Genome engineering of isogenic human ES cells to model autism disorders

Refugio A. Martinez[1],[†], Jason L. Stein[2],[†], Anne-Rachel F. Krostag[1], Angelique M. Nelson[1], John S. Marken[1], Vilas Menon[1], Ryan C. May[1], Zizhen Yao[1], Ajamete Kaykas[1], Daniel H. Geschwind[2],* and Joshua S. Grimley[1],*

[1]Allen Institute for Brain Science, Seattle, WA 98103, USA and [2]Neurogenetics Program, Department of Neurology, Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA

ABSTRACT

Isogenic pluripotent stem cells are critical tools for studying human neurological diseases by allowing one to study the effects of a mutation in a fixed genetic background. Of particular interest are the spectrum of autism disorders, some of which are monogenic such as Timothy syndrome (TS); others are multigenic such as the microdeletion and microduplication syndromes of the 16p11.2 chromosomal locus. Here, we report engineered human embryonic stem cell (hESC) lines for modeling these two disorders using locus-specific endonucleases to increase the efficiency of homology-directed repair (HDR). We developed a system to: (1) computationally identify unique transcription activator-like effector nuclease (TALEN) binding sites in the genome using a new software program, TALENSeek, (2) assemble the TALEN genes by combining golden gate cloning with modified constructs from the FLASH protocol, and (3) test the TALEN pairs in an amplification-based HDR assay that is more sensitive than the typical non-homologous end joining assay. We applied these methods to identify, construct, and test TALENs that were used with HDR donors in hESCs to generate an isogenic TS cell line in a scarless manner and to model the 16p11.2 copy number disorder without modifying genomic loci with high sequence similarity.

INTRODUCTION

Stem cell models provide a robust platform for generating the cell types of the body and understanding how they are altered in the context of disease. Human diseases can be modeled using induced pluripotent stem cells (iPSCs) derived from patients [1]. However, genetic variation across the rest of the patient's genome confounds the study of a specific mutation's effects. Engineering an isogenic mutation into healthy pluripotent stem cells allows the isolated study of that genotypic variation, holding constant any effects due to the underlying genetic background. Although the mutant gene can be introduced at a safe harbor locus if it is known to have a dominant negative phenotype [2], targeting the endogenous locus more similarly matches the genomic context of the disease [3]. The challenge of introducing a point mutation into the endogenous locus in a scarless manner has been accomplished by taking advantage of a nearby transposon-recognition site [4] or by introducing a single-stranded oligodeoxynucleotide (ssODN) donor for homology-directed repair (HDR) [3,5,6].

The genomes of pluripotent systems, such as human embryonic stem cells (hESCs) and iPSCs, are challenging to modify due to their low transfection efficiency, poor viability during transfection, and low single-cell cloning efficiency [7]. Introducing a site-specific double-strand break (DSB) using a targeted endonuclease [8–13] increases the rate of HDR by four orders of magnitude [14]. When combined with a nuclease, a donor with regions of homology flanking the cut site allows targeted editing of genomic DNA. Zinc finger nucleases, transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPR)-Cas enzymes have been used to introduce DSBs for HDR in hESCs and iPSCs [2–6]. As nuclease gene assembly is a small fraction of the time and effort required to create an engineered hES

*To whom correspondence should be addressed. Tel: +206 548 7189; Fax: +206 632 2610; Email: joshuag@alleninstitute.org
Correspondence may also be addressed to Daniel H. Geschwind. Tel: +310 794 6570; Fax: +310 267 2401; Email: dhg@mednet.ucla.edu
†These authors contributed equally to the paper as first authors.
Present Address: Ajamete Kaykas, Novartis Institute for Biomedical Research, Cambridge, MA 02139 USA.

cell line, the ease of CRISPR-Cas gene construction (12,13) relative to obligate dimeric TALENs (10,11) may not outweigh the potential off-target effects of monomeric nucleases (15–18).

Here, we present techniques for the design, assembly, and testing of site-specific TALENs for genes associated with autism spectrum disorders (ASDs; CACNA1C and KCTD13), neuronal development (FOXG1, PAX6, PVALB, SOX1, and TUBB3), and the Wnt/β-catenin signaling pathway (FOXM1 and TMEM88). We used pairs of TALENs to generate isogenic hESCs to model two known genetic causes of ASDs. The mutation of Gly406 to Arg (G406R) in the L-type voltage-sensitive calcium channel $Ca_v1.2$ (CACNA1C) is conserved in patients with Timothy syndrome (TS), a disorder that affects multiple organs including the brain, heart and digits, producing phenotypes that include autism and long QT (19). This mutation results in nearly complete loss of voltage-dependent channel activation, producing sustained inward Ca^{2+} currents. The G406R mutation was introduced in a scarless manner to match the local genetic environment of patients in a fixed hESC genetic background. KCTD13, the key gene in the 16p11.2 chromosomal locus where copy number variation (CNV) has been associated with autism-linked phenotypes (20–23), was disrupted in a manner that reduced gene expression.

We present techniques to generate isogenic engineered cell lines, including: (1) identifying unique TALEN-binding sites to loci of interest using TALENSeek software, (2) assembling the TALEN genes using the golden gate FLASH (ggFLASH) method, and (3) testing the TALEN pairs using a HDR-dependent single-stranded oligodeoxynucleotide donor assay (ssODA). The methods used here for generation of ASD models in hESCs can be applied for generation of isogenic disease models of any monogenic disorder.

MATERIALS AND METHODS

Identifying TALEN-binding sites

TALENSeek was generated as a script in the R programming language (available at http://geschwindlab.neurology. ucla.edu/protocols) calling the packages biomaRt (24) and BSgenome (Pages, H. BSgenome: Infrastructure for Biostrings-based genome data packages. R package version 1.30.0). The search for off-target binding was accomplished via inexact matching using the matchPDict() function of the Biostrings package (Pages, H., Aboyoun, P., Gentleman, R., DebRoy, S. Biostrings: String objects representing biological sequences, and matching algorithms. R package version 2.32.0). On a desktop computer, identifying TALEN-binding site with TALENSeek typically takes a few minutes per gene with the time dependent on the number of matching off-target sites identified and whether binding sites are set to match human only or mouse and human genomes simultaneously: approximately 12 min to identify one TALEN pair each at the start and end of a gene with off-target sites identified in the human genome when searching only the human genome and approximately 20 min per gene for a human and mouse genomic search without off-target site identification when searching for homologous mouse and human binding sites. In the worst case scenario,

where the program searches iteratively but does not find homologous sites in human and mouse genomes, the maximal run time is approximately eight hours. The identification of TALEN-binding sites is independent of the TALEN gene assembly; therefore, TALENSeek is compatible with gene assembly methods using the common TALEN framework and any set of repeat variable di-residues (RVDs; (10,25)).

