Genomic Editing Tools to Model Human Diseases with Isogenic Pluripotent Stem Cells

Huen Suk Kim, Jeffrey M. Bernitz, Dung-Fang Lee, and Ihor R. Lemischka

Patient-specific induced pluripotent stem cells (iPSCs) are considered a versatile resource in the field of biomedicine. As iPSCs are generated on an individual basis, iPSCs may be the optimal cellular material to use for disease modeling, drug discovery, and the development of patient-specific cellular therapies. Recently, to gain an in-depth understanding of human pathologies, patient-specific iPSCs have been used to model human diseases with some iPSC-derived cells recapitulating pathological phenotypes in vitro. However, complex multigenic diseases generally have not resulted in concise conclusions regarding the underlying mechanisms of disease, in large part due to genetic variations between disease-state and control iPSCs. To circumvent this, the use of genomic editing tools to generate perfect isogenic controls is gaining momentum. To date, DNA binding domain-based zinc finger nucleases and transcription activator-like effector nucleases have been utilized to create genetically defined conditions in patient-specific iPSCs, with some examples leading to the successful identification of novel mechanisms of disease. As the feasibility and utility of genomic editing tools in iPSCs improve, along with the introduction of the clustered regularly interspaced short palindromic repeat system, understanding the features and limitations of genomic editing tools and their applications to iPSC technology is critical to expanding the field of human disease modeling.

Introduction

HUMAN EMBRYONIC STEM CELLS (hESCs), first derived from early embryos in 1998, are not only a valuable tool for the study of early human development, but also provide an inexhaustible cell source for cell replacement therapy. While barriers of immune rejection and ethical controversy prevent the widespread clinical use of hESCs [1,2], recent alternatives that take advantage of cellular reprogramming help to overcome both these limitations. Mechanistically, reprogramming is mediated by transcription factors, as best illustrated by the ectopic expression of MyoD inducing myogenesis in fibroblasts and other somatic cells [3]. Taking advantage of this, Yamanaka and colleagues generated induced pluripotent stem cells (iPSCs) by the ectopic expression of four transcription factors: OCT4, SOX2, KLF4, and cMYC [4,5]. Since the initial report, there have been many refinements of iPSC generation techniques [6–9], and to date many cell types including motor neurons, cardiomyocytes, and various hematopoietic cells have been produced from iPSCs in vitro, with functional studies of these cells in mouse models demonstrating their potential for replacement therapy [10–13].

While iPSCs provide promise for cell replacement therapies, they also represent a powerful tool for human disease

modeling. As iPSCs are generated from patient cells, they can be used to generate specific cell types affected during disease. This would provide an unlimited source of cells for disease modeling and drug screening [14,15]. As iPSC generation has become more efficient, many patient-specific iPSCs have been derived to model disease [16]. Generally, monogenic diseases with clear causative mutations affecting well-characterized cell types have successfully recapitulated pathological phenotypes using iPSC technology [16,17]. For instance, long QT iPSCs with missense mutations in the *KCNH1* and *KCNH2* genes generate cardiomyocytes with increased depolarization and reduced potassium current, and spinal muscular atrophy iPSCs with SMN1 mutations generate fewer numbers of motor neurons with degenerated and diffuse synapses [18–20].

On the other hand, other studies using disease-specific iPSCs have not been successful at modeling diseases [17]. There are many critical reasons as to why this may be the case. First, clonal variations caused by several factors during the reprogramming process, iPSC passage number, and culture conditions can affect the epigenetic status of individual iPSC clones [21–25]. Second, in modeling diseases with sporadic or late onset, such as Alzheimer's and Parkinson's diseases (PD), in vitro assays often show insignificant differences between disease and control cells, suggesting that

specific genetic variations between individuals may work as genetic modifiers that influence susceptibility to these diseases [26–28]. It therefore becomes imperative that excess genetic variation between iPSC clones and controls should be removed to ensure more precise comparative and molecular analysis when modeling diseases [17]. To this end, generating isogenic sibling cell lines from patient iPSCs by altering only a few nucleotides is undoubtedly the most accurate way to establish a genetically defined condition.

Conventionally, homologous recombination (HR) has been a robust and frequently utilized method to modify genomic loci, most notably in the generation of diverse knockout mice. Despite requiring intensive efforts, HR has been used in hESC as well. In 2003, Zwaka and Thomson successfully disrupted the hypoxanthine phosphoribosyl-transferase (HPRT) gene and also generated OCT4-GFP reporter hESC, which were the first knockout and knock-in hESC lines, respectively [29]. The conventional gene targeting methods used to generate the Oct4-GFP hESCs established the technical framework for the generation of many other transgenic hESC lines, and the cells themselves have allowed for monitoring cell fate specification in real time during differentiation [30–33].

However, it is a prevalent idea that successful gene targeting with HR is particularly difficult in human pluripotent stem cells (hPSC). hPSCs tend to prefer to utilize nonhomologous end joining (NHEJ) to repair double-strand breaks (DSBs), as opposed to mouse ESC, which have a high propensity for HR. In addition, hPSCs do not grow well as single cells, and instead form colonies in culture making clonal selection challenging. Furthermore, they readily undergo anoikis, causing a dramatic drop in the efficacy of gene targeting when using electroporation or transfection reagents [34–36]. To facilitate low frequency of HR in hPSC, bacterial artificial chromosomes (BACs) or adeno-associated viruses gene delivery systems have been applied [37–39]. However, the most prevalent DNA delivery system in hPSCs remains intergrase-defective lentiviral vector (IDLV) [40–42]. This system, developed from the human immunodeficiency virus type I (HIV-1), has been utilized in various cell types in vitro and in vivo, and delivers genetic information in the form of episomal DNA that confers transgene expression without affecting self-renewal or pluripotency in hPSCs.

Recently, besides gene delivery systems, genetic editing tools is considered a method to increase the efficiency of modifying genomic loci, and a recent wave of studies have applied the genetic editing tools to iPSC technology. This has led to the generation of isogenic iPSC disease and control lines that have been critical to advancing the utility of iPSCs to the human disease modeling field.

Rationale of Genomic Editing Tools: Utilizing Endogenous DNA Repair Systems

Genomic editing refers to techniques that are able to change one or more nucleotides in a given gene using engineered nucleases. These nucleases induce DNA DSBs in a sequence specific manner, but rely on the endogenous cellular DNA repair machinery to repair the breaks following DNA cleavage.

There are two endogenous DNA repair systems that work in conjunction with genomic editing tools: NHEJ and HR

repair (Fig. 1) [43]. NHEJ joins two broken ends by synthesizing compatible overhangs that can then be ligated together [44]. When a DSB occurs, the two broken ends need to maintain proximity to induce subsequent repair processes. Ku proteins detect and bind to the broken region to recruit the DNA-dependent protein kinase (DNA-PK, ARTEMIS) that trims 5' and 3' overhangs to make two compatible ends to ligate together. Polymerase- μ then fills in any gaps remaining between the joined compatible ends, and finally, DNA ligase IV complex (XRCC) joins the break by ligating the two strands together. It is important to note that NHEJ is an imperfect process. Between ARTEMIS-DNA-PK recruitment and the final ligation, nucleotides may be added or deleted resulting in imperfect DNA repair (Fig. 1). When combined with genomic editing tools, NHEJ-mediated repair can induce small insertions or deletions in the target gene, which can consequently induce gene knockouts [45,46]. Additionally, using NHEJ-mediated repair large transgenes of up to 14 kb in size have been introduced into the genome [47].

Contrary to NHEJ, HR repair is precise and accurate, utilizing a homologous DNA template to guide the repair process. HR is mediated by the highly conserved enzymes RAD51 and DMC1 that catalyze the pairing and switching of homologous DNA sequences [48]. When DSB occurs, double-stranded ends of breaks are resected to become single stranded with 3' hydroxyl groups. RAD52 recombinases are then recruited and bind to the single-stranded DNA, enabling the broken DNA to invade into a homologous template. The invading sequence subsequently anneals to the complementary template forming a Holliday junction, a mobile junction between four strands of DNA. The damaged region is then extended, copying the sequence from the undamaged complementary template, thus repairing the damaged DNA strand, and the process completes with the resolution of the Holliday junction by specialized endonucleases (Fig. 1) [48]. To use HR repair with genomic editing tools, cotransfection of a donor vector containing homology arms is required. Usually, a linear donor vector also contains a selection marker such as GFP or an antibiotic resistance cassette. Additionally, single-stranded DNA oligonucleotides without selection markers are used to the target site of interest [49–51].

