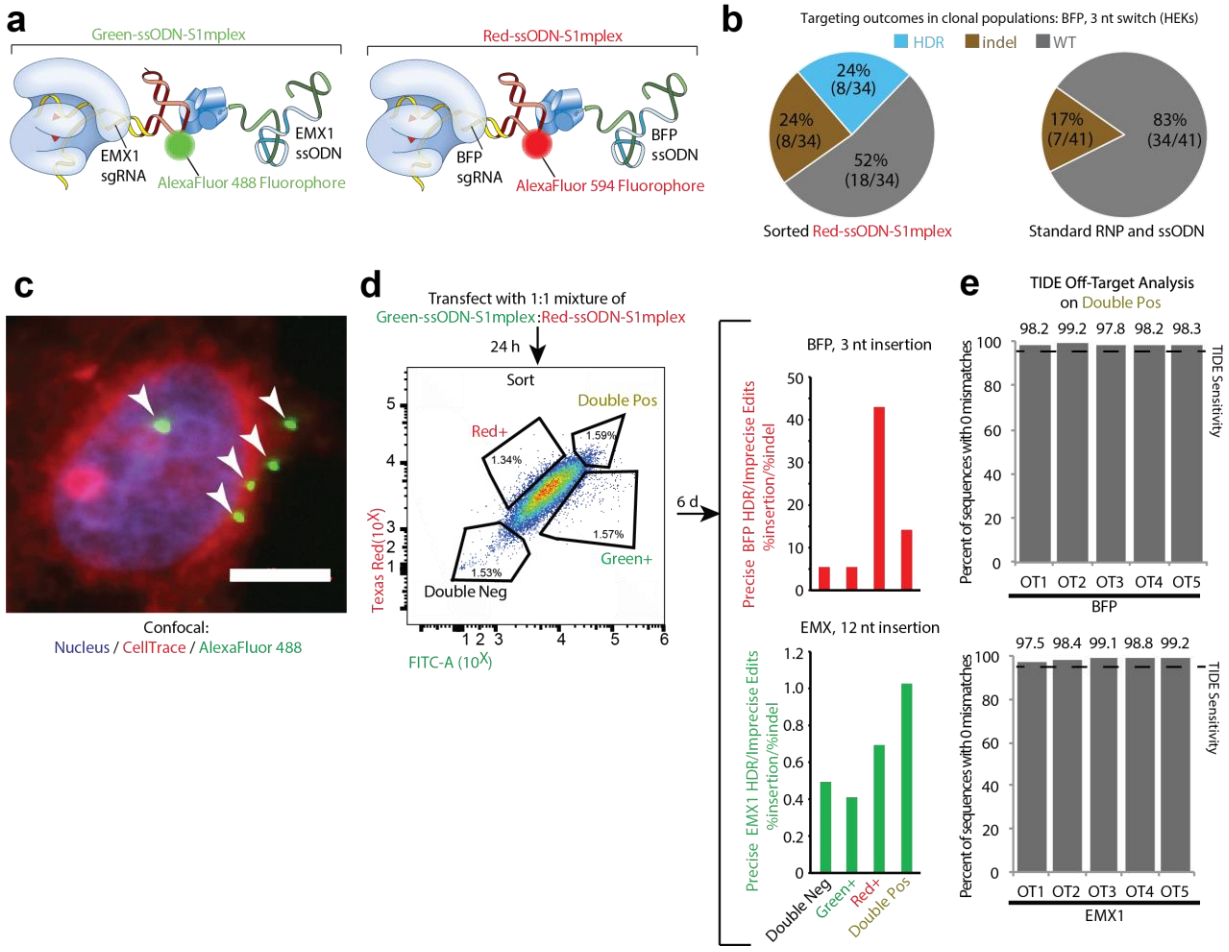

a
Green-ssODN-S1mplex
EMX1 sgRNA
AlexaFluor 488 Fluorophore
EMX1 ssODN
Red-ssODN-S1mplex
BFP sgRNA
AlexaFluor 594 Fluorophore
BFP ssODN
b
Targeting outcomes in clonal populations: BFP, 3 nt switch (HEKs)
HDR
indel
WT
24% (8/34)
24% (8/34)
52% (18/34)
Sorted Red-ssODN-S1mplex
17% (7/41)
83% (34/41)
Standard RNP and ssODN
c
Confocal:
Nucleus / CellTrace / AlexaFluor 488
d
Transfect with 1:1 mixture of
Green-ssODN-S1mplex:Red-ssODN-S1mplex
24 h
Sort
Double Pos
Red+
Green+
Double Neg
1.59%
1.34%
1.57%
1.53%
Texas Red(10X)
FITC-A (10X)
6 d
BFP, 3 nt insertion
Precise BFP HDR/Imprecise Edits %insertion/%indel
EMX, 12 nt insertion
Precise EMX1 HDR/Imprecise Edits %insertion/%indel
Double Neg
Green+
Red+
Double Pos
e
TIDE Off-Target Analysis
on Double Pos
98.2 99.2 97.8 98.2 98.3
Percent of sequences with 0 mismatches
TIDE Sensitivity
OT1 OT2 OT3 OT4 OT5
BFP
97.5 98.4 99.1 98.8 99.2
EMX1

Figure 3.6 | Fluorescent S1mplexes to enrich for gene edited cells.

(a) Schematic of simultaneous editing at two loci strategy. HEK cells were transfected simultaneously with two S1m particles, labeled with distinct fluorophores. Editing at *BFP* locus was associated with Red-ssODN-S1mplexes (AlexaFluor-594 fluorophore), while editing at *EMX1* locus was associated with Green-ssODN-S1mplexes (AlexaFluor-488 fluorophore). **(b)** Single cells sorting for enrichment of editing at *BFP* locus. In enriched S1mplex clonal populations, indels (brown) and HDR events occurred in a 1:1 ratio. In sgRNA clones, all isolated clones either had indel or wildtype genotypes. Genotypes were assayed by Sanger sequencing. No mosaic genotypes were observed. **(c)** Fluorescent S1mplexes inside the cell using confocal microscopy. Arrows denote Green-S1mplex both inside the nucleus and outside the cell (Scale bar: 10 μm). **(d)** Twenty-four hours post transfection, cells were sorted into populations that were positive for either fluorophore, both or neither. Analysis via deep sequencing was done 6 days post sorting. Top: ratio of precise (perfect sequence match to ssODN) to imprecise editing (indels) in sorted populations. Populations enriched for *BFP* targeted S1mplexes (Red+ and double positive) had elevated ratios up to 40 times as many insertions as indels. *Bottom:* ratio of precise to imprecise editing in sorted populations. Populations enriched for *EMX1* targeted S1mplexes (Red+ and double positive) had elevated ratios of precise insertions to indels. See also Table B-9. **(e)** Off-target analysis of double positive populations using TIDE at the top 5 off-target locations for each sgRNA. No modifications were detected below the TIDE limit of detection (dotted line).

CHAPTER 4: Double correction of compound heterozygote mutations by CRISPR-Cas9 gene editing efficiently rescues Pompe disease within a patient-derived stem cell model

Work in this chapter was adapted from:

Double correction of compound heterozygote mutations by CRISPR-Cas9 gene editing efficiently rescues Pompe disease within a patient-derived stem cell model

Jared Carlson-Stevermer, David Fiflis, Heidi Kletzein, Lucy Kohlenberg, Madelyn Goedland, Nadine Connor, and Krishanu Saha

Unpublished as of August 2018

Abstract

Autosomal recessive diseases frequently have diverse genetic variants, reflecting many variable mutational processes and selection pressures operating on the human population. For these diseases, compound heterozygous mutations are frequently observed within patients. Genomic medicine has thus far focused on viral gene augmentation therapy without directly modifying the two diseased alleles. Due to the challenge of producing multiplexed gene edits and the recessive nature of each allele, it has been assumed that gene editing therapies need only to correct a single mutant allele. Using CRISPR-Cas9 gene editing, here we demonstrate that gene correction of both mutant alleles within an autosomal recessive case of Pompe disease rescues the disease phenotype more effectively than correcting single mutant alleles. Gene-corrected *GAA* alleles are strongly expressed from the endogenous promoter, generating functional GAA enzyme within patient-derived induced pluripotent stem cells (iPSCs) and iPSC-differentiated cardiomyocytes. Edited cells had normal karyotypes, no detectable off-target mutations or large genetic deletions, and lower amounts of abnormal glycogen storage. Cross-correction of the diseased cells from GAA produced within edited cells was robust, and this cross-correction was equal to standard enzyme replacement therapy (ERT). These results suggest that autologous cell therapy may be feasible for autosomal recessive disorders using double-gene corrected iPSCs. Further, this work indicates that there are twice the number of alleles that could be targeted by somatic gene editing strategies, and that precise correction using repeated dosing of genome editors to correct all mutated alleles may be an effective therapeutic strategy.

Introduction

For individuals with an autosomal recessive disease (e.g., cystic fibrosis, sickle cell anemia, and Tay-Sachs disease), the mutated gene is located on one of the non-sex chromosomes (autosomes), and both alleles of the gene carry mutations. The parents of an

individual with complications from an autosomal recessive disease generally each carry one copy of the mutated gene, but do not present obvious symptoms of the disease. Therefore, it has been assumed that correcting only one allele of the mutated gene would be sufficient to rescue the disease, and that additional genomic surgery to repair the second allele may subject a patient to undue risk. Here we test this assumption with CRISPR-Cas9 gene editing to systematically correct both mutated alleles within the same cell from an autosomal recessive, infantile-onset case of Pompe disease.

Infantile-onset Pompe disease is an autosomal recessive glycogen storage disorder caused by mutations in the acid-α-glucosidase (*GAA*) gene that encodes an enzyme that breaks down glycogen within the lysosome[219] (**Figure 4.1a**). Over 400 different *GAA* mutations have been noted within ClinVar, and detailed case studies indicate a buildup of glycogen, leading to clinical complications, most prominently in cardiac and muscle tissues[107]. Left untreated, patients with infantile-onset Pompe disease typically die within the first year of life, and Pompe Disease is now frequently included within newborn screening panels[220]. Enzyme replacement therapy (ERT) using recombinant human GAA (rhGAA) is currently the only approved clinical treatment for Pompe disease[111,114]. However, patients require high levels of enzyme injected biweekly, rendering the treatment expensive and inconvenient[107]. ERT may also be less effective in a subset of patients that are cross-reactive immunologic material (CRIM) negative[221].

Newer therapies in development for Pompe disease have primarily made use of integrated viral cassettes including transgenes to express a normal GAA coding sequence from exogenous promoters. Viral particles are injected either directly into muscle or administered systemically and transported to the liver. Infected cells can generate phenotype rescue (10-fold reduction in glycogen content), but fail to correct many non-transduced cells. These therapies also require high viral loads (>10^{10} viral genomes/kg)[116,119–121]. Silencing of the viral transgene, immune response to the viral vector, and insertional oncogenesis are outstanding concerns with

these viral gene therapy approaches. Anti-sense oligonucleotides can also be introduced to correct splicing in diseased patients that possess mutations at splicing sites, but would only be beneficial to a subset of potential patients[126]. Finally, autologous cell therapy has been proposed using cells engineered to constitutively overexpress GAA[222]. None of these approaches retain endogenous *GAA* regulation nor have corrected the underlying *GAA* mutations. Correction of the mutated alleles could be a promising *in vivo* somatic gene editing strategy or alternatively, a strategy to generate gene corrected cells *ex vivo* for cell therapy.

Results

Simultaneous biallelic correction of Pompe disease iPSCs.

To explore whether multiple corrected alleles within the same cell have an epistatic effect, several clonal isogenic iPSC lines were generated by CRISPR-Cas9 gene editing of an iPSC line derived from a patient with infantile-onset Pompe disease[196]. In the cell line of interest, these compound heterozygous *GAA* mutations responsible for the disease phenotype are a deletion of a thymidine nucleotide at position 1441 (GAA:c.[1441delT], "1441delT") causing a frameshift, and premature stop codon on one allele, and a G>A conversion at nucleotide 2237 (GAA:c.[2237G>A], "2237G>A") forming an immediate stop codon on the other (**Figure 4.1b, Figure C-1a**). As mutant *GAA* alleles are on different exons, and separated by thousands of base pairs, using a single double strand break (DSB) with homology directed repair from a long plasmid or viral donor would likely be inefficient[184]. We therefore decided to use a strategy utilizing two distinct Cas9 RNPs with accompanying single stranded oligonucleotide (ssODN) templates encoding the gene correction (**Figure 4.1b, Table C-1, C-2**). We designed sgRNAs that were specific to the mutant allele by only evaluating those that contained the mutant site within the seed region of the sgRNA[19]. We also exploited a RNA aptamer based approach developed in our lab, the S1mplex[223], that was previously shown to promote precise correction

of alleles at many loci, including the GAA locus. As identifying double-corrected cells can be challenging, we complexed the specific sgRNA and ssODN pairs together as well as included a fluorophore as a marker for the presence of S1mplexes within the nucleus. Using a high-content analysis platform developed in our lab, ArrayEdit[190], we were able to track the presence of S1mplexes within the nucleus, growth rate of clones, as well as screen the pH of the lysosome (**Figure 4.1c**). This phenotypic characteristic was profiled due to the nature of Pompe disease where a buildup of glycogen in the lysosome results in a neutralization of this acidic organelle. Using this strategy, we first determined it was possible to distinguish between healthy and diseased iPSCs using LysoSensor dye (**Figure C-1b, c**). After transfection with S1mplexes, we plated compound heterozygote iPSCs on ArrayEdit and acquired images for phenotypic analysis each day (**Figure C-1d**). Using high content analysis on ArrayEdit (see **Text C-1** for details), we isolated cell lines that were corrected at the 1441delT allele and the 2237G>A allele individually (**Figure 4.1d**). These lines were termed heterozygotes and are henceforth referenced to as '[loci] corrected.' We also isolated a clone corrected at both GAA:c.[1441delT];[2237G>A] alleles which we termed a homozygote and named 'double corrected' (**Figure 4.1d**). Importantly, using ArrayEdit we were able to screen out any cell lines that harbored indels from imprecise DSB repair on either allele.

All lines remained pluripotent (**Figure 4.1e**, **Figure C-2a**), and after karyotyping each of the isolated lines, we observed no large transversions or inversions in any of the isolated lines (**Figure C-2b**). Because Cas9 can create large, unexpected indel mutations[224], we also conducted an 8 kb PCR on *GAA* that included both sgRNA targets and observed no genomic deletions between the DSBs (**Figure C-2c, d**). Sequencing of these large PCR amplicons confirmed that both alleles were present and no other sequence abnormalities were detected at the edited loci (**Figure C-2e**). Finally, chromatograms from Sanger sequencing at the top three off-target sites for each sgRNA matched the untransfected, patient-derived cell line, indicating

that none of the off-target regions were modified by our editing strategy (**Figure C-3, Table C-4**).

Regulated expression of GAA transcripts.

To evaluate whether the corrected loci are correctly expressed, we conducted qPCR at the 3` end of the GAA mRNA transcript as well as around each edited locus (**Figure 4.2a**). We observed that while compound heterozygous mutants expressed the lowest levels of *GAA* when compared to internal *GAPDH* levels (**Figure 4.2a, Figure C-4a**), there was still full-length, mature mRNA that could be used to express protein. As expected, both heterozygous correction lines displayed a recovery phenotype that expressed *GAA*. However, when we compared the heterozygous mutants to the homozygous corrected line we saw that the homozygous corrected line contained approximately a 2-fold increase in *GAA* transcripts. This supports our hypothesis that correction of both mutants may be beneficial to produce secreted *GAA* to induce cross-correction of unedited, diseased cells. We also conducted qPCR at each of edited loci and confirmed that homozygous cell lines consistently produced greater amounts of mature mRNA than any other condition (**Figure 4.2a, Figure C-4a**).

To determine the balance of corrected to diseased mutants within the mRNA from our edited lines, we conducted deep sequencing on endpoint RT-PCR samples and analyzed all lines at both corrected loci, looking for the presence of disease variants and PAM wobbles from our ssODN. In the compound heterozygous disease line we saw that at both loci, there was approximately a 1:1 balance in expression from the diseased and healthy alleles (**Figure C-4b**). This was expected, as mRNA from both mutant alleles would be expressed and similarly regulated. We then looked at the heterozygous lines and observed that the corrected locus was expressed at a higher level (~5-fold increase) than the relative unedited locus (**Figure C-4b**). This finding was supported by conducting deep sequencing at both edited loci, as sequence

data can be linked together to form a single mRNA transcript. Finally, when we looked at the homozygous corrected cells we observed data similar to that of the compound heterozygous mutant line in that both alleles are similarly expressed (**Figure C-4b**). However, when this data is overlaid with the qPCR data described previously, we can see that both alleles are expressed individually at higher levels (3-5 fold increase) than compound heterozygotes and each allele is expressed similarly to the corresponding heterozygous correction (**Figure 4.2a**). These findings suggest nonsense mediated decay of the mutant transcript[225]. At the protein level, we were able to detect high levels of mature protein using a western blot (**Figure 4.2b**). GAA was detected at the expected size in corrected lines and expression was identical to a control hPSC line. We were also able to identify precursor polypeptides, which are important to protein secretion[226], showing the edited *GAA* locus was being correctly translated and processed within edited cells. Notably, we were able to detect only small amounts of GAA protein and precursor polypeptides in the compound heterozygote iPSCs.

Complete phenotypic recovery by *GAA* correction.

After we confirmed that the molecular biology signature of the corrected lines was accurate, we explored the ability of the edited cells to clear glycogen signifying a return to a healthy phenotype. First, we obtained protein lysates from all iPSC lines and incubated them with 4-MUG. Active GAA protein cleaves this substrate and releases the fluorophore 4-MU that can be detected with a standard plate reader. As expected, we observed that the compound heterozygous cells had minimal GAA activity. Interestingly, when activity of the heterozygous and homozygous correction lines was measured all lines had approximately the same cleavage ability, consistent with the lack of symptom presentation in Pompe disease carriers (**Figure 4.2c**). As an internal assay control, we also measured neutral GAA activity and saw no difference between any of the lines, including the compound heterozygotes (**Figure C-4c**). More

importantly, we also tested the GAA activity of spent media after 24 hours of culture. Under these conditions, the homozygous lines had significantly more secretion than either heterozygous lines or a control hPSC line (**Figure 4.2d, C-4d**).

Cross-correction by GAA secretion in cardiomyocytes.

Secreted GAA may have potential benefits for a cell therapy approach where implanted, corrected tissues can rescue surrounding, unedited cells. Because Pompe disease has a significant effect on cardiac tissue within infants affected by Pompe disease, we first differentiated all iPSC lines to cardiomyocytes, using a previously described small-molecule inhibitor protocol[227]. For all, differentiations, we observed spontaneous contraction and confirmed the expression of α-actinin, a marker of cardiac lineage commitment (**Figure C-5a**). Similar to results seen in the iPSC state, differentiated correction lines still expressed and secreted active GAA, as indicated in a 4-MUG cleavage assay on cardiomyocyte protein lysates and spent culture media (**Figure C-5b-e**). It has previously been demonstrated that by culturing in medium devoid of glucose, Pompe iPSC cardiomyocytes display an accumulation of glycogen within the lysosome[196]. To test for cross-correction, we performed a medium exchange experiment wherein we took partially spent, glucose-free medium from each corrected line and used it to replace glucose-free medium on compound heterozygotes (**Figure 4.2e**). One day after this media exchange, cells were stained with LysoSensor and imaged using confocal microscopy to measure lysosome acidity as a proxy for glycogen clearance. As a control we also added rhGAA from standard ERT to compound heterozygote cultures. We again analyzed each cell individually using high content analysis and found that after 24 hours neither of the wells containing media from the heterozygote lines differed significantly from the compound heterozygote. As would be expected from the effectiveness of ERT, when compound heterozygotes were supplemented with 10 nM rhGAA, LysoSensor intensity increased

suggesting a clearance of glycogen from the lysosome. Interestingly, we also observed that when media from double corrected homozygotes or a control hPSC line was used, the recovery effect was even greater than seen using rhGAA (**Figure C-5f**). Glycogen clearance was measured again after 96 hours in culture to test if there was a cumulative effect of culture in GAA-positive media. In these conditions we found that media from heterozygote corrections did have a positive effect on glycogen clearance compared to compound heterozygotes and was similar to media from control hPSCs. Importantly, cells cultured in media from double corrections exhibited a similar glycogen clearance to those supplemented with rhGAA (**Figure 4.2f, g**). This clearance is expected to continue until normal levels of glycogen are reached[228]. To confirm this result, we looked at lysosomal size of compound heterozygotes in GAA-positive media through visualization of LAMP-1. In media from compound heterozygotes, lysosomes were enlarged, consistent with buildup of glycogen (**Figure 4.2h**). In comparison, when media was taken from double corrections or supplemented with rhGAA, lysosome size decreased and appeared as punctae. Compound heterozygotes with media from heterozygote corrections fell between these two extremes with some punctae lysosomes while others remained enlarged. Taken together, the double corrected cells cross-correct and rescue diseased cardiomyocytes quickly and effectively.

Discussion

Gene augmentation therapy by viral overexpression has long been assumed to be the most practical approach to genome surgery to address autosomal recessive disorders, especially compound heterozygous cases. In cells from Pompe diseased patients, our results indicate that either gene corrected allele is robust expressed, translated and active, and that there are additive effects to correcting both alleles. Further, many of the common Pompe disease mutations can be targeted in an allele-specific manner using Sp.Cas9 strategies (**Table**

C-6). In our strategy, transcriptional regulation is driven by the endogenous promoter, potentially correcting a number of different isoforms for GAA[229,230]. The targets for somatic cell genome editors therefore could expand from the traditional foci of liver and muscle to other tissues that may use alternate *GAA* isoforms. In contrast, in the gene augmentation approach, all cells must process a single isoform. Further, silencing from synthetic or viral elements has been observed for gene therapies such as in our hands with targeted knockin strategies that overexpress a transgene via a synthetic promoter (**Figure C-5f**). Transgene silencing raises concerns about the durability of viral gene therapies and proposed cell therapies where GAA is overexpressed from a safe harbor locus[222].

In our strategy, post-translational processing of the enzyme also appears to be intact, as the distribution of processed GAA is identical to healthy controls. In contrast, *GAA* overexpression in mammalian cells can cause cellular stress, leading to differential trafficking and processing of the nascent translated peptide[226]. The correction strategy could also employ lower titers of virus that encode the genome editor or avoid insertional oncogenesis altogether using a non-viral approach for delivery of the genome editor. Challenges with a somatic cell genome editing strategy include off target effects and other chromosomal rearrangements that could be induced by cutting the genome. Correction efficiencies may also only be lower in post-mitotic cells with some genome editors. Despite these challenges, we argue that genomic medicine should consider correcting both alleles in autosomal recessive conditions, because there could be nonlinear additive therapeutic effects within corrected cells. In diseases where non-cell autonomous mechanisms can be therapeutic (e.g., lysosomal disorders with cross-correction), these effects are likely important and suggest that correction of all alleles within a particular patient - or target tissue within the patient - should be the goal of genome editing strategies. Multiple dosing with safe and precise genome editors therefore be developed for a

greatly expanded set of targets for diseases with compound heterozygous and complex genetic causes.

Methods

Cell Culture. All hPSCs were maintained in mTeSR1 medium on Matrigel (WiCell) coated tissue culture polystyrene plates (BD Falcon). Cells were passaged every 4-5 days at a ratio of 1:8 using Versene solution (Life Technologies). Patient derived human induced pluripotent stem cell line Pompe GM04192 was a gift from the T. Kamp and M. Suzuki (UW-Madison) labs. Cardiomyocytes derived from hPSC and iPSC cultures were maintained in RPMI/B27 on Matrigel (WiCell) coated polystyrene plates (BD Falcon). All cells were maintained at 37°C in 5% CO_2, and tested for mycoplasma contamination on a monthly basis.

Creation of S1m-sgRNAs. S1m-sgRNAs were synthesized as previously described[223]. S1m gBlocks were annealed with Phusion polymerase (New England Biolabs) under the following thermocycler conditions: 98°C for 30 sec followed by 30 cycles at 98°C for 10 s, and 72°C for 15 s with a final extension at 72°C for 10 minutes. S1m cDNA was annealed with Phusion polymerase (New England Biolabs) under the following thermocycler conditions: 98°C for 30 sec followed by 30 cycles at 98°C for 10 s, 60°C for 10 s, and 72°C for 15 s with a final extension at 72°C for 10 minutes. In vitro transcription was performed with the MEGAShortscript T7 Kit (Thermo Fisher Scientific) according to manufacturer's instructions.

RNP Delivery. All hPSC transfections were performed using the 4D-Nucleofector System (Lonza) in P3 solution using protocol CA-137. 50 pmol Cas9, 60 pmol sgRNA, 50 pmol streptavidin, and 60 pmol ssODN were used to form particles per ssODN-S1mplex as described above. Cells were then harvested using TrypLE (Life Technologies) and counted. $2x10^5$ cells

per transfection were then centrifuged at 100x*g* for 3 minutes. Excess media was aspirated and cells were resuspended using 20 µL of RNP solution per condition. After nucleofection, samples were incubated in nucleocuvettes at room temperature for 15 minutes prior to plating into $2x10^4$ cells per well on ArrayEdit in mTeSR media+10 µM ROCK inhibitor. Media was changed 24 hours post transfection and replaced with mTeSR medium.

Synthesis of ArrayEdit Platform. µCP was performed using previously described methods[148]. The surface modification involved printing of an alkanethiol initiator to nucleate the polymerization of hydrophilic poly(ethylene glycol) (PEG) chains. Briefly, double sided-adhesive was attached to the bottom of a standard tissue culture plate, after which a laser cutter was used to cut out the well bottoms. Using previously described chemistry[150], patterns were transferred to gold-coated glass via a polydimethylsiloxane stamp after which the glass was submerged in a poly(ethylene glycol) (PEG) solution overnight to build PEG chains surrounding µFeatures. Standard tissue culture plates with well bottoms cut out were then fastened to processed sheets using a custom-made alignment device.

High-Content Analysis. Automated microscopy was performed using a Nikon Eclipse TI epifluorescent scope. A 15x15 grid with one µFeatrure per image was established and maintained so that each feature imaged was consistent each day. Nikon Perfect Focus was used to ensure that all colonies were in the same Z-plane and LysoSensor intensity was measured accurately. Images were processed using CellProfiler[152] to count the number of nuclei and quantify LysoSensor intensity.

Genomic Sequencing. DNA was isolated from cells using DNA QuickExtract (Epicentre) following TrypLE treatment and centrifugation. Extracted DNA was incubated at 65°C for 15 min, 68°C for 15 min, and 98°C for 10 min. Genomic PCR was performed using AccuPrime HiFi

Taq (Life Technologies) and 500 ng of genomic DNA according to manufacturer's instructions. Long (8kp) PCR reactions were thermocycled using an extension step of 10 minutes. All genomic PCR products were then submitted to the University of Wisconsin-Madison Biotechnology Center for DNA sequencing.

RT- and qPCR. RNA was isolated from cells using RNA QuickExtract (Epicentre) following manufacturer's protocol. 100 ng of extracted RNA was reverse transcribed using Superscript IV Reverse Transcriptase (Invitrogen). Endpoint PCR amplification of cDNA product was performed following manufacturer's instructions using AccuPrime HiFi Taq (Life Technologies) and 1 µl of cDNA Product. Efficacy of the endpoint PCR was performed via gel electrophoresis of PCR product in a 1% agarose gel.

The qPCR reaction was set up in triplicates for each cell line and sequence (GAPDH, 1441delT, 2237G>A, and GAA), by mixing 10 µl iTaq Universal SYBR Green Supermix (Bio-Rad), 0.5 µl sequence specific forward primer, 0.5 µl sequence specific reverse primer, 1 µl cDNA product, and 8µl water. qPCR analysis was performed in a CFX96 Real Time PCR System under the following thermocycling conditions: 95°C for 30 s followed by 35 cycles of 95°C for 5 s, and 60°C for 30 s.

Next Generation Sequencing Analysis. A custom python script was developed to perform sequence analysis. For each sample, sequences with frequency of less than 1000 were filtered from the data. Sequences in which the reads matched with primer and reverse complement subsequences classified as "target sequences". Target sequences were aligned with corresponding wildtype sequence using global pairwise sequence alignment and analyzed manually for mutants and PAM wobbles.

Western Blotting. Protein expression of GAA and β-Actin was determined in each cell line. Following cell lysis in ice-cold RIPA buffer supplemented with protease and phosphatase inhibitors and EDTA (5mM), protein concentration was determined (DC Protein Assay, BioRad). Forty µg of protein from each cell line was loaded into a 4-12% Bis-Tris precast gel (Criterion XT, BioRad) and gel electrophoresis performed. Proteins were then transferred to a nitrocellulose membrane and blocked in filtered 5% nonfat dry milk in TBS-T (Tris buffered saline, 0.15% Tween20) for 1 hour at room temperature. The membrane was then incubated overnight at 4°C with GAA (Abcam ab137068, 1:1000) and β-Actin (Millipore, MAB1501, 1:40,000) primary antibodies. Following the incubation period, the membrane was washed in TBS-T and incubated with appropriate horseradish peroxidase secondary antibodies (Goat Anti-Rabbit IgG, Abcam ab205718, 1:2000; Anti-Mouse IgG, Cell Signaling Technologies 7076 1:20,000) for 1 hour. The membrane was washed again in TBS-T, and then developed (SuperSignal West Pico Plus Chemiluminescent Substrate, Thermo Scientific) for 5 minutes using a ChemiDoc-It2 Imaging System (UVP) and imaged.

GAA Activity Assay. Acid glucosidase activity was measured by hydrolysis of 4-methylumbelliferyl-D-glucoside (4-MUG, Sigma M-9766) at pH 4 to release the fluorophore 4-methylumbelliferone (4-MU) as previously descirbed[196]. Briefly, 4-MUG was incubated with 10 µL protein lysate in 0.2M sodium acetate for one hour at 37 °C. Fluorescence from 4-MU was then measured using a Glomax plate reader (Progmega) and activity was calculated using a standard curve.

Cardiomyocyte Differentiation. hPS cell lines were differentiated into cardiomyocytes using a small molecule-directed differentiation protocol in a 12-well plate format as previously described[227]

Briefly, all adherent hPSCs and iPSCs were dissociated in TrypLE solution (Life Technologies), counted with a hemocytometer, and centrifuged at 200 x*g* for 5 minutes. Cells were plated at a density between 0.5-1x10^6 cells/well depending on cell line.