TALEN gene assembly by ggFLASH

The golden gate FLASH assembly of TALEN genes used the FLASH library of tetramer and trimer subunits (26)—with Asn-Ile, His-Asp, Asn-Asn, and Asn-Gly as the RVDs for adenosine, cytidine, guanosine, and thymidine, respectively (10)—and a modified set of repeat array receiving vectors (AI-CN441, AI-CN442, AI-CN443, and AI-CN491; Addgene plasmids 62590, 62591, 62593, and 62594, respectively; Supplementary Methods) containing obligate homodimeric FokI nuclease domains for golden gate cloning (27–30). The selected target sites from TALENSeek were entered into FLASHplasmids, a complementary program written in Java (available at http:// geschwindlab.neurology.ucla.edu/protocols) that identified the necessary components and generated a GenBank formatted file for each TALEN. The ggFLASH method and the FLASHplasmids program can each be extended to include future subunit reagents containing alternate RVDs (25).

The FLASH repeat array subunits were polymerase chain reaction (PCR)-amplified with primers specific to the location of each subunit within the final array (Supplementary Table S1). Each set of four reactions per TALEN were combined and purified; approximately 1 μg of total purified products were digested overnight with 1 μL of BbsI (Thermo Scientific; FastDigest BbsI FD1014). The appropriate receiving vector was digested with BbsI, treated with Antarctic phosphatase (New England Biolabs), and stored at −20°C for future use.

For the golden gate reaction, the reagents were mixed on ice: 4 μL of purified PCR product pool at approximately 50 ng/μL, 1 μL of digested vector at 50 ng/μL, 5 μL of nuclease-free water, 1.5 μL of 10× G buffer (Thermo Scientific), 1.5 μL of 10 mM ATP, 1 μL of BbsI, and 1 μL of T4 ligase at 400 U/μL (New England Biolabs). The reaction conditions were: (1) 25 cycles of 37°C, 3 min and 16°C, 4 min; (2) 50°C, 5 min; and (3) 80°C, 5 min (31). Then, 2.5 μL of the reaction was transformed into chemically competent DH5α cells (New England Biolabs) and placed onto plates with carbenicillin, Xgal, and IPTG. The plasmids were validated by insert check PCR and sequencing (Supplementary Table S1).

TALEN activity by ssODA and CEL-I

For ssODA, K562 cells (0.5 or 1.0×10^6; ATCC CCL-243) were electroporated (Lonza Amaxa 4D instrument with SF cell line kit) with 4 μg of each TALEN plasmid (Supplementary Table S2) and 0.3 nmol of the 120 nt ssODN HDR donor (Supplementary Table S3); the reagent amounts were consistent in the TALEN pair or ssODN donor-only controls. After 48–72 h at 37°C, genomic DNA was harvested

for junction PCR (QuickExtract; Epicentre or Gentra Pure-gene Cell; Qiagen). Each ssODN contained ataacttcgtatag-catacattatacgaagttatagatct, encoding a loxP recognition site and a BglII restriction site, flanked by 40 nt of homology on each side of the predicted cut site (32).

For the CEL-I assay, K562 cells (1×10^6) were electroporated with 5 μg of each TALEN plasmid. The untransfected control is shown only when background CEL-I activity was observed. After 48–72 h, genomic DNA was harvested for PCR. For each target, the single band product was purified, denatured, re-annealed, and incubated with Surveyor Nuclease S (Transgenomic). The reactions were quantified by capillary electrophoresis (QIAxcel Advanced, high resolution kit, OM500 method; Qiagen).

Timothy syndrome mutation in hESCs

H1 hESCs (WA01; WiCell) were maintained in mTeSR1 (StemCell Technologies) on Matrigel (BD) unless otherwise noted. At p39, cells were treated with 1 μM thiazovivin (StemRD) one day before electroporation (Neon; Invitrogen; resuspension buffer R; 100 μL electroporation tip; 1050 V, 30 ms pulse width, 2 pulses; 1×10^6 cells) as a single-cell suspension (StemPro Accutase, Life Technologies) with 4 μg of each TALEN plasmid and 0.3 nmol of the ssODN. A donor-only control included 0.6 nmol of ssODN with 1×10^6 cells. The ssODN HDR donor sequence was TCATAGGGTCATTTTTTGTACTTAAC TTGGTTCTCGGTGTCTTGTCTaGGTAAGCAGGAC CAAGGAAAAAGGTCTTGATTTTTCCATT, where the point mutation is in lowercase and the underlined nucleotides introduced an insertion-specific primer binding site while retaining exon coding identity.

The cells were treated with 2 μM thiazovivin for 24 h following electroporation. A preliminary test by 5' junction PCR one week after electroporation confirmed the targeted integration of the mutation in the cells that received the TALENs and donor, but not the ssODN-only. At that time, cells were plated at 500 cells/well into each well of 6×96-well plates with 2 μM thiazovivin. Ten days later, the cells were treated with StemPro Accutase for 2 min at 37°C, one half of the cells was replica-plated onto new 96-well plates and genomic DNA was harvested (QuickExtract) from the remaining cells in each well.

In the first round of PCR screening, DNA was combined from each well in a single row. These 48 total samples were analyzed by quantitative PCR (qPCR; LightCycler 480; Roche) in a single 96-well PCR plate for the 5' and 3' junctions using SYBR green I Taq mix (Roche). The four most promising rows were identified as those with the greatest increase in fluorescence, the correct product T_m, and the presence of a single product band at the correct size band by electrophoresis for at least one junction reaction. In the second round, each well from the selected rows was tested individually; the three best wells were combined for single-cell cloning.

The resulting clones at p46, two months after transfection, were tested by junction PCR with primers for both the desired mutation and the wild-type background. The absence of a wild-type product band, but the presence of the mutation, indicated that the clone was homozy-gous for the desired mutation. The identity of the clone was confirmed by locus amplification, blunt cloning (Zero Blunt TOPO, Life Technologies), and sequencing the insert of six bacterial colonies; all six showed the desired mutation, confirming the identification of the homozygous CACNA1C$^{G406R/G406R}$ cell line. This line was named AI07e-Timothy (33) and has a normal karyotype (Cell Line Genetics, Supplementary Figure S1). Flow cytometry of pluripotency markers (BD Pharmingen: NANOG, 560483; OCT4, 560794; SOX2, 560294) demonstrated that the clone was similar to the unmodified H1 hESC (Supplementary Figure S2).

