Whole-Genome Sequencing Analysis Reveals High Specificity of CRISPR/Cas9 and TALEN-Based Genome Editing in Human iPSCs

Cory Smith[1,2,3], **Athurva Gore**[6], **Wei Yan**[1,2], **Leire Abalde-Atristain**[1,2,4], **Zhe Li**[6], **Chaoxia He**[1,2], **Ying Wang**[1,2,5], **Robert A. Brodsky**[1], **Kun Zhang**[6], **Linzhao Cheng**[1,2,3,4,*], and **Zhaohui Ye**[1,2,*]

[1]Division of Hematology, Department of Medicine

[2]Stem Cell Program, Institute for Cell Engineering

[3]Predoctoral Training Program in Human Genetics

[4]Cellular and Molecular Medicine Graduate Program

[5]Department of Chemical and Biomolecular Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

[6]Department of Bioengineering, Institute for Genomic Medicine and Institute of Engineering in Medicine, University of California at San Diego, La Jolla, CA 92093, USA

Human iPSCs provide renewable cell sources for human biology and disease research and the potential for developing gene and cell therapy. Realization of this potential will rely in part on our ability to precisely edit or engineer the human genome in an efficient way. Recent developments in designer endonuclease technologies such as zinc finger nuclease (ZFN), transcription activator-like effector nuclease (TALEN), and clustered regulatory interspaced short palindromic repeat (CRISPR)/Cas9 endonuclease have provided ways to significantly improve genome editing efficiency in human iPSCs. These endonucleases make a double-stranded break (DSB) at a predetermined DNA sequence and trigger natural DNA repair processes such as nonhomologous end joining (NHEJ) or homologous recombination (HR) with a donor DNA template. Among these existing approaches, RNA-guided CRISPR/Cas9 is the most user-friendly and versatile system, and it has been applied in both animal models and cell lines (Cong et al., 2013; Hsu et al., 2014; Mali et al., 2013). The most commonly used system consists of a single polypeptide endonuclease Cas9 complexed with a single guide RNA (gRNA) that provides complementarity to 20-nucleotide target DNA sequence. However, the specificity and efficiency of this approach in human iPSCs have not been studied in detail (Cong et al., 2013; Ding et al., 2013; Mali et al., 2013; Yang et al., 2013). Some analyses using cancer cell lines reported higher-than-

*Correspondence: lcheng2@jhmi.edu (L.C.), zye@jhmi.edu (Z.Y.).

expected levels of off-target mutagenesis by Cas9-gRNAs (Fu et al., 2013; Hsu et al., 2013), raising concerns about the practical applicability of this approach in therapeutic contexts. Some recent studies, including one on human adult stem cells, showed a minimal level of off-target effects by CRISPR/Cas9 (Schwank et al., 2013). However, these existing analyses of off-target effects and mutational load in gene-corrected stem cells have been restricted to checking predicted off target sites and are therefore limited in scope. To assess the value of this type of gene editing approach for therapeutic applications, it is critical to rigorously examine whether it is possible to generate gene-edited cell lines with minimal mutational load.