Once tissue culture plate wells became 100% confluent (day 0), medium in each confluent well was replaced with a solution containing ml RPMI/B27-Insulin (Life Technologies), 12μM CHIR99021 (BioGems 25917), and 1 μg/ml Insulin solution (Sigma-Aldrich I9278). Exactly 24 hours later (day 1) medium in each well was removed and replaced with RPMI/B27-insulin. Exactly 48 hours later (day 3) half the well volume of spent medium was collected. To this, an equal volume of fresh RPMI/B27-Insulin was mixed. This media was then supplemented with 7.5μM IWP2 (BioGems 75844). Two days later (day 5) medium in each well was replaced with RPMI/B27-Insulin. Two days (day 7) later and every three days following, spent medium was replaced with RPMI/B27. Spontaneous contraction was generally observed between days 8-12 of differentiation.

Immunocytochemistry. Live cell imaging of lysosome intensity was done using LysoSensor Green (Life Technologies L7535). Dye was mixed in culture media at a 1:1000 dilution prior to adding media to wells. Cells were then incubated for 5 minutes in LysoSensor solution. Media was then aspirated and cells were washed 2x with PBS. All imaging was done within one hour of staining.

hPSC cultures were fixed using 4% PFA and incubated at room temperature for 10 minutes. Cells were then permeabilized using 0.05% Triton X-100 and incubated for 10 minutes. Following two washes with 5% goat serum, NANOG antibody (R&D Systems AF1997, 1:200) and TRA-1-60 antibody (Millipore MAB5360, 1:150), was added to cells and incubated overnight at 4 ºC. The next day, cells were rinsed twice with 5% goat serum and then incubated with a

donkey anti-goat secondary antibody (Life Technologies A11055 1:500) for one hour at room temperature. Cells were then washed twice with PBS and mounted for imaging.

Cardiomyocyte cultures were processed in the same manner as above. After permeabilization cells were incubated with anti-sarcomeric alpha-actinin (Abcam ab68167 1:250) overnight at 4 ºC. The next day, cells were rinsed twice with 5% goat serum and then incubated with a goat anti-rabbit secondary antibody (Santa Cruz Biotech sc-362262, 1:500)

Media Exchange. Cardiomyocytes were cultured in RPMI/B27+insulin and media was exchanged every two days. As a normal media exchange, diseased and corrected cells were introduced to RPMI/B27+insulin/-glucose. After staining, media was replaced with media from either corrected or healthy lines and cultured for an additional 24 hours. After incubation, cells were again stained with LysoSensor and imaged using confocal microscopy.

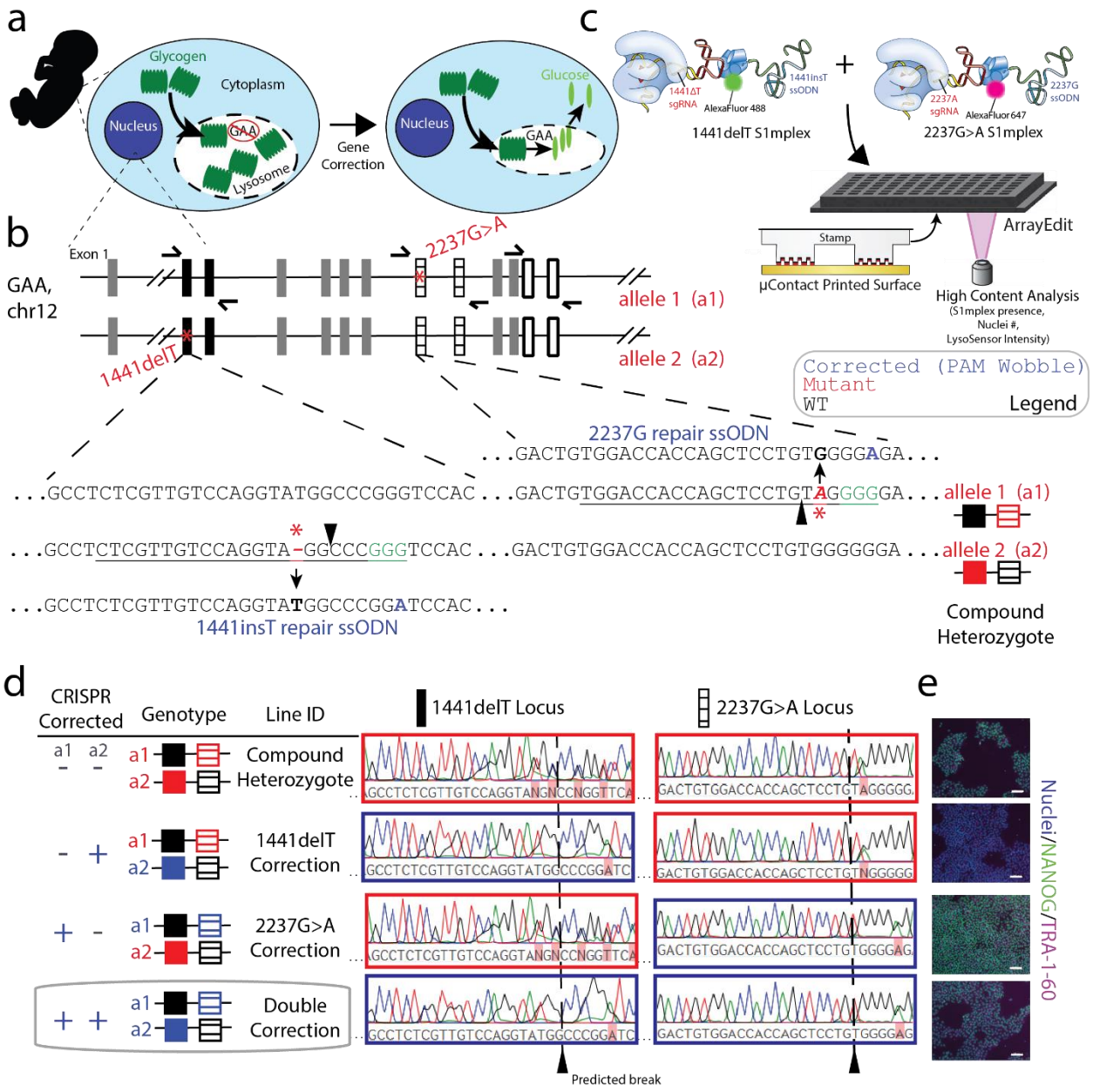

a
Glycogen
Cytoplasm
Nucleus
GAA
Lysosome
Gene Correction
Glucose
b
Exon 1
GAA, chr12
2237G>A
1441delT
allele 1 (a1)
allele 2 (a2)
c
1441ΔT sgRNA
AlexaFluor 488
1441insT ssODN
1441delT S1mplex
2237A sgRNA
AlexaFluor 647
2237G ssODN
2237G>A S1mplex
Stamp
µContact Printed Surface
ArrayEdit
High Content Analysis
(S1mplex presence, Nuclei #, LysoSensor Intensity)
Corrected (PAM Wobble)
Mutant
WT
Legend
2237G repair ssODN
...GACTGTGGACCACCAGCTCCTGTGGGGAGA...
...GCCTCTCGTTGTCCAGGTATGGCCCGGGTCCAC...GACTGTGGACCACCAGCTCCTGTAGGGGGA...
allele 1 (a1)
...GCCTCTCGTTGTCCAGGTA-GGCCCGGGTCCAC...GACTGTGGACCACCAGCTCCTGTGGGGGGA...
allele 2 (a2)
...GCCTCTCGTTGTCCAGGTATGGCCCGGATCCAC...
1441insT repair ssODN
Compound Heterozygote
d
CRISPR Corrected
Genotype
Line ID
1441delT Locus
2237G>A Locus
a1 a2
Compound Heterozygote
1441delT Correction
2237G>A Correction
Double Correction
Predicted break
e
Nuclei/NANOG/TRA-1-60

Figure 4.1 | Isolation of biallelic corrected Pompe disease iPSCs.

(**a**) Pompe disease is caused by two defective copies of acid-α-glucisdase (GAA). This enzyme is responsible for breakdown of glycogen within the lysosomes. Without GAA, glycogen build up can cause downstream health issues. After correction, *GAA* expresses a functional protein leading to a reduction in glycogen. (**b**) Schematic of editing locations within *GAA* locus. Pompe mutants are a compound heterozygous at *GAA*. Allele one contains a point mutation that causes a premature stop codon (*GAA*:c.[1441=;2237G>A]) while allele two carries a one basepair deletion (*GAA*:c.[1441delT;2237=]). sgRNAs (underline) were designed to be specific to only the diseased allele by containing mutants (red) within the seed region. ssODNs used for genomic repair contained the wildtype sequence at the mutation site as well as a silent mutation (blue) to remove the PAM site (green) to prevent re-cutting of the allele. (**c**) S1mplex design for repair of compound heterozygous mutations. S1mplexes targeting 1441delT mutant were labelled with an AlexaFluor488 compound while S1mplexes targeting the 2237G>A mutation were labelled with an AlexaFluor647. These RNP species were mixed prior to transfecting into cells and subsequently plated on ArrayEdit platform to conduct high-content analysis. (**d**) Sanger sequencing traces of all corrected lines. Parental compound heterozygous line contains mutations at both alleles. Heterozygous clones remain identical to parental line at unedited locus and contained PAM wobble on mutant allele. Homozygous line contained PAM wobble at both loci and were unedited on either healthy allele. Cas9 cut site denoted by dotted line to show that no undesired NHEJ products were observed. (**e**) Immunofluorescent imaging for pluripotency markers in isolated cell lines. All lines were over 95% positive for both NANOG and TRA-1-60. Individual images can be seen in **Figure C-2a** (*scale bars: 100μm*).

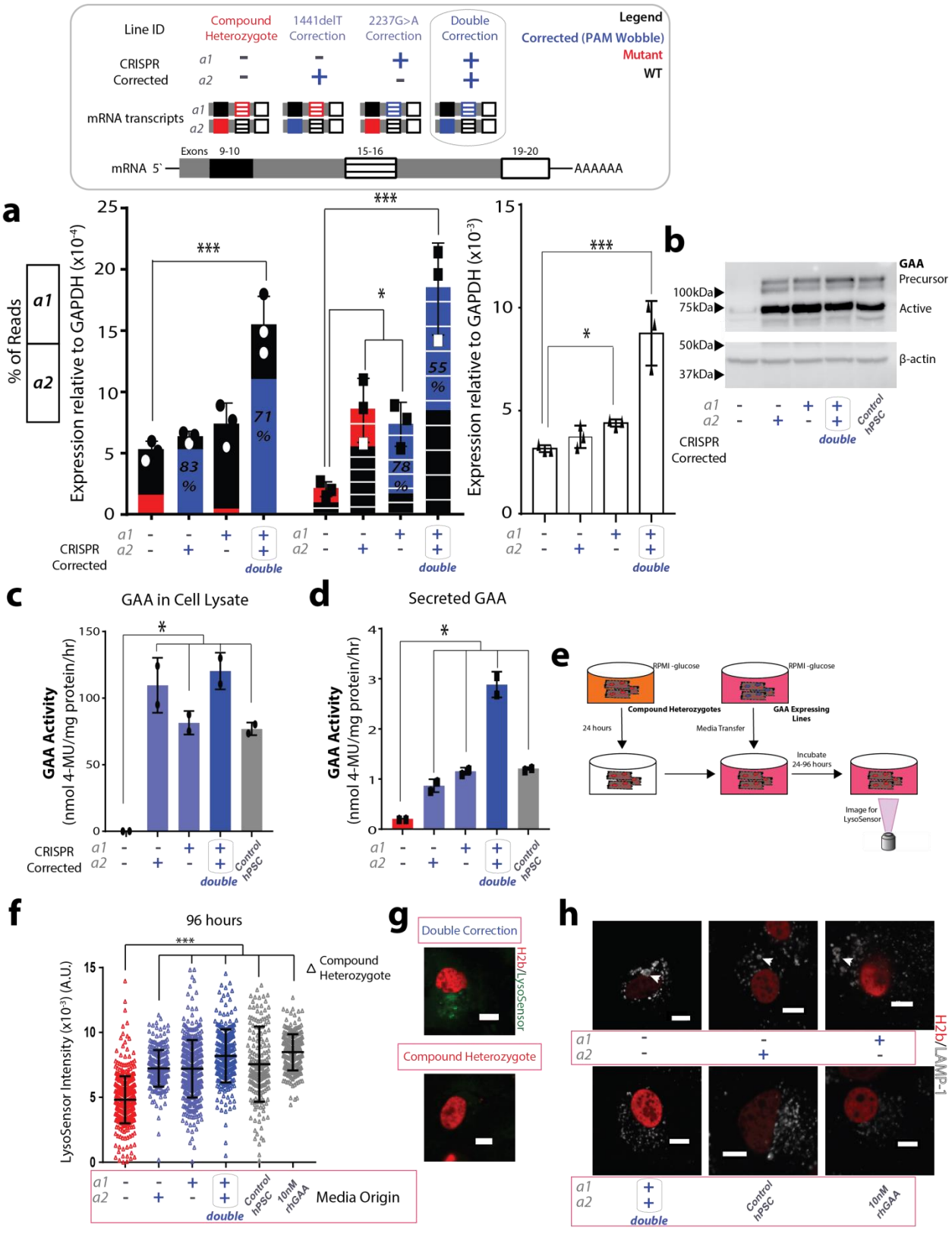

Line ID
Compound Heterozygote
1441delT Correction
2237G>A Correction
Double Correction
Legend
Corrected (PAM Wobble)
Mutant
WT
CRISPR Corrected
a1
a2
mRNA transcripts
Exons 9-10
15-16
19-20
mRNA 5`
AAAAAA
a
% of Reads
Expression relative to GAPDH (x10⁻⁴)
Expression relative to GAPDH (x10⁻³)
83 %
71 %
78 %
55 %
double
b
GAA
Precursor
Active
β-actin
100kDa
75kDa
50kDa
37kDa
Control hPSC
c
GAA in Cell Lysate
GAA Activity
(nmol 4-MU/mg protein/hr)
d
Secreted GAA
e
RPMI -glucose
Compound Heterozygotes
GAA Expressing Lines
24 hours
Media Transfer
Incubate 24-96 hours
Image for LysoSensor
f
96 hours
LysoSensor Intensity (x10⁻³) (A.U.)
Δ Compound Heterozygote
10nM rhGAA
Media Origin
g
Double Correction
H2b/LysoSensor
Compound Heterozygote
h
H2b/LAMP-1

Figure 4.2 | Phenotypic recovery of biallelic corrected iPSCs and differentiated cell types. **(a)** *Top:* Schematic of *GAA* mRNA used for RT-PCR. mRNA was analyzed at 3 locations, around the 1441delT locus (solid), the 2237G>A locus (hashed), and at the final 3` junction (outline). *Bottom:* Overlay of qPCR and NGS data around each edited loci. *Left:* Analysis around 1441delT loci (solid bars). *Middle:* Analysis around 2237G>A loci (hashed bars). Bars are color coded by sequence identity, either wildtype, mutant, or corrected, from deep sequencing analysis and ordered allele 1, allele 2, from bottom to top. Bar heights are equivalent to qPCR quantification relative to *GAPDH.* In all corrected lines the corrected allele was expressed at a higher frequency than the unedited allele. Double corrected line expressed the highest level of overall mRNA and expressed each allele at approximately equal amounts. See also Supplementary Figure 4a, b. *Right:* Quantification of total *GAA* mRNA in compound heterozygote, heterozygote, and homozygous lines via qPCR. The double correction line had significantly higher amount of mRNA than any of the other isolated line (n=3 technical replicates). This is consistent with expression from two active alleles (*p<0.05 ***p<0.005). **(b)** Western blot for GAA protein. Each of the corrected lines expressed high levels of active protein as well as detectable levels of precursor protein suggesting corrected lines were correctly regulated. Compound heterozygous cells expressed significantly lower levels of GAA protein but was still above the limit of detection. **(c)** GAA activity in cell lysate as measured by 4-MUG cleavage in acidic conditions. Compound heterozygotes was unable to cleave this substrate, showing there was little to no active protein. All corrected lines had significantly higher activity than compound heterozygotes but were indistinguishable from each other. Recovery of active protein shows that correction of mutations can recover a healthy phenotype (n=2 technical replicates). **(d)** GAA activity in spent media as measured by 4-MUG cleavage in acidic conditions. Compound heterozygotes were unable to cleave this substrate, showing there was little to no active secreted protein. All corrected lines had significantly higher activity than compound heterozygotes. Double correction cells secreted significantly more GAA than any

other line, including control hPSCs. Increased secreted GAA may lead to cross-correction in unedited cells (n=2 technical replicates). **(e)** Schematic of cross-correction experiments using gene-corrected cardiomyocytes. Compound heterozygote cardiomyocytes (red) were supplied media without glucose for 24 hours (orange). After 24 hours media was replaced with media (pink) that had previously been exposed to corrected cell lines (blue) or supplemented with rhGAA. 24-96 hours after replacement, compound heterozygotes were stained with LysoSensor and imaged using confocal microscopy for dye intensity. **(f)** Quantification of LysoSensor intensity in cross-corrected lines 96 hours post media exchange. Each triangle represents a single compound heterozygote. After 96 hours of daily media changes or supplementing with rhGAA all conditions had significant increase in dye intensity over control conditions. Importantly, double correction lines and ERT had a similar effect and demonstrated the highest level or recovery (***p<0.005). (**g**) Representative images of compound heterozygotes stained with LysoSensor in media from double corrections (top) or compound heterozygotes (bottom) (*scale bars: 10μm*). **(h)** LAMP-1 visualization of compound heterozygotes in different media conditions. Enlarged lysosomes were visible in media from compound heterozygotes and heterozygote corrections (arrowheads). Lysosomes were reduced in size after treatment with media from double corrections and control cardiomyocytes or rhGAA (*scale bars: 10μm*).

CHAPTER 5: Conclusions and Future Directions

This thesis aimed to develop new technologies to address challenges in precision genome editing and subsequently demonstrate their utility in a human pluripotent stem cell disease model. First, I presented ArrayEdit as a method to track cell populations in real time using high-content analysis in a non-destructive fashion. Use of this platform enhanced the frequency of obtaining edited clones. Second, I demonstrated a new technology, the S1mplex, to tether CRISPR ribonucleoproteins and donor DNA together. This all-in-one particle increased precision of genome editing in multiple cell types. Finally, we combined these technologies to correct mutations in an induced pluripotent stem cell model of Pompe disease, an autosomal recessive disorder. These methods and their applications demonstrated advances in the field of precision genome surgery and represent an important step toward the clinical development of genetic therapies.

ArrayEdit as a platform to identify CRISPR/Cas9 edited cells using high-content analysis.

In chapter two, I described a new platform to assay hundreds of human stem cell colonies in parallel. This platform, ArrayEdit, made use of microcontact printing techniques in combination with high-content image analysis to preferentially identify edited cell populations. In this study, a proof of concept gene, *LAMA5*, was targeted using CRISPR/Cas9 gene-editing techniques. Knockout of this extracellular matrix protein is known to cause a decrease in proliferation in human pluripotent stem cells[153]. However, due to this decrease in fitness, it is difficult to obtain edited clones as they are outcompeted by unedited cells. We found that by utilizing ArrayEdit we were able to track the phenotypic change predicted by previous knockout studies at a clonal level in hundreds of populations over an extended time course. Further, we found that when we isolated cells that demonstrated this decrease in proliferation, 82% of clones had a mutation causing a gene knockout in at least one allele and 36% carried a biallelic

mutation. This efficiency was a significant increase compared to the 2% of screened clones containing edits reported previously at this locus.

This study was among the first in the CRISPR/Cas9 literature to use a form of a forward-genetic screen, utilizing phenotypic data to predict genome editing. The use of ArrayEdit also represented a step forward from the laborious process of identifying edited clones through subculture in multi-well plates. Further, this platform could be used without fluorescence-activated cell sorting, a technology that is not immediately available to all labs, and to which not all cell types are amenable. We hypothesized that this platform could be used in combination with any gene-editing target that had a measurable phenotypic effect (i.e. growth, trafficking, morphology, delivery)[148,231,232] and could be expanded even further if Cas9 editing induced a distinct molecular signature. Future work could focus on expanding the library of phenotypic characteristics measured. Another interesting application of the ArrayEdit platform may be simultaneous editing and differentiation of hPSCs to study the effects of editing in a biologically or clinically relevant cell type. Previous work has shown that reprogramming efficiency of primary fibroblasts to a pluripotent state is increased and that neuronal differentiation can be controlled on the ArrayEdit platform[160,233]. The combination of these strategies may reduce the time from reprogramming of patient derived iPSCs to a gene-edited, autologous cell product that for *ex vivo* cell therapies.

S1mplex as an all-in-one ribonucleoprotein to promote precise genome editing.

The ArrayEdit platform described in chapter two focused on the identification of clones that had displayed a knockout phenotype at the targeted locus, an approach that is useful for identifying protein interactions and downstream effectors. However, this left us with a desire to introduce specific point mutations to correct disease-causing variants in isogenic cell lines. This is a challenge in the gene-editing field as imprecise NHEJ mutants typically outnumber precise editing outcomes. To this end, in chapter three we introduced the S1mplex, an all-in-one

ribonucleoprotein complex to promote precise editing. The S1mplex makes use of an aptamer within the sgRNA to bind streptavidin. Biotinylated species such as ssODN donor templates can then be added to the S1mplex to promote precise gene editing.

We utilized ssODN-S1mplexes in two cell types, HEKs and hPSCs to introduce a range of mutations, from a single nucleotide variation in an inserted transgene, up to an 18nt insertion in an endogenous locus. Using S1mplexes we increased the ratio of precise repair to NHEJ mutants up to 18-fold over traditional methods. Contrary to other reports, we found that the composition of the donor DNA (asymmetry, polarity, etc.) did not significantly affect the frequency at which the new DNA sequence was incorporated into the genome[183,193]. To push the capabilities of this new technology we also demonstrated the potential for multiplexed editing by tagging specific sgRNA, ssODN combinations with a distinct fluorophore attached to streptavidin. Using FACS to identify cells that had received these complexes, we were able to isolate populations that were preferentially edited at both loci simultaneously, paving the way for multiplexed editing within a single cell.

This study was supported by other technologies that came to the same conclusion that by supplying donor DNA near the cut site, frequency of precise editing was increased[214,234]. However, the S1mplex system was the first that modified only the sgRNA rather than formation of a Cas9 fusion protein or complicated chemical synthesis. In this manner, the S1mplex is a technology that can be implemented without specialized skills or equipment. To drive the adoption of this technology, there are still several open questions that could be answered regarding the boundaries of the S1mplex. For example, the tagging of endogenous loci with fluorescent proteins is often an inefficient practice and off-target integration can occur due to homologous recombination with dsDNA[235]. By creating long ssDNA that can be tethered to the S1mplex, the rate of on-target integration may increase. Also, as we have previously seen multiplexed editing at two loci, there remains a question of how many loci can be edited simultaneously. The addition of further fluorescent streptavidin molecules may allow for

combinatorial editing within a population of cells. Finally, because the S1mplex relies only on biotinylated moieties an interesting result may be seen by the addition of biotinylated small molecule inhibitors such as an NHEJ inhibitor.

Biallelic editing of a patient-derived stem cell model using S1mplexes on ArrayEdit.

After demonstrating the utility of the ArrayEdit and S1mplexes independently we combined these two technologies in chapter four to correct mutations associated with Pompe disease, an autosomal recessive glycogen storage disorder. Through high-content analysis, we were able to identify clonal populations that were corrected at either one of two distinct mutations in *GAA* within patient-derived iPSCs as well as an additional population corrected at both loci simultaneously. Further interrogation of these clonal populations also showed that multi-target S1mplex editing did not compromise the genomic integrity of cell lines and that no off-target editing at predicted sites occurred. Following editing, NGS revealed that mRNA is expressed preferentially from the corrected allele. We also observed a phenotypic rescue in all corrected lines but the effect was larger in lines that were corrected at both loci. This rescue was sustained in differentiated cardiomyocytes suggesting that this approach could be effective in *ex vivo* cell therapies. In both hPSCs and cardiomyocytes we also measured phenotypic recovery through secreted GAA in cell culture media and found correction of both loci again had a significant positive effect. Accordingly, we wondered if this secretion may have a distal effect on diseased tissues. We found that correction of both alleles had a significantly higher recovery on disease phenotype than correction of a single allele and was similar to addition of rhGAA in cell media. Future experiments may suggest gene correction as an alternative approach to enzyme replacement therapy *in vivo*.

This study was among the first to investigate the potential of simultaneous correction of multiple targets in human cells using ssODNs. This capability may play a large role in future evaluations of clinical gene editing for polygenic disorders such as hypertension and

schizophrenia[236,237]. The observation that correction of both alleles in an autosomal recessive disorder may be beneficial is also surprising and may suggest caution for clinical strategies that rely on gene knockout of mutated alleles in autosomal dominant disorders. To explore this potential, future experiments that utilize terminally differentiated mono and biallelic corrected tissues implanted into a Pompe disease mouse model could inform whether this level of correction is necessary to achieve a distal effect within the body. Alternatively, these experiments could also explore if mono-allelic correction is sufficient to recover a healthy phenotype in combination with advanced tissue engineering techniques. Use of *ex vivo* cell therapy products for Pompe disease is an attractive solution as edited cells could be rigorously screened prior to implantation for off-targeted editing and proper regulatory control. Another approach may use the S1mplex technology for *in vivo* editing in combination with novel nanoparticles that are currently being developed[238,239]. These nanoparticles can be targeted to specific tissues that promote secretion of GAA such as the liver[240]. With this advancement, gene correction of disease causing mutants could occur early in the life of patients, decreasing the need for repeated external intervention.

Towards a clinical use of genome editors to correct genetic diseases

Already, advances in the field of precision editing, controlled regulation, and targeted delivery[86,217,241] have fast tracked Cas9 to clinical trials[242,243]. However, there are still many questions that remain such as the immunogenicity of Cas9[244] as well as the potential to cause unintended consequences in targeted cells[69,70,224]. Although many challenges remain before its widespread adoption, we believe that the development of rapid, precise genome editing technologies will enhance the clinical viability of treating genetic diseases. It is our hope that this body of work represents a step in the advancement of this revolutionary technology to promote human health for all.

REFERENCES

1. Gupta, R. M. & Musunuru, K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J. Clin. Invest.* **124,** 4154–4161 (2014).
2. Ghasemi, M. & Brown, R. H. Genetics of Amyotrophic Lateral Sclerosis. *Cold Spring Harb. Perspect. Med.* a024125 (2017). doi:10.1101/cshperspect.a024125
3. Bertram, L. & Tanzi, R. E. Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. *Nat. Rev. Neurosci.* **9,** 768–778 (2008).
4. Moustafa, J. S. E.-S. & Froguel, P. From obesity genetics to the future of personalized obesity therapy. *Nat. Rev. Endocrinol.* **9,** 402–413 (2013).
5. Maeder, M. L. & Gersbach, C. A. Genome-editing Technologies for Gene and Cell Therapy. *Mol. Ther.* **24,** 430–446 (2016).
6. Hefferin, M. L. & Tomkinson, A. E. Mechanism of DNA double-strand break repair by non-homologous end joining. *DNA Repair* **4,** 639–648 (2005).
7. Liang, F., Han, M., Romanienko, P. J. & Jasin, M. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc. Natl. Acad. Sci.* **95,** 5172–5177 (1998).
8. Maddalo, D. *et al.* In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* **516,** 423–427 (2014).
9. Christie, K. A. *et al.* Towards personalised allele-specific CRISPR gene editing to treat autosomal dominant disorders. *Sci. Rep.* **7,** 16174 (2017).
10. Wu, J., Hunt, S. D., Xue, H., Liu, Y. & Darabi, R. Generation and validation of PAX7 reporter lines from human iPS cells using CRISPR/Cas9 technology. *Stem Cell Res.* **16,** 220–228 (2016).
11. Tebas, P. *et al.* Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N. Engl. J. Med.* **370,** 901–910 (2014).

12. PD-1 Knockout Engineered T Cells for Metastatic Non-small Cell Lung Cancer - Full Text View - ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/NCT02793856. (Accessed: 29th July 2018)

13. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337,** 816–821 (2012).

14. 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).

15. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339,** 819–823 (2013).

16. Gaj, T., Gersbach, C. A. & Barbas III, C. F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31,** 397–405 (2013).

17. Porteus, M. H. & Carroll, D. Gene targeting using zinc finger nucleases. *Nat. Biotechnol.* **23,** 967–973 (2005).

18. Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159,** 647–661 (2014).

19. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* **32,** 670–676 (2014).

20. Shechner, D. M., Hacisuleyman, E., Younger, S. T. & Rinn, J. L. Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat. Methods* **12,** 664–670 (2015).

21. Ma, H. *et al.* Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nat. Biotechnol.* **34,** 528–530 (2016).

22. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523,** 481–485 (2015).

23. Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* **10,** 1116–1121 (2013).

24. Hou, Z. *et al.* Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. *Proc. Natl. Acad. Sci.* 201313587 (2013). doi:10.1073/pnas.1313587110

25. Knight, S. C. *et al.* Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science* **350,** 823–826 (2015).

26. Jinek, M. *et al.* A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive bacterial Immunity. *Science* **337,** 816–822 (2012).

27. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. **823,** 823–827 (2013).

28. Cradick, T. J., Fine, E. J., Antico, C. J. & Bao, G. CRISPR/Cas9 systems targeting b-globin and CCR5 genes have substantial off-target activity. doi:10.1093/nar/gkt714

29. Duan, J. *et al.* Genome-wide identification of CRISPR/Cas9 off-targets in human genome. *Nat. Publ. Group* **24,** (2014).

30. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31,** (2013).

31. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* **32,** (2014).

32. Veres, A. *et al.* Cell Stem Cell Low Incidence of Off-Target Mutations in Individual CRISPR-Cas9 and TALEN Targeted Human Stem Cell Clones Detected by Whole-Genome Sequencing. (2014). doi:10.1016/j.stem.2014.04.020

33. Kim, D. *et al.* Digenome-seq: Genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* (2015). doi:10.1038/nmeth.3284

34. Tsai, S. Q. *et al.* CIRCLE-seq: A highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat. Methods* (2017). doi:10.1038/nmeth.4278

35. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33,** 187–198 (2015).

36. Yan, W. X. *et al.* ARTICLE BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat. Commun.* **8,** (2017).

37. Tsai, T.-L., Wang, B., Squire, M. W., Guo, L.-W. & Li, W.-J. Endothelial cells direct human mesenchymal stem cells for osteo- and chondro-lineage differentiation through endothelin-1 and AKT signaling. *Stem Cell Res. Ther.* **6,** 88 (2015).

38. Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* **33,** 73–80 (2015).

39. Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A. & Liu, D. R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol.* (2015). doi:10.1038/nchembio.1793

40. Chen, Yanhao, Liu, Xiaojian, Zhang, Yongxian, Wang, H. A Self-restricted CRISPR System to Reduce Off-target Effects. **24,** (2016).

