HHS Public Access

Author manuscript

Methods Enzymol. Author manuscript; available in PMC 2015 April 24.

Published in final edited form as:

Methods Enzymol. 2014 ; 546: 119–138. doi:10.1016/B978-0-12-801185-0.00006-4.

GENOME EDITING IN HUMAN STEM CELLS

Susan M. Byrne, **Prashant Mali**, and **George M. Church**

Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA

Abstract

The use of custom-engineered sequence-specific nucleases (including CRISPR/Cas9, ZFN, and TALEN) allows genetic changes in human cells to be easily made with much greater efficiency and precision than before. Engineered double-stranded DNA breaks can efficiently disrupt genes, or, with the right donor vector, engineer point mutations and gene insertions. However, a number of design considerations should be taken into account to ensure maximum gene targeting efficiency and specificity. This is especially true when engineering human embryonic stem (ES) or induced pluripotent stem cells (iPSCs), which are more difficult to transfect and less resilient to DNA damage than immortalized tumor cell lines. Here, we describe a protocol for easily engineering genetic changes in human iPSCs, through which we typically achieve targeting efficiencies between 1–10% without any subsequent selection steps. Since this protocol only uses the simple transient transfection of plasmids and/or single-stranded oligonucleotides, most labs will easily be able to perform it. We also describe strategies for identifying, cloning, and genotyping successfully edited cells, and how to design the optimal sgRNA target sites and donor vectors. Finally, we discuss alternative methods for gene editing including viral delivery vectors, Cas9 nickases, and orthogonal Cas9 systems.

Keywords

CRISPR; Cas9 nuclease; induced pluripotent stem cells; human genome engineering; gene targeting; transfection

Introduction

The development of sequence-specific nucleases such as zinc finger nucleases, (ZFN), transcription activator-like effector nucleases (TALEN), or CRISPR/Cas9 nucleases have enormously expanded our ability to engineer genetic changes in human cells (Joung & Sander, 2012; Mali,Yang, et al., 2013c; Urnov, Rebar, Holmes, Zhang, & Gregory, 2010). These nucleases can be custom-engineered to create double-stranded DNA (dsDNA) breaks at a desired sequence in the genome. When these are repaired using the non-homologous end joining (NHEJ) pathway, small insertion and deletion mutations (indels) are produced and disrupt genes. Alternatively, the dsDNA break can be repaired by the homologous recombination pathway – specific base pair changes or gene insertions can be formed using a homologous donor targeting vector. Of these systems, Cas9 nucleases have been favored due to their easy construction and lower toxicity in human cells (Ding et al., 2013).

Human induced pluripotent stem cells (iPSCs) have been another great breakthrough for genetic studies in human cells. Their self-renewing capability allows them to be gene-

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

targeted, cloned, genotyped, and expanded. Successfully targeted iPSC clones can then be differentiated into a variety of other cell types to analyze the effects of the induced mutations. The ability to easily genetically modify human iPSCs also holds tremendous clinical promise for generating artificial organs and safer gene therapies. However, while immortalized human tumor cell lines have been edited with almost complete efficiency (Fu, Sander, Reyon, Cascio, & Joung, 2014), much lower success rates have been achieved in human iPSCs (Mali, Yang, et al., 2013c; L. Yang et al., 2013). This difference may be due to the gross chromosomal abnormalities and an unusually robust response to DNA damage in tumor cell lines. In this article, we describe strategies to maximize the efficiency of genome editing in human iPSCs. Using these design considerations and the transient transfection protocol listed below, we typically achieve gene disruption frequencies of 1–25% and homologous gene targeting frequencies of 0.5–10% in human iPSCs without any subsequent selection steps.

Gene targeting strategies

For any gene targeting project, the structure of the gene must be considered and the nuclease targeting sites should be carefully chosen according to the experimental goals. For simple gene disruption, a single cut site can generate indel mutations using the NHEJ repair pathway. When within coding exons, such indels can cause frameshifts and disrupt protein expression. Targeting coding exons towards the beginning of the gene may be preferable, as mutations here may create more complete gene disruption and be less likely to accidentally generate truncated protein artefacts with residual biological activity. Areas possessing relatively unique genome sequences should be chosen, rather than a common domain shared by several homologous members of the same gene family (unless the goal is to target multiple members of the gene family).

Alternatively, one can design two nuclease sites to excise the intervening section of the genome. Regions from 100 bp to several kb can be excised with biallelic frequencies of over 10% (Cong et al., 2013). These junctions are often re-ligated with perfect precision between the two dsDNA break sites, although indels are also sometimes found. This strategy allows nuclease sites within introns or outside genes to be used; this is particularly useful when no satisfactory nuclease sites can be found within an exon. Again, the organization of the gene must be carefully considered to avoid alternative exon splicing events or truncated products.

When specific mutations are desired, a donor-targeting vector for homologous recombination is provided along with the nuclease elements. These donors can be single-stranded DNA oligonucleotides (ssODNs) or plasmids for engineering point mutations. Here, the nuclease site should be chosen as close to the intended mutation as possible, since homologous recombination targeting efficiencies drop precipitously as the dsDNA break becomes farther from the mutation. For ssODN donors, having the desired mutation in the center of the oligo showed the highest targeting efficiency. 90 bp ssODNs worked best, although lengths from 70 – 130 bp were able to produce targeting efficiencies >1%. The highest targeting frequencies occurred when the mutation was within 10 bp of the nuclease site; when the mutation was more than 40 bp away, gene targeting was barely detectable (Chen et al., 2011; L. Yang et al., 2013).