Custom DNA Binding Domains of Zinc Finger Nucleases and Transcription Activator-Like Effector Nucleases

The specificity of genomic editing depends on the DNA binding domains of the tools of choice. Zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) are two common options for genomic editing studies. They share two fundamental characteristics in that they contain customized DNA binding domains and a FokI nuclease, which is able to cleave nucleotides in a nonspecific manner [52]. ZFNs are the synthetic fusion of multiple zinc finger domains, each recognizing specific target DNA sequences (Fig. 2C) [53]. The zinc finger domain is the most abundant DNA recognition domain in eukaryotes, and as such the diversity of known domains enables the targeting of a broad range of sequences. An individual zinc finger domain consists of 30 amino acids,

FIG. 1. Overview of nuclease-mediated double-strand breaks (DSB) repair system. When DSBs are made by a genomic editing tool, two endogenous DNA repair systems, either nonhomologous end joining (NHEJ) or homologous recombination (HR) repair are activated. NHEJ joins two ends by DNA ligase IV (XRCC) following the addition of random nucleotides by Polymerase- μ . On the other hand, HR requires a homologous template for repair. Recombinases are recruited to break points and guide the end of single strand into a homologous template, which can be a donor vector or strands of sister chromatid. The damaged region is then repaired by copying the sequence from the undamaged complementary template. Color images available online at www.liebertpub.com/scd

which form two β -sheets and an α -helix that coordinate to stabilize a zinc ion (Fig. 2A) [54]. Each domain binds three sequential nucleotides in the major groove of double-stranded DNA through its unique α -helical motif (Fig. 2B). As multiple zinc finger domains are needed to confer ZFN specificity, the DNA binding domains of ZFNs are often composed of six to seven zinc finger domains that target sites of 18–21 nucleotides (Fig. 2C) [55]. Statistically, well-designed ZFNs should only target a single site in the genome, as a random 18 bp sequence should only be found once within 68 billion base pairs.

To date, ZFNs have been widely used in many organisms and cell lines, and these have also been utilized for a variety of therapeutic purposes including a Phase I clinical trial to disrupt the C-C chemokine receptor type 5 (CCR5) gene as treatment for HIV infection [56–60]. However, there are limitations to their applications. While the DNA binding

regions of zinc finger domains are engineered to recognize most of the 64 possible nucleotide triplets, there are specific nucleotide triplets that do not have corresponding zinc finger domains because zinc finger domains preferentially bind guanine-rich sequences [61]. Additionally, linking together multiple zinc finger domains causes interactions between the domains that reduce their DNA binding specificity, and the construction process involved in synthesizing new ZFNs is complicated and tedious [62,63].

On the other hand, TALENs represent a new genomic editing system that is considered a convenient alternative to ZFNs. The DNA binding domains of TALENs consist of conserved repeated protein modules adapted from the transcription activator-like effectors (TALEs) in various strains of *Xanthomonas*, and these are more flexible and easier to assemble than those of ZFNs. Bacterial TALEs are secreted proteins that invade plant cell nuclei to activate gene

FIG. 2. Structure of zinc finger nucleases (ZFNs) on target DNA. (A) Each zinc finger binds 3 bp of target DNA, with two pairs of cysteine and histidine residues binding a Zn^{2+} ion. (B) A set of six connected zinc fingers that recognizes nucleotide triplets (ANN; A, adenine). α -Helix of each zinc finger binds major groove of DNA [image from the RCSB PDB (www.rcsb.org) of PDB ID 2I13 [55]]. (C) The mechanism of action of ZFNs: dimers of zinc finger proteins with conjugated FokI nuclease bind target DNA. Each zinc finger domain recognizes either the left or right side of target sites. The binding domain of FokI binds the recognition site and its enzymatic domain cuts a proximal random sequence. ZFP, zinc finger protein. Color images available online at www.liebertpub.com/scd

FIG. 3. Structure of transcription activator-like effector nucleases (TALENs) on target DNA. (A) TALE domains are 34 amino acids long, and bind a single nucleotide with specificity conferred by two residues in positions 12 and 13 (RVDs). Asn/Asn (NN), Asn/Ile (NI), His/Asp (HD), and Asn/Gly (NG)—that recognize the nucleotides guanine, adenine, cytosine, and thymine respectively. More information about TALENs at the addgene website (www.addgene.org/TALEN/guide/). (B) Two compiled arrays of TALEN binding domains combine to bind the left and right side of the target locus. The FokI enzyme targets and cleaves sequences near the recognition site nonspecifically. Different color bars represent each TALE domain, with RVDs included. Color images available online at www.liebertpub.com/scd

transcription through binding to target gene promoters, aiding in the establishment of bacterial infection [64]. The TALEN DNA binding domain is a series of 33–35 amino acid repeats that form two short helices, with each repeat domain binding a single nucleotide through repeat variable di-residues (RVD) (Fig. 3A) [65]. RVDs are found at positions 12 and 13 of the amino acid chain repeat sequence and confer sequence specificity to the TALENs (Fig. 3B). There are four RVDs—Asn/Asn (NN), Asn/Ile (NI), His/Asp (HD), and Asn/Gly (NG)—that recognize the nucleotides guanine, adenine, cytosine, and thymine respectively [66]. TALENs work in pairs to target a genomic locus, like ZFNs, but maintain a 14–18 bp spacer between the two binding domains (Fig. 3B). Also like ZFNs, TALE repeats are linked together to recognize stretches of DNA sequences, but TALEs are easier to link together and linking multiple TALEs do not alter binding specificities, allowing TALENs to have longer domains with greater sequence specificity. More details regarding TALEN technology is available on the Addgene Website (www.addgene.org/TALEN/guide/).

Clustered Regularly Interspaced Short Palindromic Repeat

Clustered regularly interspaced short palindromic repeats (CRISPRs) are a novel genomic editing tool that is rapidly gaining popularity. They are adapted from RNA-based adaptive immune systems that act by destroying viruses of *Streptococcus pyogenes*. CRISPRs are comprised of a Cas9 protein and short RNA sequences (Fig. 4A) [67]. Cas9 is an endonuclease that can make a break in DNA molecules guided by short RNAs. The short RNAs include two non-coding RNAs, a CRISPR RNA (crRNA), and a transactivating crRNA (tracrRNA). The endogenous CRISPR locus contains the Cas gene and a series of 21–47 bp repeats separated by unique spacer sequences called protospacers originating from viral parasites [67]. When viruses invade *S. pyogenes*, the CRISPR nuclease detects and cleaves protospacer adjacent motifs (PAMs) in target viral DNA. These cleaved protospacer sequences are collected and integrated into the endogenous CRISPR locus so that upon secondary infection by the same virus, this integrated sequence is

FIG. 4. Schematic illustration of clustered regularly interspaced short palindromic repeats (CRISPRs) bound to target DNA. (A) DSBs mediated by Cas9:crRNA-tracrRNA. (Arrows mark location of breaks) (B) Cas9 with chimeric guide RNA (gRNA) bound to target DNA. The hairpin loop of the chimeric gRNA resembles the interaction between CRISPR RNA (crRNA) and transactivating crRNA (tracrRNA). (C) A double nicking system using Cas9 nickase (Cas9n; D10A mutation on RuvC-like domain displayed). Two gRNAs offset by generally 15 nucleotides recruit Cas9n, each nicking one strand of the target DNA. In the depicted scenario, the strand complementary to the chimeric RNA is cleaved by the D10A Cas9 nickase, generating 5' overhangs. If the enzyme was the H840A Cas9 nickase, the noncomplementary strand would be cleaved and 3' overhangs generated. Additionally, by designing chimeric RNAs 1 and 2 to be complementary to the target DNA strands opposite of those currently depicted, D10A and H840A mutants would generate 3' and 5' overhangs respectively. Color images available online at www.liebertpub.com/scd

transcribed and acts to guide the CRISPR to cleave the viral genome, which is known as a bacterial adaptive immune system [67,68].

