



PRECISE GENOME EDITING TECHNIQUES AND APPLICATIONS

EDITED BY: Zhiying Zhang, David Jay Segal and Kun Xu
PUBLISHED IN: Frontiers in Genetics



frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88963-779-9

DOI 10.3389/978-2-88963-779-9

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

PRECISE GENOME EDITING TECHNIQUES AND APPLICATIONS

Topic Editors:

Zhiying Zhang, Northwest A and F University, China

David Jay Segal, University of California, Davis, United States

Kun Xu, Northwest A and F University, China

Citation: Zhang, Z., Segal, D. J., Xu, K., eds. (2020). Precise Genome Editing Techniques and Applications. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88963-779-9

Table of Contents

- 05 Editorial: Precise Genome Editing Techniques and Applications**
Kun Xu, David Jay Segal and Zhiying Zhang
- 08 Genome Editing of Pigs for Agriculture and Biomedicine**
Huaqiang Yang and Zhenfang Wu
- 20 Trio-Based Deep Sequencing Reveals a Low Incidence of Off-Target Mutations in the Offspring of Genetically Edited Goats**
Chao Li, Shiwei Zhou, Yan Li, Guanwei Li, Yige Ding, Lan Li, Jing Liu, Lei Qu, Tad Sonstegard, Xingxu Huang, Yu Jiang, Yulin Chen, Bjoern Petersen and Xiaolong Wang
- 32 Methodologies for Improving HDR Efficiency**
Mingjie Liu, Saad Rehman, Xidian Tang, Kui Gu, Qinlei Fan, Dekun Chen and Wentao Ma
- 41 FTO Knockout Causes Chromosome Instability and G2/M Arrest in Mouse GC-1 Cells**
Tao Huang, Qiang Gao, Tongying Feng, Yi Zheng, Jiayin Guo and Wenxian Zeng
- 52 Precise and Rapid Validation of Candidate Gene by Allele Specific Knockout With CRISPR/Cas9 in Wild Mice**
Tianzhu Chao, Zhuangzhuang Liu, Yu Zhang, Lichen Zhang, Rong Huang, Le He, Yanrong Gu, Zhijun Chen, Qianqian Zheng, Lijin Shi, Wenping Zheng, Xinhui Qi, Eryan Kong, Zhongjian Zhang, Toby Lawrence, Yinming Liang and Liaoxun Lu
- 62 Guidelines for Fluorescent Guided Biallelic HDR Targeting Selection With PiggyBac System Removal for Gene Editing**
Javier Jarazo, Xiaobing Qing and Jens C. Schwamborn
- 74 Programmable Base Editing of the Sheep Genome Revealed No Genome-Wide Off-Target Mutations**
Shiwei Zhou, Bei Cai, Chong He, Ying Wang, Qiang Ding, Jiao Liu, Yao Liu, Yige Ding, Xiaoe Zhao, Guanwei Li, Chao Li, Honghao Yu, Qifang Kou, Wenzhi Niu, Bjoern Petersen, Tad Sonstegard, Baohua Ma, Yulin Chen and Xiaolong Wang
- 82 sgRNA-shRNA Structure Mediated SNP Site Editing on Porcine IGF2 Gene by CRISPR/StCas9**
Yongsen Sun, Nana Yan, Lu Mu, Bing Sun, Jingrong Deng, Yuanyuan Fang, Simin Shao, Qiang Yan, Furong Han, Zhiying Zhang and Kun Xu
- 91 Increasing Cytosine Base Editing Scope and Efficiency With Engineered Cas9-PmCDA1 Fusions and the Modified sgRNA in Rice**
Ying Wu, Wen Xu, Feipeng Wang, Si Zhao, Feng Feng, Jinling Song, Chengwei Zhang and Jinxiao Yang
- 101 Biological Characteristics of Severe Combined Immunodeficient Mice Produced by CRISPR/Cas9-Mediated Rag2 and IL2rg Mutation**
Yong Zhao, Peijuan Liu, Zhiqian Xin, Changhong Shi, Yinlan Bai, Xiuxuan Sun, Ya Zhao, Xiaoya Wang, Li Liu, Xuan Zhao, Zhinan Chen and Hai Zhang

111 *Enhancement of Precise Gene Editing by the Association of Cas9 With Homologous Recombination Factors*

Ngoc-Tung Tran, Sanum Bashir, Xun Li, Jana Rossius, Van Trung Chu, Klaus Rajewsky and Ralf Kühn

124 *Efficient and Precise CRISPR/Cas9-Mediated MECP2 Modifications in Human-Induced Pluripotent Stem Cells*

Thi Thanh Huong Le, Ngoc Tung Tran, Thi Mai Lan Dao, Dinh Dung Nguyen, Huy Duong Do, Thi Lien Ha, Ralf Kühn, Thanh Liem Nguyen, Klaus Rajewsky and Van Trung Chu



Editorial: Precise Genome Editing Techniques and Applications

Kun Xu¹, David Jay Segal^{2*} and Zhiying Zhang^{1*}

¹ College of Animal Science and Technology, Northwest A&F University, Yangling, China, ² Department of Biochemistry and Molecular Medicine, Genome Center, University of California, Davis, Davis, CA, United States

Keywords: precise genome editing, CRISPR, HDR efficiency, biallelic HDR targeting, off-target effect, animal model, base editing

Editorial on the Research Topic

Precise Genome Editing Techniques and Applications

The CRISPR/Cas system, particularly CRISPR/Cas9 (Jinek et al., 2012; Cong et al., 2013), has been developed as a robust and versatile platform for manipulating the genomes of a variety of species. In recent years, numerous reports have suggested its powerful potential application for human gene therapy and life science research, as well as animal and plant breeding. This might be evidenced by the collections in this Research Topic, “Precise Genome Editing Techniques and Applications.”

Generally, the CRISPR/Cas9 nuclease is used to cleave target genomic DNA to generate site-specific double-strand breaks (DSBs), which are predominantly repaired via non-homologous end joining (NHEJ) or, to lesser extent, by homology-directed repair (HDR). The classical NHEJ repair pathway can generate small insertions or deletions (indels), resulting in loss-of-function of targeted coding genes by introducing a frameshift in the open reading frame (ORF). NHEJ mutagenesis is a highly popular strategy for gene manipulation. In addition to the classical NHEJ, alternative or accurate NHEJ-mediated repair can achieve precise genomic DNA deletions (Guo et al., 2018; Shou et al., 2018).

Two papers in this Research Topic by Chao et al. and Zhao et al. describe the manufacture of allele-specific knockout and double gene knockout mouse models for rapid disease gene validation and human xenograft studies, respectively. *N*⁶-methyladenosine (m⁶A) is a well-established epigenetic modification on eukaryotic mRNA. An increasing number of studies have uncovered the significance of m⁶A methylation, which has given rise to the nascent field of “epitranscriptomics.” Another article in this volume (Huang et al.) describes a knock-out study in mouse spermatogonial GC-1 cells of the fat mass and obesity-associated (*Fto*) gene, which has been shown to act on the epitranscriptome as an m⁶A demethylase (Li et al., 2017; Lin et al., 2017).

On the other hand, the HDR repair pathway relies on homologous donor DNA to produce targeted gene knock-ins at the DSB site or gene replacement between two DSB sites. Precise point mutations and designed small indels can also be achieved by this method. One paper in this topic describes efforts to precisely correct the methyl-CpG binding protein 2 (MECP2) gene in the context of Rett syndrome (RTT) by CRISPR/Cas9-mediated HDR in human induced pluripotent stem cells (iPSCs). This report provides a reference for iPSC-based disease modeling and gene correction therapy (Le et al.).

Although the HDR-based genome can achieve gene insertions and precise substitutions, it is still confronted by several disadvantages during the precise editing process including low HDR efficiency, failure of biallelic targeting, complications of positive selection, and the re-deletion of selection markers.

OPEN ACCESS

Edited and reviewed by:

Youri I. Pavlov,
University of Nebraska Medical
Center, United States

*Correspondence:

David Jay Segal
djsegal@ucdavis.edu
Zhiying Zhang
zhangzhy@nwfau.edu.cn

Specialty section:

This article was submitted to
Genomic Assay Technology,
a section of the journal
Frontiers in Genetics

Received: 19 February 2020

Accepted: 01 April 2020

Published: 21 April 2020

Citation:

Xu K, Segal DJ and Zhang Z (2020)
Editorial: Precise Genome Editing
Techniques and Applications.
Front. Genet. 11:412.
doi: 10.3389/fgene.2020.00412

It has been reported that inhibiting the key molecules of the competitive NHEJ pathway, such as DNA ligase IV (LIG4) and KU70, could improve the HDR efficiency effectively (Chu et al., 2015; Maruyama et al., 2015). We have previously developed a novel sgRNA-shRNA structure transcribing multiple sgRNAs for multiplex genome targeting (Yan et al., 2016). Here, the structure was further used for simultaneous *LIG4* RNA interference and the enhanced HDR-based *IGF2* SNP editing (Sun et al.). On the other hand, the HDR pathway can be also enhanced by the association of Cas9 with a variety of homologous recombination factors, such as γ Rad52 (Shao et al., 2017), dn53BP1 (Paulsen et al., 2017; Jayavaradhan et al., 2019), hRad51 (Rees et al., 2019), and CtIP (Tran et al.). A review paper in this topic (Liu et al.) further summarizes the methodologies and other considerations for improving the HDR efficiency.

Regarding biallelic targeting, we have previously reported a novel strategy using two donors with paired selectable markers (Wu et al., 2017). However, the removal of the selection is often required to allay concerns of marker-dependent effects. There are several “pop in and out” two-step techniques for marker-free genome engineering, including the Cre/LoxP system (Zhu et al., 2015), the *piggyBac* transposon (Xie et al., 2014), and the SSA repair mechanism (Li et al., 2018). This Research Topic presents a protocol article for biallelic HDR targeting using *piggyBac*-mediated selection removal (Jarazo et al.).

The ever-expanding repertoire of CRISPR editing systems includes the widely used Cas9 of *Streptococcus pyogenes* (*SpCas9*) (Jinek et al., 2012; Cong et al., 2013), as well as *Streptococcus thermophilus* (*StCas9*) (Xu et al., 2015), and *Neisseria meningitidis* (*NmCas9*) (Hou et al., 2013). In addition, other proteins in the CRISPR family such as Cpf1/Cas12a (Zetsche et al., 2015) have been applied for genome editing. More recently, CRISPR/Cas-derived novel genome editing tools that do not create DSBs have been developed, including the cytidine

and adenine base editors (CBE and ABE) (Komor et al., 2016; Gaudelli et al., 2017), as well as prime editors (PE) (Anzalone et al., 2019). The paper by Wu et al. in this topic describes efforts to increase the CBE scope and efficiency in rice.

The rapid development of genome editing technology has provided opportunities for modifying large animal models and domestic animal breeding. Pigs serve as an important agricultural resource as well as animal models for biomedical studies. In this topic, Yang and Wu summarize the genome editing of pigs in agricultural and biomedical applications. Off-target effects are one of the major concerns for genome editing research. The last two articles (Li et al.; Zhou et al.) in this Research Topic report no obvious off-target events in the offspring of genetically edited goats. However, it remains to be determined as to whether these observations might be affected by survivorship bias, as well as differences in off-target events between human gene therapy and animal genetic breeding.

In conclusion, the articles contained within this Research Topic illustrate the mechanisms and great potential of precise genome editing techniques to further scientific inquiry and produce useful outcomes that benefit society.

AUTHOR CONTRIBUTIONS

KX wrote the manuscript draft. DS and ZZ supervised the topic and read the manuscript.

FUNDING

This work was supported by grants from the China Postdoctoral Science Foundation (2018T111111 and 2015M580887), the National Natural Science Foundation of China (NSFC, 31702099), and the National Transgenic Major Project of China (2018ZX08010-09B).

REFERENCES

- Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157. doi: 10.1038/s41586-019-1711-4
- Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., et al. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* 33, 543–548. doi: 10.1038/nbt.3198
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi: 10.1126/science.1231143
- Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., et al. (2017). Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* 551, 464–471. doi: 10.1038/nature24644
- Guo, T., Feng, Y.-L., Xiao, J.-J., Liu, Q., Sun, X.-N., Xiang, J.-F., et al. (2018). Harnessing accurate non-homologous end joining for efficient precise deletion in CRISPR/Cas9-mediated genome editing. *Genome Biol.* 19:170. doi: 10.1186/s13059-018-1518-x
- Hou, Z., Zhang, Y., Propson, N. E., Howden, S. E., Chu, L. F., Sontheimer, E. J., et al. (2013). Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15644–15649. doi: 10.1073/pnas.1313587110
- Jayavaradhan, R., Pillis, D. M., Goodman, M., Zhang, F., Zhang, Y., Andreassen, P. R., et al. (2019). CRISPR-Cas9 fusion to dominant-negative 53BP1 enhances HDR and inhibits NHEJ specifically at Cas9 target sites. *Nat. Commun.* 10:2866. doi: 10.1038/s41467-019-10735-7
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829
- Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424. doi: 10.1038/nature17946
- Li, X., Bai, Y., Cheng, X., Kalds, P. G. T., Sun, B., Wu, Y., et al. (2018). Efficient SSA-mediated precise genome editing using CRISPR/Cas9. *FEBS J* 285, 3362–3375. doi: 10.1111/febs.14626
- Li, Z., Weng, H., Su, R., Weng, X., Zuo, Z., Li, C., et al. (2017). FTO plays an oncogenic role in acute myeloid leukemia as a N(6)-methyladenosine RNA demethylase. *Cancer Cell* 31, 127–141. doi: 10.1016/j.ccell.2016.11.017
- Lin, Z., Hsu, P. J., Xing, X., Fang, J., Lu, Z., Zou, Q., et al. (2017). Mettl3-/Mettl14-mediated mRNA N(6)-methyladenosine modulates murine spermatogenesis. *Cell Res.* 27, 1216–1230. doi: 10.1038/cr.2017.117
- Maruyama, T., Dougan, S. K., Truttmann, M. C., Bilate, A. M., Ingram, J. R., and Ploegh, H. L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* 33, 538–542. doi: 10.1038/nbt.3190

- Paulsen, B. S., Mandal, P. K., Frock, R. L., Boyraz, B., Yadav, R., Upadhyayula, S., et al. (2017). Ectopic expression of RAD52 and dn53BP1 improves homology-directed repair during CRISPR-Cas9 genome editing. *Nat. Biomed. Eng.* 1, 878–888. doi: 10.1038/s41551-017-0145-2
- Rees, H. A., Yeh, W.-H., and Liu, D. R. (2019). Development of hRad51–Cas9 nickase fusions that mediate HDR without double-stranded breaks. *Nat. Commun.* 10:2212. doi: 10.1038/s41467-019-09983-4
- Shao, S., Ren, C., Liu, Z., Bai, Y., Chen, Z., Wei, Z., et al. (2017). Enhancing CRISPR/Cas9-mediated homology-directed repair in mammalian cells by expressing *Saccharomyces cerevisiae* Rad52. *Int. J. Biochem. Cell Biol.* 92, 43–52. doi: 10.1016/j.biocel.2017.09.012
- Shou, J., Li, J., Liu, Y., and Wu, Q. (2018). Precise and predictable CRISPR chromosomal rearrangements reveal principles of Cas9-mediated nucleotide insertion. *Mol. Cell* 71, 498–509 e4. doi: 10.1016/j.molcel.2018.06.021
- Wu, Y., Xu, K., Ren, C., Li, X., Lv, H., Han, F., et al. (2017). Enhanced CRISPR/Cas9-mediated biallelic genome targeting with dual surrogate reporter-integrated donors. *FEBS Lett.* 591, 903–913. doi: 10.1002/1873-3468.12599
- Xie, F., Ye, L., Chang, J. C., Beyer, A. I., Wang, J., Muench, M. O., et al. (2014). Seamless gene correction of beta-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. *Genome Res.* 24, 1526–1533. doi: 10.1101/gr.173427.114
- Xu, K., Ren, C., Liu, Z., Zhang, T., Zhang, T., Li, D., et al. (2015). Efficient genome engineering in eukaryotes using Cas9 from *Streptococcus thermophilus*. *Cell Mol. Life Sci.* 72, 383–399. doi: 10.1007/s00018-014-1679-z
- Yan, Q., Xu, K., Xing, J., Zhang, T., Wang, X., Wei, Z., et al. (2016). Multiplex CRISPR/Cas9-based genome engineering enhanced by Drosha-mediated sgRNA-shRNA structure. *Sci. Rep.* 6:38970. doi: 10.1038/srep38970
- Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771. doi: 10.1016/j.cell.2015.09.038
- Zhu, S., Rong, Z., Lu, X., Xu, Y., and Fu, X. (2015). Gene targeting through homologous recombination in monkey embryonic stem cells using CRISPR/Cas9 system. *Stem Cells Dev.* 24, 1147–1149. doi: 10.1089/scd.2014.0507

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Xu, Segal and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genome Editing of Pigs for Agriculture and Biomedicine

Huaqiang Yang* and Zhenfang Wu*

National Engineering Research Center for Breeding Swine Industry, College of Animal Science, South China Agricultural University, Guangzhou, China

OPEN ACCESS

Edited by:

Kun Xu,
Northwest A&F University, China

Reviewed by:

Dengke Pan,
Beijing Academy of Agriculture
and Forestry Sciences, China
Angelika Schnieke,
Technische Universität München,
Germany

*Correspondence:

Huaqiang Yang
yangh@scau.edu.cn
Zhenfang Wu
wzfemail@163.com

Specialty section:

This article was submitted to
Genomic Assay Technology,
a section of the journal
Frontiers in Genetics

Received: 28 April 2018

Accepted: 21 August 2018

Published: 04 September 2018

Citation:

Yang H and Wu Z (2018) Genome
Editing of Pigs for Agriculture
and Biomedicine.
Front. Genet. 9:360.
doi: 10.3389/fgene.2018.00360

Pigs serve as an important agricultural resource and animal model in biomedical studies. Efficient and precise modification of pig genome by using recently developed gene editing tools has significantly broadened the application of pig models in various research areas. The three types of site-specific nucleases, namely, zinc-finger nucleases, transcription activator-like effector nucleases, and clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein, are the main gene editing tools that can efficiently introduce predetermined modifications, including knockouts and knockins, into the pig genome. These modifications can confer desired phenotypes to pigs to improve production traits, such as optimal meat production, enhanced feed digestibility, and disease resistance. Besides, given their genetic, anatomic, and physiologic similarities to humans, pigs can also be modified to model human diseases or to serve as an organ source for xenotransplantation to save human lives. To date, many genetically modified pig models with agricultural or biomedical values have been established by using gene editing tools. These pig models are expected to accelerate research progress in related fields and benefit humans.

Keywords: pig, genome editing, ZFN, TALEN, CRISPR/Cas9, disease model, xenotransplantation, breeding

INTRODUCTION

Pigs hold great promise in agriculture and biomedicine. As an important meat source, domestic pigs provide the most commonly consumed meat worldwide. Through selective breeding, humans produce pigs that harbor desired characteristics for agriculture, albeit the selection is a long and slow process (Ruan et al., 2017). However, the process can now be substantially revolutionized and accelerated through genetic modification, including random transgenesis and gene knockouts and knockins (Gaj et al., 2013; Garas et al., 2015). With the improved efficiency of genetic modification, pig genome modification can confer any desired, predetermined genetic changes, which would take years to be realized in traditional selective breeding. Numerous economically significant characteristics, such as increased meat production (Qian et al., 2015; Wang et al., 2015, 2017b; Bi et al., 2016; Rao et al., 2016), reduced fat deposition (Zheng et al., 2017), or enhanced disease resistance (Whitworth et al., 2016; Burkard et al., 2017; Wells et al., 2017; Yang et al., 2018), have been achieved simply and efficiently through genetic modification in pigs, which can be used as valuable breeding materials to advance pig production. In biomedical research, pigs serve as an important large animal model given their advantages over other models. Compared with rodent models, pigs share a higher similarity to human beings in terms of body/organ size, lifespan, anatomy, physiology, and metabolic profile. Compared with non-human primates, pigs have low-cost and mature embryonic manipulation techniques. Pigs can be modified to carry the

same gene mutation found in humans to replicate inherited diseases (Perleberg et al., 2018), or offer organs with minimal transplant rejection during xenotransplantation (Hryhorowicz et al., 2017). Pigs bridge the gap between humans and the heavily used small rodent models to favor biomedical research ranging from basic science to translational medicine.

However, production of gene-targeted mammals other than mice remained difficult until the late 2000s because the traditional gene targeting technique developed in mice requires homologous recombination (HR) manipulation in embryonic stem (ES) cells (Mansour et al., 1988; Capecchi, 1989). The lack of bona fide germline-competent ES in large animals urged researchers to perform HR in somatic cells rather than in ES cells and then use somatic cell nuclear transfer (SCNT) to produce genetically modified large animals (Polejaeva et al., 2000). Although theoretically possible, the modification of somatic cells by using HR is extremely inefficient; thus, the generation of such cells is impractical (Wells and Prather, 2017). Therefore, only very few gene-targeted pigs were created within two decades (between the establishment of gene targeting technique prior to 1990 and the late 2010s when the novel gene targeting tools began to be used in large animals) (Dai et al., 2002; Lai et al., 2002; Rogers et al., 2008; Suzuki et al., 2012; Davis et al., 2014). This situation changed when newly developed gene targeting technologies called site-specific engineered nucleases or “gene scissors” became available. The designed engineered nucleases, such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), or clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein (Cas), are very effective in creating double-stranded breaks (DSB) at a specific locus of a genome, thereby facilitating genetic modifications, including knockouts via non-homologous end joining (NHEJ) and knockins via homology-directed repair (HDR) (Hsu et al., 2014) (**Figure 1**). Creation of an intentional DSB in the genomic target can stimulate HR by 50–1000-fold (Jasin, 1996). Therefore, a high rate of DSB formation results in a high rate of modification either in somatic cells or in embryos. With the use of these gene scissors, a large number of genetically modified pigs have been generated through SCNT of modified somatic cells or direct microinjection of engineered nucleases into the embryos. In addition to establishment of genetically modified pigs with agricultural and biomedical values, this technology might have a potentially wider range of application in pigs, such as treatment of viral infections as a therapeutic tool, and gene therapy to correct mutation in pig disease model. These areas still await further investigation.

NUCLEASE-BASED GENE EDITING TOOLS

Zinc Finger Nucleases (ZFNs)

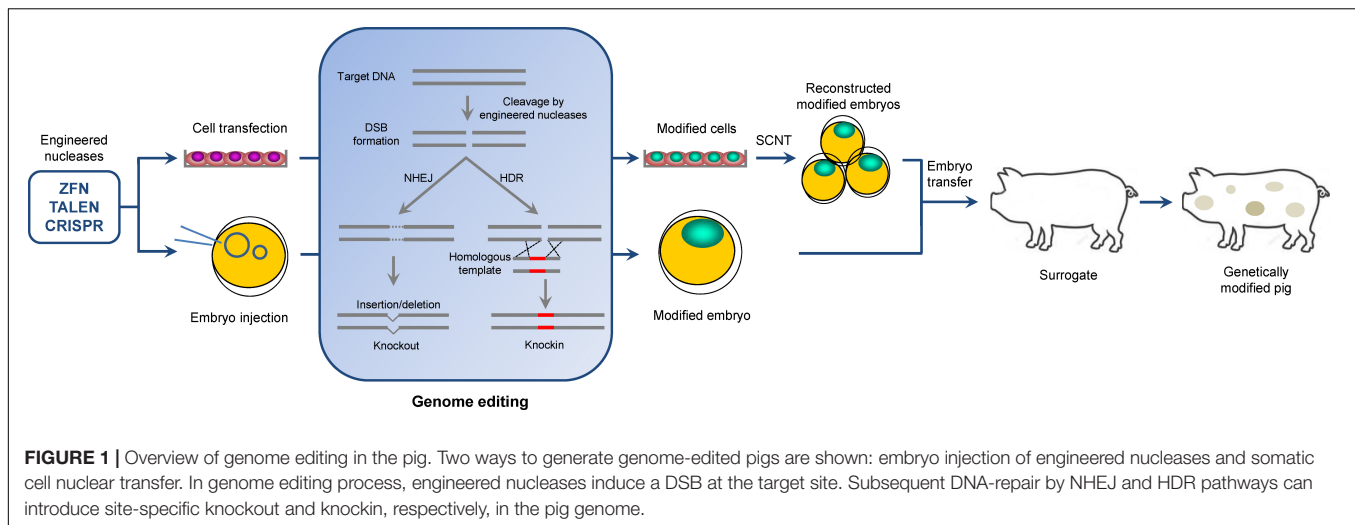
Zinc Finger Nucleases are artificial chimeric proteins consisting of a specific DNA-binding domain, which comprises tandem zinc finger-binding motifs, fused to a non-specific cleavage domain of the restriction endonuclease FokI (Kim et al., 1996; Urnov et al., 2010). Zinc finger protein characteristically consists of two

beta sheets and an alpha helix, with one or more coordinated zinc ions at their core to confer rigidity to finger (Pavletich and Pabo, 1991). Given that a zinc finger protein unit recognizes 3 bp of DNA, usually in ZFN, 3–6 zinc finger units are combined to recognize 9–18 bp DNA sequences to achieve a specific targeting. By designing two zinc finger motifs recognizing either side of 5–6 bp spacer sequences at a target region, FokI nuclease combined with zinc finger can introduce DSBs within a target region (Kim et al., 1996; Smith et al., 1999; Bibikova et al., 2003; Porteus and Baltimore, 2003; Urnov et al., 2010).

Being the early version of artificial engineered nucleases, ZFN opened a new possibility for gene targeting manipulation in pigs, although this technology still suffers from a complicated construction process and unpredictability of targeting activity. In general, a rational design and assembly of ZFN is somewhat a tough task for many laboratories (Klug, 2010; Lam et al., 2011; Chandrasegaran and Carroll, 2016). An effective ZFN reagent can only be obtained from some commercial sources at a prohibitive price or laboratories that embark in intensive work on ZFN. To the best of my knowledge, studies creating ZFN-mediated genetically modified pigs all use commercially synthesized ZFN reagents. The difficulty of generating active ZFN reagents has impeded their extensive use. Nevertheless, compared with the extremely low gene targeting efficiency of less than 10^{-6} by using conventional HR in somatic cells, ZFN can achieve approximately 1–4% gene targeting rate in selection of modified pig somatic cells (Hauschild et al., 2011; Yang et al., 2011), thereby allowing the cost-effective generation of genetically modified pigs.

Transcription Activator-Like Effector Nucleases (TALENs)

Transcription activator-like effector nucleases (TALENs) actually have a similar conceptual structure to ZFNs by comprising a DNA binding domain and a DNA cleavage domain and by acting in pairs to satisfy the requirement for dimerization (Christian et al., 2010; Miller et al., 2011). The DNA binding domain of TALENs, named transcription activator-like effector (TALE), originates from the plant-pathogenic bacterium *Xanthomonas* and includes tandem repeat modules of 34 amino acids, with each module specifying the binding to a single base pair. The repeat modules can be rearranged according to a simple cipher to target any DNA sequence (Boch et al., 2009; Moscou and Bogdanove, 2009). Unlike ZFN, TALEN reagents are easy to build by several assembly schemes and can be produced routinely for many laboratories (Cermak et al., 2011; Li et al., 2011; Morbitzer et al., 2011). Aside from their simple design and assembly, TALENs have a broad target range and a substantially improved targeting activity; thus, active TALENs may be designed for almost any DNA target in a genome. About 64% of synthesized TALENs are active in livestock fibroblasts. Three-quarters of these active TALENs demonstrate a high cleavage efficiency (19–40% NHEJ rates). Moreover, TALEN pairs efficiently induce gene knockouts after direct injection of TALEN-encoding mRNA into the cytoplasm of swine and bovine embryos, with a 29% and 43–75% knockout efficiency, respectively (Carlson et al., 2012).



Given that HR-mediated gene knockin results in precise alteration, including point mutation, DNA fragment replacement, and insertion of a new DNA sequence in a target site, knockin modification has a wide range of application prospect compared with the NHEJ-mediated gene knockout, which only causes loss in the functional phenotype of a gene. However, most of the created DSBs were repaired by NHEJ, and only few of them could be repaired by HR when a donor DNA repair template was offered (Mao et al., 2008). Therefore, HR-mediated knockin manipulation is less efficient than NHEJ-mediated knockout even in the presence of engineered nucleases. The high effectiveness of TALEN-mediated gene targeting could achieve an effective gene knockin. Using TALENs and oligonucleotide donor transfection that introduces defined nucleotide changes into multiple targets in the genome of livestock fibroblasts, Tan et al. achieved 10–64% knockin cell colonies, with up to 32% of the homozygous knockin colonies in just one round of transfection (Tan et al., 2013). Another study has established Rosa26 knockin pig models by using TALEN plasmid and long-range arm donor DNA transfection. In this study, the knockin efficiency in selected fibroblasts was as high as 31.3% (60 positive colonies of 192 selected fibroblast colonies) (Li et al., 2014).

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-Associated Protein (Cas)

Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein (Cas), originally known as a microbial adaptive immune system, has been adapted for mammalian gene editing recently. The CRISPR/Cas system is based on an adaptive immune mechanism in bacteria and archaea to defend the invasion of foreign genetic elements through DNA or RNA interference (Gasiunas et al., 2012; Jinek et al., 2012; Wiedenheft et al., 2012). Through mammalian codon optimization, CRISPR/Cas has been adapted for precise

DNA/RNA targeting and is highly efficient in mammalian cells and embryos. The most commonly used and intensively characterized CRISPR/Cas system for genome editing is the type II CRISPR system from *Streptococcus pyogenes*; this system uses a combination of Cas9 nuclease and a short guide RNA (gRNA) to target specific DNA sequences for cleavage. A 20-nucleotide gRNA complementary to the target DNA that lies immediately 5' of a PAM sequence (NGG) directs Cas9 to the target DNA and mediates cleavage of double-stranded DNA to form a DSB (Cong et al., 2013; Mali et al., 2013). Thus, CRISPR/Cas9 can achieve gene targeting in any N20-NGG site.

Since CRISPR/Cas9 first emerged, researchers have been highly impressed by its incomparable gene targeting efficiency and simple construction of customized vectors compared with previous site-specific nucleases. These characteristics render CRISPR/Cas9 highly accessible for almost any laboratories, thereby significantly contributing to the progress of research in many areas of biomedicine and agriculture. The high cleavage activity of CRISPR/Cas9 allows simultaneous targeting of multiple loci in a single cell within a single reaction (Cong et al., 2013). GGTA1/iGb3S and GGTA1/CMAH double knockout and GGTA1/iGb3S/CMAH triple knockout pigs, which have potential for xenotransplantation, were created by a single-step transfection of multiplexed sgRNA and Cas9 nuclease together with a single nuclear transfer (Li et al., 2015). We generated homozygous Pink1/Parkin double knockout pigs as Parkinson's disease models through a single transfection of CRISPR/Cas9 and SCNT. The frequency of selecting homozygous double knockout fibroblast colonies could reach up to 38.1% (Zhou et al., 2015).

GENETICALLY MODIFIED PIGS FOR AGRICULTURAL APPLICATION

Traditional selective breeding has produced a series of superior livestock varieties that demonstrate a dramatically enhanced production performance compared with their

original counterparts. However, individual trait shows only 0.5–3.0% genetic response to selection per year, and some traits such as fertility and disease resistance remain difficult to improve (Clark and Whitelaw, 2003). Furthermore, when production traits have been improved to a certain degree, their further optimization would even be more difficult and would need an even longer breeding cycle to achieve even a slight progress. Genome editing offers an alternative approach to rapidly and directly realize genetic improvement in livestock (Ruan et al., 2017). Fully improvement of individual and even multiple traits can be accomplished within only one generation. Importantly, by using genome editing, we can confer upon animals favorable genetic traits that are unavailable in natural genetic sources, thereby generating novel livestock varieties that cannot be achieved through traditional breeding.

Meat Production

Myostatin (MSTN) is a negative regulating factor of skeletal muscle mass *in vivo*. MSTN knockout mice exhibit a two–threefold increase in muscle mass due to muscle fiber hyperplasia and hypertrophy (McPherron et al., 1997). Natural mutations of MSTN have been found in several species, including cattle (Grobet et al., 1997; Kambadur et al., 1997; McPherron and Lee, 1997), sheep (Clop et al., 2006), and dogs (Mosher et al., 2007). MSTN modification is an effective approach to enhance muscle growth in various animals. Given that this gene is a “hot” candidate that is possibly beneficial in agriculture, several strains of MSTN-knockout pigs have been generated by using ZFNs (Qian et al., 2015), TALENs (Rao et al., 2016), and CRISPR/Cas9 (Wang et al., 2015, 2017b; Bi et al., 2016). These MSTN-knockout pigs demonstrate muscle hypertrophy or double-muscling (DM) phenotype, with increased muscle mass and decreased fat accumulation compared with wild-type (WT) pigs. One study showed some MSTN-mutant pigs with one extra thoracic vertebra (Qian et al., 2015). MSTN-knockout pigs are valuable breeding materials for rapid genetic improvement to produce lean meat from fat-type (indigenous) pig breeds.

However, some deleterious effects were found in some MSTN-null DM pigs. In homozygous MSTN knockout piglets of the Landrace breed, newborns have abnormal forelegs and/or hind legs and thus their motor function is severely impaired. The affected piglets usually die quickly after birth (Zou et al., 2018). MSTN possibly plays an important role in the development and function of muscles and other organs as MSTN is expressed from early embryogenesis through adulthood. It seems that MSTN modification could result in severe side effect in some pig breeds, albeit not observed in all reported MSTN-knockout pigs. Therefore, a careful selection of pig breeds which show an increased meat production but minimized collateral damage by MSTN modification is needed for establishment of MSTN-knockout DM pig breeds. A new candidate target (FBXO40) that influences muscle production has been found recently. FBXO40 knockout pigs display a muscle hypertrophy phenotype and survive normally without detectable pathological changes in major organs (Zou et al., 2018). Also of note

is that the DM animals are more susceptible to respiratory disease, lameness, stress and dystocia, thus requiring extra attention in husbandry to preserve animal welfare (Fiems, 2012).

Viral Resistance

Porcine reproductive and respiratory syndrome virus (PRRSV) is the most economically important swine disease worldwide, currently causing huge economic losses in the swine industry. Genome editing shed light on the establishment of PRRSV-resistant pigs through knockout of viral receptors in pigs. Potential PRRSV entry mediators, SIGLEC1 and CD163 were knocked out in pigs through conventional HR and CRISPR/Cas9, respectively (Prather et al., 2013; Whitworth et al., 2014). PRRSV challenge in knockout pigs demonstrated that SIGLEC1 is unnecessary for infectivity (Prather et al., 2013), whereas CD163 is the definitive receptor for PRRSV. CD163 knockout pigs are fully resistant to PRRSV challenge with no obvious PRRSV-related symptom and no detectable PRRSV antibody and RNA in the serum (Whitworth et al., 2016). Furthermore, multiple genotypes of CD163 modifications were achieved by using CRISPR/Cas9 in different laboratories; these modifications include CD163 total knockout (CD163 null) (Yang et al., 2018), domain swap of CD163 exon 7 (corresponding to SRCR 5, the PRRSV binding domain at the protein level) with human CD163L-1 exon 11 (chimeric CD163) (Wells et al., 2017), and CD163 truncation with a deletion of exon 7 (Δ SRCR5 CD163) (Burkard et al., 2017). Among these CD163-modified pigs, those with CD163 null phenotype were completely resistant to both Type 1 and Type 2 PRRSV isolates (Wells et al., 2017). The chimeric CD163 phenotype was resistant to Type 1 PRRSV but still supported the replication of the Type 2 virus (Wells et al., 2017). Our group also generated CD163 knockout pigs with null phenotype, and they demonstrated their complete resistance to the highly pathogenic PRRSV, which is the dominant circulating strain in China and other Asian countries (Yang et al., 2018).

Thermoregulation

Due to their lack of a functional UCP1 gene, pigs lack brown adipose tissue (BAT). As a result, the BAT-mediated adaptive non-shivering thermogenesis is absent in pigs (Trayhurn et al., 1989; Berg et al., 2006; Jastroch and Andersson, 2015). Newborn piglets are thus susceptible to cold stress, which may result in neonatal death. To address this issue, Zheng et al. inserted a mouse adiponectin-driven UCP1 into the porcine endogenous UCP1 locus by using a CRISPR/Cas9-mediated knockin strategy combined with SCNT. The resultant UCP1 knockin pigs showed an improved ability to maintain body temperature when acutely exposed to cold. Moreover, UCP1 prevented obesity by reducing fat deposition, an economically important trait targeted in pig breeding. UCP1 knockin pigs demonstrated reduced fat deposition through UCP1-promoted lipolysis (Zheng et al., 2017). Thus, UCP1 knockin pigs are a potentially valuable genetic resource for agricultural production on the basis of their improved thermoregulation and decreased fat deposition.

Recipients for Spermatogonial Stem Cell (SSC) Transplantation

Spermatogonial stem cells sustain normal spermatogenesis and maintain male fertility through self-renewal and differentiation. SSCs can be transplanted to generate donor-derived offspring (Brinster and Zimmermann, 1994). From the agricultural perspective, SSC transplantation is a potential tool to rapidly expand the availability of gametes from desirable superior livestock, thereby dramatically influencing production efficiency, quality, and other production traits in a population (Ehmcke et al., 2006). A recipient male lacking endogenous SSC and other germ cells but preserving intact somatic support cells is required for a successful SSC transplantation. Although chemotoxic drug or irradiation was employed to destroy spermatogenesis and cause infertility, the outcome was not ideal as manifested by the incomplete elimination of endogenous germ cells or the severe side effect on the recipient animals (Oatley, 2017; Park et al., 2017). An approach to eliminate SSC by knockout of the gene essential for SSC development has established an ideal surrogate for SSC transplantation. Park et al. generated NANOS2 knockout pigs by directly injecting CRISPR reagents into the cytoplasm of embryos. Knockout males could not produce sperm but still kept intact seminiferous tubules structure, thus had the potential to serve as an ideal SSC recipient (Park et al., 2017).

GENETICALLY MODIFIED PIGS AS DISEASE MODELS

Genome editing has extensively and intensively promoted the application of pigs as human disease models. Changing the pig genome allows these animals to resemble the mutations causing genetic disorder in humans, and pigs could phenocopy human disease manifestations more accurately than the commonly used mouse models. The suitable size and long lifespan of pig disease models also facilitates carrying out surgical manipulation closer to clinical conditions and long-term tracking and evaluation of therapeutics over clinically relevant time frames. Using the genome editing tools, genomic changes could occur not only in a single gene but also in multiple genes simultaneously with a high editing efficiency, thereby paving a way to mimic and decipher complex polygenic heredity diseases in large animal models similar to humans.

Neurodegenerative Diseases

Genetically modified pigs have been successfully used to establish animal models of neurodegenerative diseases, including Huntington's disease (HD) and Parkinson's disease (PD) (Zhou et al., 2015; Yan et al., 2018). Yan et al. reported a huntingtin (HTT) knockin pig as Huntington's disease model, in which pig pig HTT exon 1 containing 18 CAG repeats was replaced with human HTT exon 1 containing a 150-CAG repeat with CRISPR/Cas9. An expanded CAG repeat causes the HD phenotype (Mangiarini et al., 1996). The cloned HTT knockin pigs showed less weight gain compared with age- and sex-matched WT pigs, and HD-like symptoms, including deficient

motor function and respiratory difficulty usually observed in HD patients. Morphological analysis revealed neuropil and nuclear HTT aggregates in the brain of knockin pigs, as well as a marked decrease in the number of neurons in the striatum compared with that in the cortex and cerebellum; this condition is similar to the selective neurodegeneration in the striatum of brains of HD patients (Yan et al., 2018).

Parkinson's disease is the second most common form of neurodegenerative disorders. Approximately 10% of PD are familial cases with many disease-associated/caused genetic mutations were identified (Lesage and Brice, 2009). Zhou et al. (2015) reported on Pink1/Parkin double knockout pigs generated using CRISPR/Cas9 in a one-step transfection and cloning. The brains of cloned knockin pigs lost the expression of Pink1 and Parkin, but no PD-associated phenotypic changes were found (Zhou et al., 2015). These pigs are expected to show late onset of the disease because PD is a progressive disease with a mean onset at around the age of 60 in humans.

Cardiovascular Diseases

Cardiovascular diseases are the number one cause of death and disability worldwide (Joseph et al., 2017). As the most commonly used disease models, mice usually cannot accurately model the physiology and pathology of the human cardiovascular system given their significantly different heart size and rate compared with humans (Milani-Nejad and Janssen, 2014). Therefore, mouse models usually fail to predict human outcomes. For example, thiazolidinediones (TZDs), selective ligands of PPAR- γ , can sensitize insulin to treat type 2 diabetes. Although TZD-mediated activation of PPAR- γ exerts a beneficial effect on cardiovascular diseases in mouse models, the therapeutic efficacy of TZDs has been severely compromised because of the increased risk of adverse cardiovascular events in patients (Nissen and Wolski, 2007). Pig and human myocardia are highly similar. Thus, pigs are the most attractive models bridging the gap between humans and mice. Yang et al. (2011) established PPAR- γ heterozygous knockout pigs by using ZFN. Their work was the first to use the genome editing tool to knockout an endogenous gene of large animals (Yang et al., 2011).

The loss of function of low-density lipoprotein receptor (LDLR) and Apolipoprotein E (ApoE) has been implicated in the progression of atherosclerosis, the primary culprit for cardiovascular diseases (Mahley, 1988; Hasler-Rapacz et al., 1998; Sehaye et al., 2000; Rader et al., 2003). Given that the lipoprotein profiles and metabolism in mice differ from those in humans, atherosclerosis pig models may benefit atherosclerosis research. ApoE/LDLR double knockout pigs have been recently established by using CRISPR/Cas9 (Huang et al., 2017). These pigs show an abnormal lipid metabolism related to atherosclerosis. Prior to that, several strains of LDLR knockout pigs, including Yucatan miniature pigs engineered using the traditional HR (Davis et al., 2014) and Ossabaw pigs engineered by using TALENs (Carlson et al., 2012), were generated. Among the models, LDLR knockout Yucatan miniature pigs, which are used to model atherosclerosis, have been well characterized. LDLR-deficient pigs had considerably elevated levels in total and LDL cholesterol

fed a standard diet, resulting in atherosclerotic lesions in the coronary arteries and abdominal aorta that resemble human familial hypercholesterolemia and atherosclerosis in a short time (Davis et al., 2014). Recently, LDLR-deficient Yucatan pigs were used to test the effect of bempedoic acid on lowering LDL cholesterol and in attenuating atherosclerosis, indicating the value of LDLR-deficient pigs in preclinical evaluation of therapeutics (Burke et al., 2018).

Cancer

Acquired mutations are the most common causes of cancer. Instability of genome, activation of oncogenes, and inactivation of tumor-suppressor genes all can result in different types of cancer. An optimal cancer animal model should be established by inducing genetically defined tumors in a tissue-specific manner rather than by genetically engineering the animal. Wang et al. (2017a) established a Cre-dependent inducible Cas9-expression pig with CRISPR/Cas9-mediated knockin in the Rosa26 locus. This pig model allows *ex vivo* genome editing in isolated pig cells and *in vivo* genome editing by introducing a corresponding gRNA. To mimic an abnormal oncogenic EML4-ALK fusion gene identified in a subset of non-small-cell lung cancers (NSCLCs) (Soda et al., 2007), two gRNAs targeting EML4 and ALK were introduced into the isolated pig fibroblasts via lentivirus. Fibroblasts with oncogenic EML4-ALK fusion gene arising through a paired CRISPR/Cas9-mediated genome inversion could be generated. Moreover, the study investigated the feasibility of lung cancer induction *in vivo* through intranasal delivery of multiplexed gRNAs targeting tumor suppressor genes TP53, PTEN, APC, BRCA1, and BRCA2, as well as oncogene KRAS. Three months after gRNA administration, the Cre-dependent Cas9-expressing pigs presented signs of pneumonopathy, and morphological analysis of lung tissue showed a pathological feature similar to the human adenocarcinoma. The target genes harbored insertions/deletions close to the CRISPR-cleavage site, indicating a loss-of-function mutation in the tumor suppressor genes. For the oncogene KRAS, gain-of-function mutations induced by CRISPR/Cas9 were observed, with genotypes similar to that in the potent oncogenic mutations in human lung tumors (Wang et al., 2017a). The inducible Cas9 pig models offer an ideal platform for inducing tumor-associated somatic mutations *in situ* to model human cancer. Apart from this model, several other pig cancer models were produced, including knockout of tumor suppressors P53 and RUNX3 by TALENs and CRISPR/Cas9 as the germline mutations, respectively (Kang et al., 2016; Shen et al., 2017). The carcinogenic phenotypes of these animals require further investigation.