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Alternatively, plasmid targeting vectors for homologous recombination can be used to generate desired point mutations, as well as larger "knock-in" gene insertions. Since the presence of a dsDNA break drastically increases the homologous recombination efficiency, shorter homology arms of 0.4–0.8 kb can be used (rather than the 2–14 kb arms used in conventional gene targeting vectors without nucleases), although increased homology may still improve targeting of difficult constructs (Beumer, Trautman, Mukherjee, & Carroll, 2013; Hendel et al., 2014; Hockemeyer et al., 2009; Orlando et al., 2010). Again, the dsDNA break must be positioned within 200 bp of the mutation, and gene targeting efficiency decreases with larger transgene insertions (Guye, Li, Wroblewska, Duportet, & Weiss, 2013; Moehle et al., 2007; Urnov et al., 2010).

Choice of nuclease targeting sites

The *S. pyogenes* Cas9 nuclease (SpCas9) targets a 20 bp dsDNA sequence specified by the single guide RNA (sgRNA) next to a 3′ protospacer adjacent motif (PAM) of NGG, although PAM sequences of NAG can be targeted as well (Jinek et al., 2012; Mali, Aach, et al., 2013a; Mali, Yang, et al., 2013c). Upon binding to the sgRNA and complementary DNA targeting site, the Cas9 nuclease generates a blunt-ended, dsDNA break three base pairs upstream of the PAM. Cas9-sgRNA complexes can potentially tolerate 1–6 bp mismatches between the sgRNA and the target sequence, creating off-target cuts in genomic DNA. Although a "seed" sequence of the 8–13 nucleotides closest to the PAM appears to be more important for Cas9 nuclease specificity, mismatches can sometimes be tolerated here as well (Jinek et al., 2012; Mali, Aach, et al., 2013a). Off-target Cas9 nuclease activity can also occur with small indel mismatches (Lin et al., 2014).

Several online tools and algorithms are available to identify specific nuclease targeting sites, including: the CRISPR Design Tool (crispr.mit.edu) (Hsu et al., 2013); ZiFiT targeter (zifit.partners.org/ZiFiT) (Fu et al., 2014); CasFinder (arep.med.harvard.edu/CasFinder/) (Aach, Mali, & Church, 2014); and E-Crisp (www.e-crisp.org/E-CRISP/) (Heigwer, Kerr, & Boutros, 2014). In addition, specific Cas9 sgRNA targets for disrupting human exons can be found from published sets of sgRNA screening libraries (Aach et al., 2014; Shalem et al., 2014; Wang, Wei, Sabatini, & Lander, 2014). These algorithms are constantly being refined to incorporate further discoveries about Cas9 targeting specificity.

The nuclease activity among different sgRNAs can also vary widely. Cas9 nuclease activity is positively correlated with areas of open chromatin (Kuscu, Arslan, Singh, Thorpe, & Adli, 2014; L. Yang et al., 2013); however, substantial variations in activity can still be found among neighboring sgRNAs in the same locus. Other characteristics associated with higher levels of sgRNA activity are: targeting sequences with between 20–80% GC content, sgRNAs targeting the non-transcribed strand, and purines in the last four bases of the spacer sequence (Wang et al., 2014). While these criteria were statistically significant, they still could not account for all of the observed variation in sgRNA activity.

Initial constructs used the human U6 polymerase III promoter to express the sgRNA due to its specific initiation and termination sites and its ubiquitous expression in human cells. Since the U6 promoter requires a G to initiate transcription, this led to the early restriction

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

that only sequences fitting the form $GN_{20}GG$ could be targeted (Mali, Yang, et al., 2013c). However, subsequent studies showed that up to 10 extra nucleotides could be added to the 5′ end of the sgRNA while retaining similar levels of nuclease activity, and that these sgRNA extensions were being processed off (Mali, Aach, et al., 2013a; Ran, Hsu, Lin, et al., 2013a). Thus, any 20 bp sequence next to a PAM can be targeted, although an extra G is still required in the sgRNA expression construct to initiate transcription when the U6 promoter is used. Truncated sgRNAs with up to 3 base pairs missing from the 5′ end have been shown to increase specificity without much loss in activity, although truncations beyond 3 bp ablated activity (Fu et al., 2014). Appending up to 40 extra bp at the 3′ end of the sgRNA construct, after the hairpin backbone, resulted in slightly higher sgRNA activity, possibly due to increased half-life of the longer sgRNA (Mali, Aach, et al., 2013a). Other promoters besides U6, such as H1 or pol-II, may also be used to express the sgRNA. The sgRNA constructs may also be transfected into cells as linear PCR products rather than plasmids (Ran, Hsu, Wright, et al., 2013b).

Due to the ease of cloning sgRNAs, and the ongoing questions regarding sgRNA specificity and activity, we recommend that users select a few sgRNA target sites and test them empirically. While it is important to try to select sgRNAs that are as specific as possible, a perfectly unique sequence may not exist suitably close to your desired mutation. Alternative approaches are further discussed below.

Experimental procedures

With this protocol, we can consistently introduce plasmid DNA into human iPSCs with 60–70% transfection efficiency. While we have also had success using ZFNs and TALENs to edit iPSC genomes, the ease of cloning sgRNAs has made CRISPR/Cas the preferred method in our lab. Without any selection scheme, our overall gene disruption efficiencies using a single sgRNA in human iPSCs ranges from 1–25%, depending on the particular sgRNA used. Once the plasmids and cells are ready, the nucleofection process takes a few hours. After nucleofection, it takes 5–10 days of culture for the transient Cas9 transfection to subside and protein expression to turn over. Then, the potentially edited iPSCs can be cloned by single-cell FACS sorting. Eight days after sorting, individual iPSC have formed stable colonies, which can be further expanded and genotyped.

While this protocol focuses on human iPSC, it can be adapted for use in other cell types, using culture conditions and nucleofection protocols suitable for that cell type (although the amounts of plasmid/ssODN and promoters for Cas9 expression may need adjustment). Overall gene disruption efficiencies greater than 60% have thus been achieved in immortalized tumor cell lines.