41. Hemphill, J., Borchardt, E. K., Brown, K., Asokan, A. & Deiters, A. Optical Control of CRISPR/Cas9 Gene Editing. doi:10.1021/ja512664v

42. Richter, F. *et al.* Engineering of temperature- and light-switchable Cas9 variants. *Nucleic Acids Res.* **44,** 10003–10014 (2016).

43. Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. doi:10.1016/j.cell.2013.08.021

44. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. doi:10.1038/nature14592

45. Hou, Z. *et al.* Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. doi:10.1073/pnas.1313587110

46. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. doi:10.1038/nature14299

47. Nishimasu, H. *et al.* Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **156,** 935–949 (2014).

48. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513,** (2014).

49. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351,** 84–8 (2016).

50. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects CRISPR–Cas9 nucleases enable highly efficient genome editing in a wide variety of organisms Alteration of SpCas9 DNA contacts. *Nature* **529,** (2016).

51. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nat. Publ. Group* **556,** (2018).

52. Singh, D., Sternberg, S. H., Fei, J., Ha, T. & Doudna, J. A. Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *bioRxiv* **7,** 048371 (2016).

53. Kim, S., Bae, T., Hwang, J. & Kim, J.-S. Rescue of high-specificity Cas9 variants using sgRNAs with matched 5’ nucleotides. doi:10.1186/s13059-017-1355-3

54. Cox, D. B. T. *et al.* RNA editing with CRISPR-Cas13. *Science* eaaq0180 (2017). doi:10.1126/science.aaq0180

55. Gootenberg, J. S. *et al.* Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science* eaaq0179 (2018). doi:10.1126/science.aaq0179

56. Young, C. S. *et al.* A Single CRISPR-Cas9 Deletion Strategy that Targets the Majority of DMD Patients Restores Dystrophin Function in hiPSC-Derived Muscle Cells. *Cell Stem Cell* **0,** (2016).

57. 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,** (2015).

58. Yu, C. *et al.* Small molecules enhance crispr genome editing in pluripotent stem cells. *Cell Stem Cell* **16,** 142–147 (2015).

59. Canny, M. D. *et al.* Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR-Cas9 genome-editing efficiency. *Nat. Biotechnol.* **36,** 95–102 (2018).

60. Yang, D. *et al.* Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with customizable endonucleases. *Nat. Publ. Group* (2016). doi:10.1038/srep21264

61. 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,** (2014).

62. Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G. F. & Chin, L. Post-translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair. *CellReports* **14,** 1555–1566 (2016).

63. 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,** (2016).

64. Liang, X., Potter, J., Kumar, S., Ravinder, N. & Chesnut, J. D. Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *J. Biotechnol.* **241,** 136–146 (2017).

65. Renaud, J.-B. *et al.* Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases. *Cell Rep.* **14,** 2263–2272 (2016).

66. Lee, K. *et al.* Synthetically modified guide RNA and donor DNA are a versatile platform for CRISPR-Cas9 engineering. *eLife* **6,** e25312 (2017).

67. Ma, M. *et al.* LETTER TO THE EDITOR Efficient generation of mice carrying homozygous double-floxp alleles using the Cas9-Avidin/Biotin-donor DNA system. *Nat. Publ. Group* **27,** (2017).

68. Carlson-Stevermer, J. *et al.* Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing. doi:10.1038/s41467-017-01875-9

69. Janic, A. *et al.* DNA repair processes are critical mediators of p53-dependent tumor suppression. *Nat. Med.* 1 (2018). doi:10.1038/s41591-018-0043-5

70. Ihry, R. J. *et al.* p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nat. Med.* 1 (2018). doi:10.1038/s41591-018-0050-6

71. Richardson, C. D. *et al.* CRISPR–Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. *Nat. Genet.* 1 (2018). doi:10.1038/s41588-018-0174-0

72. Truong, L. N. *et al.* Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proc. Natl. Acad. Sci.* **110,** 7720–7725 (2013).

73. Bae, S., Kweon, J., Kim, H. S. & Kim, J. S. Microhomology-based choice of Cas9 nuclease target sites. *Nature Methods* (2014). doi:10.1038/nmeth.3015

74. Lemos, B. R. *et al.* CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. doi:10.1073/pnas.1716855115

75. Yao, X. *et al.* Homology-mediated end joining-based targeted integration using CRISPR/Cas9. *Nat. Publ. Group* **27,** (2017).

76. Suzuki, K. *et al.* In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540,** 144–149 (2016).

77. Susani, L. *et al.* Correction of a Recessive Genetic Defect by CRISPR-Cas9-Mediated Endogenous Repair. *CRISPR J.* **1,** 230–238 (2018).

78. Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System Cpf1 is a RNA-guided DNA nuclease that provides immunity in bacteria and can be adapted for genome editing in mammalian cells. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163,** 759–771 (2015).

79. Kim, D. *et al.* Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* **34,** (2016).

80. Moreno-Mateos, M. A. *et al.* CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing. *Antonio J Giraldez* **679,**

81. Zhou, M. *et al.* Seamless Genetic Conversion of *SMN2* to *SMN1* via CRISPR/Cpf1 and Single-Stranded Oligodeoxynucleotides in Spinal Muscular Atrophy Patient-Specific Induced Pluripotent Stem Cells. *Hum. Gene Ther.* hum.2017.255 (2018). doi:10.1089/hum.2017.255

82. Li, S. Y. *et al.* CRISPR-Cas12a has both cis- and trans-cleavage activities on single-stranded DNA. *Cell Res.* 1–3 (2018). doi:10.1038/s41422-018-0022-x

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

84. Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551,** 464–471 (2017).

85. Bill Kim, Y. *et al.* Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Publ. Group* **35,** (2017).

86. Justin Eyquem *et al.* Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection. *Nature* **543,** 113–117 (2017).

87. Liu, N. *et al.* Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch. *Cell* **0,** (2018).

88. Gori, J. L. *et al.* Delivery and Specificity of CRISPR/Cas9 Genome Editing Technologies for Human Gene Therapy. *Hum. Gene Ther.* **26,** 443–451 (2015).

89. Nishiyama, J., Mikuni, T. & Yasuda, R. Virus-Mediated Genome Editing via Homology-Directed Repair in Mitotic and Postmitotic Cells in Mammalian Brain. *Neuron* **96,** 755–768.e5 (2017).

90. Gaj, T. *et al.* In vivo genome editing improves motor function and extends survival in a mouse model of ALS. *Sci. Adv.* **3,** eaar3952 (2017).

91. Ginn, S. L., Amaya, A. K., Alexander, I. E., Edelstein, M. & Abedi, M. R. Gene therapy clinical trials worldwide to 2017: An update. *J. Gene Med.* **20,** e3015

92. Swiech, L. *et al.* In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. (2014). doi:10.1038/nbt.3055

93. Ertl, H. C. J. Preclinical models to assess the immunogenicity of AAV vectors. *Cell. Immunol.* (2017). doi:10.1016/j.cellimm.2017.11.006

94. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31,** 822–826 (2013).

95. Yin, H., Kauffman, K. J. & Anderson, D. G. Delivery technologies for genome editing. *Nat. Rev. Drug Discov.* **16,** 387–399 (2017).

96. Glass, Z., Lee, M., Li, Y. & Xu, Q. Engineering the Delivery System for CRISPR-Based Genome Editing. *Trends Biotechnol.* **36,** 173–185 (2018).

97. Alsaiari, S. K. *et al.* Endosomal Escape and Delivery of CRISPR/Cas9 Genome Editing Machinery Enabled by Nanoscale Zeolitic Imidazolate Framework. doi:10.1021/jacs.7b11754

98. Mout, R. *et al.* Direct Cytosolic Delivery of CRISPR/Cas9- Ribonucleoprotein for Efficient Gene Editing. doi:10.1021/acsnano.6b07600

99. Sun, W. *et al.* Drug Delivery Hot Paper Self-Assembled DNANanoclews for the Efficient Delivery of CRISPR–Cas9 for Genome Editing.

100. Yu Xiquan Liang Huimin Xie Shantanu Kumar Namritha Ravinder Jason Potter Xavier de Mollerat du Jeu Jonathan Chesnut, X. D. Improved delivery of Cas9 protein/gRNA complexes using lipofectamine CRISPRMAX. *Biotechnol. Lett.* **38,** (2064).

101. Rouet, R. *et al.* Receptor-Mediated Delivery of CRISPR-Cas9 Endonuclease for Cell Type Specific Gene Editing. *J. Am. Chem. Soc.* jacs.8b01551 (2018). doi:10.1021/jacs.8b01551

102. Toscano, A. & Schoser, B. Enzyme replacement therapy in late-onset Pompe disease: a systematic literature review. *J. Neurol.* **260,** 951–959 (2013).

103. Slonim, A. E. *et al.* Modification of the natural history of adult-onset acid maltase deficiency by nutrition and exercise therapy. *Muscle Nerve* **35,** 70–77 (2007).

104. Case, L. E. & Kishnani, P. S. Physical therapy management of Pompe disease. *Genet. Med.* **8,** 318–327 (2006).

105. Nicolino, M. *et al.* Clinical outcomes after long-term treatment with alglucosidase alfa in infants and children with advanced Pompe disease. *Genet. Med.* **11,** 210–219 (2009).

106. Kishnani, P. S. *et al.* Recombinant human acid α-glucosidase Major clinical benefits in infantile-onset Pompe disease. *Neurology* **68,** 99–109 (2007).

107. Lim, J.-A., Li, L. & Raben, N. Pompe disease: from pathophysiology to therapy and back again. *Front. Aging Neurosci.* **6,** 177 (2014).

108. Richard, E., Douillard-Guilloux, G. & Caillaud, C. New insights into therapeutic options for Pompe disease. *IUBMB Life* **63,** 979–986 (2011).

109. Shea, L. & Raben, N. Autophagy in skeletal muscle: implications for Pompe disease. *Int. J. Clin. Pharmacol. Ther.* **47,** S42–S47 (2009).

110. Raben, N. *et al.* Replacing acid α-glucosidase in Pompe disease: recombinant and transgenic enzymes are equipotent, but neither completely clears glycogen from type II muscle fibers. *Mol. Ther.* **11,** 48–56 (2005).

111. Raben, N. *et al.* Enzyme replacement therapy in the mouse model of Pompe disease. *Mol. Genet. Metab.* **80,** 159–169 (2003).

112. Ziegler, R. J. *et al.* Ability of Adeno-Associated Virus Serotype 8-Mediated Hepatic Expression of Acid α-Glucosidase to Correct the Biochemical and Motor Function

Deficits of Presymptomatic and Symptomatic Pompe Mice. *Hum. Gene Ther.* **19,** 609–621 (2008).

113. Franco, L. M. *et al.* Evasion of Immune Responses to Introduced Human Acid α-Glucosidase by Liver-Restricted Expression in Glycogen Storage Disease Type II. *Mol. Ther.* **12,** 876–884 (2005).

114. Sun, B. *et al.* Enhanced Response to Enzyme Replacement Therapy in Pompe Disease after the Induction of Immune Tolerance. *Am. J. Hum. Genet.* **81,** 1042–1049 (2007).

115. Sun, B. *et al.* Immunomodulatory Gene Therapy Prevents Antibody Formation and Lethal Hypersensitivity Reactions in Murine Pompe Disease. *Mol. Ther.* **18,** 353–360 (2010).

116. Han, S. *et al.* Low-Dose Liver-Targeted Gene Therapy for Pompe Disease Enhances Therapeutic Efficacy of ERT via Immune Tolerance Induction. *Mol. Ther. - Methods Clin. Dev.* **4,** 126–136 (2017).

117. Raben, N. *et al.* Conditional tissue-specific expression of the acid α-glucosidase (GAA) gene in the GAA knockout mice: implications for therapy. *Hum. Mol. Genet.* **10,** 2039–2047 (2001).

118. Amalfitano, A. *et al.* Systemic correction of the muscle disorder glycogen storage disease type II after hepatic targeting of a modified adenovirus vector encoding human acid-α-glucosidase. *Proc. Natl. Acad. Sci.* **96,** 8861–8866 (1999).

119. Martin-Touaux, E. *et al.* Muscle as a putative producer of acid α-glucosidase for glycogenosis type II gene therapy. *Hum. Mol. Genet.* **11,** 1637–1645 (2002).

120. Ding, E. y. *et al.* Long-Term Efficacy after [E1-, polymerase-] Adenovirus-Mediated Transfer of Human Acid-α-Glucosidase Gene into Glycogen Storage Disease Type II Knockout Mice. *Hum. Gene Ther.* **12,** 955–965 (2001).

121. Xu, F. *et al.* Glycogen storage in multiple muscles of old GSD-II mice can be rapidly cleared after a single intravenous injection with a modified adenoviral vector expressing hGAA. *J. Gene Med.* **7,** 171–178 (2005).

122. Cresawn, K. O. *et al.* Impact of Humoral Immune Response on Distribution and Efficacy of Recombinant Adeno-Associated Virus-Derived Acid α-Glucosidase in a Model of Glycogen Storage Disease Type II. *Hum. Gene Ther.* **16,** 68–80 (2005).

123. Til, N. P. van *et al.* Lentiviral gene therapy of murine hematopoietic stem cells ameliorates the Pompe disease phenotype. *Blood* **115,** 5329–5337 (2010).

124. Douillard-Guilloux, G., Richard, E., Batista, L. & Caillaud, C. Partial phenotypic correction and immune tolerance induction to enzyme replacement therapy after hematopoietic stem cell gene transfer of α-glucosidase in Pompe disease. *J. Gene Med.* **11,** 279–287 (2009).

125. van der Wal, E. *et al.* GAA Deficiency in Pompe Disease Is Alleviated by Exon Inclusion in iPSC-Derived Skeletal Muscle Cells. *Mol. Ther. - Nucleic Acids* **7,** 101–115 (2017).

126. van der Wal, E., Bergsma, A. J., Pijnenburg, J. M., van der Ploeg, A. T. & Pijnappel, W. W. M. P. Antisense Oligonucleotides Promote Exon Inclusion and Correct the Common c.-32-13T>G GAA Splicing Variant in Pompe Disease. *Mol. Ther. - Nucleic Acids* **7,** 90–100 (2017).

127. Kasap, C., Elemento, O. & Kapoor, T. M. DrugTargetSeqR: a genomics- and CRISPR-Cas9–based method to analyze drug targets. *Nat. Chem. Biol.* **10,** 626–628 (2014).

128. Shi, J. *et al.* Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat. Biotechnol.* **33,** 661–667 (2015).

129. Smurnyy, Y. *et al.* DNA sequencing and CRISPR-Cas9 gene editing for target validation in mammalian cells. *Nat. Chem. Biol.* **10,** 623–625 (2014).

130. Hsu, P. D., Lander, E. S. & Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157,** 1262–1278 (2014).

131. Doudna JA. Genomic engineering and the future of medicine. *JAMA* **313,** 791–792 (2015).

132. Sternberg, S. H. & Doudna, J. A. Expanding the Biologist's Toolkit with CRISPR-Cas9. *Mol. Cell* **58,** 568–574 (2015).

133. Baltimore, D. *et al.* A prudent path forward for genomic engineering and germline gene modification. *Science* **348,** 36–38 (2015).

134. Ding, Q. *et al.* Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell* **12,** 393–394 (2013).

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

136. Miyaoka, Y. *et al.* Isolation of single-base genome-edited human iPS cells without antibiotic selection. *Nat. Methods* **11,** 291–293 (2014).

137. Yang, L. *et al.* Optimization of scarless human stem cell genome editing. *Nucleic Acids Res.* **41,** 9049–9061 (2013).

138. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11,** 783–784 (2014).

139. Shalem, O. *et al.* Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **343,** 84–87 (2014).

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

141. Bosley, K. S. *et al.* CRISPR germline engineering—the community speaks. *Nat. Biotechnol.* **33,** 478–486 (2015).

142. Merkle, F. T. *et al.* Efficient CRISPR-Cas9-Mediated Generation of Knockin Human Pluripotent Stem Cells Lacking Undesired Mutations at the Targeted Locus. *Cell Rep.* **11,** 875–883 (2015).

143. Byrne, S. M., Ortiz, L., Mali, P., Aach, J. & Church, G. M. Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. *Nucleic Acids Res.* **43,** e21 (2015).

144. González, F. *et al.* An iCRISPR Platform for Rapid, Multiplexable, and Inducible Genome Editing in Human Pluripotent Stem Cells. *Cell Stem Cell* **15,** 215–226 (2014).

145. Liang, X. *et al.* Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J. Biotechnol.* **208,** 44–53 (2015).

146. 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 (2014).

147. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155,** 1479–1491 (2013).

148. Harkness, T. *et al.* High-content imaging with micropatterned multiwell plates reveals influence of cell geometry and cytoskeleton on chromatin dynamics. *Biotechnol. J.* **10,** 1555–1567 (2015).

149. Schumann, K. *et al.* Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proc. Natl. Acad. Sci. U. S. A.* (2015). doi:10.1073/pnas.1512503112

150. Sha, J., Lippmann, E. S., McNulty, J., Ma, Y. & Ashton, R. S. Sequential Nucleophilic Substitutions Permit Orthogonal Click Functionalization of Multicomponent PEG Brushes. *Biomacromolecules* **14,** 3294–3303 (2013).

151. Bae, S., Kweon, J., Kim, H. S. & Kim, J.-S. Microhomology-based choice of Cas9 nuclease target sites. *Nat. Methods* **11,** 705–706 (2014).

152. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7,** R100 (2006).

153. Laperle, A. *et al.* α-5 Laminin Synthesized by Human Pluripotent Stem Cells Promotes Self-Renewal. *Stem Cell Rep.* **5,** 195–206 (2015).

154. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31,** 827–832 (2013).

155. Xu, H. *et al.* Sequence determinants of improved CRISPR sgRNA design. *Genome Res.* gr.191452.115 (2015). doi:10.1101/gr.191452.115

156. D'Astolfo, D. S. *et al.* Efficient intracellular delivery of native proteins. *Cell* **161,** 674–690 (2015).

157. Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* **33,** 73–80 (2015).

158. Chiappini, C. *et al.* Biodegradable silicon nanoneedles delivering nucleic acids intracellularly induce localized in vivo neovascularization. *Nat. Mater.* **14,** 532–539 (2015).

159. Taylor, D. L. & Haskins, J. R. *High Content Screening: A Powerful Approach to Systems Cell Biology and Drug Discovery*. (Springer Science & Business Media, 2007).

160. Knight, G. T., Sha, J. & Ashton, R. S. Micropatterned, clickable culture substrates enable in situ spatiotemporal control of human PSC-derived neural tissue morphology. *Chem. Commun.* (2015). doi:10.1039/C4CC08665A

161. Singh, S., Carpenter, A. E. & Genovesio, A. Increasing the Content of High-Content Screening An Overview. *J. Biomol. Screen.* **19,** 640–650 (2014).

162. Nazareth, E. J. P. *et al.* High-throughput fingerprinting of human pluripotent stem cell fate responses and lineage bias. *Nat. Methods* **10,** 1225–1231 (2013).

163. McNulty, J. D. *et al.* High-precision robotic microcontact printing (R-μCP) utilizing a vision guided selectively compliant articulated robotic arm. *Lab. Chip* **14,** 1923–1930 (2014).

164. Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P. & Lu, T. K. Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells. *Mol. Cell* **54,** 698–710 (2014).

165. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* gr.171322.113 (2014). doi:10.1101/gr.171322.113

166. Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* **33,** 139–142 (2015).

167. Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A. & Liu, D. R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol.* **11,** 316–318 (2015).

168. Hemphill, J., Borchardt, E. K., Brown, K., Asokan, A. & Deiters, A. Optical Control of CRISPR/Cas9 Gene Editing. *J. Am. Chem. Soc.* **137,** 5642–5645 (2015).

169. Nihongaki, Y., Kawano, F., Nakajima, T. & Sato, M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat. Biotechnol.* **33,** 755–760 (2015).

170. Ma, H. *et al.* Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 3002–3007 (2015).

171. Chen, G. *et al.* Chemically defined conditions for human iPSC derivation and culture. *Nat. Methods* **8,** 424–429 (2011).

172. Prediger, E. Simplifying CRISPR. *Genet. Eng. Biotechnol. News* **35,** 18–19 (2015).

173. Dever, D. P. *et al.* CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells. *Nature* **539,** 384–389 (2016).

174. DeWitt, M. A. *et al.* Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Sci. Transl. Med.* **8,** 360ra134–360ra134 (2016).

175. Ravin, S. S. D. *et al.* CRISPR-Cas9 gene repair of hematopoietic stem cells from patients with X-linked chronic granulomatous disease. *Sci. Transl. Med.* **9,** eaah3480 (2017).

176. Ma, H. *et al.* Correction of a pathogenic gene mutation in human embryos. *Nature* **548,** 413–419 (2017).

177. Li, K., Wang, G., Andersen, T., Zhou, P. & Pu, W. T. Optimization of Genome Engineering Approaches with the CRISPR/Cas9 System. *PLOS ONE* **9,** e105779 (2014).

178. Bae, S., Kweon, J., Kim, H. S. & Kim, J.-S. Microhomology-based choice of Cas9 nuclease target sites. *Nat. Methods* **11,** 705–706 (2014).

179. Merkle, F. T. *et al.* Efficient CRISPR-Cas9-mediated generation of knockin human pluripotent stem cells lacking undesired mutations at the targeted locus. *Cell Rep.* **11,** 875–883 (2015).

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

181. 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).

182. Song, J. *et al.* RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. *Nat. Commun.* **7,** 10548 (2016).

183. Liang, X., Potter, J., Kumar, S., Ravinder, N. & Chesnut, J. D. Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *J. Biotechnol.* **241,** 136–146 (2017).

184. Gaj, T. *et al.* Targeted gene knock-in by homology-directed genome editing using Cas9 ribonucleoprotein and AAV donor delivery. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx154

185. Duda, K. *et al.* High-efficiency genome editing via 2A-coupled co-expression of fluorescent proteins and zinc finger nucleases or CRISPR/Cas9 nickase pairs. *Nucleic Acids Res.* **42,** e84 (2014).

186. Lonowski, L. A. *et al.* Genome editing using FACS enrichment of nuclease-expressing cells and indel detection by amplicon analysis. *Nat. Protoc.* **12,** 581–603 (2017).

187. Steyer, B. *et al.* Scarless Genome Editing of Human Pluripotent Stem Cells via Transient Puromycin Selection. *Stem Cell Rep.* **10,** 642–654 (2018).

188. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517,** 583–588 (2015).

189. Leppek, K. & Stoecklin, G. An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. *Nucleic Acids Res.* gkt956 (2013). doi:10.1093/nar/gkt956

190. Carlson-Stevermer, J. *et al.* High-Content Analysis of CRISPR-Cas9 Gene-Edited Human Embryonic Stem Cells. *Stem Cell Rep.* **6,** 109–120 (2016).

191. Le Trong, I. *et al.* Streptavidin and its biotin complex at atomic resolution. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 813–821 (2011).

192. Nishimasu, H. *et al.* Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **156,** 935–949 (2014).

193. 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).

194. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31,** 839–843 (2013).

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

196. Raval, K. K. *et al.* Pompe Disease Results in a Golgi-based Glycosylation Deficit in Human Induced Pluripotent Stem Cell-derived Cardiomyocytes. *J. Biol. Chem.* **290,** 3121–3136 (2015).

197. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42,** e168 (2014).

198. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24,** 1012–1019 (2014).

199. Lee, K. *et al.* Synthetically modified guide RNA and donor DNA are a versatile platform for CRISPR-Cas9 engineering. *eLife* **6,** e25312 (2017).

200. Titus, S. *et al.* Impairment of BRCA1-Related DNA Double-Strand Break Repair Leads to Ovarian Aging in Mice and Humans. *Sci. Transl. Med.* **5,** 172ra21–172ra21 (2013).

201. Gundry, M. C. *et al.* Highly efficient genome editing of murine and human hematopoietic progenitor cells by CRISPR/Cas9. *Cell Rep.* **17,** 1453–1461 (2016).

202. Miyaoka, Y. *et al.* Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci. Rep.* **6,** (2016).

203. Soldner, F. & Jaenisch, R. In Vitro Modeling of Complex Neurological Diseases. in *Genome Editing in Neurosciences* 1–19 (Springer, Cham, 2017).

204. Soldner, F. *et al.* Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression. *Nature* **533,** 95–99 (2016).

205. Papapetrou, E. P. Patient-derived induced pluripotent stem cells in cancer research and precision oncology. *Nat. Med.* **22,** 1392–1401 (2016).

206. Tabebordbar, M., Cheng, J. & Wagers, A. J. Therapeutic Gene Editing in Muscles and Muscle Stem Cells. in *Genome Editing in Neurosciences* 103–123 (Springer, Cham, 2017).

207. Li, H. *et al.* In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature* **475,** 217–221 (2011).

208. Paulk, N. K. *et al.* Adeno-associated virus gene repair corrects a mouse model of hereditary tyrosinemia in vivo. *Hepatology* **51,** 1200–1208 (2010).

209. Yin, H. *et al.* Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat. Biotechnol.* **advance online publication,** (2014).

210. Li, H. *et al.* Design and specificity of long ssDNA donors for CRISPR-based knock-in. *bioRxiv* 178905 (2017). doi:10.1101/178905

211. Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44,** D862–D868 (2016).

212. Sevier, C. S. & Kaiser, C. A. Formation and transfer of disulphide bonds in living cells. *Nat. Rev. Mol. Cell Biol.* **3,** 836–847 (2002).

213. Bao, G., Mitragotri, S. & Tong, S. Multifunctional Nanoparticles for Drug Delivery and Molecular Imaging. *Annu. Rev. Biomed. Eng.* **15,** 253–282 (2013).

214. Ma, M. *et al.* Efficient generation of mice carrying homozygous double-floxp alleles using the Cas9-Avidin/Biotin-donor DNA system. *Cell Res.* **27,** 578–581 (2017).

215. Liu, X. *et al.* In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell* **170,** 1028–1043.e19 (2017).

216. Ma, H. *et al.* Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nat. Biotechnol.* **34,** 528–530 (2016).

217. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529,** 490–495 (2016).

218. Petronzelli, F. *et al.* Therapeutic Use of Avidin Is Not Hampered by Antiavidin Antibodies in Humans. *Cancer Biother. Radiopharm.* **25,** 563–570 (2010).

219. Kishnani, P. S. & Howell, R. R. Pompe disease in infants and children. *J. Pediatr.* **144,** S35–S43 (2004).

220. Bodamer, O. A., Scott, C. R., Giugliani, R. & Group, on behalf of the P. D. N. S. W. Newborn Screening for Pompe Disease. *Pediatrics* **140,** S4–S13 (2017).

221. Berrier, K. L. *et al.* CRIM-negative infantile Pompe disease: characterization of immune responses in patients treated with ERT monotherapy. *Genet. Med.* (2015). doi:10.1038/gim.2015.6

222. van der Wal, E. *et al.* Large-Scale Expansion of Human iPSC-Derived Skeletal Muscle Cells for Disease Modeling and Cell-Based Therapeutic Strategies. *Stem Cell Rep.* doi:10.1016/j.stemcr.2018.04.002

223. Carlson-Stevermer, J. *et al.* Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing. *Nat. Commun.* **8,** 1711 (2017).

224. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* (2018). doi:10.1038/nbt.4192

225. Popp, M. W. & Maquat, L. E. Leveraging Rules of Nonsense-Mediated mRNA Decay for Genome Engineering and Personalized Medicine. *Cell* **165,** 1319–1322 (2016).

226. Migani, D., Smales, C. M. & Bracewell, D. G. Effects of lysosomal biotherapeutic recombinant protein expression on cell stress and protease and general host cell protein release in Chinese hamster ovary cells. *Biotechnol. Prog.* **33,** 666–676

227. Lian, X. *et al.* Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proc. Natl. Acad. Sci.* (2012). doi:10.1073/pnas.1200250109

228. Yang, H. W. *et al.* Recombinant Human Acid α-Glucosidase Corrects Acidα-Glucosidase-Deficient Human Fibroblasts, Quail Fibroblasts, and Quail Myoblasts. *Pediatr. Res.* **43,** 374–380 (1998).

229. Herzog, A. *et al.* A cross-sectional single-centre study on the spectrum of Pompe disease, German patients: molecular analysis of the GAA gene, manifestation and genotype-phenotype correlations. *Orphanet J. Rare Dis.* **7,** 35 (2012).

230. Fukuhara, Y. *et al.* A molecular analysis of the GAA gene and clinical spectrum in 38 patients with Pompe disease in Japan. *Mol. Genet. Metab. Rep.* **14,** 3–9 (2018).

231. Kuchenov, D. *et al.* High-Content Imaging Platform for Profiling Intracellular Signaling Network Activity in Living Cells. *Cell Chem. Biol.* **23,** 1550–1559 (2016).

232. Kang, J. *et al.* Improving drug discovery with high-content phenotypic screens by systematic selection of reporter cell lines. *Nat. Biotechnol.* **34,** 70–77 (2016).

233. Harkness, T. *et al.* High-content imaging with micropatterned multiwell plates reveals influence of cell geometry and cytoskeleton on chromatin dynamics. *Biotechnol. J.* **10,** 1555–1567 (2015).

234. Lee, K. *et al.* Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nat. Biomed. Eng.* **1,** 889 (2017).

235. Zelensky, A. N., Schimmel, J., Kool, H., Kanaar, R. & Tijsterman, M. Inactivation of Pol θ and C-NHEJ eliminates off-target integration of exogenous DNA. *Nat. Commun.* **8,** 66 (2017).

236. Lvovs, D., Favorova, O. O. & Favorov, A. V. A Polygenic Approach to the Study of Polygenic Diseases. *Acta Naturae* **4,** 59–71 (2012).

237. McCarroll, S. A. & Hyman, S. E. Progress in the Genetics of Polygenic Brain Disorders: Significant New Challenges for Neurobiology. *Neuron* **80,** 578–587 (2013).

238. Finn, J. D. *et al.* A Single Administration of CRISPR/Cas9 Lipid Nanoparticles Achieves Robust and Persistent In Vivo Genome Editing. *Cell Rep.* **22,** 2227–2235 (2018).

239. Mout, R. *et al.* Direct Cytosolic Delivery of CRISPR/Cas9-Ribonucleoprotein for Efficient Gene Editing. *ACS Nano* **11,** 2452–2458 (2017).

240. Puzzo, F. *et al.* Rescue of Pompe disease in mice by AAV-mediated liver delivery of secretable acid α-glucosidase. *Sci. Transl. Med.* **9,** eaam6375 (2017).

241. Sun, D. *et al.* Targeted Multifunctional Lipid ECO Plasmid DNA Nanoparticles as Efficient Non-viral Gene Therapy for Leber's Congenital Amaurosis. *Mol. Ther. - Nucleic Acids* **7,** 42–52 (2017).