Immunodeficiency

An animal with severe combined immune deficiency (SCID) not only mimics human diseases but also serves as a valuable research tool for cancer, stem cell, cell therapy, and organ transplantation. The interleukin-2 receptor gamma (IL2RG) knockout pigs that were generated using conventional HR or ZFN exhibited X-linked SCID, in which T and NK cells were

absent (Suzuki et al., 2012; Watanabe et al., 2013). In generation of ZFN-mediated IL2RG knockout pigs, ZFN-encoding mRNA was transfected into male porcine fibroblasts to target IL2RG. The researchers obtained 1 ZFN-induced knockout cell line from 192 single cell-derived cell lines obtained by limiting dilution (0.5% targeting efficiency) (Watanabe et al., 2013). In the allogeneic bone marrow transplantation, the lymphoid lineage of the SCID pigs was reconstituted by donor cells and survival time was prolonged through restoration of immune function (Suzuki et al., 2012). This SCID pig serves as an essential preclinical model to evaluate stem cell therapy. In 2014, two laboratories separately established RAG1/2 knockout pigs by using TALENs (Huang et al., 2014; Lee et al., 2014). RAG knockout pigs exhibited an SCID phenotype and lacked T and B cells. In one study, human induced pluripotent cells were injected into these SCID pigs, which developed teratomas that represent a wide range of human tissues.

In another study, B cell-deficient pigs were generated by applying CRISPR/Cas9 to target the IgM heavy chain gene, which is crucial in B cell development and differentiation (Chen et al., 2015). The cloned modified pigs manifested a depletion of both B cells and antibody in their blood and thus can be used to model human B cell deficiency. Moreover, the pig models can be further engineered for large-scale production of therapeutic humanized polyclonal antibodies for clinical use.

GENETICALLY MODIFIED PIGS FOR XENOTRANSPLANTATION

Domestic pigs are the most suitable donors for xenotransplantation to alleviate the growing shortage of allogeneic donor organs for clinical transplantation to treat patients with end-stage organ failure. Pigs can provide size-matched organs at an affordable cost. However, the high immune incompatibility between the donor and the recipient is the major barrier for xenotransplantation. The immunological hurdles to xenotransplantation include but not limited to hyperacute rejection, delayed xenograft rejection, acute cellular rejection, and chronic rejection (Yang and Sykes, 2007; Hryhorowicz et al., 2017). Advances in genome editing enable genetic modifications in pigs to reduce cross-species immune barrier and prevent xenograft rejection.

The host normally destroys xenografts within minutes or hours through hyperacute xenograft rejection, which inevitably leads to failure. The main reason for hyperacute rejection of xenograft is the presence of naturally occurring antibodies in human plasma that recognizes the Galactose $\alpha(1-3)$ Galactose (Gal $\alpha(1-3)$ Gal) antigen on the surface of porcine endothelial cells. Synthesis of Gal $\alpha(1-3)$ Gal is catalyzed by the enzyme α -1,3-galactosyltransferase (GGTA1), which is present in pigs but absent in humans (Good et al., 1992; Galili, 1993). As early as 2002, pigs with heterozygously knockout GGTA1 were produced using HR (Dai et al., 2002; Lai et al., 2002), and homozygous GGTA1 knockout (GTKO) piglets were born by further screening of homozygous knockout cells from the heterozygous pigs (Phelps et al., 2003;

Kolber-Simonds et al., 2014). Heart xenotransplantation from GTKO pigs into immune-suppressed baboons showed a mean graft survival period of 99 days, and the longest surviving graft functioned in the recipient for 179 days (Kuwaki et al., 2005). Generation of GTKO pigs indicates that hyperacute xenograft rejection has been overcome to a great extent. A series of GTKO pigs with different genetic backgrounds has been recently established by using genome editing tools (Hauschild et al., 2011; Xin et al., 2013; Bao et al., 2014; Feng et al., 2016; Petersen et al., 2016; Chuang et al., 2017). Furthermore, other xenoreactive antigens present in pigs but absent in humans have been identified; these antigens include Neu5Gc antigen (*N*-glycolylneuraminic acid) catalyzed by cytidine monophosphate-*N*-acetylneuraminic acid hydroxylase (CMAH) and a glycan produced by β 1,4-*N*-acetylgalactosaminyltransferase (B4GALNT2) (Byrne et al., 2015, 2018). In addition, corresponding knockout pigs with double (GGTA1/CMAH) and triple (GGTA1/CMAH/B4GALNT2) gene inactivation were established (Lutz et al., 2013; Estrada et al., 2015; Miyagawa et al., 2015; Gao et al., 2017). Based on the GTKO pigs, numerous other genetic modifications were performed to further overcome xenogeneic barriers; these approaches mainly include individual or combined overexpression of transgenes, such as CD46, CD55, CD59, CD39, thrombomodulin, heme oxygenase 1, A20, HLA-E, and CD47, to prevent complement activity, delayed xenograft rejection, or/and acute cellular rejection (Fischer et al., 2016; Hryhorowicz et al., 2017; Laird et al., 2017). These pigs were produced through transgenesis and thus excluded in this review for a deep discussion. These multi-modified pigs may exhibit further immune tolerance to attenuate xenograft injury.

Besides the concern on pig-to-human immune barrier, another notable issue in xenotransplant is the risk of cross-species transmission of porcine endogenous retroviruses (PERVs), which are dormant (inactive) endogenous retroviruses constituting an integral part of the porcine genome; PERVs may be reactivated by certain factors or changes in the environment and thus become infectious (Patience et al., 1997; van der Laan et al., 2000). PERVs in xenografts possess the potential to become pathogenic and infectious in the recipient. Elimination of PERV in porcine genome is difficult because they are integrated in multiple locations in the genome. Yang et al. (2015) used CRISPR/Cas9 to disrupt all 62 copies of the PERV pol gene in the immortal pig cell line PK15; by using the PERV knockout cells, they demonstrated a >1000-fold reduction in PERV transmission to human cells (Yang et al., 2015). The same group has recently cloned PERV inactivated pigs combining CRISPR/Cas9 and SCNT, in which all 25 copies of functional PERVs were inactivated (Niu et al., 2017). The use of PERV knockout pigs addressed the safety concern in clinical xenotransplantation. Moreover, the powerful ability of the CRISPR tool to target dozens of genomic sites simultaneously provides infinite possibility to create animals harboring complex modifications.

In this section, we present an emerging cutting-edge technology, that is, growing of humanized organ in pigs or called xeno-generation; this technology is greatly promoted with the use of genome editing. Compared with directly modifying pig as

an organ donor, xeno-generation includes interspecies blastocyst complementation combining donor (human) pluripotent stem cells and organogenesis-disabled hosts (pigs), allowing the enrichment of donor cells in a target organ to form a chimeric animal harboring a humanized target organ, which can be finally used for transplantation (Wu et al., 2016). In this regard, organogenesis-disabled pigs can be simply realized by genome editing when the gene controlling the development of the target organ is identified. Matsunari et al. (2013) demonstrated the feasibility of isogenic organ generation using blastocyst complementation in pigs. In which, the transgene *Hes1* under the *Pdx1* promoter (*Pdx1*-*Hes1*) was used to suppress pancreatic development, resulting in the creation of apaucipancic pigs compatible for pancreatogenesis derived from donor cells. Wu et al. (2017a) reported on an interspecies chimerism that human pluripotent stem cells could integrate and differentiate in a pig embryo, constituting a big step toward xenogeneic organ generation. Furthermore, this group disabled pancreatogenesis in pigs through knockout of *PDX1* using CRISPR/Cas9, creating a suitable platform for realizing human organogenesis in pig (Wu et al., 2017b).

CONCLUSIVE THOUGHTS

Comparison of Three Engineered Nucleases

Engineered nuclease-based genome editing has revolutionized the creation of genetically modified pigs, thus expanding their utilization in diverse research fields. As the prelude of next-generation genome editing technology, ZFN surprised researchers for opening an effective approach to modify a target genome site. The gene targeting efficiency of ZFN, although limited, remains considerably higher than that of traditional HR. TALEN quickly overtook ZFN technology when it first appeared due to its higher efficiency in gene targeting, greater flexibility in targeting specific sequences, and ease of construction. Moreover, the emergence of CRISPR/Cas system represents a major leap for remarkably efficient specific genetic modification in mammalian cells and zygotes. In addition, the design and production of gRNA are through a quick and simple procedure by the *in vitro* transcription of synthetic DNA oligonucleotides or by the cloning of oligonucleotides into expression vectors; this process offers a clear advantage over the production of ZFN and TALEN. However, few concerns emerged in CRISPR/Cas, such as requirement of a PAM sequence adjacent to the 3' end of the target sequence and a high frequency of off-target cleavage. The requirement of a specific PAM sequence often restricts the range of targetable sequences. The commonly used Cas9 protein derived from *S. pyogenes* utilizes NGG as the PAM sequence, but recent exploitation of Cas9 orthologs from other bacterial species or redesigned/evolved Cas9 can recognize different PAM sequences, thereby increasing flexibility in genome editing. A study reported on the fusion of Cas9 with a programmable DNA-binding domain at an improved precision and increased targeting range. The Cas9 fusion protein was equally efficient for a range of PAMs, including NAG, NGA,

NGC, and NGG (Bolukbasi et al., 2015). Another study used phage-assisted continuous evolution to generate an expanded PAM SpCas9 variant that can recognize a broad range of PAM sequences, including NG, GAA, and GAT. Moreover, the SpCas9 variant demonstrated a considerably greater DNA specificity than the original SpCas9 (Hu et al., 2018). The current Cas9 family expands the DNA targeting scope of CRISPR systems and is suitable at nearly any genomic locus. With regard to off-target, a typical TALEN target sequence usually covers 30 nt, which is unique within the genome, whereas CRISPR/Cas recognizes 20 nt target and can allow multiple mismatches in the guide sequence, thereby increasing the likelihood of off-target effects (Fu et al., 2013; Hsu et al., 2013). A possible solution is the use of Cas-nickases guided by a pair of gRNAs targeting the opposite strands for cooperative genome editing, as the longer target site increases the on-target precision (Mali et al., 2013; Ran et al., 2013). In addition, various strategies have been developed to improve the targeting specificity; these strategies include optimized design of gRNA, Cas9 enzyme engineering, and off-target detection assays (Tycko et al., 2016). Notably, two recent engineered enzymes (eSpCas9 and SpCas9-HF) have elegantly increased SpCas9 specificity by reducing tolerance for mismatched DNA binding (Kleinstiver et al., 2016; Slaymaker et al., 2016).

Approaches in Pig Genome Editing: SCNT Versus Embryo Injections

Somatic cell nuclear transfer or cloning involves screening of somatic cells (typically fetal fibroblasts), which carry the intended genetic alterations, and the nuclear transfer of the modified cells in a cloning process. Engineered nuclease can be easily applied to create either NHEJ- or HDR-induced mutations within a donor cell *in vitro* through a pre-screening or selection strategy, which enables enrichment of cells carrying the desired mutation (Zhou et al., 2015). An alternative to SCNT is the method involving direct gene editing in single-cell embryos. The mRNA of editors (for knockout) or together with donor DNA (for knockin) can be microinjected into the cytoplasm or pronucleus of zygotes, which are then transferred into the synchronized surrogates to generate edited animals. This procedure is vastly simple compared with SCNT (Lillico et al., 2013).

The major advantage of SCNT over direct embryo injection is the predictable genotype of the founder pigs. By contrast, pigs generated via embryo injection usually contain mosaic genotypes with multiple modification types in different cells, and several cycles of breeding are usually needed to produce homozygously modified pigs with identical genotype. In some situations, the chimeric founder includes intact WT germ cells and thus cannot generate genetically modified progeny. However, SCNT of genome-edited cells suffer from an impaired embryonic development probably due to the off-target effect or other unidentified toxicity of the genome editing nucleases. Therefore, genome-edited somatic cells usually have a relatively lower cloning efficiency than WT and randomly integrated transgenic cells. Taking into account the aspects

mentioned above, embryos injection is currently the preferable ways for production of CRISPR/Cas9-edited pigs in many laboratories. The extremely high editing activity and low toxicity to embryos during microinjection of CRISPR/Cas9 mRNA cocktail can reduce mosaicism and even generate homozygously knockout founder pigs with identical mutation (Whitworth et al., 2017). However, knockin manipulation remains a limitation because few knockin pigs generated through zygote injection have been reported to date (Peng et al., 2015; Zhou et al., 2016). The field still awaits for the optimal strategies that enhance HDR, such as drug treatment, optimal design of donor, and new generation of editing tools.

Other Emerging Genome Editing Nucleases

The number of reported genetically unique large animals dramatically increases as a result of the extensive use of genome editing tools. Moreover, the types of endonucleases used in genome editing are rapidly increasing. In the Cas9 system, many Cas9-like nucleases were developed given the natural diversity of bacterial CRISPR systems. Cpf1, a putative Class 2 CRISPR effector, mediates target DNA editing with distinct features from Cas9 (Zetsche et al., 2015). In contrast to Cas9 which generates blunt ends, Cpf1 generates a 5-nt staggered cut with a 5' overhang, which is particularly advantageous in facilitating a NHEJ-based gene insertion (knockin) into a genome. Recently, the CRISPR/Cpf1-mediated dystrophin knock-out pigs, and phospholamban knock-in pigs with a 3-nt deletion in the presence of a single-stranded oligo donor, have been established as a Duchenne muscular dystrophy (DMD) and dilated cardiomyopathy (DCM) models, respectively. In this study, the CRISPR/Cpf1 induced 41.8% knockout rate and 2% knockin rate in the selected fibroblast colonies (Wu et al., 2018). A hybrid enzyme combining the Cas9-nickase and PmCDA1, an activation-induced cytidine deaminase could perform targeted nucleotide substitution (C-U) without the use of template DNA, providing a novel route for point mutation (Nishida et al., 2016). A CRISPR system (Cas13a) that targets RNA has also been developed recently (Abudayyeh et al., 2017). In addition to the CRISPR system, Xu et al. (2016) designed a structure-guided endonuclease (SGN) consisting of flap endonuclease-1 that recognizes the 3' flap structure and the cleavage domain of Fok I, which cleaves DNA strands. A guide DNA complementary to the target with an unpaired 3' end is needed to form a 3' flap structure. SGN recognizes and cleaves the target DNA on the basis of the 3' flap structure of a double-flap complex formed between the target and the guide DNA. The SGN offers a strategy for a structure-based recognition, capture, and editing of any desired target DNA, thereby expanding the toolkit for genetic modification (Xu et al., 2016). Taken together, these systems are expected to substantially broaden the application of artificial engineered nucleases and facilitate the establishment of excellent pig models with desirable genotypes in agriculture and biomedicine.

AUTHOR CONTRIBUTIONS

HY wrote the initial draft of the manuscript and worked on subsequent revisions. ZW worked on revising the manuscript.

REFERENCES

- Abudayyeh, O. O., Gootenberg, J. S., Essletzbichler, P., Han, S., Joung, J., Belanto, J. J., et al. (2017). RNA targeting with CRISPR–Cas13. *Nature* 550, 280–284. doi: 10.1038/nature24049
- Bao, L., Chen, H., Jong, U., Rim, C., Li, W., Lin, X., et al. (2014). Generation of GGTA1 biallelic knockout pigs via zinc-finger nucleases and somatic cell nuclear transfer. *Sci. China Life Sci.* 57, 263–268. doi: 10.1007/s11427-013-4601-2
- Berg, F., Gustafson, U., and Andersson, L. (2006). The uncoupling protein 1 gene (UCP1) is disrupted in the pig lineage: a genetic explanation for poor thermoregulation in piglets. *PLoS Genet.* 2:e129. doi: 10.1371/journal.pgen.0020129
- Bi, Y., Hua, Z., Liu, X., Hua, W., Ren, H., Xiao, H., et al. (2016). Isozygous and selectable marker-free MSTN knockout cloned pigs generated by the combined use of CRISPR/Cas9 and Cre/LoxP. *Sci. Rep.* 6:31729. doi: 10.1038/srep31729
- Bibikova, M., Beumer, K., Trautman, J. K., and Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. *Science* 300:64. doi: 10.1126/science.1079512
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., et al. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509–1512. doi: 10.1126/science.1178811
- Bolukbasi, M. F., Gupta, A., Oikemus, S., Derr, A. G., Garber, M., Brodsky, M. H., et al. (2015). DNA-binding-domain fusions enhance the targeting range and precision of Cas9. *Nat. Methods* 12, 1150–1156. doi: 10.1038/nmeth.3624
- Brinster, R. L., and Zimmermann, J. W. (1994). Spermatogenesis following male germ-cell transplantation. *Proc. Natl. Acad. Sci. U.S.A.* 91, 11298–11302. doi: 10.1073/pnas.91.24.11298
- Burkard, C., Lillico, S. G., Reid, E., Jackson, B., Mileham, A. J., Ait-Ali, T., et al. (2017). Precision engineering for PRRSV resistance in pigs: macrophages from genome edited pigs lacking CD163 SRCR5 domain are fully resistant to both PRRSV genotypes while maintaining biological function. *PLoS Pathog.* 13:e1006206. doi: 10.1371/journal.ppat.1006206
- Burke, A. C., Telford, D. E., Sutherland, B. G., Edwards, J. Y., Sawyez, C. G., Barrett, P. H. R., et al. (2018). Bempedoic acid lowers low-density lipoprotein cholesterol and attenuates atherosclerosis in low-density lipoprotein receptor-deficient (LDLR^{+/−} and LDLR^{−/−}) Yucatan miniature pigs. *Arterioscler. Thromb. Vasc. Biol.* 38, 1178–1190. doi: 10.1161/ATVBAHA.117.310676
- Byrne, G., Ahmad-Villiers, S., Du, Z., and McGregor, C. (2018). B4GALNT2 and xenotransplantation: a newly appreciated xenogeneic antigen. *Xenotransplantation* doi: 10.1111/xen.12394 [Epub ahead of print].
- Byrne, G. W., McGregor, C. G. A., and Breimer, M. E. (2015). Recent investigations into pig antigen and anti-pig antibody expression. *Int. J. Surg.* 23, 223–228. doi: 10.1016/j.ijsu.2015.07.724
- Capecchi, M. R. (1989). Altering the genome by homologous recombination. *Science* 244, 1288–1292. doi: 10.1126/science.2660260
- Carlson, D. F., Tan, W., Lillico, S. G., Stverakova, D., Proudfoot, C., Christian, M., et al. (2012). Efficient TALEN-mediated gene knockout in livestock. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17382–17387. doi: 10.1073/pnas.1211446109
- Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., et al. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* 39:e82. doi: 10.1093/nar/gkr218
- Chandrasegaran, S., and Carroll, D. (2016). Origins of programmable nucleases for genome engineering. *J. Mol. Biol.* 428, 963–989. doi: 10.1016/j.jmb.2015.10.014
- Chen, F., Wang, Y., Yuan, Y., Zhang, W., Ren, Z., Jin, Y., et al. (2015). Generation of B cell-deficient pigs by highly efficient CRISPR/Cas9-mediated gene targeting. *J. Genet. Genomics* 42, 437–444. doi: 10.1016/j.jgg.2015.05.002
- Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., et al. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186, 757–761. doi: 10.1534/genetics.110.120717
- Chuang, C. K., Chen, C. H., Huang, C. L., Su, Y. H., Peng, S. H., Lin, T. Y., et al. (2017). Generation of GGTA1 mutant pigs by direct pronuclear microinjection of CRISPR/Cas9 plasmid vectors. *Anim. Biotechnol.* 28, 174–181. doi: 10.1080/10495398.2016.1246453
- Clark, J., and Whitelaw, B. (2003). A future for transgenic livestock. *Nat. Rev. Genet.* 4, 825–833. doi: 10.1038/nrg1183
- Clop, A., Marcq, F., Takeda, H., Pirottin, D., Tordoir, X., Bibé, B., et al. (2006). A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat. Genet.* 38, 813–818. doi: 10.1038/ng1810
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi: 10.1126/science.1231143
- Dai, Y., Vaught, T. D., Boone, J., Chen, S. H., Phelps, C. J., Ball, S., et al. (2002). Targeted disruption of the alpha1,3-galactosyltransferase gene in cloned pigs. *Nat. Biotechnol.* 20, 251–255. doi: 10.1038/nbt0302-251
- Davis, B. T., Wang, X. J., Rohret, J. A., Struzynski, J. T., Merricks, E. P., Bellinger, D. A., et al. (2014). Targeted disruption of LDLR causes hypercholesterolemia and atherosclerosis in Yucatan miniature pigs. *PLoS One* 9:e93457. doi: 10.1371/journal.pone.0093457
- Ehmcke, J., Hübner, K., Schöler, H. R., and Schlatt, S. (2006). Spermatogonia: origin, physiology and prospects for conservation and manipulation of the male germ line. *Reprod. Fertil. Dev.* 18, 7–12. doi: 10.1071/RD05119
- Estrada, J. L., Martens, G., Li, P., Adams, A., Newell, K. A., Ford, M. L., et al. (2015). Evaluation of human and non-human primate antibody binding to pig cells lacking GGTA1/CMAH/β4GalNT2 genes. *Xenotransplantation* 22, 194–202. doi: 10.1111/xen.12161
- Feng, C., Li, X., Cui, H., Long, C., Liu, X., Tian, X., et al. (2016). Highly efficient generation of GGTA1 knockout pigs using a combination of TALEN mRNA and magnetic beads with somatic cell nuclear transfer. *J. Integr. Agric.* 15, 1540–1549. doi: 10.1016/S2095-3119(16)61347-3
- Fiems, L. O. (2012). Double muscling in cattle: genes, husbandry, carcasses and meat. *Animals (Basel)* 2, 472–506. doi: 10.3390/ani2030472
- Fischer, K., Kraner-Scheiber, S., Petersen, B., Rieblinger, B., Buermann, A., Flisikowska, T., et al. (2016). Efficient production of multi-modified pigs for xenotransplantation by ‘combining’ gene stacking and gene editing. *Sci. Rep.* 6:29081. doi: 10.1038/srep29081
- Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., et al. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 31, 822–826. doi: 10.1038/nbt.2623
- Gaj, T., Gersbach, C. A., and Barbas, C. F. III (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31, 397–405. doi: 10.1016/j.tibtech.2013.04.004
- Galili, U. (1993). Interaction of the natural anti-Gal antibody with alpha-galactosyl epitopes: a major obstacle for xenotransplantation in humans. *Immunol. Today* 14, 480–482. doi: 10.1016/0167-5699(93)90261-I
- Gao, H., Zhao, C., Xiang, X., Li, Y., Zhao, Y., Li, Z., et al. (2017). Production of α1,3-galactosyltransferase and cytidine monophosphate-N-acetylneuraminic acid hydroxylase gene double-deficient pigs by CRISPR/Cas9 and handmade cloning. *J. Reprod. Dev.* 63, 17–26. doi: 10.1262/jrd.2016-079
- Garas, L. C., Murray, J. D., and Maga, E. A. (2015). Genetically engineered livestock: ethical use for food and medical models. *Annu. Rev. Anim. Biosci.* 3, 559–575. doi: 10.1146/annurev-animal-022114-110739
- Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2579–E2586. doi: 10.1073/pnas.1208507109
- Good, A. H., Cooper, D. K., Malcolm, A. J., Ippolito, R. M., Koren, E., Neethling, F. A., et al. (1992). Identification of carbohydrate structures that bind human antiporcine antibodies: implications for discordant xenografting in humans. *Transplant. Proc.* 24, 559–562.

FUNDING

This work was supported by a grant from the National Natural Science Foundation of China (31772555).

- Grobet, L., Martin, L. J., Poncelet, D., Pirotin, D., Brouwers, B., and Riquet, J. (1997). A deletion in the bovine myostatin gene causes the double-muscling phenotype in cattle. *Nat. Genet.* 17, 71–74. doi: 10.1038/ng0997-71
- Hasler-Rapacz, J., Ellegren, H., Fridolfsson, A. K., Kirkpatrick, B., Kirk, S., Andersson, L., et al. (1998). Identification of a mutation in the low density lipoprotein receptor gene associated with recessive familial hypercholesterolemia in swine. *Am. J. Med. Genet.* 76, 379–386.
- Hauschild, J., Petersen, B., Santiago, Y., Queisser, A. L., Carnwath, J. W., Lucas-Hahn, A., et al. (2011). Efficient generation of a biallelic knockout in pigs using zinc-finger nucleases. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12013–12017. doi: 10.1073/pnas.1106422108
- Hryhorowicz, M., Zeyland, J., Słomski, R., and Lipiński, D. (2017). Genetically modified pigs as organ donors for xenotransplantation. *Mol. Biotechnol.* 59, 435–444. doi: 10.1007/s12033-017-0024-9
- Hsu, P. D., Lander, E. S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262–1278. doi: 10.1016/j.cell.2014.05.010
- Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31, 827–832. doi: 10.1038/nbt.2647
- Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., et al. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 556, 57–63. doi: 10.1038/nature26155
- Huang, J., Guo, X., Fan, N., Song, J., Zhao, B., Ouyang, Z., et al. (2014). RAG1/2 knockout pigs with severe combined immunodeficiency. *J. Immunol.* 193, 1496–1503. doi: 10.4049/jimmunol.1400915
- Huang, L., Hua, Z., Xiao, H., Cheng, Y., Xu, K., Gao, Q., et al. (2017). CRISPR/Cas9-mediated ApoE^{-/-} and LDLR^{-/-} double gene knockout in pigs elevates serum LDL-C and TC levels. *Oncotarget* 8, 37751–37760. doi: 10.18632/oncotarget.17154
- Jasin, M. (1996). Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.* 12, 224–228. doi: 10.1016/0168-9525(96)10019-6
- Jastroch, M., and Andersson, L. (2015). When pigs fly, UCP1 makes heat. *Mol. Metab.* 4, 359–362. doi: 10.1016/j.molmet.2015.02.005
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829
- Joseph, P., Leong, D., McKee, M., Anand, S. S., Schwalm, J. D., Teo, K., et al. (2017). Reducing the global burden of cardiovascular disease, part 1: the epidemiology and risk factors. *Circ. Res.* 121, 677–694. doi: 10.1161/CIRCRESAHA.117.308903
- Kambadur, R., Sharma, M., Smith, T. P., and Bass, J. J. (1997). Mutations in myostatin (GDF8) in double-muscling belgian blue and piedmontese cattle. *Genome Res.* 7, 910–916. doi: 10.1101/gr.7.9.910
- Kang, J. T., Ryu, J., Cho, B., Lee, E. J., Yun, Y. J., Ahn, S., et al. (2016). Generation of RUNX3 knockout pigs using CRISPR/Cas9-mediated gene targeting. *Reprod. Domest. Anim.* 51, 970–978. doi: 10.1111/rda.12775
- Kim, Y. G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U.S.A.* 93, 1156–1160. doi: 10.1073/pnas.93.3.1156
- Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., et al. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529, 490–495. doi: 10.1038/nature16526
- Klug, A. (2010). The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu. Rev. Biochem.* 79, 213–231. doi: 10.1146/annurev-biochem-010909-095056
- Kolber-Simonds, D., Lai, L., Watt, S. R., Denaro, M., Arn, S., Augenstein, M. L., et al. (2014). Production of alpha-1,3-galactosyltransferase null pigs by means of nuclear transfer with fibroblasts bearing loss of heterozygosity mutations. *Proc. Natl. Acad. Sci. U.S.A.* 101, 7335–7340. doi: 10.1073/pnas.0307819101
- Kuwaki, K., Tseng, Y. L., Dor, F. J., Shimizu, A., Houser, S. L., Sanderson, T. M., et al. (2005). Heart transplantation in baboons using α 1,3-galactosyltransferase gene-knockout pigs as donors: initial experience. *Nat. Med.* 11, 29–31. doi: 10.1038/nm1171
- Lai, L., Kolber-Simonds, D., Park, K. W., Cheong, H. T., Greenstein, J. L., Im, G. S., et al. (2002). Production of alpha-1,3-galactosyltransferase knockout pigs by nuclear transfer cloning. *Science* 295, 1089–1092. doi: 10.1126/science.1068228
- Laird, C. T., Burdorf, L., French, B. M., Kubicki, N., Cheng, X., Braileanu, G., et al. (2017). Transgenic expression of human leukocyte antigen-E attenuates GalKO.hCD46 porcine lung xenograft injury. *Xenotransplantation* 24:e12294. doi: 10.1111/xen.12294
- Lam, K. N., van Bakel, H., Cote, A. G., van der Ven, A., and Hughes, T. R. (2011). Sequence specificity is obtained from the majority of modular C2H2 zinc-finger arrays. *Nucleic Acids Res.* 39, 4680–4690. doi: 10.1093/nar/gkq1303
- Lee, K., Kwon, D. N., Ezashi, T., Choi, Y. J., Park, C., Ericsson, A. C., et al. (2014). Engraftment of human iPS cells and allogeneic porcine cells into pigs with inactivated RAG2 and accompanying severe combined immunodeficiency. *Proc. Natl. Acad. Sci. U.S.A.* 111, 7260–7265. doi: 10.1073/pnas.1406376111
- Lesage, S., and Brice, A. (2009). Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Hum. Mol. Genet.* 18, R48–R59. doi: 10.1093/hmg/ddp012
- Li, P., Estrada, J. L., Burlak, C., Montgomery, J., Butler, J. R., Santos, R. M., et al. (2015). Efficient generation of genetically distinct pigs in a single pregnancy using multiplexed single-guide RNA and carbohydrate selection. *Xenotransplantation* 22, 20–31. doi: 10.1111/xen.12131
- Li, P., Huang, S., Zhao, X., Wright, D. A., Carpenter, S., Spalding, M. H., et al. (2011). Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res.* 39, 6315–6325. doi: 10.1093/nar/gkr188
- Li, X., Yang, Y., Bu, L., Guo, X., Tang, C., Song, J., et al. (2014). Rosa26-targeted swine models for stable gene over-expression and Cre-mediated lineage tracing. *Cell Res.* 24, 501–504. doi: 10.1038/cr.2014.15
- Lillico, S. G., Proudfoot, C., Carlson, D. F., Stverakova, D., Neil, C., Blain, C., et al. (2013). Live pigs produced from genome edited zygotes. *Sci. Rep.* 3:2847. doi: 10.1038/srep02847
- Lutz, A. J., Li, P., Estrada, J. L., Sidner, R. A., Chihara, R. K., Downey, S. M., et al. (2013). Double knockout pigs deficient in N-glycolylneuraminic acid and galactose α -1,3-galactose reduce the humoral barrier to xenotransplantation. *Xenotransplantation* 20, 27–35. doi: 10.1111/xen.12019
- Mahley, R. W. (1988). Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. *Science* 240, 622–630. doi: 10.1126/science.3283935
- Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., et al. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826. doi: 10.1126/science.1232033
- Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetherington, C., et al. (1996). Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* 87, 493–506. doi: 10.1016/S0092-8674(00)81369-0
- Mansour, S. L., Thomas, K. R., and Capecchi, M. R. (1988). Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* 336, 348–352. doi: 10.1038/336348a0
- Mao, Z., Bozzella, M., Seluanov, A., and Gorbunova, V. (2008). Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst)* 7, 1765–1771. doi: 10.1016/j.dnarep.2008.06.018
- Matsunari, H., Nagashima, H., Watanabe, M., Umeyama, K., Nakano, K., Nagaya, M., et al. (2013). Blastocyst complementation generates exogenic pancreas in vivo in pancreatic cloned pigs. *Proc. Natl. Acad. Sci. U.S.A.* 110, 4557–4562. doi: 10.1073/pnas.1222902110
- McPherron, A. C., Lawler, A. M., and Lee, S. J. (1997). Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. *Nature* 387, 83–90. doi: 10.1038/387083a0
- McPherron, A. C., and Lee, S. J. (1997). Double muscling in cattle due to mutations in the myostatin gene. *Proc. Natl. Acad. Sci. U.S.A.* 94, 12457–12461. doi: 10.1073/pnas.94.23.12457
- Milani-Nejad, N., and Janssen, P. M. (2014). Small and large animal models in cardiac contraction research: advantages and disadvantages. *Pharmacol. Ther.* 141, 235–249. doi: 10.1016/j.pharmthera.2013.10.007
- Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., et al. (2011). A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* 29, 143–148. doi: 10.1038/nbt.1755
- Miyagawa, S., Matsunari, H., Watanabe, M., Nakano, K., Umeyama, K., Sakai, R., et al. (2015). Generation of α 1,3-galactosyltransferase and cytidine

- monophospho-N-acetylneuraminic acid hydroxylase gene double-knockout pigs. *J. Reprod. Dev.* 61, 449–457. doi: 10.1262/jrd.2015-058
- Morbiter, R., Elsaesser, J., Hausner, J., and Lahaye, T. (2011). Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res.* 39, 5790–5799. doi: 10.1093/nar/gkr151
- Moscou, M. J., and Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science* 326:1501. doi: 10.1126/science.1178817
- Mosher, D. S., Quignon, P., Bustamante, C. D., Sutter, N. B., Mellersh, C. S., Parker, H. G., et al. (2007). A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. *PLoS Genet.* 3:e79. doi: 10.1371/journal.pgen.0030079
- Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353:aaf8729. doi: 10.1126/science.aaf8729
- Nissen, S. E., and Wolski, K. (2007). Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N. Engl. J. Med.* 356, 2457–2471. doi: 10.1056/NEJMoa072761
- Niu, D., Wei, H. J., Lin, L., George, H., Wang, T., Lee, I. H., et al. (2017). Inactivation of porcine endogenous retrovirus in pigs using CRISPR-Cas9. *Science* 357, 1303–1307. doi: 10.1126/science.aan4187
- Oatley, J. M. (2017). Recent advances for spermatogonial stem cell transplantation in livestock. *Reprod. Fertil. Dev.* 30, 44–49. doi: 10.1071/RD17418
- Park, K. E., Kaucher, A. V., Powell, A., Waqas, M. S., Sandmaier, S. E., Oatley, M. J., et al. (2017). Generation of germline ablated male pigs by CRISPR/Cas9 editing of the NANOS2 gene. *Sci. Rep.* 7:40176. doi: 10.1038/srep40176
- Patience, C., Takeuchi, Y., and Weiss, R. A. (1997). Infection of human cells by an endogenous retrovirus of pigs. *Nat. Med.* 3, 282–286. doi: 10.1038/nm0397-282
- Pavletich, N. P., and Pabo, C. O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252, 809–817. doi: 10.1126/science.2028256
- Peng, J., Wang, Y., Jiang, J., Zhou, X., Song, L., Wang, L., et al. (2015). Production of human albumin in pigs through CRISPR/Cas9-mediated knockin of human cDNA into swine albumin locus in the zygotes. *Sci. Rep.* 5:16705. doi: 10.1038/srep16705
- Perleberg, C., Kind, A., and Schnieke, A. (2018). Genetically engineered pigs as models for human disease. *Dis. Model Mech.* 11:dmm030783. doi: 10.1242/dmm.030783
- Petersen, B., Frenzel, A., Lucas-Hahn, A., Herrmann, D., Hassel, P., Klein, S., et al. (2016). Efficient production of biallelic GGTA1 knockout pigs by cytoplasmic microinjection of CRISPR/Cas9 into zygotes. *Xenotransplantation* 23, 338–346. doi: 10.1111/xen.12258
- Phelps, C. J., Koike, C., Vaught, T. D., Boone, J., Wells, K. D., Chen, S. H., et al. (2003). Production of alpha 1,3-galactosyltransferase-deficient pigs. *Science* 299, 411–414. doi: 10.1126/science.1078942
- Polejaeva, I. A., Chen, S. H., Vaught, T. D., Page, R. L., Mullins, J., Ball, S., et al. (2000). Cloned pigs produced by nuclear transfer from adult somatic cells. *Nature* 407, 86–90. doi: 10.1038/35024082
- Porteus, M. H., and Baltimore, D. (2003). Chimeric nucleases stimulate gene targeting in human cells. *Science* 300:763. doi: 10.1126/science.1078395
- Prather, R. S., Rowland, R. R., Ewen, C., Tribble, B., Kerrigan, M., Bawa, B., et al. (2013). An intact sialoadhesin (Sn/SIGLEC1/CD169) is not required for attachment/internalization of the porcine reproductive and respiratory syndrome virus. *J. Virol.* 87, 9538–9546. doi: 10.1128/JVI.00177-13
- Qian, L., Tang, M., Yang, J., Wang, Q., Cai, C., Jiang, S., et al. (2015). Targeted mutations in myostatin by zinc-finger nucleases result in double-muscling phenotype in *Meishan* pigs. *Sci. Rep.* 5:14435. doi: 10.1038/srep14435
- Rader, D. J., Cohen, J., and Hobbs, H. H. (2003). Monogenic hypercholesterolemia: new insights in pathogenesis and treatment. *J. Clin. Invest.* 111, 1795–1803. doi: 10.1172/JCI18925
- Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., et al. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154, 1380–1389. doi: 10.1016/j.cell.2013.08.021
- Rao, S., Fujimura, T., Matsunari, H., Sakuma, T., Nakano, K., Watanabe, M., et al. (2016). Efficient modification of the myostatin gene in porcine somatic cells and generation of knockout piglets. *Mol. Reprod. Dev.* 83, 61–70. doi: 10.1002/mrd.22591
- Rogers, C. S., Hao, Y., Rokhlina, T., Samuel, M., Stoltz, D. A., Li, Y., et al. (2008). Production of CFTR-null and CFTR-DeltaF508 heterozygous pigs by adeno-associated virus-mediated gene targeting and somatic cell nuclear transfer. *J. Clin. Invest.* 118, 1571–1577. doi: 10.1172/JCI34773
- Ruan, J., Xu, J., Chen-Tsai, R. Y., and Li, K. (2017). Genome editing in livestock: are we ready for a revolution in animal breeding industry? *Transgenic Res.* 26, 715–726. doi: 10.1007/s11248-017-0049-7
- Sehayek, E., Shefer, S., Nguyen, L. B., Ono, J. G., Merkel, M., and Breslow, J. L. (2000). Apolipoprotein E regulates dietary cholesterol absorption and biliary cholesterol excretion: studies in C57BL/6 apolipoprotein E knockout mice. *Proc. Natl. Acad. Sci. U.S.A.* 97, 3433–3437. doi: 10.1073/pnas.050016197
- Shen, Y., Xu, K., Yuan, Z., Guo, J., Zhao, H., Zhang, X., et al. (2017). Efficient generation of P53 biallelic knockout Diannan miniature pigs via TALENs and somatic cell nuclear transfer. *J. Transl. Med.* 15:224. doi: 10.1186/s12967-017-1327-0
- Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science* 351, 84–88. doi: 10.1126/science.aad5227
- Smith, J., Berg, J. M., and Chandrasegaran, S. (1999). A detailed study of the substrate specificity of a chimeric restriction enzyme. *Nucleic Acids Res.* 27, 674–681. doi: 10.1093/nar/27.2.674
- Soda, M., Choi, Y. L., Enomoto, M., Takada, S., Yamashita, Y., Ishikawa, S., et al. (2007). Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* 448, 561–566. doi: 10.1038/nature05945
- Suzuki, S., Iwamoto, M., Saito, Y., Fuchimoto, D., Sembon, S., Suzuki, M., et al. (2012). Il2rg gene-targeted severe combined immunodeficiency pigs. *Cell Stem Cell* 10, 753–758. doi: 10.1016/j.stem.2012.04.021
- Tan, W., Carlson, D. F., Lancto, C. A., Garbe, J. R., Webster, D. A., and Hackett, P. B. (2013). Efficient nonmeiotic allele introgression in livestock using custom endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16526–16531. doi: 10.1073/pnas.1310478110
- Trayhurn, P., Temple, N. J., and Van Aerde, J. (1989). Evidence from immunoblotting studies on uncoupling protein that brown adipose tissue is not present in the domestic pig. *Can. J. Physiol. Pharmacol.* 67, 1480–1485.
- Tycko, J., Myer, V. E., and Hsu, P. D. (2016). Methods for optimizing CRISPR-Cas9 genome editing specificity. *Mol. Cell* 63, 355–370. doi: 10.1016/j.molcel.2016.07.004
- Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., and Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* 11, 636–646. doi: 10.1038/nrg2842
- van der Laan, L. J., Lockett, C., Griffeth, B. C., Frasier, F. S., Wilson, C. A., Onions, D. E., et al. (2000). Infection by porcine endogenous retrovirus after islet xenotransplantation in SCID mice. *Nature* 407, 90–94. doi: 10.1038/35024089
- Wang, K., Jin, Q., Ruan, D., Yang, Y., Liu, Q., Wu, H., et al. (2017a). Cre-dependent Cas9-expressing pigs enable efficient in vivo genome editing. *Genome Res.* 27, 2061–2071. doi: 10.1101/gr.222521.117
- Wang, K., Tang, X., Xie, Z., Zou, X., Li, M., Yuan, H., et al. (2017b). CRISPR/Cas9-mediated knockout of myostatin in Chinese indigenous Erhualian pigs. *Transgenic Res.* 26, 799–805. doi: 10.1007/s11248-017-0044-z
- Wang, K., Ouyang, H., Xie, Z., Yao, C., Guo, N., Li, M., et al. (2015). Efficient generation of myostatin mutations in pigs using the CRISPR/Cas9 system. *Sci. Rep.* 5:16623. doi: 10.1038/srep16623
- Watanabe, M., Nakano, K., Matsunari, H., Matsuda, T., Maehara, M., Kanai, T., et al. (2013). Generation of interleukin-2 receptor gamma gene knockout pigs from somatic cells genetically modified by zinc finger nuclease-encoding mRNA. *PLoS One* 8:e76478. doi: 10.1371/journal.pone.0076478
- Wells, K. D., Bardot, R., Whitworth, K. M., Tribble, B. R., Fang, Y., Mileham, A., et al. (2017). Replacement of porcine CD163 scavenger receptor cysteine-rich domain 5 with a CD163-like homolog confers resistance of pigs to genotype 1 but not genotype 2 porcine reproductive and respiratory syndrome virus. *J. Virol.* 91, e1521–e1516. doi: 10.1128/JVI.01521-16
- Wells, K. D., and Prather, R. S. (2017). Genome-editing technologies to improve research, reproduction, and production in pigs. *Mol. Reprod. Dev.* 84, 1012–1017. doi: 10.1002/mrd.22812
- Whitworth, K. M., Benne, J. A., Spate, L. D., Murphy, S. L., Samuel, M. S., Murphy, C. N., et al. (2017). Zygote injection of CRISPR/Cas9 RNA successfully modifies the target gene without delaying blastocyst development or altering the sex ratio in pigs. *Transgenic Res.* 26, 97–107. doi: 10.1007/s11248-016-9989-6

- Whitworth, K. M., Lee, K., Benne, J. A., Beaton, B. P., Spate, L. D., Murphy, S. L., et al. (2014). Use of the CRISPR/Cas9 system to produce genetically engineered pigs from in vitro-derived oocytes and embryos. *Biol. Reprod.* 91:78. doi: 10.1095/biolreprod.114.121723
- Whitworth, K. M., Rowland, R. R., Ewen, C. L., Triple, B. R., Kerrigan, M. A., Cino-Ozuna, A. G., et al. (2016). Gene-edited pigs are protected from porcine reproductive and respiratory syndrome virus. *Nat. Biotechnol.* 34, 20–22. doi: 10.1038/nbt.3434
- Wiedenheft, B., Sternberg, S. H., and Doudna, J. A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482, 331–338. doi: 10.1038/nature10886
- Wu, H., Liu, Q., Shi, H., Xie, J., Zhang, Q., Ouyang, Z., et al. (2018). Engineering CRISPR/Cpf1 with tRNA promotes genome editing capability in mammalian systems. *Cell Mol. Life Sci.* doi: 10.1007/s00018-018-2810-3 [Epub ahead of print].
- Wu, J., Platero Luengo, A., Gil, M. A., Suzuki, K., Cuello, C., Morales Valencia, M., et al. (2016). Generation of human organs in pigs via interspecies blastocyst complementation. *Reprod. Domest. Anim.* 51, 18–24. doi: 10.1111/rda.12796
- Wu, J., Platero-Luengo, A., Sakurai, M., Sugawara, A., Gil, M. A., Yamauchi, T., et al. (2017a). Interspecies chimerism with mammalian pluripotent stem cells. *Cell* 168, 473–486. doi: 10.1016/j.cell.2016.12.036
- Wu, J., Vilarino, M., Suzuki, K., Okamura, D., Bogliotti, Y. S., Park, I., et al. (2017b). CRISPR-Cas9 mediated one-step disabling of pancreatogenesis in pigs. *Sci. Rep.* 7:10487. doi: 10.1038/s41598-017-08596-5
- Xin, J., Yang, H., Fan, N., Zhao, B., Ouyang, Z., Liu, Z., et al. (2013). Highly efficient generation of GGTA1 biallelic knockout inbred mini-pigs with TALENs. *PLoS One* 8:e84250. doi: 10.1371/journal.pone.0084250
- Xu, S., Cao, S., Zou, B., Yue, Y., Gu, C., Chen, X., et al. (2016). An alternative novel tool for DNA editing without target sequence limitation: the structure-guided nuclease. *Genome Biol.* 17:186. doi: 10.1186/s13059-016-1038-5
- Yan, S., Tu, Z., Liu, Z., Fan, N., Yang, H., Yang, S., et al. (2018). A Huntingtin knockin pig model recapitulates features of selective neurodegeneration in huntington's disease. *Cell* 173, 989.e13–1002.e1002. doi: 10.1016/j.cell.2018.03.005
- Yang, D., Yang, H., Li, W., Zhao, B., Ouyang, Z., Liu, Z., et al. (2011). Generation of PPAR γ mono-allelic knockout pigs via zinc-finger nucleases and nuclear transfer cloning. *Cell Res.* 21, 979–982. doi: 10.1038/cr.2011.70
- Yang, H., Zhang, J., Zhang, X., Shi, J., Pan, Y., Zhou, R., et al. (2018). CD163 knockout pigs are fully resistant to highly pathogenic porcine reproductive and respiratory syndrome virus. *Antiviral Res.* 151, 63–70. doi: 10.1016/j.antiviral.2018.01.004
- Yang, L., Güell, M., Niu, D., George, H., Lesha, E., Grishin, D., et al. (2015). Genome-wide inactivation of porcine endogenous retroviruses (PERVs). *Science* 350, 1101–1104. doi: 10.1126/science.aad1191
- Yang, Y. G., and Sykes, M. (2007). Xenotransplantation: current status and a perspective on the future. *Nat. Rev. Immunol.* 7, 519–531. doi: 10.1038/nri2099
- Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771. doi: 10.1016/j.cell.2015.09.038
- Zheng, Q., Lin, J., Huang, J., Zhang, H., Zhang, R., Zhang, X., et al. (2017). Reconstitution of UCP1 using CRISPR/Cas9 in the white adipose tissue of pigs decreases fat deposition and improves thermogenic capacity. *Proc. Natl. Acad. Sci. U.S.A.* 114, E9474–E9482. doi: 10.1073/pnas.1707853114
- Zhou, X., Wang, L., Du, Y., Xie, F., Li, L., Liu, Y., et al. (2016). Efficient generation of gene-modified pigs harboring precise orthologous human mutation via CRISPR/Cas9-induced homology-directed repair in zygotes. *Hum. Mutat.* 37, 110–118. doi: 10.1002/humu.22913
- Zhou, X., Xin, J., Fan, N., Zou, Q., Huang, J., Ouyang, Z., et al. (2015). Generation of CRISPR/Cas9-mediated gene-targeted pigs via somatic cell nuclear transfer. *Cell Mol. Life Sci.* 72, 1175–1184. doi: 10.1007/s00018-014-1744-7
- Zou, Y., Li, Z., Zou, Y., Hao, H., Li, N., and Li, Q. (2018). An FBXO40 knockout generated by CRISPR/Cas9 causes muscle hypertrophy in pigs without detectable pathological effects. *Biochem. Biophys. Res. Commun.* 498, 940–945. doi: 10.1016/j.bbrc.2018.03.085

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Yang and Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Trio-Based Deep Sequencing Reveals a Low Incidence of Off-Target Mutations in the Offspring of Genetically Edited Goats

OPEN ACCESS

Edited by:

Zhiying Zhang,
Northwest A&F University, China

Reviewed by:

Alejo Menchaca,
Fundación IRAUy, Uruguay
Rui Chen,
Baylor College of Medicine,
United States
Robin Ketteler,
University College London,
United Kingdom

*Correspondence:

Bjoern Petersen
bjoern.petersen@fli.de
Xiaolong Wang
xiaolongwang@nwfafu.edu.cn

† These authors have contributed
equally to this work as co-first authors

Specialty section:

This article was submitted to
Genomic Assay Technology,
a section of the journal
Frontiers in Genetics

Received: 02 August 2018

Accepted: 18 September 2018

Published: 04 October 2018

Citation:

Li C, Zhou S, Li Y, Li G, Ding Y,
Li L, Liu J, Qu L, Sonstegard T,
Huang X, Jiang Y, Chen Y, Petersen B
and Wang X (2018) Trio-Based Deep
Sequencing Reveals a Low Incidence
of Off-Target Mutations
in the Offspring of Genetically Edited
Goats. *Front. Genet.* 9:449.
doi: 10.3389/fgene.2018.00449

Chao Li^{††}, Shiwei Zhou^{††}, Yan Li^{††}, Guanwei Li¹, Yige Ding¹, Lan Li¹, Jing Liu¹, Lei Qu²,
Tad Sonstegard³, Xingxu Huang⁴, Yu Jiang¹, Yulin Chen¹, Bjoern Petersen^{5*} and
Xiaolong Wang^{1*}

¹ College of Animal Science and Technology, Northwest A&F University, Yangling, China, ² Life Science Research Center, Yulin University, Yulin, China, ³ Recombinetics, Saint Paul, MN, United States, ⁴ School of Life Sciences and Technology, ShanghaiTech University, Shanghai, China, ⁵ Institut für Nutztiergenetik, Friedrich-Loeffler-Institut, Neustadt an der Weinstraße, Germany

Unintended off-target mutations induced by CRISPR/Cas9 nucleases may result in unwanted consequences, which will impede the efficient applicability of this technology for genetic improvement. We have recently edited the goat genome through CRISPR/Cas9 by targeting *MSTN* and *FGF5*, which increased muscle fiber diameter and hair fiber length, respectively. Using family trio-based sequencing that allow better discrimination of variant origins, we herein generated offspring from edited goats, and sequenced the members of four family trios (gene-edited goats and their offspring) to an average of $\sim 36.8\times$ coverage. This data was to systematically examined for mutation profiles using a stringent pipeline that comprehensively analyzed the sequence data for *de novo* single nucleotide variants, indels, and structural variants from the genome. Our results revealed that the incidence of *de novo* mutations in the offspring was equivalent to normal populations. We further conducted RNA sequencing using muscle and skin tissues from the offspring and control animals, the differentially expressed genes (DEGs) were related to muscle fiber development in muscles, skin development, and immune responses in skin tissues. Furthermore, in contrast to recently reports of Cas9 triggered p53 expression alterations in cultured cells, we provide primary evidence to show that Cas9-mediated genetic modification does not induce apparent p53 expression changes in animal tissues. This work provides adequate molecular evidence to support the reliability of conducting Cas9-mediated genome editing in large animal models for biomedicine and agriculture.