KCTD13 disruption in hESCs

H1 cells at p36 were treated with 1 μM thiazovivin one day prior to electroporation (3×10^6 cells; conditions as per TS hESC) as single cells (Accutase) with 0.5 μg of each TALEN plasmid and 2 μg of the HDR donor plasmid AI-CN590 (Addgene plasmid 62596; Supplementary Methods). Cells were plated into a single well of a six-well plate with CF-1 MEFs (GlobalStem), hES media (DMEMF12 with 20% KSR; Life Technologies), and 2 μM thiazovivin. After five days in hES media, cells were passaged onto DR4 MEFs (GlobalStem) and maintained for three days prior to a five day treatment with 1 μg/ml puromycin (Life Technologies). Individual colonies were picked within the following three days. At p43, five weeks after transfection, transgene integration was assayed by PCR (Supplementary Table S1). The heterozygous and homozygous clones, AI09e-KCTD13a and AI10e-KCTD13b respectively, have normal karyotypes (Cell Line Genetics, Supplementary Figure S3) and the pluripotency marker expression pattern of hESCs (Supplementary Figure S4). Following reverse transcription of RNA from modified and parental hESCs, the cDNA was assayed by qPCR with primers for KCTD13 (CTGCCGGAGAGTACGA-GAGA, GGACAGCGTCTCCCTTTTT) and the reference HSP90AB1 (CCTCACTAATGACTGGGAAGAC, CAACTCATCACAGCTGTCCATGA).

Off-target analyses in engineered hESCs

Human genomic sequences similar to the desired TALEN-binding sites used in generating the KCTD13 disruption and TS mutation cell lines were identified computationally (Supplementary Table S4). The search reported loci where two TALENs, either as heterodimeric or homodimeric pairs, beginning with thymine and spaced 12–20 bp apart might bind and modify the genome. As the Asn-Asn RVD was used for targeting guanosine, the presence of adenosine instead of guanosine at the same position in the binding site was not counted as a mismatch.

RESULTS

TALEN-binding sites identified

The first procedure in generating isogenic cell lines is the identification of unique TALEN-binding sites. Such sites must satisfy a set of rules influencing their binding potential and specificity (26). For generation of TALEN-binding

sites anywhere in the human genome (34–37), we wrote a software program in the R programming language called TALENSeek. TALENSeek finds TALEN pair binding sites near a genomic locus of interest or the 5′ and 3′ ends of a user-specified gene in the genome and also generates homology sequences used for HDR donor construction. The program inputs are the HUGO gene name of interest (www.genenames.org), the length of the TALEN-binding sites and the spacer separating them, the number of mismatches allowed when searching for off-target binding sites in the genome, and the desired length of homology surrounding the TALEN-induced DSB. In addition, the program allows for discovery of unique homologous TALEN-binding sites that are conserved in both human and mouse genomes.

TALENSeek first creates the super-isoform, the furthest 5′ and 3′ length across all isoforms defined in ENSEMBL, for the specified gene (Figure 1a). Near the transcriptional start and stop sites, an acceptable TALEN-binding site pair is sought where the pairs are separated by 15–30 bp and each 14–19 bp binding site begins with a T. The program searches iteratively into the gene until an acceptable site is found (Figure 1b and c). This site is then searched for uniqueness in the genome, allowing for a number of user-defined mismatches. If a homologous sequence in human and mouse is specified, the binding site identified in the human genome is tested to be within 250 bp of the start or end site of that same gene in the mouse genome (Figure 1d). The user has the option to test if the TALEN-binding site is unique in each genome. The output includes a .bed file for visualization in the UCSC genome browser (38) that identifies the individual TALEN-binding sites in the human and mouse genomes for each specified target (Figure 1e).

TALENSeek has some unique features in comparison to other TALEN-binding site detection algorithms currently available (Supplementary Table S5). Notably, TALENSeek allows for binding site detection near a gene of interest by simply entering the gene symbol rather than finding and copying each sequence of interest. In addition, TALENSeek iteratively searches for binding sites which do not have off-target sites (with the user inputted number of mismatches). Other programs with off-target search availability (34,36,39–41) all require first running a TALEN identification program prior to the separate off-target search. Lastly, TALENSeek offers the user the ability to design a TALEN that homologously targets a gene of interest in both human and mouse. This unique feature can allow the user to save both reagent costs and time by limiting the number of TALENs produced as well as likely having a more consistent binding specificity for cross-species experiments. TALENSeek does not implement an empirically determined binding affinity algorithm as previous algorithms (29) were not supported by subsequent literature (26). A recent binding site algorithm provides a large data set to empirically determine binding site scores (42). This is very encouraging and we would urge users to utilize both tools for their specific advantages.

ggFLASH assembly of TALEN genes

The assembly of a TALEN-coding gene is non-trivial as each TALEN typically contains 12–20 repeats of 34 amino acids that are conserved in 32 positions per repeat (10). Of the numerous methods to construct TALEN genes (10,26,28–30,37,43,44), including high-throughput methods that use liquid handling robotics (37,44), the FLASH assembly method (26) is attractive in its speed, robust assembly, and library of component plasmids. The library contains plasmids encoding up to four TALEN repeat array subunits, but the FLASH method of TALEN assembly requires microgram quantities of plasmids. The ggFLASH assembly method combines the FLASH plasmid library with the ligation specificity and small amounts of materials needed in golden gate cloning (Figure 2a; (27–30)).

Instead of large-scale preparations of the plasmid library members, the desired plasmids are PCR-amplified with primers specific to that subunit's position in the array. For example, targeting 18 bp requires assembly of 16.5 repeats using four PCR-amplified tetramers and a vector encoding the initial thymine recognition element and final half-repeat. One-pot golden gate cloning is most efficient when a single restriction enzyme can be used to digest all of the components. The four receiving vectors shared by the REAL and FLASH methods use BsmBI restriction sites for insertion of the repeat array (43); however, this site is present within the FLASH tetramer repeat subunits. To create a system compatible with golden gate cloning while continuing to use the FLASH plasmid library, the four array receiving vectors were modified by removing the two BbsI sites and replacing the BsmBI array insertion site with a lacZ expression cassette flanked by two new BbsI sites. Following golden gate assembly and transformation, twelve white colonies are typically screened by PCR; constructs with an insert of the appropriate size are sequenced, and 60% of the sequenced clones are correct. At least one clone is correct in >90% of the twelve picked colonies per construct; this efficiency is similar to the FLASH protocol (26). Errors introduced by PCR amplification of the 0.3–0.4 kb subunits were present in <10% of sequenced clones.