In vitro, Cas9 of the CRISPR system requires complementary base pairing between crRNA, tracrRNA, and target DNA to efficiently cleave its target (Fig. 4A). A conserved PAM sequence, NGG, on the target DNA strand acts to recruit the Cas9-crRNA-tracrRNA complex to the targets seed region [69–71]. PAMs are short, but when mutated the complex loses target affinity and the efficiency of cleavage is reduced [70]. The seed region consists of approximately seven bases near the PAM sequence, and is where pairing between the crRNA and target DNA occurs (Fig. 4A, B) [70,71]. The seed region helps to determine DNA binding specificity, and mutations in this region prevent Cas9-mediated cleavage of the target DNA. The PAMs specify the cleavage sites on each target DNA strand, which is cleaved by separate domains of Cas9—the HNH domain (a zinc finger domain having two pairs of histidines and one asparagine) cleaves the target strand complementary to the crRNA, and the RuvC-like domain (a conserved domain in Holliday junction resolvase) cleaves the noncomplementary strand, inducing a DSB in the target DNA (Fig. 4A, B) [69].

The CRISPR system has recently been reengineered to increase its specificity and feasibility of application. To this end, the crRNA and tracrRNA have been fused into a single chimeric guide RNA (gRNA) [70], which contains a protospacer and a hairpin loop to mimic the natural base pairing interaction between tracrRNA and crRNA (Fig. 4B). In cleavage assays that check the feasibility of the chimeric

gRNA, it functions with efficiency similar to the smaller tracrRNA-crRNA duplex and the positions of the cleavage sites on target DNA are identical [70].

Limitations of Genomic Editing Systems

Despite the fact that each of the three genomic editing systems increases the efficiency of NHEJ or HR-mediated gene alterations, there is an inevitable disadvantage to engineered nucleases (Table 1). Ideally, DNA binding domain-based genomic editing systems have high specificity to target sites since the target sequence is flanked by two DNA binding domains that bind two unique sequences in the genome. However, the specificity of DNA binding domain relies on its binding affinity to DNA. It has been reported that high DNA binding affinities can cause imperfect specificity that can tolerate up to three nucleotide mutations [72]. As a consequence, nucleases often make off-target DSBs leading to mutations in unwanted sites [72]. Pattanayak et al. tested the off-target effects of ZFNs in vitro to elucidate ZFN binding specificities and its mutational tolerance. Interestingly, they observed that ZFNs engineered with more zinc finger domains, allowing for greater sequence specificity, still generated mutations in multiple off-target sites, suggesting that different strategies are required for optimizing DNA binding affinities to reduce off-target effects [72].

Epigenetic modification also contributes to the specificity of genomic editing systems. It has been shown that sites where engineered nucleases can cleave may not be consistent from one cell line to another due to different epigenetic marks

TABLE 1. COMPARISON OF THREE GENOMIC EDITING SYSTEMS

	<i>ZFNs</i>	<i>TALENs</i>	<i>CRISPR</i>
Features	Zinc finger-based DNA binding domain with restriction enzyme.	TALE-based DNA binding domain with restriction enzyme.	RNA-enzyme (Cas9) complex binds DNA in sequence-specific manner.
Components	Two arrays of zinc finger binding domains. FokI endonuclease.	Two arrays of TALE domains. FokI endonuclease.	gRNA. Cas9.
The mechanism of action	Each domain of ZFNs binds three paired nucleotides in the major groove of DNA, two multiple zinc fingers bind target sequence with specificity. Functions in dimmers. FokI detects conservative sequence and cut near the conserved sequence without specificity.	TALE domains bind a single nucleotide through repeat variable di-residue. Functions in dimmers. Same mechanism ZFN.	PAM site (NGG) and seed region detected by gRNA recruit Cas9. Recruited Cas9 makes DSB 3 bp upstream of PAM site.
Advantage	Improved ZFNs technology available commercially. Widely used and established feasibility.	Binding domains are more flexible and easier to assemble than ZFNs. Higher sequence specificity than ZFNs.	Easy to construct. High efficiency and specificity.
Disadvantage	Construction process is complicated. Some specific nucleotide triples do not have corresponding zinc finger domains.	Affected by DNA methylation status.	High off-target effects.

CRISPR, clustered regularly interspaced short palindromic repeat; DSB, double-strand breaks; gRNA, guide RNA; PAM, protospacer adjacent motifs; ZFN, zinc finger nucleases.

and chromatin structure of the disparate cells [73–76]. Bultmann et al. investigated how epigenetic modifications affect the targeting efficiency of TALENs at the OCT4 locus, observing that the active OCT4 promoter in ESCs was successfully targeted by TALENs while the silenced OCT4 promoter in neural stem cells inhibited TALEN binding [77]. The use of small molecule epigenetic modifiers, such as valproic acid or 5-aza-2'-deoxycytidine, to alter epigenetic status has been suggested to overcome this limitation [78].

The recent introduction of CRISPRs as an effective and convenient tool for gene editing has shown that they are an advantageous alternative to ZFNs and TALENs. The specificity of CRISPRs relies on a single easily engineered gRNA, and Hsu et al. have demonstrated that CRISPR-mediated cleavage is resistant to DNA methylation [79]. Additionally, with the use of chimeric gRNAs, in conjunction with human-optimized Cas9 proteins, the CRISPR system now functions with a high degree of efficiency in human cells, and it can be manipulated to target multiple genetic loci [80,81]. However, this high Cas9 efficiency has demonstrated significant increases in off-target frequency when compared with ZFNs or TALENs due to the leniency of CRISPRs toward mismatched base-pairing between gRNA and target sequences [82,83].

Recently, a double nicking system was reported to reduce off-target activity. In a manner similar to dimerized ZFNs and TALENs, when two gRNAs designed with appropriate offsets between target sequences are used in combination with a Cas9 nickase, the two separate complexes can successfully induce DSBs and mediate NHEJ or HR repair at higher frequencies than wild-type Cas9 [70,84] (Fig. 4C). As mentioned above, the Cas9 enzyme has two enzymatic domains, HNH and RuvC, which cleave the complementary and noncomplementary strands to the gRNA, respectively (Fig. 4B). By inducing mutations in the catalytic residues of either the RuvC (Asp10Ala, D10A) or HNH (His840Ala, H840A) domains, Cas9 becomes a nickase capable of cleaving only one strand of the target DNA (Fig. 4C). When two pairs of gRNAs that target opposite DNA strands are used, DSB can occur with considerably reduced off-target activity ranging from 50 to 1,500-fold depending on the cell line used [84]. Recently, Shen et al. demonstrated that both the D10A and H840A nickases, when used with two appropriately offset gRNAs, are also capable of reducing off-target effects *in vivo* [85].

While the new double nicking system does not completely eliminate all off-target effects of the CRISPR system, it is regarded as a promising genome editing tool with demonstrated widespread applications from gene manipulation in mouse embryos to systemic and genome wide screening in human cell lines [86–88]. As CRISPRs show high targeting efficiency using both HR and NHEJ in ESC and iPSCs [89], it is anticipated that soon CRISPRs will also be utilized for modeling diseases.

Generation of Disease-Specific hPSCs with Genomic Editing Systems

Even though hPSCs are regarded as a powerful resource for cell replacement therapy, due to the limited feasibility of current cell replacement therapy options, disease modeling with iPSCs has recently become an especially attractive branch of research. To date, ZFNs have been the primary genomic editing

tool used to model diseases with hPSCs, in part due to their availability before other genomic editing tools were established. Today there are currently only a few disease modeling studies that take advantage of TALENs, and no reports using CRISPRs (Table 2). The first report of ZFN application in hESCs was in 2007, and demonstrated that ZFN-mediated DSBs increase site-specific integration of vector sequences in hPSCs [90]. The CCR5 locus was chosen as the target site because its homozygous null mutations are tolerated in human cells. A donor vector with CCR5 homology arms containing the selection markers GFP or puromycin successfully targeted the CCR5 locus when delivered together with the ZFN expression vector into hESC through IDLV [90]. This demonstrated that ZFN-mediated HR efficiency is higher than standard HR-mediated integration alone [90]. GFP-positive clones maintained their pluripotent phenotype and were capable of generating GFP-positive neural progenitors in culture [90]. While this report was the first to show the utility of ZFNs for gene targeting in PSC, they did not demonstrate that their results were not due to the existence of random insertions. Additionally, they found that NHEJ-mediated mutations occurred in the CCR5 locus even if the donor vector was transferred [90]. These limitations raised by this article have been discussed in many subsequent articles using ZFNs or TALENs in hPSCs, which in turn has led to further advancement of ZFN and TALEN application methods.