1. Human iPSC culture and passaging

A number of different human iPSC lines are available from cell line resources such as Coriell (coriell.org), ATCC (atcc.org), and the Harvard Stem Cell Institute (hsci.harvard.edu), among many others. Furthermore, numerous academic and commercial facilities offer iPSC derivation services. Detailed protocols for culturing and passaging human ES and iPSC lines are available elsewhere (wicell.org, stembook.org). Here, we have

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

used iPSC derived from open-consented participants in the Personal Genome Project (J.-H. Lee et al., 2009), but this protocol is widely applicable to any human ES or iPS cell line. Cells used for gene targeting should be of a low passage number and free of karyotypic abnormalities. Cells should exhibit normal iPSC morphology and express pluripotency markers such as Tra-1/60 and SSEA4.

Human iPSCs for genome engineering are cultured under feeder-free conditions, in the defined mTesr-1 medium (StemCell Technologies) on Matrigel-coated tissue culture plates (BD). We have found lower transfection efficiencies (40–60%) when transfecting iPSCs growing on irradiated mouse embryonic fibroblasts (MEF), due to incomplete separation of the iPSCs from the MEFs immediately before nucleofection.

2. Preparation of plasmids for transient transfection

An increasingly wide selection of plasmids for ZFN, TALEN, and CRISPR/Cas9 genome editing, with instructions for cloning, are available from the Addgene plasmid repository (www.addgene.org/CRISPR/). This protocol was specifically developed with the plasmids to express human-codon optimized SpCas9 and sgRNAs from (Mali, Yang, et al., 2013c). However, an EF1α promoter was used to express Cas9 instead of the CMV promoter in iPSCs, as it produced a 5-fold increase in gene disruption efficiency.

Plasmid donor vectors containing homology arms can be easily cloned using isothermal assembly or synthesized as gene fragments (Integrated DNA Technologies). Homology arm sequences should ideally be cloned from the cell line being targeted to obtain identical (isogenic) sequences. Any polymorphic differences between the targeting vector and the genomic locus can decrease gene targeting frequencies (Deyle, Li, Ren, & Russell, 2013). All plasmids for nucleofection into iPSCs should be endotoxin-free (Qiagen Endo-free Plasmid Maxi Kit) and at a concentration greater than 2 mg/ml, so as not to dilute the nucleofection buffer. Oligo donors (ssODN) should be HPLC-purified and resuspended in sterile distilled water.

3. Nucleofection protocol

This protocol uses the Amaxa 4D-Nucleofector X Unit (Lonza), but we have also gotten good transfection efficiencies in human iPSCs from the Neon Transfection system (Life Technologies). Traditional electroporation methods will produce lower transfection efficiencies in iPSCs, which will lower the overall gene targeting efficiency. The amounts listed below are for the 20 μl Nucleocuvette strips; if using the 100 μl single Nucleocuvettes, increase all quantities five-fold. A control reaction transfecting a fluorescent protein-expressing plasmid can be used to verify nucleofection efficiency.

Expand hiPS cells under feeder-free conditions in mTesr-1 medium on tissue culture plates coated with ES-qualified Matrigel (BD) according to the manufacturer's instructions. Each nucleofection reaction will need 0.5×10^6 cells, although a range of $0.2 – 2 \times 10^6$ iPSCs per reaction can be used. Depending on the number of reactions, 6-well plates or 10 cm dishes of cultured iPSC may be required.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Prepare Matrigel-coated 24-well tissue culture plates, one well per nucleofection reaction. Additional Matrigel-coated 96-well flat-bottom tissue culture plates may also be prepared to culture aliquots of transfected cells for analysis.

Pretreat hiPSC cultures with 10 μM ROCK inhibitor (Y-27632) (R&D Systems, EMD Millipore, or other source) in mTesr-1 for at least 30 minutes before nucleofection. Prepare additional mTesr with 10 μM ROCK inhibitor for use throughout the nucleofection procedure. Cells treated with ROCK inhibitor should display the characteristic change in morphology of colonies with jagged edges.

Combine Nucleofector solution P3 with supplement according to manufacturer's instructions (Lonza). For each nucleofection reaction, dilute and combine the DNA mixtures in Nucleofector solution P3 (with supplement) to a final volume of 10 μl. Each nucleofection should contain 0.5 μg of Cas9-expressing plasmid and 1–1.5 μg of sgRNA-expressing plasmids. (When multiple sgRNA-expressing plasmids are used, mix them in equal amounts for a total of 1–1.5 μg plasmid.) If a plasmid targeting vector is being used, include 2 μg per nucleofection reaction. If an ssODN donor is being used, include up to 200 pmol per nucleofection reaction. DNA stock solutions must be concentrated enough such that the total volume of DNA does not exceed 10% of the nucleofection reaction (2 μl for a 20 μl Nucleocuvette). DNA amounts exceeding 4 μg per nucleofection may have an adverse effect on iPSC viability.

Remove the mTesr with Rock inhibitor media from the cells and incubate with Accutase dissociating enzyme (EMD Millipore, StemCell Technologies, or other source) for 5–10 min. Once iPSCs have detached, add an equal volume of mTesr with Rock inhibitor, and pipet to achieve a single cell suspension. Centrifuge the cells at 110 × g for 3 min. at room temperature. Resuspend cell pellet in mTesr with Rock inhibitor and count live cells.

Centrifuge the required number of iPSCs at 110 × g for 3 min. Aspirate off the media. Resuspend cell pellet in 10 μl Nuclofector solution P3 (with supplement) for each reaction.

For each reaction, promptly combine 10 μl of DNA mixture with 10 μl of resuspended cells and transfer the whole 20 μl into a Nucleocuvette. Ensure that the sample is at the bottom of the cuvette.