Keywords: genome editing, CRISPR/Cas9, whole genome sequencing, off-target, *de novo* mutation

INTRODUCTION

Recent advances in genome editing using the type II bacterial clustered, regularly interspaced, palindromic repeats (CRISPR)-associated (Cas) system have enabled efficient genetic modification in the genomes of many organisms, including large animal models for biomedicine or agricultural purposes. Within the CRISPR/Cas9 system, the Cas9 from *Streptococcus pyogenes* recognizes a 5'-NGG-3' PAM sequence on the non-target DNA strand, and allows complementation for 20-base-pair of target DNA sequence (Mali and Church, 2013). However, unwanted off-target mutations and chromosomal translocations are potential drawbacks, raising concerns about the precision of the CRISPR/Cas system, which would prohibit its use in correcting human genetic diseases, and for optimal commercialization within livestock genetic improvement programs (Garas et al., 2015; Ruan et al., 2017).

Off-target detection in advance is a challenge, as the existing proven analysis methods depend largely on amplification and sequencing of pre-selected off-target sites, identified by several bioinformatics tools [e.g., CasOT (Xiao et al., 2014), and CT-Finder (Zhu et al., 2016)]. This approach can be more difficult to implement when the analysis aims to interrogate all possible non-unique matches and allowed mismatches distal from the PAM sequences. Compared with Sanger sequencing or short-reads deep sequencing of pre-selected off-target PCR amplicons, whole genome sequencing (WGS) is a less biased assessment of off-target mutations caused by Cas9. WGS is able to fully characterize the genome-wide mutation profiles, which not only include small insertion and deletions (indels) and SNPs but also structural variants such as inversions, rearrangements, duplications, and major deletions (Zischewski et al., 2017). This approach has been used to screen for off-target mutations induced by CRISPR/Cas9 in human cells (Smith et al., 2014; Kim et al., 2015), mice (Veres et al., 2014; Iyer et al., 2015), and plants (Zhang et al., 2017). Screening for off-target mutagenesis in gene-edited animals is rare and will be highly important in farm animal since gene editing brings the commercial benefits of improving the genetics of livestock, and also serves as a research model for biomedical studies. In addition, investigating the mutation profiles in the offspring of edited animals will provide fundamental evidence to support the reliability of the CRISPR/Cas9 system.

Genetically modified goats were successfully generated through multiplex injection of four sgRNAs targeting two functional genes (*MSTN* and *FGF5*) and Cas9 mRNA in one-cell-stage embryos (Wang et al., 2015). The desired phenotypes were independently observed in the edited animals. For example, disruption of *MSTN* caused increased muscle mass (Wang et al., 2018b), while disruption of *FGF5* increased the number of secondary hair follicles and enhanced fiber length (Wang et al., 2016). Although healthy edited goats with ideal phenotypes were generated by this effort, the rate of genome-wide off-target mutations in the edited animals and their progenies have not been well documented. As such, 11 goats from four family trios were sequenced at a high coverage ($>36.8\times$), and the mutational profiles of these animals were systematically

characterized to determine rates of off-target activity. The *de novo* mutation rate in the offspring were determined to be largely equivalent to the mutation rates of other populations such as human and cattle. Together with our previous results using trio-based WGS to show a low incidence of off-target mutations in gene-edited sheep (Wang et al., 2018a), this work confirms the reliability of a multiplex-based CRISPR/Cas9 approach for the production of offspring from genetically modified large animal models that are intended for bio-medical studies and food production.

MATERIALS AND METHODS

Animals

The genetically edited animals (Shaanbei Cashmere goat) used in the present study were generated previously (Wang et al., 2015). Two edited males (#009 and #040) were selected and mated with either edited (#076, #082, #073) or wildtype (#1392) female animals (same breed) after puberty (1.5 years old). All animals involved in this study were maintained at the Research Farm of Yulin University, Yulin, China. All protocols involving animals were approved by the College of Animal Science and Technology, Northwest A&F University (Approval ID: 2014ZX08008002). The pedigree information of these four family trios was validated by estimating the kinship coefficient according to a previous study (Manichaikul et al., 2010), and were present in **Figure 1A**. Additional WGS data from two goat trios were selected as control trios for the present study (unpublished data).

Whole-Genome Sequencing and Data Analyses

A total of 11 goats from four family trios were chosen for WGS (**Figure 1A**). For each animal, genomic DNA was extracted from peripheral venous blood samples with a Qiagen DNeasy Blood and Tissue Kit (Qiagen). To construct the WGS library, 1 μ g of genomic DNA was fragmented to around 300 bp by ultrasonication using a Covaris S2 system. Then, the sheared DNA fragments were used for library construction using an Illumina TruSeq DNA library preparation kit at Novogene¹. The final quality-ensured libraries were sequenced on an Illumina HiSeq 3000 for 125-bp paired-end reads. The raw sequencing reads were first filtered to remove low quality paired reads with the following criteria: (1) reads with $>10\%$ N bases, (2) reads with $>50\%$ bases with a sequencing quality of <3 , and (3) reads with residual length of <40 bases after the adaptor sequences were trimmed. All reads that passed the quality control procedures were converted into FASTQ files.

To analyze the mutational classes in all the mutations, 12 different mutations were categorized into nine classes (T $>$ G, T $>$ C, T $>$ A, G $>$ T, G $>$ A, C $>$ T, A $>$ T, A $>$ G, and A $>$ C), and the base changes were

¹<http://www.novogene.com>

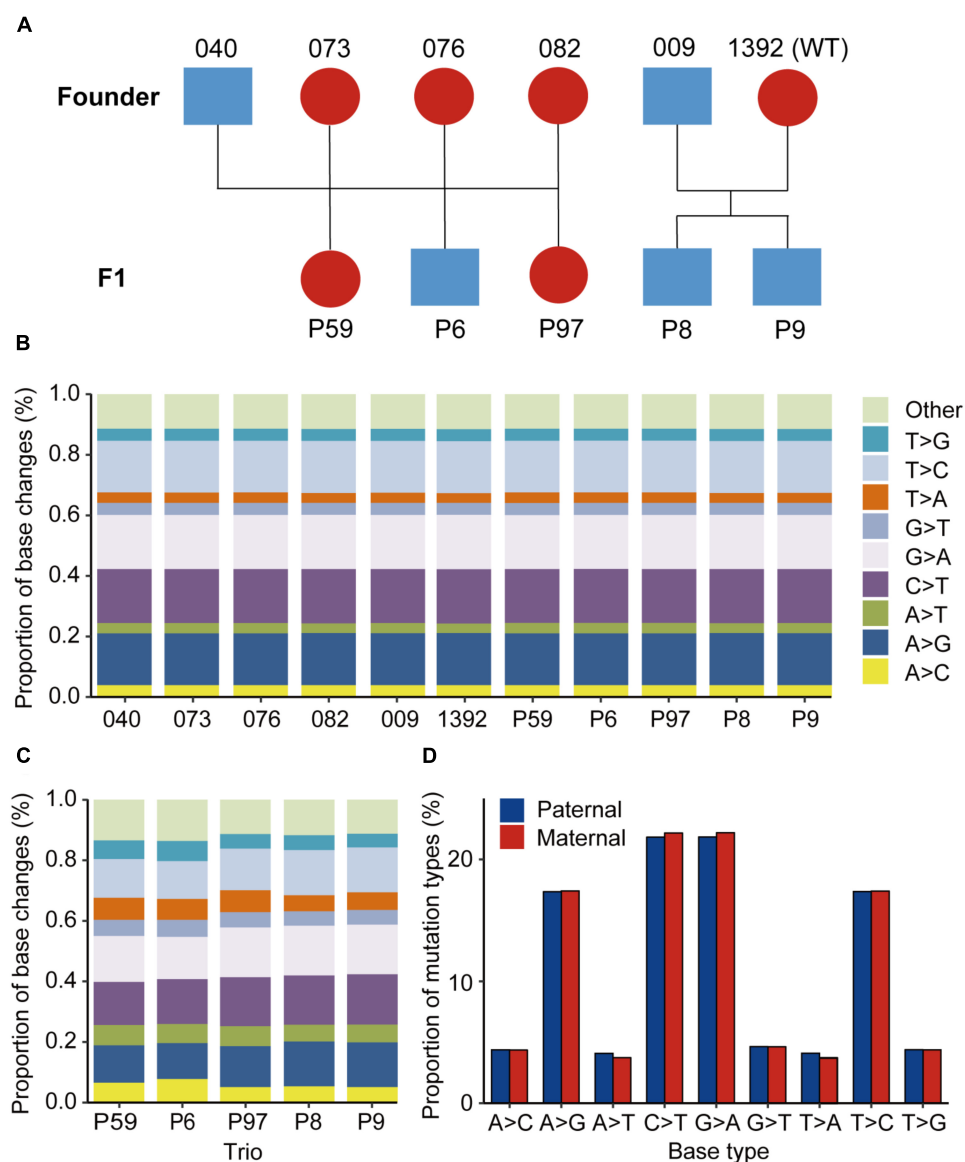


FIGURE 1 | Animals used and the mutational spectra in the edited goat genome. **(A)** Animals used in this study. **(B)** Proportion of base changes in each animal in this study. Blue color represents male animals, while red color represents female animals. **(C)** Proportion of base changes in each family trio. **(D)** Nucleotide substitution ratios from paternal or maternal origins in the F1 progenies.

measured in each animal and each family trio (Jónsson et al., 2017).

Identification of *de novo* Variants

All reads after quality control were mapped to the goat reference sequence assembly ARS1 (Bickhart et al., 2017), using BWA (v0.7.13) tools (Li and Durbin, 2009) with default parameters. Local realignment and base quality recalibration were conducted using the Genome Analysis Toolkit (GATK, v2.7-2) (McKenna et al., 2010). Both single nucleotide variants (SNVs) and indels (2–100 bp) were called using GATK and Samtools (Li et al., 2009; McKenna et al., 2010).

De novo SNVs and indels for each F1 animal were extracted according to the following criteria: (1) SNVs were independently identified by GATK and Samtools within a single trio, (2) choosing the overlapped SNVs that were identified by GATK and Samtools, and selecting the SNVs that were specific in each F1 animal; (3) filtering SNVs that exist in a goat SNP database ($n = 234$, 11 populations including 30 cashmere goats, >79 million SNPs, unpublished data); (4) filtering SNVs with a read depth in parents <12, and in F1 animals <1/10 of the sum of the coverage in both parents (Allen et al., 2013; Jónsson et al., 2017); (5) the normalized phred-scaled likelihood (PL) scores for the genotypes (AA, AB, and BB) of F1 animals (A, reference allele; B, alternate

allele), the PL scores for each genotype in founders should be >20, 0, and >0, the PL scores for each genotype in both parents should be 0, >20, and >20 (Allen et al., 2013); (6) filtering SNVs with >10% average soft clipping per read; (7) manual examination to remove mis-aligned or miscalled SNVs/indels. Copy number variations (CNVs) were identified using a CNVcaller tool (Wang X. et al., 2017), we developed recently. The scanning window was defined as 400 bp, and the effective window value was set as 7, then the low-quality CNVs were removed manually. Structural variations (SVs) were independently detected using BreakDancer (Chen et al., 2009) and Lumpy (Layer et al., 2014) with suggested parameters. The detailed filtering pipelines for *de novo* SNVs are summarized in Table 1.

Prediction of Off-Target Sites

The putative off-target sites in the goat genome that might be recognized by the sgRNAs targeting the *MSTN* and *FGF5* genes were predicted by CasOT (Xiao et al., 2014), and Cas-OFFinder (Bae et al., 2014). The potential off-target sites were defined as up to five mismatches according to a recent study (Boyle et al., 2017). The detailed information of predicted off-target sites is summarized in Supplementary Table S1.

Estimation of Mutation Rate

The estimation of mutation rate per base pair per generation was calculated according to a recent study (Jónsson et al., 2017). Briefly, we retrieved short read sequences (.bam file) were retrieved by averaging the coverage in 10,000 base windows and the sequences from autosome genome within 13× to 130× coverage were selected. This resulted in 245,722 effective windows or 2,457,220,000 base pairs (R) within our coverage

range. We then estimated the mutation rate per base pair per generation for each F1 animal by dividing the average number of *de novo* mutations (μ_{α}) by twice the R account.

μg^​=2μ​α​×R

Validation of Edited Sites, SNVs, and Indels

PCR-based Sanger sequencing was conducted to validate the genetic regions with editing or the existence of *de novo* variants (SNVs or indels) identified by WGS. Primers for amplifying the edited sites or the regions encompassing *de novo* variants are listed in Supplementary Tables S2, S3. Procedures for the purification of PCR products, the T7E1 cleavage assay, and Sanger sequencing were conducted according to our previously report (Wang et al., 2015).

RNA Sequencing and Data Analysis

Muscle from *MSTN*-edited and skin tissues from *FGF5*-edited F1 animals (*n* = 3) and wild-type animals (*n* = 3) at the same age (4-month) on the same farm, were collected for RNA-seq analysis. RNA extraction and sequencing were performed as described previously (Wang L. et al., 2017). Total RNA was isolated using Trizol Reagent (Invitrogen) and then treated with RNase-free DNase I (Qiagen) according to the manufacturer's instructions. The quality and concentration of the total RNA were determined using an Agilent 2100 Bioanalyzer (Agilent). From each sample, 12 RNA libraries were constructed and oligo (dT)s were used to isolate poly (A) mRNA. The mRNA was fragmented and reverse transcribed using random primers. Second-strand cDNAs were synthesized using RNase H and

TABLE 1 | Summary of filtering process for SNVs and indels in the F1 progenies.

Variants	P59	P6	P97	P8	P9
SNVs					
Called by GATK	11,809,830	11,863,185	11,764,624	11,649,340	11,665,474
Called by Samtools	12,188,260	12,213,943	12,130,082	12,021,749	12,031,956
GATK+Samtools	1404	1461	9610	5839	7099
Filtering using the goat SNP databases (<i>n</i> = 233, > 79 million SNPs)	1081	1174	3018	1526	1818
Reads depth and allelic balance filtering	136	166	166	139	127
After PL scores, soft clipping, and ambiguous bases filtering	53	70	68	50	42
After manual check	18	18	24	11	14
Validated by Sanger sequencing	11/15	7/9	–	–	5/7
Mutation rate (× 10 ^{−8})	1.08	1.42	1.38	1.02	0.85
Indels					
Called by GATK	1,796,995	1,813,872	1,781,374	1,706,893	1,713,903
Called by Samtools	1,579,898	1,589,589	1,546,312	1,408,611	1,420,070
GATK+Samtools	815	898	3314	1198	1583
Filtering using the goat SNP databases (<i>n</i> = 233, > 79 million SNPs)	187	233	562	181	222
Reads depth and allelic balance filtering	20	29	26	16	22
After PL scores, soft clipping, and ambiguous bases filtering	7	14	7	2	7
After manual check	3	8	5	1	2
Validated by Sanger sequencing	1/2	5/8	4/5	–	2/2

DNA polymerase I. The double-strand cDNAs were then purified using the QiaQuick PCR extraction kit. The required fragments were purified via agarose gel electrophoresis and were enriched through PCR amplification. Finally, the amplified fragments were sequenced using Illumina HiSeq™ 3000 system at Novogene (see text footnote 1), according to the manufacturer's instructions.

Among the raw data from RNA-seq, the sequencing adaptors, reads with unknown nucleotides larger than 5%, and the bases with low quality (more than half of the bases' qualities were less than 10) were removed. The remaining clean data was mapped to the currently available goat genome sequence assembly (ARS1) (Bickhart et al., 2017) using TopHat2 (Kim et al., 2013), to screen differentially expressed genes (DEGs). Counts for each gene were computed by means of the HTSeq Python package (Anders et al., 2015), and DEGs between the F1 progenies and control groups were determined with the EdgeR Bioconductor package using the classic method (Robinson et al., 2010). Gene Ontology (GO) functional enrichment analysis was conducted using g:Profiler to identify the functional categories enriched in DEGs (Reimand et al., 2016). The default settings were used, and GO terms with corrected *P*-value of less than 0.05 were considered significantly enriched.

The same RNA from muscle or skin tissues was used for qPCR analyses, to validate the RNA-seq results. First strand cDNA synthesized using the Thermo Scientific RevertAid First Strand cDNA Synthesis kit (#K1622, Thermo Fisher Scientific, United States) under the manual instructions, and was then subjected to quantification using a standard SYBR Premix Ex Taq (Tli RNaseH Plus) kit (#RR420A, Takara, China) on the Bio-Rad CFX96 Real-Time System. The primers for eight genes used for this study are listed in **Supplementary Table S4**. Biological and technical replicates were performed in triplicate for each sample. Gene relative expression was calculated using $2^{-\Delta\Delta C_t}$ method, quantified relative to the housekeeping gene *GAPDH*.

Immunofluorescent Staining and Western Blotting

The biopsied tissues were immediately fixed in 4% paraformaldehyde at 4°C overnight. The fixed tissues were then embedded in paraffin using standard immunohistochemical protocols. The immunofluorescence staining was conducted with anti-p53 (Cell Signaling Technology) primary antibody, the sections were then counterstained with Hoechst 33342 and analyzed by confocal laser microscopy. We extracted total protein from muscles, and then quantified the protein using the Bradford assay. Equal amounts of soluble protein were separated by SDS/PAGE and transferred onto a polyvinylidene difluoride membrane (PVDF, Roche). Immunoblotting was conducted using antibodies specific for phospho-p53 (Ser15) (Cell Signaling Technology, 1:1000), and β -actin (Proteintech, 1:1000). Primary antibodies were visualized using a fluorescence imager system (Sagecreation).

RESULTS

Generation of F1 Progenies

Cas9 mRNA and single guide RNAs targeting two functional genes, *MSTN* and *FGF5*, were multiplex-injected into one-cell-stage goat embryos, to generate animals with gene disruption (Wang et al., 2015). From this treatment, edited goats with improved phenotypes for muscling and fiber length were successfully obtained (Wang et al., 2016, 2018b). We then selected five founder animals (#9, #040, #076, #082, and #073) and one wildtype individual (#1392) for natural breeding, and obtained five F1 progenies (**Figure 1A**). We next genotyped the targeted sites in the progenies and their parents through Sanger sequencing. The genotypes of mutations at on-targeting site were validated by both WGS data (**Supplementary Figure S1**) and Sanger sequencing (**Supplementary Figure S2**). The sequencing data confirmed that the edited sites in the founder animals are successfully transmitted to their offspring, except the F1 animal #P59 was wild-type at the *MSTN*_sg1 locus, most likely because its dam is wildtype at this site (**Supplementary Figure S2**). In particular, mutations in #009 were transmitted to the twin progeny #P8 and #P9, even though #009 was mated with a wildtype female goat.

Deep Sequencing of Family-Trio Individuals

These 11 animals representing four family trios were subject to WGS variant analyses. The kinship coefficient values in each animal was used to ensure the correct pedigree information (**Supplementary Table S5**). WGS of 11 animal genomic DNAs yielded a total of 722 Gb of raw data, and produced between 516 and 944 million sequence reads mapped per animal (**Supplementary Table S6**). Over 99.02% of the generated sequence reads were mapped, indicating that high quality sequences were obtained. After alignment to the goat reference genome (ARS1) (Bickhart et al., 2017), an average of $\sim 36.8 \times (25.0\text{--}47.8 \times)$ sequencing depth were generated for further analysis (**Supplementary Table S6**).

Of all the SNVs identified by GATK, the mutation spectrums in each animal and each trio were analyzed. It was found that the C > T, A > G, G > A, and T > C substitutions are predominant in all the mutation types in each sequenced animal, and each base change type represents >17% of base changes (**Figure 1B**). Additionally, the proportion of base changes in parents and offspring is non-significant in each trio ($p = 0.326$, Student's *t*-test) (**Supplementary Figure S3**). The base changes in each family trio were further examined, and no significant changes were found among the trios used for sequencing (**Figure 1C**). Moreover, we observed that the differences in nucleotide substitution patterns between paternal and maternal mutations in F1 animals were non-significant (**Figure 1D**).

Identification and Validation of *de novo* SNVs

To detect SNVs that may be derived from Cas9-mediated genetic modification, we employed a series of stringent variant filtering

procedures (**Figure 2A** and **Table 1**). We initially called >11.6 million SNVs by GATK, and >12.0 million SNVs by Samtools independently in each progeny. We then selected the specific SNVs in each F1 animal, and chose the SNVs that were both identities by GATK and Samtools. Next, any SNVs that already existed in the goat SNP database (294 individuals from 11 wild and domestic populations, >79 million SNPs) were removed, and further filtering procedures included read depth, phred-scaled

likelihood (PL) scores and manual examination according to a previous study (Allen et al., 2013). After manual check, 18, 24, 18, 11, and 14 SNVs remained in the F1 progenies P6, P97, P59, P8, and P9, respectively (**Table 1** and **Supplementary Table S4**). These *de novo* SNVs were distributed in all chromosomes, and did not no cluster near the gene target sites (**Figure 2B**). We selected the SNVs from P6, P59, and P9 for PCR amplification followed by Sanger sequence validation, which confirmed that

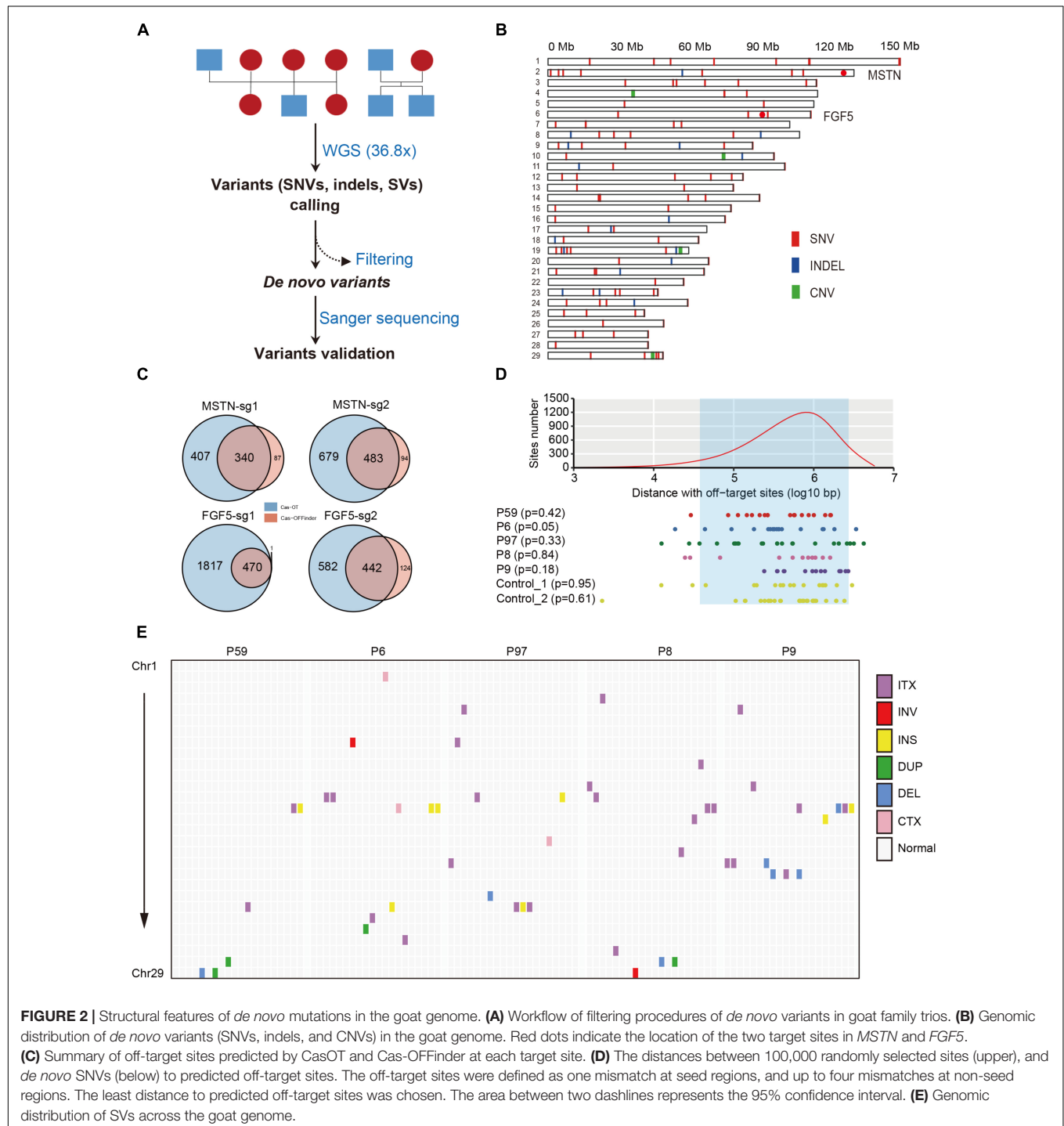


FIGURE 2 | Structural features of *de novo* mutations in the goat genome. **(A)** Workflow of filtering procedures of *de novo* variants in goat family trios. **(B)** Genomic distribution of *de novo* variants (SNVs, indels, and CNVs) in the goat genome. Red dots indicate the location of the two target sites in *MSTN* and *FGF5*. **(C)** Summary of off-target sites predicted by CasOT and Cas-OFFinder at each target site. **(D)** The distances between 100,000 randomly selected sites (upper), and *de novo* SNVs (below) to predicted off-target sites. The off-target sites were defined as one mismatch at seed regions, and up to four mismatches at non-seed regions. The least distance to predicted off-target sites was chosen. The area between two dashlines represents the 95% confidence interval. **(E)** Genomic distribution of SVs across the goat genome.

over 70% of SNVs truly exist (Table 1 and Supplementary Table S2). The germline *de novo* mutation rate (per base pair per generation) in these F1 progenies was estimated to range from 0.85×10^{-8} – 1.42×10^{-8} base substitutions per site per generation (Table 1). We next predicted the genome-wide off-target sites using two programs, Cas-OT and Cas-OFFinder, we show that the vast majority of the off-targets identified by Cas-OFFinder were also included in the off-targets predicted by Cas-OT (Figure 2C). To ensure the reliability of predicted off-targets, we chose the overlapped off-targets for further analysis. The distance between 100,000 randomly selected SNV sites, and *de novo* SNVs to 534 predicted off-target sites were simulated (five mismatches) (Supplementary Table S1 and Supplementary Figure S4), and no significant effects between the random selected SNVs and these *de novo* SNVs were observed from five F1 progenies, and two F1 animals from the control trios ($p > 0.05$, Kolmogorov–Smirnov test) (Figure 2D). Together, these results indicated that *de novo* SNVs in the F1 progenies resulted from normal spontaneous mutagenesis rather than from CRISPR/Cas-mediated gene editing.

Identification of Indels, CNVs, and SVs

Next, a comparative analysis within each trio was performed to identify for Cas9-induced small indels, given the possible likely outcomes of Cas9 induced double-strand breaks (DSB) repaired via non-homologous end-joining (NHEJ). Similar filtering procedures were used to screen candidate *de novo* indels as were conducted for *de novo* SNVs. After stringent indel filtering procedures including read depth, PL value and manual examination, a total of 19 indels were determined as *de novo* indels in all the five F1 animals (Table 2 and Supplementary Table S3). PCR-based Sanger sequencing confirmed the existence of 13 indels (Table 1 and Supplementary Table S3).

We next examined whether the large-scale genomic alterations (CNVs and SVs) could be attributed to Cas9 nucleases. CNVcaller (Wang X. et al., 2017) was used to search for CNVs, and after filtering CNVs by its genotypes and the effective window value, only four candidate CNVs were left in the F1 animals (Figure 2B and Table 2). A number of SVs were identified using BreakDancer (Chen et al., 2009) and Lumpy (Layer et al., 2014), and only a few remained after filtering and were considered as candidate *de novo* SVs in each animal (Figure 2E, Table 2, and Supplementary Table S7).

Analyzing of Off-Target Mutations

To assess the off-target effects in F1 animals, we identified, using Cas-OFFinder, potential off-target sites with up to 3-nucleotide mismatches and NRG PAM sites in the goat genome. We then determined whether the *de novo* mutations, as well as the mutations that shared in parents and progenies were within the potential off-target sites, merely two indels were determined as off-target mutations (Supplementary Table S8). Sanger sequencing further validated these two variants (Supplementary Figure S5), indicating the off-target mutations are low in the offspring of edited animals and is guide RNA specific.

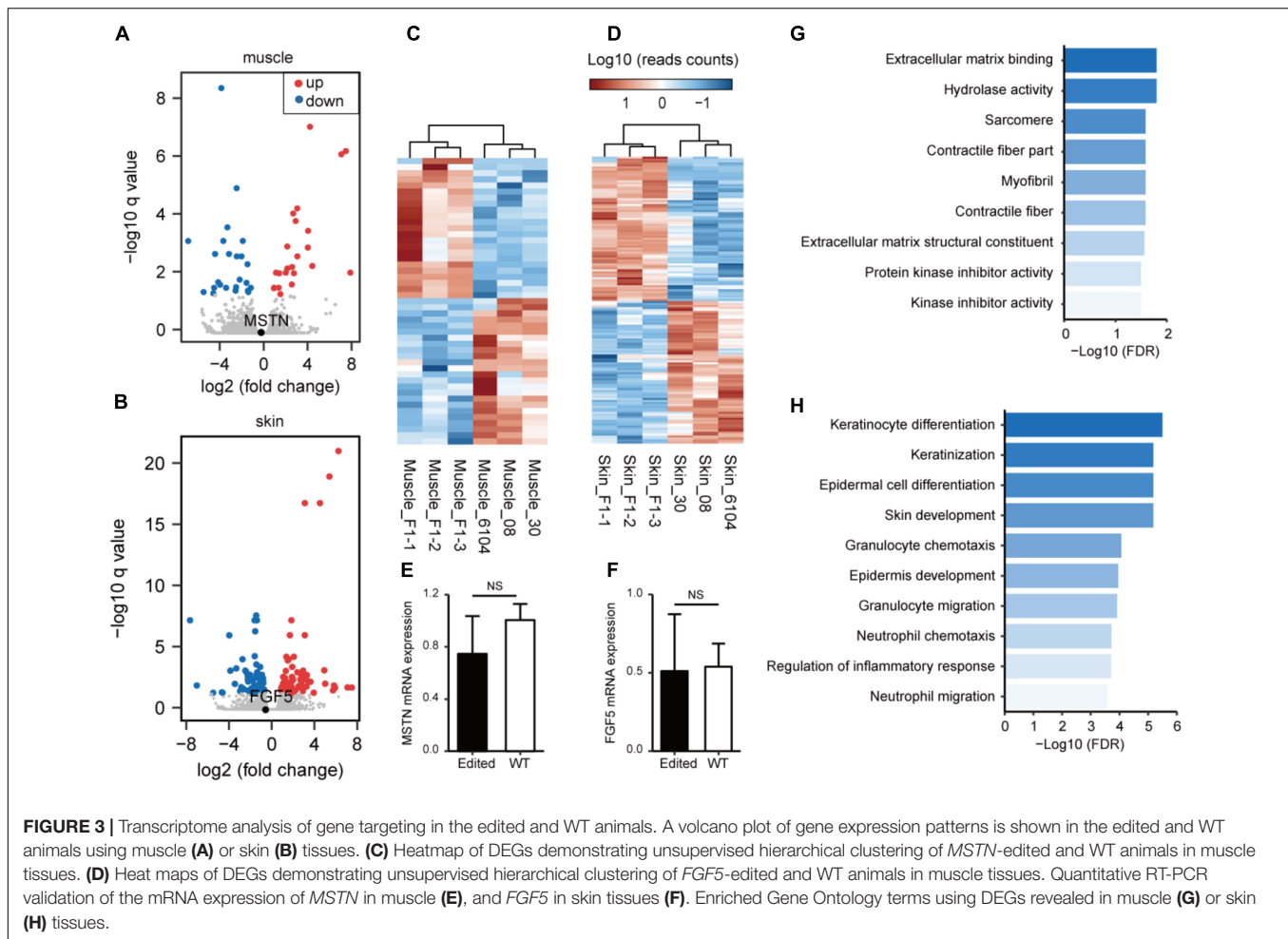
RNA-seq Analyses of Edited Animals

We have recently analyzed the transcriptome profiles using muscle tissues from *MSTN* and/or *FGF5*-edited cashmere goats (Wang L. et al., 2017), and showed that the *MSTN*-disruption resulted in substantial changes in genes involved in lipid metabolism and biosynthesis. To better understand the transcriptional consequences of gene disruption in the genome of F1 progenies, we conducted transcriptome sequencing (RNA-seq) analysis in the edited progenies and WT animals using muscle or skin tissues. The volcano plot demonstrated that the expression of *MSTN* did not change significantly (Figure 3A). However, the disruption of *MSTN* resulted in 43 (23 up-regulated and 20 down-regulated) genes with significantly changed expression (Figure 3C and Supplementary Table S9). Some of these genes are known to be associated with muscle developmental including *FMOD*, *ARG2*, *TNMD*, *CSRP3*, *PCK2*, *EGR1*, and *TNC*. Meanwhile, disruption of *FGF5* led to the identification of 140 DEGs (74 up-regulated and 66 down-regulated) in the skins of F1 progenies and control animals (Figures 3B,D and Supplementary Table S10). Key regulators related to skin/hair follicle development such as *AQP3*, *AQP5*, *SPINK7*, and *WIF1* were involved, indicating that *FGF5* disruption may stimulate hair follicle functions resulting in increased fiber length (Wang et al., 2016). We performed qPCR on ten DEGs (including *MSTN* and *FGF5*) using RNA isolated from tissues of the same individuals. The validation results revealed a similar correlation between the RNA-seq and qPCR results (Figures 3E,F and Supplementary Figure S6), suggestion the reliability of RNA-seq analyses.

Subsequently, we performed GO enrichment analysis to predict the over-represented GO terms associated the DEGs

TABLE 2 | Identified candidate CNV and SVs in the F1 progenies.

Variants	P59	P6	P97	P8	P9
CNV					
Called by CNVcaller	1426	1016	2835	418	307
Filtering by genotype and effective window	6	8	15	5	2
Candidate <i>de novo</i> CNVs	1	1	2	0	0
SVs					
(1) Called by BreakDancer, and specific in F1 animals	1570	1411	1633	2058	2107
After removal of common SVs in every two animals, and the read depth <50%, manual check	3	9	13	12	13
(2) Called by Lumpy	3609	3116	3406	2697	3082
After removal of common SVs in every two animals, and the read depth <50%, manual check	3	2	0	3	1



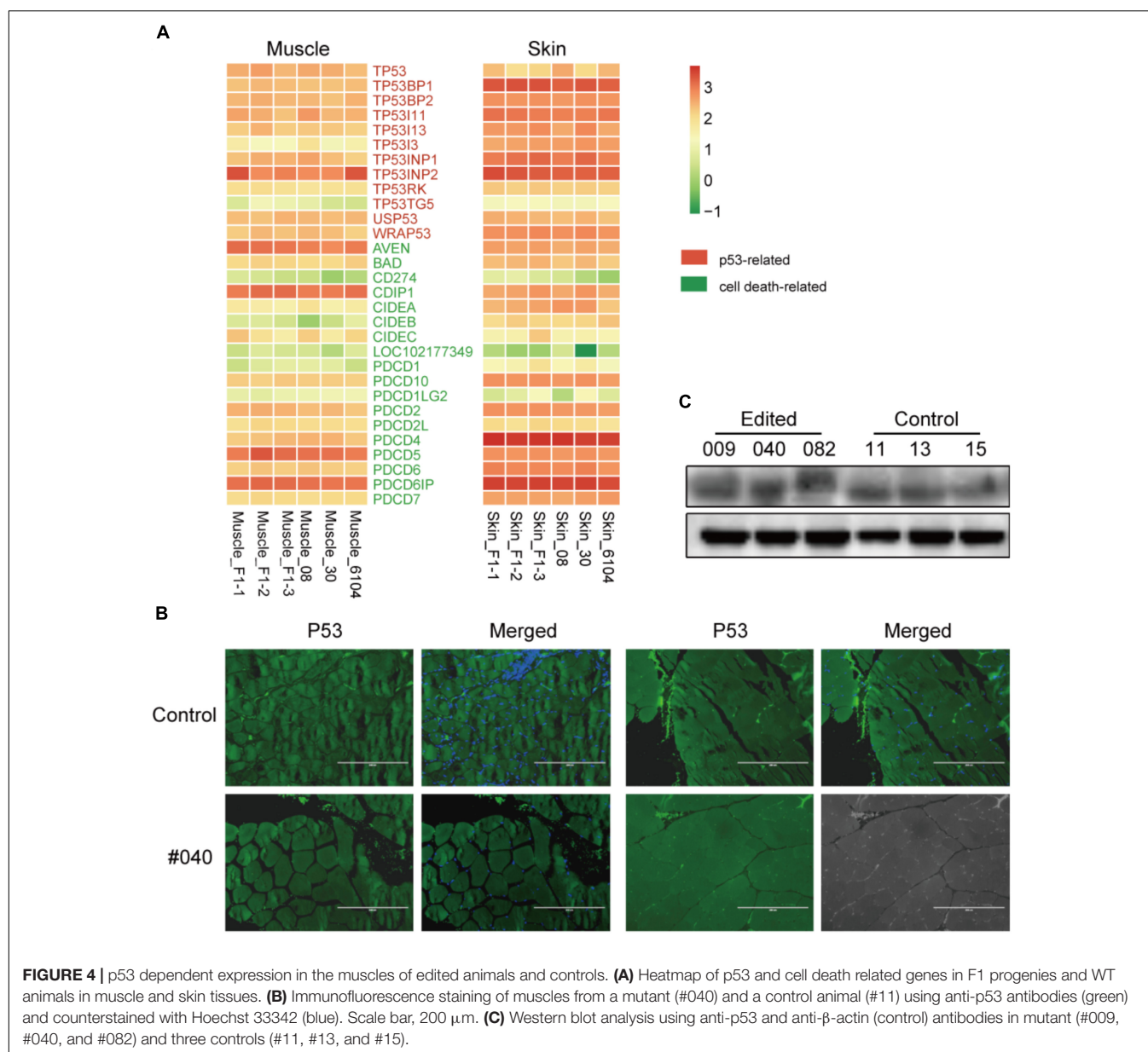
identified in muscle and/or skin tissues. The DEGs identified in muscles exhibited significant over-representation of GO terms related to muscle fiber development (such as “sarcomere,” “contractile fiber,” and “myofibril”) (Figure 3G). The DEGs identified in skins are significantly more enriched in GO terms related to skin and hair follicle development (“skin development,” “keratinization,” “keratinocyte differentiation,” and “epidermis development”) and immune responses (“granulocyte chemotaxis and migration,” “leukocyte chemotaxis and migration,” and “myeloid leukocyte migration”) (Figure 3H).