242. Cyranoski, D. CRISPR gene-editing tested in a person for the first time. *Nature* (2016). doi:10.1038/nature.2016.20988

243. Reardon, S. First CRISPR clinical trial gets green light from US panel. *Nat. News* doi:10.1038/nature.2016.20137

244. Charlesworth, C. T. *et al.* Identification of Pre-Existing Adaptive Immunity to Cas9 Proteins in Humans. *bioRxiv* 243345 (2018). doi:10.1101/243345

245. 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).

246. Sonoda, E., Hochegger, H., Saberi, A., Taniguchi, Y. & Takeda, S. Differential usage of non-homologous end-joining and homologous recombination in double strand break repair. *DNA Repair* **5,** 1021–1029 (2006).

247. Chen, F. *et al.* High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat. Methods* **8,** 753–755 (2011).

248. Hockemeyer, D. *et al.* Genetic engineering of human pluripotent cells using TALE nucleases. *Nat. Biotechnol.* **29,** 731–734 (2011).

249. Zwaka, T. P. & Thomson, J. A. Homologous recombination in human embryonic stem cells. *Nat. Biotechnol.* **21,** 319–321 (2003).

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

251. Shalem, O. *et al.* Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **343,** 84–87 (2014).

252. Lakshmipathy, U. *et al.* Efficient Transfection of Embryonic and Adult Stem Cells. *STEM CELLS* **22,** 531–543 (2004).

253. Cordie, T. *et al.* Nanofibrous Electrospun Polymers for Reprogramming Human Cells. *Cell. Mol. Bioeng.* **7,** 379–393 (2014).

254. Qiu, P. *et al.* Mutation detection using Surveyor nuclease. *BioTechniques* **36,** 702–707 (2004).

255. Zhu, X. *et al.* An Efficient Genotyping Method for Genome-modified Animals and Human Cells Generated with CRISPR/Cas9 System. *Sci. Rep.* **4,** (2014).

256. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6,** 1621–1624 (2012).

257. Wittwer, C. T., Reed, G. H., Gundry, C. N., Vandersteen, J. G. & Pryor, R. J. High-Resolution Genotyping by Amplicon Melting Analysis Using LCGreen. *Clin. Chem.* **49,** 853–860 (2003).

258. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74,** 5463–5467 (1977).

259. Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* (2013). doi:10.1038/nbt.2507

260. Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* (2013). doi:10.1038/nbt.2501

261. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32,** 279–284 (2014).

262. Hockemeyer, D. *et al.* Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol.* **27,** 851–857 (2009).

APPENDICES

Appendix A: Supplemental Tables and Figures for Chapter 2

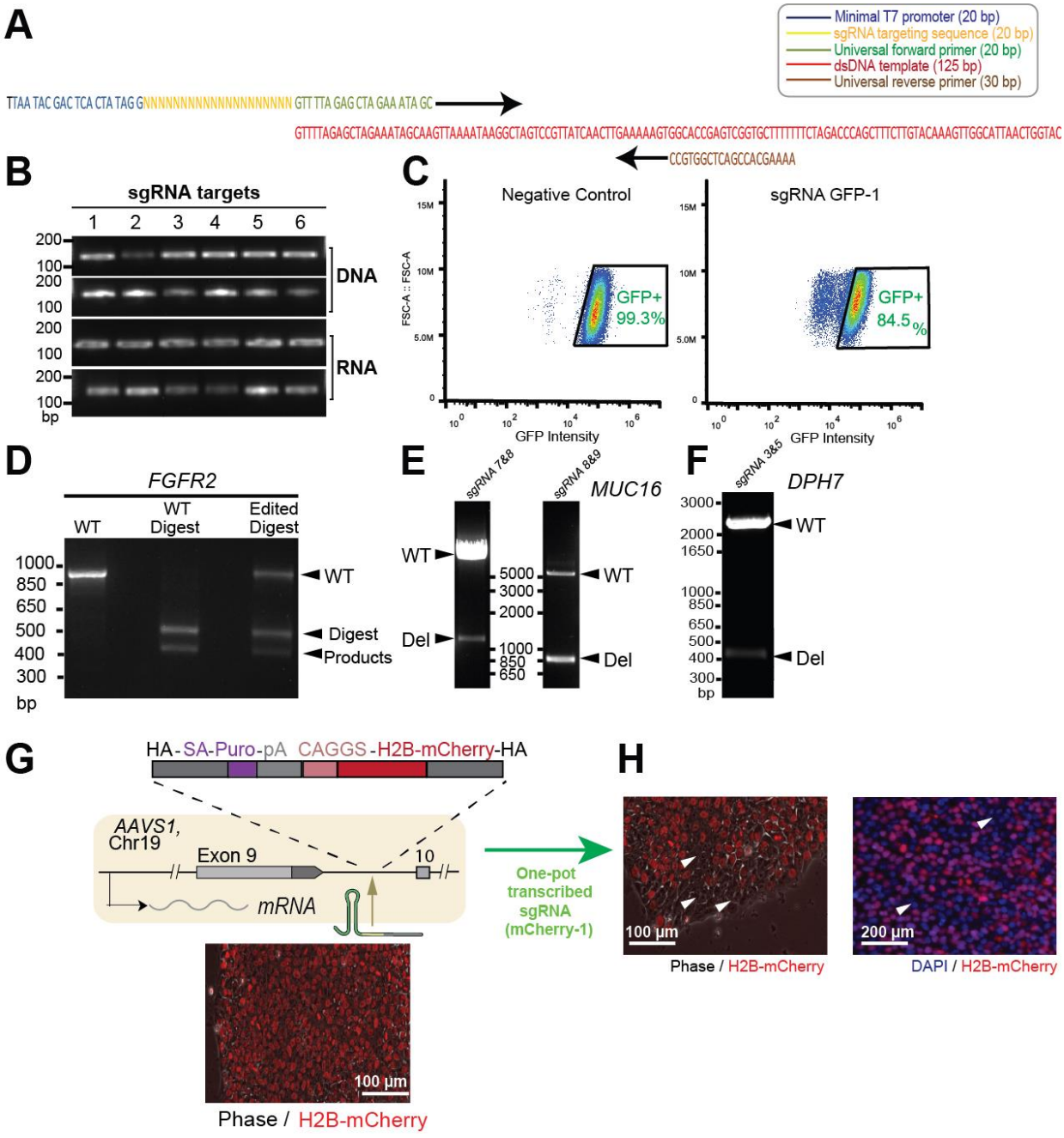
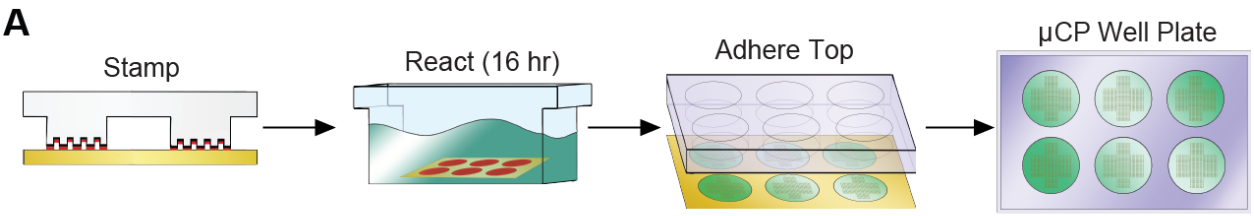
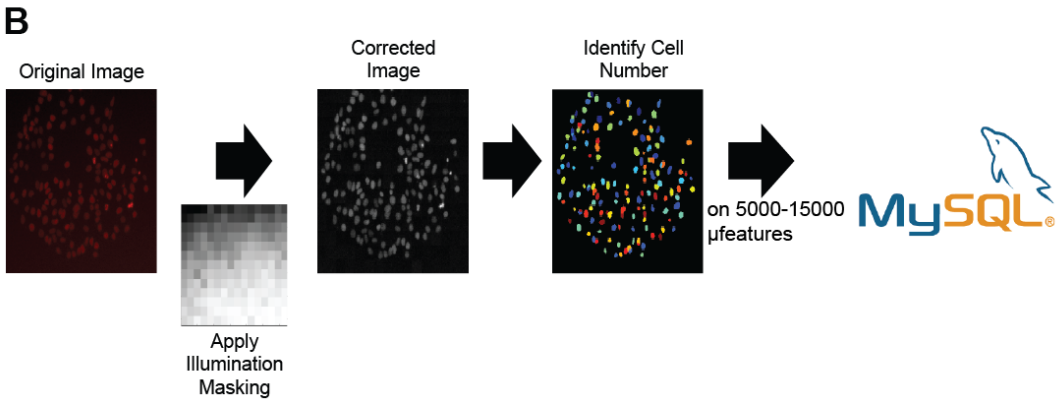

A
Minimal T7 promoter (20 bp)
sgRNA targeting sequence (20 bp)
Universal forward primer (20 bp)
dsDNA template (125 bp)
Universal reverse primer (30 bp)
TTAA TAC GAC TCA CTA TAG GNNNNNNNNNNNNNNNNNNN GTT TTA GAG CTA GAA ATA GC
GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTAACTGGTAC
CCGTGGCTCAGCCACGAAAA
B
sgRNA targets
1 2 3 4 5 6
200
100
200
100
200
100
200
100
bp
DNA
RNA
C
Negative Control
sgRNA GFP-1
FSC-A :: FSC-A
GFP+ 99.3%
GFP+ 84.5%
GFP Intensity
GFP Intensity
D
FGFR2
WT
WT Digest
Edited Digest
1000
850
650
500
400
300
bp
WT
Digest Products
E
sgRNA 7&8
sgRNA 8&9
MUC16
WT
Del
5000
3000
2000
1000
850
650
WT
Del
F
sgRNA 3&5
DPH7
3000
2000
1650
1000
850
650
500
400
300
bp
WT
Del
G
HA-SA-Puro-pA CAGGS-H2B-mCherry-HA
AAVS1, Chr19
Exon 9
10
mRNA
One-pot transcribed sgRNA (mCherry-1)
100 μm
Phase / H2B-mCherry
H
100 μm
Phase / H2B-mCherry
200 μm
DAPI / H2B-mCherry

Figure A-1 | Synthesis and functionality of "one-pot" transcribed single guide RNAs.

(A) Schematic of the PCR step for "one-pot" transcription of sgRNA. The forward primer defines the genomic target (yellow) of editing by Cas9, while the reverse primer encodes the constant regions of the sgRNA. Truncated T7 promoter (blue) is used for in vitro transcription. (**B)** Gels with each column showing DNA and RNA products from one-pot reactions performed twice with six distinct primer sets (targets 1-3: *MCHERRY-1-3*; 4-6: *GFP-1-3*). Each reaction occurred in one well of a 96-well plate. *Top two rows:* Agarose gel of DNA generated by PCR using primers and template shown in part (A). Products are of consistent length and yield, regardless of target. *Bottom two rows:* Denaturing agarose gel following one-pot transcription of DNA shown in top. Products are of a consistent length and yield. **(C)** Flow cytometry results knocking out a constitutive GFP reporter in human embryonic kidney (HEK) cells using sgRNA GFP-1 (>10,000 events; replicate experiments). **(D)** Restriction fragment length polymorphism assay of gene editing at the *FGFR2* locus. Gene modification removes an endogenous PvuI site in *FGFR2*. WT denotes wildtype sequence. WT digest: WT PCR digested using PvuI. WT band is completely removed in digested lane. Edited digest denotes hESCs transfected with sgRNA FGFR2-1. WT band remains suggesting the removal of the PvuI cut site via gene editing. **(E, F)** Large-scale genome edits in HEK cells in (E) *MUC16* and (F) *DPH7*. The presence of a WT band and a smaller second band show that 2-4 kb have been cut out of a population in the analyzed cells. **(G)** Generation of WA09-H2B-mCherry line. *Top*: Donor plasmid for histone2B-mCherry insertion used. Constitutive expression from a safe-harbour, intronic *AAVS1* locus was employed: CAAGS: synthetic promoter, Puro: puromycin resistance, SA: splice acceptor, HA: homology arm; pA: poly(A) tail. *Bottom*: Overlay of phase and red fluorescence channels of WA09 hESCs with transgene. All nuclei contain red fluorescence. **(H)** Images of WA09-H2B-mCherry hESCs on standard culture plates four days after electroporation with Cas9-encoding plasmid and one-pot sgRNAs against mCherry transgene. Arrowheads: cells that lost mCherry fluorescence due to gene-editing.





C

Frame	Day2Count	Day3Count	Day4Count	Day5Count	Day6Count
216	1	1	2	2	1
227	5	13	19	43	77
247	4	11	1	4	7
255	4	8	12	37	83
261	3	7	5	8	13
267	10	6	8	10	10
269	3	1	7	4	3
279	3	7	8	10	7
280	1	3	2	3	14
289	2	7	8	8	6
296	9	7	16	11	16

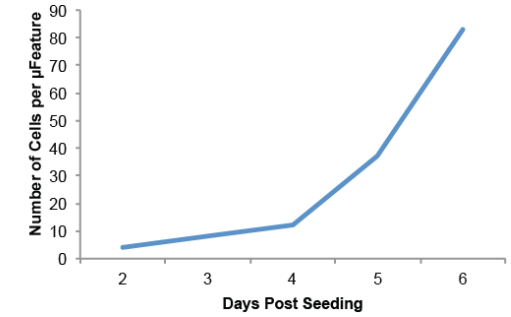
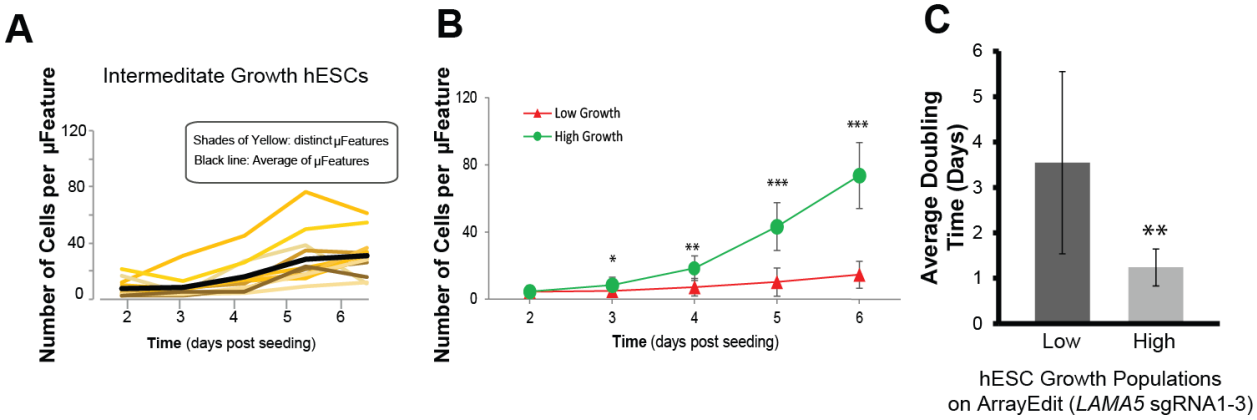


Figure A-2 | Microcontact printing permits isolation of gene-edited hESCs on ArrayEdit. **(A)** Schematic of μcontact printing methodology. Alkanethiol self-assembled monolayers are deposited onto gold-coated glass via a polydimethylsiloxane (PDMS) stamp. Stamped glass was then submerged in poly (ethylene-glycol) (PEG) solution overnight. Finally, reacted glass was attached to a standard well plate. **(B)** Detail of image analysis pipeline. First, an illumination mask is applied in CellProfiler to even out bright and dark areas. Second, the number of nuclei is identified. For each μFeature, that number, as well as spatial location, is exported to a MySQL database. **(C)** Representative table of growth data from MySQL Workbench. **(D)** Example of tracking growth data using highlighted data (blue in part c) from one μFeature over time.



D

hESC Growth Population on ArrayEdit (*LAMA5* sgRNA1-3)	***LAMA5* around sgRNA1 target**	
	TTCAGCGTGTCCATCTCGATGCAGCCCCGGTAGCCGGGGAAGCGAAGCAGGGGAGGGGGCTGTGGG	**WT**
Intermediate Growth Clone 1	TTCAGCGTGTCCATCTCGATGCAGCCCCNGTAGCCGGGGAAGCGAAGCAGGGGAGGGGNCTGTGGG	
Intermediate Growth Clone 2	TTCAGCGTGTCCATCTCGATGCAGCCCCGGTAGCCGGGGAAGCGAAGCAGGGGAGGGGGCTGTGGG	
Intermediate Growth Clone 3	TTCAGCGTGTCCATCTCGATGCAGCCCCGGTAGCCGGGGAAGCGAAGCAGGGGAGGGGGCTGTGGG	
Intermediate Growth Clone 4	TTCAGCGTGTCCATCTCGATGCAGCCCCGGTAGCCGGGGAAGCGAAGCAGGGGAGGGGGCTGTGGG	
Intermediate Growth Clone 5	TTCAGCGTGTCCATCTCGATGCAGCCCCGGTAGCCGGGGAAGCGAAGCAGGGGAGGGGGCTGTGGG	
High Growth Clone 1	TTCAGCGTGTCCATCTCGATGCAGCCCCNGTAGCCGGGGAAGCGAAGCAGGGGAGGGGNCTGTGGG	
High Growth Clone 2	TTCAGCGTGTCCATCTCGATGCAGCCCCGGTAGCCGGGGAAGCGAAGCAGGGGAGGGGGCTGTGGG	
High Growth Clone 3	TTCAGCGTGTCCATCTCGATGCAGCCCCGGTAGCCGGGGAAGCGAAGCAGGGGAGGGGGCTGTGGG	
High Growth Clone 4	TTCAGCGTGTCCATCTCGATGCAGCCCCGGTAGCCGGGGAAGCGAAGCAGGGGAGGGGGCTGTGGG	
High Growth Clone 5	TTCAGCGTGTCCATCTCGATGCAGCCCCGGTAGCCGGGGAAGCGAAGCAGGGGAGGGGGCTGTGGG	

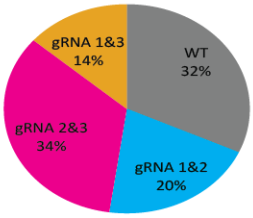
E

Population	Wildtype	Single Allelic Edit	Biallelic Edit
High	12/12 (100%)	0/12 (0%)	0/12 (0%)
Medium	12/12 (100%)	0/12 (0%)	0/12 (0%)
Low	4/22 (18%)	10/22 (46%)	8/22 (36%)

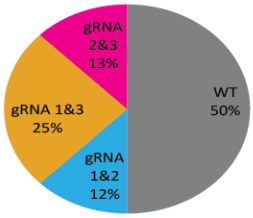
Figure A-3 | LAMA5-edited hESC lines grouped by high-content growth analysis on ArrayEdit.

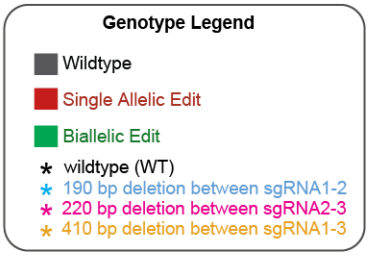
(A) Growth curves for intermediate growth rate grouped cells within 12 µFeatures on ArrayEdit over 5 consecutive days. See Figure 3E for curves of the low and high growth populations. Yellow lines are individual populations. Black line is the average of all populations in the intermediate group. **(B)** Growth curves of the high and low growth populations. Data are represented as mean ± 95% C.I. from 12 independent µFeatures on ArrayEdit. Starting at day 3, the populations have significantly different number of cells (Student's t-test, *p<0.05, **p<5x10^{-5}, ***p<5x10^{-10}) **(C)** Doubling times for both high and low growth population observed on ArrayEdit. Data are represented as mean ± 95% C.I. from 12 independent µFeatures on ArrayEdit. Low growth populations have significantly higher doubling times than high growing populations (p<5x10^{-5}, Student's t-test). **(D)** Sanger sequencing analysis of representative intermediate and high growth hESC lines isolated from ArrayEdit. Wildtype is denoted on top, and the hESC lines are below. sgRNA targets are denoted in red and PAM in blue. No edits were observed in these cells, as well as all of the other 14 clones isolated from the intermediate and high growth populations. **(E)** Table of observed genotypes in the high, intermediate, and low growth groups. No edits were observed in either the high or intermediate growth groups.

A cut alleles - observed



B cut alleles - expected





C

Predicted top off-target site of sgRNA2

```
          TTATTCTGGGTGGGTCCATGGCCCGGGTAGGAGCCCTGAGTGAGTGTCCATTCCAAGGGGACGAGT   WT
C8        TTATTCTGGGTGGGTCCATGGCCCGGGTAGGAGCCCTGAGTGAGTGTCCATTCCAAGGGGACGAGT
A10       TTATTCTGGGTGGGTCCATGGCCCGGGTAGGAGCCCTGAGTGAGTGTCCATTCCAAGGGGACGAGT
A13       TTATTCTGGGTGGGTCCATGGCCCGGGTAGGAGCCCTGAGTGAGTGTCCATTCCAAGGGGACGAGT
A20       TTATTCTGGGTGGGTCCATGGCCCGGGTAGGAGCCCTGAGTGAGTGTCCATTCCAAGGGGACGAGT
G20       TTATTCTGGGTGGGTCCATGGCCCGGGTAGGAGCCCTGAGTGAGTGTCCATTCCAAGGGGACGAGT
```

Predicted top off-target site of sgRNA3

```
          TGGGACTAGGAGCAGAAGCCTGCCAGAAAAAGGCCCTCCCAAGGCTTTGGCCTATAAGACAGTGGA   WT
C8        TGGGACTAGGAGCAGAAGCCTGCCAGAAAAAGGCCCTCCCAAGGCTTTGGCCTATAAGACAGTGGA
A10       TGGGACTAGGAGCAGAAGCCTGCCAGAAAAAGGCCCTCCCAAGGCTTTGGCCTATAAGACAGTGGA
A13       TGGGACTAGGAGCAGAAGCCTGCCAGAAAAAGGCCCTCCCAAGGCTTTGGCCTATAAGACAGTGGA
A20       TGGGACTAGGAGCAGAAGCCTGCCAGAAAAAGGCCCTCCCAAGGCTTTGGCCTATAAGACAGTGGA
G20       TGGGACTAGGAGCAGAAGCCTGCCAGAAAAAGGCCCTCCCAAGGCTTTGGCCTATAAGACAGTGGA
```

Predicted off-target site of sgRNA1

```
          GTGCAAAGATGAGTGGGACCAATCCCCGGGGCAGCGAAGCAGAGAGAAGGAGCCTGGCCTAGGACT   WT
C8        GTGCAAAGATGAGTGGGACCAATCCCCGGGGCAGCGAAGCAGAGAGAAGGAGCCTGGCCTAGGACT
A10       GTGCAAAGATGAGTGGGACCAATCCCCGGGGCAGCGAAGCAGAGAGAAGGAGCCTGGCCTAGGACT
A13       GTGCAAAGATGAGTGGGACCAATCCCCGGGGCAGCGAAGCAGAGAGAAGGAGCCTGGCCTAGGACT
A20       GTGCAAAGATGAGTGGGACCAATCCCCGGGGCAGCGAAGCAGAGAGAAGGAGCCTGGCCTAGGACT
G20       GTGCAAAGATGAGTGGGACCAATCCCCGGGGCAGCGAAGCAGAGAGAAGGAGCCTGGCCTAGGACT
```

Predicted off-target site of sgRNA2

```
          GCTCTGAACTGCGGCCTCTGTTTCAGGTACGGCTCCTACCTGGTCTGGAAAGAGCTGGGAGGCTT   WT
C8        GCTCTGAACTGCGGCCTCTGTTTCAGGTACGGCTCCTACCTGGTCTGGAAAGAGCTGGGAGGCTT
A10       GCTCTGAACTGCGGCCTCTGTTTCAGGTACGGCTCCTACCTGGTCTGGAAAGAGCTGGGAGGCTT
A13       GCTCTGAACTGCGGCCTCTGTTTCAGGTACGGCTCCTACCTGGTCTGGAAAGAGCTGGGAGGCTT
A20       GCTCTGAACTGCGGCCTCTGTTTCAGGTACGGCTCCTACCTGGTCTGGAAAGAGCTGGGAGGCTT
G20       GCTCTGAACTGCGGCCTCTGTTTCAGGTACGGCTCCTACCTGGTCTGGAAAGAGCTGGGAGGCTT
```

Predicted off-target site of sgRNA2

```
          TACAGGGATGACAGTCTCCACCTTACGGGAGCCTCCTACCCAGGGGATGGAGTGAGACAGGCAGG   WT
C8        TACAGGGATGACAGTCTCCACCTTACGGGAGCCTCCTACCCAGGGGATGGAGTGAGACAGGCAGG
A10       TACAGGGATGACAGTCTCCACCTTACGGGAGCCTCCTACCCAGGGGATGGAGTGAGACAGGCAGG
A13       TACAGGGATGACAGTCTCCACCTTACGGGAGCCTCCTACCCAGGGGATGGAGTGAGACAGGCAGG
A20       TACAGGGATGACAGTCTCCACCTTACGGGAGCCTCCTACCCAGGGGATGGAGTGAGACAGGCAGG
G20       TACAGGGATGACAGTCTCCACCTTACGGGAGCCTCCTACCCAGGGGATGGAGTGAGACAGGCAGG
```

Predicted off-target site of sgRNA3

```
          GCCTCCCACCACTGGGGCCGGAAAAAGGCCGACCCTCTGGAGGGGCCAAGGGGGAGGGGTAGCAG   WT
C8        GCCTCCCACCACTGGGGCCGGAAAAAGGCCGACCCTCTGGAGGGGCCAAGGGGGAGGGGTAGCAG
A10       GCCTCCCACCACTGGGGCCGGAAAAAGGCCGACCCTCTGGAGGGGCCAAGGGGGAGGGGTAGCAG
A13       GCCTCCCACCACTGGGGCCGGAAAAAGGCCGACCCTCTGGAGGGGCCAAGGGGGAGGGGTAGCAG
A20       GCCTCCCACCACTGGGGCCGGAAAAAGGCCGACCCTCTGGAGGGGCCAAGGGGGAGGGGTAGCAG
G20       GCCTCCCACCACTGGGGCCGGAAAAAGGCCGACCCTCTGGAGGGGCCAAGGGGGAGGGGTAGCAG
```

Predicted off-target site of sgRNA3

```
          TAATTGAGGATGGAAATCCAGAAAAAGGCCGCCCCTCAGCAGCAGGGGATAACGAGGACCCTTTC   WT
C8        TAATTGAGGATGGAAATCCAGAAAAAGGCCGCCCCTCAGCAGCAGGGGATAACGAGGACCCTTTC
A10       TAATTGAGGATGGAAATCCAGAAAAAGGCCGCCCCTCAGCAGCAGGGGATAACGAGGACCCTTTC
A13       TAATTGAGGATGGAAATCCAGAAAAAGGCCGCCCCTCAGCAGCAGGGGATAACGAGGACCCTTTC
A20       TAATTGAGGATGGAAATCCAGAAAAAGGCCGCCCCTCAGCAGCAGGGGATAACGAGGACCCTTTC
G20       TAATTGAGGATGGAAATCCAGAAAAAGGCCGCCCCTCAGCAGCAGGGGATAACGAGGACCCTTTC
```

Figure A-4 | Genotypes observed in edited hESC lines isolated from ArrayEdit.

(A) Summary of genotypes obtained from hESC lines isolated after gene-editing using ArrayEdit or standard procedures (from screening gels in Figures 4C; see text), to edit *LAMA5*. **(B)** Calculated distribution of cut alleles from a probability tree. Construction of the tree made the following assumptions: all sgRNAs cut at the same frequency and that any sgRNA that enters the cell will generate edits. **(C)** Sanger sequencing around loci of top predicted off-target sites of LAMA5 sgRNAs 1, 2 and 3 as determined by bioinformatics. Potential off-target sites are colored. All sequenced clones show perfect alignment to reference genotype at the off-target sites.

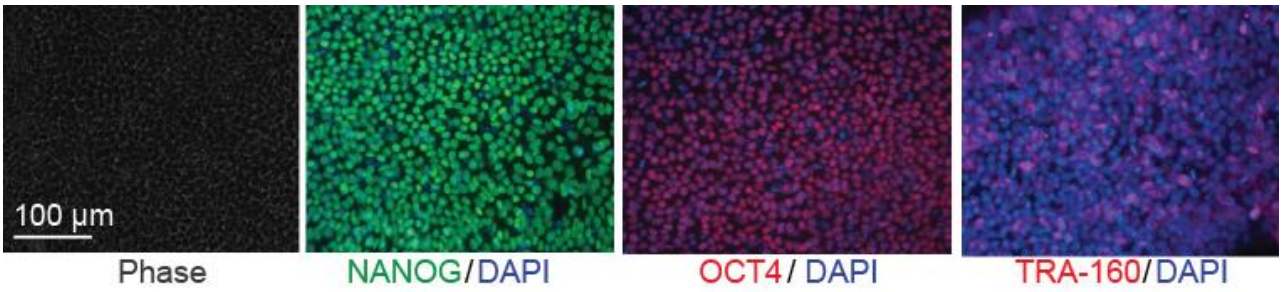


Figure A-5 | Pluripotency characterization of hESCs after isolation from ArrayEdit. Immunocytochemical assay for pluripotency proteins (Oct4, Sox2, TRA-1-60) within hESCs after isolation from ArrayEdit. Live, gene-edited hESCs isolated from platform retain pluripotent marker expression over >10 passages.

Table A-1 | Primer Sequences Used In This Study, Related To Figure 1. All sequences are listed as 5’ to 3’.