Place Nucleocuvette into the Nucleofector device and run program CB-150.

Add 80 μl mTesr with Rock inhibitor medium into each nucleocuvette well and pipet once or twice to resuspend cells. Transfer each reaction into one well of a Matrigel-coated 24-well plate containing 1 ml warm mTesr with Rock inhibitor medium. Alternatively, the nucleofected cells may also be distributed among one 24-well and one or two 96-wells, if analysis at intermediate time points is desired. (If Matrigel-coated 96-well plates are used, an optional centrifugation step (70 × g, 3 min, room temperature) can help plate the cells.) A high plating density post nucleofection is important for cell survival.

24 hours post-nucleofection, iPSCs transfected with a fluorescent protein-expressing plasmid may be examined to assess the transfection efficiency. Change the media to

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

mTesr-1 without Rock inhibitor. Since the iPSCs were plated at a high density, they may appear confluent. As most of the Cas9-induced cell death occurs between 1 and 2 days post nucleofection, we advise waiting until 2 days post nucleofection to passage the iPSCs. Transfected iPSC can then be propagated using regular iPSC culture protocols. After 4 or 5 days post nucleofection, the transient transfection will have subsided, and the cell population can be assayed for gene editing efficiency.

4. Verification of successful cutting and gene targeting

As the isolation and genotyping of edited iPSC clones can be time consuming, laborious, and expensive, it is desirable to have intermediate ways to verify successful gene disruption and evaluate gene targeting efficiency. Examining a portion of the targeted cell population will help estimate how many clones should be genotyped and provide guidance for troubleshooting.

If the gene being disrupted or inserted is expressed by human iPSCs, the most straightforward assay is to check for expression of that protein by microscopy or flow cytometry. If the targeted gene is not expressed or lacks a convenient stain, then a control reaction using an sgRNA that does target an easily detectable expressed gene can be used to troubleshoot the overall protocol and vectors, although individual sgRNA activities may still vary widely.

If a gene segment is being inserted into the genome, a dilution PCR for the inserted segment can be done on genomic DNA from the edited cell population; however, care must be taken to ensure that the PCR reaction does not simply amplify residual amounts of the transfected donor fragment itself (De Semir & Aran, 2003). PCR primers designed to anneal to genomic DNA sequences outside of the targeted homology region may be used to ensure that only integrated segments are detected. Alternatively, a control targeting vector can be constructed with the same homology arms as the insertion targeting vector, except that a constitutively-expressed fluorescent protein cassette is being inserted into the genome. This may provide a quick estimate of “knock-in” insertion frequencies at that locus using the same sgRNA and homology arms.

To directly measure the extent of gene disruption at a particular locus, a mismatch-specific endonuclease assay – either T7 endonuclease I (New England Biolabs), or Cel-1 Surveyor nuclease (Transgenomic) – is commonly used (H. J. Kim, Lee, Kim, Cho, & Kim, 2009; Qiu et al., 2004). These assays involve PCR-amplifying a short region (roughly 500 bp) around the intended sgRNA targeting site from the genomic DNA of the population of potentially edited cells. These PCR products are melted and re-annealed. Any mutations at the intended nuclease site will form a mismatch in the dsDNA, which will be recognized and cleaved by the mismatch-specific endonuclease. Cleaved PCR products can then be analyzed and quantitated by gel electrophoresis. If a restriction enzyme site is inserted or removed at the intended sgRNA targeting site, a restriction fragment length polymorphism (RFLP) assay may also assess Cas9 nuclease activity; here, PCR products around the intended sgRNA site are digested with the restriction enzyme to generate cleaved fragments (Chen et al., 2011).

Author Manuscript

While the endonuclease assays offer a rapid and cheap measure of gene disruption activity, the endonuclease digestion reaction can be sensitive to buffer and incubation conditions and the limit of detection is around 1–3% of sequences. We prefer a next-generation sequencing-based assay that has a much lower limit of detection (< 0.1%) and provides additional sequence information about the edited sgRNA site (L. Yang et al., 2013). Here, a 100–200 bp region around the edited sgRNA targeting site is PCR amplified and sequenced on a MiSeq system (Illumina). The initial set of genome-specific PCR primers are designed with the requisite MiSeq adaptor sequences appended to the 5′ end. Then, a second round of nested PCR with standard index primers incorporates the barcodes (ScriptSeq from Epicentre or Nextera from Illumina). A detailed protocol with primer sequences has been published (L. Yang, Mali, Kim-Kiselak, & Church, 2014). While each MiSeq run (150 bp, paired-end) can be expensive, up to ~200 different samples can be barcoded, pooled, and sequenced in parallel to reduce costs (L. Yang et al., 2013). The resulting next-generation sequencing data can be analyzed by the online CRISPR Genome Analyzer platform, which accepts the sequencing reads, the genomic sequence being targeted, and a donor sequence for homologous recombination (if applicable), and calculates the rate of indels and successful homologous recombination (crispr-ga.net) (Guell, Yang, & Church, 2014).

5. Cloning by single cell FACS sorting

Several days post nucleofection, after the transiently transfected plasmids have been lost and the Cas9 nuclease activity has subsided, targeted iPSCs may be selected and cloned to generate a culture of successfully targeted cells. As was done for traditional gene targeting without nucleases, if a positive selection marker for antibiotic resistance has been integrated into the genome (such as those for neomycin, hygromycin, or puromycin), that antibiotic may be added to the culture to remove unrecombined antibiotic-sensitive cells. Emerging antibiotic-resistant stem cell clones can then be individually picked by hand and cultured.