Since two recent studies described that CRISPR/Cas9-mediated genetic modification in normal cultured cells derived a p53-dependent toxic response (Haapaniemi et al., 2018; Ihry et al., 2018), which may raise tumor risks. To validate this scenario in edited animals, we first investigated the expression of p53 and cell death-related genes retrieved from the RNA-seq data in skin and muscles, no significant gene expression changes were observed (Figure 4A). We further assessed the protein expression of p53 in the muscle of *MSTN*-disrupted animals. The immunofluorescence staining indicated no apparent changes of p53 expression in muscles between an edited animal (#040) and controls (#11) (Figure 4B). Western blotting further verified the p53 expression in

muscles (Figure 4C). These results are consistent with the p53 expression in mice (personal communication), and suggested that Cas9-mediated modification did not induce apparent p53 expression changes in animal tissues. This result was probably due to p53-dependent molecular response in animal bodies may be differed with that in cultured cells (Haapaniemi et al., 2018), or the cellular response to DNA damage was repaired during the embryonic development process.

DISCUSSION

Deep sequencing is able to fully characterize the mutational profiles in genomes, and is used to detect mutational changes in genetically edited organisms (Huang et al., 2016). In this study, through sequencing four family trios at a high coverage, the *de novo* variants in F1 animals that could be attributed to the engineered nucleases were determined to be neglectable, representing a low incidence of CRISPR/Cas9-mediated off-target mutations. On the other hand, our results further demonstrate the reliability of WGS in documenting mutations induced by genome editing.



Previous studies have shown that the *de novo* mutations exhibited variant type preferences and discriminative parental origins, and the mutational signatures were influenced by multiple factors including nucleotide type, sequence context, replication timing, and epigenetics. In this study, the mutational spectrum in genomes other than human and mice are reported. The C-T and A-G transitions were predominant in goat genomes, the enrichment of C > T transitions at CpG dinucleotides could reflect spontaneous domination of methylated cytosine to thymine (Bolli et al., 2014). We observed that the proportion of base changes was not significant in the parents or in the F1 progenies (Figure 1C), indicating the mutation profiles are largely stable in the edited animals and their offspring.

Mutation rate is a key parameter for calibrating the timescale of sequence divergence. The estimated average germline *de novo*

SNV rate (per base pair per generation) in the present study is 1.15×10^{-8} , which is equivalent to the average germline mutation rate in serval trio-based human populations including CEU (1.2×10^{-8}) and YRI (1.0×10^{-8}) (Consortium et al., 2010), Icelanders (1.29×10^{-8}) (Jónsson et al., 2017), Danish (1.28×10^{-8}) (Maretty et al., 2017), a large cattle population (1.2×10^{-8}) (Harland et al., 2017), two goat trios (unpublished data), as well as three family trios of gene-edited sheep¹⁶ (Figure 5). These findings further supported the conclusion that the *de novo* SNVs in the F1 animals are generated naturally rather than induced by genetic modification.

The CRISPR based approach relies on micro-injection of recombinant Cas9 mRNA/protein and guide RNAs and often results in off-target mutagenesis (Fu et al., 2013; Cho et al., 2014). Off-target sites predicted by different computational programs

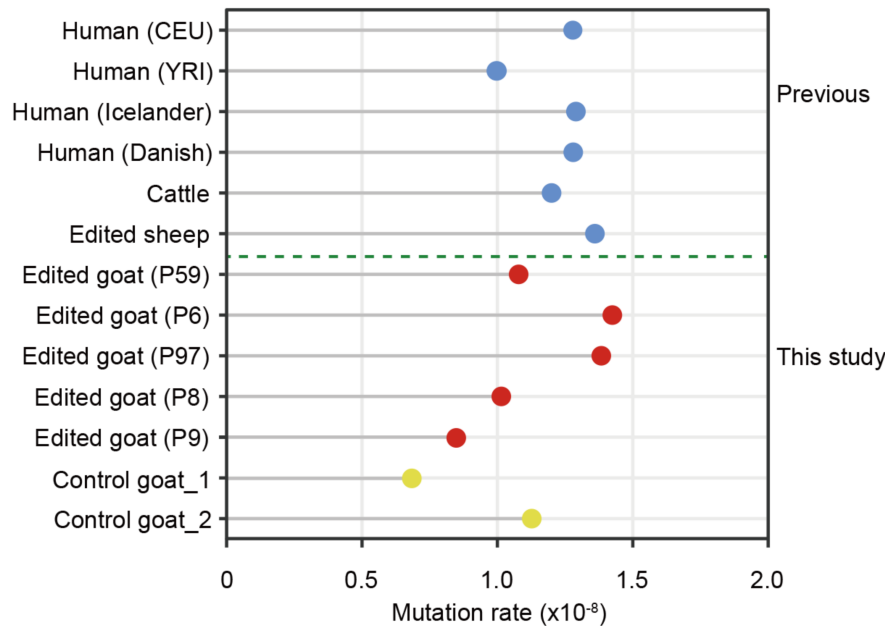


FIGURE 5 | The mutation rates in edited goats and other populations.

may have less overlaps (Tsai et al., 2015). In the present study, Cas-OT recognized most of the off-targets that predicted by Cas-OFFinder, indicating the overlapped off-target sites are most likely represent the bona fide off-target sites for further analysis. Furthermore, our work presents trio-based WGS analysis to examine genome-wide *de novo* variants that may be induced by genetic modification in the F1 animals. Consistent with the off-target mutations observed in human cells (Veres et al., 2014; Yang et al., 2014), and mice (Iyer et al., 2015), we observed low incident nuclease induced mutations in large animal models through deep sequencing. Therefore, supported the reliability of CRISPR approach for the production of viable animals.

We have previously demonstrated that disruption of *MSTN* in goats resulted in increased body weight and enlarged myofiber diameters in muscles (Wang et al., 2018b), we also show disruption of the *FGF5* genes led to longer hair fibers in goats (Wang et al., 2016). To test the effect of Cas9-modification on global transcriptional status in the F1 progenies, we conducted RNA-seq on muscle and skin tissues to independently characterize the transcriptional consequences and genetic mechanism by knockout *MSTN* and *FGF5* in F1 progenies. Inconsistent with previous studies (Wang et al., 2016, 2018b), we did not observe significantly expression changes of both *MSTN* and *FGF5* with RNA-seq and validated by qPCR (Figures 3E,F). The plausible reason for this difference is animals from two generations were used for analyses. However, in the present study, we did identify a list of DEGs that are known to be associated with muscle development (e.g., *FMOD*, *ARG2*, *TNMD*, *CSRP3*, *PCK2*, *EGR1*, and *TNC*) in the muscle tissues of *MSTN*-disrupted animals, or skin and hair follicle development (*AQP3*, *AQP5*, *SPINK7*, and *WIF1*) in the skin tissues of *FGF5*-disrupted animals.

Interestingly, functional enrichment analyses indicated that the DEGs are over-represented in GO terms associated with muscle fiber development in *MSTN*-disrupted goats, and GO terms related to skin development and immune responses in *FGF5*-edited animals. *MSTN* is primarily thought to inhibit muscle differentiation and growth (Poncelet, 1997; Mosher et al., 2007), while *FGF5* represses hair growth by blocking dermal papilla cell activation (Hebert et al., 1994; Ota et al., 2002). Therefore, it is plausible that disruption of these key regulators triggers multiple functional regulatory genes at post-transcriptional levels and eventually resulting in observed phenotypic changes.

CONCLUSION

In summary, the present study provides a comprehensive analysis of genomes from edited animals and their progenies through deep sequencing. We provide sufficient evidence to show that the incidence of *de novo* mutations is low not only in edited founder animals (Wang et al., 2018a), but also in the F1 progenies, and their mutation rate is not different from what normally occurs in wild type animals as spontaneous mutations. This study will serve as a valuable resource for evaluating the reliability of the CRISPR-based genome editing technologies in the engineering the genomes of large animals.

DATA AVAILABILITY

All relevant results are within the paper and its **Supplementary data files**. The raw WGS data of 11 animals are available at the NCBI-SRA under accession nos. SRR6378093, SRR6378094,

SRR6378095, SRR6378096, SRR6378097, SRR6378098, SRR6378099, SRR6378100, SRR6378101, SRR6378102, and SRR6378103. Transcriptomic data are available at NCBI-SRA under accession nos. SRR6411189, SRR6411190, SRR6411191, SRR6411192, SRR6411193, SRR6411194, SRR6411195, SRR6411196, SRR6411197, SRR6411198, SRR6411199, SRR6411200, and SRX743626.

AUTHOR CONTRIBUTIONS

XW, TS, BP, XH, YJ, and YC conceived the research plans. CL, SZ, YL, GL, YD, LL, and JL performed the experiments. LQ provided the samples. XW, TS, and BP wrote the article.

REFERENCES

- Allen, A. S., Berkovic, S. F., Cossette, P., Delanty, N., Dlugos, D., Eichler, E. E., et al. (2013). De novo mutations in epileptic encephalopathies. *Nature* 501, 217–221. doi: 10.1038/nature12439
- Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq—a python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638
- Bae, S., Park, J., and Kim, J.-S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 30, 1473–1475. doi: 10.1093/bioinformatics/btu048
- Bickhart, D. M., Rosen, B. D., Koren, S., Sayre, B. L., Hastie, A. R., Chan, S., et al. (2017). Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat. Genet.* 49, 643–650. doi: 10.1038/ng.3802
- Bolli, N., Avetloiseau, H., Wedge, D. C., Loo, P. V., Alexandrov, L. B., Martincorena, I., et al. (2014). Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat. Commun.* 5:2997. doi: 10.1038/ncomms3997
- Boyle, E. A., Andreasson, J. O., Chircus, L. M., Sternberg, S. H., Wu, M. J., Guegler, C. K., et al. (2017). High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc. Natl. Acad. Sci. U.S.A.* 114, 5461–5466. doi: 10.1073/pnas.1700557114
- Chen, K., Wallis, J. W., McLellan, M. D., Larson, D. E., Kalicki, J. M., Pohl, C. S., et al. (2009). BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat. Methods* 6, 677–681. doi: 10.1038/nmeth.1363
- Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., et al. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* 24, 132–141. doi: 10.1101/gr.162339.113
- Consortium, T. G. P., Altshuler, D. L., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., et al. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073. doi: 10.1038/nature09534
- Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., et al. (2013). High frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 31, 822–826. doi: 10.1038/nbt.2623
- Garas, L. C., Murray, J. D., and Maga, E. A. (2015). Genetically engineered livestock: ethical use for food and medical models. *Annu. Rev. Anim. Biosci.* 3, 559–575. doi: 10.1146/annurev-animal-022114-110739
- Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., and Taipale, J. (2018). CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* 24, 927–930. doi: 10.1038/s41591-018-0049-z
- Harland, C., Charlier, C., Karim, L., Cambisano, N., Deckers, M., Mni, M., et al. (2017). Frequency of mosaicism points towards mutation-prone early cleavage cell divisions. *bioRxiv* [Preprint]. doi: 10.1101/079863
- Hebert, J., Rosenquist, T., Gotz, J., and Martin, G. (1994). FGF5 as a regulator of the hair growth cycle: evidence from targeted and spontaneous mutations. *Cell* 78, 1017–1025. doi: 10.1016/0092-8674(94)90276-3

FUNDING

This study was supported by grants from National Natural Science Foundation of China (31772571 and 31572369), and local grants (2017NY-072 and 2018KJXX-009), as well as by China Agriculture Research System (CARS-39). XW is a Tang Scholar at Northwest A&F University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00449/full#supplementary-material>

- Huang, S., Weigel, D., Beachy, R. N., and Li, J. (2016). A proposed regulatory framework for genome-edited crops. *Nat. Genet.* 48, 109–111. doi: 10.1038/ng.3484
- Ihry, R. J., Worringer, K. A., Salick, M. R., Frias, E., Ho, D., Theriault, K., et al. (2018). p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med.* 24, 939–946. doi: 10.1038/s41591-018-0050-6
- Iyer, V., Shen, B., Zhang, W., Hodgkins, A., Keane, T., Huang, X., et al. (2015). Off-target mutations are rare in Cas9-modified mice. *Nat. Methods* 12:479. doi: 10.1038/nmeth.3408
- Jónsson, H., Sulem, P., Kehr, B., Kristmundsdottir, S., Zink, F., Hjartarson, E., et al. (2017). Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* 549, 519–522. doi: 10.1038/nature24018
- Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H. R., et al. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* 12, 237–243. doi: 10.1038/nmeth.3284
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14:R36. doi: 10.1186/gb-2013-14-4-r36
- Lamas-Toranzo, I., Guerrero-Sánchez, J., Miralles-Bover, H., Alegre-Cid, G., Pericuesta, E., and Bermejo-Ivarez, P. (2017). CRISPR is knocking on barn door. *Reprod. Domest. Anim.* 52, 39–47. doi: 10.1111/rda.13047
- Layer, R. M., Chiang, C., Quinlan, A. R., and Hall, I. M. (2014). LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* 15:R84. doi: 10.1186/gb-2014-15-6-r84
- Li, H., and Durbin, R. (2009). *Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform*. Oxford: Oxford University Press. doi: 10.1186/gb-2014-15-6-r84
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Mali, P., and Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826. doi: 10.1126/science.1232033
- Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., and Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics* 26, 2867–2873. doi: 10.1093/bioinformatics/btq559
- Maretty, L., Jensen, J. M., Petersen, B., Sibbesen, J. A., Liu, S., Villesen, P., et al. (2017). Sequencing and de novo assembly of 150 genomes from Denmark as a population reference. *Nature* 548, 87–91. doi: 10.1038/nature23264
- Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Mosher, D. S., Quignon, P., Bustamante, C. D., Sutter, N. B., Mellersh, C. S., Parker, H. G., et al. (2007). A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. *PLoS Genet.* 3:e79. doi: 10.1371/journal.pgen.0030079
- Ota, Y., Saitoh, Y., Suzuki, S., Ozawa, K., Kawano, M., and Imamura, T. (2002). Fibroblast growth factor 5 inhibits hair growth by blocking dermal papilla cell

- activation. *Biochem. Biophys. Res. Commun.* 290, 169–176. doi: 10.1006/bbrc.2001.6140
- Poncellet, D. (1997). A deletion in the bovine myostatin gene causes the double-muscling phenotype in cattle. *Nat. Genet.* 17, 71–74. doi: 10.1038/ng0997-71
- Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., et al. (2016). g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* 44, W83–W89. doi: 10.1093/nar/gkw199
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Ruan, J., Xu, J., Chen-Tsai, R. Y., and Li, K. (2017). Genome editing in livestock: are we ready for a revolution in animal breeding industry? *Transgenic Res.* 26, 715–726. doi: 10.1007/s11248-017-0049-7
- Smith, C., Gore, A., Yan, W., Abaldearistain, L., Li, Z., He, C., et al. (2014). Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-Based genome Editing in human iPSCs. *Cell Stem Cell* 15, 12–13. doi: 10.1016/j.stem.2014.06.011
- Tsai, S. Q., Zheng, Z., Nguyen, N. T., Liebers, M., Topkar, V. V., Thapar, V., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* 33, 187–197. doi: 10.1038/nbt.3117
- Veres, A., Gosis, B. S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., et al. (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. doi: 10.1016/j.stem.2014.04.020
- Wang, L., Cai, B., Zhou, S., Zhu, H., Qu, L., Wang, X., et al. (2017). RNA-seq reveals transcriptome changes in goats following myostatin gene knockout. *PLoS One* 12:e0187966. doi: 10.1371/journal.pone.0187966
- Wang, X., Cai, B., Zhou, J., Zhu, H., Niu, Y., Ma, B., et al. (2016). Disruption of FGF5 in cashmere goats using CRISPR/Cas9 results in more secondary hair follicles and longer fibers. *PLoS One* 11:e0164640. doi: 10.1371/journal.pone.0164640
- Wang, X., Liu, J., Niu, Y., Li, Y., Zhou, S., Li, C., et al. (2018a). Low incidence of SNVs and indels in trio genomes of Cas9-mediated multiplex edited sheep. *BMC Genomics* 19:397. doi: 10.1186/s12864-018-4712-z
- Wang, X., Niu, Y., Zhou, J., Zhu, H., Ma, B., Yu, H., et al. (2018b). CRISPR/Cas9-mediated MSTN disruption and heritable mutagenesis in goats causes increased body mass. *Anim. Genet.* 49, 43–51. doi: 10.1111/age.12626
- Wang, X., Yu, H., Lei, A., Zhou, J., Zeng, W., Zhu, H., et al. (2015). Generation of gene-modified goats targeting MSTN and FGF5 via zygote injection of CRISPR/Cas9 system. *Sci. Rep.* 5:13878. doi: 10.1038/srep13878
- Wang, X., Zheng, Z., Cai, Y., Chen, T., Li, C., Fu, W., et al. (2017). CNVcaller: highly efficient and widely applicable software for detecting copy number variations in large populations. *Gigascience* 6, 1–12. doi: 10.1093/gigascience/gix115
- Xiao, A., Cheng, Z., Kong, L., Zhu, Z., Lin, S., Gao, G., et al. (2014). CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics* 30, 1180–1182. doi: 10.1093/bioinformatics/btt764
- Yang, L., Grishin, D., Wang, G., Aach, J., Zhang, C. Z., Chari, R., et al. (2014). Targeted and genome-wide sequencing reveal single nucleotide variations impacting specificity of Cas9 in human stem cells. *Nat. Commun.* 5:5507. doi: 10.1038/ncomms6507
- Zhang, Q., Xing, H.-L., Wang, Z.-P., Zhang, H.-Y., Yang, F., Zhou, Y., et al. (2017). High-frequency off-target mutagenesis induced by CRISPR/Cas9 in Arabidopsis and its prevention by improving specificity of the tools. *bioRxiv* [Preprint]. doi: 10.1101/203489
- Zhu, H., Misel, L., Graham, M., Robinson, M. L., and Liang, C. (2016). CT-Finder: a web service for CRISPR optimal target prediction and visualization. *Sci. Rep.* 6:25516. doi: 10.1038/srep25516
- Zischewski, J., Fischer, R., and Bortesi, L. (2017). Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnol. Adv.* 35, 95–104. doi: 10.1016/j.biotechadv.2016.12.003

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor declared a shared affiliation, though no other collaboration, with the authors XW, CL, SZ, YL, GL, YD, LL, JL, YJ, and YC at time of review.

Copyright © 2018 Li, Zhou, Li, Li, Ding, Li, Liu, Qu, Sonstegard, Huang, Jiang, Chen, Petersen and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Methodologies for Improving HDR Efficiency

Mingjie Liu¹, Saad Rehman¹, Xidian Tang¹, Kui Gu¹, Qinlei Fan², Dekun Chen^{1*} and Wentao Ma^{1*}

¹ College of Veterinary Medicine, Northwest A&F University, Xianyang, China, ² China Animal Health and Epidemiology Center, Qingdao, China

OPEN ACCESS

Edited by:

David Jay Segal,
University of California, Davis,
United States

Reviewed by:

James Carney,
Sandia National Laboratories (SNL),
United States
Alex Michael Ward,
Sangamo BioSciences, United States

*Correspondence:

Wentao Ma
mawentao@nwsuaf.edu.cn
Dekun Chen
cdk@nwsuaf.edu.cn

Specialty section:

This article was submitted to
Genomic Assay Technology,
a section of the journal
Frontiers in Genetics

Received: 31 August 2018

Accepted: 11 December 2018

Published: 07 January 2019

Citation:

Liu M, Rehman S, Tang X, Gu K,
Fan Q, Chen D and Ma W (2019)
Methodologies for Improving HDR
Efficiency. *Front. Genet.* 9:691.
doi: 10.3389/fgene.2018.00691

Clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein 9 (Cas9) is a precise genome manipulating technology that can be programmed to induce double-strand break (DSB) in the genome wherever needed. After nuclease cleavage, DSBs can be repaired by non-homologous end joining (NHEJ) or homology-directed repair (HDR) pathway. For producing targeted gene knock-in or other specific mutations, DSBs should be repaired by the HDR pathway. While NHEJ can cause various length insertions/deletion mutations (indels), which can lead the targeted gene to lose its function by shifting the open reading frame (ORF). Furthermore, HDR has low efficiency compared with the NHEJ pathway. In order to modify the gene precisely, numerous methods arose by inhibiting NHEJ or enhancing HDR, such as chemical modulation, synchronized expression, and overlapping homology arm. Here we focus on the efficiency and other considerations of these methodologies.

Keywords: CRISPR-Cas9, HDR, NHEJ, HDR enhancement, DSB, cell arrest, NHEJ inhibitors

OVERVIEW OF CRISPR-CAS9 SYSTEM

Clustered regularly interspaced short palindromic repeats (CRISPR) represents a family of DNA sequences in bacteria and archaea (Barrangou, 2015). This family is characterized by direct palindromic repeats, where sequences are the same in both directions, varying in size from 21 to 37 bp (Barrangou and Marraffini, 2014), interspaced by spacers, which have fragments gathered from viruses or phages that previously tried to infect the cell (Horvath and Barrangou, 2010; Morange, 2015). To appreciate this characteristic array, it is termed CRISPR (Jansen et al., 2002; Mojica and Rodriguez-Valera, 2016). CRISPR-associated (*cas*) genes are invariably located adjacent to a CRISPR locus (Jansen et al., 2002). The CRISPR-Cas system can be grouped into three types: type I, type II, and type III. In addition, there are 12 subtypes of the CRISPR-Cas system, which are based on their exclusive genetic content and structural differences (Makarova et al., 2015). *Cas1* and *cas2* are universal across types and subtypes, whereas *cas3*, *cas9*, *cas10* are signature genes for type I, type II, and type III, respectively (Makarova et al., 2011). Here in this review we only focus on type II. The CRISPR-Cas system functions as a defense system in bacteria and archaea against bacteriophage infection, conjugation, and natural transformation by degrading foreign nucleic acid that enters the cell (Marraffini, 2015). The CRISPR-Cas system involves three distinct mechanistic stages: adaptation, biogenesis, and interference (Marraffini and Sontheimer, 2010). The adaptation stage involves the integration of fragments of foreign DNA (termed “protospacers,” captured, excised, and inserted by Cas proteins) into the CRISPR array as new spacers. New spacers are usually added at the beginning of the CRISPR locus next to the leader sequence, creating a chronological record of viral infections (Sorek et al., 2013) and protecting the cell from further

infection. During the biogenesis stage, the CRISPR array is transcribed as a single long transcript (termed “pre-crRNA”) containing much of the CRISPR array (Marraffini and Sontheimer, 2010) and is then processed and matured to produce CRISPR RNAs (crRNAs) with only one spacer sequence. As for the interference stage, the spacers in these crRNAs guide Cas proteins to foreign DNAs and cleave them (van der Oost et al., 2009; Wiedenheft et al., 2012; Barrangou, 2013). The type II CRISPR-Cas system needs only *cas9* to execute immunity in the presence of an existing targeting spacer sequence (Sapranaukas et al., 2011). It requires two small RNAs: the crRNA and the trans-encoded crRNA (tracrRNA) (Deltcheva et al., 2011). TracrRNA forms a secondary structure that interacts with Cas9 protein and it has a complementary region that enables itself to bind to pre-crRNA (Anders et al., 2014; Jinek et al., 2014; Nishimasu et al., 2014). The dsRNA formed between pre-crRNA and tracrRNA is then handled by RNase III to form mature crRNA guides that are used in genome editing. When crRNA and tracrRNA are combined together, they are collectively termed as guide RNA (gRNA) (Jinek et al., 2012). Another important short (3–5 bp) DNA termed protospacer adjacent motif (PAM) is required for targeting. PAM is a component of the invading virus or plasmid, but it is not a component of the bacterial CRISPR locus. Cas9 will not successfully bind to or cleave the target DNA sequence if it is not followed by the PAM sequence (Mojica et al., 2009). The first step in target recognition is the transient binding of Cas9 to PAM sequences within the target DNA, which promotes the unwinding of the two DNA strands immediately upstream of the PAM (Sternberg et al., 2014), the spacer sequence of the crRNA binds with the unwound DNA (6–8 bp in length), then forms an RNA-DNA heteroduplex and triggers cleavage at the targeted site (Sternberg et al., 2014; Szczelkun et al., 2014). After recognition, the CRISPR-Cas9 system introduces a crRNA-specific DSB in the target sequence, which is further resolved either by homology-directed repair (HDR) or non-homologous end joining (NHEJ).

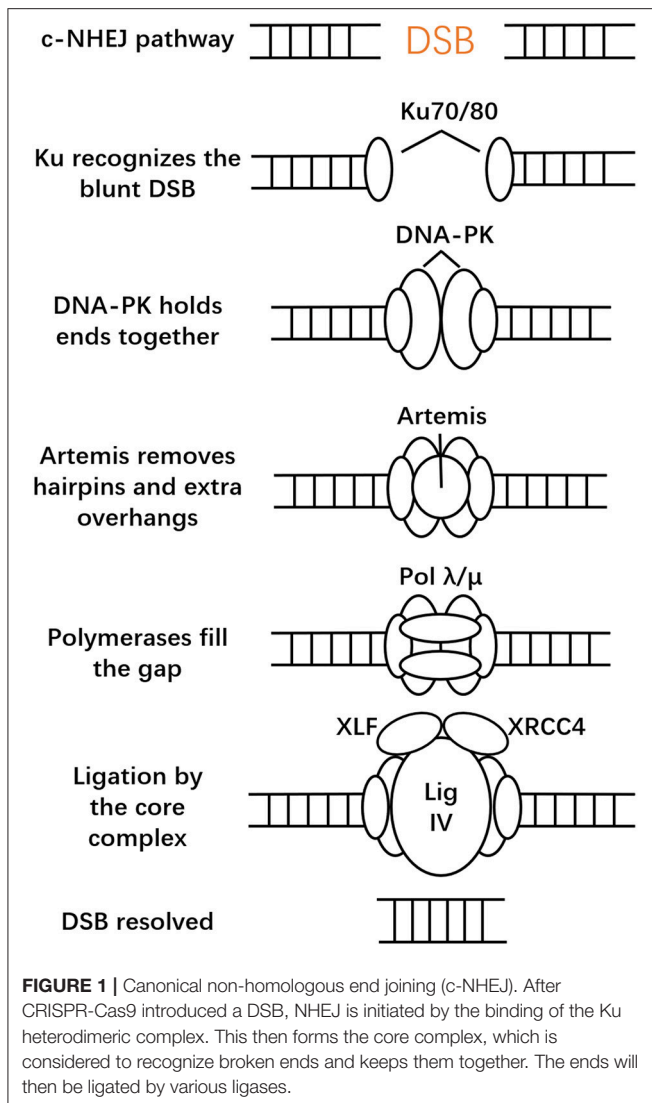
NHEJ PATHWAY

Typically, cells employ two main mechanisms to repair DSBs: classical NHEJ and HDR (Symington and Gautier, 2011). There are also many alternative error-prone DSB repair pathways: single-strand annealing (SSA) and breakage-induced replication (BIR) (Pardo et al., 2009; Jasin and Rothstein, 2013). SSA does not require a homologous template, and rejoining DNA ends with direct sequence repeats (Symington, 2014). BIR repairs one-ended DSBs, a process that is caused by the collapse of a replication fork (Mayle et al., 2015). When DSBs occur in cells, the first reaction is usually carried out in an NHEJ manner. Compared to other DNA repair and DNA recombination pathways, the NHEJ pathway is a robust, error-prone but predominant and fast pathway with high flexibility. It can recognize diverse end structures at DSBs and accomplish diverse repair results (Aravind and Koonin, 2001; Gu and Lieber, 2008; Salsman and Deltcheva, 2017). NHEJ can be classified into two types: canonical NHEJ (c-NHEJ) and alternative NHEJ

(alt-NHEJ), also called microhomology-mediated end-joining (MMEJ) (Bae et al., 2014). c-NHEJ is active throughout the cell cycle and stabilizes the DSB from translocations (Roth et al., 1995; Soutoglou et al., 2007). Based on different DNA ends, NHEJ is capable of employing different strategies. The whole process deals with assembling the core complex, which recognizes broken ends and keeps them together so that the following processing factors can act (Waters et al., 2014). The core complex is considered to include the Ku heterodimer (Ku80/70), the DNA-dependent protein kinase catalytic subunit (DNA-PKcs), DNA ligase IV, the X-ray repair cross-complementing protein 4 (XRCC4), the XRCC4-like factor (XLF, or Cernunnos). Ku is a heterodimer, composed of two subunits (70 and 83 kD), that recognizes and binds to blunt DSBs first (Walker et al., 2001). In c-NHEJ, Ku recruits DNA-PKcs to the DSB site and forms a very stable complex that remains bound to the end (Weterings et al., 2003). Their assembly activates the kinase activity of DNA-PK and orchestrates c-NHEJ (Davis et al., 2014; Radhakrishnan et al., 2014). DNA-PK phosphorylates a host of DNA damage response proteins and thus regulates c-NHEJ and DSB processing and recruits Artemis nuclease (Moshous et al., 2001). However, DNA-PK mostly phosphorylates itself, which is crucial in DSB processing (Neal et al., 2014). Artemis has 5′ to 3′ single-stranded DNA exonuclease activity and DNA-PKcs-dependent 5′ and 3′ endonuclease activity on hairpins and single-stranded overhangs (Moshous et al., 2001). Ku and DNA-PKcs alone can also promote multiple DNA end-processing activities at the break site. The X family of DNA polymerases (pol mu and pol lambda) adds missing nucleotides at the DSB ends (Daley et al., 2005; Paull, 2005). Next, the DSBs will be ligated by Ligase IV/XRCC4/XLF, which is regulated by DNA-PK. Ligase IV/XRCC4/XLF forms an extended filament that wraps and stabilizes DNA and stimulates ligation (Tsai et al., 2007; Andres et al., 2012). Recent research also showed that a newly identified PAXX (a paralog of XRCC4 and XLF), a member of the XRCC4 superfamily, is another important mediator of c-NHEJ, which interacts directly with Ku. In most cases DNA is repaired via the c-NHEJ pathway and its efficiency can approach nearly 90% (Yang et al., 2013; Dow et al., 2015), which constitutes the basis of CRISPR/Cas9 technology (Vartak and Raghavan, 2015). The NHEJ process is illustrated in **Figure 1**.

HDR PATHWAY

HDR is a faithful repair pathway. It comes into action mainly in the S- or G2-phase of the cell cycle and requires homologous DNA sequences. Homologous recombination is the desired mechanism for precise genome editing, which only happens in the presence of a homologous duplex template to repair the broken site. When DSB occurs, pathway choice depends on end resection (Symington and Gautier, 2011). The MRE11-RAD50-NBS1 (MRN, MRX in yeast) complex recognizes dsDNA and first creates a nick 15–20 bp from the 5′-ends of the DSB (Symington, 2014). Exonucleases such as SGS1-DNA2 and EXO1 complete the resection step (Kim and Mirkin, 2018). It then moves into flanking dsDNA regions and recruits ataxia telangiectasia mutated (ATM) kinase, the key upstream kinase of DSB signaling



(Falck et al., 2005), and interacts with CtIP (Makharashvili and Paull, 2015). MRN also tethers DNA ends, which increases its local concentration and thus facilitates ATM activation (Dupre et al., 2006).

The MRN complex consists of three subunits. MRE11 is a Mn^{2+} dependent nuclease involved in homologous recombination, telomere maintenance, and DNA DSB repair (Paull, 2015). SAE2 activates MRE11 for its dsDNA-specific endonuclease activity (Cannavo and Cejka, 2014) and regulates the resection step during appropriate stages of the cell cycle (Mathiasen and Lisby, 2014). RAD50 belongs to the structural maintenance of chromosomes (SMC) family, and contains ATPase activity (de Jager et al., 2001). RAD50 becomes dimerized and its DNA-binding activity is activated after ATP binding. Two MRE11 genes will then bind to the ATPase heads of the RAD50 homodimer, enabling itself to interact with RAD50 (Williams et al., 2008). RAD50 forms the core of MRN and uses its extended coiled-coil domain to tether DSB ends during HR (Williams et al.,

2010; Hohl et al., 2011). NBS1 contains a Fork-Head associated (FHA) domain and BRCA1 C-terminal (BRCT) domain, binds MRE11 and recruits ATM, linking the core MRN activities to DNA damage response (DDR) proteins domains at its N terminus (Glover et al., 2004; Williams et al., 2009).

Histone variant H2AX is phosphorylated by ATM, which becomes γ H2AX throughout the area surrounding the breakage within seconds after damage occurs (Rogakou et al., 1998). This sets off elaborate ubiquitylation and SUMOylation cascades to promote recruitment of BRCA1 (Morris et al., 2009) and 53BP1 (Stewart, 2009) but it is not crucial for the activation of ATM substrates such as CHK2 and p53 (Kang et al., 2005). Then MDC1, a large nuclear factor, directly binds to γ H2AX and functions as a molecular scaffold that interacts with ATM and NBS1, promoting further MRN accumulation. In addition, MDC1 also helps ATM spread on DSB-flanking chromatin and furthers H2AX activation (Spycher et al., 2008). It also mediates the accumulation of many DDR factors, including 53BP1 and BRCA1 (Wang et al., 2002; Stucki and Jackson, 2006; Kim et al., 2007). ATM leads to phosphorylation of DDR cascades such as BRCA1, Chk2, p53, etc., (Shiloh, 2003; Lavin, 2008). DSBs also activate ataxia telangiectasia and RAD3-related protein (ATR). Full ATR activation requires not only itself and DNA damage sensors but also proteins that function as signal transducers and effectors, such as RPA, RAD17, TopBP1, Claspin, and Chk1. ssDNA overhangs will be coated by replication protein A (RPA) rapidly, and the ssDNA-RPA complex acts as a scaffold to attract ATR/ATR-interacting protein (ATRIP) (Zou and Elledge, 2003) and other DNA damage checkpoint kinases to trigger DDR (Chen and Wold, 2014). It will then be replaced with adenosine triphosphate (ATP)-dependent recombinase RAD51 (San Filippo et al., 2008) with the help of BRCA1 and BRCA2 as described above (Prakash et al., 2015). ssDNA can be generated by nuclease resection, such as the MRN-C-terminal binding protein interacting protein (CtIP) for short resection, and EXO1/BLM for long resection (Mladenov et al., 2016). In mammals, resection depends on CtIP and needs to be phosphorylated by CDK first (Huertas and Jackson, 2009). BRCA1 contributes to HR by colocalizing with MRN after DNA damage occurs and interacts directly with the resection factor CtIP (Sartori et al., 2007). BRCA1 assists RAD51 binding to ssDNA by evicting RPA (Zelensky et al., 2014) and promotes the recruitment of BRCA2 to DSBs through the bridging protein PALB2 (Sy et al., 2009). BRCA1 also appears to inhibit the resection suppressor 53BP1 (Bunting et al., 2010). RAD51 is a DNA strand-exchange protein that exists in mammalian cells and forms a filament referred to as the presynaptic complex (van der Heijden et al., 2007; Hilario et al., 2009). The assembly of a RAD51 nucleoprotein filament promotes homologous search by locating and pairing the 3'-overhang with a homologous duplex DNA and catalyzing strand invasion (termed single-end invasion, SEI) (Morrical, 2015; Ma et al., 2017). The two ends of the DSB are identical, but one end serves as the "first end," which searches for the homologous sequence and forms a displacement loops (D-loops) structure while the other end waits for the latter process (Kim and Mirkin, 2018). Besides RAD51, DNA strand exchange also requires RAD54 and RDH54/TID1, which performs this step by

stabilizing RAD51-ssDNA presynaptic filaments (Mazin et al., 2003).

Resolution of the exchanged DNA strands includes the Holliday Junction (HJ) pathway and the synthesis-dependent strand annealing (SDSA) pathway. Dissolution is the primary pathway for HJ resolution, which involves the BLM helicase-Topoisomerase III α -RMI1-RMI2 (BTR) complex. The BTR complex promotes branch migration of Holliday junctions (Karow et al., 2000) and also acts to suppress crossing over during homologous recombination (Wu and Hickson, 2003). Thus, this dissolution pathway gives rise exclusively to non-crossovers. The other pathways use structure-selective resolvases (SLX-MUS complex and GEN1) to process the exchange intermediates and can produce both crossover and non-crossover products (West et al., 2015). The SLX1 and MUS81-EME1 nucleases bind in close proximity on the SLX4 scaffold and process HJs (Castor et al., 2013). SLX1 catalyzes the initial incision and MUS81 introduces the second cut on the opposing strand (Wyatt and West, 2014). GEN1 is a member of the RAD2/XPG family and can only access and cleave recombination intermediates when the nuclear membrane breaks down (Rass et al., 2010). GEN1 first forms a dimeric complex that contains the two active sites and then performs a dual symmetric incision at HJs, generating nicked duplex products that can be ligated.

The SDSA pathway also begins with the generation of a D-loop structure like the HJ pathway but also includes DNA synthesis in the 3'-direction, which extends the heteroduplex (Daley et al., 2014). The translocating D-loop then collapses, and the other resected DSB end will anneal to this extended DSB end. Both ends will go through replicative extension and ligation, which generates non-crossover products.

When it comes to single-stranded template repair (SSTR), the repair mechanism is quite different from the dsDNA repair template scenario. Richardson et al. found that human Cas9-induced SSTR requires the Fanconi anemia (FA) pathway, which was previously implicated in responses to interstrand crosslinks rather than nuclease-induced breaks (Richardson et al., 2018). They confirmed that SSTR is RAD51-independent while dsDNA donor HDR is RAD51-dependent. After FA pathway knockdown, the efficiency of SSTR decreased while simultaneously the levels of NHEJ increased, and the total editing stayed relatively constant. This means that the FA pathway can drive the repair events from NHEJ to SSTR. Additionally, FA pathway knockdown specifically inhibits SSTR and has no effect on NHEJ. RAD51C and XRCC3 are required for SSTR, but RAD51B and XRCC2 are not. They also found that FANCD2, a central player in the FA pathway, enriched even in the absence of an exogenous homology donor. In short, SSTR is much more efficient than HDR from a dsDNA donor but still needs future investigations. The HDR pathway is demonstrated in **Figure 2**.

FAVORING THE HDR PATHWAY USING CHEMICAL AND GENETIC MODULATION

DSBs caused by Cas9 can go through both the NHEJ and HDR pathways, but in most cases they are handled by NHEJ (Frit

et al., 2014), so it seems reasonable to inhibit key enzymes (e.g., DNA ligase IV) of the NHEJ pathway. Maruyama et al. (2015) investigated the effect of SCR7, a putative inhibitor of ligase IV, which targets the DNA binding domain of ligase IV, impeding its ability to bind to DSBs (Srivastava et al., 2012) in human epithelial (A549) and melanoma (MelJuSo) cell lines. Results showed that SCR7 promotes a 21 bp precise insertion (ssDNA donor with two 100 bp homology arms) in A549 cells 3-fold at 0.01 μ M and 19-fold for MelJuSo at 1 μ M. They also assessed the effect of SCR7 on insertion of a \sim 800 bp fragment (ssDNA donor with \sim 80 bp homology flanking sequence on both sides of the DSB) into a murine bone marrow-derived dendritic cell line (DC2.4 cells). After treating the DC2.4 cells with 1 μ M SCR7, the efficiency of insertion increased by \sim 13-fold. It is worth mentioning that SCR7 affects lymphocyte development and can induce apoptosis. As additional approaches to DNA ligase IV inhibition, Chu et al. used the adenovirus 4 (Ad4) E1B55K and E4orf6 proteins to suppress NHEJ. These two proteins can mediate the ubiquitination and proteasomal degradation of DNA ligase IV (Forrester et al., 2011). Results showed that HDR efficiency was enhanced up to 7-fold (5 to 36%) by the Ad4 protein in human HEK293 cells. And in a mouse Burkitt lymphoma cell line, the addition of Ad4 proteins reduced transfection efficiency from 40 to 27%, but promoted HDR by 5-fold (Chu et al., 2015). Yu et al. identified two small molecules (L755507 and Brefeldin A) that could improve HDR efficiency. L755507, a β 3-adrenergic receptor agonist, can increase HDR by 3-fold at 5 μ M in mouse ESCs. Brefeldin A, an inhibitor of intracellular protein transport from the ER to the Golgi apparatus, promotes HDR by 2-fold at 0.1 μ M in mouse ESCs (Yu et al., 2015).

Pinder et al. identified that RS-1 can enhance HDR between 3- and 6-fold, varying with the locus and transfection factor in HEK-293A human embryonic kidney and U2OS osteosarcoma cell lines (Pinder et al., 2015). RS-1 is a compound that stabilizes the association between Rad51 and DNA. They also found that an optimized ratio for Cas9/gRNA to homology donor plasmid doubled the HDR efficiency. Upon BRCA1 over-expression, HDR is increased by 2- to 3-fold.

Besides these inhibiting molecules, HDR can also be promoted using small interfering RNA (siRNA) to inhibit the expression of Ku protein, which is the pioneer protein in the NHEJ pathway. Li et al. assessed this method on pig fetal fibroblasts (Li et al., 2018). The result showed that by inhibiting Ku70 or Ku80 expression, HR can be promoted by 1.6- to 3-fold, as well as SSA and ssODN-mediated repairs. Yu et al. constructed a Rad51 and a Rad50 co-expression vector to evaluate its performance (Yu et al., 2011). It was determined that HR efficiency increased 110–245%. Chu et al. used short hairpin RNA (shRNA) in HEK293 cells to suppress key NHEJ pathway proteins such as KU70, KU80, and DNA ligase IV. Results showed that HDR efficiency was enhanced from 5 to 8–14% when transfected with single shRNAs against KU70, KU80, or DNA ligase IV. Furthermore, they found that the expression of the target gene diminished in the cells undergoing NHEJ blockade, which may be caused by local chromatin remodeling through extended DNA damage (Chu et al., 2015).

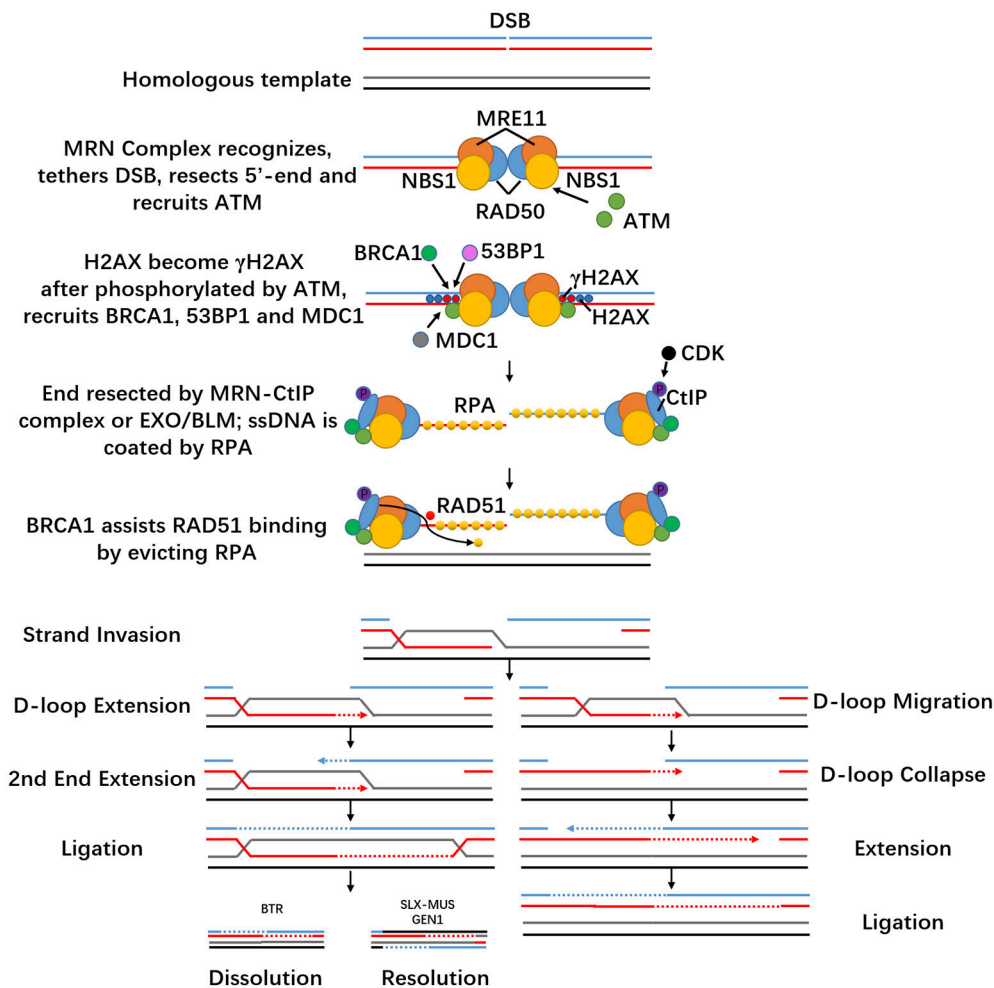


FIGURE 2 | Homology-directed repair HDR. When DSB happens in the S- or G2-phase of the cell cycle and homologous sequence exists near the DSB, DSB can be handled through the HDR pathway if the ends of the DSB are resected. Ends will be coated with various proteins and then invade homologous duplex DNA to form an exchange intermediate: the D-loop structure. Most D-loop structures will be extended by DNA synthesis (dashed arrow). The second end pairs to the D-loop and starts extension. This pathway is called the double Holliday junction pathway. Ligation generates the characteristic double Holliday junction, which may be cleaved by HJ resolvases into either crossover or non-crossover products. The synthesis-dependent strand annealing pathway is illustrated on the right. After D-loop formation, replication and branch migration take place which can lead to D-loop translocation. The translocating D-loop is unstable and collapses easily. After collapse, the extended first end may anneal to complementary ssDNA in the resected second end. Replicative extension of both ends and ligation generates non-crossover products.