The ggFLASH method was used to assemble TALEN constructs to ten loci in the human genome using either four tetramer repeats or three such repeats and a trimer to target 18 or 17 bp, respectively (Figure 2a). These loci were identified by TALENSeek to nine genes associated with neural disorders (CACNA1C and KCTD13) or development (FOXG1, FOXM1, PAX6, PVALB, SOX1, TMEM88, and TUBB3). For modeling TS, the TALEN pairs target Gly406 of CACNA1C (CACNA1C-Gly406; (19)); for the 16p11.2 microdeletion and microduplication syndrome the target is the N-terminus of the key gene (KCTD13-N).

Homology-directed repair assay for nuclease activity

A rapid assay to qualitatively assess HDR activity is necessary prior to spending weeks to months generating a transgenic hESC line (3,5,6). Following design and assembly of the TALEN pairs, the endonucleases were tested in K562 cells due to their high transfection efficiency prior to genome engineering of hESCs. Engineered nucleases are often tested by a non-homologous end joining (NHEJ) assay with the CEL-I endonuclease (45) or with a frame-shifted nonfunctional green fluorescent protein either integrated (9) or expressed episomally (46). Assessment of HDR activ-
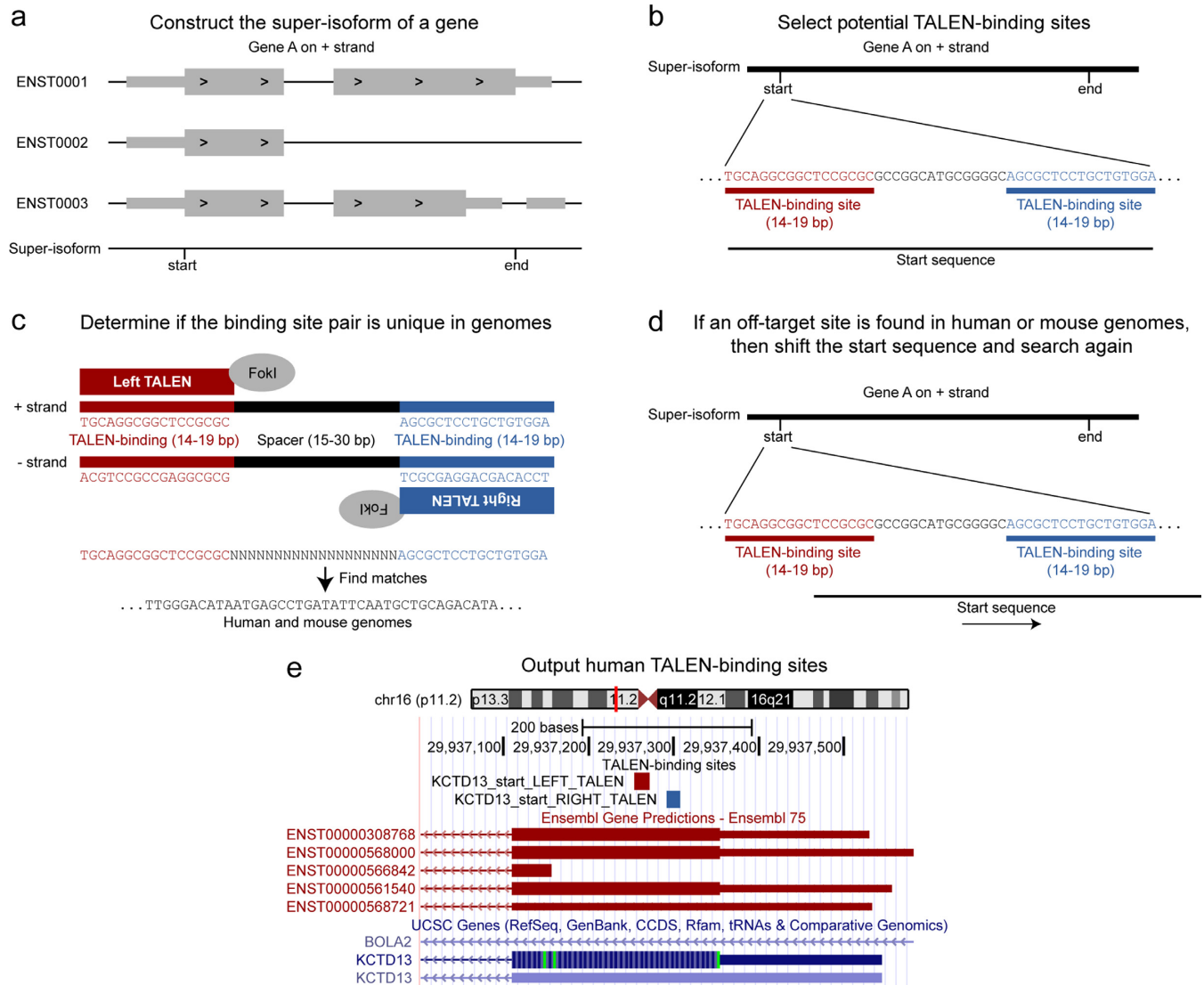
Figure 1. Description of TALENSeek, an automated TALEN-binding site detection algorithm. (**a**) Automated annotation of genomic regions is used to identify start and end sites of a gene of interest using the farthest protein coding regions across all isoforms (a super-isoform). (**b**) Potential TALEN-binding sites are identified. (**c**) The sites are evaluated for uniqueness across the genome. (**d**) If a binding site is not unique, the search continues further into the gene and the algorithm iterates from step (b). (**e**) The output .bed file allows visualization of the TALEN-binding sites in the UCSC genome browser, shown with human genome build GRCh37 (hg19).

ity has been done by detecting a restriction fragment length polymorphism (RFLP) following introduction of a restriction endonuclease recognition site (32), but this requires integration at sufficient levels to detect digestion products. A sensitive HDR assay was desired that would confirm the targeted integration of a homologous donor.