Alternatively, human iPSCs may be cloned by FACS sorting individual cells into separate wells of a 96-well plate. To preserve the viability of the dissociated single iPSC, a cocktail of small molecule inhibitors (termed SMC4, from Biovision) is added to the culture (Valamehr et al., 2012). We find that the viability of isolated iPSCs is further enhanced by sorting the cells (previously cultured in feeder-free mTesr-1 media) onto a feeder layer of irradiated MEFs in human ES cell medium. Eight days after FACS sorting, colony formation should be apparent from the individually sorted iPSC, and the SMC4 inhibitors can be removed from the ES cell medium. Our detailed protocol for FACS-sorting targeted human iPSCs has been published (L. Yang et al., 2014). We usually achieve 20–60% iPSC survival and colony formation post-sort. The gene targeting efficiency in the iPSC population (measured as described in the previous section) can be used to estimate the number of wells needed for sorting to obtain a successfully targeted viable clone. The iPSC colonies may then be cultured and expanded as usual on a MEF feeder layer for a few passages before being transitioned to feeder-free iPSC conditions. A portion of each potentially targeted iPSC clone may be taken for genomic DNA extraction and genotyping.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

6. Genotyping of clones

Once potentially targeted iPSC clones have been expanded, they must then be genotyped to identify successful gene targeting. While the use of custom-engineered nucleases greatly increases the frequency of correctly targeted events, incorrect mutations still sometimes occur, including partial integrations, random integrations, homology arm duplications, and incorporation of plasmid backbone sequences. In addition, since the use of nucleases allows for potential targeting of both alleles, a genotyping scheme must be able to detect whether the targeted mutation is homozygous or heterozygous.

Typically, genomic DNA is purified from a portion of each expanded clone (while freezing or continuing to expand the remaining culture). For simple gene disruptions or small bp changes, PCR amplification and Sanger sequencing of the targeted locus would suffice. Heterozygous base pair changes will be apparent as a double peak on the Sanger sequencing trace. Heterozygous indels can similarly be identified through programs that deconvolute a biallelic Sanger sequencing trace (Mutation Surveyor by Softgenetics). Alternatively, the biallelic PCR product can be subcloned into a plasmid vector (TOPO from Life Technologies) for each allele to be sequenced in a separate reaction. Any potential off-target nuclease sites may also be genotyped in this manner to check for mutations.

To genotype larger gene deletions, a PCR reaction with primers that span the two nuclease sites can be sequenced. A second PCR reaction with primers located within the two nuclease sites can identify any un-excised alleles, and determine whether the gene deletion is homozygous or heterozygous.

For targeted “knock-in” gene insertions, one must not only ensure that the entire transgene has been incorporated into the genome, but also that both homology arms have been recombined to the correct site, without recombination into other areas or duplication of the homology arms. Southern blot screening has traditionally been used to determine this, using probes specific to the target locus outside of the homology arm regions. While non-radioactive Southern blot protocols now exist, this screening still requires a relatively large amount of genomic DNA, unique restriction enzyme patterns, and probes verified beforehand. A faster alternative is to use a series of PCR reactions to confirm complete and correct integration of the “knock-in” construct into the targeted locus. One set of PCR primers that spans the inserted gene can confirm complete insertion, while two other sets of primers that span each of the homology arms (with one primer annealing outside of the homology arm region) can confirm proper recombination on each end.

New screening techniques have been developed to genotype very long constructs or homology arms. Fluorescence in situ hybridization (FISH) can measure the copy number of long homology arms to distinguish between correct targeting events (where copy number is maintained) and random gene integration (where an extra copy of the homology arm is added) (Y. Yang & Seed, 2003). Single-molecule real-time (SMRT) DNA sequencing is capable of producing longer read lengths than Sanger sequencing, and has been used for genotyping nuclease-edited human cell lines with an average sequencing read length approaching 3 kb and ability to detect mutations down to 0.01% (Hendel et al., 2014).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

7. Verify iPSC pluripotency and quality

Once successfully targeted iPSC clones have been identified through genotyping, they should be examined to confirm that they have not lost pluripotency or gained chromosomal abnormalities through the process. These checks are standard practice for any iPSC culture, even without nuclease-mediated gene targeting, as there is always a background level of differentiation and chromosomal rearrangement (Martí et al., 2013). However, these checks are particularly important when gene-targeted clones have been derived from a single iPS cell. Many academic stem cell core facilities and commercial suppliers offer these iPSC characterization services.

First, human iPSC can be stained for the expression of pluripotency markers both extracellular (Tra-1/60 or Tra-1/81, SSEA4) and intracellular (Oct4, Nanog) by either immunohistochemistry or flow cytometry. Gene-targeted iPSCs should also retain normal colony morphology. Second, cells should be karyotyped to ensure a normal chromosome number and lack of aberrant translocations. Finally, a teratoma assay is performed where iPSCs are injected into immunocompromised mice. The resulting iPSC-derived teratomas are histologically examined for generation of all three germ layers (mesoderm, ectoderm, endoderm).

Alternative approaches

1. Low transfection

Using the above protocol, we typically achieve transfection efficiencies around 60–70% and gene targeting/disruption efficiencies around 1–25%. However, several strategies can be used to enrich for targeted clones in cases of low transfection or gene editing efficiencies. Positive selection markers and antibiotic selection schemes may still be used to select for rare gene insertions, although gene targeting frequencies around 1% are generally high enough for genotypic screening without the use of a positive or negative selection marker.

In case of low transfection efficiencies, the Cas9 nuclease and sgRNA construct can be expressed from the same plasmid, such that both components of the Crispr system are co-delivered into any transfected cells. Furthermore, several constructs have been developed to express multiple sgRNAs from the same plasmid rather than separate plasmids (Cong et al., 2013; Tsai et al., 2014).