TIMED DELIVERY OF THE CRISPR-CAS9 SYSTEM

While HDR is typically restricted to the S and G2 phases of the cell cycle, its efficiency can be increased by synchronizing and capturing cells at the S and G2 phases or using timed delivery. Lin et al. combined cell cycle synchronization techniques, using chemical inhibitors to arrest the cells at specific phases of the cell cycle with direct nucleofection of pre-assembled Cas9 ribonucleoprotein (RNP) in HEK293T cells (Lin et al., 2014). Results showed that using lower Cas9/RNP concentrations and cell cycle arrest can improve HDR efficiency to 31% (3.4-fold) at maximum. Consistent with this strategy, Yang et al.

successfully enhanced HDR by three- to six-fold using the microtubule polymerization inhibitor nocodazole or ABT-751 (Yang et al., 2016). As for non-proliferating cells, BRCA1 is inhibited by the 53BP1 (Escribano-Diaz et al., 2013) and KEAP1-CUL3 complexes (Orthwein et al., 2015), and RIF1 thus will not bind to DSB. In addition, CtIP can function normally after being phosphorylated by CDK, but CDK is absent when cells stay in the G0/G1 phase (Escribano-Diaz et al., 2013). In order to overcome this inhibition, Orthwein et al. overexpressed mutated activated CtIP. This depleted the 53BP1 and KEAP1-CUL3 complexes simultaneously, which successfully activated the HDR pathway in G1 cells (Orthwein et al., 2015).

ENHANCING HDR BY USING OVERLAPPING SEQUENCES

Several types of donor DNA have been used, such as plasmid DNA and synthetic oligonucleotides (Carroll and Beumer, 2014). Rational design of homology repair templates strongly enhances HDR efficiency (Renaud et al., 2016). By using a linear repair template with homologous flanks in zebrafish, HDR can increase by almost 10-fold (Irion et al., 2014). Chu et al. assessed the influence of the lengths of homology regions of the repair template on HDR efficiency (Chu et al., 2015). A homologous template is the most important component of HDR-mediated genome editing. It usually contains intended mutations or insertions flanked by homologous regions. Templates can be plasmids (up to kilobases modification) or single-stranded oligodeoxynucleotides (ssODN) (50–100 nt modification). It is suggested that sequencing around the interested region should be carried out because cell-specific mutations and single nucleotide polymorphisms (SNPs) can influence gRNA targeting up to 6-fold, merely one mismatch in 100 bases (Tham et al., 2016). This problem can be overcome by amplifying the homology arm from the genomic DNA extracted from target cells or by synthesizing consulting sequence analysis (Salsman and Deltaire, 2017). It is important to remember that DSB should always be as close as possible to the region of homology, within 10 nt up and to a maximum of 100 nt (Elliott et al., 1998). Furthermore, having each homology arm about 50–100% the size of the payload that can promote HDR (Li et al., 2014).

ENHANCING HDR BY USING MODIFIED CAS9

As mentioned above, DSB ends must be resected so that they can enter the HDR pathway. CtIP, a key protein in early steps of DSB resection, is essential for HDR initiation. In order to ensure the presence of functional activated CtIP, Charpentier et al. fused a minimal N-terminal fragment of CtIP to the Cas9 protein (Charpentier et al., 2018). Forcing CtIP to the DNA cleavage site, through fusion to either catalytically dead Cas9 (dCas9) or Cas9 together with 800 bp homology arms, obtained a 2-fold increase in the efficiency of HDR in human fibroblast RG37DR cells, iPS cells, and rat zygotes. However, the expression of dCas9-CtIP is not sufficient by itself to stimulate integration. In addition, the patterns of indels induced by modified Cas9 were different from Cas9. This may be because the modified Cas9 induced a different balance of the end-joining pathway.

As for point mutation, current approaches to target base correction are inefficient and typically induce an abundance of random insertions and deletions at the target locus. Alexis et al. reported a powerful approach called a “base-editing (BE) system” to introduce specific point mutations without introducing DSB or a donor template by linking deaminases and dCas9 together (Komor et al., 2016). dCas9 helps deaminase to locate and

deaminases modify cytidine to uridine. This conversion will then be repaired through various pathways (Hess et al., 2017). BE2 (optimized BE system) results from the addition of the uracil DNA glycosylase inhibitor, which increases the base-editing efficiency of the C>T substitution 3-fold (Hess et al., 2016). Additionally, the BE3 system was realized by changing the dCas9 to Cas9 D10A. This improvement achieved a 6-fold increase in efficiency over BE2 but exhibited a slightly increased indel frequency as nicks can lead to NHEJ at a low rate (Certo et al., 2011). Target-activation-induced cytidine deaminase (AID), a similar system, uses nickase Cas9 D10A to recruit the cytidine deaminase PmCDA1 to the target, achieving a mutation frequency of 35%. Adding UGI can obtain 2- to 3-fold increase in efficiency and reduction in deletions (Nishida et al., 2016). In addition, targeted AID-mediated mutagenesis (TAM) (Ma et al., 2016) and CRISPR-X (Hess et al., 2016) can generate transitions and transversions. These systems can both achieve the precision editing of single C>T bases with a low rate of indels as well as sequence diversification. By further Cas9 engineering, the inhibition of HR during the G1 phase will be overcome.

The CRISPR-Cas system has been fully studied and adapted for various applications over the decades, which gives us the ability to manipulate the genome as we desire. It has certain limitations, such as off-target effects (which can be overcome by rational sgRNA design; Doench et al., 2016) and low efficiency, which can be improved by utilizing the methodologies as described above. These promising strategies have proven their enhancement in the HDR pathway more than once with results varying from 2- to 30-fold. Combining various approaches can be a potential method of maximizing the rates of HDR. Here we only reviewed NHEJ inhibition by using inhibitors or hindering certain gene expression with siRNA or shRNA, CRISPR-Cas delivery in the G2/S phase, adding homologous arms in donor templates and using modified Cas9. These tactics surely make the CRISPR-Cas system more efficient. There is no doubt that more and more ways to boost CRISPR-Cas are imminent.

AUTHOR CONTRIBUTIONS

ML, WM, and DC designed the structure of this review. ML wrote the first version of the manuscript. SR, XT, KG, and QF helped to revise the manuscript. All authors have reviewed the final version of the manuscript.

FUNDING

This research was supported by the Qinghai Province Major R&D and Transformation Project (2018-NK-125), Xianyang Science and Technology Major Project (2017K01-34), Key Industrial Innovation Chains of Shaanxi Province (2018ZDCXL-NY-01-06), and the PhD research startup fund of the Northwest Agriculture and Forestry University (00500/Z109021716).

REFERENCES

- Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513, 569–573. doi: 10.1038/nature13579
- Andres, S. N., Vergnes, A., Ristic, D., Wyman, C., Modesti, M., and Junop, M. (2012). A human XRCC4-XLF complex bridges DNA. *Nucleic Acids Res.* 40, 1868–1878. doi: 10.1093/nar/gks022
- Aravind, L., and Koonin, E. V. (2001). Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res.* 11, 1365–1374. doi: 10.1101/gr.181001
- Bae, S., Kweon, J., Kim, H. S., and Kim, J. S. (2014). Microhomology-based choice of Cas9 nuclease target sites. *Nat. Methods* 11, 705–706. doi: 10.1038/nmeth.3015
- Barrangou, R. (2013). CRISPR-Cas systems and RNA-guided interference. *Wiley Interdiscip. Rev. RNA* 4, 267–278. doi: 10.1002/wrna.1159
- Barrangou, R. (2015). The roles of CRISPR-Cas systems in adaptive immunity and beyond. *Curr. Opin. Immunol.* 32, 36–41. doi: 10.1016/j.coi.2014.12.008
- Barrangou, R., and Marraffini, L. A. (2014). CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol. Cell* 54, 234–244. doi: 10.1016/j.molcel.2014.03.011
- Bunting, S. F., Callen, E., Wong, N., Chen, H. T., Polato, F., Gunn, A., et al. (2010). 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell* 141, 243–254. doi: 10.1016/j.cell.2010.03.012
- Cannavo, E., and Cejka, P. (2014). Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks. *Nature* 514, 122–125. doi: 10.1038/nature13771
- Carroll, D., and Beumer, K. J. (2014). Genome engineering with TALENs and ZFNs: repair pathways and donor design. *Methods* 69, 137–141. doi: 10.1016/j.ymeth.2014.03.026
- Castor, D., Nair, N., Declais, A. C., Lachaud, C., Toth, R., Macartney, T. J., et al. (2013). Cooperative control of holliday junction resolution and DNA repair by the SLX1 and MUS81-EME1 nucleases. *Mol. Cell* 52, 221–233. doi: 10.1016/j.molcel.2013.08.036
- Certo, M. T., Ryu, B. Y., Annis, J. E., Garibov, M., Jarjour, J., Rawlings, D. J., et al. (2011). Tracking genome engineering outcome at individual DNA breakpoints. *Nat. Methods* 8, 671–676. doi: 10.1038/nmeth.1648
- Charpentier, M., Khedher, A. H. Y., Menoret, S., Brion, A., Lamribet, K., Dardillac, E., et al. (2018). CtfP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nat. Commun.* 9:1133. doi: 10.1038/s41467-018-03475-7
- Chen, R., and Wold, M. S. (2014). Replication protein a: single-stranded DNA's first responder: dynamic DNA-interactions allow replication protein A to direct single-strand DNA intermediates into different pathways for synthesis or repair. *Bioessays* 36, 1156–1161. doi: 10.1002/bies.201400107
- Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., et al. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* 33, 543–548. doi: 10.1038/nbt.3198
- Daley, J. M., Gaines, W. A., Kwon, Y., and Sung, P. (2014). Regulation of DNA pairing in homologous recombination. *Cold Spring Harb. Perspect. Biol.* 6:a017954. doi: 10.1101/cshperspect.a017954
- Daley, J. M., Laan, R. L., Suresh, A., and Wilson, T. E. (2005). DNA joint dependence of pol X family polymerase action in nonhomologous end joining. *J. Biol. Chem.* 280, 29030–29037. doi: 10.1074/jbc.M505277200
- Davis, A. J., Chen, B. P., and Chen, D. J. (2014). DNA-PK: a dynamic enzyme in a versatile DSB repair pathway. *DNA Repair* 7, 21–29. doi: 10.1016/j.dnarep.2014.02.020
- de Jager, M., van Noort, J., van Gent, D. C., Dekker, C., Kanaar, R., and Wyman, C. (2001). Human Rad50/Mre11 is a flexible complex that can tether DNA ends. *Mol. Cell* 8, 1129–1135. doi: 10.1016/S1097-2765(01)00381-1
- Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., et al. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602–607. doi: 10.1038/nature09886
- Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34, 184–191. doi: 10.1038/nbt.3437
- Dow, L. E., Fisher, J., O'Rourke, K. P., Muley, A., Kastenhuber, E. R., Livshits, G., et al. (2015). Inducible *in vivo* genome editing with CRISPR-Cas9. *Nat. Biotechnol.* 33, 390–394. doi: 10.1038/nbt.3155
- Dupre, A., Boyer-Chatenet, L., and Gautier, J. (2006). Two-step activation of ATM by DNA and the Mre11-Rad50-Nbs1 complex. *Nat. Struct. Mol. Biol.* 13, 451–457. doi: 10.1038/nsmb1090
- Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J. A., and Jasin, M. (1998). Gene conversion tracts from double-strand break repair in mammalian cells. *Mol. Cell Biol.* 18, 93–101.
- Escobedo-Diaz, C., Orthwein, A., Fradet-Turcotte, A., Xing, M., Young, J. T., Tkac, J., et al. (2013). A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. *Mol. Cell* 49, 872–883. doi: 10.1016/j.molcel.2013.01.001
- Falck, J., Coates, J., and Jackson, S. P. (2005). Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature* 434, 605–611. doi: 10.1038/nature03442
- Forrester, N. A., Sedgwick, G. G., Thomas, A., Blackford, A. N., Speiseder, T., Dobner, T., et al. (2011). Serotype-specific inactivation of the cellular DNA damage response during adenovirus infection. *J. Virol.* 85, 2201–2211. doi: 10.1128/JVI.01748-10
- Frit, P., Barboule, N., Yuan, Y., Gomez, D., and Calsou, P. (2014). Alternative end-joining pathway(s): bricolage at DNA breaks. *DNA Repair* 17, 81–97. doi: 10.1016/j.dnarep.2014.02.007
- Glover, J. N., Williams, R. S., and Lee, M. S. (2004). Interactions between BRCT repeats and phosphoproteins: tangled up in two. *Trends Biochem. Sci.* 29, 579–585. doi: 10.1016/j.tibs.2004.09.010
- Gu, J., and Lieber, M. R. (2008). Mechanistic flexibility as a conserved theme across 3 billion years of nonhomologous DNA end-joining. *Genes Dev.* 22, 411–415. doi: 10.1101/gad.1646608
- Hess, G. T., Fresard, L., Han, K., Lee, C. H., Li, A., Cimprich, K. A., et al. (2016). Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* 13, 1036–1042. doi: 10.1038/nmeth.4038
- Hess, G. T., Tycko, J., Yao, D., and Bassik, M. C. (2017). Methods and applications of CRISPR-mediated base editing in eukaryotic genomes. *Mol. Cell* 68, 26–43. doi: 10.1016/j.molcel.2017.09.029
- Hilario, J., Amitani, I., Baskin, R. J., and Kowalczykowski, S. C. (2009). Direct imaging of human Rad51 nucleoprotein dynamics on individual DNA molecules. *Proc. Natl. Acad. Sci. U.S.A.* 106, 361–368. doi: 10.1073/pnas.0811965106
- Hohl, M., Kwon, Y., Galvan, S. M., Xue, X., Tous, C., Aguilera, A., et al. (2011). The Rad50 coiled-coil domain is indispensable for Mre11 complex functions. *Nat. Struct. Mol. Biol.* 18, 1124–1131. doi: 10.1038/nsmb.2116
- Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327, 167–170. doi: 10.1126/science.1179555
- Huertas, P., and Jackson, S. P. (2009). Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. *J. Biol. Chem.* 284, 9558–9565. doi: 10.1074/jbc.M808906200
- Irion, U., Krauss, J., and Nusslein-Volhard, C. (2014). Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. *Development* 141, 4827–4830. doi: 10.1242/dev.115584
- Jansen, R., Embden, J. D., Gastra, W., and Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* 43, 1565–1575. doi: 10.1046/j.1365-2958.2002.02839.x
- Jasin, M., and Rothstein, R. (2013). Repair of strand breaks by homologous recombination. *Cold Spring Harb. Perspect. Biol.* 5:a012740. doi: 10.1101/cshperspect.a012740
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829
- Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., et al. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343:1247997. doi: 10.1126/science.1247997
- Kang, J., Ferguson, D., Song, H., Bassing, C., Eckersdorff, M., Alt, F. W., et al. (2005). Functional interaction of H2AX, NBS1, and p53 in ATM-dependent DNA damage responses and tumor suppression. *Mol. Cell Biol.* 25, 661–670. doi: 10.1128/MCB.25.2.661-670.2005

- Karow, J. K., Constantinou, A., Li, J. L., West, S. C., and Hickson, I. D. (2000). The Bloom's syndrome gene product promotes branch migration of holliday junctions. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6504–6508. doi: 10.1073/pnas.100448097
- Kim, H., Chen, J., and Yu, X. (2007). Ubiquitin-binding protein RAP80 mediates BRCA1-dependent DNA damage response. *Science* 316, 1202–1205. doi: 10.1126/science.1139621
- Kim, K. P., and Mirkin, E. V. (2018). So similar yet so different: the two ends of a double strand break. *Mutat. Res.* 809, 70–80. doi: 10.1016/j.mrfmmm.2017.06.007
- Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424. doi: 10.1038/nature17946
- Lavin, M. F. (2008). Ataxia-telangiectasia: from a rare disorder to a paradigm for cell signalling and cancer. *Nat. Rev. Mol. Cell Biol.* 9, 759–769. doi: 10.1038/nrm2514
- Li, G., Liu, D., Zhang, X., Quan, R., Zhong, C., Mo, J., et al. (2018). Suppressing Ku70/Ku80 expression elevates homology-directed repair efficiency in primary fibroblasts. *Int. J. Biochem. Cell Biol.* 99, 154–160. doi: 10.1016/j.biocel.2018.04.011
- Li, K., Wang, G., Andersen, T., Zhou, P., and Pu, W. T. (2014). Optimization of genome engineering approaches with the CRISPR/Cas9 system. *PLoS ONE* 9:e105779. doi: 10.1371/journal.pone.0105779
- Lin, S., Staahl, B. T., Alla, R. K., and Doudna, J. A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* 3:e04766. doi: 10.7554/eLife.04766
- Ma, C. J., Gibb, B., Kwon, Y., Sung, P., and Greene, E. C. (2017). Protein dynamics of human RPA and RAD51 on ssDNA during assembly and disassembly of the RAD51 filament. *Nucleic Acids Res.* 45, 749–761. doi: 10.1093/nar/gkw1125
- Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., and Chang, X. (2016). Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* 13, 1029–1035. doi: 10.1038/nmeth.4027
- Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., et al. (2011). Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* 9, 467–477. doi: 10.1038/nrmicro2577
- Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736. doi: 10.1038/nrmicro3569
- Makharashvili, N., and Paull, T. T. (2015). CtIP: A DNA damage response protein at the intersection of DNA metabolism. *DNA Repair* 32, 75–81. doi: 10.1016/j.dnarep.2015.04.016
- Marraffini, L. A. (2015). CRISPR-Cas immunity in prokaryotes. *Nature* 526, 55–61. doi: 10.1038/nature15386
- Marraffini, L. A., and Sontheimer, E. J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.* 11, 181–190. doi: 10.1038/nrg2749
- Maruyama, T., Dougan, S. K., Truttmann, M. C., Bilate, A. M., Ingram, J. R., and Ploegh, H. L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* 33, 538–542. doi: 10.1038/nbt.3190
- Mathiasen, D. P., and Lisby, M. (2014). Cell cycle regulation of homologous recombination in *Saccharomyces cerevisiae*. *FEBS Microbiol. Rev.* 38, 172–184. doi: 10.1111/1574-6976.12066
- Mayle, R., Campbell, I. M., Beck, C. R., Yu, Y., Wilson, M., Shaw, C. A., et al. (2015). DNA REPAIR. Mus81 and converging forks limit the mutagenicity of replication fork breakage. *Science* 349, 742–747. doi: 10.1126/science.aaa8391
- Mazin, A. V., Alexeev, A. A., and Kowalczykowski, S. C. (2003). A novel function of Rad54 protein. stabilization of the Rad51 nucleoprotein filament. *J. Biol. Chem.* 278, 14029–14036. doi: 10.1074/jbc.M212779200
- Mladenov, E., Magin, S., Soni, A., and Iliakis, G. (2016). DNA double-strand-break repair in higher eukaryotes and its role in genomic instability and cancer: cell cycle and proliferation-dependent regulation. *Semin. Cancer Biol.* 37–38, 51–64. doi: 10.1016/j.semcancer.2016.03.003
- Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155, 733–740. doi: 10.1099/mic.0.023960-0
- Mojica, F. J., and Rodriguez-Valera, F. (2016). The discovery of CRISPR in archaea and bacteria. *FEBS J.* 283, 3162–3169. doi: 10.1111/febs.13766
- Morange, M. (2015). What history tells us XXXVII. CRISPR-Cas: the discovery of an immune system in prokaryotes. *J. Biosci.* 40, 221–223. doi: 10.1007/s12038-015-9532-6
- Morrill, S. W. (2015). DNA-pairing and annealing processes in homologous recombination and homology-directed repair. *Cold Spring Harb. Perspect. Biol.* 7:a016444. doi: 10.1101/cshperspect.a016444
- Morris, J. R., Boutell, C., Keppler, M., Densham, R., Weekes, D., Alamshah, A., et al. (2009). The SUMO modification pathway is involved in the BRCA1 response to genotoxic stress. *Nature* 462, 886–890. doi: 10.1038/nature08593
- Moshous, D., Callebaut, I., de Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., et al. (2001). Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell* 105, 177–186. doi: 10.1016/S0092-8674(01)00309-9
- Neal, J. A., Sugiman-Marangos, S., VanderVere-Carozza, P., Wagner, M., Turchi, J., Lees-Miller, S. P., et al. (2014). Unraveling the complexities of DNA-dependent protein kinase autophosphorylation. *Mol. Cell Biol.* 34, 2162–2175. doi: 10.1128/MCB.01554-13
- Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353:aaf8729. doi: 10.1126/science.aaf8729
- Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., et al. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156, 935–949. doi: 10.1016/j.cell.2014.02.001
- Orthwein, A., Noordermeer, S. M., Wilson, M. D., Landry, S., Enchev, R. I., Sherker, A., et al. (2015). A mechanism for the suppression of homologous recombination in G1 cells. *Nature* 528, 422–426. doi: 10.1038/nature16142
- Pardo, B., Gomez-Gonzalez, B., and Aguilera, A. (2009). DNA repair in mammalian cells: DNA double-strand break repair: how to fix a broken relationship. *Cell Mol. Life Sci.* 66, 1039–1056. doi: 10.1007/s00018-009-8740-3
- Paull, T. T. (2005). Saving the ends for last: the role of pol mu in DNA end joining. *Mol. Cell* 19, 294–296. doi: 10.1016/j.molcel.2005.07.008
- Paull, T. T. (2015). Mechanisms of ATM Activation. *Annu. Rev. Biochem.* 84, 711–738. doi: 10.1146/annurev-biochem-060614-034335
- Pinder, J., Salsman, J., and Deltre, G. (2015). Nuclear domain 'knock-in' screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. *Nucleic Acids Res.* 43, 9379–9392. doi: 10.1093/nar/gkv993
- Prakash, R., Zhang, Y., Feng, W., and Jasin, M. (2015). Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins. *Cold Spring Harb. Perspect. Biol.* 7:a016600. doi: 10.1101/cshperspect.a016600
- Radhakrishnan, S. K., Jette, N., and Lees-Miller, S. P. (2014). Non-homologous end joining: emerging themes and unanswered questions. *DNA Repair* 17, 2–8. doi: 10.1016/j.dnarep.2014.01.009
- Rass, U., Compton, S. A., Matos, J., Singleton, M. R., Ip, S. C., Blanco, M. G., et al. (2010). Mechanism of Holliday junction resolution by the human GEN1 protein. *Genes Dev.* 24, 1559–1569. doi: 10.1101/gad.585310
- Renaud, J. B., Boix, C., Charpentier, M., De Cian, A., Cochenne, J., Duvernois-Berthet, E., et al. (2016). Improved genome editing efficiency and flexibility using modified oligonucleotides with TALEN and CRISPR-Cas9 Nucleases. *Cell Rep.* 14, 2263–2272. doi: 10.1016/j.celrep.2016.02.018
- Richardson, C. D., Kazane, K. R., Feng, S. J., Zelin, E., Bray, N. L., Schafer, A. J., et al. (2018). CRISPR-Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. *Nat. Genet.* 50, 1132–1139. doi: 10.1038/s41588-018-0174-0
- Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S., and Bonner, W. M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.* 273, 5858–5868.
- Roth, D. B., Lindahl, T., and Gellert, M. (1995). Repair and recombination. how to make ends meet. *Curr. Biol.* 5, 496–499.
- Salsman, J., and Deltre, G. (2017). Precision genome editing in the CRISPR era. *Biochem. Cell Biol.* 95, 187–201. doi: 10.1139/bcb-2016-0137
- San Filippo, J., Sung, P., and Klein, H. (2008). Mechanism of eukaryotic homologous recombination. *Annu. Rev. Biochem.* 77, 229–257. doi: 10.1146/annurev-biochem.77.061306.125255
- Sapranaukas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* 39, 9275–9282. doi: 10.1093/nar/gkr606

- Sartori, A. A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., et al. (2007). Human CtIP promotes DNA end resection. *Nature* 450, 509–514. doi: 10.1038/nature06337
- Shiloh, Y. (2003). ATM and related protein kinases: safeguarding genome integrity. *Nat. Rev. Cancer* 3, 155–168. doi: 10.1038/nrc1011
- Sorek, R., Lawrence, C. M., and Wiedenheft, B. (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu. Rev. Biochem.* 82, 237–266. doi: 10.1146/annurev-biochem-072911-172315
- Soutoglou, E., Dorn, J. F., Sengupta, K., Jasin, M., Nussenzweig, A., Ried, T., et al. (2007). Positional stability of single double-strand breaks in mammalian cells. *Nat. Cell Biol.* 9, 675–682. doi: 10.1038/ncb1591
- Spycher, C., Miller, E. S., Townsend, K., Pavic, L., Morrice, N. A., Janscak, P., et al. (2008). Constitutive phosphorylation of MDC1 physically links the MRE11-RAD50-NBS1 complex to damaged chromatin. *J. Cell Biol.* 181, 227–240. doi: 10.1083/jcb.200709008
- Srivastava, M., Nambiar, M., Sharma, S., Karki, S. S., Goldsmith, G., Hegde, M., et al. (2012). An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression. *Cell* 151, 1474–1487. doi: 10.1016/j.cell.2012.11.054
- Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., and Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507, 62–67. doi: 10.1038/nature13011
- Stewart, G. S. (2009). Solving the RIDDLE of 53BP1 recruitment to sites of damage. *Cell Cycle* 8, 1532–1538. doi: 10.4161/cc.8.10.8351
- Stucki, M., and Jackson, S. P. (2006). gammaH2AX and MDC1: anchoring the DNA-damage-response machinery to broken chromosomes. *DNA Repair* 5, 534–543. doi: 10.1016/j.dnarep.2006.01.012
- Sy, S. M., Huen, M. S., and Chen, J. (2009). PALB2 is an integral component of the BRCA complex required for homologous recombination repair. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7155–7160. doi: 10.1073/pnas.0811159106
- Symington, L. S. (2014). End resection at double-strand breaks: mechanism and regulation. *Cold Spring Harb. Perspect. Biol.* 6:a016436. doi: 10.1101/cshperspect.a016436
- Symington, L. S., and Gautier, J. (2011). Double-strand break end resection and repair pathway choice. *Annu. Rev. Genet.* 45, 247–271. doi: 10.1146/annurev-genet-110410-132435
- Szczelkun, M. D., Tikhomirova, M. S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., et al. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9798–9803. doi: 10.1073/pnas.1402597111
- Tham, K. C., Kanaar, R., and Lebbink, J. H. (2016). Mismatch repair and homologous recombination. *DNA Repair* 38, 75–83. doi: 10.1016/j.dnarep.2015.11.010
- Tsai, C. J., Kim, S. A., and Chu, G. (2007). Cernunnos/XLF promotes the ligation of mismatched and noncohesive DNA ends. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7851–7856. doi: 10.1073/pnas.0702620104
- van der Heijden, T., Seidel, R., Modesti, M., Kanaar, R., Wyman, C., and Dekker, C. (2007). Real-time assembly and disassembly of human RAD51 filaments on individual DNA molecules. *Nucleic Acids Res.* 35, 5646–5657. doi: 10.1093/nar/gkm629
- van der Oost, J., Jore, M. M., Westra, E. R., Lundgren, M., and Brouns, S. J. (2009). CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.* 34, 401–407. doi: 10.1016/j.tibs.2009.05.002
- Vartak, S. V., and Raghavan, S. C. (2015). Inhibition of nonhomologous end joining to increase the specificity of CRISPR/Cas9 genome editing. *FEBS J.* 282, 4289–4294. doi: 10.1111/febs.13416
- Walker, J. R., Corpina, R. A., and Goldberg, J. (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* 412, 607–614. doi: 10.1038/35088000
- Wang, B., Matsuoka, S., Carpenter, P. B., and Elledge, S. J. (2002). 53BP1, a mediator of the DNA damage checkpoint. *Science* 298, 1435–1438. doi: 10.1126/science.1076182
- Waters, C. A., Strande, N. T., Wyatt, D. W., Pryor, J. M., and Ramsden, D. A. (2014). Nonhomologous end joining: a good solution for bad ends. *DNA Repair* 17, 39–51. doi: 10.1016/j.dnarep.2014.02.008
- West, S. C., Blanco, M. G., Chan, Y. W., Matos, J., Sarbajna, S., and Wyatt, H. D. (2015). Resolution of recombination intermediates: mechanisms and regulation. *Cold Spring Harb. Symp. Q. Biol.* 80, 103–109. doi: 10.1101/sqb.2015.80.027649
- Weterings, E., Verkaik, N. S., Bruggenwirth, H. T., Hoeijmakers, J. H., and van Gent, D. C. (2003). The role of DNA dependent protein kinase in synopsis of DNA ends. *Nucleic Acids Res.* 31, 7238–7246. doi: 10.1093/nar/gkg889
- Wiedenheft, B., Sternberg, S. H., and Doudna, J. A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482, 331–338. doi: 10.1038/nature10886
- Williams, G. J., Lees-Miller, S. P., and Tainer, J. A. (2010). Mre11-Rad50-Nbs1 conformations and the control of sensing, signaling, and effector responses at DNA double-strand breaks. *DNA Repair* 9, 1299–1306. doi: 10.1016/j.dnarep.2010.10.001
- Williams, R. S., Dodson, G. E., Limbo, O., Yamada, Y., Williams, J. S., Guenther, G., et al. (2009). Nbs1 flexibly tethers Ctp1 and Mre11-Rad50 to coordinate DNA double-strand break processing and repair. *Cell* 139, 87–99. doi: 10.1016/j.cell.2009.07.033
- Williams, R. S., Moncalian, G., Williams, J. S., Yamada, Y., Limbo, O., Shin, D. S., et al. (2008). Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair. *Cell* 135, 97–109. doi: 10.1016/j.cell.2008.08.017
- Wu, L., and Hickson, I. D. (2003). The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature* 426, 870–874. doi: 10.1038/nature02253
- Wyatt, H. D., and West, S. C. (2014). Holliday junction resolvases. *Cold Spring Harb. Perspect. Biol.* 6:a023192. doi: 10.1101/cshperspect.a023192
- Yang, D., Scavuzzo, M. A., Chmielowiec, J., Sharp, R., Bajic, A., and Borowiak, M. (2016). Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with customizable endonucleases. *Sci. Rep.* 6:21264. doi: 10.1038/srep21264
- Yang, H., Wang, H., Shivalila, C. S., Cheng, A. W., Shi, L., and Jaenisch, R. (2013). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* 154, 1370–1379. doi: 10.1016/j.cell.2013.08.022
- Yu, C., Liu, Y., Ma, T., Liu, K., Xu, S., Zhang, Y., et al. (2015). Small molecules enhance CRISPR genome editing in pluripotent stem cells. *Cell Stem Cell* 16, 142–147. doi: 10.1016/j.stem.2015.01.003
- Yu, S., Song, Z., Luo, J., Dai, Y., and Li, N. (2011). Over-expression of RAD51 or RAD54 but not RAD51/4 enhances extra-chromosomal homologous recombination in the human sarcoma (HT-1080) cell line. *J. Biotechnol.* 154, 21–24. doi: 10.1016/j.jbiotec.2011.03.023
- Zelensky, A., Kanaar, R., and Wyman, C. (2014). Mediators of homologous DNA pairing. *Cold Spring Harb. Perspect. Biol.* 6:a016451. doi: 10.1101/cshperspect.a016451
- Zou, L., and Elledge, S. J. (2003). Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* 300, 1542–1548. doi: 10.1126/science.1083430

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Liu, Rehman, Tang, Gu, Fan, Chen and Ma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



FTO Knockout Causes Chromosome Instability and G2/M Arrest in Mouse GC-1 Cells

Tao Huang, Qiang Gao, Tongying Feng, Yi Zheng, Jiayin Guo and Wenxian Zeng*

Laboratory of Reproductive Biology and Cell Engineering, College of Animal Science and Technology, Northwest A&F University, Xianyang, China

OPEN ACCESS

Edited by:

David Jay Segal,
University of California, Davis,
United States

Reviewed by:

Huabing Li,
Shanghai Jiao Tong University School
of Medicine, China
Milind Ratnaparkhe,
ICAR-Indian Institute of Soybean
Research, India
Masoud Zamani Esteki,
Maastricht University Medical Centre,
Netherlands

*Correspondence:

Wenxian Zeng
zengwenxian2015@126.com

Specialty section:

This article was submitted to
Genomic Assay Technology,
a section of the journal
Frontiers in Genetics

Received: 31 August 2018

Accepted: 22 December 2018

Published: 21 January 2019

Citation:

Huang T, Gao Q, Feng T, Zheng Y,
Guo J and Zeng W (2019) FTO
Knockout Causes Chromosome
Instability and G2/M Arrest in Mouse
GC-1 Cells. *Front. Genet.* 9:732.
doi: 10.3389/fgene.2018.00732

N^6 -methyladenosine (m^6A) is the most abundant modification on eukaryotic mRNA. m^6A plays important roles in the regulation of post-transcriptional RNA splicing, translation, and degradation. Increasing studies have uncovered the significance of m^6A in various biological processes such as stem cell fate determination, carcinogenesis, adipogenesis, stress response, etc, which put forwards a novel conception called epitranscriptome. However, functions of the fat mass and obesity-associated protein (FTO), the first characterized m^6A demethylase, in spermatogenesis remains obscure. Here we reported that depletion of FTO by CRISPR/Cas9 induces chromosome instability and G2/M arrest in mouse spermatogonia, which was partially rescued by expression of wild type FTO but not demethylase inactivated FTO. FTO depletion significantly decreased the expression of mitotic checkpoint complex and G2/M regulators. We further demonstrated that the m^6A modification on Mad1, Mad2, Bub1b, Cdk1, and Ccnb2 were directly targeted by FTO. Therefore, FTO regulates cell cycle and mitosis checkpoint in spermatogonia because of its m^6A demethylase activity. The findings give novel insights into the role of RNA methylation in spermatogenesis.

Keywords: N^6 -methyladenosine, FTO, spermatogonia, cell cycle, chromosome instability, mitotic checkpoint

INTRODUCTION

Life-long male fertility relies on spermatogenesis that is responsible for the generation of millions of sperm (Fok et al., 2014). Spermatogenesis is a complex developmental process that consists of three stages: mitosis of spermatogonia, meiosis of spermatocyte and transformation of sperm from haploid spermatids (Kanatsu-Shinohara and Shinohara, 2013). Thus, spermatogonia are the cornerstone of sperm production (Hamra et al., 2004). Nevertheless, the underlying mechanism regulating spermatogonial proliferation and differentiation remains largely elusive.

Over 100 different types of chemical modifications have been found in RNAs, among which the N^6 -methyladenosine (m^6A) is mostly prevalent in eukaryotes (Niu et al., 2013). In general, m^6A is mainly enriched near the stop codon, within the consensus motif DRACH (D = A, G, U; R = A, G; H = A, C, U) (Dominissini et al., 2012). In most species, m^6A is installed by the “writer” complex that is composed of METTL3, METTL14, WTAP, and several unknown components (Liu et al., 2014). Interestingly, m^6A can be erased by the demethylase FTO and ALKBH5 (Zheng et al., 2013), and is recognized by the YTH-domain-containing proteins (Zhu et al., 2014). Growing evidences have indicated that m^6A is involved in post-transcriptional processes including translation, mRNA degradation, alternative splicing, and microRNA maturation, thus affecting gene expression

(Wang et al., 2014, 2015; Alarcon et al., 2015). Recent studies have elucidated the significance of m^6A in the regulation of stem cell fate determination, pre-adipocytes differentiation, DNA damage response, self-renewal of neural stem cells, and T cell homeostasis (Batista et al., 2014; Zhao et al., 2014; Li H.B. et al., 2017; Xiang et al., 2017; Wang et al., 2018).

The importance of m^6A in spermatogenesis have been preliminarily revealed. Depletion of METTL3 leads to inhibition of spermatogonial differentiation and arrest of meiosis initiation, resulting in infertility (Xu et al., 2017). Furthermore, double knockout of METTL3 and METTL14 in advanced germ cells with Stra8-Cre disrupts spermiogenesis (Lin et al., 2017). Knockout of ALKBH5 causes severe apoptosis in spermatogonia and spermatocytes (Zheng et al., 2013). Consistently, YTHDC2 deficient leads to arrest of meiosis at zygotene spermatocytes (Hsu et al., 2017). These evidences provide the proof of concept with respect to the involvement of m^6A in spermatogenesis.

The fat mass and obesity-associated (*Fto*) gene, located at chromosome 16 in humans, encodes FTO protein, which belongs to the α -ketoglutarate-dependent dioxygenase alkB family. Loss of FTO leads to postnatal growth retardation and a significant reduction in adipose tissue and lean body mass (Fischer et al., 2009). It has been sufficiently demonstrated that FTO possesses the activity of m^6A demethylase, which regulates pre-adipocyte differentiation, the leukemic oncogene-mediated cell transformation, and tumorigenesis of glioblastoma stem cells (Li L. et al., 2017; Li Z.J. et al., 2017). Recent studies have revealed the significance of FTO in the regulation of neural development and stress response (Du et al., 2018; Engel et al., 2018). Interestingly, two missense mutations in *Fto* are associated with the reduced semen quality in azoospermic patients (Landfors et al., 2016). However, little is known about the functions of FTO in spermatogenesis.

The aim of the present study was to gain more insights into the role of FTO in spermatogonia division. To this end, we employed the mouse GC-1 spermatogonial cell line as a research model and performed loss-of-function study by CRISPR/Cas9. We found that knockout of *FTO* triggered abnormal chromosome segregation and cell cycle arrest. This phenotype could be partially rescued by wild-type FTO but not mutant FTO. FTO depletion elevated the m^6A level of core mitosis checkpoint complex (MCC) components and G2/M regulators. Therefore, FTO regulates cell cycle and mitosis checkpoint in spermatogonia because of its m^6A demethylase activity.

MATERIALS AND METHODS

Cell Culture and Plasmid Transfection

The mouse spermatogonia cell line (GC-1) were maintained in Dulbecco's Modified Eagle's Medium (GE) with 10% fetal bovine serum (Gibco), 100 U/ml penicillin and 0.1 mg/ml streptomycin (PS) and incubated at 37°C with 5% CO₂. For plasmid transfection, cells were seeded to 6-well plate (2 × 10⁵ cells per plate) and cultured overnight.

Plasmids were transfected to cells using TurboFect™ Transfection Reagent (Thermo Fisher Scientific™) following the instructions. Twenty-four hours post-transfection, cells were subjected to puromycin (2 µg/ml, Sigma) selection for 2 days.

Antibodies

The primary and secondary antibodies were purchased from commercial sources as follows: Mouse anti-FTO, Mouse anti-Mad2, Mouse anti-Cdc20, Mouse anti-Bub1, Mouse anti-Bub1b, Mouse anti-Bub3, Mouse anti Tubulin (Santa Cruz Biotechnology), Rabbit anti m^6A (Synaptic Systems), Rabbit anti-Actin (Sigma-Aldrich). HRP-goat anti rabbit IgG (CWbio) and HRP-goat anti mouse IgG (CWbio).

Vectors Construction

For knocking out FTO in GC-1 cells, the following sgRNAs were designed and synthesized, sg-FTO1U: 5'-ACCGCCGTCCTGCGATGATGAAG-3', sg-FTO1D: 5'-AAACCTTCATCATCGCAGGACGG-3', sg-FTO2U: 5'-ACCGGAACCTCTGCCATGCACAG-3', sg-FTO2D: 5'-AAACCTGTGCATG GCAGAGTTC-3'. The PGL3-U6-PGK plasmid (gifted from Shanghai Tech University) was used as the backbone. Plasmid was ligated with annealed sgRNAs via T4 ligase (Thermo Fisher Scientific). For the FTO rescue experiment, total RNA was extracted from GC-1 cells using RNAiso plus Reagent (Takara Clontech). cDNA was synthesized by the first strand cDNA synthesis kit (Takara Clontech) following the manufacturer's instructions. The following primers were designed and synthesized for the amplification of FTO CDS, FTO-res-F: 5'-GAATCTAGAATGAAGCGCGTCCAGAC-3', FTO-res-R: 5'-GGAGAATTCTGCTGGAAGCAAGATCCTAG-3'. PCR products were purified by the PCR clean-up Kit (Axgen). CD513B plasmid and purified PCR products were digested by restriction enzymes *Eco*RI and *Xba*I (NEB), following by ligation using the T4 ligase.

For the FTO mutant experiment, the following primers were designed and synthesized. FTO-mut-1F: 5'-GAATCTAGAATGAAGCGCGTCCAGAC-3', FTO-mut-1R: 5'-GCGTGAGTGGAACTAAACGCAGGCTGTGA GCCAGC-3', FTO-mut-2F: 5'-GCTGGCTCACAGCCTGCG TTTAGTTCCACTACCG-3', FTO-mut-2R: 5'-GGAGAAT TCTGCTGGAAGCAAGATCCTAG-3'. cDNA of FTO was used as the PCR template. PCR products were purified using Gel Extraction Kit (Omega) following by recombinant using Neotec reagent. Recombined fragments were purified and ligated with CD513-B1 plasmids.

T7E1 Assay

Genomic DNA was extracted using phenol-chloroform followed by ethyl alcohol precipitation. For indels detection, following primers were designed and synthesized, FTO-F: 5'-CCAGTGTCTCGCATCCTCATC-3', FTO-R: 5'-TTACTCATCCTCAGAGCCTCAGA-3'. PCR products were purified using PCR clean up Kit. The purified DNA was annealed following by digested by T7 endonuclease (NEB). After digestion

at 37°C for 30 min, DNA was analyzed by the agarose gel electrophoresis. Image J was used to calculate the cleavage efficiency.

Establishment of the FTO^{-/-} Cell Strain

Plasmids expressing cas9 and sgRNAs were co-transfected to spermatogonia using the TurboFectTM Transfection Reagent as previously described. Twenty-four hours post-transfection, cells were screened using 2 µg/ml puromycin for 2 days. The residual cells were suspended to 300 cells/ml and seeded to the 100-mm-dish. After 7 days of culture, mono clones were observed under the microscope. Monoclonal cells were picked and transferred to the 96-well plate (one clone per well) followed by a 7-day culture. Subsequently, genomic DNA of each cell clone were extracted using the QuickExtractTM DNA Extraction Solution 1.0 (Epicenter) following the manufacturer's instructions. The DNA fragments containing sgRNA target sites were amplified using PCR followed by Sanger sequencing. Cell strains harboring frameshift mutations within *Fto* locus in di-alleles were considered as the *Fto*^{-/-} cell strain.

m⁶A Dot Blot

Total RNA was extracted from cells using Trizol reagent (TAKARA). mRNA was isolated and purified using Poly Attract mRNA Isolation System III with Magnetic Stand (Promega) following the manufacturer's instructions. For m⁶A dot blot, mRNA was hybridized onto the Hybond-N+ membrane (GE Healthcare). After crosslinking at 80°C for 30 min, the membrane was blocked with 5% non-fat milk (Bio-Rad) for 1 h, incubated with rabbit anti-m⁶A antibody (1:1000, Synaptic Systems) at 4°C overnight. Then the membrane was incubated with HRP-conjugated goat anti-rabbit IgG at room temperature for 2 h. After being incubated with Immobilon Western Chemiluminescent HRP Substrate (Millipore), the immunocomplex was photographed using the ECL imaging system (Bio-Rad). Finally, the membrane was stained with 0.02% methylene blue to eliminate the difference in mRNA amount. Relative m⁶A level was quantified via gray intensity analysis using ImageJ.