Reverse primer, RNAtracrR AAAAGCACCGACTCGGTGCC
dsDNA Template GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG GCACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTAACTGGTAC
Genomic Amplification
mCherry AAGGGCGAGGAGGATAACATGG

TTGTACAGCTCGTCCATGCCG

MUC16

AGTGAGGACACTCAGAACCAC

ATGTTGGGGTCTCAGATGGTG

DPH7

ACTGCAGCCCCGTTGTTAAG

AAAGGTCTTCAAGATCCCAACA

CDH5

GACTCCTCCAGCTTCACCA

CACGGACGCATTGAACAACC

CDH20

ATCCAGACAGTGAGTGCGGT

CCTCAGGAGCCAAGCTGTAG

FOXP2

AGCATCTGCTCAGCCTTCAG

TTGAGGCAGCGATTGGACAG

NANOG

AAGGCCTCAGCACCTACCTA

GAAGGTTCCCAGTCGGGTTC

OCT4

GTGGTCCGAGTGTGGTT

GAAAGGAGACCCAGCAG

PAX6

AAGCAAAATAGCCCAGTATAAG

TATGTTATCGTTGGTACAGACC

POU4F1

CCTCACCCGCATATGCACAG

CCCGGACGGCATGTTCA

LAMA5 Genomic

CCTCTGCCCTCAGGACTCT

GGCCAGCTCCAGATCATTGT

LAMA5 Off-target 1

TCAGAGCAGTTGCCTTCCCA

TTTGGCATCCTTCGGCCTT

LAMA5 Off-target 2

GAGTCTCGGCAGATGGTAGC

TCAGCGTGAGCACACTCATT

LAMA5 Off-target 3

CTCTACACCAAAGAGCGGCA

GCAATGCCAGATGGGTCAAC

LAMA5 Off-target 4
ACACTGCATTCTGAAGCCCA

GCTGGCCGGCTGAGG

LAMA5 Off-target 5
TCCGCAAGGGTTTCTTGGTA

TTTCAGGAGCTGTTCTGGGAC

LAMA5 Off-target 6
CAGGCCAAATGTCCCATCCT

ACAGTACCAGGCTTCGTGTG

LAMA5 Off-target 7
ACATGGGCCATAGGAGATGC

CCCAGGTGGGACTCTCAAAC

LAMA5 Off-target 8
TCCCTCCCTTTGATGGGTTG

CCGCTGATTGAAAGCCTCTT

LAMA5 Off-target 9
CCGCTACTTGCCAAACACAT

CTCCATGTCCAGGGTTCTGC

Table A-2 | Genomic targets for sgRNAs produced using one-pot transcription. All sequences are listed as 5' to 3'. (Labels for each sgRNA target in text are in parentheses.) Synthesized forward primers were `TTAATACGACTCACTATAGG-Target_sequence_below-GTTTTAGAGCTAGAAATAGC`.

LAMA5

GTAGCCGGGGAAGCGAAGCA (LAMA5-1)

GCTCACGGACGGCTCCTACC (LAMA5-2)

GCAGTGGGACGGCCTTTTTC (LAMA5-3)

MCHERRY

GGAGCCGTACATGAACTGAG (mCherry-1)

CTCGAACTCGTGGCCGTTCA (mCherry-2)

GCGGTCTGGGTGCCCTCGTA

CGCCCTCGATCTCGAACTCG

GGGCGAGGGCCGCCCCTACG

AGTAGTCGGGGATGTCGGCG (mCherry-3)

GCGTTCGTACTGTTCCACGA

CTGTCCCCTCAGTTCATGTA

GFP

GACCAGGATGGGCACCACCC (GFP-1)

GCACGGGGCCGTCGCCGATG (GFP-2)

GTGAACCGCATCGAGCTGAA (GFP-3)

CTCGAACTCGTGGCCGTTCA

NANOG

GTCTTCAGGTTGCATGTTCA

AAGTCACTGGCAGGAGAATT

GTATTCGTATTGTTTGGGAT

GTGCTAATCTTTGTAGAAAG

CAGTCGGATGCTTCAAAGCA

CTTCAAAGCAAGGCAAGCTT

AGAGAAGAGTGTCGCAAAAA

GATGCCTGGTGAACCCGACT

CTTTACTCTTCCTACCACCA (NANOG-1)

SOX2

GTGCCGTTAATGGCCGTGCC

GTTCCGGCACCTCGGCGCCG

GTGCCCGGCACGGCCATTAA

ACAACTACGGAAAATAAAGG

CAACTACGGAAAATAAAGGG

OCT4

ATCGCTTGCCCTTCTGGCGC

TGCTTGATCGCTTGCCCTTC

GAACTTAATCCCAAAAACCC

CCTGTCCCCCATTCCTAGAA

TCTCTTTCGGGCCTGCACGA

CAGCGACTATGCACAACGAG (OCT4-1)

FGFR2

GGCCGGTGAGGCGATCGCTC

GTGGCCTCGGTACCGGAGTG (FGFR2-1)

CACCAGAGCGATCGCCTCAC

PAX6

TTATTATAGAAATCATTCTG

CGAGACAGATTACTGTCCGA (PAX6-1)

OTX2

GCCTAGCAGTAAAGAGACAT

ACCAATGTCTCTTTACTGCT

MUC16

GTCACCAACCGTGATACAGC

GCTATCGTCTTTGGTTCACC

GTGGGGACTTAGTCAGTGTT

GACCCCACTGGTCACTATTT

GAAAACAGATGGGGTTGTCC

GCCAATGAGGATGTCGTTTC

GGTTTTACTGAAACCCCTGA (MUC16-7)

TCACCCCACTGACAACCACG (MUC16-8)

CGGGAGCTCTGACATCCTGA (MUC16-9)

CDC42

AAGTGTGTTGTTGTGGGCGA

GGTGGAGAACCATATACTCT

TCTGTTTGTGGATAACTCAG

CACAACAAACAAATTTCCAT

GCAGTCACAGTTATGATTGG

ATTTCTAGTCTGTTTTTCAG

DPH7

GCACCACTCCACCGAGTCCG

TTCTTTGGACCTCGACCAGA

ATCCATACAACTGCTCCGCC (DPH7-3)

TTTCCCAGTGGACCAATCTA

ATGGCAGGCACATCAATTCG (DPH7-5)

GGCGACGATGGCCTTCTGAG

CCAGGATGTGCTCCCGATGA

ACGAAACATGAAGCAGCCGT

GGTCTCTTGCCTCGGTGCCT

GTTCACGCCACAGTCACGTG

CDH20

TTACCTTGCCGACATACAAA

AGAAATGCTGACCATTGCGT (CDH20-1)

CDH5

ACTGGCCTGTGTTCACGCAT

FMR1

CCTTCCTTCCAATAATTCAA

FOXP2

GGACTCATCTCCATTCCACC

GAPDH

TATTGGGCGCCTGGTCACCA (GAPDH-1)

ACTGTGGCGTGATGGCCGCG

POU4F1

GGGCCACCTGTCGCACCCCG

Table A-3 | Nucleic acid-related reagent cost estimates for implementing ArrayEdit. Standard protocol uses cloning of sgRNA into a plasmid backbone (e.g. Mali et al., Science, 2013).

	ArrayEdit	Standard
gBlock	$2.11	--
Forward Primer	$12.00	$12.00
Reverse Primer	--	$12.00
Polymerase	$0.84	$0.84
IVT	$2.12	--
Restriction Enzyme	--	$0.31
Gibson Assembly	--	$15.90
Competent Bacteria	--	$24.00
Miniprep	--	$1.63
Reagent Cost	$17.07	$66.68
Number of uses	20	2
Cost per experiment	**$0.85**	**$33.34**

Code A-1

Below is an example of a code used in MySQL Workbench for implementation of ArrayEdit. Comments are preceded by a # and in grey.

```
#select all tables from each day of experiment, Image_Frame tables contain data about each image in series

SELECT Day6.Image_Frame_TxRed as Frame, Day2.Image_Count_H9_H2B as Day2Count,
Day3.Image_Count_H9_H2B as Day3Count, Day4.Image_Count_H9_H2B as Day4Count,
Day5.Image_Count_H9_H2B as Day5Count, Day6.Image_Count_H9_H2B as Day6Count

FROM cellprofilerdb1.JCS2_5_LamA5KOPer_Image Day2


#join together each day on the basis of the frame. Frame is consistent from day today as it conveys spatial
information within the well. Uses one day as master table to join other tables to, decreasing the likelihood of error

join cellprofilerdb1.JCS2_5_LamA5KOPer_Image Day3 on Day2.Image_Frame_TxRed=Day3.Image_Frame_TxRed

join cellprofilerdb1.JCS2_5_LamA5KOPer_Image Day4 on Day2.Image_Frame_TxRed=Day4.Image_Frame_TxRed

join cellprofilerdb1.JCS2_5_LamA5KOPer_Image Day5 on Day2.Image_Frame_TxRed=Day5.Image_Frame_TxRed

join cellprofilerdb1.JCS2_5_LamA5KOPer_Image Day6 on Day2.Image_Frame_TxRed=Day6.Image_Frame_TxRed

#when using 6-well plates, each well must be imaged individually and is named by the position of the well.  This
check ensures that only like wells will be joined together across days. Wildcards are used in case a day was
misnamed with extra spaces or trailing characters.  Standard notation is DayX-WellXX.

where Day2.Image_FileName_TxRed like '%2-WellB2%' and Day3.Image_FileName_TxRed like '%3-WellB2%' and
Day4.Image_FileName_TxRed like '%4-WellB2%' and Day5.Image_FileName_TxRed like '%5-WellB2%'and
Day6.Image_FileName_TxRed like '%6-WellB2%'


#Discard any wells that are not of interested of may produce misleading data due to the presence of 0's.

and Day2.Image_Count_H9_H2B != 0 and Day3.Image_Count_H9_H2B!=0 and Day4.Image_Count_H9_H2B!=0
and Day5.Image_Count_H9_H2B!=0 and Day6.Image_Count_H9_H2B!=0

order by Day2.Image_Frame_TxRed;
```

Appendix B: Supplemental Tables and Figures for Chapter 3

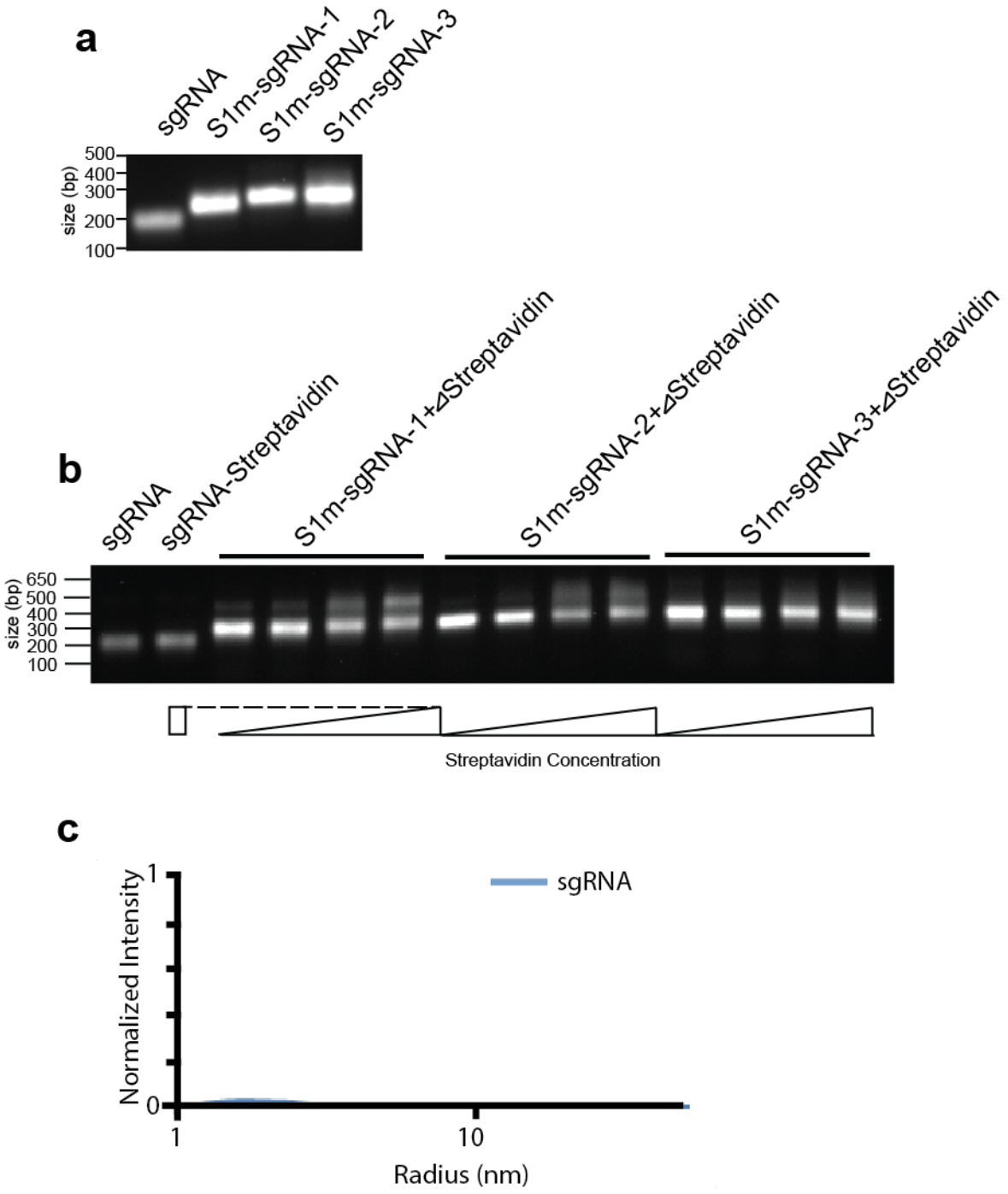

a
sgRNA
S1m-sgRNA-1
S1m-sgRNA-2
S1m-sgRNA-3
size (bp)
500
400
300
200
100
b
sgRNA
sgRNA-Streptavidin
S1m-sgRNA-1+⊿Streptavidin
S1m-sgRNA-2+⊿Streptavidin
S1m-sgRNA-3+⊿Streptavidin
size (bp)
650
500
400
300
200
100
Streptavidin Concentration
c
sgRNA
Normalized Intensity
1
0
1
10
Radius (nm)

Figure B-1 | Creation and in vitro testing of S1m-sgRNA.

(a) *In vitro* transcription of S1m-sgRNAs compared to standard sgRNAs. S1m-sgRNAs are larger than sgRNAs due to the insertion of S1m stem loop. **(b)** *In vitro* complexes of sgRNAs and streptavidin. Lane 1: sgRNA. Lane 2: sgRNA-streptavidin. Addition of streptavidin did not shift the electrophoretic front. Lanes 3-6, 7-10, 11-14: Progressive ratios of each S1m-sgRNA streptavidin (0, 0.1, 1, and 10 molar equivalents of streptavidin). As streptavidin concentration was increased the electrophoretic front of S1m-sgRNAs was slowed. The presence of several bands may be due to multiple S1m-sgRNAs binding to a single streptavidin. **(c)** DLS trace of free sgRNA alone in solution. Nucleic acids are not robustly detected by this DLS instrument.

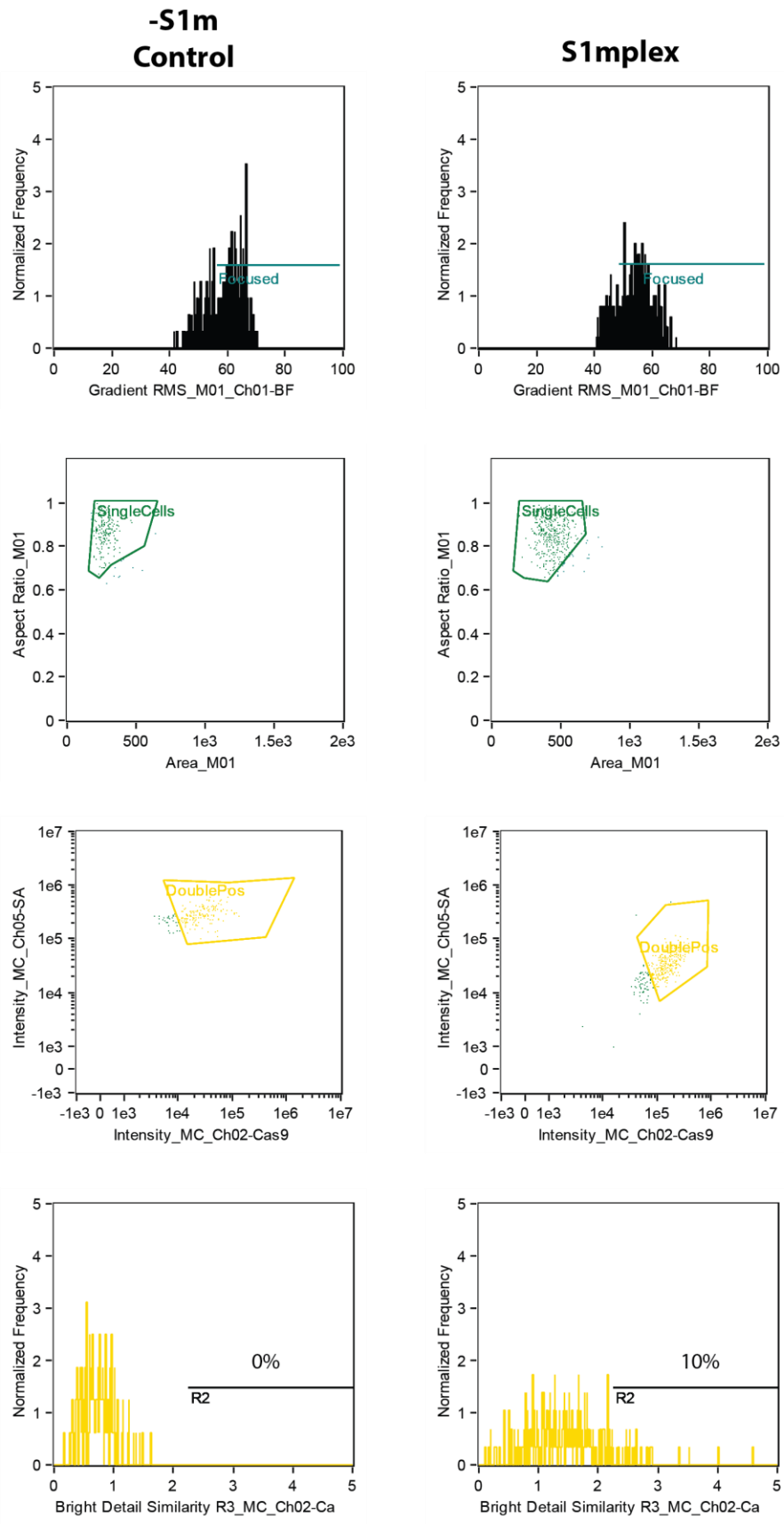

-S1m
Control
S1mplex
Normalized Frequency
Gradient RMS_M01_Ch01-BF
Focused
SingleCells
Aspect Ratio_M01
Area_M01
DoublePos
Intensity_MC_Ch05-SA
Intensity_MC_Ch02-Cas9
0%
10%
R2
Bright Detail Similarity R3_MC_Ch02-Ca

Figure B-2 | Multispectral imaging flow cytometric analysis.

Gating strategy used by multispectral imaging flow cytometry analysis for colocalization of Cas9 and streptavidin. hPSCs were first gated to be in focus. Next, the aspect ratio and area measured from brightfield (Ch1) images were plotted against each other to identify single cells. Then, intensities of Cas9 (Ch2) and streptavidin (Ch5) were plotted to identify double positive cells. Finally, IDEAS software wizard identified cells with high similarity between Ch2 and Ch5 (Bright Detail Similarity).

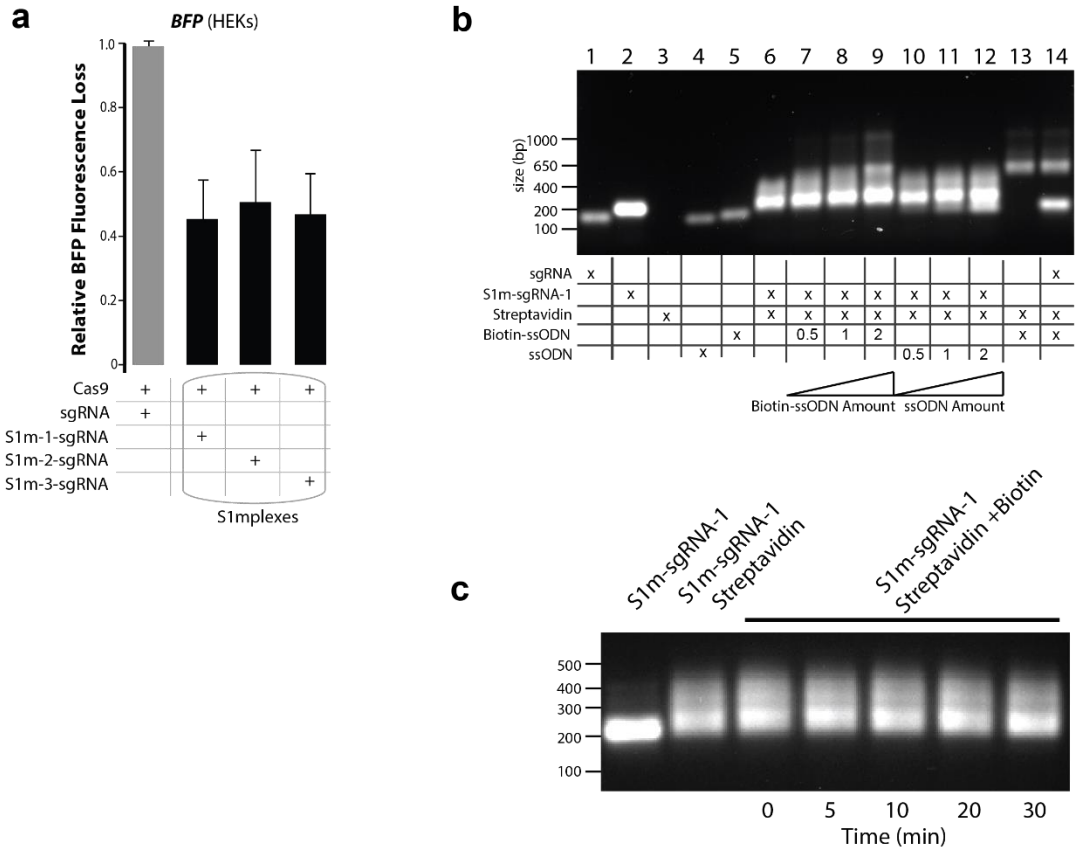

a
BFP (HEKs)
Relative BFP Fluorescence Loss
1.0
0.8
0.6
0.4
0.2
0
Cas9
sgRNA
S1m-1-sgRNA
S1m-2-sgRNA
S1m-3-sgRNA
S1mplexes
b
1 2 3 4 5 6 7 8 9 10 11 12 13 14
size (bp)
1000
650
400
200
100
sgRNA
S1m-sgRNA-1
Streptavidin
Biotin-ssODN
ssODN
0.5
1
2
Biotin-ssODN Amount
ssODN Amount
c
S1m-sgRNA-1
S1m-sgRNA-1
Streptavidin
S1m-sgRNA-1
Streptavidin+Biotin
500
400
300
200
100
0
5
10
20
30
Time (min)

Figure B-3 | Gene editing activity of S1m-sgRNAs and in vitro assembly of S1m-sgRNA-1 with streptavidin and ssODNs.

(a) Knockout of integrated BFP fluorescence in human embryonic kidney (HEK) cells. When transfected together with a plasmid encoding *Sp*.Cas9, S1m-sgRNAs induced ~50% the level of NHEJ as sgRNA as measured by the loss of fluorescence five days post transfection. **(b)** *In vitro* tertiary complexes of S1m-sgRNA-1, streptavidin, and ssODN. Lanes 1-5: Components of ssODN-S1mplex particles ran individually. Lanes 6-9: complexes of S1m-sgRNA-1, streptavidin, and biotinylated ssODNs. Numbers represent relative stoichiometry between components ran on gel. Major bands showing the complex of all three components can be seen. Lanes 10-12: complexes of S1m-sgRNA-1, streptavidin, and free ssODNs. ssODNs do not interfere with the binary S1m-sgRNA-1/streptavidin complex. Lane 13-14: complexes of streptavidin and biotin-ssODNs, with free sgRNAs. None of the typical S1m-sgRNA-1-streptavidin complexes can be seen in this lane. **(c)** Competition of biotin with S1m-sgRNA-1 after binding to streptavidin protein. 4-fold excess biotin was added to S1m-sgRNA-1-streptavidin complexes and incubated for 0, 5, 10, 20, and 30 minutes. No significant change was seen even after 30 minutes of competition.

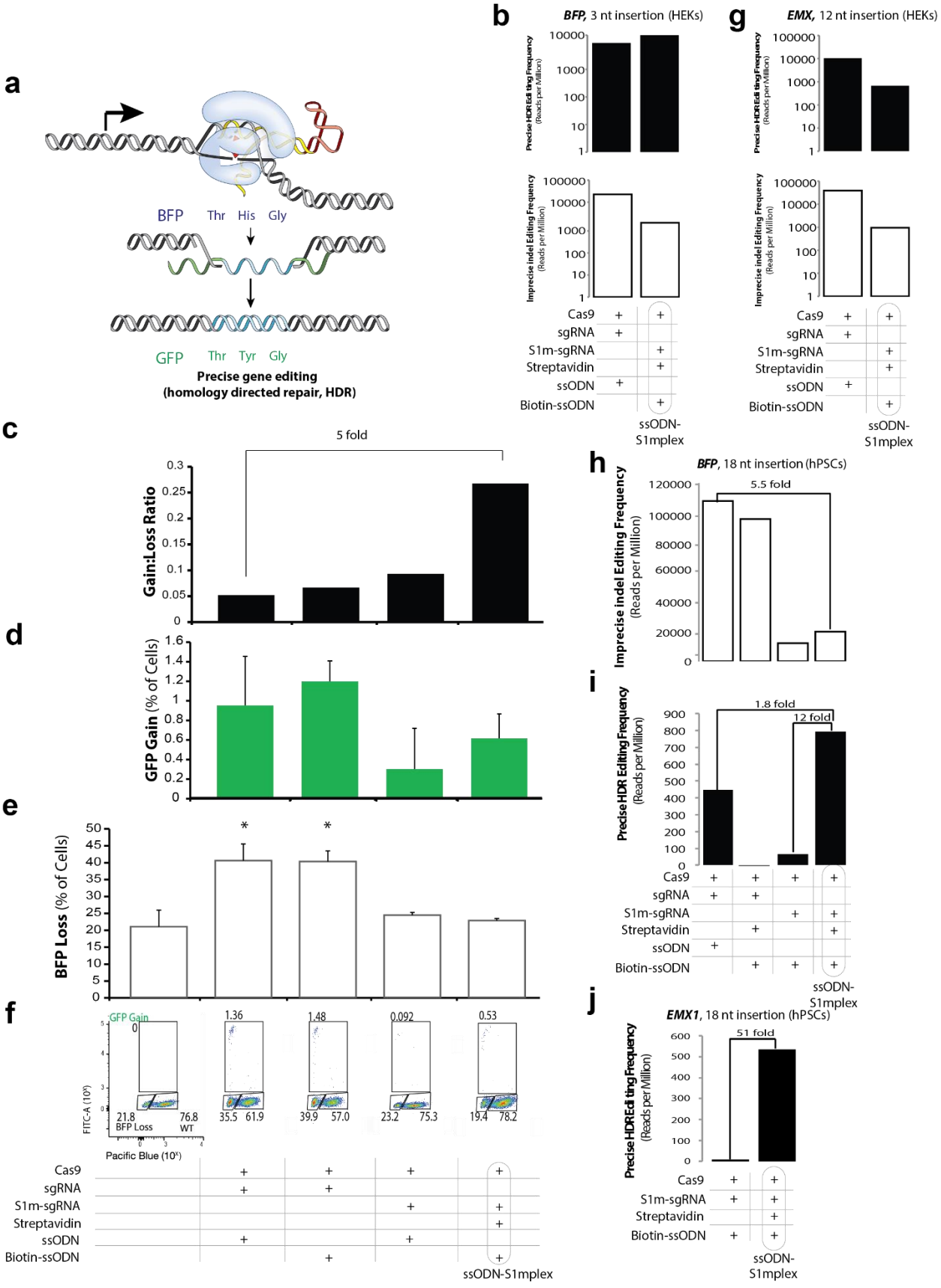

a
BFP Thr His Gly
GFP Thr Tyr Gly
Precise gene editing
(homology directed repair, HDR)
b
BFP, 3 nt insertion (HEKs)
Precise HDR Editing Frequency (Reads per Million)
Imprecise indel Editing Frequency (Reads per Million)
Cas9
sgRNA
S1m-sgRNA
Streptavidin
ssODN
Biotin-ssODN
ssODN-S1mplex
g
EMX, 12 nt insertion (HEKs)
c
5 fold
Gain:Loss Ratio
d
GFP Gain (% of Cells)
e
BFP Loss (% of Cells)
f
GFP Gain
1.36
1.48
0.092
0.53
21.8 BFP Loss
76.8 WT
35.5 61.9
39.9 57.0
23.2 75.3
19.4 78.2
FITC-A (10³)
Pacific Blue (10³)
ssODN-S1mplex
h
BFP, 18 nt insertion (hPSCs)
5.5 fold
i
1.8 fold
12 fold
j
EMX1, 18 nt insertion (hPSCs)
51 fold

Figure B-4 | Increased ratio of precise to imprecise editing at the *BFP* locus using ssODN-S1mplexes.

(a) Schematic showing HDR conversion of *BFP* locus to GFP using ssODN-S1mplex particles. **(b)** *Top:* Number of precise insertion reads at *BFP* locus. S1mplexes had higher levels of insertion than sgRNA RNPs. *Bottom:* Number of indel reads at *BFP* locus. S1mplexes had lower levels of imprecise edits than sgRNA RNPs. **(c)** Ratio of GFP gain to BFP loss using four different editing methods. ssODN-S1mplexes had 5-fold higher ratio of gain:loss than free ssODN and gRNA. **(d)** ssODN-S1mplexes had similar levels of GFP gain when compared to free ssODN and gRNA. Using S1m-sgRNAs with free ssODN had low levels of HDR (n=3 biological replicates). **(e)** Standard sgRNAs caused elevated levels of BFP loss when compared to the control cell lines. S1m-sgRNAs did not result in significant levels of NHEJ (n=3 biological replicates). **(f)** Representative flow cytometry plots from all conditions. **(g)** *Top:* Number of precise insertion reads at *EMX1* locus. S1mplexes had lower levels of insertion than sgRNA RNPs. *Bottom:* Number of indel reads at *EMX1* locus. S1mplexes had significantly lower levels of imprecise edits than sgRNA RNPs. **(h)** ssODN-S1mplexes decrease the frequency of imprecise NHEJ editing events 5.5-fold compared to RNPs. **(i)** Frequency of precise insertions in *BFP* locus measured by deep sequencing. ssODN-S1mplexes had 1.8 times more insertions than standard sgRNA RNPs and 12 times more insertions than S1m-sgRNAs without complexed ssODNs due to lack of streptavidin. When sgRNAs were transfected with biotin-ssODNs there were nearly no precise insertions. **(j)** Frequency of precise insertions at *EMX1* locus. Addition of streptavidin to bind S1m-sgRNAs and biotin-ssODN resulted in a 51-fold increase in the frequency of precise editing reads.

See also Table B-8 for absolute levels of editing.