Transiently transfected cells may be enriched using a fluorescent or antibiotic resistance selection marker that is either co-transfected or co-expressed with the Cas9 nuclease. Human iPSCs electroporated with a Cas9-T2A-GFP fusion protein can be FACS-sorted 24–48 hr post transfection (Ding et al., 2013). Cas9-T2A-Puromycin resistance constructs are also available, although the drug selection may need to be carefully optimized to match the period of transient Cas9 and resistance marker expression (Ran, Hsu, Wright, et al., 2013b). Alternatively, a reporter construct plasmid could be co-transfected that possesses the sgRNA targeting site upstream of a fluorescent protein such that Cas9 nuclease editing brings the fluorescent protein in frame for expression. Cells with active Cas9-sgRNA complexes can then be enriched by flow cytometry (Ramakrishna et al., 2014).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2. Viral vectors

Viral vectors have also been used as alternatives to transient transfection. Lentiviral vectors are commonly used to introduce Cas9 and sgRNA components into a wide variety of cell types, both dividing and non-dividing. As these retroviruses integrate the genetic construct into the chromosome, they are particularly useful when sustained nuclease activity is desired, including CRISPR library screens that require almost complete gene disruption efficiencies (Shalem et al., 2014; Wang et al., 2014). However, lentiviral vectors are limited by the size of insert that can practically be packaged into the capsid (roughly 7.5 kb) (Yacoub, Romanowska, Haritonova, & Foerster, 2007) and tend to recombine out repetitive DNA sequences (Holkers et al., 2012; L. Yang et al., 2013). Integrase-deficient lentiviral vectors (IDLV), that deliver gene constructs which do not integrate into the genome and are gradually lost through cell division, have also been used to deliver nucleases and homologous donor templates to edit human stem cells (Joglekar et al., 2013; Lombardo et al., 2007).

Adenoviral vectors have been a popular approach for administering gene therapies *in vivo* as well as gene editing human stem cells *in vitro* as they can transduce a wide variety of dividing and non-dividing cells. The helper-dependent (or high-capacity) adenoviral vectors, where the viral genes have been removed, can deliver constructs up to 37 kb (Gonçalves & de Vries, 2006). Their linear dsDNA genome is generally not integrated into the chromosome, but maintained episomally until lost through cell division. Adenoviral vectors were better than lentiviral vectors at introducing constructs with repetitive elements such as TALENs into cells (Holkers et al., 2012). The higher packaging capacity of adenoviral vectors also allowed the 4.1 kb *S. pyogenes* Cas9 nuclease gene to be efficiently delivered into human cells (Maggio et al., 2014).

Recombinant Adeno-Associated viruses (rAAV) have also been used for gene targeting a wide variety of cell types. Even without sequence-specific nucleases, they can induce higher rates of homologous recombination than a transfected plasmid (Russell & Hirata, 1998); however, like conventional plasmid targeting vectors, gene targeting with AAV vectors can be further enhanced by introducing a dsDNA break (Hirsch & Samulski, 2014). AAV vectors possess a single-stranded DNA genome with a packaging capacity limited to around 4.5 kb, although they may self-assemble or be directed to form concatamers, thereby producing longer constructs (Hirsch et al., 2013).

3. Off-targets

When targeting a particular human genomic locus, one can often not find a perfectly specific sgRNA site that possesses a unique 13 bp "seed" sequence not found alongside any other active PAM site. Even then, Cas9 nuclease activity may still occur at off-target sites containing 1–2 mismatches in the seed sequence. Of an initial set of 190,000 sgRNA sequences designed to target human exons, 99.96% were later computationally found to have off-target sites containing at least one mismatch in the seed sequence next to a NGG or NAG PAM (Mali, Aach, et al., 2013a).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

However, the frequency of off-target nuclease activity is a dose-dependent function of the on-target activity – one can titrate down the on-target Cas9 nuclease activity to remove the off-target activity (Fu et al., 2013). The initial reports studying off-target Cas9 nuclease activity were done on human immortalized tumor cell lines, which were able to disrupt genes with very high rates of on-target activity (40–80%) leading to substantial rates of off-target activity at certain sites (Fu et al., 2013; Hsu et al., 2013; Kuscu et al., 2014; Pattanayak et al., 2013). For cells with a lower rate of on-target nuclease activity, the off-target sites may be less of a concern. In mouse ES cells, a Cas9-sgRNA complex bound to hundreds of off-target sites, but nuclease activity was only found at one similar off-target site (Wu et al., 2014). With the above protocol transiently transfecting Cas9 nuclease and sgRNA into human iPSC, we typically find low off-target gene disruption frequencies around 0.1–0.2%, even with identical seed sequences. Robust tumor cell lines allow for very high on- and off-target gene disruption frequencies, whereas off-target mutations may be less of an issue for cells with a lower rate of on-target activity like human iPSCs.

Since neither the on-target nor off-target Cas9 nuclease activities can currently be completely predicted through computational analysis, we recommend that any close off-target sgRNA sites also be checked when assessing gene disruption frequency in the cell population and when genotyping successfully targeted clones. Off-target gene disruption frequencies can be lowered by titrating down Cas9 nuclease activity (by transfecting a smaller quantity of plasmid, or expressing the Cas9 nuclease under a weaker promoter), although this will also decrease the on-target nuclease activity.