Western Blot Assay

Cells were lysed with RIPA buffer containing 1% PMSF followed by ultrasonication. Cell lysates were incubated on ice for 30 min, centrifuged at 10,000 g for 10 min. The supernatants were collected and the protein concentration was detected using a BCA detection Kit. Equal amount of proteins was loaded to the polyacrylamide gel. The proteins were separated through SDS-PAGE using the electrophoresis apparatus (Bio-Rad). After electrophoresis, the proteins were transferred to the PVDF membrane (Millipore, IBFP0785C) using a semi-dry transfer instrument (Bio-Rad). The membranes were blocked with 5% non-fat milk for 1 h at room temperature, incubated with primary antibodies at 4°C overnight. Subsequently, the membranes were washed with PBST and incubated with HRP-conjugated secondary antibodies for 1 h at room temperature. After washing, the membranes were incubated with the Immobilon Western

Chemiluminescent HRP Substrate (Millipore, United States) and photographed using the ECL imaging system (Bio-Rad, United States).

Flow Cytometric Analysis

For cell cycle analysis, cells were suspended in 75% cold ethanol and treated with 0.1% Triton X-100 and 100 µg/ml RNase at 37°C for 30 min. Subsequently, the cells were stained with 50 µg/ml PI for 2 h and analyzed by flow cytometry. For cell clustering analysis, cells were fixed in cold 70% ethanol, permeabilized with 0.1% Triton X-100. Then the cells were stained with 4',6-diamidino-2-phenylindole (DAPI, Thermo Fisher Scientific) for 30 min and analyzed by flow cytometry.

Quantitative Real-Time PCR

Cells were lysed with Trizol reagent (TAKARA). Total RNA was isolated by chloroform followed by precipitating with isopropanol. cDNA was synthesized with the PrimeScriptTM RT reagent Kit (TAKARA) following the manufacturer's instructions. Primers designed and synthesized for RT-qPCR were listed in **Supplementary Table S1**. Quantitative PCR was performed using the SYBR Green II PCR Mix (TAKARA) and the IQ5 (Bio-Rad).

Chromosome Spread Assays

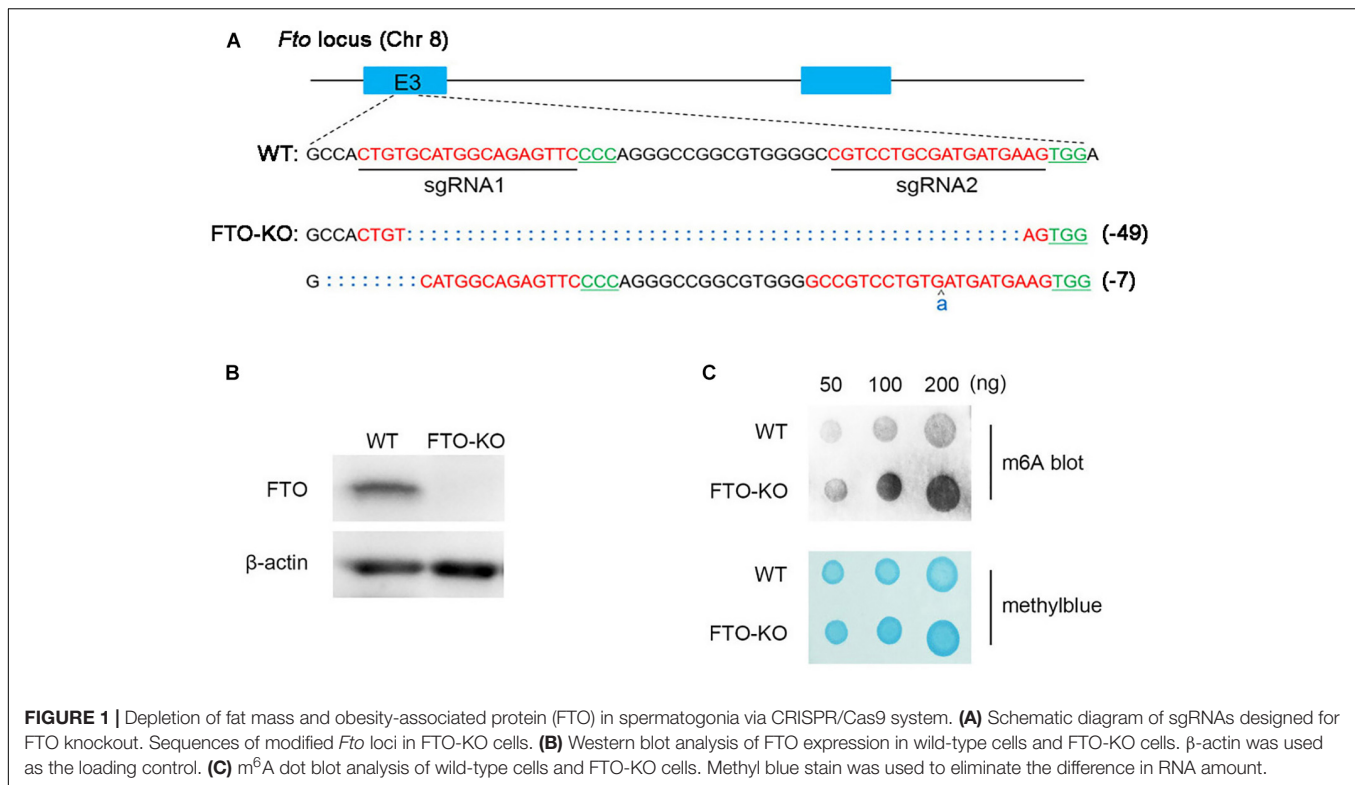
Wild-type and FTO-KO cells were cultured in complete medium to 70% confluence and treated with 50 ng/µL nocodazole for 16 h. Cells were collected and subjected to hypotonical swell in 75 mM KCl at 37°C for 30 min. Subsequently, cells were fixed in Carnoy's fluid (methanol: acetic acid 3:1) at room temperature for 30 min. Cells were dropped onto pre-cooling glass slides and air dried. Slides were stained with Hoechst 33342 (1:500) and photographed under the fluorescence microscope. For each biological repetition, chromosome number of 150 cells were counted and analyzed.

Immunofluorescence

For immunofluorescence analysis, cells were fixed in 4% paraformaldehyde/PBS for 30 min, permeabilized in 0.5% Triton X-100/PBS and blocked with 5% bovine serum albumin (BSA). After washed with PBS for three times, the cells were incubated with rabbit anti-CREST antibody (1:200) and mouse anti β-tubulin (1:200) antibody at 4°C overnight. Then the cells were washed for another three times with PBS and incubated with FITC-conjugated goat anti rabbit and rhodamine red-conjugated goat anti-mouse secondary antibodies (1:2000) at room temperature for 1 h. Cells were washed in PBS for three times and counterstaining with DAPI. Images were photographed under an inverted fluorescence microscope (Olympus, IX71).

m⁶A-IP-qPCR

Total RNA was extracted from cells using the RNAiso plus reagent (TAKARA). mRNA was isolated using the PolyATtract[®]mRNA Isolation Systems (Promega, Z5310) following the manufacturer's instructions. The m⁶A-IP was performed as previously described



(Dominissini et al., 2012). In brief, 3 μ g mRNA was mixed with 12.5 μ L of rabbit anti m^6A antibody (0.5 mg/mL, Synaptic Systems, 202003), 5 \times IP buffer (50 mM Tris-HCl, pH 7.4, 750 mM NaCl, and 0.5% NP-40), RNA inhibitor and DEPC-treated nuclease free water to make 500 μ L of IP mixture. Protein A beads were washed with wash buffer (IP buffer mixed with RNA inhibitor) for three times, and then blocked with 0.5 mg/mL BSA. After blocking, the beads were incubated with IP mixture and rotated at 4°C overnight followed by extensive washing. Bound RNA was eluted using 100 μ L elution buffer (1 \times IP buffer, 6.7 mM m^6A). For m^6A level measurement, 40 ng of IP-RNA and Input-RNA was used for cDNA synthesis. The m^6A^+ mRNA level was finally determined by real-time quantitative PCR.

RNA-Decay Assay

WT cells and FTO-KO cells were treated with 5 μ g/mL actinomycin D for 0, 3, and 6 h, respectively. Cells were harvested and subjected to RNA extraction. Real-time quantitative PCR were used to analyze the mRNA level of target genes in each group.

Statistical Analysis

All data were collected from at least three independent experiments. Data were analyzed using two-tailed Student's *t*-test or one-way ANOVA followed by a Duncan's multiple range test (SPSS 22 for windows). Significance were presented as * p < 0.05, ** p < 0.01, and *** p < 0.001. Error bars represented SEM of the mean.

RESULTS

Depletion of FTO in Spermatogonia Subjected to CRISPR-Cas9 Gene Editing

The endogenous FTO of spermatogonia was abrogated using the CRISPR-Cas9 genome editing technique with *Fto*-specific sgRNAs. Two sgRNAs targeting exon 3 of *Fto* were designed and synthesized (Figure 1A). The CRISPR-Cas9 mutations resulted in 7 and 49 nucleotide deletions in the two alleles of the *Fto* gene, respectively, and thus induced the frameshift mutations at the target sites (Figure 1A). Western blot analysis showed that the FTO protein was completely absent in the KO cells (Figure 1B). We next detected total m^6A level of mRNA extracted from wild type cells and FTO-KO cells. The blot signal strength showed significant increases in mRNA of FTO depletion cells (Figure 1C), indicating that depletion of FTO elevated the m^6A level in spermatogonia.

FTO Depletion Induces Formation of Multinuclear Giant Cells

CCK-8 assay was used to detect cell viability. FTO depletion did not affect cell viability (Figure 2A). The morphology in the FTO-KO cells was markedly different from the Cas9-transfected controls, which was characterized by an increasing population of cells with large cell sizes and spreading areas (Figure 2B). In general, a small number of giant cells can be found in the wild type cells, which was considered as the binucleated spermatocytes. Un-expectedly, in the FTO-KO cells the rate and size of giant cells dramatically increased

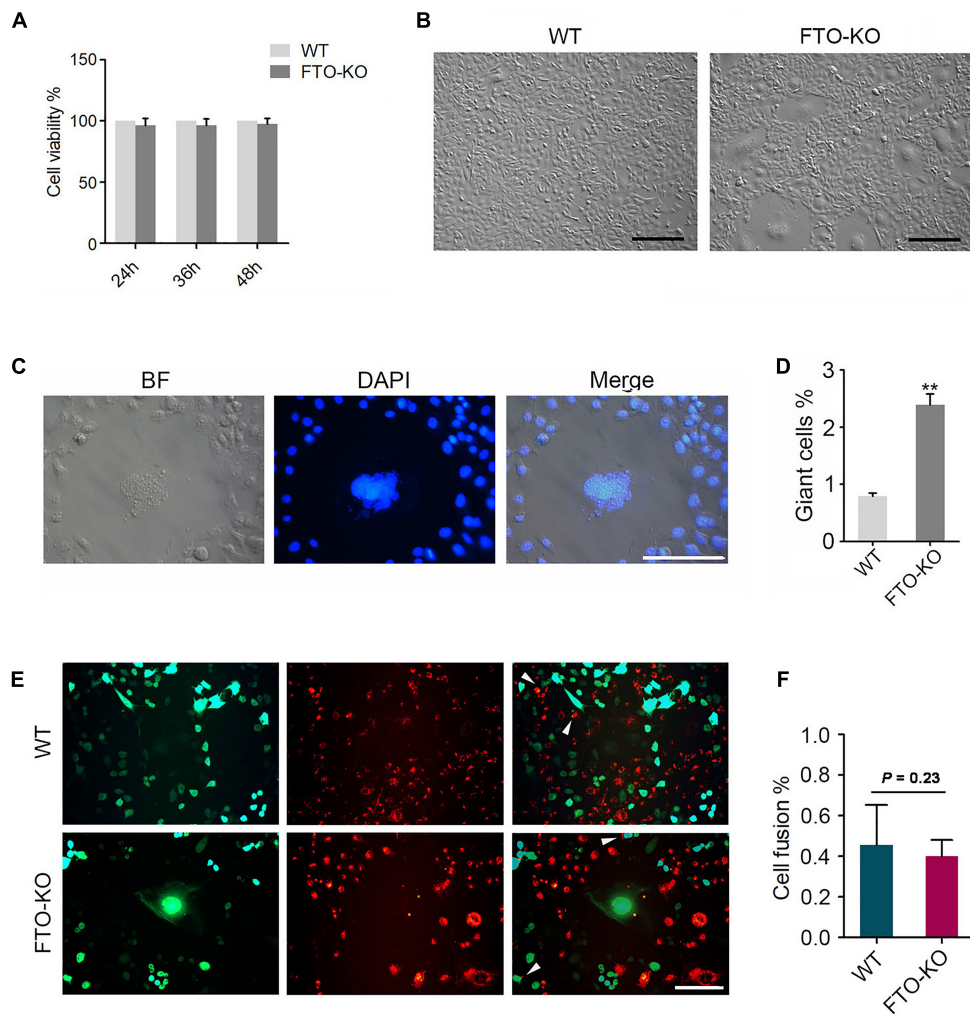


FIGURE 2 | Phenotypes of FTO depletion in spermatogonia. **(A)** Effect of FTO depletion on cell viability. **(B)** Image of morphology of wild-type cells and FTO-KO cells. Bar = 100 μ M. **(C)** Morphology of nuclei of giant cells and normal cells. **(D)** Percentage of giant cells were shown in column diagram. Data were represented by the mean \pm SEM, $n = 3$, ** $p < 0.01$. **(E)** Wild-type cells and FTO-KO cells were stained with red and green dye, respectively, and were co-cultured for 48 h. Representative image of cell fusion status. White arrows indicate the fused cells. Scale bar = 20 μ m. **(F)** Percentage of cell fusion were shown in column diagram. Data were represented by the mean \pm SEM, $n = 3$.

(Figure 2D). To further investigate the giant cells in detail, we stained cell nuclei using DAPI. As shown in Figure 2C, giant cells were aneuploidies that contained large and irregular nuclei. Flow cytometry analysis showed that the ratio of aneuploidies was significantly increased in FTO-KO cells compared with WT cells (Supplementary Figure S1). These results suggested that FTO deletion caused aneuploidy formation in spermatogonia.

FTO Depletion Suppresses Chromosome Segregation

Either cell fusion or abnormal chromosome segregation probably leads to aneuploidy formation. To determine whether the increase in aneuploidy proportion was caused by cell fusion, we stained the cells using the double fluorescent tracer assay.

The cells stained with red dye and the one stained with green dye were mixed and cultured for 48 h, the fused cells showed bifluorescence (Figure 2E). However, the proportion of fused cells in FTO-KO cells was not different with that in wild-type cells (Figure 2F), indicating that the giant cells were not induced by cell fusion.

To investigate whether the aneuploidy was induced by abnormal chromosome segregation, we counted the chromosomes in wild-type cells and FTO-KO cells through the chromosome spreading assay. Interestingly, chromosome number in FTO-KO cells showed a significant increase, compared with wild-type cells (Figures 3A,B). The mitotic checkpoint complex (MCC) is the effector of the spindle assembly checkpoint (SAC) that prevents cells from undergoing cytokinesis when the spindle is assembled improperly with chromosome at metaphase (Lara-Gonzalez et al., 2012). Previous

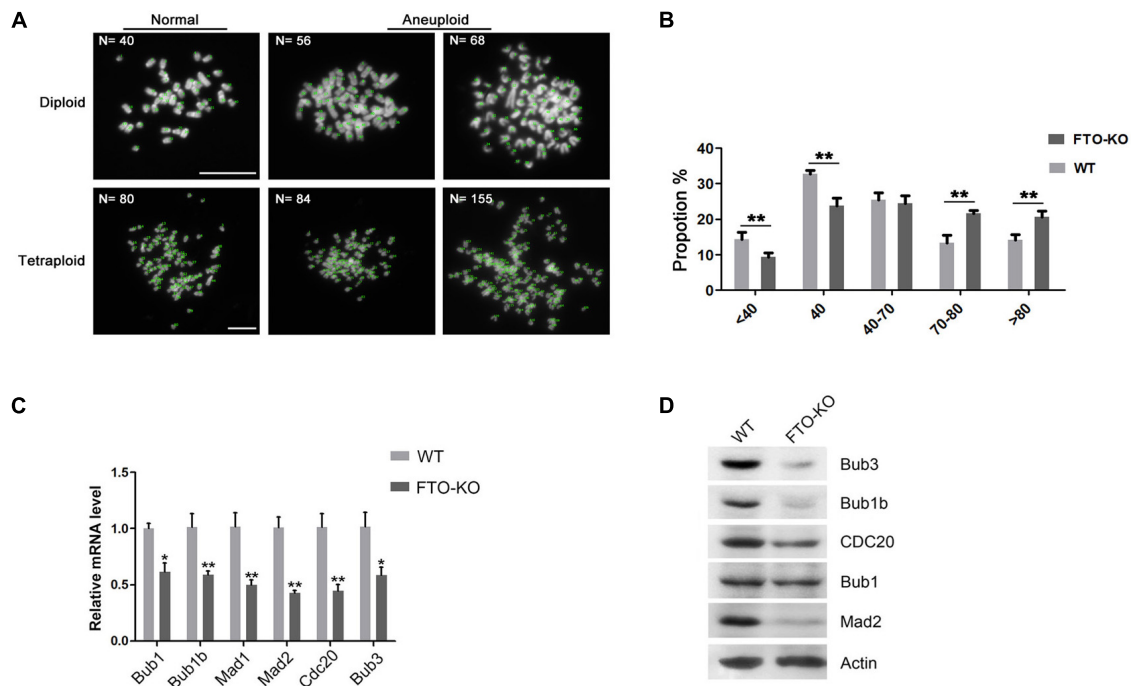


FIGURE 3 | Effect of FTO depletion on chromosome segregation. **(A)** Representative images of normal diploidy and aneuploidy chromosome spread in FTO-KO cells. **(B)** Chromosome numbers from individual metaphase spreads of FTO-KO and wild-type cells. **(C)** Relative mRNA expression of MCC components. Data were represented by the mean \pm SEM, $n = 3$, * $p < 0.05$, ** $p < 0.01$. **(D)** Western blot analysis of MCC components expression in wild-type cells and FTO-KO cells. β -actin was used as the loading control.

studies have reported that dys-regulation of MCC components resulted in chromosomal instability and aneuploidy (Kapanidou et al., 2015). Here, we hypothesized that FTO depletion induced the formation of aneuploidy due to aberrant expression of MCC. To verify it, we detected the expression of the core MCC components Mad1, Mad2, Bub1, Bub1b, Bub3, and Cdc20. Interestingly, the expression of all detected MCC components significantly decreased in FTO-KO cells both in mRNA and protein levels (Figures 3C,D). These results suggested that FTO deletion suppressed chromosome segregation and induced aneuploidy formation through up-regulation of MCC expression.

FTO Depletion Arrests G2/M Transition

Previous studies have shown that m⁶A methylation is correlated with cell cycle progress during oocyte meiotic maturation (Qi et al., 2016). Therefore, we presumed that FTO regulates cell cycle in spermatogonia. To this end, we analyzed the cell cycle by flow cytometry. Interestingly, we found that the proportion of G2 stage cells significantly increased in FTO-KO cells, compared with wild-type cells (Figures 4A,B). We next detected the expression of core regulatory proteins involved in G2/M transition. As shown in Figures 4C,D, the expression of CDK1 and CCNB2 was significantly down-regulated in FTO-KO cells, indicating that FTO modulated G2/M transition through regulating the expression of Cdk1/Ccnb2 complex.

FTO Regulates Cell Cycle and Aneuploidy Formation Through the m⁶A Demethylase Activity

Previous studies have reported that mutation of the critical amino acid residue 313R to A (R313A) in the catalytic center of FTO protein can completely ablate its m⁶A demethylase activity (Zhao et al., 2014). Hence, to investigate whether the FTO knockout phenotype in spermatogonia is due to its m⁶A demethylase activity, we constructed two lentivirus vectors that expressed wild-type FTO (named FTO-wt) and R313A mutant FTO (named FTO-mut), respectively (Figure 5A). We next established three cell lines by transfection of the FTO-wt, the FTO-mut and the control (GFP) lentivirus to the FTO-KO cells, respectively. Western blot analysis showed that the FTO expression was rescued in FTO-wt and FTO-mut cells, but not control cells (Figure 5B). We next detected the proportion of aneuploidy and G2 stage cells in the three cell lines. Interestingly, the rate of aneuploidy and G2 stage cells in FTO-wt group was significantly less than those in the FTO-mut and control cells, indicating that the FTO depletion phenotype could be partially rescued by wild-type FTO but not mutant FTO (Figures 5C–F). These results suggested that FTO regulated cell cycle in spermatogonia mainly through its m⁶A demethylase activity.

To verify whether FTO deletion leads to increase of m⁶A level in the transcripts of target genes, we performed m⁶A-IP-qPCR. As shown in Figure 6A, in the m⁶A-IP transcripts, the abundance of Bub1b, Mad1, Mad2, Cdk1 and Ccnb2 was

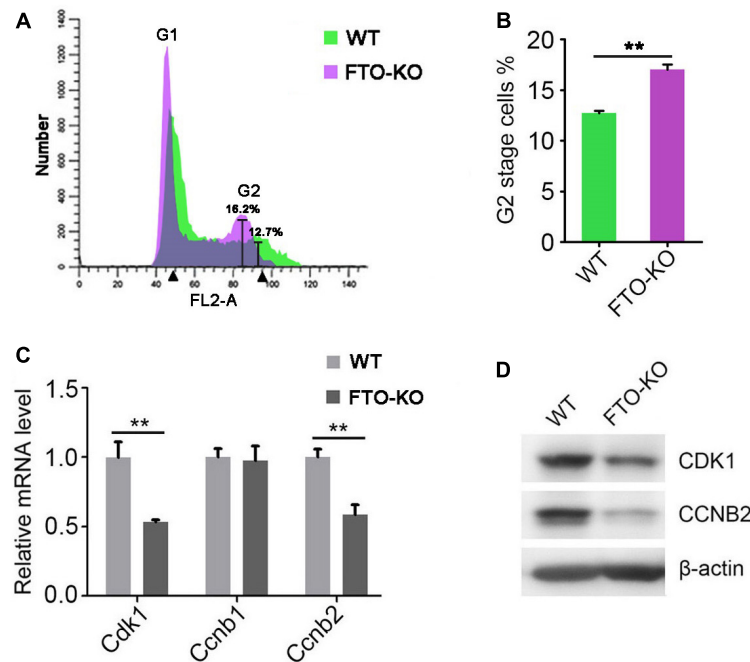


FIGURE 4 | Effect of FTO depletion on cell cycle. **(A)** Cell cycle of FTO-KO and wild-type cells were analyzed by flow cytometry. **(B)** Percentage of G2 stage cells were shown in column diagram. Data were represented by the mean \pm SEM, $n = 3$, $**p < 0.01$. **(C)** Relative mRNA level of genes involved in G2/M transition. Data were represented by the mean \pm SEM, $n = 3$, $**p < 0.01$. **(D)** Western blot analysis of proteins involved in G2/M transition. β -actin was used as the loading control.

significantly up-regulated in FTO depletion group, while Bub1, Cdc20, and Bub3 were undetectable under the sensitivity of q-PCR, suggesting that m^6A level in the transcripts of Bub1b, Mad1, Mad2, Cdk1, and Ccnb2 is elevated due to FTO knockout.

To further demonstrate whether the increased m^6A level accelerates the degradation of target mRNAs, we performed an RNA decay assay. Cells were treated with 5 μ g/mL actinomycin D for 0, 3, and 6 h and harvested for RNA extraction. The remained mRNA level was normalized by real-time quantitative PCR. As shown in **Figure 6B**, mRNA stability of Mad1, Mad2, Bub1b, CDK1, and Ccnb2 were significantly decreased after FTO depletion. These results suggested that FTO regulated the expression of target transcripts through the regulation of RNA stability.

Together, these data indicate that FTO directly regulates the expression of the core MCC components and G2/M regulators through the m^6A /RNA decay pathway, thus regulating cell cycle and mitosis checkpoint in spermatogonia.

DISCUSSION

Spermatogenesis is a highly dynamic developmental process involving intricate regulation of gene expression. The significance of m^6A in spermatogenesis has increasingly been unraveled. FTO, the first discovered m^6A demethylase, regulates RNA splicing, stability or translation, thereby making a difference in cell fate determination (Li L. et al., 2017). However, FTO function in spermatogonia remain unclear. Here, we established a Fto-null

mouse spermatogonial cell line using CRISPR/Cas9 system. We found that FTO deletion led to aneuploidy formation and G2/M arrest. We further demonstrated that FTO demethylated five transcripts of core MCC components and G2/M regulators. The findings suggest that FTO regulates chromosome segregation and cell cycle progression via m^6A demethylase activity.

Accurate segregation of duplicated chromosomes is indispensable for the cell division. The accurate chromosome segregation relies on precise temporal regulation of sequential processes including the orientation of bipolar spindle, the attachment of kinetochore and microtubules and the separation of daughter cells during cytokinesis (Thompson et al., 2010). Error occurred at any step may lead to chromosome mis-segregation and aneuploidy formation (Meraldi, 2016). The mitotic checkpoint is a safeguard mechanism against the aneuploidy formation (London and Biggins, 2014). When chromosomes fails to assemble with spindle, the checkpoint activates to inhibit the downstream anaphase promoting complex (APC/C), resulting in prevention of cells from entry the next cell cycle (Thompson et al., 2010). The importance of MCC in the regulation of chromosome instability have been largely reported (Lara-Gonzalez et al., 2012). Chromosome segregation errors in mitosis are the most common cause for aneuploidies formation *in vitro*, as well as in clinical cancer samples (van Jaarsveld and Kops, 2016). Hence, the MCC components have been selected as promising targets for the therapy of cancers (Tanaka and Hirota, 2016). In the present study, we found that FTO regulates the expression of core MCC components, thus regulating chromosome segregation. The roles of FTO and

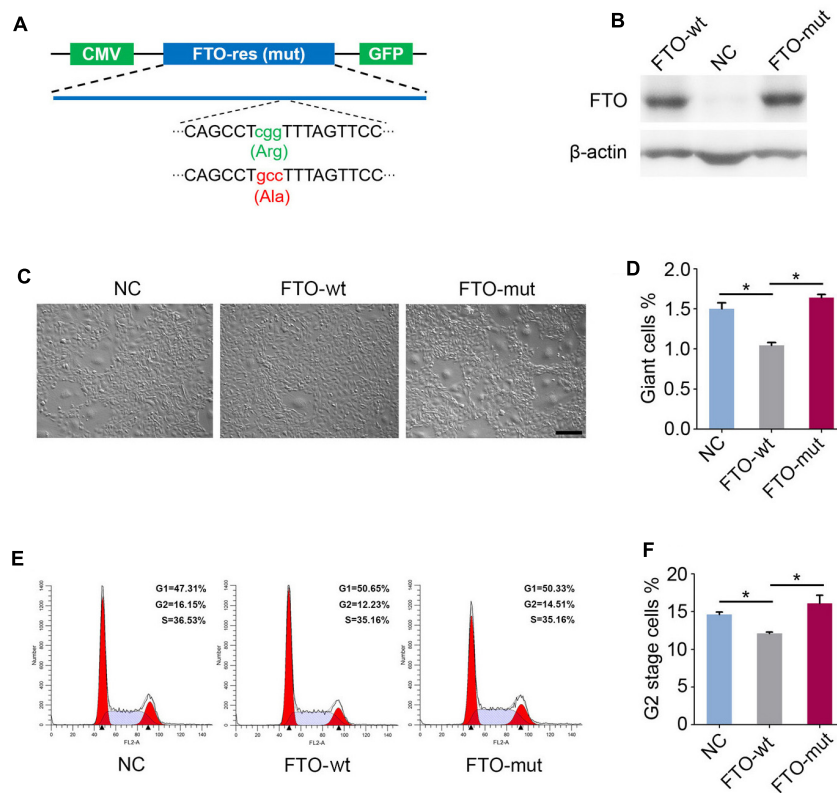


FIGURE 5 | Fat mass and obesity-associated protein regulates giant cell formation and G2/M transition dependent on the m^6A demethylase activity. **(A)** Schematic diagram of FTO-wt and FTO-mut plasmids. Green and red letters represent the mutation of 313R to A. **(B)** Western blot analysis of FTO expression in the FTO-wt cells and the FTO-mut cells. β -actin was used as the loading control. **(C)** Morphology of FTO-wt cells and FTO-mut cells. Bar = 100 μ m. **(D)** Percentage of giant cells were shown in column diagram. Data were represented by the mean \pm SEM, $n = 3$, $**p < 0.01$. **(E)** Cell cycle of FTO-wt and FTO-mut cells were analyzed by flow cytometry. **(F)** Percentage of G2 stage cells were shown in column diagram. Data were represented by the mean \pm SEM, $n = 3$, $*p < 0.05$.

m^6A in the regulation of chromosome stability have not been reported yet. Therefore, our results suggested that FTO may play important roles in the progression of seminoma for the first time. It will be interesting to deeply investigate the functions of FTO in seminoma carcinogenesis.

The Cdk1/Ccnb complex is the main composition of the maturation-promoting factor (MPF) that triggers the G2/M transition (Adhikari and Liu, 2014). The role of Cdk1/Ccnb in the regulation of metaphase arrest during oogenesis have been well documented (Turner, 2015). In contrast, studies on the function of MPF in spermatogenesis are limited. Clement et al. (2015) reported that decreased expression of Cdk1 caused late meiotic arrest and infertility in mice. Recent studies showed that FTO regulates the expression of CDK2 and CCNB2, thus affecting cell cycle progression during adipogenesis (Wu et al., 2018). In the present study, we showed that FTO depletion led to decreased expression of the Cdk1 and Ccnb2, resulting in G2/M arrest in spermatogonia.

A few studies have reported the significance of m^6A relative proteins in spermatogenesis (Zheng et al., 2013; Hsu et al., 2017; Lin et al., 2017). The underlying mechanisms of how m^6A regulates the spermatogenesis remain obscure. Previous studies mainly analyzed the global m^6A methylome, combined the

variation of m^6A peaks with the differentiation in global expression or splicing of transcriptome, thus determined the function of m^6A in stability or splicing of target transcripts (Xu et al., 2017). In the present study, we detected the m^6A level of target genes via the m^6A -IP-qPCR assay, which can precisely reveal the m^6A variation on target transcripts. We found that three transcripts of the core MCC components (Bub1b, Mad1, and Mad2) and two transcripts of the G2/M regulatory proteins (Cdk1 and Ccnb2), were directly targeted by FTO. We also demonstrated that the increased m^6A level retarded the stability of target transcripts. Recent reports have shown that FTO simultaneously demethylates m^6A and m^6A_m in mammalian cells. To further elucidate the mechanism by which FTO knockout leads to the phenotypes, it will be interesting to detect m^6A_m through the miCLIP-seq (Mauer et al., 2017; Wei et al., 2018). Additionally, our results showed that m^6A modification did not occur in the transcripts of Bub1, Bub3, and Cdc20, indicating that FTO regulates the three genes through other pathways. It is no doubt that changes on gene expression can be global after FTO depletion. Though the target genes we focused on are directly associated with the phenotypes, contributions of other differentially expressed genes through other pathways should not be ignored. The mechanisms of m^6A

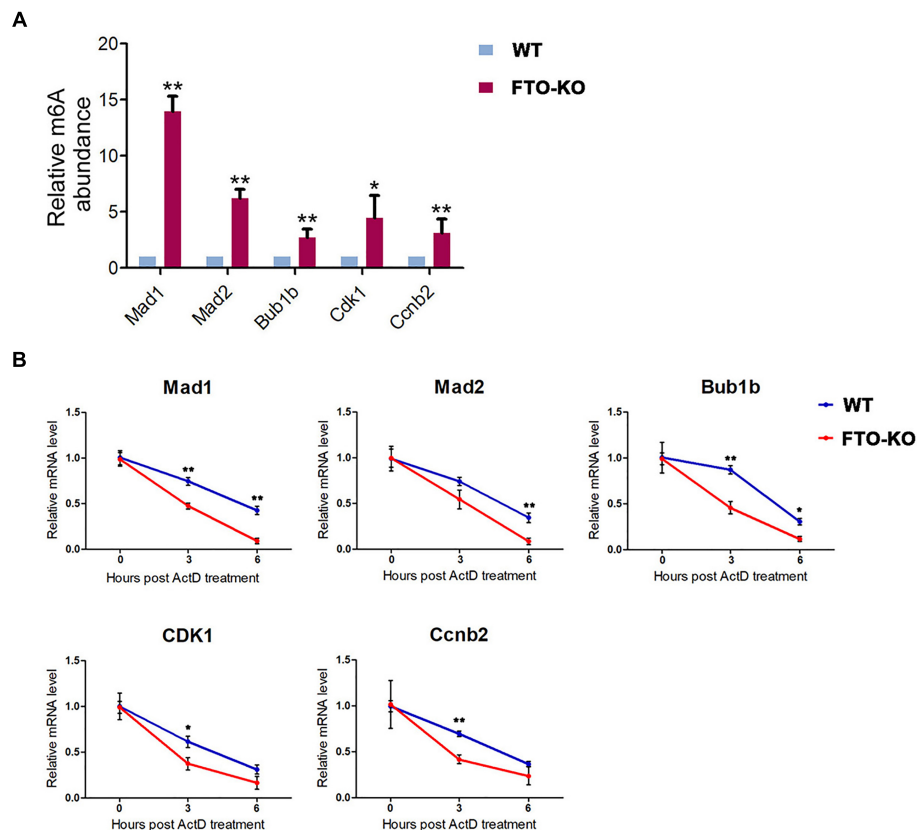


FIGURE 6 | Fat mass and obesity-associated protein targets the m⁶A modification on the key components of MCC and G2/M regulators. **(A)** mRNA isolated from wild-type cells and FTO-KO cells were subjected to m⁶A-IP. The m⁶A positive mRNA were detected via quantitative PCR. Y axes represent the relative mRNA level in IP-mRNA compared to input. Data were represented by the mean \pm SEM, $n = 3$, * $p < 0.05$, ** $p < 0.01$. **(B)** Cells were treated with 5 μ g/mL actinomycin and harvested for RNA extraction. Target mRNA level were quantitated by q-PCR. X axes represent hours post treatment. Y axes represent the remained mRNA level. Data were represented by the mean \pm SEM, $n = 3$, * $p < 0.05$, ** $p < 0.01$.

on the regulation of gene expression are comprehensive. Changes in m⁶A level can make differences in mRNA decay, splicing or translation, which relies on the recognition of different readers. Here we mainly focused on the effects of FTO on the decay of target transcripts. Hence, to gain deeper insights into the regulatory role of FTO to the phenotype, RNA-seq combined with splicing analysis and translation efficiency assay will be necessary.

Male infertility has been becoming a worldwide problem in recent years (Mascarenhas et al., 2012). To understand the underlying mechanisms of spermatogenesis is important for the precise therapy of male infertility. As spermatogonia are the precursor of male germ cells, to elucidate the regulation of spermatogonia homeostasis is important for understanding male infertility. The present study first revealed the role of RNA demethylase FTO in the regulation of chromosome instability and cell cycle progression in spermatogonia, thus giving novel insights into the role of RNA methylation in spermatogenesis and potentially, in seminoma progression. Our studies are limited to the functions of FTO in immortalized cell line. It will be important to generate conditional knockout mice to gain better understandings of the roles of FTO plays in spermatogenesis.

CONCLUSION

In conclusion, knockout of FTO triggered aberrant chromosome segregation and cell cycle arrest, which could be partially rescued by wild-type FTO but not mutant FTO. FTO depletion elevated the m⁶A level of core MCC components and G2/M regulators. Therefore, FTO regulates cell cycle and mitosis checkpoint in spermatogonia through the m⁶A/mRNA degradation pathway. Our findings give novel insights into the role of RNA methylation in spermatogenesis.

DATA AVAILABILITY STATEMENT

All datasets (generated/analyzed) for this study are included in the manuscript.

AUTHOR CONTRIBUTIONS

TH and WZ conceived and designed the experiments. TH, QG, TE, and JG performed the experiments. TH analyzed the data. TH, WZ, and YZ wrote the manuscript.

FUNDING

This study was supported in part by the National Natural Science Foundation of China (Grant No. 31572401) to WZ.

ACKNOWLEDGMENTS

We thank Yungui Yang and Ying Yang from the Beijing Institute of Genome Research for the advices on m⁶A-IP. Thanks to

Yinghua Lv for the technical support. We also thank Jiaying Li from the Northwest A&F University for the analysis of flow cytometry. Thanks to all the members of Zeng laboratory for the helpful discussion.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00732/full#supplementary-material>

REFERENCES

- Adhikari, D., and Liu, K. (2014). The regulation of maturation promoting factor during prophase I arrest and meiotic entry in mammalian oocytes. *Mol. Cell. Endocrinol.* 382, 480–487. doi: 10.1016/j.mce.2013.07.027
- Alarcon, C. R., Lee, H., Goodarzi, H., Halberg, N., and Tavazoie, S. F. (2015). N6-methyladenosine marks primary microRNAs for processing. *Nature* 519, 482–485. doi: 10.1038/nature14281
- Batista, P. J., Molin, B., Wang, J., Qu, K., Zhang, J., Li, L., et al. (2014). m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* 15, 707–719. doi: 10.1016/j.stem.2014.09.019
- Clement, T. M., Inselman, A. L., Goulding, E. H., Willis, W. D., and Eddy, E. M. (2015). Disrupting cyclin dependent kinase 1 in spermatocytes causes late meiotic arrest and infertility in mice. *Biol. Reprod.* 93:137. doi: 10.1095/biolreprod.115.134940
- Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485, 201–206. doi: 10.1038/nature11112
- Du, K., Zhang, L., Lee, T., and Sun, T. (2018). m(6)A RNA methylation controls neural development and is involved in human diseases. *Mol. Neurobiol.* doi: 10.1007/s12035-018-1138-1 [Epub ahead of print].
- Engel, M., Eggert, C., Kaplick, P. M., Eder, M., Roh, S., Tietze, L., et al. (2018). The role of m(6)A/m-RNA methylation in stress response regulation. *Neuron* 99:e389. doi: 10.1016/j.neuron.2018.07.009
- Fischer, J., Koch, L., Emmerling, C., Vierkotten, J., Peters, T., Bruning, J. C., et al. (2009). Inactivation of the Fto gene protects from obesity. *Nature* 458, 894–898. doi: 10.1038/nature07848
- Fok, K. L., Chen, H., Ruan, Y. C., and Chan, H. C. (2014). Novel regulators of spermatogenesis. *Semin. Cell Dev. Biol.* 29, 31–42. doi: 10.1016/j.semcdb.2014.02.008
- Hamra, F. K., Schultz, N., Chapman, K. M., Grellhes, D. M., Cronkhite, J. T., Hammer, R. E., et al. (2004). Defining the spermatogonial stem cell. *Dev. Biol.* 269, 393–410. doi: 10.1016/j.ydbio.2004.01.027
- Hsu, P. J., Zhu, Y. F., Ma, H. H., Guo, Y. H., Shi, X. D., Liu, Y. Y., et al. (2017). Ythdc2 is an N6-methyladenosine binding protein that regulates mammalian spermatogenesis. *Cell Res.* 27, 1115–1127. doi: 10.1038/cr.2017.99
- Kanatsu-Shinohara, M., and Shinohara, T. (2013). Spermatogonial stem cell self-renewal and development. *Annu. Rev. Cell. Dev. Biol.* 29, 163–187. doi: 10.1146/annurev-cellbio-101512-122353
- Kapanidou, M., Lee, S., and Bolanos-Garcia, V. M. (2015). BubR1 kinase: protection against aneuploidy and premature aging. *Trends Mol. Med.* 21, 364–372. doi: 10.1016/j.molmed.2015.04.003
- Landfors, M., Nakken, S., Fusser, M., Dahl, J. A., Klungland, A., and Fedorcsak, P. (2016). Sequencing of FTO and ALKBH5 in men undergoing infertility work-up identifies an infertility-associated variant and two missense mutations. *Fertil. Steril.* 105:e1175. doi: 10.1016/j.fertnstert.2016.01.002
- Lara-Gonzalez, P., Westhorpe, F. G., and Taylor, S. S. (2012). The spindle assembly checkpoint. *Curr. Biol.* 22, R966–R980. doi: 10.1016/j.cub.2012.10.006
- Li, H. B., Tong, J., Zhu, S., Batista, P. J., Duffy, E. E., Zhao, J., et al. (2017). m(6)A mRNA methylation controls T cell homeostasis by targeting the IL-7/STAT5/SOCS pathways. *Nature* 548, 338–342. doi: 10.1038/nature23450
- Li, L., Zang, L., Zhang, F., Chen, J., Shen, H., Shu, L., et al. (2017). Fat mass and obesity-associated (FTO) protein regulates adult neurogenesis. *Hum. Mol. Genet.* 26, 2398–2411. doi: 10.1093/hmg/ddx128
- Li, Z. J., Weng, H. Y., Su, R., Weng, X. C., Zuo, Z. X., Li, C. Y., et al. (2017). FTO plays an oncogenic role in acute myeloid leukemia as a N6-Methyladenosine RNA demethylase. *Cancer Cell* 31, 127–141. doi: 10.1016/j.ccell.2016.11.017
- Lin, Z., Hsu, P. J., Xing, X., Fang, J., Lu, Z., Zou, Q., et al. (2017). Mettl3-/Mettl14-mediated mRNA N(6)-methyladenosine modulates murine spermatogenesis. *Cell Res.* 27, 1216–1230. doi: 10.1038/cr.2017.117
- Liu, J. Z., Yue, Y. N., Han, D. L., Wang, X., Fu, Y., Zhang, L., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* 10, 93–95. doi: 10.1038/nchembio.1432
- London, N., and Biggins, S. (2014). Signalling dynamics in the spindle checkpoint response. *Nat. Rev. Mol. Cell Biol.* 15, 735–747. doi: 10.1038/nrm3888
- Mascarenhas, M. N., Flaxman, S. R., Boerma, T., Vanderpoel, S., and Stevens, G. A. (2012). National, regional, and global trends in infertility prevalence since 1990: a systematic analysis of 277 health surveys. *PLoS Med.* 9:e1001356. doi: 10.1371/journal.pmed.1001356
- Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A. V., Patil, D. P., et al. (2017). Reversible methylation of m(6)Am in the 5' cap controls mRNA stability. *Nature* 541, 371–375. doi: 10.1038/nature21022
- Meraldi, P. (2016). Centrosomes in spindle organization and chromosome segregation: a mechanistic view. *Chromosom. Res.* 24, 19–34. doi: 10.1007/s10577-015-9508-2
- Niu, Y., Zhao, X., Wu, Y. S., Li, M. M., Wang, X. J., and Yang, Y. G. (2013). N6-methyl-adenosine (m6A) in RNA: an old modification with a novel epigenetic function. *Genomics Proteomics Bioinformatics* 11, 8–17. doi: 10.1016/j.gpb.2012.12.002
- Qi, S. T., Ma, J. Y., Wang, Z. B., Guo, L., Hou, Y., and Sun, Q. Y. (2016). N6-Methyladenosine sequencing highlights the involvement of mRNA methylation in oocyte meiotic maturation and embryo development by regulating translation in *Xenopus laevis*. *J. Biol. Chem.* 291, 23020–23026. doi: 10.1074/jbc.M116.748889
- Tanaka, K., and Hirota, T. (2016). Chromosomal instability: a common feature and a therapeutic target of cancer. *Biochim. Biophys. Acta Rev. Cancer* 1866, 64–75. doi: 10.1016/j.bbcan.2016.06.002
- Thompson, S. L., Bakhoun, S. F., and Compton, D. A. (2010). Mechanisms of chromosomal instability. *Curr. Biol.* 20, R285–R295. doi: 10.1016/j.cub.2010.01.034
- Turner, J. M. (2015). Meiotic silencing in mammals. *Annu. Rev. Genet.* 49, 395–412. doi: 10.1146/annurev-genet-112414-055145
- van Jaarsveld, R. H., and Kops, G. J. P. L. (2016). Difference Makers: chromosomal instability versus Aneuploidy in Cancer. *Trends Cancer* 2, 561–571. doi: 10.1016/j.trecan.2016.09.003
- Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., et al. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 505, 117–120. doi: 10.1038/nature12730
- Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z., Han, D., Ma, H., et al. (2015). N(6)-methyladenosine modulates messenger RNA translation efficiency. *Cell* 161, 1388–1399. doi: 10.1016/j.cell.2015.05.014
- Wang, Y., Li, Y., Yue, M. H., Wang, J., Kumar, S., Wechsler-Reya, R. J., et al. (2018). N6-methyladenosine RNA modification regulates embryonic neural stem cell

- self-renewal through histone modifications. *Nat. Neurosci.* 21, 1139–1139. doi: 10.1038/s41593-018-0169-2
- Wei, J., Liu, F., Lu, Z., Fei, Q., Ai, Y., He, P. C., et al. (2018). Differential m(6)A, m(6)Am, and m(1)A demethylation mediated by FTO in the cell nucleus and cytoplasm. *Mol. Cell.* 71:e975. doi: 10.1016/j.molcel.2018.08.011
- Wu, R., Liu, Y., Yao, Y., Zhao, Y., Bi, Z., Jiang, Q., et al. (2018). FTO regulates adipogenesis by controlling cell cycle progression via m(6)A-YTHDF2 dependent mechanism. *Biochim. Biophys. Acta. Mol. Cell. Biol. Lipids* 1863, 1323–1330. doi: 10.1016/j.bbalip.2018.08.008
- Xiang, Y., Laurent, B., Hsu, C. H., Nachtergaele, S., Lu, Z., Sheng, W., et al. (2017). RNA m(6)A methylation regulates the ultraviolet-induced DNA damage response. *Nature* 543, 573–576. doi: 10.1038/nature21671
- Xu, K., Yang, Y., Feng, G. H., Sun, B. F., Chen, J. Q., Li, Y. F., et al. (2017). Mettl3-mediated m(6)A regulates spermatogonial differentiation and meiosis initiation. *Cell Res.* 27, 1100–1114. doi: 10.1038/cr.2017.100
- Zhao, X., Yang, Y., Sun, B. F., Shi, Y., Yang, X., Xiao, W., et al. (2014). FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. *Cell Res.* 24, 1403–1419. doi: 10.1038/cr.2014.151
- Zheng, G. Q., Dahl, J. A., Niu, Y. M., Fedorcsak, P., Huang, C. M., Li, C. J., et al. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* 49, 18–29. doi: 10.1016/j.molcel.2012.10.015
- Zhu, T., Roundtree, I. A., Wang, P., Wang, X., Wang, L., Sun, C., et al. (2014). Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine. *Cell Res.* 24, 1493–1496. doi: 10.1038/cr.2014.152

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Huang, Gao, Feng, Zheng, Guo and Zeng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Precise and Rapid Validation of Candidate Gene by Allele Specific Knockout With CRISPR/Cas9 in Wild Mice

Tianzhu Chao^{1,2†}, Zhuangzhuang Liu^{1,2†}, Yu Zhang^{1,2}, Lichen Zhang^{1,2}, Rong Huang^{1,2}, Le He^{1,2}, Yanrong Gu^{1,2}, Zhijun Chen^{1,2}, Qianqian Zheng^{1,2}, Lijin Shi^{1,3}, Wenping Zheng^{1,2}, Xinhui Qi^{1,2}, Eryan Kong¹, Zhongjian Zhang¹, Toby Lawrence⁴, Yinming Liang^{1,2,3*} and Liaoxun Lu^{1,2,3*}

OPEN ACCESS

Edited by:

Zhiying Zhang,
Northwest A&F University, China

Reviewed by:

Shun Li,
Fudan University, China
Ashwin S. Shetty,
Harvard University, United States
Feng Gu,
Wenzhou Medical University, China

*Correspondence:

Yinming Liang
yinming.liang@foxmail.com
Liaoxun Lu
luliaofox@foxmail.com

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Genomic Assay Technology,
a section of the journal
Frontiers in Genetics

Received: 26 August 2018

Accepted: 04 February 2019

Published: 19 February 2019

Citation:

Chao T, Liu Z, Zhang Y, Zhang L,
Huang R, He L, Gu Y, Chen Z,
Zheng Q, Shi L, Zheng W, Qi X,
Kong E, Zhang Z, Lawrence T, Liang Y
and Lu L (2019) Precise and Rapid
Validation of Candidate Gene by Allele
Specific Knockout With CRISPR/Cas9
in Wild Mice. *Front. Genet.* 10:124.
doi: 10.3389/fgene.2019.00124

¹ Institute of Psychiatry and Neuroscience, Xinxiang Medical University, Xinxiang, China, ² Henan Key Laboratory of Immunology and Targeted Therapy, School of Laboratory Medicine, Xinxiang Medical University, Xinxiang, China, ³ Laboratory of Genetic Regulators in the Immune System, Henan Collaborative Innovation Center of Molecular Diagnosis and Laboratory Medicine, School of Laboratory Medicine, Xinxiang Medical University, Xinxiang, China, ⁴ Centre for Inflammation Biology and Cancer Immunology, King's College London, London, United Kingdom

It is a tempting goal to identify causative genes underlying phenotypic differences among inbred strains of mice, which is a huge reservoir of genetic resources to understand mammalian pathophysiology. In particular, the wild-derived mouse strains harbor enormous genetic variations that have been acquired during evolutionary divergence over 100s of 1000s of years. However, validating the genetic variation in non-classical strains was extremely difficult, until the advent of CRISPR/Cas9 genome editing tools. In this study, we first describe a T cell phenotype in both wild-derived PWD/PhJ parental mice and F1 hybrids, from a cross to C57BL/6 (B6) mice, and we isolate a genetic locus on Chr2, using linkage mapping and chromosome substitution mice. Importantly, we validate the identification of the functional gene controlling this T cell phenotype, *Cd44*, by allele specific knockout of the PWD copy, leaving the B6 copy completely intact. Our experiments using F1 mice with a dominant phenotype, allowed rapid validation of candidate genes by designing sgRNA PAM sequences that only target the DNA of the PWD genome. We obtained 10 animals derived from B6 eggs fertilized with PWD sperm cells which were subjected to microinjection of CRISPR/Cas9 gene targeting machinery. In the newborns of F1 hybrids, 80% ($n = 10$) had allele specific knockout of the candidate gene *Cd44* of PWD origin, and no mice showed mistargeting of the B6 copy. In the resultant allele-specific knockout F1 mice, we observe full recovery of T cell phenotype. Therefore, our study provided a precise and rapid approach to functionally validate genes that could facilitate gene discovery in classic mouse genetics. More importantly, as we succeeded in genetic manipulation of mice, allele specific knockout could provide the possibility to inactivate disease alleles while keeping the normal allele of the gene intact in human cells.