The ssODA is an amplification-based assay that only produces a junction PCR product if an exogenous sequence is integrated in the target site. Cells are transfected with a TALEN pair and a ssODN donor containing 40 nt of exogenous sequence flanked on each end by 40 nt of homology (Figure 2b; (32)). The exogenous sequence is used as a primer-binding site for junction PCR that produces a product only upon integration of the ssODN (Figure 2c).

In K562, the ssODA both corroborated and was more sensitive than the CEL-I assay and the ssODA was insensi-

tive to background polymorphisms (Figure 2d and Supplementary Figures S5–S8). At three loci in K562 cells, weaker HDR activity of ssODN donors was observed in the absence of engineered nucleases (Supplementary Figures S5a, S6a, S7b). Such background activity has been reported in mammalian cells previously (47), but was not observed during the preparation of the TS stem cell line or in similar work in iPSCs (6). Of the ten TALEN pairs designed by TALENSeek and assembled by ggFLASH, 90% (9/10) were active by ssODA but only 29% (2/7) were conclusively active by CEL-I. The inactivity of the FOXG1-N TALEN pair, which had the highest guanosine content (44%) of all targeted TALEN-binding sites, could be due to the use of the Asn-Asn RVD (10), the local chromatin structure (48), or the DNA methylation (37,48) of this gene associated with telencephalon development (49) in the bone-marrow de-
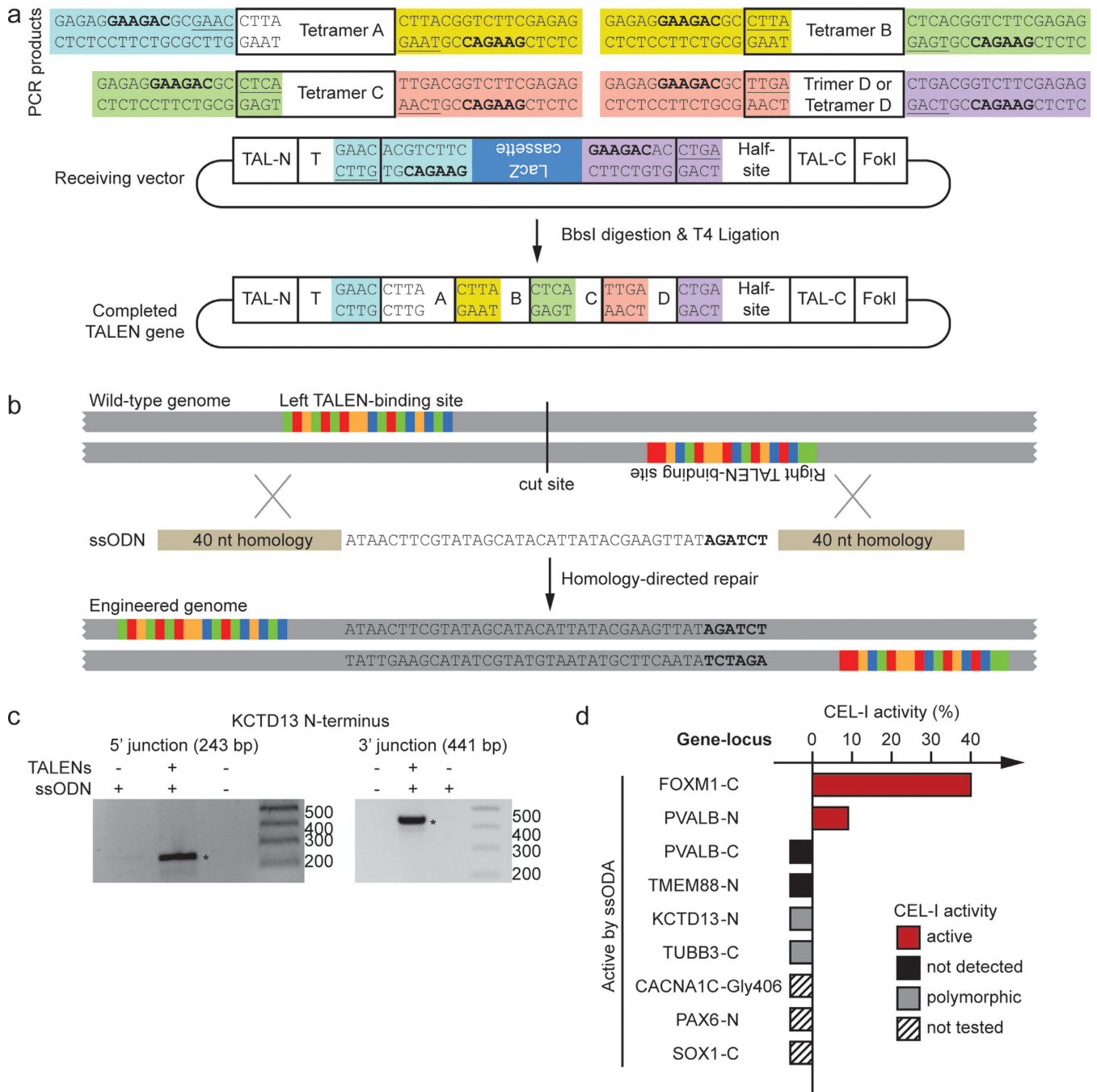
Figure 2. TALEN gene assembly and HDR activity assay. (**a**) TALEN assembly by ggFLASH. The PCR products for the four subunits include pairs of BbsI recognition sequences (bold) and overhangs (underlined). The repeat array receiving vectors each contain a LacZ expression cassette flanked by BbsI sites. (**b**) In the ssODA, a LoxP sequence and BglII recognition site (bold) are introduced with a ssODN donor. Schematics in (a) and (b) are not to scale. (**c**) Integration of a ssODN donor into the KCTD13 N-terminus of K562 cells required the presence of the appropriate TALEN pair. An asterisk denotes the desired product band. (**d**) CEL-I activity for TALENs that were active by ssODA in K562 cells.

rived K562 cell line. The CACNA1C-Gly406 and KCTD13-N TALEN pairs active by ssODA were each used with an appropriate homologous donor to generate hESCs to model ASDs.