4. Cas9 nickases

When further specificity is required, two Cas9 nickases may be used to generate the dsDNA break instead of a single Cas9 nuclease. In the nickase version of Cas9 (D10A), the RuvC endonuclease-like domain has been mutated, such that only a single-stranded DNA break is made on the complementary DNA strand (Jinek et al., 2012). (Gene targeting with an alternate Cas9 nickase where the HNH endonuclease-like domain has been mutated so only the non-complementary strand is cleaved (H840A) has been less well characterized.) Two sgRNAs designed to target opposite strands at the same locus can be combined to generate an offset dsDNA break. Any off-target single-stranded DNA nicks will be unlikely to be repaired by NHEJ and result in very low indel rates (Cong et al., 2013; Mali, Aach, et al., 2013a). A single-stranded DNA nick was sufficient to induce homologous recombination in a human tumor cell line, but not in a human ES cell line (Hsu et al., 2013). Offset dsDNA breaks generated by Cas9 nickases may be especially necessary for targeting genes that have many conserved family members, or for therapeutic applications that require more than a single accurately targeted cell clone.

For gene disruption using paired Cas9 nickases, the highest rate of indel formation was achieved using two offset sgRNAs where the double nicks resulted in a 5′ DNA overhang. Indel formation was greatest with a 20–50 bp 5′ overhang, although detectable up to 130 bp (Cho et al., 2013; Mali, Aach, et al., 2013a; Ran, Hsu, Lin, et al., 2013a). Genomic deletions could also be made with the Cas9 nickase and four sgRNAs that generated two offset dsDNA breaks. A 5′ DNA overhang produced by a double nick also showed a higher ratio

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of homologous recombination to NHEJ compared to a blunt dsDNA break, although the overall rate of homologous recombination was still higher with the Cas9 nuclease. Recently, further specificity has been achieved using catalytically-inactive nuclease-null Cas9 proteins fused to a FokI homodimer nuclease domain – a pair of sgRNAs can bring the attached FokI domains together at the target site to dimerize and generate a dsDNA break provided that they have the appropriate orientation (PAM sites facing outward) and spacing (14–25 bp apart depending on the fusion construct) (Guilinger, Thompson, & Liu, 2014; Tsai et al., 2014).

5. Orthogonal Cas9 systems

While the Cas9 nuclease from *S. pyogenes* has been most commonly used, Cas9 nucleases derived from other bacteria are also capable of editing human genomes. Cas9 nucleases from *S. thermophilus*, *N. meningitidis*, and *T. denticola* recognize different PAM sequences, thereby expanding the set of potentially targetable sgRNA sites (Aach et al., 2014; Esvelt et al., 2013; Hou et al., 2013). Specific unique sgRNA backbones have been developed for *S. pyogenes*, *S. thermophilus*, and *N. meningitidis*, which allow these three Cas9 systems to be used simultaneously in an orthogonal manner.

In addition to nuclease and nickase activity, the easily programmable DNA binding ability of Cas9 has been adapted for many other functions. A nuclease-null Cas9 can be used by itself to repress gene expression, or be fused to transcriptional activator domains, repressor domains, epigenetic regulators, or fluorescent proteins (Mali, Esvelt, & Church, 2013b).

The technology for genetic engineering has progressed rapidly in the past few years, and will certainly continue to improve. The ability to easily and efficiently edit human genomes using custom-engineered nucleases has already greatly expanded studies of gene function and holds great potential for constructing modified human iPSCs and safer gene therapies.

References

Aach J, Mali P, Church GM. CasFinder: Flexible algorithm for identifying specific Cas9 targets in genomes. bioRxiv. 201410.1101/005074

Beumer KJ, Trautman JK, Mukherjee K, Carroll D. Donor DNA Utilization during Gene Targeting with Zinc-finger Nucleases. G3 Genes|Genomes|Genetics. 2013; 3:657–664.

Chen F, Pruett-Miller SM, Huang Y, Gjoka M, Duda K, Taunton J, et al. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. Nature Methods. 2011; 8:753–755. [PubMed: 21765410]

Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, Kim JS. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Research. 201310.1101/gr.162339.113

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science. 2013; 339:819–823. [PubMed: 23287718]

De Semir D, Aran JM. Misleading Gene Conversion Frequencies Due to a PCR Artifact Using Small Fragment Homologous Replacement. Oligonucleotides. 2003; 13:261–269. [PubMed: 15000840]

Deyle DR, Li LB, Ren G, Russell DW. The effects of polymorphisms on human gene targeting. Nucleic Acids Research. 2013; 42:3119–3124. [PubMed: 24371280]

Ding Q, Regan SN, Xia Y, Oostrom LA, Cowan CA, Musunuru K. Enhanced Efficiency of Human Pluripotent Stem Cell Genome Editing through Replacing TALENs with CRISPRs. Stem Cell. 2013; 12:393–394.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nature Methods. 2013; 10:1116–1121. [PubMed: 24076762]

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature Biotechnology. 2013; 31:822–826.

Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nature Biotechnology. 2014; 32:279–284.

Gonçalves MAFV, de Vries AAF. Adenovirus: from foe to friend. Reviews in Medical Virology. 2006; 16:167–186. [PubMed: 16710837]

Guell M, Yang L, Church GM. Genome Editing Assessment using CRISPR Genome Analyzer (CRISPR-GA). Bioinformatics. 2014 submitted.

Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nature Biotechnology. 2014; 32:577–582.

Guye P, Li Y, Wroblewska L, Duportet X, Weiss R. Rapid, modular and reliable construction of complex mammalian gene circuits. Nucleic Acids Research. 2013; 41:e156. [PubMed: 23847100]

Heigwer F, Kerr G, Boutros M. E-CRISP: fast CRISPR target site identification. Nature Methods. 2014; 11:122–123. [PubMed: 24481216]

Hendel A, Kildebeck EJ, Fine EJ, Clark JT, Punjya N, Sebastiano V, et al. Quantifying genome-editing outcomes at endogenous loci with SMRT Sequencing. Cell Reports. 2014; 7:293–305. [PubMed: 24685129]

Hirsch ML, Samulski RJ. AAV-mediated gene editing via double-strand break repair. Methods in Molecular Biology. 2014; 1114:291–307. [PubMed: 24557911]

Hirsch ML, Li C, Bellon I, Yin C, Chavala S, Pryadkina M, et al. Oversized AAV Transductifon is mediated via a DNA-PKcs-independent, Rad51C-dependent repair pathway. Molecular Therapy. 2013; 21:2205–2216. [PubMed: 23939025]

Hockemeyer D, Soldner F, Beard C, Gao Q, Mitalipova M, DeKelver RC, et al. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nature Biotechnology. 2009; 27:851–857.