Keywords: CD44, allele specific knockout, CRISPR/Cas9, functional genomics, wild mice

INTRODUCTION

Wild mice refer to both inbred lines and individual animals from natural house mouse populations, both of which harbor enormous genetic variations. Such models are particularly useful for study of genetic and environmental factors contributing to host immune response to pathogens (Rosshart et al., 2017). C57BL/6 and PWD/Ph strains are wild-derived mouse strains representing two major subspecies of house mouse, namely *M. m. domesticus* and *M. m. musculus* (Guenet and Bonhomme, 2003). The *M. m. domesticus* derived B6 mice and *M. m. musculus* derived PWK mice have highly diverged genomes with over 17 million single nucleotide polymorphisms (SNPs) which far outnumbers the genetic variations between classical laboratory strains (Keane et al., 2011). Over 90% of the genomic composition of classical laboratory mouse strains are mainly derived from subspecies *M. m. domesticus*, and remnants of the *M. m. musculus* genome are extremely rare (Yang et al., 2011). Therefore, from a genetics perspective, it is interesting to analyze phenotype of the *M. m. musculus* derived strain and harness the functional genetic variations which could not be found in classical mice (Gregorova and Forejt, 2000). We compared phenotype of T lymphocytes between B6 and PWD strains in an attempt to search for genetic factors contributing to T cell biology, which is essential to understand host defense against infection and cancer (Malissen and Bongrand, 2015). We found that a typical subset of naïve CD4 T cells expressing high levels of CD62L and low levels of CD44, was absent in the PWD strain. Both CD4 and CD8 T cells express higher levels of CD44 on cell surface in PWD mice. To map the genetic factor(s) responsible for this phenotype, we crossed B6 and PWD mice to generate F1 hybrids, interestingly the F1 mice had an identical phenotype to PWD mice, suggesting a dominant effect of these gene(s). We then backcrossed F1 mice to B6 to segregate the causative genetic alleles. Indeed, in the backcrossed population, we found 50% mice (275 out of 559) displaying the PWD T cell phenotype.

By means of genome wide scanning with genetic markers, we identified a single locus on Chr2 of PWD mice. Since the PWD strain was involved in a very particular genetic resource, named chromosome substitution strains, which are available for rapid validation of genetic mapping, we used the C57BL/6J-Chr2^{PWD/Ph}/ForeJ strain which carries the entire Chr2 from PWD on the pure B6 background (Gregorova et al., 2008). We found that such B6.PWD-Chr2 mice have an identical T cell phenotype to PWD mice. In further fine mapping experiments, we found *Cd44 per se* was among the candidate genes in the mapped locus which was responsible for the T cell phenotype. CD44 is a cell surface marker for memory T cells and regulates memory cell survival (Baaten et al., 2010). In further experiments, we sought to inactivate the PWD derived *Cd44* locus in F1 hybrids via CRISPR/Cas9 genome editing, as previous studies showed that PAM sequences were necessary to cleave target DNA and allele specific modification of DNA sequence could be performed in mice (Hsu et al., 2013; Wu et al., 2013). To perform functional gene validation involving wild mice in our study, we employed allele specific genome editing which was also reported in human iPSCs and rats, we first analyzed the

sequences between PWD mice and B6 mice (Yoshimi et al., 2014; Smith et al., 2015). In the coding sequence of *Cd44*, two SNPs constitute CRISPR/Cas9 PAM sequences only for PWD mice which enables allele specific knockout of *Cd44*. Therefore, we could validate the functional relevance of this gene in F1 hybrids by specific knockout of the PWD allele. We designed two sgRNAs which were co-injected into B6 eggs fertilized by PWD sperms along with the CRISPR/Cas9 machinery. The results showed that among the 10 newborns, 80% of the mice carried ORF loss mutations of *Cd44* from PWD origin, with no mistargeting of the B6 *Cd44* allele. The CRISPR/Cas9 engineered F1 mutant mice had a phenotype identical to B6 mice. Therefore, we validated the causative gene for this phenotype, and our study provided a strategy for functional gene identification via allele-specific knockout, which could be useful for forward genetic studies in mice.

MATERIALS AND METHODS

Animals

PWD/PhJ and Chromosome substitution mice C57BL/6J-Chr2^{PWD/Ph}/ForeJ strain (B6.PWD-Chr2, Stock No: 005995) were purchased from the Jackson Laboratory¹ via distribution by Shanghai MAOSHENG YAN Biologic Science & Technology Co., Ltd., ICR outbred foster mice and C57BL/6 mice were purchased from Beijing Vital River Laboratory Animal Technology Co., Ltd., and all animal procedures were performed according to guidelines approved by the committee on animal care at Xinxiang Medical University.

Generation of CD44 Allele Specific Knockout F1 Mice

Single nucleotide polymorphisms or SNPs information of *Cd44* gene between PWK and C57BL/6 mice were analyzed from Mouse Genomes Project². The SNPs were validated in PWD mice by Sanger sequencing before design of allele specific knockout sgRNAs. The sgRNA (Shao et al., 2014) and Cas9 mRNA (Liang et al., 2015) were produced by *in vitro* transcription (IVT) as described previously. 4-week old female C57BL/6N mice were intra-peritoneally injected Pregnant Mare Serum Gonadotropin (PMSG) at 5 p.m. followed by Human Chorionic Gonadotropin (hCG, 10 units/mouse) 48 h later and then mated with 12-week old male PWD mice immediately. Fertilized embryos were collected in the next morning and Cas9 mRNA (50 ng/μL) and sgRNA (50 ng/μL) were microinjected into the cytoplasm of fertilized embryos by using a standard microinjection system (Eppendorf TransferMan® 4r, Eppendorf, Germany). Survived eggs were cultured at 37°C in 5% CO₂ over-night and in the next day were transferred into the oviductal ampullas of the surrogate ICR mice.

Genomic DNA of F1 newborn tails were subjected to PCR analysis by using Phusion High-Fidelity DNA

¹www.jax.org

²<https://www.sanger.ac.uk/science/data/mouse-genomes-project>

Polymerase (Thermo Fisher Scientific) with 5'-FAM-labeled primers to amplify two loci targeted by sgRNAs (Primer pair 1: GCTTTCTGGGGTGCTCTTCT; AGAGTA TGTGGGTGAAGGGG. Primer pair 2: TGGATGTGAGATTGG GTCGAAG; GGCAGCATGTGTCGAGAATTAC). The PCR products were run on an ABI 3730 DNA analyzer and data analyzed by GeneMapper software V3.1. The positions of the peaks indicate the lengths of PCR products (Velasco et al., 2007). For sequencing, PCR products were further cloned into T-vector (Tiangen, China). In general, 10 colonies were picked from each agar plate and were proceeded to Sanger sequencing.

Immunophenotyping by Flow Cytometry

The splenocytes and thymocytes of mice were stained with monoclonal antibody mixture and analyzed by flow cytometry. For activated T cell analysis, *in vitro* stimulation with anti-CD3 (3 µg/mL, 145-2C11, BD Biosciences) and anti-CD28 (1 µg/mL, 37.51, BD Biosciences) was performed in total splenocytes. The antibody labeling experiments were done as described in our previous studies for mouse immunophenotyping (Liang et al., 2013). In brief for splenocytes, 1 million cells were stained in 50 µL with antibody mixes CD3 (Alexa700), CD4 (eFlour450), CD8 (PE), CD25 (PE-Cy5.5), CD44 (Brilliant Violet 605), CD45 (APC-eFlour780), CD62L (APC), and TCRβ (FITC). Live cells were gated by Sytox blue staining and acquired on the FACS Canto flow cytometer (BD, United States). For immunophenotyping of the thymocytes, 2 million cells were stained in 50 µL with antibody mixture CD4 (PE), CD8 (Brilliant Violet 421), CD25 (APC), CD44 (Brilliant Violet 605) and acquired on the FACS Canto flow cytometer (BD, United States). The FACS data was analyzed using Flowjo software version 10.0.

QTL Mapping

The inbred strain C57BL/6 female mice and PWD/PhJ male mice were crossed to obtain F1 hybrids. The resultant F1 mice were crossed to C57BL/6 mice to establish N2 backcross population for linkage mapping. Hundred and twenty short random repeat (STR) markers were used to construct a genetic map using the function *est.map* from the R/qtl package. The phenotypic data and genetic map were analyzed with R/qtl using a standard interval mapping method (Broman et al., 2003). Initial interval mapping was performed with the R/qtl function *scannone* using EM algorithm (Lander and Botstein, 1989), and a significant threshold at $p = 0.05$ for the selected T cell trait was determined with 1000 permutation. Next, refining the localization of QTL with logarithm of odds (LOD) score above threshold. To fine-map the major QTL implicated in mouse T cell phenotype, CD44 expression in CD4 T cells, we selected markers flanking the candidate gene on Chr2, 3 STR markers D2Mit75, D2Mit97 and D2Mit100 were used to screen the N2 individuals for recombinants with 491 mice. Then 37 recombinant mice were genotyped with 6 additional STR markers to narrow the mapped interval. The primers of the markers used for genotyping are listed in **Supplementary Table S2**.

Statistical Analysis

GraphPad Prism software (version 7.0) was used for data analysis and statistical significance was assessed by unpaired, two-tailed Student's *t*-test. Data were presented as Mean ± SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

RESULTS

Wild Derived PWD/PhJ Mice Display a Distinct T Cell Phenotype From C57BL/6 Mice

T lymphocytes are pivotal components in the immune system of mammals to fight against pathogens and tumors and their development is under sophisticated molecular control which requires genetic dissection (Liang et al., 2013; Wang et al., 2017). In both CD4 and CD8 T cells, which are the two major subsets of T lymphocytes, CD44 expression on the cell surface is used as important marker to monitor T cell activation (Huet et al., 1989; DeGrendele et al., 1997). In a standard B6 mouse, assessment of CD44 expression on the cell surface helps define activated memory T cells. *In vitro*, T cell receptor (TCR) stimulation with anti-CD3 and anti-CD28, results in high CD44 expression (Flaherty and Reynolds, 2015). Interestingly, T cells from PWD mice displayed a distinct CD44 expression pattern in steady-state compared to B6 mice, and an increase of effector memory T cells upon TCR (CD3 and CD28) stimulation. As shown in **Figures 1A,B**, the CD44^{high}CD62L^{high} compartment among CD4-positive T cells, termed central memory cells (R2), was strikingly increased in PWD mice compared to B6 counterparts. While CD44^{low}CD62L^{high} naïve T cells (R1), conversely, were dramatically reduced. In addition, age matched PWD mice had significantly less CD44^{high}CD62L^{low} effector memory CD4 T cells (R3). Among CD8 T cells of PWD mice, the CD44^{high}CD62L^{high} compartment was also dramatically increased in frequency (R5), while CD44^{low}CD62L^{high} naïve T cells (R4), were again significantly reduced (**Figure 1C**). Next, we compared the mean fluorescent intensity (MFI) of CD44 by surface staining of CD4 and CD8 T cells; the MFI of CD44 in PWD CD4 T cells was 2.3-fold higher than in B6 mice, and in CD8 T cells CD44 expression was increased 5.8-fold (**Figures 1D,E**). Since CD44 expression can be induced by TCR stimulation, we analyzed the CD44 levels of CD4 and CD8 T cells between PWD and B6 mice following TCR stimulation. Notably, CD4 and CD8 T cells from PWD and B6 mice both showed a dramatic increase of CD44 expression 24 h after TCR stimulation, which could facilitate T cell migration to sites of inflammation. PWD mice had more CD44^{high}CD62L^{low} effector memory cells among both CD4 and CD8 T cells (R6, R7), following *in vitro* activation by TCR stimulation (**Figures 1F,G**). Unexpectedly, the MFI of CD44 on both CD4 and CD8 T cells was 3-fold higher in PWD mice following TCR stimulation (**Figures 1H,I**). These results showed that PWD mice had significantly higher expression of CD44 in steady-state, and more importantly, CD44 upregulation following TCR stimulation was more potent in T cells of PWD mice.

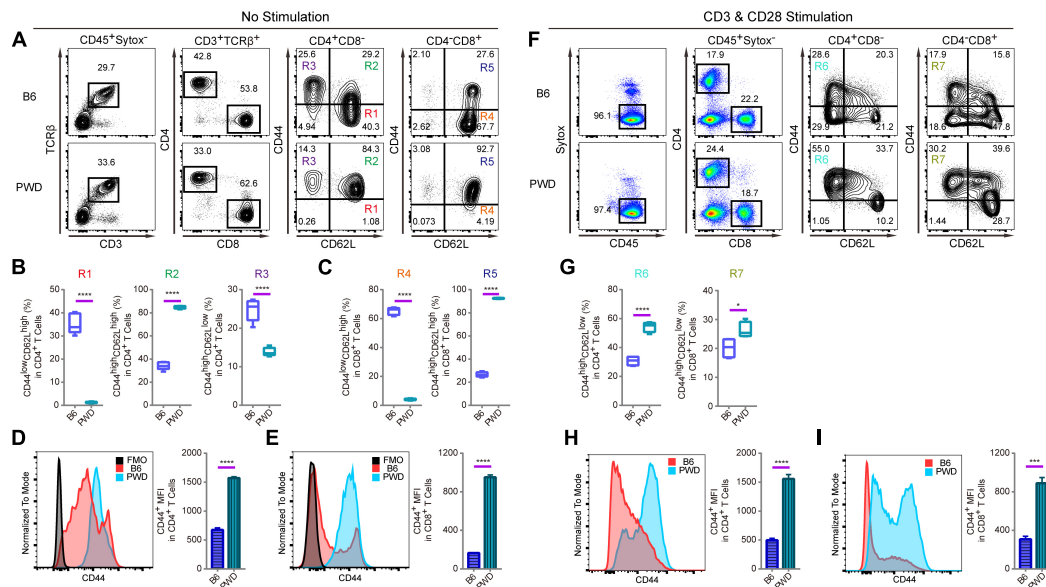


FIGURE 1 | T cell phenotyping of B6 and PWD mice with splenocytes. For steady state analysis of splenocytes *ex vivo*, surface staining with antibodies labeling CD3 and TCR β was used to define T cells in which CD4 and CD8 T cell subsets were further analyzed. For activated T cell analysis, *in vitro* stimulation with anti-CD3 and anti-CD28 was performed in total splenocytes. CD4 and CD8 T cells were analyzed for naïve and memory T cell frequencies in percentage and mean fluorescence intensity or MFI. **(A)** The gating method for T cells in the spleen of B6 and PWD mice. **(B)** Frequency of CD44^{low}CD62L^{high} cells (R1), CD44^{high}CD62L^{high} cells (R2) and CD44^{high}CD62L^{low} cells (R3) in CD4 T cells (CD4⁺CD8⁻ cells) of B6 and PWD mice. **(C)** Frequency of CD44^{low}CD62L^{high} cells (R4) and CD44^{high}CD62L^{high} cells (R5) in CD8 T cells (CD4⁻CD8⁺ cells) of B6 and PWD mice. **(D)** MFI of CD44 in CD4 T cells of B6 and PWD mice, FMO control was used as negative control by staining all the surface labeling antibodies except CD44. **(E)** MFI of CD44 in CD8 T cells of B6 and PWD mice, FMO control was used as negative control by staining all the surface labeling antibodies except CD44. **(F)** The gating method for immune cells in the spleen of B6 and PWD mice after anti-CD3 (3 μ g/mL, coated) and anti-CD28 (1 μ g/mL, soluble) stimulation. **(G)** Frequency of CD44^{high}CD62L^{low} cells in CD4 T cells (R6) and CD8 T cells (R7) of B6 and PWD mice after anti-CD3 (3 μ g/mL, coated) and anti-CD28 (1 μ g/mL, soluble) stimulation. **(H)** MFI of CD44 in CD4 T cells of B6 and PWD mice after anti-CD3 (3 μ g/mL, coated) and anti-CD28 (1 μ g/mL, soluble) stimulation. **(I)** MFI of CD44 positive cells in CD8 T cells of B6 and PWD mice after anti-CD3 (3 μ g/mL, coated) and anti-CD28 (1 μ g/mL, soluble) stimulation. Representative FACS data were from two independent experiments involving at least six animals for each group of mice. Data were analyzed by two-tailed Student's *t*-test. Data were presented as Mean \pm SEM. **p* < 0.05, ****p* < 0.001, *****p* < 0.0001.

The Splenic CD4 T Cell Phenotype of PWD Mice Is Dominant in F1 Hybrids

Since CD44 is implicated in T cell mobilization and differentiation into immunotolerant T cells, we analyzed the inheritance pattern of the T cell phenotype observed in PWD mice, prior to genetic mapping of the chromosomal region containing the causative gene(s) (DeGrendele et al., 1997; Wu et al., 2014). In F1 hybrids derived from B6 and PWD crosses, the CD4 T cells had significantly higher CD44 expression in *ex vivo* assays, closely resembling the phenotype of parental PWD mice with respect to R1 and R2 compartments, shown in **Figure 2A**. Furthermore, both F1 and PWD parental strains had dramatically different distributions of naïve and central memory T cells (**Figure 2B**). CD44 expression by CD4 T cells, was significantly higher in the PWD parental strain and F1 mice than that of B6 mice, and this increase in CD44 was not distinguishable between male and female F1 mice (**Figure 2C**). Since the increased CD44 expression on CD4 T cells was observed in both F1 and parental PWD mice, we regard this phenotype as “dominant,” even though the phenotype in CD8 T cells was less stringent (data not shown). We further analyzed the CD44 expression in F1 hybrids in thymocytes which give

rise to T cells. CD4⁻CD8⁻ double negative (DN) thymocytes, which are the progenitors of T cells, cells were further divided into DN1, DN2, DN3, and DN4 populations (**Figure 2D**). As shown in **Figures 2E,F**, F1 mice either displayed an intermediate phenotype, between the two parental strains of B6 and PWD, or no significant difference from B6 mice, for multiple parameters including the thymocyte number and CD44 expression. In more complete comparisons between F1 and parental mice for CD44 expression in various populations in thymus, we found that F1 phenotype was generally intermediate (**Figure 2G**). These results indicated that the dominant PWD phenotype in CD4 T cells of increased CD44 expression, was restricted to T cells in the periphery. Therefore, we used CD4 T cell phenotype to perform further genetic mapping experiments to facilitate identification of causative gene.

A Single Genetic Locus on Chr2 Controls the PWD CD4 T Cell Phenotype

To map the genetic factor(s) regulating the T cell phenotype we observed in PWD mice, we constructed a backcross pedigree for linkage mapping by crossing F1 hybrids to B6 mice (Mihola et al., 2009; Su et al., 2009). In the

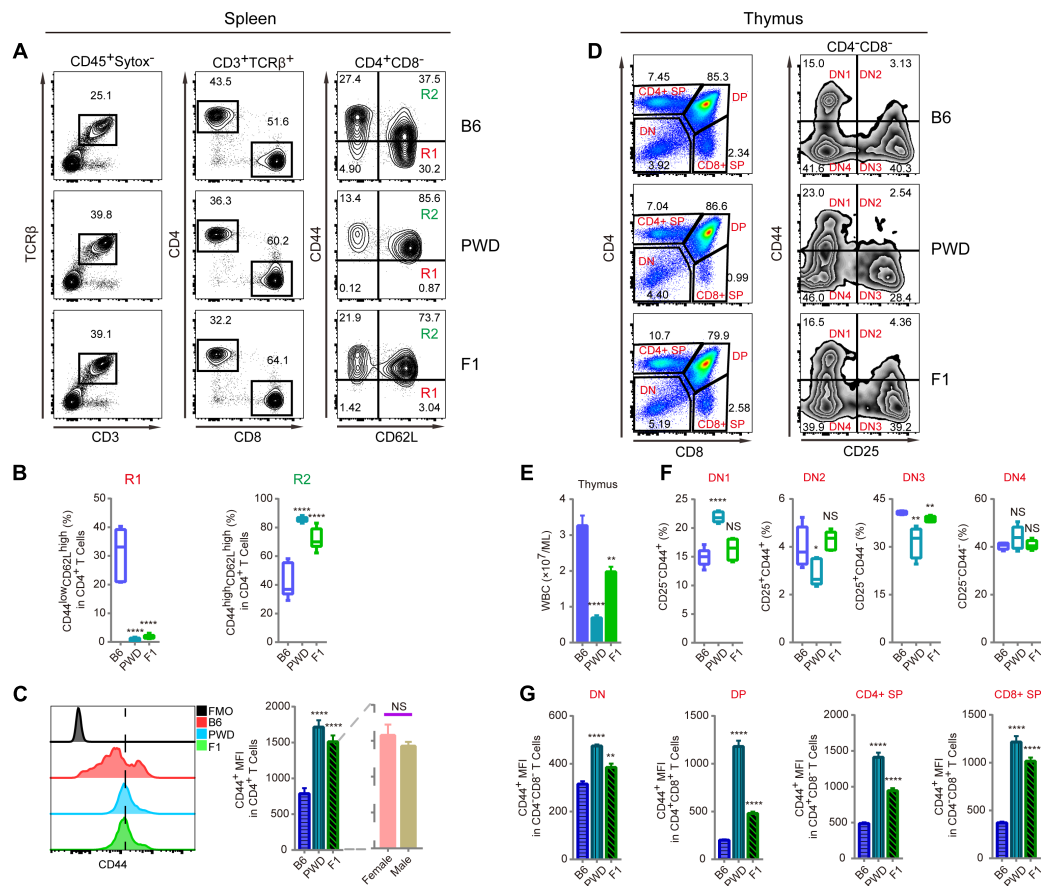


FIGURE 2 | T cell phenotyping of B6, PWD and F1 mice for spleen and thymus. For peripheral T cell analysis, splenocytes were analyzed ex vivo for CD44 and CD62L expression in CD4 T cells. For thymocytes analysis, CD4 and CD8 T single positive T cells and their precursors namely CD4 and CD8 double positive and double negative cells were analyzed. Inside the double negative compartment, further analysis with CD25 and CD44 were used to define each precursor population. **(A)** The gating method for T cells in the spleen of B6, PWD and F1 mice. **(B)** Frequency of CD44^{low}CD62L^{high} cells (R1) and CD44^{high}CD62L^{high} cells (R2) in CD4⁺CD8⁻ cells of B6, PWD and F1 splenocytes. **(C)** Overlay of CD44 MFI in CD4⁺ T cells of B6, PWD and F1 splenocytes, FMO control was used as negative control by staining all the surface labeling antibodies except CD44. **(D)** The gating method for immune cells in the thymus of B6, PWD and F1 mice. **(E)** Absolute thymocyte count of B6, PWD and F1 mice. **(F)** Frequency of DN1-DN4 cells in DN cells of B6, PWD and F1 thymocytes. **(G)** MFI of CD44 positive cells in DN, DP, CD4⁺ SP and CD8⁺ SP cells of B6, PWD and F1 thymocytes. Representative FACS data were from two independent experiments involving at least seven animals for each group of mice. Data were analyzed by two-tailed Student's *t*-test. Data were presented as Mean ± SEM. ***p* < 0.01, *****p* < 0.0001.

resultant backcross N2 animals, we observed obvious phenotype segregation in 1:1 ratio, suggesting a single gene or closely linked genes were responsible for the phenotype. Among the 559 N2 mice, 275 mice had high CD44 expression, another 284 animals had low CD44 expression, sex ratio was close to 1:1. Genome wide scanning with 120 microsatellite markers was performed initially with 68 N2 mice (Figure 3A). We found the highest LOD score was located at Chr2 D2Mit395 (Figures 3B,C). From the chromosome substitution (CS) mouse resource or consomic mice, the C57BL/6J-Chr2^{PWD/Ph/ForeJ} (B6.PWD-Chr2, Stock No: 005995) strain carries a complete Chr2 from PWD and the rest of the genome is of C57BL/6 origin. In these CS mice we confirmed the T cell phenotype, originally found in PWD mice. In the B6.PWD-Chr2 CS strain, CD44 expression on CD4 T cells was comparable to the PWD mice and significantly higher than the B6 mice (Supplementary Figures S1A,B).

Therefore, we mapped the genetic locus for this phenotype to Chr2, and further validated the localization of the causative locus in CS mice. Following initial linkage mapping, fine mapping using an additional 491 N2 mice, further identified that a 1 cM chromosomal segment between D2Mit300 and D2Mit127 was responsible for this phenotype (Figure 3D and Supplementary Table S1). Within this chromosomal segment, there were 21 protein coding genes. Referring to Immunological Genome Project database (Heng et al., 2008; Benoist et al., 2012), 8 genes were expressed in T cells which included Cd44, Cat, Caprin, Trim44, Ldlrad3, Traf6, Rag1, and Rag2 (Supplementary Table S1). We suspected *Cd44* as the candidate gene, since the T cell phenotype was defined by changes in CD44 expression, and more notably among the candidate genes only *Cd44* mRNA expression was higher in PWD CD4 T cells than the B6 controls (Immgen database, and data not shown).

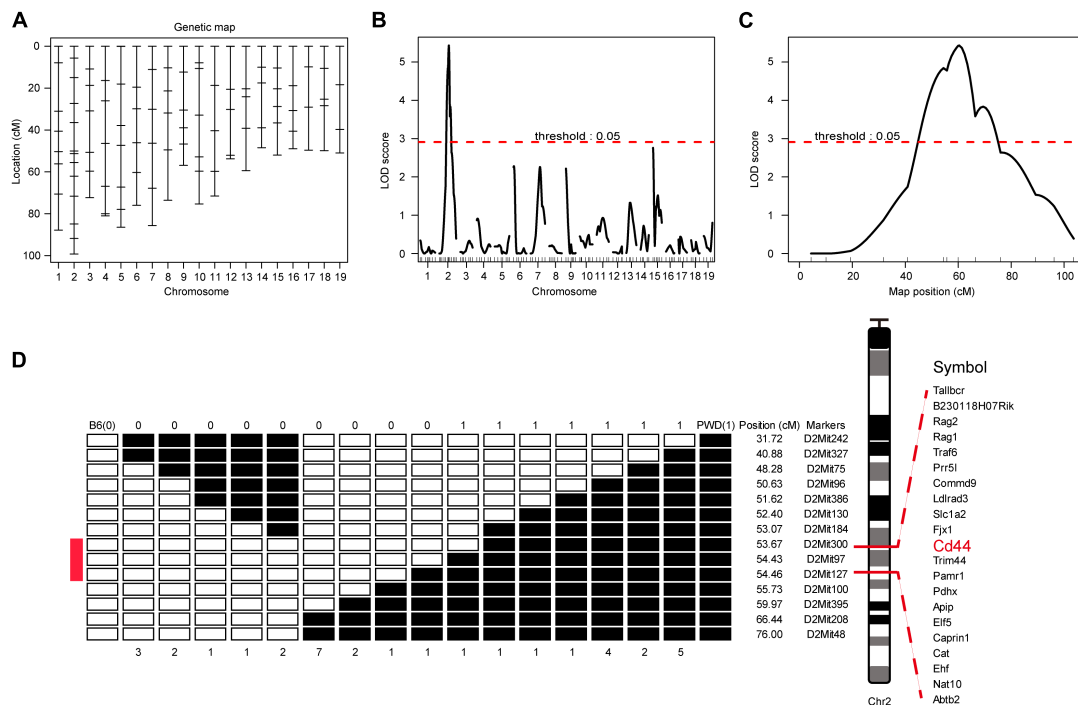


FIGURE 3 | Genome-wide scanning to map QTL influencing the CD4 T phenotype. The dominant phenotype of CD4 T cells was used to categorize mice with PWD phenotype or B6 phenotype by absence or presence of the “naïve” T cells which were $CD44^{low}CD62L^{high}$ (refer to **Figure 2A**, R1). **(A)** The genetic marker distribution throughout the mouse genome to analyze N2 population with 120 STRs. **(B)** LOD score curves of the initial mapping with 68 animals (37 phenotypic mice, and 31 non-phenotypic mice), and chromosome 1–19 were represented numerically on the X-axis. The relative width of the space allotted for each chromosome reflects the number of microsatellite markers typed for that chromosome. The Y-axis represents the LOD score. The horizontal dashed lines denote genome-wide empirical thresholds for significant ($p = 0.05$) linkage. **(C)** Interval mapping identified one major QTL in the position from 47.42 to 66.42 cM on mouse chromosome 2. **(D)** Fine-mapping of the major Chr2 QTL in an N2 population of 491 individuals. Open box denotes B6 genotype of the STR markers, and filled box represents PWD genotype. “0” was used to represent B6 phenotype and “1” for PWD phenotype. Inside the fine mapped interval between D2Mit300 and D2Mit127, candidate genes were listed.

Allele Specific Knockout of PWD *Cd44* in F1 Mice by CRISPR/Cas9

We suspected that *Cd44* *per se* was responsible for the dominant T cell phenotype in PWD and F1 mice, based on fine mapping which resulted in 21 candidate genes and 2 differentially expressed genes between the parental strains (**Supplementary Table S1**). To confirm the dominant effect of the *Cd44* allele originating from PWD mice on the T cell phenotype we observed in F1 mice, we designed allele specific knockout (ASK) to inactivate only the PWD copy of *Cd44* gene. CRISPR/Cas9 genome editing requires guide RNA and PAM sequence to target specific DNA elements. In the absence of PAM, genome editing efficiency was not detectable. To validate that PWD *Cd44* itself determines the T cell phenotype we observed in B6/PWD F1 hybrids, we sought to inactivate only the PWD copy of the gene, since the phenotype was dominant. The exon and intron structure of murine *Cd44* and SNPs between B6 and PWK, which is closely related to PWD, are depicted in **Figure 4A**. The whole genome sequence and SNPs for PWK, but not PWD, mice are available from the Mouse Genomes Project³. The CD44

protein coding sequences of B6 and PWK were aligned and 8 out of 19 exons had SNPs (**Figure 4A**). Among these SNPs we selected those that formed PAM (5'-NGG-3') sequences which were only existing in the PWK mice, and such SNPs were later confirmed in PWD mice by Sanger sequencing (**Figure 4B**). The PWD copy of *Cd44* were sequenced for cDNA, and we found 2 amino acid absence in comparison to B6 protein sequence (**Supplementary Data Sheet S1**). The guide RNA was prepared by IVT, and potential off-targets were analyzed using the CRISPOR software, as described previously (Luo et al., 2018). Before microinjection of the CRISPR/Cas9 machinery designed to specifically target *Cd44* of PWD mice, fertilized eggs were prepared by superovulation of B6 female mice aged 4–6 weeks and fertilization with sperm cells from PWD male mice aged 10–12 weeks. Two sets of allele specific guide RNAs and Cas9 mRNA were co-injected (**Figure 4C**). The engineered eggs were developed in ICR outbred foster mice and genotyped using mouse tail tip DNA from newborns by fluorescent PCR and capillary gel electrophoresis (**Figure 4D**). As shown in **Figure 4E**, in total 170 fertilized eggs were subject to microinjection and 111 live eggs were transplanted, resulting in 8 mice with Indels. In theory, the Indels should only occur for the PWD allele of *Cd44*, which possesses PAM sequences that are absent in B6 mice.

³www.sanger.ac.uk/sanger/Mouse_SnpViewer

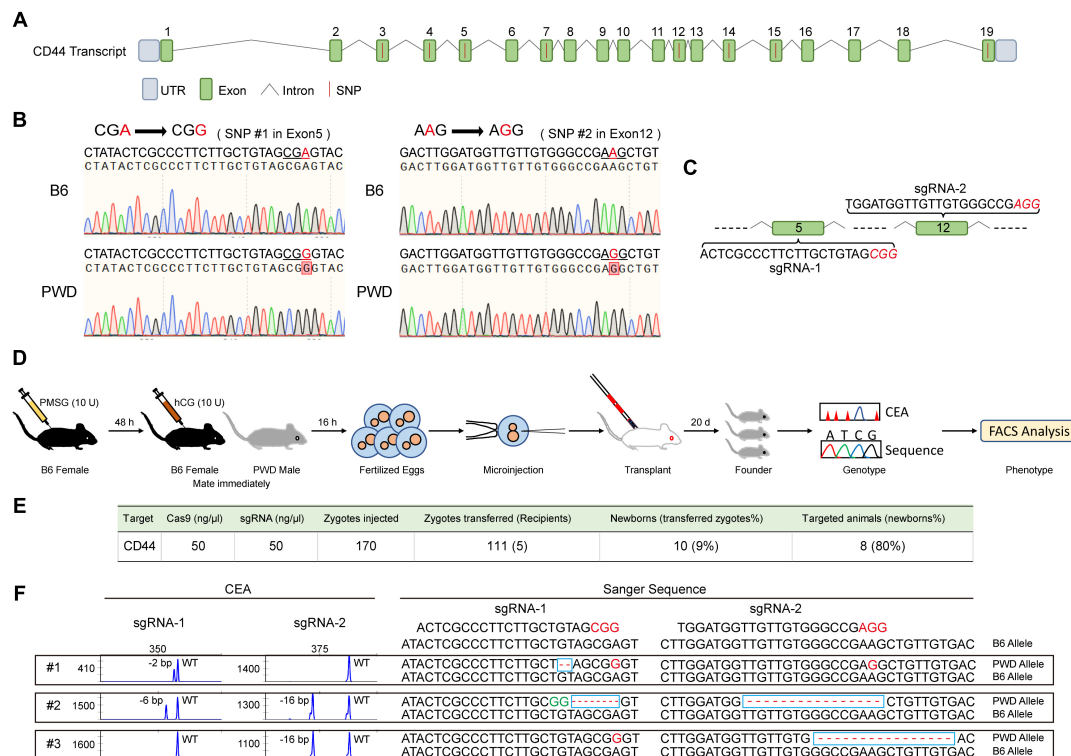


FIGURE 4 | Generation and genotyping of PWD CD44 allele specific knockout in F1 mice. **(A)** The distribution of SNPs in *Cd44* exons of the whole genome sequenced PWK strain which is a close relative of PWD strain. **(B)** Two PAM sequences (5'-NGG-3') that only existed in the PWD mice were selected and confirmed by Sanger sequencing. **(C)** SgRNA sequences in *Cd44* exon5 and exon12 which were selected for allele specific knockout. **(D)** Workflow of producing PWD CD44 allele specific knockout in F1 mice. **(E)** Efficiency of CRISPR/Cas9 system in specific inactivation of CD44 from PWD origin. **(F)** The genotyping of F1-CD44^{ASK} mice by using capillary electrophoresis analysis (left) and Sanger sequencing (right).

To validate the consequence of genome editing by allele specific targeting, we sequenced all the DNA samples from mice which carried Indel mutations by TA cloning of the PCR products and Sanger sequencing (Luo et al., 2018). Indeed, the Indel mutations were stringently restricted to the PWD allele, and all the B6 alleles tested were completely intact (Figure 4F and Supplementary Table S3). By selecting PAM sequences specific to the allele of PWD, we obtained animals carrying only the B6 copy of *Cd44* gene in F1 hybrids which were used for further phenotyping.

Allele Specific Knockout of PWD *Cd44* Rescues the T Cell Phenotype in Hybrid F1 Mice

The T cell phenotype we observed in PWD and F1 mice was comparable, furthermore, the phenotype was consistent in N2 backcross mice and distributed in mendelian ratio. We mapped the causative PWD gene to Chr2 and further validated the phenotype using chromosome substitution mice, which have a clean genomic background identical to B6 animals, except for the donor chromosome. CRISPR/Cas9 genome editing allowed us to obtain F1 animals that were specifically deficient in the *Cd44* allele of PWD origin. In such allele specific knockout (ASK) mice, we tested the T cell phenotype and found that the phenotype in ASK mice differed significantly from that

of PWD and F1 mice, and resembled the B6 phenotype. As shown in Figures 5A–C, comparison of the T cell phenotype between F1 mice and *Cd44*^{ASK} F1 mice for CD4 and CD8 subsets revealed that inactivation of PWD CD44 resulted in significant alteration, notably increased frequency of CD44^{low}CD62L^{high} naïve CD4 (R1) and CD8 T cells (R3) and conversely, decrease of CD44^{high}CD62L^{high} central memory CD4 (R2) and CD8 T cell (R4) frequencies. *Cd44*^{ASK} F1 mice, that were deficient in only the PWD copy of *Cd44*, had significantly decreased CD44 MFI in both CD4 and CD8 T cells (Figures 5D,E). We performed further experiments to compare *Cd44*^{ASK} F1 mice and B6 mice, interestingly both CD4 and CD8 T cells had decreased CD44 expression in allele specific knockout F1 mice, to a level comparable to B6 T cells (Supplementary Figures S2A–C). Therefore, allele specific knockout of PWD CD44 in F1 mice was sufficient to restore the T cell phenotype to that of parental B6 mice. This allele specific knockout strategy demonstrated that a candidate gene from a specific parental origin can be precisely targeted in hybrid mice.

DISCUSSION

Mouse genetics contributes enormously to understanding of human pathophysiology (Demant, 2003; Nguyen and Xu, 2008).