Integration Methods of Genomic Editing Tools in hPSCs: Safe Harbor Locus

Most studies using genomic editing tools to alter gene function in hPSCs use one of three gene targeting methods: integration into a safe harbor locus, HR-mediated direct correction, and induction of gene mutations. Integration approaches that utilize a safe harbor locus refer to a method that targets a specific genomic site where integration of a transgene would allow for both active transgene transcription, and no downstream consequences caused by the disruption of the endogenous locus [91]. The AAVS1 locus (also known as PPP1R12C) on chromosome 19 in humans has long been used as a safe harbor region for gene therapy [92,93]. The AAVS1 locus has an open chromatin structure so that integrating transgenes can stably integrate into the AAVS1 site and show robust expression [94]. Taking advantage of the AAVS1 safe harbor locus to generate isogenic hPSCs for modeling diseases has been achieved and has been utilized most when mutations leading to disease are due to large deletions or other large genomic alterations. Zou et al. in 2011 used ZFN-mediated HR to insert corrected gp91phox (NOX2) into the AAVS1 locus to recover the enzyme deficiency for X-linked chronic granulomatous disease [95]. Another article using ZFNs to model α -thalassemia also utilized the AAVS1 integration method to compensate for large deletion sites found in endogenous α -globin [96].

However, there are significant concerns regarding the therapeutic use of ZFNs at safe harbor loci. iPSC could acquire genetic and epigenetic alterations from genomic editing through off-target integration, or generation of nicks in the DNA backbone [95]. Zou et al. showed that 25% of clones contained random insertions, and 50% contained NHEJ-mediated mutations. Second, safe harbor loci can be affected by integrated sequences. Sequences that integrate within close proximity to

TABLE 2. EXAMPLES OF HUMAN DISEASE-SPECIFIC INDUCED PLURIPOTENT STEM CELLS APPLIED TO GENOMIC EDITING TOOLS

	Disease	Target gene (mutation type)	Targeting method	Pathological phenotype recovered in isogenic control	Reference
Genomic editing tool ZFN	X-linked chronic granulomatous disease	<i>gp91phox</i> (deletion)	AAVS1 integration	Gp91 phox expression	95
	α -thalassemia	<i>α-globin</i> (deletion)	AAVS1 integration	Low α -globin expression ^a	103
	Sickle cell anemia	<i>β-globin</i> (missense)	HR-mediated correction	Abnormal β -globin expression ^a	105,106
	Paroxysmal nocturnal hemoglobinuria	<i>PIG-A</i>	HR-mediated Mutagenesis	Polymers of mutant AAT proteins	106
	Cancer	<i>Translocation</i>	HR-mediated Mutagenesis	Not determined	116
TALENs	Parkinson's disease	<i>Synuclein</i> (missense)	HR correction	The accumulation of reactive oxygen species/reactive nitrogen species ^b	121,122
	Tauopathy	<i>MAPT gene</i> (missense)		Predispose TAU proteolyzed by caspase, which leads to axon degradation and short neurite ^b	119
	HIV	<i>C-C chemokine receptor type 5</i>	HR-mediated KO	Inhibition of HIV entry	123
	β -thalassemia	<i>β-globin</i> (missense)	HR-mediated correction	Abnormal β -globin expression ^a	102
	Sickle cell anemia	<i>β-globin</i> (missense)	HR-mediated correction	Abnormal β -globin expression ^a	103
Genomic editing tool without nuclease	Cancer	<i>Translocation</i>	HR-mediated Mutagenesis	Not determined	109
	Lesch Nyhan-Syndrom	<i>HPRT1</i>	NHEJ-mediated KO	Decreased neurite length	110
	Hepatitis C virus infection	<i>SORT1</i>	Overexpression	Increased SORT1 decreased apoB ^b and resulted in low cholesterol ^b	120
	α -1 antitrypsin (AAT) deficiency,	<i>α-1 antitrypsin</i> (AAT)	HR-mediated correction	AAT accumulation	124
Bacterial artificial chromosome Adeno-associated virus	Genetic instability syndromes	<i>p53 and ATM</i>	HR-mediated KO		37
	Down syndrome	Chromosome 21	TK-neo mediated spontaneous chromosome loss	Improved cell growth and more endothelial cells in vivo	38

^aPathological phenotype is not clear due to lack of differentiation protocol.

^bThe mechanism of pathological phenotypes was demonstrated using corrected patient iPSCs. apoB, apolipoprotein B; HPRT1, hypoxanthine phosphoribosyltransferase; HR, homologous recombination; iPSCs, induced pluripotent stem cells; HIV, human immunodeficiency virus; KO, knockout; NHEJ, nonhomologous end joining; TALENs, transcription activator-like effector nucleases.

the AAVS1 gene locus are able to activate inappropriate AAVS1 expression in clones upon erythroid differentiation [96].

Correcting Mutations for Targeting Monogenic Diseases

The most prevalent disorders targeted by gene therapy and disease-modeling studies are thalassemia and sickle cell anemia (SCD) as these are well-characterized monogenic disorders [97]. Because these diseases are caused by missense mutations, HR-mediated gene correction is used to target the mutation. SCD is caused by a missense mutation in codon 6 of β -globin (HBB) causing a glutamine to valine change in the amino acid sequence, and two groups have reported the genetic correction of this mutation in human iPSCs (hiPSCs) through HR-based direct correction. Zou et al. showed homozygous correction of mutant β -globin alleles by a ZFN in conjunction with HR using vectors designed to integrate into intron 1 of the HBB gene, and demonstrated expression of wild-type β -globin in red blood cells derived from the corrected patient iPSCs [98]. A second group followed a similar strategy for modeling SCD, correcting the mutation with a cassette containing a drug resistance marker that allowed for easy selection of corrected clones [99]. While the second report found no evidence of off-target effects, Zou et al. raised the possibility that selection cassettes could affect corrected gene expression. Indeed, when the selection cassette was excised via Cre-mediated recombination, the expression of the corrected allele was diminished or lost during *in vitro* differentiation [98]. They reasoned that residual LoxP sequences left after Cre-mediated selection marker excision resulted in splicing defects, interfering with corrected HBB gene expression. Also, it has since been hypothesized that Cre-mediated excision of selection cassettes may also affect cis regulatory elements by mutating GATA binding sites found in the 3' enhancer of the HBB gene, leaving questions and concerns for the use of HR-mediated ZFN repair [98–101].

TALENs have also been applied to correct mutations for modeling SCD and thalassemia. β -Thalassemia iPSCs and their TALEN-corrected controls were used to recapitulate β -Thalassemia *in vitro* to demonstrate that restored β -globin expression reverses the phenotype [102]. Additionally, Sun and Zhao used TALENs to generate isogenic iPSC for modeling SCD instead of ZFNs to avoid off-target effects and potential genomic instability caused by the low specificity of ZFNs [103]. To avoid residual genomic scars left behind from removed selection markers, the authors used PiggyBack transposons as they do not leave and residual sequences after excision, unlike loxP sites [103]. The use of PiggyBack transposons resulted in 60% of clones with corrected mutations, with only two heterozygous clones out of 48 corrected clones [103]. Furthermore, no off-target effects, translocations, or alterations were found, confirming the specificity of TALENs and the advantage of PiggyBack vectors to target sites of interest [103–105].

Generating Mutations in Wild-Type hPSC with Genomic Editing Systems

Genomic editing tools have been utilized not only for correcting mutations, but also for inducing mutations or

deletions in hPSCs. In 2009, Zou et al. directly targeted the endogenous PIG-A locus, frequently found mutated in hematopoietic stem cells from patients with paroxysmal nocturnal hemoglobinuria. Using ZFNs and induced HR-mediated mutagenic insertion, they found a 200-fold increased targeting efficiency and a decreased random insertion frequency than would be found by HR alone [106]. Additionally, ZFNs have also been reported for the use of inducing genomic translocations. Brunet et al. used ZFN and TALENs in hESCs to generate genomic translocations, which are often associated with many cancers [107–109]. While modeling of a specific cancer with chromosomal translocations was not performed, characteristics of translocations in human tumors such as deletions, insertions, and microhomology at t(19;X), t(11;22)(q24;q12), and t(2;5)(p23;q35) translocations were recapitulated, opening up the possibility of modeling cancer associated translocations with hESCs [108,109].