Holkers M, Maggio I, Liu J, Janssen JM, Miselli F, Mussolino C, et al. Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. Nucleic Acids Research. 2012; 41:e63. [PubMed: 23275534]

Hou Z, Zhang Y, Propson NE, Howden SE, Chu LF, Sontheimer EJ, Thomson JA. Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110:15644–15649. [PubMed: 23940360]

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nature Biotechnology. 2013; 31:827–832.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science. 2012; 337:816–821. [PubMed: 22745249]

Joglekar AV, Hollis RP, Kuftinec G, Senadheera S, Chan R, Kohn DB. Integrase-defective Lentiviral Vectors as a Delivery Platform for Targeted Modification of Adenosine Deaminase Locus. Molecular Therapy. 2013; 21:1705–1717. [PubMed: 23857176]

Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. Nature Reviews Molecular Cell Biology. 2012; 14:49–55.

Kim HJ, Lee HJ, Kim H, Cho SW, Kim JS. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. Genome Research. 2009; 19:1279–1288. [PubMed: 19470664]

Kuscu C, Arslan S, Singh R, Thorpe J, Adli M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nature Biotechnology. 201410.1038/nbt.2916

Lee JH, Park IH, Gao Y, Li JB, Li Z, Daley GQ, et al. A Robust Approach to Identifying Tissue-Specific Gene Expression Regulatory Variants Using Personalized Human Induced Pluripotent Stem Cells. PLoS Genetics. 2009; 5:e1000718. [PubMed: 19911041]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Lin Y, Cradick TJ, Brown MT, Deshmukh H, Ranjan P, Sarode N, et al. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. Nucleic Acids Research. 201410.1093/nar/gku402

Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, et al. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nature Biotechnology. 2007; 25:1298–1306.

Maggio I, Holkers M, Liu J, Janssen JM, Chen X, Gonçalves MAFV. Adenoviral vector delivery of RNA-guided CRISPR/Cas9 nuclease complexes induces targeted mutagenesis in a diverse array of human cells. Scientific Reports. 2014; 410.1038/srep05105

Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nature Biotechnology. 2013a; 31:833–838.

Mali P, Esvelt KM, Church GM. Cas9 as a versatile tool for engineering biology. Nature Methods. 2013b; 10:957–963. [PubMed: 24076990]

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013c; 339:823–826. [PubMed: 23287722]

Martí M, Mulero L, Pardo C, Morera C, Carrió M, Laricchia-Robbio L, et al. Characterization of pluripotent stem cells. Nature Protocols. 2013; 8:223–253.

Moehle EA, Rock JM, Lee YL, Jouvenot Y, DeKelver RC, Gregory PD, et al. Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104:3055–3060. [PubMed: 17360608]

Orlando SJ, Santiago Y, DeKelver RC, Freyvert Y, Boydston EA, Moehle EA, et al. Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. Nucleic Acids Research. 2010; 38:e152. [PubMed: 20530528]

Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. High-throughput profiling of off-target DNA cleavage reveals rNA-programmed Cas9 nuclease specificity. Nature Biotechnology. 2013; 31:839–843.

Qiu P, Shandilya H, D'Alessio JM, O'Connor K, Durocher J, Gerard GF. Mutation detection using Surveyor nuclease. BioTechniques. 2004; 36:702–707. [PubMed: 15088388]

Ramakrishna S, Cho SW, Kim S, Song M, Gopalappa R, Kim JS, Kim H. Surrogate reporter-based enrichment of cells containing RNA-guided Cas9 nuclease-induced mutations. Nature Communications. 2014; 5:3378.

Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, et al. Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Cell. 2013a; 154:1380–1389. [PubMed: 23992846]

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nature Protocols. 2013b; 8:2281–2308.

Russell DW, Hirata RK. Human gene targeting by viral vectors. Nature Genetics. 1998; 18:325–330. [PubMed: 9537413]

Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, et al. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Science. 2014; 343:84–87. [PubMed: 24336571]

Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nature Biotechnology. 2014; 32:569–576.

Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. Nature reviews Genetics. 2010; 11:636–646.

Valamehr B, Abujarour R, Robinson M, Le T, Robbins D, Shoemaker D, Flynn P. A novel platform to enable the high-throughput derivation and characterization of feeder-free human iPSCs. Scientific Reports. 2012; 2:213. [PubMed: 22355727]

Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. Science. 2014; 343:80–84. [PubMed: 24336569]

Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nature Biotechnology. 201410.1038/nbt.2889

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Yacoub NA, Romanowska M, Haritonova N, Foerster J. Optimized production and concentration of lentiviral vectors containing large inserts. The Journal of Gene Medicine. 2007; 9:579–584. [PubMed: 17533614]

Yang L, Guell M, Byrne S, Yang JL, De Los Angeles A, Mali P, et al. Optimization of scarless human stem cell genome editing. Nucleic Acids Research. 2013; 41:9049–9061. [PubMed: 23907390]

Yang L, Mali P, Kim-Kiselak C, Church G. CRISPR-Cas-Mediated Targeted Genome Editing in Human Cells. Methods in Molecular Biology. 2014; 1114:245–267. [PubMed: 24557908]

Yang Y, Seed B. Site-specific gene targeting in mouse embryonic stem cells with intact bacterial artificial chromosomes. Nature Biotechnology. 2003; 21:447–451.