hESC models of ASDs

The CACNA1C G406R mutation that is conserved in patients with TS was introduced into the non-disease H1 hESC background to generate an isogenic cell line. To generate this mutation in a scarless manner, hESCs cells were transfected with a TALEN pair and an 88 nt ssODN. This ssODN included 40 nt of homology on both the 5′ and
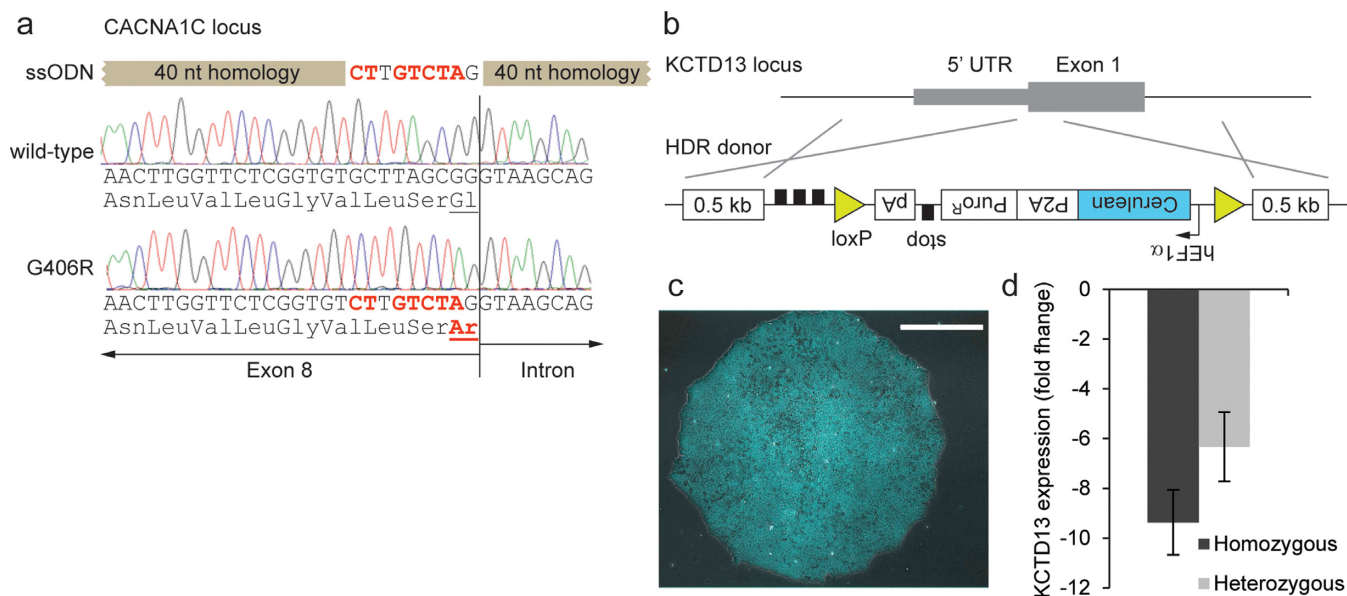
Figure 3. Engineered H1 hESCs for modeling human disease. (**a**) Aligned sequencing chromatograms showing the CACNA1C G406R mutation with the incorporation of the desired mutations (red nucleotides) from the ssODN. The first two codons of residue 406 are underlined. (**b**) The disruption cassette that was introduced immediately 3′ of the start codon of KCTD13 by HDR. (**c**) Colony of engineered KCTD13 heterozygous clone; scale bar 500 μm. (**d**) KCTD13 gene expression in the homozygous, dark grey, and heterozygous, light grey, KCTD13 disruption cell lines normalized to the expression in the parental H1 hESC ($n = 3$, error bars represent 1 standard deviation).

3′ ends of a central region including the G to A mutation encoding Arg406 and six silent mutations to create a unique primer-binding site (Figure 3a). The splice junction and the coding sequence of other residues in exon 8 were retained. Following a round of enrichment and single-cell cloning, the targeted loci of six selected clones were analyzed by sequencing blunt-cloned PCR products. One homozygous clone (CACNA1C$^{G406R/G406R}$) with targeted integration was identified.

Isogenic disease models will be useful tools for studying ASDs of both single gene origin, such as TS, and larger genotypes such as the CNVs of the 29 gene region of chromosome 16p11.2. Patients with 16p11.2 microdeletion or microduplication have a range of neurocognitive disorders and macroencephaly or microencephaly, respectively ([50]). Recent work has demonstrated that KCTD13, a single gene in this locus, may be the causative gene for the 16p11.2 disorders; in zebrafish, the CNV of this gene matched the observed neuroanatomical phenotypes of the regional CNV ([20]).

To disrupt the expression of KCTD13, an homologous donor plasmid was designed to insert stop codons in all three frames and a selection and screening cassette expressing puromycin resistance and the cerulean fluorescent protein ([51]) under the control of the constitutive hEF1α promoter (Figure 3b). The transgenes were flanked by 470 bp of homology on each side of the start codon. H1 hESCs were transfected with the KCTD13-N TALENs targeting the 5′ terminus of the coding region and the HDR plasmid. Following puromycin selection, twelve fluorescent colonies (Figure 3c) were selected, expanded, and analyzed by PCR for targeted integration. Eight colonies were heterozygous for the targeted integration of the transgenes, one colony was homozygous and three lacked the transgenes at the

desired locus. In undifferentiated hESCs, the heterozygous and homozygous clones with the inserted disruption elements each had the phenotype of healthy hESCs (Supplementary Figures S2 and S4) with a decreased expression of KCTD13 relative to unmodified H1 cells (Figure 3d).

In addition to the desired HDR activity of the TALENs and the plasmid or ssODN donor, the exogenous endonucleases used to introduce the site-specific DSB could also modify the genome by NHEJ activity at loci with similar sequences. The most similar loci to the KCTD13 and CACNA1C TALEN-binding sites were nine loci with four or five mismatches and seven loci with six or seven mismatches, respectively (Supplementary Table S4). Fifteen of these sixteen sites were amplified by PCR and sequenced in unmodified H1 and the corresponding engineered hESCs. No modifications were identified at these loci in these clones relative to H1.

DISCUSSION

Isogenic cell lines with a fixed genetic background are necessary for the study of human disease ([3]). In addition to generating isogenic hESCs for the study of two autism spectrum disorders, we report methods to identify unique TALEN-binding sites in the human genome using TALENSeek software, construct the genes using golden gate FLASH, and test the nuclease pairs with an amplification-based homology-directed repair assay at the loci of interest. TALENSeek, which can be run in a timely manner on a desktop computer, identifies TALEN-binding sites in the human genome, has the ability to find unique homologous binding sites in the human and mouse genomes, and is modifiable to find TALEN-binding sites in any species with an available complete genome sequence.