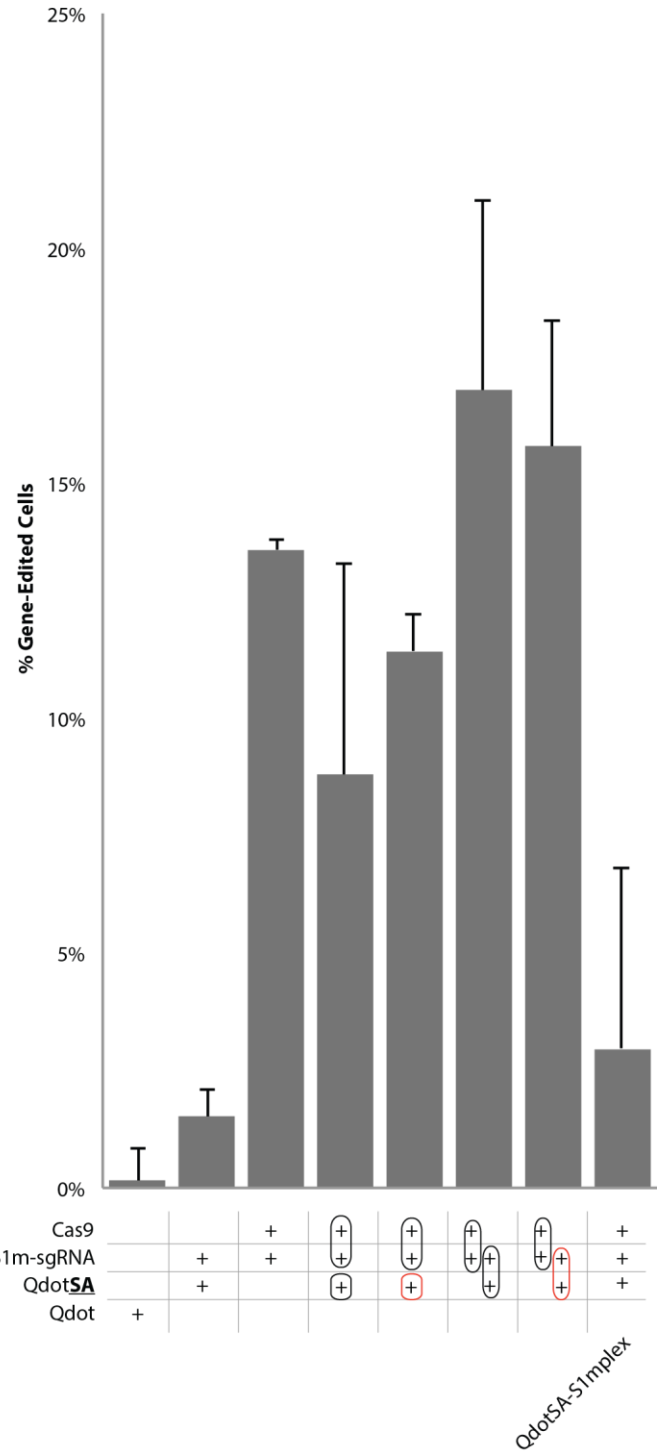

25%
20%
15%
10%
5%
0%
% Gene-Edited Cells
Cas9
S1m-sgRNA
QdotSA
Qdot
QdotSA-S1mplex

Figure B-5 | Gene-editing using various combinations of components with QdotSA.

Conjugation of S1mplexes to QdotSA significantly lowers gene editing efficiency with H2B-mCherry HEKs. Editing efficiency is lower even if QdotSA is transfected separately from the S1mplexes without complexation. Bounding boxes denote separate transfections and red bounding boxes denote separate transfection of reagents in the red box 5 hours later. Immediate application of the QdotSA can moderately interfere with the activity of the RNP, but these interference effects are abrogated if QdotSA is added 5 hours later. All RNP activity is abrogated by assembly with the QdotSA (last column) (n=3 technical replicates).

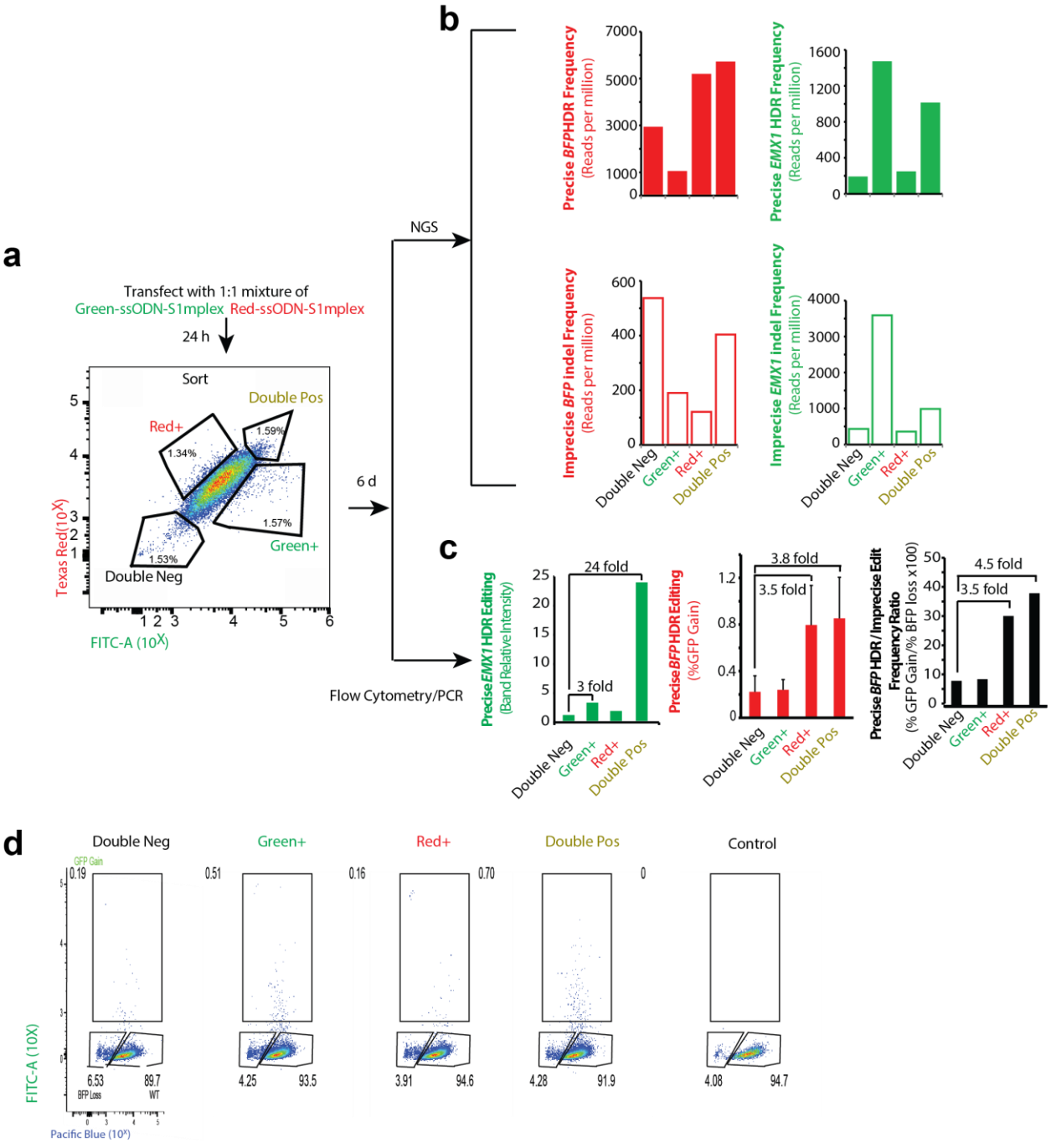

a
Transfect with 1:1 mixture of
Green-ssODN-S1mplex Red-ssODN-S1mplex
24 h
Sort
Double Pos
Red+
Green+
Double Neg
1.34%
1.59%
1.57%
1.53%
Texas Red(10X)
FITC-A (10X)
6 d
NGS
Flow Cytometry/PCR
b
Precise BFP HDR Frequency
(Reads per million)
Precise EMX1 HDR Frequency
(Reads per million)
Imprecise BFP indel Frequency
(Reads per million)
Imprecise EMX1 indel Frequency
(Reads per million)
Double Neg
Green+
Red+
Double Pos
c
Precise EMX1 HDR Editing
(Band Relative Intensity)
24 fold
3 fold
Precise BFP HDR Editing
(%GFP Gain)
3.8 fold
3.5 fold
Precise BFP HDR / Imprecise Edit
Frequency Ratio
(% GFP Gain/% BFP loss x100)
4.5 fold
3.5 fold
d
Double Neg
Green+
Red+
Double Pos
Control
GFP Gain
FITC-A (10X)
Pacific Blue (10X)
0.19
0.51
0.16
0.70
0
6.53
89.7
BFP Loss
WT
4.25
93.5
3.91
94.6
4.28
91.9
4.08
94.7

Figure B-6 | Multiplexed editing with fluorescent ssODN-S1mplexes.

(a) Flow cytometry plot of fluorescent S1mplexes 24 hours post transfection, HEKs were sorted into populations that were positive for either fluorophore, both or neither. Analysis was done 6 days post sorting. **(b)** Precise and imprecise editing frequency measured by deep sequencing for all four sorted populations. Precise editing was enriched in populations specific for an ssODN-S1mplex. Imprecise editing was not uniformly higher in sorted populations for a specific sgRNA. **(c)** Precise and imprecise editing measured by conventional methods for all four sorted populations. Populations specific for editing at a single locus as well as the double positive population were enriched for edits at that locus and not the other (n=3 biological replicates). **(d)** Representative flow cytometry plots related to Figure 3.6d and parts a-c above.

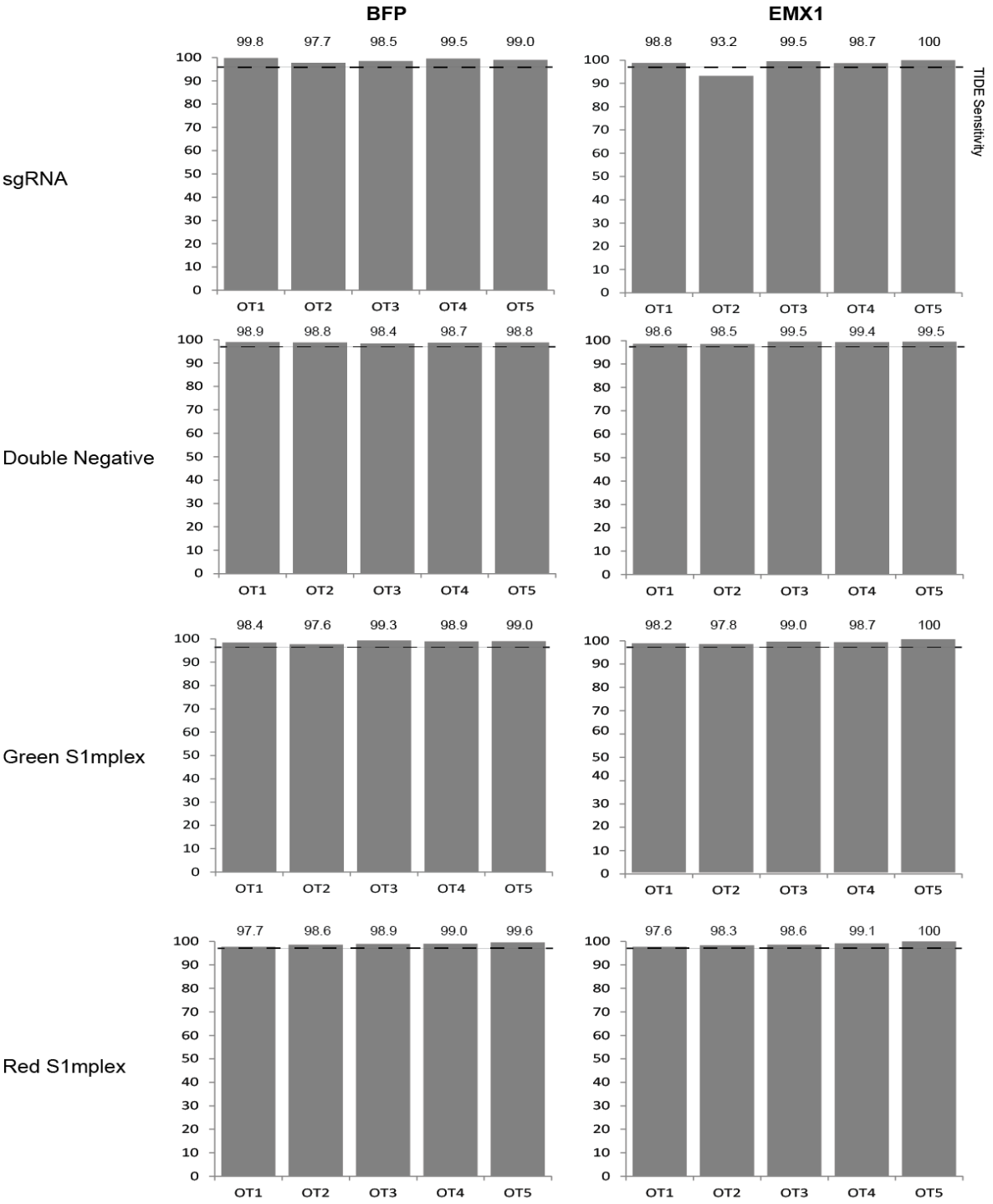

BFP
EMX1
TIDE Sensitivity
sgRNA
Double Negative
Green S1mplex
Red S1mplex
99.8 97.7 98.5 99.5 99.0
98.8 93.2 99.5 98.7 100
98.9 98.8 98.4 98.7 98.8
98.6 98.5 99.5 99.4 99.5
98.4 97.6 99.3 98.9 99.0
98.2 97.8 99.0 98.7 100
97.7 98.6 98.9 99.0 99.6
97.6 98.3 98.6 99.1 100
OT1 OT2 OT3 OT4 OT5

Figure B-7 | Off-target analysis of sorted S1mplex populations.

Off-target analysis using TIDE software at the top 5 predicted off-target sites within the human genome at the *BFP* and *EMX1* loci. Y axis indicates the percentage of cells with 0 mismatches from the parental sequence (perfect matches in sequencing reads). None of the sorted S1mplex populations showed off-target effects above the limit of detection. The unsorted sgRNA RNP population had a small proportion of cells that may have been edited at OT-2 of the *EMX1* off-target sites.

Text B-1

In parallel with the deep sequencing assays (**Table B-9)**, flow cytometry and PCR-based assays were used to assay for precise and imprecise gene editing with the S1mplexes. For the ssODN-S1mplexes, we found that ssODN-S1mplexes significantly decreased imprecise NHEJ gene editing as measured by loss of BFP (**Figure B-4e**), consistent with the results from the sgRNA variant editing experiments (**Figure B-3a**). At the same time, ssODN-S1mplexes produced robust levels of precise editing as indicated by the gain of GFP (**Figure B-4d**). The ratio of precisely-edited GFP cells to improperly-edited non-fluorescent cells increased 5-fold with ssODN-S1mplexes (**Figure B-4c**). Importantly, these increases were abrogated by omitting any of the critical components of the complex (either S1m aptamer, streptavidin or biotin) in the formulation. Compared to the approximately 2-fold decrease in NHEJ editing by utilizing a S1m-sgRNA (**Figure B-3a**), the 5-fold increase in the precise: imprecise ratio could not be attributed solely to the presences of a S1m aptamer.

For the fluorescent ssODN-S1mplexes, we also analyzed each sorted population using conventional flow cytometry and PCR-based methods. At the *BFP* locus each ssODN-S1mplex sorted population had similar levels of BFP-negative cells as the control (**Figure B-6d**). However, when we analyzed GFP fluorescence gain (indicating precise editing via HDR), cells that were positive for complexes targeted against the *BFP* locus (Red+ and the double positive fractions) had increased levels of precise editing in comparison to the other two sorted cell populations (**Figure B-6c**). It was observed that the Red+ population had 3.5-fold higher precise-to-imprecise editing ratio while the double positive had a 4.5-fold increase over the double negative population, respectively (**Figure B-6c**). As seen from deep sequencing data (**Figure B-6b**), these ratios are likely underrepresented in flow cytometry assays, due to the presence of BFP-negative cells in the starting population (4.08% in **Figure B-6d**, right). We next tested for the insertion of the 12 nt sequence into the *EMX1* locus by creating a PCR primer that

was specific to only the insert. Populations specific for *EMX1* editing (Green+ and double positive fractions) displayed strong bands at the expected size. The Green+ population had a 3-fold increase in intensity when compared background levels, while the double positive population increased HDR 24-fold (**Figure B-6c**, left).

Overall, the trend of increased precise editing relative to imprecise editing with the flow cytometry and PCR assays are consistent with the deep sequencing results in **Figures 3.3** and **3.6 and displayed in Table B-9**. Deep sequencing of the target locus, particularly when using ssODNs, allows for faithful determination of precise repair across the entire homology region, which is difficult to determine with the flow cytometry of reporter gene disruption (e.g., BFP loss/GFP gain in **Figure B-4** and **B-6** and with traffic light reporters[245]) and PCR-based assays involving insertion of unique sequences (e.g., unique primer as in **Figure B-6** or restriction site[195]). The use of large homology arms, such as in plasmid-based and virally-delivered donors used previously[173,245], that extend past the region of interest may contain also mismatches that remain undetected even after HDR. Further, while high targeted insertion rates have been reported with plasmid-based transfection methods[180,245], these reports focus on selected cells from the transfected cell population and do not quantify the levels of NHEJ of the entire cell population. For RNP delivery with ssODNs to human cells[173,174], deep sequencing results have at best generated approximately one precise edit to each imprecise edit detected per million reads of the full transfected population.

Table B-1 | Primers used to create sgRNA and S1m-sgRNAvariants.

S1m Construct Name	Sequence (5′ to 3′)
S1m_StemLoop1_F	GTTTAAGAGCTATGCTGCGAATACGAGATGCGGCCGCCGACCAGAATCATGCAAGTGCGTAAGATAGTCGCGGGTCGGCGGCCGCATCTCGTATTC
S1m_StemLoop1_R	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAAACTTGCTATGCTGCGAATACGAGCCGCCGACCCG
S1m_StemLoop2_F	GTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTCGAATACGAGATGCGGCCGCCGACCAGA
S1m_StemLoop2_R	AAAAAAAGCACCGACTCGGTGCCACTTTTTCCGAATACGAGatgcggCCGCCGACCCGCGACTATCTTACGCACTTGCATGATTCTGGTCGGCGGC
S1m_StemLoop3_F	GTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCCGAA
S1m_StemLoop3_R	AAAAAAACGAATACGAGATGCGGCCGCCGACCCGCGACTATCTTACGCACTTGCATGATTCTGGTCGGCGGCCGCATCTCGTATTCGGCACCGACT
S1m1 Forward	TTAATACGACTCACTATAGGNNNNNNNNNNNNNNNNNNNGTTTAAGAGCTATGCTGCGA
S1m2_3 Forward	TTAATACGACTCACTATAGGNNNNNNNNNNNNNNNNNNNGTTTAAGAGCTATGCTGGAA
RNATracR	AAAAGCACCGACTCGGTGCC
S1m3_Reverse	AAAACGAATACGAGCCGCCG

Table B-2 | Protospacer and respective PAMs used for genomic targeting.

sgRNA Name	Sequence (5′ to 3′)	PAM
BFP (BFP → GFP)	GCTGAAGCACTGCACGCCAT	GGG
EMX1 (EMX1_21)	GTCACCTCCAATGACTAGGG	TGG
mCherry (mCherry_15)	GGAGCCGTACATGAACTGAG	GGG
GAA ΔT	CTCGTTGTCCAGGTAGGCCC	GGG

Table B-3 | Forward and reverse primers for genomic loci.

Genomic Primer	Forward (5′ to 3′)	Reverse (5′ to 3′)
EMX1	CCATCCCCTTCTGTGAATGT	GGAGATTGGAGACACGGAGA
EMX1 Symmetric	TCCACCTTGGCTTGGCTTTG	CCCTCCACCAGTACCCCAC
mCherry Interior	AAGGGCGAGGAGGATAACATGG	TTGTACAGCTCGTCCATGCCG
EMX1 Insertion	CCAATGACAAGCTTGCTAGC	GGAGATTGGAGACACGGAGA
GAA ΔT	AGCTGCTCATTGACCTCCAG	CAATCCACATGCCGTCGAAG

Table B-4 | ssODNs used to direct HDR after DSB formation.

ssODN Donor	Sequence (5' to 3')
BFP → GFP NT	TCATGTGGTCGGGGTAGCGGCTGAAGCACTGCACGCCATGGGTCAGGGTGGTCACGAGGGTG GGCCAGGGCACCGGCAGCTTGCCGGTGGTGCAGATGAA
BFP → GFP 5PCBio NT	5Biotin/TCATGTGGTCGGGGTAGCGGCTGAAGCACTGCACGCCATGGGTCAGGGTGGTCA CGAGGGTGGGCCAGGGCACCGGCAGCTTGCCGGTGGTGCAGATGAA
EMX1 NT	AAGCAGCACTCTGCCCTCGTGGGTTTGTGGTTGCCCACCGCTAGCAAGCTTGTCATTGGAGG TGACATCGATGTCCTCCCCATTGGCCTG
EMX1 5PCBio NT	5Biotin/AAGCAGCACTCTGCCCTCGTGGGTTTGTGGTTGCCCACCGCTAGCAAGCTTGTC ATTGGAGGTGACATCGATGTCCTCCCCATTGGCCTG
3' BFP → GFP -30+67 Antisense	GGCATGGCGGACTTGAAGAAGTCGTGCTGCTTCATGTGGTCGGGGTAGCGGCTGAAGCACTG CACGCCGTACGTCAGGGTGGTCACGAGGGTGGGCC/3Biotin
5' BFP → GFP -30+67 Antisense	5Biotin/GGCATGGCGGACTTGAAGAAGTCGTGCTGCTTCATGTGGTCGGGGTAGCGGCTG AAGCACTGCACGCCGTACGTCAGGGTGGTCACGAGGGTGGGCC
3' BFP → GFP -30+67 Sense	GGCCCACCCTCGTGACCACCCTGACGTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCAC ATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCC/3Biotin
5' BFP → GFP -30+67 Sense	5Biotin/GGCCCACCCTCGTGACCACCCTGACGTACGGCGTGCAGTGCTTCAGCCGCTACC CCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCC
3' BFP → GFP -67+30 Antisense	GGTCGGGGTAGCGGCTGAAGCACTGCACGCCGTACGTCAGGGTGGTCACGAGGGTGGGCCAG GGCACCGGCAGCTTGCCGGTGGTGCAGATGAACTT/3Biotin
5' BFP → GFP -67+30 Antisense	5Biotin/GGTCGGGGTAGCGGCTGAAGCACTGCACGCCGTACGTCAGGGTGGTCACGAGGG TGGGCCAGGGCACCGGCAGCTTGCCGGTGGTGCAGATGAACTT
3' BFP → GFP -67+30 Sense	AAGTTCATCTGCACCACCGGCAAGCTGCCGGTGCCCTGGCCCACCCTCGTGACCACCCTGAC GTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACC/3Biotin
5' BFP → GFP -67+30 Sense	5Biotin/AAGTTCATCTGCACCACCGGCAAGCTGCCGGTGCCCTGGCCCACCCTCGTGACC ACCCTGACGTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACC
3' GAA Insert T -34+66 PAM	CTCCCCACTGCAGCCTCTCGTTGTCCAGGTATGGCCCGGGTCCACTGCCTTCCCCGACTTCA CCAACCCCACAGCCCTGGCCTGGTGGGAGGACATGGTG/3Biotin
5' GAA Insert T -34+66 PAM	5Biotin/CTCCCCACTGCAGCCTCTCGTTGTCCAGGTATGGCCCGGGTCCACTGCCTTCCC CGACTTCACCAACCCCACAGCCCTGGCCTGGTGGGAGGACATGGTG

3’ GAA Insert T -34+66 NonPAM	CACCATGTCCTCCCACCAGGCCAGGGCTGTGGGGTTGGTGAAGTCGGGGAAGGCAGTGGACC CGGGCCATACCTGGACAACGAGAGGCTGCAGTGGGGAG/3Biotin
5’ GAA Insert T -34+66 NonPAM	5Biotin/CACCATGTCCTCCCACCAGGCCAGGGCTGTGGGGTTGGTGAAGTCGGGGAAGGC AGTGGACCCGGGCCATACCTGGACAACGAGAGGCTGCAGTGGGGAG

Table B-5 | Off-target sequences and corresponding genomic locus for each sgRNA used. Mismatches from protospacer are labelled in red.

sgRNA Target Sequence	Off-Target Sequence		PAM	Locus
BFP→ GFP GCTGAAGCACTGCACGCCAT	OT1	GCAGAAGCACTGCAAGCCAT	CAG	chr17:+39786906
	OT2	TCTGAAGTGCTGCACGCCAT	CAG	chr2:-238397265
	OT3	GTGGAAGCACTGCAAGCCAT	TGG	chr7:-11228464
	OT4	GGTGGAGCAGGGCACGCCAT	CAG	chr9:+109114765
	OT5	GAAGAAGCACTGCACCCCAT	CAG	chr13:-75660548
EMX1 GTCACCTCCAATGACTAGGG	OT1	AGGACCACCAATGACTAGGG	CAG	chr3:-64303990
	OT2	ACCACCTGTAATGACTAGGG	TAG	chr4:-149749778
	OT3	GGAGCCTCCAGTGACTAGGG	GAG	chr17:-38423030
	OT4	GTGAACTACAGTGACTAGGG	TGG	chr8:+112210096
	OT5	CTGGCCTCCAAAGACTAGGG	GAG	chr15:-75011931

Table B-6 | Forward and reverse primers used to amplify off-target genomic loci.

Off-Target Primer	Forward (5' to 3')	Reverse (5' to 3')
BFP OT1	TTTCCTAGCAAGCAGACTCAGA	AGCTGTCCTTTGTCCCATTGA
BFP OT2	TCTCCATGCCCTCCTTTCCAT	GGATGTAGTCCATGATCTTCCCC
BFP OT3	TCCCAGAATGTGAAAGTGGAGG	CTGTGGGCTTTCCTCAGCTC
BFP OT4	GCTGACTAACGTCCACTGCT	TGGACCTATGTTTTTCTTCGTCAC
BFP OT5	AAAGTCTGTGGCCTTGTGAGA	AACCCTACCCCCTACCTGAA
EMX1 OT1	TTCCCCAGGTAGTTGCTGTTC	TCTGCACATGTCCCAACTGTC
EMX1 OT2	ATCCGTACCTAACCATGACCC	GCACAGATCTTGGTGGCTTT
EMX1 OT3	GGCTGGGTTTCCCAAACGTA	CAAACTGCTGTGTTGGGTGG
EMX1 OT4	ACTTGGAAGGGTCCACACAA	CCTTGAATAGAGCATTTTTCCCCA
EMX1 OT5	TCCTACCCTTGGATGGGGTT	GGGCTACACGGTCCCTAAAG

Table B-7 | Next generation sequencing primers including adapters and barcode around DSB site.

NGS Primer	Sequence (5' to 3')
BFP→ GFP F	TCATCTGCACCACCGGCAAG
BFP→ GFP F1	AATGATACGGCGACCACCGAGATCTACACTATAGCCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTCAT CTGCACCACCGGCAAG
BFP→ GFP F2	AATGATACGGCGACCACCGAGATCTACACATAGAGGCACACTCTTTCCCTACACGACGCTCTTCCGATCTTCAT CTGCACCACCGGCAAG
BFP→ GFP F3	AATGATACGGCGACCACCGAGATCTACACCCTATCCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTCAT CTGCACCACCGGCAAG
BFP→ GFP R	GAAGTCGTGCTGCTTCATGTGG
BFP→ GFP R1	CAAGCAGAAGACGGCATACGAGATCTGAAGCTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCGAAGTCGTG CTGCTTCATGTGG
BFP→ GFP R2	CAAGCAGAAGACGGCATACGAGATTAATGCGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCGAAGTCGTG CTGCTTCATGTGG
EMX1 F	CTCCCATCACATCAACCGGTGG
EMX1 F1	AATGATACGGCGACCACCGAGATCTACACTATAGCCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCTCC CATCACATCAACCGGTGG
EMX1 F2	AATGATACGGCGACCACCGAGATCTACACATAGAGGCACACTCTTTCCCTACACGACGCTCTTCCGATCTCTCC CATCACATCAACCGGTGG
EMX1 F3	AATGATACGGCGACCACCGAGATCTACACCCTATCCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCTCC CATCACATCAACCGGTGG
EMX1 R	CACTCTGCCCTCGTGGGTTT
EMX1 R1	CAAGCAGAAGACGGCATACGAGATCGCTCATTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCCACTCTGCC CTCGTGGGTTT
EMX1 R2	CAAGCAGAAGACGGCATACGAGATGAGATTCCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCCACTCTGCC CTCGTGGGTTT
GAA F	AATGATACGGCGACCACCGAGATCTACACTATAGCCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAGTG GGGCTTCCATGCAG

GAA R	CAAGCAGAAGACGGCATACGAGATGAGATTCCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCGGTTGGTGA AGTCGGGGAAG

Table B-8 | Absolute levels of editing in this study and prior studies to date. Prior studies were selected based on a comparable number of base pair changes within the ssODN template and use of deep or Sanger sequencing of the edited locus. Deep sequencing assays were performed shortly after RNP delivery without significant selection, while Sanger sequencing occurred after clonal isolation and expansion without drug selection.