Most recently, Greber and coworkers used TALENs in hESCs to model Lesch Nyhan-Syndrom (LNS). To do so, TALENs were used to knock out the HPRT1 gene through NHEJ, as mutations in this region are known to cause the disease [110]. The targeting efficiency of HPRT1 mutant clones was found at 15% with the use of two selection markers. The use of the two antibiotic selection markers was critical, as targeted mutation rates without the additional selective pressure were low at 0.5%. As a result, out of five clones generated, three were homozygous and two were heterozygous for the induced mutation. Finally, when the HPRT1 mutant cell lines were differentiated into RN3A+ central nervous system neurons, the knock-out cell lines successfully recapitulated phenotypes of LNS, including impaired neuronal differentiation and short neurite length. This demonstrated the usefulness of NHEJ-mediated KO system with TALENs in hPSCs for human disease modeling and also argued that conventional lipofection delivery methods are capable of generating sufficient numbers of clones [110].

As discussed above, there are some important concerns regarding the use of genomic editing tools. These include unwanted mutations in the genome due to off-target effect, the impact of selection markers on the expression of neighboring genes, or the effect of residual LoxP sites left after Cre-mediated excision. Thus, it is extremely important to ensure that there are no lesions or residues left behind by genomic editing tools, which can only be ensured through deep sequencing or global genomic analysis. However, it is nonetheless clear that ZFNs and TALENs significantly increase targeting efficiency either through the use of safe harbor loci, HR correction, or knock-out methods. Recent reports show that perfect isogenic cell lines can be successfully engineered, and they do not contain unexpected genomic mutations [99,102,103,107]. This implies that as ZFNs and TALENs continue to be optimized for use in hPSCs, perfect isogenic cell lines will be generated more efficiently and with greater ease, allowing for better models of genetic diseases.

Modeling Complex Diseases with Isogenic hPSCs

Unlike modeling early onset and monogenic diseases that manifest robust pathological phenotypes, when complex diseases such as PD are modeled with patient and wild-type iPSCs, the patient iPSCs tend to show subtle phenotypic

changes or unpredictable background effects [111–113]. This is a major challenge that stem cell technology needs to overcome for the comprehensive study of complex diseases. One conceivable solution is to generate perfect isogenic wild-type iPSCs directly from patient iPSCs. Genomic editing is inevitably required for modeling these diseases with iPSC technology, because, fundamentally, to create a disease in a dish an underlying mutation should be the only difference between disease and control cells. Indeed, this perspective has caused many to model complex diseases with isogenic iPSCs in combination with ZFNs or TALENs, with recent reports demonstrating mechanisms of pathogenesis or elucidating novel therapeutic targets, highlighting the unique therapeutic potential of iPSCs.

Previously, Soldner et al. applied ZFN technology to target a missense mutation in α -synuclein, a protein frequently mutated in familial PD [114]. They corrected A53T mutations in patient iPSCs and induced A53T or E46T mutations in wild-type iPSCs to model PD. Although these lines could successfully differentiate into dopaminergic tyrosine hydroxylase expressing neurons, they did not address a pathological phenotype at the time [114]. A recent report modeling PD with the same isogenic iPSCs was not only able to recapitulate a pathological phenotype, but also linked interactions between genetic mutations to environmental factors known to contribute to PD [115]. They found that the cells with the A53T mutation were more susceptible to toxin-mediated apoptosis than control cells. Furthermore, it was found that this sensitivity to cell death in A53T was mediated by the MEF2-PGC1 α pathway [115]. They noted the accumulation of reactive oxygen species/reactive nitrogen species caused by A53T mutant-mediated protein aggregation, mitochondrial toxins, or both, resulting in the sulfonated nitrosylation of MEF2C that causes neurons to undergo apoptosis [115]. Importantly, they used hiPSCs and isogenic iPSCs for high throughput screening studies to resolve the pathway, highlighting the feasibility of using iPSCs for modeling multi-hit diseases [115].

Recently, modeling complex diseases using iPSCs in conjunction with genomic editing tools has successfully elucidated new mechanisms of disease irresolvable by mouse modeling or human autopsy samples alone [116–118]. When modeling Tauopathy, a neurodegenerative disorder leading to Alzheimer's disease and frontotemporal dementia, isogenic TAU cell lines created by ZFNs elucidated a new pathological phenomena where the mutant microtubule-associated protein tau protein is predisposed to caspase cleavage, leading to axon degradation and shortened neurites [119].

Additionally, using TALEN-generated isogenic iPSCs, Ding et al. clarified data from conflicting reports about the effect of SORT1 on low-density lipoprotein and apolipoprotein B (ApoB) secretion from the liver, key components required for hepatitis C virus (HCV) infection [120]. While the SORT1 gene has been identified as a regulator of ApoB in human genetic studies, additional studies in mice and humans provide conflicting evidence as to whether SORT1 functions to increase or decrease ApoB expression [121,122]. Ding et al. demonstrated with isogenic lines that increased SORT1 expression resulted in decreased ApoB and lower cholesterol levels, which is consistent with human genomic studies, suggesting that modification of SORT1 may be a tractable

therapy for HCV infection, and a promising tool to study the underlying mechanisms in a human context.

Conclusions

Since the initial generation of hiPSCs in 2007, the derivation of patient-specific iPSCs for modeling diseases has become more frequent and prevalent. Current studies have taken this a step further, and have focused on demonstrating the feasibility of genetic correction using patient-derived iPSCs. Isogenic cells generated by ZFNs, TALENs, and CRISPRs will undoubtedly help to overcome variability found between iPSC clones and the intrinsic genetic background of patients. However, to generate perfect isogenic iPSCs, many factors need to be considered including selection markers, donor vectors, culture conditions, gene correction strategy, off-target effects, and genomic integrity. In addition, it is of note that ZFNs and TALENs each have their own strengths and limitations that should be considered (Table 1). Both are capable of producing off-target effects, and both show reduced efficiency when targeting methylated genomic loci. Furthermore, even though many isogenic iPSCs have been derived with ZFNs and TALENs, using isogenic patient-specific iPSCs generated by genomic editing tools to study molecular and pathological mechanism is still in the early stages, and continued evaluation of its feasibility is needed.

CRISPRs are the newest genetic editing tool for modeling diseases. Easy to construct and insensitive to DNA methylation, with increasing targeting efficiency in human cells, CRISPRs are rapidly gaining popularity for gene editing in human cells. Although off-target rates of CRISPRs cannot be disregarded, and their use has not been reported in hPSCs for disease modeling purposes to date, with the introduction of the new CRISPR double nicking system, CRISPRs may prove to be an effective genomic editing tool for modeling diseases with hiPSCs. Nevertheless, as time progresses and the technologies advance, ZFNs and TALENs, in addition to CRISPRs, will all aid in improving hPSC-based studies that will hopefully help to both elucidate biological mechanisms and provide novel clinical therapies.

Acknowledgments

This study was funded by grants from the National Institutes of Health (NIH) to I.R.L. (R01GM078465) and NIH Pathway to Independence Award (K99CA181496) to D.F.L.

Author Disclosure Statement

The authors declare they have no competing financial interests.

References

1. Thomson JA, J Itskovitz-Eldor, SS Shapiro, MA Waknitz, JJ Swiergiel, VS Marshall and JM Jones. (1998). Embryonic stem cell lines derived from human blastocysts. *Science* 282:1145–1147.
2. Cheloufi S and K Hochedlinger. (2012). Reproductive biology: stem cells bear eggs. *Nature* 491:535–536.
3. Weintraub H, VJ Dwarki, I Verma, R Davis, S Hollenberg, L Snider, A Lassar and SJ Tapscott. (1991). Muscle-specific transcriptional activation by MyoD. *Genes Dev* 5:1377–1386.