There are several advantages of the golden gate FLASH TALEN assembly method that produces correct TALEN genes at the same efficiency as the original FLASH method (26). Firstly, only nanogram quantities of PCR products are required instead of microgram quantities of plasmids; artifacts from amplification of the trimer or tetramer subunits are observed in <10% of sequenced clones. The lacZ cassette in each receiving vector enables targeted investigations of plasmids that have an inserted array. Finally, the unique position-specific overhangs necessary for the golden gate cloning reaction provide internal primer-binding sites for sequencing of the repeat array, obviating the need for long sequencing reads from outside of the array.

This represents a bioinformatic, cloning, and testing set of tools to design, assemble, and assess the homology-directed repair activity of any TALEN pair. We utilized these tools and methods to generate hESC lines that disrupt the expression of KCTD13 and scarlessly introduce the TS point mutation without modifying the genome at similar loci. These reagents and techniques could be applied to correct the Arg406 mutation in iPSCs from patients with TS (52), create a KCTD13 reporter cell line to monitor intracellular protein levels in live cells, or create isogenic cell lines for studying other human diseases.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Robinton,D.A. and Daley,G.Q. (2012) The promise of induced pluripotent stem cells in research and therapy. *Nature*, **481**, 295–305.
2. Wang,Y., Liang,P., Lan,F., Wu,H., Lisowski,L., Gu,M., Hu,S., Kay,M.A., Urnov,F.D., Shinnawi,R. *et al.* (2014) Genome editing of isogenic human induced pluripotent stem cells recapitulates long QT phenotype for drug testing. *J. Am. Coll. Cardiol.*, **64**, 451–459.
3. Soldner,F., Laganière,J., Cheng,A.W., Hockemeyer,D., Gao,Q., Alagappan,R., Khurana,V., Golbe,L.I., Myers,R.H., Lindquist,S. *et al.* (2011) Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell*, **146**, 318–331.
4. Yusa,K., Rashid,S.T., Strick-Marchand,H., Varela,I., Liu,P.-Q., Paschon,D.E., Miranda,E., Ordóñez,A., Hannan,N.R.F., Rouhani,F.J. *et al.* (2011) Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. *Nature*, **478**, 391–394.
5. Ding,Q., Lee,Y.K., Schaefer,E.A., Peters,D.T., Veres,A., Kim,K., Kuperwasser,N., Motola,D.L., Meissner,T.B., Hendriks,W.T. *et al.* (2013) A TALEN genome-editing system for generating human stem cell-based disease models. *Cell Stem Cell*, **12**, 238–251.
6. Miyaoka,Y., Chan,A.H., Judge,L.M., Yoo,J., Huang,M., Nguyen,T.D., Lizarraga,P.P., So,P.-L. and Conklin,B.R. (2014) Isolation of single-base genome-edited human iPS cells without antibiotic selection. *Nat. Methods*, **11**, 291–293.
7. Irion,S., Luche,H., Gadue,P., Fehling,H.J., Kennedy,M. and Keller,G. (2007) Identification and targeting of the ROSA26 locus in human embryonic stem cells. *Nat. Biotechnol.*, **25**, 1477–1482.
8. Rouet,P., Smih,F. and Jasin,M. (1994) Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.*, **14**, 8096–8106.
9. Urnov,F.D., Miller,J.C., Lee,Y.-L., Beausejour,C.M., Rock,J.M., Augustus,S., Jamieson,A.C., Porteus,M.H., Gregory,P.D. and Holmes,M.C. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, **435**, 646–651.
10. Boch,J., Scholze,H., Schornack,S., Landgraf,A., Hahn,S., Kay,S., Lahaye,T., Nickstadt,A. and Bonas,U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509–1512.
11. Miller,J.C., Tan,S., Qiao,G., Barlow,K.A., Wang,J., Xia,D.F., Meng,X., Paschon,D.E., Leung,E., Hinkley,S.J. *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.*, **29**, 143–148.
12. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
13. Mali,P., Esvelt,K.M. and Church,G.M. (2013) Cas9 as a versatile tool for engineering biology. *Nat. Methods*, **10**, 957–963.
14. Porteus,M.H. and Baltimore,D. (2003) Gene targeting in human cells. *Science*, **300**, 763.
15. Fu,Y., Foden,J.A., Khayter,C., Maeder,M.L., Reyon,D., Joung,J.K. and Sander,J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.*, **31**, 822–826.
16. Wu,X., Scott,D.A., Kriz,A.J., Chiu,A.C., Hsu,P.D., Dadon,D.B., Cheng,A.W., Trevino,A.E., Konermann,S., Chen,S. *et al.* (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.*, **32**, 670–676.
17. Kuscu,C., Arslan,S., Singh,R., Thorpe,J. and Adli,M. (2014) Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.*, **32**, 677–683.
18. Veres,A., Gosis,B.S., Ding,Q., Collins,R., Ragavendran,A., Brand,H., Erdin,S., Talkowski,M.E. and Musunuru,K. (2014) Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell*, **15**, 27–30.
19. Splawski,I., Timothy,K.W., Sharpe,L.M., Decher,N., Kumar,P., Bloise,R., Napolitano,C., Schwartz,P.J., Joseph,R.M., Condouris,K. *et al.* (2004) Ca(V)1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. *Cell*, **119**, 19–31.
20. Golzio,C., Willer,J., Talkowski,M.E., Oh,E.C., Taniguchi,Y., Jacquemont,S., Reymond,A., Sun,M., Sawa,A., Gusella,J.F. *et al.* (2012) KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature*, **485**, 363–367.
21. Sanders,S.J., Ercan-Sencicek,A.G., Hus,V., Luo,R., Murtha,M.T., Moreno-De-Luca,D., Chu,S.H., Moreau,M.P., Gupta,A.R., Thomson,S.A. *et al.* (2011) Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron*, **70**, 863–885.
22. Levy,D., Ronemus,M., Yamrom,B., Lee,Y.H., Leotta,A., Kendall,J., Marks,S., Lakshmi,B., Pai,D., Ye,K. *et al.* (2011) Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron*, **70**, 886–897.
23. Weiss,L.A., Shen,Y., Korn,J.M., Arking,D.E., Miller,D.T., Fossdal,R., Saemundsen,E., Stefansson,H., Ferreira,M.A.R., Green,T. *et al.* (2008) Association between microdeletion and microduplication at 16p11.2 and autism. *N. Eng. J. Med.*, **358**, 667–675.
24. Durinck,S., Moreau,Y., Kasprzyk,A., Davis,S., De Moor,B., Brazma,A. and Huber,W. (2005) BioMart and Bioconductor: a

powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21**, 3439–3440.