Experiment	**Cell type**	**Gene**	**# of base pair changes**	**% of HDR** (absolute percent of deep sequencing reads)	**% of NHEJ** (absolute percent of deep sequencing reads)	**Reference**
Figure 3.3, averaged	HEK	*BFP, EMX1*	3, 12	0.53%	0.16 %	This study
Figure 3.3, averaged	hPSC	*BFP, EMX1*	18	0.34%	1.89%	This study
Figure 3.4, averaged	hPSC	*BFP, GAA*	2-3	2.0%	0.58%	This study
Figure 3.6, averaged	HEK	*BFP and EMX1*	3 and 12	0.34%	0.14%	This study
Averaged across experiments involving HDR as assayed by deep sequencing (Figs. 3.3, 3.4, 3.6)	HEK, hPSC	*BFP, GAA, EMX1*	2 to 18	1.6%	0.67%	This study
Small base pair changes	HEK, hPSC	*BFP, GAA*	<5	1.8%	0.75%	This study
HDR as assayed by Sanger sequencing of enriched clones (Fig. 3.6b)	HEK	*BFP*	3	24%	24%	This study
Figure 1	hPSC	*SCN8A, SCN1B, CHD2, PCDH19, HPRT1,*	N/A	N/A	1-17.5%	A. M. Tidball et. al. *Stem Cell Reports*, 2017

		SMC1A				
Extended data Figure 1f	hPSC	*MYBPC3*	4	1.5%	3.8%	H. Ma et. al. *Nature,* 548, 413–419, 2017
Extended data Table 1	hPSC	*APP, PSEN1*	2	0.3-6.0%	Not reported	Paquet, D. *et al. Nature* **533,** 125–129 (2016).
Figure 3a	hPSC	*CCR5*	2	0.2-1.6%	1-1.8%	Yang, L. *et al. Nucl. Acids Res.* (2013). doi:10.1093/nar/gkt555
Table S1	hPSC	*AKT2*	1	6.4%*	Not reported	Ding, Q. *et al. Cell Stem Cell* **12,** 393–394 (2013).
Figure 2f	hPSC	*SOD1*	1	0.35-3.1%	27-32%	Yu, C. *et al. Cell Stem Cell* **16,** 142–147 (2015).
Figure 5	HEK293T	*RBM20, ATP7B*	1	0.09-0.6%	3-9%	Miyaoka, Y. *et al. Scientific Reports* **6,** srep23549 (2016).
Figure 5	hPSC	*RBM20*	1	0.005-0.1%	1-10%	Miyaoka, Y. *et al. Scientific Reports* **6,** srep23549 (2016).
Figure 2c, Supplementary Table 12 before drug selection	HCT116	*ERCC3*	2	0.3-5.2%	7.1-12.6%	Smurnyy, Y. et al. Nat Chem Biol 10, 623–625 (2014).
Figure 2a, unexpanded cells	HSPC	*HBB*	1	6-11%	10-30%	DeWitt, M. A. *et al. Science Translational Medicine* **8,** 360ra134-(2016).
Figure 3.4c	hPSC	*EGFP*	2	2.5-4.5%**	30-60%**	Howden, S. E. *et al. Stem Cell Reports* doi:10.1016/j.stemcr.2016.07.001
Table B-2	HSPC	*CD45*	2	24%	16%***	Gundry, M. C. *et al. Cell Rep* 17, 1453–

						1461, 2016
Figure 3c	HEK	*BFP*	3	9-63%	25%-90%	Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. *Nat Biotech* 34, 339–344, 2016
Figure 6, 7	HEK	*EMX1*	4	43-66%	2-31%****	Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. *Nat Biotech* 34, 339–344, 2016
Figure 3	HEK	*EMX1*	12	0-6%*****	Not reported	Lin, S. et al. eLife 2014;3:e04766
Figure 4	hPSC	*EMX1*	12	0-1.6%	3-44%	Lin, S. *et al.* eLife 2014;3:e04766

*enriched by flow cytometry for Cas9-GFP expressing cells; 10 of 94 clones with monoallelic mutation introduced and 1 of 94 clones with biallelic HDR editing, resulting in 12 out of 188 alleles = 6.4%

** with plasmid HDR donors; assayed by flow cytometry

***Only the most common indels are reported. Reported frequencies for all alleles add up to 85% in this table.

**** ssODN most similar to work involved in this paper induced 2-9% HDR measured by Restriction Fragment Length Polymorphism (RFLP) assay.

***** ssODN similar to work in the paper was below limit of detection using Restriction Fragment Length Polymorphism (RFLP) assay.

Table B-9 | Absolute levels of editing for each experimental replicate assayed by deep sequencing. Analysis of HDR and NHEJ rates following deep sequencing in each experimental condition and replicate presented in this work.

Experiment	Condition	Cell Line	Locus	Number of base pair changes	HDR1	HDR2	HDR3	NHEJ1	NHEJ2	NHEJ3
Figure 3.3,										
also Figure	ssODN-S1m-sgRNA-1									
B-4	S1mplex	hPSC	BFP	3	0.49%			0.43%		
	ssODN-S1m- sgRNA-2									
	S1mplex	hPSC	BFP	3	0.76%			0.68%		
	ssODN-S1m- sgRNA-3									
	S1mplex	hPSC	BFP	3	0.30%			0.35%		
	Standard sgRNA									
	ssODN	hPSC	BFP	3	1.3%			3.9%		
	Standard sgRNA									
	ssODN	HEK	BFP	3	0.52%			2.2%		
	ssODN-S1mplex	HEK	BFP	3	0.99%			0.23%		
	Standard sgRNA									
	ssODN	HEK	EMX1	12	1.0%			3.9%		
	ssODN-S1mplex	HEK	EMX1	12	0.07%			0.09%		
	Standard sgRNA									
	ssODN	hPSC	BFP	18	0.045%			11%		
	sgRNA Bio-ssODN	hPSC	BFP	18	0.002%			9.8%		
	S1m Bio-ssODN (-SA)	hPSC	BFP	18	0.009%			1.3%		
	ssODN-S1mplex	hPSC	BFP	18	0.079%			2.0%		
	S1m Bio-ssODN (-SA)	hPSC	EMX1	18	0.001%			1.8%		
	ssODN-S1mplex	hPSC	EMX1	18	0.052%			6.0%		
Figure 3.4	5' -67 NonPAM	hPSC	BFP	3	1.0%	1.1%	3.6%	0.51%	0.22%	0.94%
	3'-67 NonPAM	hPSC	BFP	3	1.0%	0.97%	5.9%	0.18%	0.15%	1.7%
	5' -30 NonPAM	hPSC	BFP	3	1.1%	1.2%	3.1%	0.11%	0.30%	1.4%
	3' -30 NonPAM	hPSC	BFP	3	1.0%	1.3%	3.4%	0.31%	0.44%	1.5%
	5' -67 PAM	hPSC	BFP	3	1.3%	1.1%	2.8%	0.37%	0.13%	1.4%
	3' -67 PAM	hPSC	BFP	3	1.0%	1.2%	5.0%	0.34%	0.34%	2.0%
	5' -30 PAM	hPSC	BFP	3	1.1%	1.1%	6.6%	0.27%	0.25%	1.4%
	3' -30 PAM	hPSC	BFP	3	1.0%	1.0%	3.4%	0.20%	0.24%	1.4%
	RNP Control	hPSC	BFP	3	3.1%	1.1%	1.4%	2.9%	2.4%	4.6%
	RNP Control	hPSC	GAA	2	1.7%	2.7%		2.2%	3.7%	

Experiment	Condition	Cell Line	Locus	Number of base pair changes	HDR1	HDR2	HDR3	NHEJ1	NHEJ2	NHEJ3
	3' -34 PAM	hPSC	GAA	2	1.2%	3.8%		0.18%	0.46%	
	5' -34 PAM	hPSC	GAA	2	1.4%	2.6%		0.40%	0.37%	
	3' -34 NonPAM	hPSC	GAA	2	1.1%	1.0%		0.17%	0.33%	
	5' -34 NonPAM	hPSC	GAA	2	1.1%	1.1%		0.18%	0.27%	
Figure 3.6d, also Figure B-6b *Multiplexed editing for both EMX and BFP*	Double Negative	HEK	EMX1	12	0.02%			0.04%		
	Green+	HEK	EMX1	12	0.15%			0.36%		
	Red+	HEK	EMX1	12	0.02%			0.04%		
	Double Positive	HEK	EMX1	12	0.10%			0.13%		
	Double Negative	HEK	BFP	3	0.29%			0.05%		
	Green+	HEK	BFP	3	0.11%			0.02%		
	Red+	HEK	BFP	3	0.52%			0.01%		
	Double Positive	HEK	BFP	3	0.57%			0.04%		

All deep sequencing data and Sanger sequencing traces from PCR amplicons are available at Bioproject PRJNA381066.

Appendix C: Supplemental Tables and Figures for Chapter 4

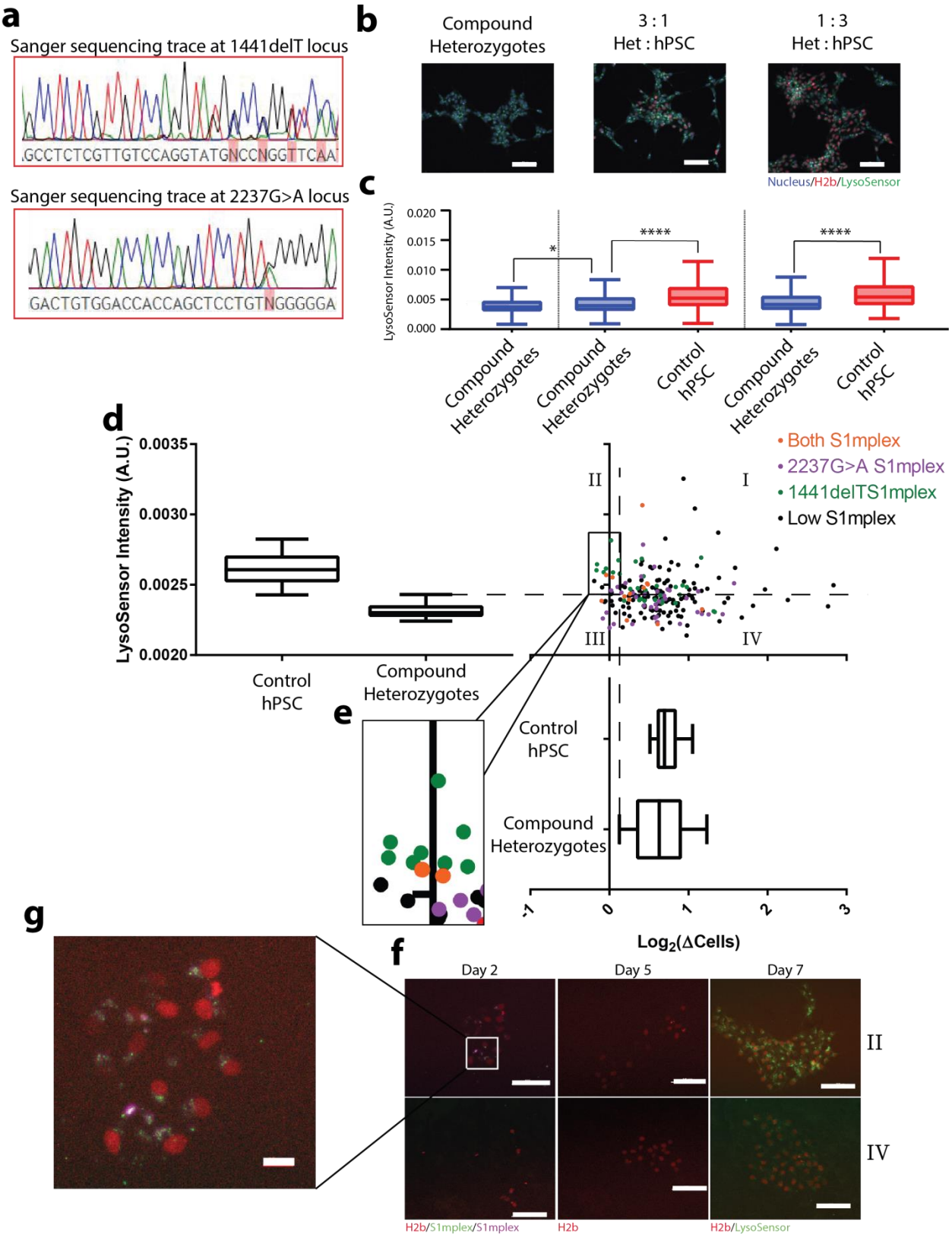

a
Sanger sequencing trace at 1441delT locus
GCCTCTCGTTGTCCAGGTATGNCCNGGTTCAA
Sanger sequencing trace at 2237G>A locus
GACTGTGGACCACCAGCTCCTGTNGGGGGA
b
Compound Heterozygotes
3 : 1 Het : hPSC
1 : 3 Het : hPSC
Nucleus/H2b/LysoSensor
c
LysoSensor Intensity (A.U.)
0.020
0.015
0.010
0.005
0.000
*
****
****
Compound Heterozygotes
Compound Heterozygotes
Control hPSC
Compound Heterozygotes
Control hPSC
d
LysoSensor Intensity (A.U.)
0.0035
0.0030
0.0025
0.0020
Control hPSC
Compound Heterozygotes
II
I
III
IV
• Both S1mplex
• 2237G>A S1mplex
• 1441delTS1mplex
• Low S1mplex
e
Control hPSC
Compound Heterozygotes
-1
0
1
2
3
Log2(ΔCells)
g
f
Day 2
Day 5
Day 7
II
IV
H2b/S1mplex/S1mplex
H2b
H2b/LysoSensor

Figure C-1 | Use of ArrayEdit for isolation of biallelic edited iPSCs.

(a) Sequence traces of compound heterozygote iPSC line. 1441delT mutation causes a breakdown of sequence trace. Comparatively, a single point mutation demonstrates heterozygosity 2237G>A locus. (**b)** Representative images of hPSCs following LysoSensor Green staining. Compound heterozygotes were either cultured alone, or at the indicated ratio with hPS-H2b labelled cells to identify each cell line (*scale bar: 100μm*). **(c)** Quantification of LysoSensor intensity in each co-culture condition. LysoSensor intensity was measured on a per-cell basis using confocal microscopy. hPS cells had a significantly higher intensity that compound heterozygote cells in all conditions. Compound heterozygote cells co-cultured with hPSCs also had an increased lysosomal intensity when compared to those cultured alone. (**p*<0.05, *****p*<0.001). **(d)** *Left:* LysoSensor quantification per μFeature of two mock transfections after 7 days of growth. Dye in hPSCs was significantly more intense than compound heterozygote iPSCs on ArrayEdit. *Bottom:* Growth rate per μFeature following mock transfection. Growth rates were calculated by measuring the per-day change in the number of cells of the μFeature. Features were graphed as an average of these per day changes. *Top Right:* Individual μFeatures graphed on the axes described above. Each μFeature is color coded by the presence of S1mplex within the nucleus on Day 1. Dashed lines indicate regions of interest. **(e)** Magnification of quadrant II from part D. μFeatures in this region were selected for genomic analysis to isolate edited clones. (**f)** Representative μFeatures from ArrayEdit representing populations from quadrants 2 and 4 from part D. S1mplexes can be seen in the nucleus as late as day 2 and form the basis of which feature to select. (*scale bar: 100μm*) (**g)** Magnification of Day 2 colony. S1mplexes can be seen throughout the image. (*scale bar: 25μm)*

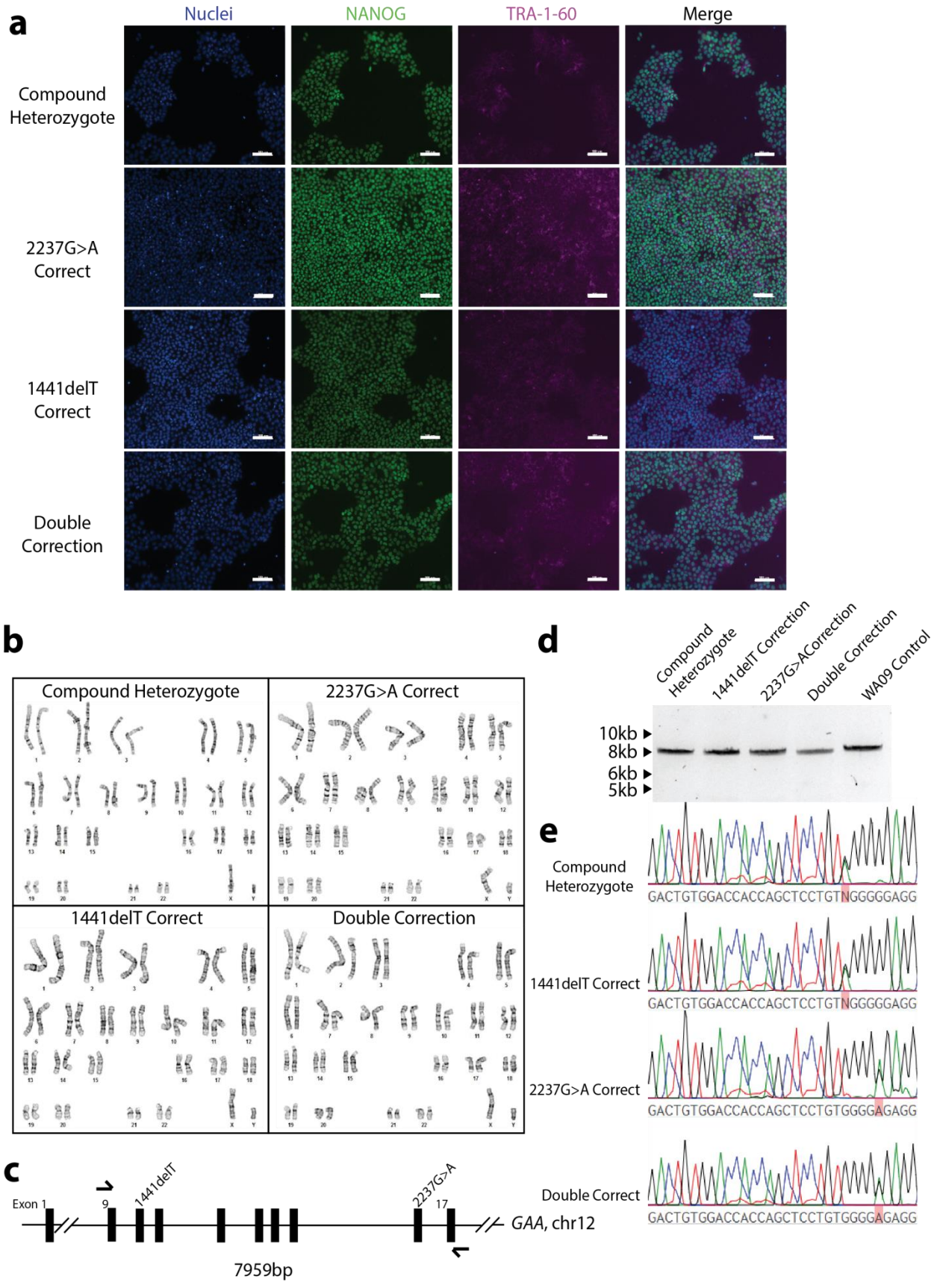

a
Nuclei
NANOG
TRA-1-60
Merge
Compound Heterozygote
2237G>A Correct
1441delT Correct
Double Correction
b
Compound Heterozygote
2237G>A Correct
1441delT Correct
Double Correction
c
Exon 1
9
1441delT
17
2237G>A
GAA, chr12
7959bp
d
Compound Heterozygote
1441delT Correction
2237G>ACorrection
Double Correction
WA09 Control
10kb
8kb
6kb
5kb
e
Compound Heterozygote
GACTGTGGACCACCAGCTCCTGTNGGGGGAGG
1441delT Correct
GACTGTGGACCACCAGCTCCTGTNGGGGGAGG
2237G>A Correct
GACTGTGGACCACCAGCTCCTGTGGGGAGAGG
Double Correct
GACTGTGGACCACCAGCTCCTGTGGGGAGAGG

Figure C-2 | Genome editing with S1mplexes does not cause genomic instability.

(a) Immunocytochemistry of pluripotency markers in gene edited lines. All lines were positive for pluripotency markers NANOG and TRA-1-60 (*scale bar: 100μm*). (**b)** Karyotypes of all isolated gene-edited lines as well as compound heterozygotes. No abnormalities were detected at a band resolution of 500. (**c)** PCR schematic for long PCR covering both Cas9 cut sites. Expected product is 7959bp in length. **(d)** Gel analysis of long PCR described in (c) in each isolated cell line. No significant deviances from expected length were detected and no other significant bands were observed. **(e)** Sanger sequence traces of long PCR shown in (d). SNPs were observed showing that PCR products were a result of amplification from both alleles within the cells.

2237 OT-1		
	Compound Heterozygote	CCTCCCTTCCTAGACCACCAGCTCCTGCAGGAGGGCTTGG
	1441delT Correction	CCTCCCTTCCTAGACCACCAGCTCCTGCAGGAGGGCTTGG
	2237G>A Correction	CCTCCCTTCCTAGACCACCAGCTCCTGCAGGAGGGCTTGG
	Double Correction	CCTCCCTTCCTAGACCACCAGCTCCTGCAGGAGGGCTTGG
2237 OT-2		
	Compound Heterozygote	GCCCCTGCCTCTACAGGAGCAGGTGGTGAGGATGGCTCCG
	1441delT Correction	GCCCCTGCCTCTACAGGAGCAGGTGGTGAGGATGGCTCCG
	2237G>A Correction	GCCCCTGCCTCTACAGGAGCAGGTGGTGAGGATGGCTCCG
	Double Correction	GCCCCTGCCTCTACAGGAGCAGGTGGTGAGGATGGCTCCG
2237 OT-3		
	Compound Heterozygote	CACGCTGCCACGACAGGAGCTGCTGGTCAACTCCTCCTCT
	1441delT Correction	CACGCTGCCACGACAGGAGCTGCTGGTCAACTCCTCCTCT
	2237G>A Correction	CACGCTGCCACGACAGGAGCTGCTGGTCAACTCCTCCTCT
	Double Correction	CACGCTGCCACGACAGGAGCTGCTGGTCAACTCCTCCTCT
1441 OT-1		
	Compound Heterozygote	CCCCGTATCCCTGGTTGTCCAGGTGGGCCCTGGGAGAACA
	1441delT Correction	CCCCGTATCCCTGGTTGTCCAGGTGGGCCCTGGGAGAACA
	2237G>A Correction	CCCCGTATCCCTGGTTGTCCAGGTGGGCCCTGGGAGAACA
	Double Correction	CCCCGTATCCCTGGTTGTCCAGGTGGGCCCTGGGAGAACA
1441 OT-2		
	Compound Heterozygote	CAGCTGCCGTCTCGATGGCCAGGTAGGCCTGGGCAAGGAC
	1441delT Correction	CAGCTGCCGTCTCGATGGCCAGGTAGGCCTGGGCAAGGAC
	2237G>A Correction	CAGCTGCCGTCTCGATGGCCAGGTAGGCCTGGGCAAGGAC
	Double Correction	CAGCTGCCGTCTCGATGGCCAGGTAGGCCTGGGCAAGGAC
1441 OT-3		
	Compound Heterozygote	AATTAGGCTAGGGCCTACCTGGTCAATAATGAAATAATTG
	1441delT Correction	AATTAGGCTAGGGCCTACCTGGTCAATAATGAAATAATTG
	2237G>A Correction	AATTAGGCTAGGGCCTACCTGGTCAATAATGAAATAATTG
	Double Correction	AATTAGGCTAGGGCCTACCTGGTCAATAATGAAATAATTG

Figure C-3 | Off-target analysis of S1mplex corrected lines.

Off-target analysis of top 3 off-target sites for both sgRNAs. Sanger sequencing was unable to detect any deviations at off-target sites in any of the corrected lines in comparison to compound heterozygotes.

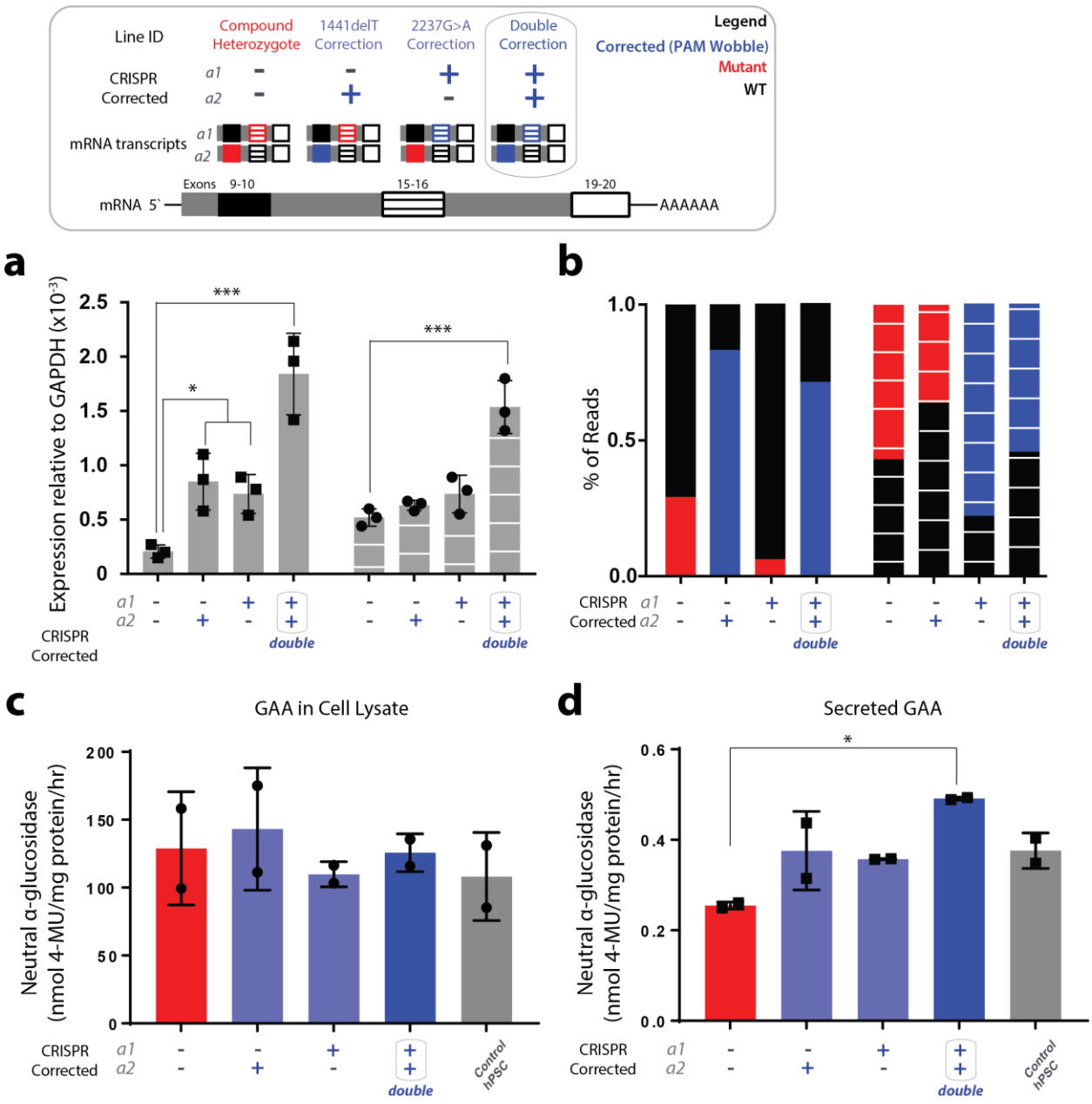

Line ID
Compound Heterozygote
1441delT Correction
2237G>A Correction
Double Correction
Legend
Corrected (PAM Wobble)
Mutant
WT
CRISPR Corrected
a1
a2
mRNA transcripts
Exons 9-10
15-16
19-20
mRNA 5`
AAAAAA
a
Expression relative to GAPDH (x10⁻³)
CRISPR Corrected
double
b
% of Reads
c
GAA in Cell Lysate
Neutral α-glucosidase (nmol 4-MU/mg protein/hr)
Control hPSC
d
Secreted GAA

Figure C-4 | Restoration of wildtype genotype increases mRNA expression from corrected allele.

(a) qPCR analysis around both mutated alleles. Expression was increased in heterozygous lines at the 1447delT locus (solid) by 4-fold and in homozygous corrections by 10-fold over compound heterozygotes. In comparison expression at the 2237G>A locus (hashed) was not significantly different between compound heterozygote and heterozygous lines. Homozygous correction lines saw a 3-fold increase in expression. (*p<0.05, ***p<0.005, n=3 technical replicates). **(b)** Next generation sequencing from RT-PCR around both diseased loci in isolated cell lines. Reads were classified as WT, mutant, or corrected (PAM wobble) and mapped to either allele 1 (top bar) or allele 2 (bottom bar). When neither allele was corrected, both alleles were expressed at approximately the same rate. However, when either mutation was corrected the corresponding allele was expressed at a higher rate than the one that still possessed a mutation. When both alleles were corrected the fraction of reads making up the population was evenly distributed. Observations at individual alleles were consistent across both assayed loci. **(c)** GAA activity in cell lysate as measured by 4-MUG cleavage in neutral conditions. All lines had equal activity compared to compound heterozygotes. (n=2 technical replicates). **(d)** GAA activity in culture media as measured by 4-MUG cleavage in neutral conditions. Double mutation cells had a 2-fold increase in activity over compound heterozygotes (*p<0.05, n=2 technical replicates).

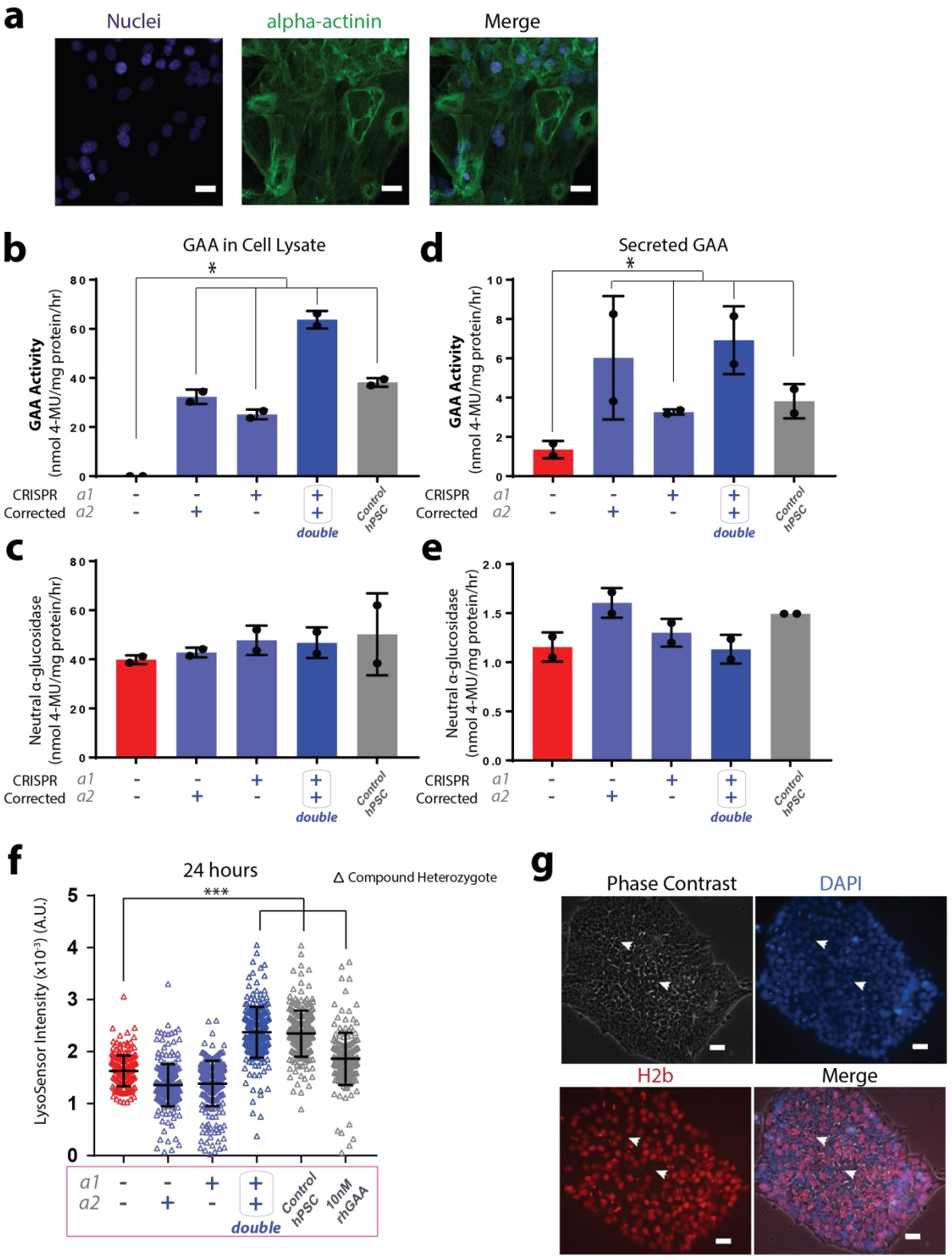

a
Nuclei
alpha-actinin
Merge
b
GAA in Cell Lysate
GAA Activity
(nmol 4-MU/mg protein/hr)
CRISPR a1
Corrected a2
double
Control hPSC
d
Secreted GAA
c
Neutral α-glucosidase
e
f
24 hours
Δ Compound Heterozygote
LysoSensor Intensity (x10³) (A.U.)
10nM rhGAA
g
Phase Contrast
DAPI
H2b
Merge

Figure C-5 | Phenotypic recovery of GAA activity in differentiated cardiomyocytes.