4. Takahashi K and S Yamanaka. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126:663–676.
5. Takahashi K, K Tanabe, M Ohnuki, M Narita, T Ichisaka, K Tomoda and S Yamanaka. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131:861–872.
6. Su RJ, DJ Baylink, A Neises, JB Kiroyan, X Meng, KJ Payne, B Tschudy-Seney, Y Duan, N Appleby, et al. (2013). Efficient generation of integration-free iPSCs from human adult peripheral blood using BCL-XL together with Yamanaka factors. *PLoS One* 8:e64496.
7. Luo M, T Ling, W Xie, H Sun, Y Zhou, Q Zhu, M Shen, L Zong, G Lyu, et al. (2013). NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells. *Stem Cells* 31:1278–1286.
8. Habib O, G Habib, HW Choi, KS Hong, JT Do, SH Moon and HM Chung. (2013). An improved method for the derivation of high quality iPSCs in the absence of c-Myc. *Exp Cell Res* 319:3190–3200.
9. Chen G, DR Gulbranson, Z Hou, JM Bolin, V Ruotti, MD Probasco, K Smuga-Otto, SE Howden, NR Diol, et al. (2011). Chemically defined conditions for human iPSC derivation and culture. *Nat Methods* 8:424–429.
10. Karumbayaram S, BG Novitch, M Patterson, JA Umbach, L Richter, A Lindgren, AE Conway, AT Clark, SA Goldman, et al. (2009). Directed differentiation of human-induced pluripotent stem cells generates active motor neurons. *Stem Cells* 27:806–811.
11. Ye L, S Zhang, L Greder, J Dutton, SA Keirstead, M Lepley, L Zhang, D Kaufman and J Zhang. (2013). Effective cardiac myocyte differentiation of human induced pluripotent stem cells requires VEGF. *PLoS One* 8: e53764.
12. Kumano K, S Arai and M Kurokawa. (2013). Generation of iPS cells from normal and malignant hematopoietic cells. *Stem Cells* 31:2015–2023.
13. Chou BK, Z Ye and L Cheng. (2013). Generation and homing of iPSC-derived hematopoietic cells in vivo. *Mol Ther* 21:1292–1293.
14. Carvajal-Vergara X, A Sevilla, SL D'Souza, YS Ang, C Schaniel, DF Lee, L Yang, AD Kaplan, ED Adler, et al. (2010). Patient-specific induced pluripotent stem-cell-derived models of LEOPARD syndrome. *Nature* 465:7299.
15. Lee G, EP Papapetrou, H Kim, SM Chambers, MJ Tomishima, CA Fasano, YM Ganat, J Menon, F Shimizu, et al. (2009). Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. *Nature* 461:402–406.
16. Park I-H, N Arora, H Huo, N Maherali, T Ahfeldt, A Shimamura, MW Lensch, C Cowan, K Hochedlinger and GQ Daley. (2008). Disease-specific induced pluripotent stem cells. *Cell* 134:877–886.
17. Onder TT and GQ Daley. (2012). New lessons learned from disease modeling with induced pluripotent stem cells. *Curr Opin Genet Dev* 22:500–508.
18. Moretti A, M Bellin, A Welling, CB Jung, JT Lam, L Bott-Flügel, T Dorn, A Goedel, C Höhnke, et al. (2010). Patient-specific induced pluripotent stem-cell models for long-QT syndrome. *N Engl J Med* 363:1397–1409.
19. Itzhaki I, L Maizels, I Huber, L Zwi-Dantsis, O Caspi, A Winterstern, O Feldman, A Gepstein, G Arbel, et al. (2011). Modelling the long QT syndrome with induced pluripotent stem cells. *Nature* 471:225–229.
20. Ebert AD, J Yu, FF Rose, VB Mattis, CL Lorson, JA Thomson and CN Svendsen. (2009). Induced pluripotent stem cells from a spinal muscular atrophy patient. *Nature* 457:277–280.
21. Kim K, A Doi, B Wen, K Ng, R Zhao, P Cahan, J Kim, MJ Aryee, H Ji, et al. (2010). Epigenetic memory in induced pluripotent stem cells. *Nature* 467:285–290.
22. Gore A, Z Li, HL Fung, JE Young, S Agarwal, J Antosiewicz-Bourget, I Canto, A Giorgetti, MA Israel, et al. (2011). Somatic coding mutations in human induced pluripotent stem cells. *Nature* 471:63–67.
23. Hussein SM, NN Batada, S Vuoristo, RW Ching, R Autio, E Närvä, S Ng, M Sourour, R Härmäläinen, et al. (2011). Copy number variation and selection during reprogramming to pluripotency. *Nature* 471:58–62.
24. Müller LU, MD Milsom, CE Harris, R Vyas, KM Brumme, K Parmar, LA Moreau, A Schambach, IH Park, et al. Daley, David A. Williams. (2012). Overcoming reprogramming resistance of Fanconi anemia cells. *Blood* 119:5449–5457.
25. Raya A, I Rodríguez-Pizà, G Guenechea, R Vassena, S Navarro, MJ Barrero, A Consiglio, M Castellà, P Río, et al. (2009). Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. *Nature* 460:53–59.
26. Israel MA, SH Yuan, C Bardy, SM Reyna, Y Mu, C Herrera, MP Hefferan, S Van Gorp, KL Nazor, et al. (2012). Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. *Nature* 482:216–220.
27. Soldner F, D Hockemeyer, C Beard, Q Gao, GW Bell, EG Cook, G Hargus, A Blak, O Cooper, et al. (2009). Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell* 136:964–977.
28. Kondo T, M Asai, K Tsukita, Y Kutoku, Y Ohsawa, Y Sunada, K Imamura, N Egawa, N Yahata, et al. (2013). Modeling Alzheimer's disease with iPSCs reveals stress phenotypes associated with intracellular A β and differential drug responsiveness. *Cell Stem Cell* 12:487–496.
29. Zwaka T and J Thomson. (2003). Homologous recombination in human embryonic stem cells. *Nat Biotechnol* 21: 319–321.
30. Xue H, S Wu, ST Papadeas, S Spusta, AM Swistowska, CC MacArthur, MP Mattson, NJ Maragakis, MR Capocchi, et al. (2009). A targeted neuroglial reporter line generated by homologous recombination in human embryonic stem cells. *Stem Cells* 27:1836–1846.
31. Du ZW, BY Hu, M Ayala, B Sauer and SC Zhang. (2009). Cre recombination-mediated cassette exchange for building versatile transgenic human embryonic stem cells lines. *Stem Cells* 27:1032–1041.
32. Hockemeyer D, F Soldner, C Beard, Q Gao, M Mitalipova, RC DeKever, GE Katibah, R Amora, EA Boydston, et al. (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol* 27:851–857.
33. Hockemeyer D, H Wang, S Kiani, CS Lai, Q Gao, JP Cassidy, GJ Cost, L Zhang, Y Santiago, et al. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* 29:731–734.
34. Pan H, W Zhang, W Zhang and GH Liu. (2012). Find and replace: editing human genome in pluripotent stem cells. *Protein Cell* 2:950–956.
35. Collin J and M Lako. (2011). Concise review: putting a finger on stem cell biology: zinc finger nuclease-driven