25. Streubel,J., Blücher,C., Landgraf,A. and Boch,J. (2012) TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.*, **30**, 593–595.

26. Reyon,D., Tsai,S.Q., Khayter,C., Foden,J.A., Sander,J.D. and Joung,J.K. (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.*, **30**, 460–465.

27. Engler,C., Kandzia,R. and Marillonnet,S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PloS One*, **3**, e3647.

28. Geißler,R., Scholze,H., Hahn,S., Streubel,J., Bonas,U., Behrens,S.-E. and Boch,J. (2011) Transcriptional activators of human genes with programmable DNA-specificity. *PloS One*, **6**, e19509.

29. Cermak,T., Doyle,E.L., Christian,M., Wang,L., Zhang,Y., Schmidt,C., Baller,J.A., Somia,N.V., Bogdanove,A.J. and Voytas,D.F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.*, **39**, e82.

30. Sanjana,N.E., Cong,L., Zhou,Y., Cunniff,M.M., Feng,G. and Zhang,F. (2012) A transcription activator-like effector toolbox for genome engineering. *Nat. Protoc.*, **7**, 171–192.

31. Engler,C., Gruetzner,R., Kandzia,R. and Marillonnet,S. (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PloS One*, **4**, e5553.

32. Chen,F., Pruett-Miller,S.M., Huang,Y., Gjoka,M., Duda,K., Taunton,J., Collingwood,T.N., Frodin,M. and Davis,G.D. (2011) High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat. Methods*, **8**, 753–755.

33. Luong,M.X., Auerbach,J., Crook,J.M., Daheron,L., Hei,D., Lomax,G., Loring,J.F., Ludwig,T., Schlaeger,T.M., Smith,K.P. *et al.* (2011) A call for standardized naming and reporting of human ESC and iPSC lines. *Cell Stem Cell*, **8**, 357–359.

34. Doyle,E.L., Booher,N.J., Standage,D.S., Voytas,D.F., Brendel,V.P., Vandyk,J.K. and Bogdanove,A.J. (2012) TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res.*, **40**, W117–W122.

35. Neff,K.L., Argue,D.P., Ma,A.C., Lee,H.B., Clark,K.J. and Ekker,S.C. (2013) Mojo Hand, a TALEN design tool for genome editing applications. *BMC Bioinformatics*, **14**, doi:10.1186/1471-2105-14-1.

36. Heigwer,F., Kerr,G., Walther,N., Glaeser,K., Pelz,O., Breinig,M. and Boutros,M. (2013) E-TALEN: a web tool to design TALENs for genome engineering. *Nucleic Acids Res.*, **41**, e190.

37. Kim,Y., Kweon,J., Kim,A., Chon,J.K., Yoo,J.Y., Kim,H.J., Kim,S., Lee,C., Jeong,E., Chung,E. *et al.* (2013) A library of TAL effector nucleases spanning the human genome. *Nat. Biotechnol.*, **31**, 251–258.

38. Karolchik,D., Barber,G.P., Casper,J., Clawson,H., Cline,M.S., Diekhans,M., Dreszer,T.R., Fujita,P.a., Guruvadoo,L., Haeussler,M. *et al.* (2014) The UCSC genome browser database: 2014 update. *Nucleic Acids Res.*, **42**, D764–D770.

39. Fine,E.J., Cradick,T.J., Zhao,C.L., Lin,Y. and Bao,G. (2014) An online bioinformatics tool predicts zinc finger and TALE nuclease off-target cleavage. *Nucleic Acids Res.*, **42**, e42.

40. Li,L., Piatek,M.J., Atef,A., Piatek,A., Wibowo,A., Fang,X., Sabir,J.S., Zhu,J.K. and Mahfouz,M.M. (2012) Rapid and highly efficient construction of TALE-based transcriptional regulators and nucleases for genome modification. *Plant Mol. Biol.*, **78**, 407–416.

41. Grau,J., Boch,J. and Posch,S. (2013) TALENoffer: genome-wide TALEN off-target prediction. *Bioinformatics*, **29**, 2931–2932.

42. Lin,Y., Fine,E.J., Zheng,Z., Antico,C.J., Voit,R.A., Porteus,M.H., Cradick,T.J. and Bao,G. (2014) SAPTA: a new design tool for improving TALE nuclease activity. *Nucleic Acids Res.*, **42**, e47.

43. Sander,J.D., Cade,L., Khayter,C., Reyon,D., Peterson,R.T., Joung,J.K. and Yeh,J.R. (2011) Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat. Biotechnol.*, **29**, 697–698.

44. Schmid-Burgk,J.L., Schmidt,T., Kaiser,V., Honing,K. and Hornung,V. (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat. Biotechnol.*, **31**, 76–81.

45. Qiu,P., Shandilya,H., D'Alessio,J.M., O'Connor,K., Durocher,J. and Gerard,G.F. (2004) Mutation detection using Surveyor nuclease. *Biotechniques*, **36**, 702–707.

46. Kim,H., Um,E., Cho,S.-R., Jung,C., Kim,H. and Kim,J.-S. (2011) Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat. Methods*, **8**, 941–943.

47. Igoucheva,O., Alexeev,V. and Yoon,K. (2001) Targeted gene correction by small single-stranded oligonucleotides in mammalian cells. *Gene Ther.*, **8**, 391–399.

48. Bultmann,S., Morbitzer,R., Schmidt,C.S., Thanisch,K., Spada,F., Elsaesser,J., Lahaye,T. and Leonhardt,H. (2012) Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res.*, **40**, 5368–5377.

49. Hébert,J.M. and Fishell,G. (2008) The genetics of early telencephalon patterning: some assembly required. *Nat. Rev. Neurosci.*, **9**, 678–685.

50. Jacquemont,S., Reymond,A., Zufferey,F., Harewood,L., Walters,R.G., Kutalik,Z., Martinet,D., Shen,Y., Valsesia,A., Beckmann,N.D. *et al.* (2011) Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature*, **478**, 97–102.

51. Rizzo,M.A., Springer,G.H., Granada,B. and Piston,D.W. (2004) An improved cyan fluorescent protein variant useful for FRET. *Nat. Biotechnol.*, **22**, 445–449.

52. Paşca,S.P., Portmann,T., Voineagu,I., Yazawa,M., Shcheglovitov,A., Paşca,A.M., Cord,B., Palmer,T.D., Chikahisa,S., Nishino,S. *et al.* (2011) Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome. *Nat. Med.*, **17**, 1657–1662.