To this end, we have conducted whole-genome sequencing of four iPSC clones successfully targeted at the AAVS1 locus, a "safe harbor" in the human genome that is used for stable transgene expression in a variety of contexts. To generate the lines, we used an integration-free human iPSC line, BC1, whose genomic integrity has been characterized in detail by next-generation sequencing (Cheng et al., 2012) and targeted a GFP expression cassette into the AAVS1 site with either a previously reported Cas9-gRNA combination or a pair of improved heterodimeric TALENs (Mali et al., 2013; Yan et al., 2013) (Table S1 and Supplemental Experimental Procedures available online). Twenty days after transfection of the donor plasmid and either the TALENs or Cas9-gRNA into BC1, we harvested four clones with confirmed targeted integration (hCas9-C4, hCas9-C16, TALEN-C3, and TALEN-C6; Table S1 and Supplemental Experimental Procedures) and the parental BC1 iPSCs for whole-genome sequencing. The sequencing reads, ranging from 83 Gbps to 100 Gbps from each targeted clone, were first aligned to the human hg19 reference genome to enable identification of single-nucleotide variants (SNVs) and small indels (Table S1). Our analysis identified 4.2 million SNVs and 500,000 indels in each genome (Table S1) in comparison to the hg19 reference genome, suggesting that it is a rigorous data set that covers the genome in sufficient depth to detect sequence variants. The "germ-line" variants (present in BC1 parental iPSCs and different from hg19) were readily detectable in each targeted cell line (80%%–88%%), indicating that the sensitivity of variant detection in our analysis is high (Table S1). The variations from each targeted clone were then compared to the BC1 parental iPSCs to enable the generation of a list of potential variations arising during the gene editing process, which we then confirmed using genomic PCR and Sanger sequencing. We confirmed 62 out of 69 SNVs tested for an overall confirmation rate of 90%, and based on that we estimate that the total SNVs in the four iPSC clones range between 217 and 281 and that the total indels range between 7 and 12 (Table S1). Overall the genomic variation levels in TALEN- and Cas9-targeted groups were comparable. One important consideration is how many of these detected SNVs and indels were the results of off-target mutagenesis by the engineered endonucleases. To address this question, we generated a list of 3,665 (Cas9) and 238 (TALEN) putative off-target positions by using the EMBOSS fuzznuc software package. Each candidate SNV and indel was compared to this list and none of them are within a potential off-target region (Table S1), consistent with previous analyses looking at predicted off-target sites. Our analysis also shows that each SNV and indel is unique and that none of them occurred in more than one cell line. The absence of recurring mutations and the fact that none of the mutations resides in any putative off-target site by bioinformatic prediction strongly suggest that these mutations were

randomly accumulated during regular cell expansion and are not direct results of off-target activities by Cas9 or TALENs.

Our results from whole-genome sequencing analysis of Cas9- and TALEN-targeted human iPSC clones demonstrate that these engineered endonucleases provide efficient genome-editing tools with high specificity. It remains to be clarified whether the higher off-target rates observed in cancer cell lines are due to the overexpression of gRNAs and Cas9 protein and/or due to exacerbated and faulty DNA repair in these cell types. The higher specificity observed in human iPSCs, combined with the rapid development of next-generation sequencing technology, makes it possible to characterize and isolate high quality genome-edited stem cell clones with minimal mutational load. The guiding principle established with human iPSCs will likely be applicable to other types of stem cells and come with improvements in gene transfer and targeting efficiencies. Our current study of gene targeting in human iPSCs will help to establish better models for human biology and disease research and to provide proof-of-principle for future gene therapy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Cheng L, Hansen NF, Zhao L, Du Y, Zou C, Donovan FX, Chou BK, Zhou G, Li S, Dowey SN, et al. NISC Comparative Sequencing Program. Cell Stem Cell. 2012; 10:337–344. [PubMed: 22385660]

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. Science. 2013; 339:819–823. [PubMed: 23287718]

Ding Q, Regan SN, Xia Y, Oostrom LA, Cowan CA, Musunuru K. Cell Stem Cell. 2013; 12:393–394. [PubMed: 23561441]

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. Nat Biotechnol. 2013; 31:822–826. [PubMed: 23792628]

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, et al. Nat Biotechnol. 2013; 31:827–832. [PubMed: 23873081]

Hsu PD, Lander ES, Zhang F. Cell. 2014; 157:1262–1278. [PubMed: 24906146]

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. Science. 2013; 339:823–826. [PubMed: 23287722]

Schwank G, Koo BK, Sasselli V, Dekkers JF, Heo I, Demircan T, Sasaki N, Boymans S, Cuppen E, van der Ent CK, et al. Cell Stem Cell. 2013; 13:653–658. [PubMed: 24315439]

Yan W, Smith C, Cheng L. Sci Rep. 2013; 3:2376. [PubMed: 23921522]

Yang L, Guell M, Byrne S, Yang JL, De Los Angeles A, Mali P, Aach J, Kim-Kiselak C, Briggs AW, Rios X, et al. Nucleic Acids Res. 2013; 41:9049–9061. [PubMed: 23907390]