- targeted genetic editing in human pluripotent stem cells. *Stem Cells* 29:1021–1033.
36. Wang X, G Lin, K Martins-Taylor, H Zeng and RH Xu. (2009). Inhibition of caspase-mediated anoikis is critical for basic fibroblast growth factor-sustained culture of human pluripotent stem cells. *J Biol Chem* 284:34054–34064.
37. Song H, SK Chung and Y Xu. (2010). Modeling disease in human ESCs using an efficient BAC-based homologous recombination system. *Cell Stem Cell* 6:80–89.
38. Li LB, KH Chang, PR Wang, RK Hirata, T Papayannopoulou and DW Russell. (2012). Trisomy correction in Down syndrome induced pluripotent stem cells. *Cell Stem Cell* 11:615–619.
39. Hockemeyer D and R Jaenisch. (2010). Gene targeting in human pluripotent cells. *Cold Spring Harb Symp Quant Biol* 75:201–209.
40. Gropp M, P Itsykson, O Singer, HT Ben, E Reinhartz, E Galun and BE Reubinoff. (2003). Stable genetic modification of human embryonic stem cells by lentiviral vectors. *Mol Ther* 7:281–287.
41. Zaehres H, MW Lensch, L Daheron, SA Stewart, EJ Itskovitz and GQ Daley. (2005). High-efficiency RNA interference in human embryonic stem cells. *Stem Cells* 23:299–305.
42. Xiong C, DQ Tang, CQ Xie, L Zhang, KF Xu, WE Thompson, W Chou, GH Gibbons, LJ Chang, LJ Yang and YE Chen. (2005). Genetic engineering of human embryonic stem cells with lentiviral vectors. *Stem Cells Dev* 14:367–377.
43. Wei C, J Liu, Z Yu, B Zhang, G Gao and R Jiao. (2013). TALEN or Cas9-rapid, efficient and specific choices for genome modifications. *J Genet Genomics* 40:281–289.
44. Lieber MR, Y Ma, U Pannicke and K Schwarz. (2003). Mechanism and regulation of human non-homologous DNA end-joining. *Nat Rev Mol Cell Biol* 4:712–720.
45. Santiago Y, E Chan, PQ Liu, S Orlando, L Zhang, FD Urnov, MC Holmes, D Guschin, A Waite, et al. (2008). Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc Natl Acad Sci U S A* 105:5809–5814.
46. Lee HJ, E Kim and JS Kim. (2010). Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res* 20:81–89.
47. Lee HJ, J Kweon, E Kim, S Kim and JS Kim. (2012). Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome Res* 22:539–548.
48. San Filippo J, P Sung and H Klein. (2008). Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem* 77:229–257.
49. Ma S, X Wang, Y Liu, J Gao, S Zhang, R Shi, J Chang, P Zhao and Q Xia. (2014). Multiplex genomic structure variation mediated by TALEN and ssODN. *BMC Genomics* 15:41.
50. Yang L, M Guell, S Byrne, JL Yang, A De Los Angeles, P Mali, J Aach, C Kim-Kiselak, AW Briggs, et al. (2013). Optimization of scarless human stem cell genome editing. *Nucleic Acids Res* 41:9049–9061.
51. Papaioannou I, JP Simons and JS Owen. (2012). Oligonucleotide-directed gene-editing technology: mechanisms and future prospects. *Expert Opin Biol Ther* 12:329–342.
52. Gaj T, CA Gersbach and CF Barbas III. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* 31:397–405.
53. Schierling B, N Dannemann, L Gabsalilow, W Wende, T Cathomen and A Pingoud. (2012). A novel zinc-finger nuclease platform with a sequence-specific cleavage module. *Nucleic Acids Res* 40:2623–2638.
54. Wilson JH. (2003). Pointing fingers at the limiting step in gene targeting. *Nat Biotechnol* 21:759–760.
55. Segal DJ, JW Crotty, MS Bhakta, CF Barbas and NC Horton. (2006). Structure of Aart, a designed six-finger zinc finger peptide, bound to DNA. *J Mol Biol* 363:405–421.
56. Liu X, Y Wang, W Guo, B Chang, J Liu, Z Guo, F Quan and Y Zhang. (2013). Zinc-finger nickase-mediated insertion of the lysostaphin gene into the beta-casein locus in cloned cows. *Nat Commun* 4:2565.
57. Baltes NJ, J Gil-Humanes, T Cermak, PA Atkins and DF Voytas. (2014). DNA Replicons for plant genome engineering. *Plant Cell* 26:151–163.
58. Wang L, K Xu, J Lin, S Shao, T Zhang, H Xu, Z Wei and Z Zhang. (2013). A novel genetic system based on zinc finger nucleases for the identification of interactions between proteins in vivo. *PLoS One* 8:e85650.
59. Hofer U, JE Henley, CM Exline, O Mulhern, E Lopez and PM Cannon. (2013). Pre-clinical modeling of CCR5 knockout in human hematopoietic stem cells by zinc finger nucleases using humanized mice. *J Infect Dis* 208 Suppl 2:S160–S164.
60. Perez EE, J Wang, JC Miller, Y Jouvenot, KA Kim, O Liu, N Wang, G Lee, VV Bartsevich, et al. (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol* 26:808–816.
61. Isalan M. (2011). Zinc-finger nucleases: how to play two good hands. *Nat Methods* 9:32–34.
62. Urnov FD, EJ Rebar, MC Holmes, HS Zhang and PD Gregory. (2010). Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* 11:636–646.
63. Gonzalez B, LJ Schwimmer, RP Fuller, Y Ye, L Asawa-pornmongkol and CF Barbas 3rd. (2010). Modular system for the construction of zinc-finger libraries and proteins. *Nat Protoc* 5:791–810.
64. Kay S, S Hahn, E Marois, G Hause and U Bonas. (2007). A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* 318:648–651.
65. Bogdanove AJ and DF Voytas. (2011). TAL effectors: customizable proteins for DNA targeting. *Science* 333:1843–1846.
66. Deng D, C Yan, X Pan, M Mahfouz, J Wang, JK Zhu, Y Shi and N Yan. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 335:720–723.
67. Wiedenheft B, SH Sternberg and JA Doudna. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482:331–338.
68. Garneau JE, MÈ Dupuis, M Villion, DA Romero, R Barrangou, P Boyaval, C Fremaux, P Horvath, AH Magadán and S Moineau. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468:67–71.
69. Gasiunas G, R Barrangou, P Horvath and V Siksnys. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* 109:E2579–E2586.
70. Jinek M, K Chylinski, I Fonfara, M Hauer, JA Doudna and E Charpentier. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–821.

71. Semenova E, MM Jore, KA Datsenko, A Semenova, ER Westra, B Wanner, J van der Oost, SJ Brouns and K Severinov. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* 108:10098–10103.
72. Pattanayak V, CL Ramirez, JK Joung and DR Liu. (2011). Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods* 8:765–770.
73. Maeder ML, S Thibodeau-Beganny, A Osiak, DA Wright, RM Anthony, M Eichinger, T Jiang, JE Foley, RJ Winfrey, et al. (2008). Rapid “open-source” engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell* 31:294–301.
74. de Groote ML, PJ Verschure and MG Rots. (2012). Epigenetic editing: targeted rewriting of epigenetic marks to modulate expression of selected target genes. *Nucleic Acids Res* 40:10596–10613.
75. Mussolino C, R Morbitzer, F Lütge, N Dannemann, T Lahaye and T Cathomen. (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* 39:9283–9293.
76. Chen S, G Oikonomou, CN Chiu, BJ Niles, J Liu, DA Lee, I Antoshechkin and DA Prober. (2013). A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly. *Nucleic Acids Res* 41:2769–2778.
77. Bultmann S, R Morbitzer, CS Schmidt, K Thanisch, F Spada, J Elsaesser, T Lahaye and H Leonhardt. (2012). Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res* 40:5368–5377.
78. Valton J, A Dupuy, F Daboussi, S Thomas, A Maréchal, R Macmaster, K Melliand, A Juillerat and P Duchateau. (2012). Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. *J Biol Chem* 287:38427–38432.
79. Hsu PD, DA Scott, JA Weinstein, FA Ran, S Konermann, V Agarwala, Y Li, EJ Fine, X Wu, et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31:827–832.
80. Cong L, FA Ran, D Cox, S Lin, R Barretto, N Habib, PD Hsu, X Wu, W Jiang, LA Marraffini and F Zhang. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339:819–823.
81. Burgess DJ. (2013). Technology: a CRISPR genome-editing tool. *Nat Rev Genet* 14:80.
82. Fu Y, JA Foden, C Khayter, ML Maeder, D Reyon, JK Joung and JD Sander. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 31:822–826.
83. Cho SW, S Kim, Y Kim, J Kweon, HS Kim, S Bae and JS Kim. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* 24:132–141.
84. Ran FA, PD Hsu, CY Lin, JS Gootenberg, S Konermann, AE Trevino, DA Scott, A Inoue, S Matoba, Y Zhang and F Zhang. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154:1380–1389.
85. Shen B, W Zhang, J Zhang, J Zhou, J Wang, L Chen, L Wang, A Hodgkins, V Iyer, X Huang and WC Skarnes. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods* 11:399–402.
86. Wang H, H Yang, CS Shivalila, MM Dawlaty, AW Cheng, F Zhang and R Jaenisch. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153:910–918.
87. Shalem O, NE Sanjana, E Hartenian, X Shi, DA Scott, TS Mikkelsen, D Heckl, BL Ebert, DE Root, JG Doench and F Zhang. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343:84–87.
88. Wang T, JJ Wei, DM Sabatini and ES Lander. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343:80–84.
89. Ding Q, SN Regan, Y Xia, LA Oostrom, CA Cowan and K Musunuru. (2013). Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell* 12:393–394.
90. Lombardo A, P Genovese, CM Beausejour, et al. (2007). Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol* 25:1298–1306.
91. DeKolver RC, VM Choi, EA Moehle, DE Paschon, D Hockemeyer, SH Meijnsing, Y Sancak, X Cui, EJ Steine, et al. (2010). Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Res* 20:1133–1142.
92. Ogata T, T Kozuka and T Kanda. (2003). Identification of an insulator in AAVS1, a preferred region for integration of adeno-associated virus DNA. *J Virol* 77:9000–9007.
93. Henckaerts E, N Dutheil, N Zeltner, S Kattman, E Kohlbrenner, P Ward, N Clément, P Rebollo, M Kennedy, GM Keller and RM Linden. (2009). Site-specific integration of adeno-associated virus involves partial duplication of the target locus. *Proc Natl Acad Sci U S A* 106:7571–7576.
94. Lombardo A, D Cesana, P Genovese, B Di Stefano, E Provati, DF Colombo, M Neri, Z Magnani, A Cantore, et al. (2011). Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat Methods* 8:861–869.
95. Zou J, CL Sweeney, BK Chou, U Choi, J Pan, H Wang, SN Dowey, L Cheng and HL Malech. (2011). Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease-mediated safe harbor targeting. *Blood* 117:5561–5572.
96. Chang CJ and EE Bouhassira. (2012). Zinc-finger nuclease-mediated correction of α -thalassemia in iPS cells. *Blood* 120:3906–3914.
97. Persons DA and AW Nienhuis. (2000). Gene therapy for the hemoglobin disorders: past, present, and future. *Proc Natl Acad Sci U S A* 97:5022–5024.
98. Zou J, P Mali, X Huang, SN Dowey and L Cheng. (2011). Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. *Blood* 118:4599–4608.
99. Sebastiano V, ML Maeder, JF Angstman, B Haddad, C Khayter, DT Yeo, MJ Goodwin, JS Hawkins, CL Ramirez, et al. (2011). In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. *Stem Cells* 29:1717–1726.
100. Liu Q, J Bungert and JD Engel. (1997). Mutation of gene-proximal regulatory elements disrupts human epsilon-, gamma-, and beta-globin expression in yeast artificial