(a) Immunocytochemistry for sarcomeric α-actinin in differentiated compound heterozygotes. Striation pattern indicative of cardiomyocyte differentiation can be observed (*scale bar: 25μm*). **(b)** GAA activity in cell lysate of cardiomyocytes as measured by 4-MUG cleavage in acidic conditions. Compound heterozygotes were unable to cleave this substrate, showing there was little to no active protein. All corrected lines had significantly higher activity than compound heterozygotes but were indistinguishable from each other. Recovery of active protein is preserved through directed differentiation (n=2 technical replicates). **(c)** GAA activity in cell lysate as measured by 4-MUG cleavage in neutral conditions. All lines had equal activity compared to compound heterozygotes. (n=2 technical replicates). **(d)** GAA activity in spent media as measured by 4-MUG cleavage in acidic conditions. Compound heterozygotes were unable to cleave this substrate, showing there was little to no active secreted protein. All corrected lines had significantly higher activity than compound heterozygotes. Double correction cells secreted significantly more GAA than any other line, including control hPSCs. Increased secreted GAA may lead to a distal recovery effect in differentiated cell types (n=2 technical replicates). **(e)** GAA activity in spent media as measured by 4-MUG cleavage in neutral conditions. All lines had equal activity compared to compound heterozygotes. (n=2 technical replicates). **(f)** Quantification of LysoSensor intensity in cross-corrected lines 24 hours post media exchange. Each triangle represents a single compound heterozygote identified using CellProfiler. After 24 hours, only double corrected and control hPSC media and rhGAA treatment had a significant improvement in glycogen clearance over compound heterozygotes (***p<0.005). **(g)** Silencing of active transgenes at *AAVS1* locus. A CAGGS-H2b-mCherry reporter was integrated into the safe-harbor locus *AAVS1* promoting constitutive expression of this fluorophore. After 30 passages this expression was silenced in a significant number of hPSCs (arrowheads) (*scale bar: 25μm*).

Text C-1

Isolation of corrected iPSCs

On day one post plating we measured the presence of S1mplex within the nucleus as well as identified µFeatures that contained only one cell to ensure clonal populations. On days two through six we measured the number of cells to obtain a growth rate via day over day change. Finally, on day 7 we measured the number of cells as well as stained cells with LysoSensor to identify populations that may have been edited using phenotypic recovery as a marker. We also mock transfected WA09 and Pompe iPSCs and plated them on ArrayEdit and subjected them to the same high content analysis as a control. After 7 days we plotted each individual µFeature by its Lysosensor intensity and growth rate and color coded each feature by the presence of S1mplexes on day 1. We observed a large population of clones that grew slower than the mock transfected compound heterozygote colonies suggesting that that population may undergo editing events as seen previously on ArrayEdit. By comparing LysoSensor intensity we also observed that many of the µFeatures within the wells had higher, and therefore more acidic organelles than mock transfected compound heterozygote iPSCs. In fact, many of the µFeatures had similar intensities to control hPSC colonies. By combining these data with the presence of S1mplexes we were able to select colonies that were potentially preferably edited at either loci individually or both simultaneously. Using this knowledge, we isolated colonies of interest for expansion and analysis.

Following expansion of selected clones, we analyzed each one at both loci for the correction of mutations. We also designed the introduction of a PAM codon wobble to ensure that supplied donor DNA was the source for repair. When we looked at colonies that only had the presence of one S1mplex on day 1 we obtained clones that were edited at the specified allele. Further, colonies that were positive for one S1mplex were not observed to be edited at the other locus. We next analyzed clones positive for both S1mplexes and managed to obtain a

clone that was edited at both alleles and also contained the PAM wobble. There was also one colony that contained one PAM mutation while the other allele was repaired but did not introduce the novel mutation. Importantly, across all screened clones we did not obtain any that contained indel products. We then selected one clone from each population (edited at either allele individually, or both) to assay for phenotypic recovery.

Table C-1: Protospacer and respective PAMs used for genomic targeting.

Name	Protospacer	PAM
1441delT sgRNA	CTCGTTGTCCAGGTAGGCCC	GGG
2237G>A sgRNA	TGGACCACCAGCTCCTGTAG	GGG

Table C-2: ssODNs used to direct HDR after DSB Formation.

Name	Sequence (5`-3`)
1441insT ssODN	CTTCCATGCAGGCCCTGGGTGGGGCCGGGTCTCCCCACTGCAGC CTCTCGTTGTCCAGGTATGGCCCGGATCCACTGCCTTCCCCGACTTCACCAACCCC
2237A>G ssODN	TGCCCATCCCCCTTGCAGGTTCCCCAAGGACTCTAGCACCTGGA CTGTGGACCACCAGCTCCTGTGGGGAGAGGCCCTGCTCATCACCCCAGTGCTCCAG

Table C-3: Forward and reverse primers for genomic loci.

Name	Primer F (5`-3')	Primer R (5`-3`)
1441delT genomic	AGCTGCTCATTGACCTCCAG	CAATCCACATGCCGTCGAAG
2237G>A genomic	AATTCAGCCTCTTCCTGTGC	CATACGTTCCTCTTTCCGCC
Full length genomic	TGACAGGTTTCCCTCTTCCCAG	TTGATAACCTACACTGCGGGGG
1441delT qPCR/NGS	AGTGGGGCTTCCATGCAG	GGTTGGTGAAGTCGGGGAAG
2237G>A qPCR/NGS	CCAAGGACTCTAGCACCTGGAC	GGGAAGTAGCCAGTCACTTCGG

Table C-4: Off-target sequences and corresponding genomic locus for each sgRNA used. Mismatches from protospacer are labelled in red.

sgRNA Target Sequnce	Off-target sequence		Chromosome location
2237G>A TGGACCACCAGCTCCTGTAG	OT1	T**A**GACCACCAGCTCCTG**C**AG	chr8:-42696136
	OT2	**CTC**ACCACC**T**GCTCCTGTAG	chr9:-123379574
	OT3	T**T**GACCA**G**CAGCTCCTGT**C**G	chr15:-77699091
1441delT TGGACCACCAGCTCCTGTAG	OT1	CT**G**GTTGTCCAGGT**G**GGCCC	chr19:+9976610
	OT2	CTCG**A**TG**G**CCAGGTAGGCC**T**	chr9:+113788274
	OT3	**A**T**TA**TTG**A**CCAGGTAGGCCC	chr20:-42195228

Table C-5: Forward and reverse primers used to amplify off-target genomic loci.

Off-Target Primer	Primer F (5`-3')	Primer R (5`-3`)
2237-OT1	CCCTCCTCTGTGTGCCATTA	GTGCCATATTTTGGGGACCAC
2237-OT2	GGGGCATGGTCAGATGATGG	CACAGAAATTCCTGAGGCCAAC
2237-OT3	GGAGAGGCTGACCTTCATGG	TCGTGCTTTCCTGACCATCG
1441-OT1	AGTGTGCTTCCACTGTCGTT	GTGCGGGTAACCTTCTCCAT
1441-OT2	TTCCTCTGCTGCTGAGTTGG	GCCGATTAAAAGGCTGTCGC
1441-OT3	AGAGCCCTGGAGGTCATTGT	CTGTCTGGCCTCTGAATCGG

Table C-6: Subset of mutations in *GAA* and accompanying allele-specific sgRNAs.

Red denotes location of mutations relative to wildtype for allele-specific sgRNAs.

Mutation	sgRNA Sequence	Allele Frequency ($x10^{-5}$)
c.118C>T	GAGGAGCCACTCAGCTCTCAGGG	0.86
c.258dupC	ATCGAAGCGGCTGTTGGGGGGGG	2.65
c.525delT	CTGGACGTGATGATGGAGAC-GAG	7.04
c.1822C>T	AGTGGCCGGCGTATCAGCCGTGG	2.76
c.1827delC	TGCTGGCCACGGCCGATA-GCCGG	3.75
c.1930_1936dupGCCGACG	AAGCCGCAGACGTCGGCCGTCGG	1.17
c.2242dupG	CACCAGCTCCTGTAGGGGGGAGG	1.66
c.2560C>T	ACCAAGGGTGGGGAGGCCTGAGG	21.5
c.2662G>T	TAACACGATCGTGAATTAGCTGG	1.65

Appendix D: Genome editing in human pluripotent stem cells

Work in this chapter was adapted from:

Genome editing in human pluripotent stem cells

Jared Carlson-Stevermer and Krishanu Saha

Methods in Molecular Biology 1590 p.165-174, 2017

Introduction

Nuclease based gene-editing relies on endogenous DNA damage repair processes to edit desired sequences in the genome. Double strand breaks (DSBs) generated by nucleases are typically repaired via one of two processes: homology-directed recombination (HDR) and non-homologous end joining (NHEJ)[246]. HDR using donor DNA templates or single-stranded oligos (ssODNs) results in error-free incorporation of foreign DNA sequences. NHEJ, in contrast, generates a variety of insertions and deletions at the DSB site, which in some cases results in frameshifts causing premature stop codons to generate knockout cell lines.

Currently, the most common method to generate gene-edited hPSCs uses a clustered regularly interspaced palindromic repeat (CRISPR)/CRISPR associated protein (Cas9) system[135]. Engineering of the type II bacterial CRISPR associated (Cas) system into human cells[135] has dramatically increased the rate of which genetic engineering has progressed. Cas9 is an RNA-guided endonuclease that uses short guide RNAs (gRNAs) to selectively create double strand breaks (DSBs) at any point in the genome that is flanked by a 3' protospacer adjacent motif (PAM). When DSBs are created in cells, endogenous repair mechanisms are activated, resulting in repair from a donor via HDR or NHEJ. This method of gene editing was previously shown using zinc-finger nucleases[247] and TALENs[248] in hPSCs. HDR mediated insertion of transgenes occurs naturally in hPSCs[249] but occurs at a much higher rate when DSBs are introduced. Transgenes can be stably inserted into the genome by HDR if a donor is supplied that contains the gene surrounded by regions homologous to the area where it should be inserted. Single nucleotide polymorphisms (SNP) can also be introduced by transfecting with a ssODN that is homologous to the region around the cut site but for the point mutation of interest[247 250].

There are currently three methods in which the necessary components can be introduced into the cell: (*i*) transient transfection of two plasmids, one containing the coding

sequence for Cas9 and the other expressing the gRNA[135 250], (*ii*) viral integration of the Cas9 coding sequence and transient expression of the gRNA [144] and, (*iii*) viral integration of both Cas9 and gRNA sequences[138 140 251]. Delivery of nucleic acids to hPSCs is typically limited using standard lipofection reagents, so most studies utilize electroporation or nucleofection[252]. Efficiency of the two plasmid system has been increased by utilizing a Cas9-2A-GFP plasmid[134] which can be used to enrich the cell population expressing Cas9 via fluorescence-activated cell sorting (FACS) of GFP-expressing cells. Using these strategies, we have been able to insert transgenes (**Figure D-1**)[253] and detect frame-shifting indels causing a knockout phenotype in over 25% of cultured hPSCs.

Materials

- Dulbecco's Phosphate-Buffered Saline, without calcium, without magnesium (DPBS, Life Technologies)
- mTeSR1 media (STEMCELL Technologies)
- Y-27632 (ROCK Inhibitor, STEMCELL Technologies)
- Cas9-2A-GFP Plasmid (Addgene #44719)
- gRNA vector (Addgene #41824)
- DMEM/F12 (Life Technologies)
- Accutase (STEMCELL Technologies)
- Microcentrifuge Tubes (Fischer Scientific or other source)
- 15mL Conical Tubes (BD Biosciences or other source)
- Standard micropipettes with P2.5, P20, P200, P1000 + tips
- Hemocytometer (Hausser Scientific or other source)
- Bio-Rad cuvette, 0.4 cm (1 per transfection)
- Matrigel-coated tissue culture plates (Corning)
- Electroporator (Bio-Rad or equivalent for mammalian cells)
- QuickExtract (Epicenter)
- AccuPrime Taq High Fidelity (Life Technologies)
- Forward and reverse primers for genomic locus (IDT or other source)

Transfection Protocol

Set Up

1. Seed hPSCs at a density such that they are about 70-80% confluent in 2-3 days (*see* Note 1).
2. 24 hours prior to transfection, change media to include 10 µM ROCK Inhibitor (*see* Note 2).
3. Prepare Matrigel plates for subsequent culture of electroporated cells. Cells may be plated at a density of 10^4-10^5 cells per cm^2.
4. Prepare plasmid stocks for electroporation (*see* Note 3).
5. Warm Accutase and cell culture reagents to 37^{o}C prior to starting experiments.

Experimental

1. Prepare electroporation solution containing plasmids and media:
2. Determine plasmid volume for each plasmid. 2.5 µg of Cas9 plasmid and 1.0 µg of gRNA plasmid are needed for every 1x10^6 cells to be transfected (*see* Note 4). If using a donor DNA template, add 2.0 µg per 1x10^6 cells. If using a ssODN template, add 1µL of a 10 µM stock (*see* Note 5) [250]. For protocol modifications for ZFN and TALEN strategies, *see* Note 6.
3. Pipette 400 µL mTeSR1 media into a to microcentrifuge tube.
4. Thaw and gently vortex plasmid stocks.
5. Add plasmids to media.
6. Plasmid solution should be kept on ice.
7. Aspirate media from the hPSC wells and wash with 1-2 mL of DPBS.
8. Harvest cells via Accutase enzymatic detachment:
9. Add enough enzyme to cover well and place in 37^{o}C incubator for 5-7 minutes.
10. Dislodge cells by tapping the culture plate, then collect cells from each well using a P1000 pipette and add cell suspension to a 15 mL conical tube.
11. Gently pipette up and down to dissociate any remaining clumps (*see* Note 7).
12. Immediately rinse each well with an additional 0.5 volumes of culture media or DMEM/F12 and add to the conical tube to collect any remaining cells and neutralize Accutase by dilution.
13. Obtain cell count via hemocytometer or flow cytometry-based methods.
14. Remove sufficient volume to obtain the desired cell count per transfection (0.5-5x10^6 per transfection has been shown to be effective) and place into a 15 mL conical tube.
15. Pellet cells by centrifuging 5 min at 200 x *g*.
16. Aspirate supernatant and resuspend pellet in the cold electroporation solution created in step 1 (*see* Note 8).
17. Transfer solution into a 0.4 cm cuvette (*see* Notes 9, 10).
18. Pulse cuvette in electroporator. High efficiency has been observed using an exponential decay waveform, 250V, 750 µF capacitance, and infinite resistance (*see* Notes 11, 12, 13).
19. Using a P200 pipette remove electroporation solution from cuvette and transfer into a 15 mL conical tube, containing sufficient volume of mTeSR + 10 µM ROCK Inhibitor medium for desired number of wells (*see* Note 14).

20. Wash cuvette with a 200µL aliquot of mTeSR media to remove any remaining cells, add wash to the 15 mL conical tube.
21. Plate cells into culture dish.
22. At 24 hours post electroporation change media with mTeSR1 + 10 µM ROCK Inhibitor.
23. At 48 hours post electroporation, ROCK Inhibitor may be removed and selection agents (if any) may be added.

Assay Efficiency of NHEJ-mediated Indels

1. Allow cells to grow to ~80% confluency.
2. Passage wells at a 1:2 split ratio.
3. Allow to grow until all wells are again ~80% confluent with daily media changes of mTeSR without ROCK inhibitor.
4. Harvest genomic DNA using QuickExtract solution or equivalent extraction method (*see* Note 15).
5. Run one of the following assays for detection of NHEJ: T7 Endonuclease (**Figure D-2**) or SURVEYOR [254] (*see* Note 16), PAGE gel [255], MiSeq [256], HRMA [257]. Mutational frequencies can vary between assays, as do the lower limits of resolution for each assay. T7 Endonuclease assays require ~5% NHEJ to be resolved whereas MiSeq can detect a single cell in which a mutation has occurred. The sensitivity of other assays fall between these two extremes.

Isolate Clonal Populations via FACS sorting

1. Schedule a time to use the FACS sorting facility 2-3 days after electroporation.
2. Follow electroporation protocol. In step 1 of the electroporation protocol, substitute Cas9 plasmid with Cas9-2A-GFP plasmid. Alternatively, in step 1 of the electroporation protocol, add 250 ng of GFP plasmid per $1x10^6$ cells (Addgene #13031).
3. At 24-48 hours post electroporation, confirm transfection success using fluorescent microscopy.
4. Add 10 µM ROCK Inhibitor to transfected wells one hour prior to dissociation (*see* Note 17).
5. Obtain a single cell dissociation using Accutase as described in step 3 of electroporation protocol.
6. Centrifuge for 5 min at 200 x *g*.
7. Resuspend in 2-3 mL of mTeSR media + 10 µM ROCK inhibitor and pass through a 40 µm cell strainer cap into a FACS tube.
8. Keep cells and collection tubes on ice (*see* Note 18).
9. Using FACS, collect cells that are expressing GFP. Ideally, use cells that were mock electroporated (no plasmid) to set up gates for fluorescence (**Figure D-3**).
10. Transfer cells into fresh wells of a 6-well plate at a density of ~1000 cells/cm^2 in mTeSR1 with 10uM ROCK inhibitor.

11. Allow cells to grow for 7-10 days while performing daily media changes with mTeSR1 with 10uM ROCK inhibitor (*see* Note 19). Other cloning methods can also be successful (*see* Note 20)
12. When colonies begin to form, use a P200 pipette to lightly scrape at colonies of interest while slowly releasing the plunger (*see* Note 21).
13. Transfer each clone to a single well of a 24-well plate.
14. When each clone is individually confluent (typically 4-5 days after step 13), passage at a 1:3 ratio into 3 wells of a 12 well plate.
15. From one well, harvest genomic DNA using QuickExtract solution (*see* Note 15) or equivalent extraction method.
16. Individually freeze other two wells (see Chapter 12 in this book) in liquid nitrogen.
17. Prepare harvested genomic DNA for Sanger sequencing[258]. After receiving sequencing results, thaw clones that contain desired mutations (see Chapter 11 in this book).

Notes

1. Cells must be in an actively dividing state
2. The longer that the cells are in ROCK Inhibitor, the greater the cell attachment following electroporation (18-24 hours is recommended). The minimum amount of time for cells to be in ROCK inhibitor that has an appreciable effect is 1 hour.
3. Concentration will depend on efficiency of plasmid isolation protocol. Ideal values will be greater than 1 µg/mL as determined by NanoDrop or equivalent method. Plasmid purification protocol must contain an endotoxin removal step.
4. An alternative to transfecting plasmids is to directly transfect the RNA and Cas9 protein required. The advantage of this alternative is that there is less of a lag time until results can be collected. In this alternative, generate gRNAs via *in vitro transcription* [144 259] off of the created gRNA plasmid using the MEGAshortscript T7 kit (Ambion). Cas9 coding sequence can be cloned into a bacterial expression vector and expressed in a bacterial host before being purified via an attached His tag [13].
5. Optimal ssODN design is necessary for successful HDR. Some of the considerations that must be taken into account are the mutation of the PAM sequence via codon wobble to produce a silent mutation[260]. Without this modification, the Cas9-gRNA complex will cleave the ssODN. Secondly, design intended point mutations in the center of ssODN with ~40 bp flanking homology on either side[137]. It is also advisable to design your ssODN to either introduce or remove a restriction enzyme site[261]. This consideration will allow for facile analysis of clones, without having to perform one of the NHEJ assays mentioned in step 4.
6. Both zinc finger nucleases (ZFNs) and TALENs remain viable options for gene editing. If using a either method, electroporate using 0.5 µg of nuclease encoding plasmid and 4 µg of donor DNA per $1x10^6$ cell [248 262]. All other steps of the protocol remain the same.
7. Single cell dissociations are important as electroporation of cell clumps decreases efficiency.
8. It is important to remove as much supernatant as possible as excess volume during electroporation will decrease efficiency.

9. Bubbles must be avoided, as they will cause increased cell death during electroporation.
10. Avoid touching the metal sides of cuvette as oils may transfer and cause waveform to be disrupted.
11. Individual machines vary, so these values will have to be optimized on a per-machine basis. High voltage electroporations will increase efficiency while also increasing cell death. We recommend starting at our settings at changing the voltage in increments of ±25 V while holding the capacitance constant. Then increment capacitance ±50 µF at optimum voltage.
12. Make a note of time constant (TC) values. Highly variable values may indicate a malfunctioning machine.
13. Variability between stem cell lines, especially between embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs), can have a drastic effect of cell viability. If you are experiencing high amounts of cell death go back and reoptimize electroporation conditions for each cell line. This is especially necessary if transitioning conditions from an ESC line to an iPSC line.
14. We generally electroporate enough cells to provide 6 replicates in a full 6 well plate. Using this setup 12 mL of mTeSR would be needed per electroporation.
15. To perform an extraction using QuickExtract cells are first trypsinized, pelleted, and resuspended in 500 µL of QuickExtract before being transferred in to a microcentrifuge tube. Then place microcentrifuge tube in a heat block at 65℃ for 15min, 68℃ for 15min, and 98℃ for 10 min.
16. SURVEYOR from Transgenomic is similar to the T7 Endonuclease assay. We have found that reagents can be exchanged between the two kits.
17. ROCK inhibitor may be left in culture from pre-electroporation through FACS without decrease in viability or proliferation rate.
18. Sorting cells into collection tubes may be beneficial to sorting directly into culture plates. Sorting into plates causes cell death due to slamming into the plastic on bottom of plate. Sorting parameters need also to be modified for hPSCs. For a BD Aria, we use a nozzle size of 100 µm and a pressure of 20 psi.
19. Do not allow colonies to grow too large or merge.
20. Single cell cloning can also be done in a high-throughput method if mutations happen at a frequency of <1%. This method, known as sib-cloning, combines a TaqMan PCR assay with fluorescent probes that distinguish between mutant and wild type products. Mutant frequencies as low as 0.1% can be detected and enriched by selection and propagation to a further round of cloning[136].
21. Be sure to remove entire colony from culture plate. If cells are detached they may float through media and contaminate other clones being isolated. To avoid this contamination, change media between each clone being selected.
22. FACS sorting at clonal density culture can survive without differentiation 48-72 hours without changing media. This simulates the use of conditioned media and decreases stress on single cells. After this period change media daily. When colonies form after 7-10 days use a P200 pipette set to 200 µL and, with the plunger depressed, gently scrape at the colony edge causing it to lift off the plate. Simultaneously, slowly release the plunger to suck the entire colony into pipette tip. Transfer into new well and gently

pipette up and down 3-4 times to dissociate the colony. Empirically, this dissociation dramatically speeds up the rate of proliferation. Change media in new well every 24 hours. Wait until cells are ~80% confluent, usually about four days, and then passage 1:3 into 3 wells. Wait another 4-5 days or until wells are again ~80% confluent before freezing 2 wells. 2 wells are frozen to provide a backup in the event one sample does not survive the freeze/thaw cycle. Collected genomic DNA from final well should be PCR amplified using AccuPrime Taq with primers for the locus of interest and then Sanger sequenced using the same forward primer as used for amplification.

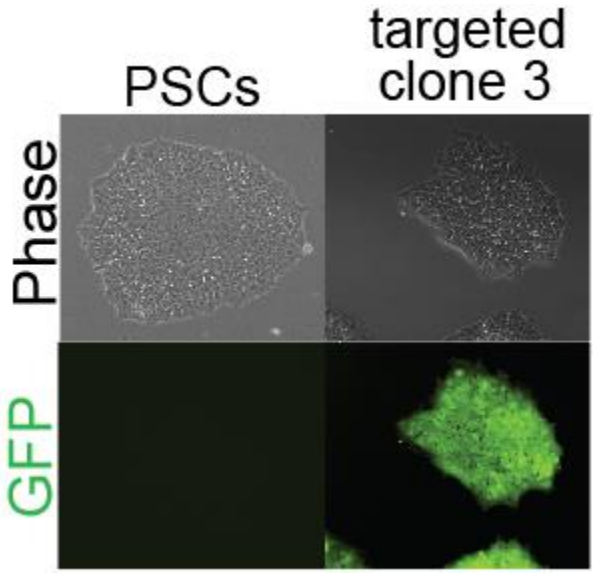


Figure D-1 | Electroporation of Cas9 plasmid and *AAVS1* T2 gRNA plasmid (Addgene #41818), and GFP donor plasmid (Addgene #28014) yielded target GFP cell lines. This example uses HDR targeting techniques.

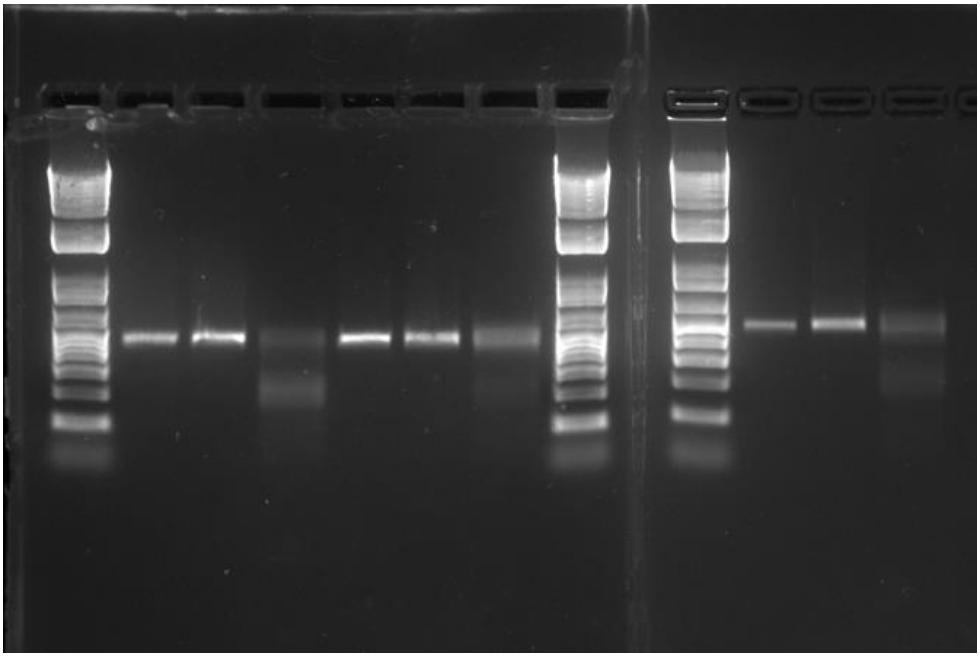


Figure D-2 | Representative image from 3 T7 assays done in parallel. Each sample is from a different gRNA targeted against a mCherry transgene. 3 lanes of each genomic PCR product are run: lanes 2, 5, and 10 are purified genomic PCR product, lanes 3, 6, and 11 are PCR product after hybridization to form heterodimers and in lanes 4, 7, and 12 the heterodimers have been cleaved with T7 endonuclease. Primers: Forward: AAGGGCGAGGAGGATAACATGG; Reverse: TTGTACAGCTCGTCCATGCCG.

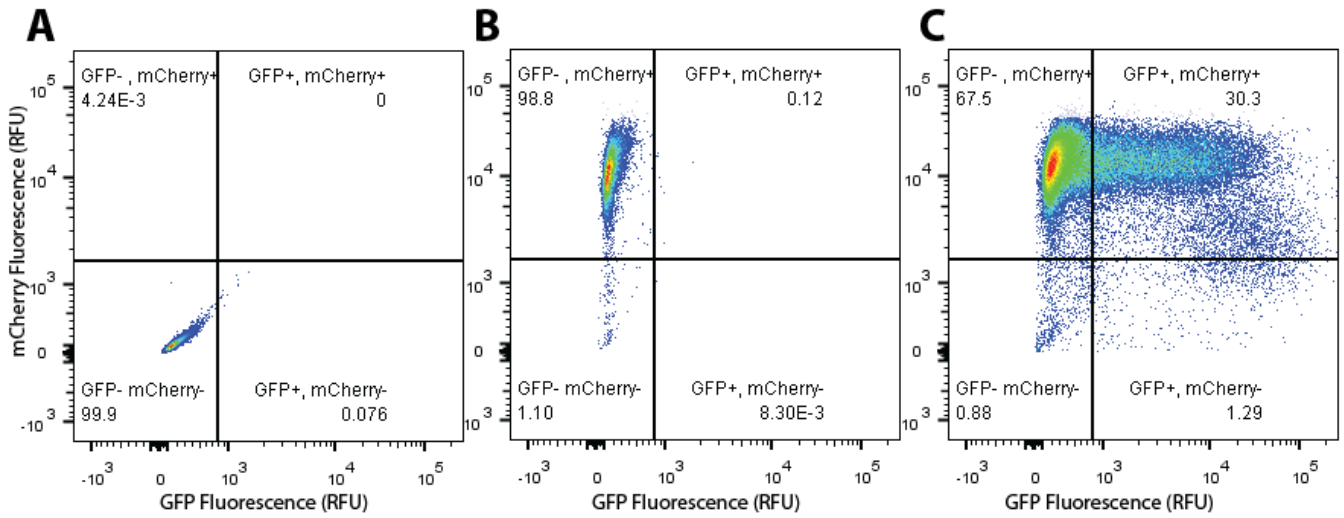


Figure D-3 | Representative flow cytometry plot for sorting Cas9- and GFP- expressing hPSCs. (**A**) Non-fluorescent hPSC negative control (**B**) Mock electroporated H9-H2B-mCherry expressing cells. (**C**) Cas9, eGFP-N1 co-electroporated hPSCs. This method shows >30% efficiency of introduction into hPSCs.