- chromosome transgenic mice. *Proc Natl Acad Sci U S A* 94:169–174.
101. Rubin JE, P Pasceri, X Wu, P Leboulch and J Ellis. (2000). Locus control region activity by 5'HS3 requires a functional interaction with beta-globin gene regulatory elements: expression of novel beta/gamma-globin hybrid transgenes. *Blood* 95:3242–3249.
 102. Ma N, B Liao, H Zhang, L Wang, Y Shan, Y Xue, K Huang, S Chen, X Zhou, et al. (2013). Transcription activator-like effector nuclease (TALEN)-mediated gene correction in integration-free β -thalassemia induced pluripotent stem cells. *J Biol Chem* 288:34671–34679.
 103. Sun N and H Zhao. (2013). Seamless correction of the sickle cell disease mutation of the HBB gene in human induced pluripotent stem cells using TALENs. *Biotechnol Bioeng* 111:1048–1053.
 104. Owens JB, D Mauro, I Stoytchev, MS Bhakta, MS Kim, DJ Segal and S Moisyadi. (2013). Transcription activator like effector (TALE)-directed piggyBac transposition in human cells. *Nucleic Acids Res* 41:9197–9207.
 105. Kettlun C, DL Galvan, AL George Jr., A Kaja and MH Wilson. (2011). Manipulating piggyBac transposon chromosomal integration site selection in human cells. *Mol Ther* 19:1636–1644.
 106. Zou J, ML Maeder, P Mali, SM Pruett-Miller, S Thibodeau-Beganny, BK Chou, G Chen, Z Ye, IH Park, et al. (2009). Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* 5:97–110.
 107. Weinstock DM, B Elliott and M Jasin. (2006). A model of oncogenic rearrangements: differences between chromosomal translocation mechanisms and simple double-strand break repair. *Blood* 107:777–780.
 108. Brunet E, D Simsek, M Tomishima, R DeKolver, VM Choi, P Gregory, F Urnov, DM Weinstock and M Jasin. (2009). Chromosomal translocations induced at specified loci in human stem cells. *Proc Natl Acad Sci U S A* 106:10620–10625.
 109. Piganeau M, H Ghezraoui, A De Cian, L Guittat, M Tomishima, L Perrouault, O René, GE Katibah, L Zhang, et al. (2013). Cancer translocations in human cells induced by zinc finger and TALE nucleases. *Genome Res* 23:1182–1193.
 110. Frank S, BV Skryabin and B Greber. (2013). A modified TALEN-based system for robust generation of knock-out human pluripotent stem cell lines and disease models. *BMC Genomics* 14:773.
 111. Khandekar N, KH Lie, PS Sachdev and KS Sidhu. (2012). Amyloid precursor proteins, neural differentiation of pluripotent stem cells and its relevance to Alzheimer's disease. *Stem Cells Dev* 21:997–1006.
 112. Young JE and LS Goldstein. (2012). Alzheimer's disease in a dish: promises and challenges of human stem cell models. *Hum Mol Genet* 21:82–89.
 113. Jung YW, E Hysolli, KY Kim, Y Tanaka and IH Park. (2012). Human induced pluripotent stem cells and neurodegenerative disease: prospects for novel therapies. *Curr Opin Neurol* 25:125–130.
 114. Soldner F, J Laganière, AW Cheng, D Hockemeyer, Q Gao, R Alagappan, V Khurana, LI Golbe, RH Myers, et al. (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell* 146:318–331.
 115. Ryan SD, N Dolatabadi, SF Chan, X Zhang, MW Akhtar, J Parker, F Soldner, CR Sunico, S Nagar, et al. (2013). Isogenic human iPSC Parkinson's model shows nitrosative stress-induced dysfunction in MEF2-PGC1 α transcription. *Cell* 155:1351–1364.
 116. Ashe KH and KR Zahs. (2010). Probing the biology of Alzheimer's disease in mice. *Neuron* 66:631–645.
 117. Huang Y and L Mucke. (2012). Alzheimer mechanisms and therapeutic strategies. *Cell* 148:1204–1222.
 118. Ballatore C, VM Lee and JQ Trojanowski. (2007). Tau mediated neurodegeneration in Alzheimer's disease and related disorders. *Nat Rev Neurosci* 8:663–672.
 119. Fong H, C Wang, J Knoferle, D Walker, ME Balestra, LM Tong, L Leung, KL Ring, WW Seeley, et al. (2013). Genetic correction of tauopathy phenotypes in neurons derived from human induced pluripotent stem cells. *Stem Cell Reports* 1:226–234.
 120. Ding Q, YK Lee, EA Schaefer, DT Peters, A Veres, K Kim, N Kuperwasser, DL Motola, TB Meissner, et al. (2013). A TALEN genome-editing system for generating human stem cell-based disease models. *Cell Stem Cell* 12:238–251.
 121. Musunuru K, A Strong, M Frank-Kamenetsky, NE Lee, T Ahfeldt, KV Sachs, X Li, H Li, N Kuperwasser, et al. (2010). From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* 466:714–719.
 122. Kjolby M, OM Andersen, T Breiderhoff, AW Fjorback, KM Pedersen, P Madsen, P Jansen, J Heeren, TE Willnow and A Nykjaer. (2010). Sort1, encoded by the cardiovascular risk locus 1p13.3, is a regulator of hepatic lipoprotein export. *Cell Metab* 12:213–223.
 123. Yao Y, B Nashun, T Zhou, L Qin, L Qin, S Zhao, J Xu, MA Esteban and X Chen. (2012). Generation of CD34+ cells from CCR5-disrupted human embryonic and induced pluripotent stem cells. *Hum Gene Ther* 23:238–242.
 124. Choi SM, Y Kim, JS Shim, JT Park, RH Wang, SD Leach, JO Liu, C Deng, Z Ye and YY Jang. (2013). Efficient drug screening and gene correction for treating liver disease using patient-specific stem cells. *Hepatology* 57:2458–2468.

Address correspondence to:

Ihor R. Lemischka

Department of Developmental and Regenerative Biology

The Black Family Stem Cell Institute

Icahn School of Medicine at Mount Sinai

New York, NY 10029

E-mail: ihor.lemischka@mssm.edu

Received for publication April 2, 2014

Accepted after revision July 30, 2014

Prepublished on Liebert Instant Online July 30, 2014