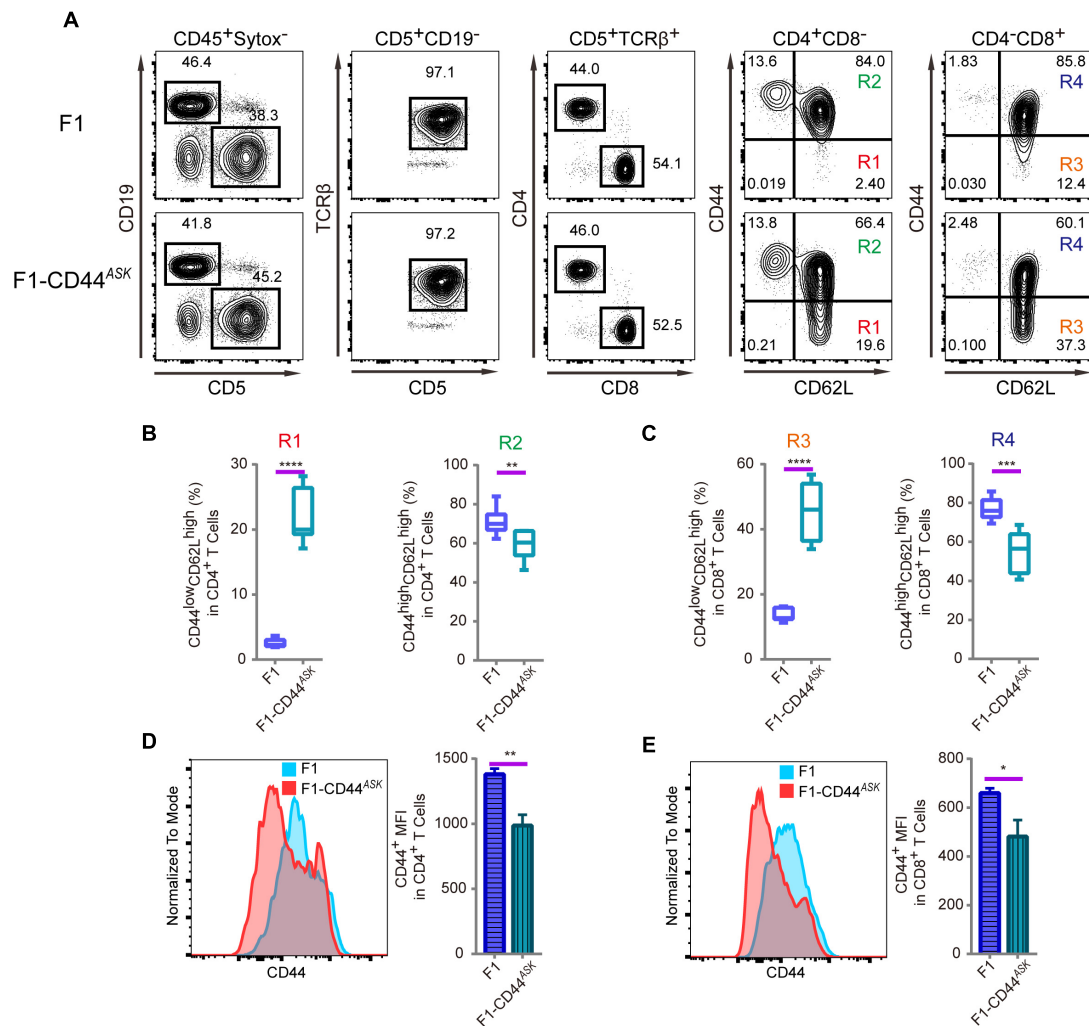


FIGURE 5 | Immunophenotyping of T cells in splenocytes of F1 and F1-CD44^{ASK} mice. In the F1 mice, consequence of the allele specific knockout of CD44 was analyzed by assessing the alteration of phenotype which was observed in PWD parental strain. **(A)** The gating method for T cells in the spleen of F1 and F1-CD44^{ASK} mice. **(B)** Frequency of CD44^{low}CD62L^{high} cells (R1) and CD44^{high}CD62L^{high} cells (R2) in CD4⁺CD8⁻ cells of F1 and F1-CD44^{ASK} mice. **(C)** Frequency of CD44^{low}CD62L^{high} cells (R3) and CD44^{high}CD62L^{high} cells (R4) in CD4⁺CD8⁺ cells of F1 and F1-CD44^{ASK} mice. **(D)** MFI of CD44 positive cells in CD4⁺ cells of F1 and F1-CD44^{ASK} mice. **(E)** MFI of CD44 positive cells in CD8⁺ cells of F1 and F1-CD44^{ASK} mice. Representative FACS data were from two independent experiments involving at least six mice for each genotype. Data were analyzed by two-tailed Student's *t*-test. Data were presented as Mean ± SEM. **p* < 0.05, ***p* < 0.01, ****p* < 0.001, *****p* < 0.0001.

Functional mirroring of human immune cells in mice has led to discovery of novel cellular and genetic components of the immune system (Liang et al., 2013; Wang et al., 2016). Gene discovery via forward genetic mapping in mice has been highly successful in the last three decades, as genome wide scanning revealed thousands of genetic loci implicated in various diseases and immunological traits (Peters et al., 2007; Siggs, 2014). However, identification of the causative gene in the mapped loci, remains extremely difficult, due to extensive linkage disequilibrium and low efficiency in validation of candidate genes (Peters et al., 2007). The development of chromosome substitution mice provides a straight forward and rapid method for validation of mapped loci, maintaining a homogeneous genomic background identical to the reference C57BL/6 mice,

however, cloning the causative gene is still challenging. In this study, we first discovered a T cell phenotype which was dramatically different between B6 and PWD mice, and further mapped locus on Chr2 responsible for the phenotypic difference. To validate the causative role of the mapped locus, we analyzed the chromosome substitution line B6.PWD-Chr2 and found the same phenotype as observed in PWD parental mice. Such results excluded genetic factors outside Chr2 that could co-contribute to the T cell phenotype we observed in PWD mice.

Fine mapping can be achieved with the N2 progeny by increasing the number of samples, to analyze more recombinants, or increasing the generations of backcrossing to generate more recombination flanking the causative gene(s) (Jeffs et al., 2000). In the initial mapping of the dominant locus, we used 68 mice,

and for fine mapping we analyzed 491 animals that gave rise to an interval confining the causative gene inside a 1 cM segment on Chr2 of PWD strain. Among the candidate genes, we highly suspected *Cd44* itself as causative gene, as the CD44 protein was more abundantly expressed in CD4 T cells of PWD mice (Heng et al., 2008). Our experiments reflect numerous other studies using forward genetics in mice to identify genetic variation in inbred lines based on phenotyping and to test for candidate genes via linkage mapping, however, the challenge to confirm the role of the PWD allele in regulating T cell phenotype lied in obtaining selective knockout of this allele and test its consequence on a background that should otherwise maintain the phenotype. Knockout of the gene on a B6 background does not provide direct evidence of the functional role, since B6 mice themselves do not have the PWD phenotype. We found that F1 hybrids keep the PWD phenotype, therefore we could test by specific CRISPR/Cas9 mediated targeting of PWD allele of candidate gene *Cd44*.

We have established a procedure for producing allele specific knockout of a candidate gene by designing sgRNA with PAM sequences that only exist on the PWD background. Even though both copies of the gene from B6 and PWD parents were exposed to CRISPR/Cas9 machinery, precise and allele specific targeting of only the PWD allele of *Cd44* was achieved in our study, and it is interesting to note that such experiments could be further improved by applying variants of Cas9 nuclease to obtain higher fidelity in genome editing and non-canonical PAM sequences could be applied to fit broader range of genomic contexts (Zhang et al., 2014; Kleinstiver et al., 2016). In the F1 animals which were mutated only in the PWD copy of *Cd44* had an altered phenotype compared to B6 parental mice. Therefore, by allele specific targeting of PWD CD44, we confirmed the functional role of CD44 itself in maintaining the CD44^{high}CD62L^{high} population only found in PWD mice. The molecular mechanisms behind how PWD CD44 contributes to this T cell phenotype, still remain to be elucidated, which could provide further hints to uncover new roles for CD44 since this molecule has been already found crucial for regulatory T cell development (Wu et al., 2014). More importantly, our approach of allele specific knockout could provide a new strategy to inactivate disease alleles while keeping the normal allele of the gene intact.

REFERENCES

- Baaten, B. J., Li, C. R., Deiro, M. F., Lin, M. M., Linton, P. J., and Bradley, L. M. (2010). CD44 regulates survival and memory development in Th1 cells. *Immunity* 32, 104–115. doi: 10.1016/j.immuni.2009.10.011
- Benoist, C., Lanier, L., Merad, M., Mathis, D., and Immunological Genome Project (2012). Consortium biology in immunology: the perspective from the immunological genome project. *Nat. Rev. Immunol.* 12, 734–740. doi: 10.1038/nri3300
- Broman, K. W., Wu, H., Sen, S., and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19, 889–890. doi: 10.1093/bioinformatics/btg112
- DeGrendele, H. C., Estess, P., and Siegelman, M. H. (1997). Requirement for CD44 in activated T cell extravasation into an inflammatory site. *Science* 278, 672–675. doi: 10.1126/science.278.5338.672
- Demant, P. (2003). Cancer susceptibility in the mouse: genetics, biology and implications for human cancer. *Nat. Rev. Genet.* 4, 721–734. doi: 10.1038/nrg1157
- Flaherty, S., and Reynolds, J. M. (2015). Mouse naive CD4 + T cell isolation and in vitro differentiation into T cell subsets. *J. Vis. Exp.* 98:52739. doi: 10.3791/52739
- Gregorova, S., Divina, P., Storchova, R., Trachtulec, Z., Fotopulosova, V., Svenson, K. L., et al. (2008). Mouse consomic strains: exploiting genetic divergence between *Mus m. musculus* and *Mus m. domesticus* subspecies. *Genome Res.* 18, 509–515. doi: 10.1101/gr.7160508
- Gregorova, S., and Forejt, J. (2000). PWD/Ph and PWK/Ph inbred mouse strains of *Mus m. musculus* subspecies—a valuable resource of phenotypic variations and genomic polymorphisms. *Folia Biol.* 46, 31–41.
- Guenet, J. L., and Bonhomme, F. (2003). Wild mice: an ever-increasing contribution to a popular mammalian model. *Trends Genet.* 19, 24–31. doi: 10.1016/S0168-9525(02)00007-0

DATA AVAILABILITY STATEMENT

All datasets for this study are included in the manuscript and the **Supplementary Files**.

AUTHOR CONTRIBUTIONS

YL designed the project and wrote the manuscript. LL supervised the project, analyzed the data, and prepared the figures. TC and ZL performed the experiments and analyzed the data. YZ established the N2 population and performed the initial QTL mapping. ZL, RH, LH, YG, ZC, QZ, LS, WZ, and XQ were involved in preparation of mRNA, mouse embryos and genotyping. EK and ZZ assisted in supervising the project. TL contributed to writing and revising the manuscript. All authors read and approved the final manuscript.

FUNDING

The work was supported by NSFC No. 31400759 and No. 81471595 to YL, NSFC No. 81501342 to LL, and by the Foundation of Henan Educational Committee No. 16HASTIT030 to YL.

ACKNOWLEDGMENTS

We would like to thank Assegai Medical Laboratory Xinxiang for assistance with the capillary array electrophoresis-based genotyping, and we are also like to thank the core facility of flow cytometry in Xinxiang Medical University for technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00124/full#supplementary-material>

- Heng, T. S., Painter, M. W., and Immunological Genome Project Consortium (2008). The immunological genome project: networks of gene expression in immune cells. *Nat. Immunol.* 9, 1091–1094. doi: 10.1038/ni1008-1091
- Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31, 827–832. doi: 10.1038/nbt.2647
- Huet, S., Groux, H., Caillou, B., Valentin, H., Prieur, A. M., and Bernard, A. (1989). CD44 contributes to T cell activation. *J. Immunol.* 143, 798–801.
- Jeffs, B., Negrin, C. D., Graham, D., Clark, J. S., Anderson, N. H., Gauguier, D., et al. (2000). Applicability of a "speed" congenic strategy to dissect blood pressure quantitative trait loci on rat chromosome 2. *Hypertension* 35, 179–187. doi: 10.1161/hyp.35.1.179
- Keane, T. M., Goodstadt, L., Danecek, P., White, M. A., Wong, K., Yalcin, B., et al. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477, 289–294. doi: 10.1038/nature10413
- Kleistiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., et al. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529, 490–495. doi: 10.1038/nature16526
- Lander, E. S., and Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185–199.
- Liang, P., Xu, Y., Zhang, X., Ding, C., Huang, R., Zhang, Z., et al. (2015). CRISPR/Cas9-mediated gene editing in human triploid zygotes. *Protein Cell* 6, 363–372. doi: 10.1007/s13238-015-0153-5
- Liang, Y., Cucchetti, M., Roncagalli, R., Yokosuka, T., Malzac, A., Bertoso, E., et al. (2013). The lymphoid lineage-specific actin-uncapping protein Rltpr is essential for costimulation via CD28 and the development of regulatory T cells. *Nat. Immunol.* 14, 858–866. doi: 10.1038/ni.2634
- Luo, J., Lu, L., Gu, Y., Huang, R., Gui, L., Li, S., et al. (2018). Speed genome editing by transient CRISPR/Cas9 targeting and large DNA fragment deletion. *J. Biotechnol.* 281, 11–20. doi: 10.1016/j.jbiotec.2018.06.308
- Malissen, B., and Bongrand, P. (2015). Early T cell activation: integrating biochemical, structural, and biophysical cues. *Annu. Rev. Immunol.* 33, 539–561. doi: 10.1146/annurev-immunol-032414-112158
- Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J. C., and Forejt, J. (2009). A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* 323, 373–375. doi: 10.1126/science.1163601
- Nguyen, D., and Xu, T. (2008). The expanding role of mouse genetics for understanding human biology and disease. *Dis. Model. Mech.* 1, 56–66. doi: 10.1242/dmm.000232
- Peters, L. L., Robledo, R. F., Bult, C. J., Churchill, G. A., Paigen, B. J., and Svenson, K. L. (2007). The mouse as a model for human biology: a resource guide for complex trait analysis. *Nat. Rev. Genet.* 8, 58–69. doi: 10.1038/nrg2025
- Rosshart, S. P., Vassallo, B. G., Angeletti, D., Hutchinson, D. S., Morgan, A. P., Takeda, K., et al. (2017). Wild mouse gut microbiota promotes host fitness and improves disease resistance. *Cell* 171, 1015–1028. e13. doi: 10.1016/j.cell.2017.09.016
- Shao, Y., Guan, Y., Wang, L., Qiu, Z., Liu, M., Chen, Y., et al. (2014). CRISPR/Cas-mediated genome editing in the rat via direct injection of one-cell embryos. *Nat. Protoc.* 9, 2493–2512. doi: 10.1038/nprot.2014.171
- Siggs, O. M. (2014). Dissecting mammalian immunity through mutation. *Immunol. Cell Biol.* 92, 392–399. doi: 10.1038/icb.2014.8
- Smith, C., Abalde-Atristain, L., He, C., Brodsky, B. R., Braunstein, E. M., Chaudhari, P., et al. (2015). Efficient and allele-specific genome editing of disease loci in human iPSCs. *Mol. Ther.* 23, 570–577. doi: 10.1038/mt.2014.226
- Su, Z., Ishimori, N., Chen, Y., Leiter, E. H., Churchill, G. A., Paigen, B., et al. (2009). Four additional mouse crosses improve the lipid QTL landscape and identify Lpiga as a QTL gene. *J. Lipid Res.* 50, 2083–2094. doi: 10.1194/jlr.M900076-JLR200
- Velasco, E., Infante, M., Duran, M., Perez-Cabornero, L., Sanz, D. J., Esteban-Cardenosa, E., et al. (2007). Heteroduplex analysis by capillary array electrophoresis for rapid mutation detection in large multiexon genes. *Nat. Protoc.* 2, 237–246. doi: 10.1038/nprot.2006.482
- Wang, H., Li, S., Chao, T., Wang, X., Shi, L., Zhang, L., et al. (2017). A point mutation in the extracellular domain of CD4 completely abolishes CD4 T cell development in C57BL/6 mouse. *Mol. Immunol.* 92, 12–20. doi: 10.1016/j.molimm.2017.09.015
- Wang, Y., Ma, C. S., Ling, Y., Bousfiha, A., Camcioglu, Y., Jacquot, S., et al. (2016). Dual T cell- and B cell-intrinsic deficiency in humans with biallelic RLTPR mutations. *J. Exp. Med.* 213, 2413–2435. doi: 10.1084/jem.2016.0576
- Wu, C., Thalhamer, T., Franca, R. F., Xiao, S., Wang, C., Hotta, C., et al. (2014). Galectin-9-CD44 interaction enhances stability and function of adaptive regulatory T cells. *Immunity* 41, 270–282. doi: 10.1016/j.immuni.2014.06.011
- Wu, Y., Liang, D., Wang, Y., Bai, M., Tang, W., Bao, S., et al. (2013). Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell* 13, 659–662. doi: 10.1016/j.stem.2013.10.016
- Yang, H., Wang, J. R., Didion, J. P., Buus, R. J., Bell, T. A., Welch, C. E., et al. (2011). Subspecific origin and haplotype diversity in the laboratory mouse. *Nat. Genet.* 43, 648–655. doi: 10.1038/ng.847
- Yoshimi, K., Kaneko, T., Voigt, B., and Mashimo, T. (2014). Allele-specific genome editing and correction of disease-associated phenotypes in rats using the CRISPR-Cas platform. *Nat. Commun.* 5:4240. doi: 10.1038/ncomms5240
- Zhang, Y., Ge, X., Yang, F., Zhang, L., Zheng, J., Tan, X., et al. (2014). Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Sci. Rep.* 4:5405. doi: 10.1038/srep05405

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Chao, Liu, Zhang, Zhang, Huang, He, Gu, Chen, Zheng, Shi, Zheng, Qi, Kong, Zhang, Lawrence, Liang and Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Guidelines for Fluorescent Guided Biallelic HDR Targeting Selection With PiggyBac System Removal for Gene Editing

Javier Jarazo[†], Xiaobing Qing[†] and Jens C. Schwamborn*

Developmental and Cellular Biology, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Belvaux, Luxembourg

OPEN ACCESS

Edited by:

David Jay Segal,
University of California, Davis,
United States

Reviewed by:

Amar M. Singh,
University of Georgia, United States
Michael Tsang,
University of Pittsburgh, United States

*Correspondence:

Jens C. Schwamborn
jens.schwamborn@uni.lu

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Genomic Assay Technology,
a section of the journal
Frontiers in Genetics

Received: 30 November 2018

Accepted: 22 February 2019

Published: 13 March 2019

Citation:

Jarazo J, Qing X and
Schwamborn JC (2019) Guidelines
for Fluorescent Guided Biallelic HDR
Targeting Selection With PiggyBac
System Removal for Gene Editing.
Front. Genet. 10:190.
doi: 10.3389/fgene.2019.00190

The development of new and easy-to-use nucleases, such as CRISPR/Cas9, made tools for gene editing widely accessible to the scientific community. Cas9-based gene editing protocols are robust for creating knock-out models, but the generation of single nucleotide transitions or transversions remains challenging. This is mainly due to the low frequency of homology directed repair, which leads to the screening of a high number of clones to identify positive events. Moreover, lack of simultaneous biallelic modifications, frequently results in second-allele indels. For example, while one allele might undergo homology directed repair, the second can undergo non-homologous end joining repair. Here we present a step-wise protocol for biallelic gene editing. It uses two donors carrying a combination of fluorescent reporters alongside homology arms directed to the same genomic region for biallelic targeting. These homology arms carry the desired composite of modifications to be introduced (homozygous or heterozygous changes). Plus, the backbone of the plasmid carries a third fluorescent reporter for negative selection (to discard random integration events). Fluorescent selection of non-random biallelic targeted clones can be performed by microscopy guided picking or cell sorting (FACS). The positive selection module (PSM), carrying the fluorescence reporter and an antibiotic resistance, is flanked by inverted terminal repeats (ITR) that are recognized by transposase. Upon purification of the clones correctly modified, transfection of the excision-only transposase allows the removal of the PSM resulting in the integration of only the desired modifications.

Keywords: CRISPR, biallelic, HDR, genome editing, iPSC

INTRODUCTION

Disease modeling in vitro had a technological leap with the advent of biotechnology tools such as the induction to pluripotency states, and the targeted nucleotide modifications by gene editing techniques (Hockemeyer and Jaenisch, 2016). Combining both techniques allows us to validate the effect of disease causing point mutations, as well as the influence of risk variants in the context of a human cell model (Jehuda et al., 2018). Moreover, it can help in the assessment of disease modifiers by introducing mutations that lead to phospho-mimetic or phospho-null protein variants

identifying them as novel targets for drug development, without the influence of exogenous or overexpressed sequences (Chen and Cole, 2015).

Even though CRISPR/Cas9 represents the democratization of gene editing tools for most research labs (Jasin and Haber, 2016), certain aspects of the process demonstrated to be cumbersome in practice, such as the number of clones to be screened, reduced biallelic targeting or high on target non-homologous end joining (NHEJ). We previously reported the concept of circumventing these issues by using two constructs targeting the same genomic region but having different positive selection modules (PSM) (Arias-Fuenzalida et al., 2017). These PSM have different fluorescent proteins (namely EGFP and dTomato) allowing the identification of a correct knock-in in both alleles simultaneously. Compared to other systems using only an antibiotic resistance (e.g., puromycin) in the PSM, the use of fluorescent proteins circumvents clones that underwent NHEJ repair in the second allele. The PSM is surrounded by transposon inverted terminal repeats (ITRs) of the piggyBac transposon system for removal of the PSM after selection. The transposase enzyme recognizes these ITRs, excising the sequenced flanked by them reconstituting a TTAA motif in the host genome (Yusa et al., 2011). The use of excision only variants prevents the reintegration of the transposon in the genome (Li et al., 2013).

The previous reported workflow faces challenges when trying to edit genomic regions that present a high density of repetitive elements since it increases the chances of having homologous recombination in other genomic regions (Saito and Adachi, 2016). In our previous work we modeled the influence of the different types of repetitive elements and showed that the presence of repetitive elements of the family Short Interspersed Nuclear Elements (SINE) in the homology arms present higher frequency of random integration (Arias-Fuenzalida et al., 2017). Our model matched the observations previously reported (Ishii et al., 2014). Due to the high content of repetitive elements in mammalian genomes mainly coming from transposable elements integrated during evolution (de Koning et al., 2011), it is in some cases difficult to design homology arms of an appropriate size that are free of repetitive elements. The donor plasmids in the presented design carry negative selection modules for the identification and exclusion of random integration events. As random integration events could occur excluding the negative selection module, we explore adapting the genome engineering pipeline to perform fluorescent-microscopy guided colony picking. The clones selected and picked carry the EGFP+dTomato+ BFP- fluorescent phenotype. Colonies will have to be PCR screened for detecting the presence of the backbone of the donor construct before continuing with the rest of the workflow. Here we present a detailed protocol for this process.

STEPWISE PROTOCOL

Please read the entire process before starting since elements listed in the reagents table (**Supplementary Table 1**) are only those specific for this pipeline (summarized in **Figure 1**), and common cell culture reagents are not described in detail. Please also use

as a reference the **Supplementary Table 2** containing the list of primers used in the protocol.

CONCRETE EXAMPLE

In order to easily understand the pipeline of work here presented, we provide an example of a particular Single Nucleotide Polymorphism (SNP) we have edited. The SNP rs45539432¹ is a transition (c.1366C > T, NM_032409.2) in the PINK1 gene that generates a premature stop codon (p.Gln456Ter, NP_115785.1). This SNP has been linked to early-onset Parkinson's disease (Hedrich et al., 2006). In this case, the mutations to be corrected are homozygous, hence the design of the homology arms for both donors (carrying EGFP or dTomato in the PSM) is identical. For doing biallelic targeting of heterozygous modifications one of the donors (either the EGFP or dTomato one) should not have the SNP, hence a different homology arm would have to be generated.

IN SILICO WORK

One of the first steps in designing your plasmids for gene editing is the identification of the region of interest to be edited (**Figure 1A**). It is important to assess if the gene to be modified presents splicing variants that might show unexpected effects of the modification when performing downstream assays for phenotyping. The in silico work is required for designing the donors, the sgRNA and the oligonucleotides used to generate the constructs or to screen the editing process (**Figure 1A**).

Designing of the Donors

For designing the donors, the identification of the Base to Edit (BTE) and the context of the genomic region allows the user to screen for the presence of repetitive elements that could define the boundaries of the homology arms (**Figure 2A**). Considering a broader genomic region around the desired site for introducing the mutation helps the user to create the entire pipeline for screening the editing process. We recommend the usage of a sequence editor software (SES) such as ApE² or SnapGene³ for working with the sequences over all the steps of the design. The steps required for designing the donors can be summarized in: identification of the region to be edited, evaluation of the presence of repetitive elements, identification of a TTAA site and design of primers for generating the arms.

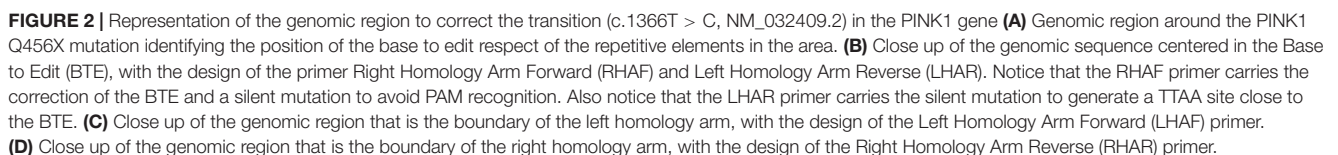
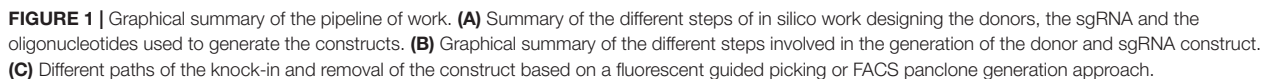
Identification of the Region to Be Edited

1. These recommendations and steps are based on the introduction of SNPs in coding regions of the genome. However this technique could also be applied for the introduction of SNPs in non-coding regions.
2. Identify in the genome the position of the BTE (**Figure 2A**). In case no detailed information is available about the SNP desired to introduce, and only the amino acid change in the

¹https://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=45539432

²<http://jorgensen.biology.utah.edu/wayned/ap/>

³<http://www.snapgene.com/>



protein sequence is known, we recommend following the steps mentioned on **Box 1**.

3. Centered on BTE, select a genomic region that expands 3 kbp upstream and downstream, and transfer this information into a SES.

Evaluation of the Presence of Repetitive Elements

1. Upload your sequence in the Repeat Masker tool⁴ for identifying the presence of repetitive elements in the selected region. Repetitive elements should be avoided as much as possible in designing the homology arms.
2. Upload the information of repetitive elements into your region of interest using a SES.
3. Label those regions in your sequence of interest with identifiers of the different types of repetitive elements (**Figure 2A**).

Identification of a TTAA Region in the Vicinity of the BTE

1. Search in the vicinity of the BTE for the presence of a TTAA site, preferably in an exonic region. Ideally the distance between the TTAA site should and the BTE should not be more than 50 bp, making it easier to introduce the SNP with the primers used to amplify the arm in one step. If non TTAA can be found in the vicinity, check if any permutation of TTAA can be achieved by introducing a silent mutation (**Figure 2B**).
2. Detect the presence of potential silent mutations that can generate the formation of a TTAA site. In order to speed up the process consider using restriction site detection software⁵ and check for restriction enzymes that cut in TTAA site (like MSE I). This point mutation would need to be introduced in the homology arms used in the donor. Control that the introduction of this silent mutation is not drastically changing the codon frequency usage⁶.
3. Select an area of 1 kb upstream and downstream of the TTAA to define the homology arms (Note: In case that the surrounding genomic sequence has presence of repetitive elements reduce the size of the arms to avoid the inclusion of them. Based on the linear optimization calculated in our previous paper a set of guidelines is described in **Box 2**) (**Figure 2B**).

Designing of the sgRNAs

Selection of the Guides

1. Upload a sequence of 100 bp upstream and downstream of your BTE to an online tool such as the one provided by the Broad Institute which ranks the sgRNAs based on the on and off targets⁷.
2. Select the best five sgRNA. Selection of the guides should be ideally performed in the exonic area of the gene since silent mutations in the PAM region will need to be introduced.

⁴<http://repeatmasker.org/>

⁵<http://resfinder.appspot.com/>

⁶<https://www.genscript.com/tools/codon-frequency-table>

⁷<https://portals.broadinstitute.org/gpp/public/analysis-tools/ssgRNA-design>

BOX 1 | Guidelines for identifying the base to edit (BTE) in the genomic sequence.

- Use the Nucleotide search tool available in Pubmed to identify the region desired (<https://www.ncbi.nlm.nih.gov/nucleotide/>) Introduce your gene name. Select the RefSeq option and download the NM_file.
- Open the file in the software and select the coding region (CDS).
- Then translate your selection and find the amino acid that you want to change and the surrounding sequence of the codon you would like to change.
- Translate that region and identify the position of the amino acid you want to change with the SNP.
- We recommend using the SNP website from Pubmed for evaluating the role of the Base to Edit (BTE). If it is an annotated SNP this website would also provide the roles in the different splicing variants of the mRNA.

BOX 2 | Rough guidelines for reducing the size of the homology arms to avoid the repetitive elements (RepEI).

- Do not reduce the size of the homology arm to less than 200 bp even if repetitive elements are present.
- Reduce the size of the arm if repetitive elements of the SINE family are present.
- If there are elements of repeated bases present within the closes 500 bp to the BTE, do not reduce the size of the arms.

Moreover, the distance between the double strand break (DSB) and the BTE should not be more than 25 bp. In the case a microhomology-mediated end-joining (MMEJ) repair occurs, the integration of the selection cassette can happen without the SNP, if this one is designed outside this limit (Nakade et al., 2014).

3. As previously reported, it is recommended to pick the guides that hit the reading DNA strand of the gene. It is reported to increase efficiency since the RNA polymerase dislodges the bound Cas9 allowing the access to the cell's repair mechanism (Clarke et al., 2018).

Oligonucleotides Design

Primer Design to Obtain the Arms

1. Use a primer generator tool, such as the one available in the NCBI platform Primer-BLAST⁸.
2. Define sequencing primers (SEQPRA) that can extract 2 kbp upstream and downstream of the BTE. These primers will be used to extract the genomic region to be cloned in a TOPO vector (see section "Preparation of the homology arms template").
3. Based on the analysis of the presence of repetitive elements, define primers that can ideally extract 1 kb upstream and downstream of the TTAA. In case of repetitive elements, proceed as mentioned in the "Identification of a TTAA region in the vicinity of the BTE" section. The primers to be generated will define the boundaries of your homology arms, namely Left Homology Arm Forward (LHAF), Left

⁸<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>

Homology Arm Reverse (LHAR), Right Homology Arm Forward (RHAF), and Right Homology Arm Reverse (RHAR) (**Figures 2B–D**). Each primer possess a homology region to the genomic DNA and a homology to do donor plasmid for assembly. For the homology region in the genomic DNA, consider an amount of bases in the border of your homology arm that generates an oligo with a T_m of 60°C (this can be assessed in the SES) (**Figures 2B–D**). For the homology region in the donor, overhangs (of 20 bp in length) will need to be added to the designed oligos matching the splitted scaffold after digestion with the restriction enzyme Hpa I (**Figures 3A–C**). These primers will be used to perform Gibson's assembly (Gibson, 2011) of the homology arms into the donor scaffold (**Figure 3D**). The assembly of the homology arms is performed in the TTAA splitting point of the ITR of the donor (**Figure 3B**).

- At this point of the primer design consider all the base pair changes that need to be introduced in the donor: BTE, silent mutation of the PAMs, and silent mutation to generate a TTAA (if needed) (**Figure 2B**). In case that the generation of a silent mutation is needed to produce a TTAA site, we recommend to introduce it with either the LHAR or RHAF primers, depending which is the permutation available: if NTAA or TNAA should be permuted, changes must be made in the LHAR; and if TTNA or TTAN should be changed, the modifications must be in the RHAF. If close to the extremes of the homology arms, consider using these primers for introducing the point mutations (**Figure 2B**).
- Four primers will need to be generated (one per extreme of each homology arm) (**Figures 2, 3E–H**):

Left Homology Arm Forward (LHAF)

AAGCTTGGATCCCCTAGGTT (+ sequence into your left homology arm) 3' end

Left Homology Arm Reverse (LHAR) at the splitting point of TTAA

CAGACTATCTTTCTAGGGTT (+ sequence into your left homology arm TTAA site) 3' end

Right Homology Arm Forward (RHAF) at the splitting point of TTAA

ATGATTATCTTTCTAGGGTT (+ sequence into your right homology arm TTAA site) 3' end

Right Homology Arm Reverse (RHAR)

GCATACGCGTATACTAGGTT (+ sequence into your right homology arm) 3' end

Primer Design to Introduce the SNPs

The donor would need to have not only the modification of the BTE that shall be introduced but also the modification of the PAM and a silent mutation to generate a TTAA site (if needed). We recommend to introduce a silent mutation for the PAM of at least 2 different sgRNAs. We recommend to clone the extracted amplicon with the surrounding genomic region to the BTE in a TOPO vector (Zero Blunt™ TOPO™ PCR Cloning Kit, Thermo Fisher Scientific) for doing the steps for inserting the SNPs (see section "Preparation of the homology arms template"). If the SNP are close to the extremes of the arms, they can be introduced with the primers for generating the homology arms (see the previous

section). If not, insertion of the SNP could be performed by Site Directed Mutagenesis (SDM, e.g., Q5® Site-Directed Mutagenesis Kit, NEB). Primers would have to be designed to introduce these mutations. Alternatively, this process can be outsourced to a de novo DNA sequence synthesis company (e.g., GeneArt® Gene Synthesis, Thermo Fischer Scientific). In that case, the best option could be to synthesize the left homology arm and the right homology arm independently to then be ligated with the donor scaffold.

Oligonucleotide Design for sgRNA

- We recommend using the plasmid BB-CBh-hSpCas9 (also known as pX330; Addgene #42230) generated by the Zhang lab. The protocol used for cloning the sgRNA into the pX330 plasmid has been explained in detailed previously (Ran et al., 2013). Briefly, sgRNA selected from the in silico tool has to be modified in order to be incorporated in the plasmid and its complement has to be also designed (**Figure 1A**).
- Selection of only the 20 bp sequence (non-inclusion of the PAM in sgRNA).
5'-NNNNNNNNNNNNNNNNNNNNNNNGG-3'
3'-NNNNNNNNNNNNNNNNNNNNNNNCC-5'
- Addition of sticky ends and 5' G (in case your selected sgRNA starts with a G it is not necessary to add this base).
5'-CACCGNNNNNNNNNNNNNNNNNNNNNNNN-3'
3'-CNNNNNNNNNNNNNNNNNNNNNNNNNCAAA-5'

Oligonucleotide Design for Validating the Knock-In (VKI)

In order to validate the knock-in, a set of primers has to be designed to verify the right and left junctions between the PSM and the genomic DNA. The left junction forward primer (VKI Primer 1) and the right junction reverse primer (VKI Primer 4) depend on the genomic region of interest (**Figure 1A**). We recommend designing oligonucleotides in the genomic region at a distance of 500 bp from the junction between the homology arms and genomic DNA for VKI Primer 1 and 4. Alternatively, the same SEQPRA designed in section "Primer design to obtain arms" could be used. For the left junction reverse primer (VKI Primer 2) and the right junction forward primer (VKI Primer 3) we recommend using:

VKI Primer 2

5'-AGATGTCCTAAATGCACAGCG-3'

VKI Primer 3

5'-CGTCAATTTTACGCATGATTATCTTTAAC-3'

Plus, a set primers to obtain an amplicon expanding from the PSM to the backbone of the plasmid to identify random integration events that might have left out the BFP during the integration process. For detecting the presence of the backbone of the plasmid, we recommend designing the left backbone forward primer (VKI Primer 5) and the right backbone reverse primer (VKI Primer 6) (**Figure 1A**):

VKI Primer 5

5'-GCTGCCTATCAGAAGGTGGTG-3'

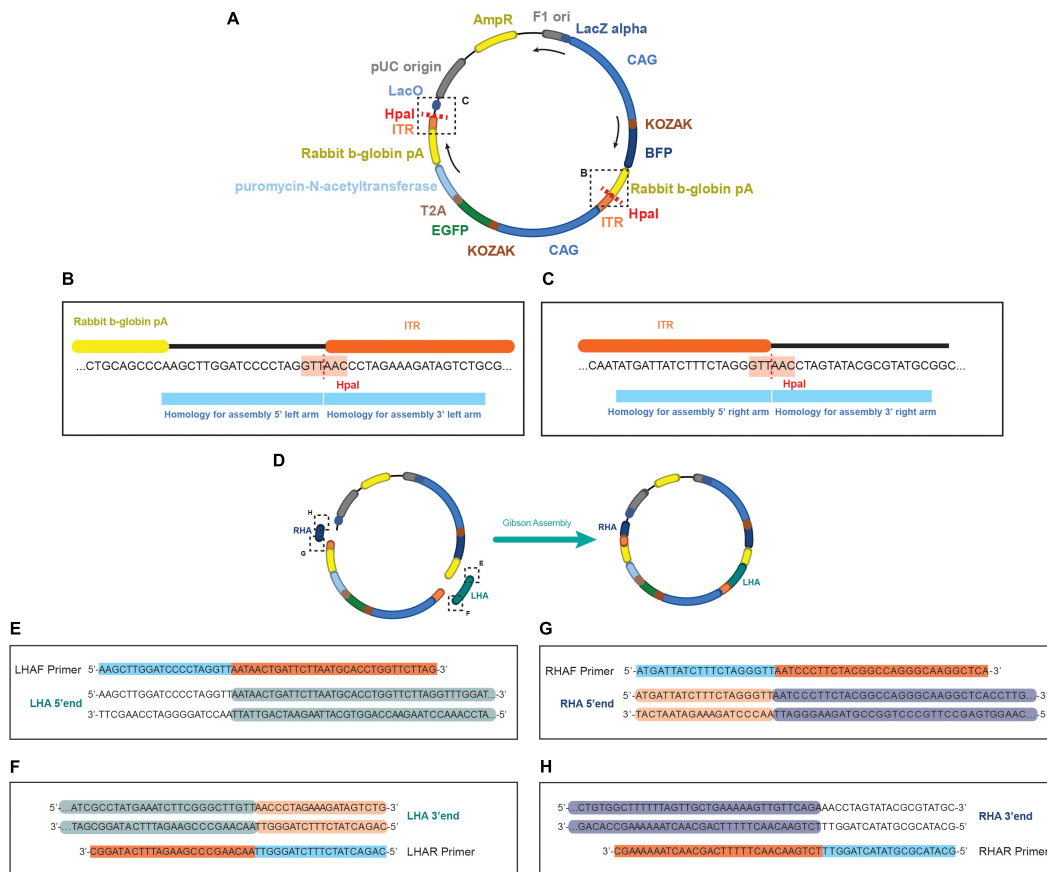


FIGURE 3 | Schematics of the EGFP donor plasmid and the homology in the primers for performing the assembly. **(A)** Representation of the pDONOR-tagBFP-PSM-EGFP with the restriction sites of HpaI. **(B)** Close up of the HpaI region where the left homology arm will be assembled. **(C)** Close up of the HpaI region where the right homology arm will be assembled. **(D)** Representation of the donor after assembly. On each primer from **(E–H)** it is represented in light blue the homology to the donor and in orange the homology to the genomic region of the example. Every double stranded DNA section in **E–H** represents the sequence of the homology arms to be assembled, showing in green **(E,F)** the genomic sequence incorporated in the left arm, and in dark blue **(G,H)** the genomic sequence incorporated in the right arm. In light orange **(F,G)** the homology in the arms to the ITR sequence of the donor, and unlabeled **(E,H)** the backbone of the donor plasmid.

VKI Primer 6

5'-GCAGCCACTGGTAACAGGAT-3'

Oligonucleotide Design for Final Sequencing

An oligonucleotide (SEQPR) at a distance of around 100 bp from the BTE has to be designed for doing the final sequencing of the edited clone.

BENCH WORK

The steps performed in this section are summarized in **Figure 1B**

Generation of the Guides

This protocol established by the Zhang lab has been explained in detail previously (Ran et al., 2013). Here we summarize the steps needed for the generation of the sgRNAs.

Preparation of px330 Scaffold

1. Digest the vector px330 with Bpil (FastDigest, Thermo Fisher Scientific) for 3 h at 37°C for complete digestion. Prepare a maximum of 1 µg of DNA per single reaction:

Solution	Volume (µl)
px330	Maximum 1 µg per reaction
Bpil FD	1
Fast Digest buffer (10x)	2.5
H2O	q.s.
Total	25

2. Column purify the digestion product (e.g., with QIAquick PCR Purification Kit, Qiagen) assume a 50% lost in column purification and elute in a volume of nuclease free water.
3. Determine purified plasmid concentration (e.g., by NanoDrop Spectrophotometer).

Annealing of sgRNA Oligonucleotides

1. Prepare annealing buffer (AB) using the T4 DNA ligase kit (NEB) at a ratio 1:7 (T4 ligation buffer: H₂O). The mix can be stored at -20°C .
2. Example 100 μL of T4 ligation buffer +700 μL of nuclease free water.
3. Prepare annealing reaction.

sgRNA oligo forward (100 μM)	10 μL
sgRNA oligo reverse (100 μM)	10 μL
Annealing buffer	80 μL
Total	100 μL

4. Annealing. Use a ramp protocol for annealing in thermocycler.

	Temperature (C)	time	delta T (C)
1	95	5 minutes	na
2 (loop 14x)	95	1 min	-5°C
3	25	HOLD	Na

Ligation of Annealed Oligonucleotides and px330

1. Set the ligation reaction as bellow.

Solution	Volume (μL)
Digested px330	Total 112 ng
annealed oligos	2
T4 ligase	1
T4 ligase buffer	2
H ₂ O	q.s.
Total	20

2. Incubate at RT for 1 h.
3. Transform One ShotTM TOP10 Chemically Competent *E. coli* (Thermo Fisher Scientific) with 10 μL of ligation following manufacturer's protocol.
4. Grow in 1 LB agar plate (supplemented with ampicillin) per sample using 450 μL of the culture tube.

Picking of Colonies and Sequencing

1. From the transformed plate pick 5 colonies per sample and grow in a volume of 2 ml LB supplemented with ampicillin overnight in shaking conditions. The plates can be stores at 4°C awaiting screening results.
2. The following day from the same tube per sample make a glycerol stock using 500 μL of culture plus 500 μL of 50% glycerol, and purified the plasmid with a miniprep kit (e.g., QIAprep Spin Miniprep Kit, Qiagen) using 1 mL of culture according to manufacturer instructions but eluting in 50 μL of nuclease free water.
3. Send 20 μL of the miniprep sample to sequence with the following primer:

5'-GAGGGCCTATTTCCCATGATTCC-3'

Generation of the Donor

Preparation of the Donor Scaffold DNA

1. Prepare the donor scaffold DNA with a maxiprep (e.g., HiSpeed Plasmid Maxi Kit, Qiagen).
2. Determine plasmid concentration (e.g., by NanoDrop Spectrophotometer).
3. Digest the DNA with HpaI enzyme (NEB). Bulk preparations are recommended.

Reagent	Volume for 1 reaction
DNA	Volume of 1 μg (per each reaction)
HpaI (5k unit/ml)	1 μl
NEB buffer 4 (10 \times)	2 μl
H ₂ O	Bring up to 20 μl
Total	20 μl

4. Incubate for 2 h at 37°C in incubator.
5. Purify the digestion product using a column purification kit (e.g., with QIAquick PCR Purification Kit, Qiagen).

Preparation of the Homology Arms Template

If the homology arm generation was outsourced to a de novo DNA sequence synthesis company skip this section.

1. Define the parental line you want to edit and purify genomic DNA. Karyotype the line before you start with the editing process, and if possible sequence specific regions that might be interesting for your specific disease modeling to discard that the parental has unexpected alterations from the beginning.
2. Using as template the parental line you want to edit, amplify the genomic DNA with the SEQPRA primers designed in step 2 of "Primer design to obtain the arms." For this step use a high fidelity polymerase (e.g., PrimeSTAR[®] GXL DNA Polymerase, Takara).

Reagent	Volume for 1 reaction
Genomic DNA	Volume of 100–200 ng
GXL polymerase (10U/ μl)	1 μl
GXL polymerase buffer (5x)	10 μl
dNTPs (2.5mM)	4 μl
Primer forward (10 μM)	2 μl
Primer reverse (10 μM)	2 μl
H ₂ O	to 50 μl
Total	50 μl

3. Define the amplification PCR protocol depending on enzyme extension temperature, enzyme synthesis speed and template length. Normally 30 cycles of amplification is enough.
4. Verify the amplified product by agarose gel electrophoresis. Use only 5 μl .

5. Purify the rest of the PCR product using a column purification kit (e.g., with QIAquick PCR Purification Kit, Qiagen).
6. Clone the PCR product of one reaction into a TOPO vector (Zero Blunt™ TOPO™ PCR Cloning Kit, Thermo Fisher Scientific) following manufacturer's instructions. For choosing the right TOPO consider the terminal ending activity of the polymerase. Most high fidelity polymerases generate blunt end products.
7. After cloning, the TOPO-gDNA vector should be restriction mapped and sequenced. This allows to define if there are SNP variants in your region of interest. Special consideration should be placed in Cas9 binding sites and TTAA region.

Introduction of Mutations

As explained previously in the section “Primer design to introduce the SNPs,” SNPs to be introduced in the region of interest should be performed on the TOPO vector generated in the previous section using SDM or introduced with the primers designed to obtain the arms (see step 4 of “Primer design to obtain the arms”).

Preparation of the Homology Arms for Assembly

The assembly of the donor DNA is performed using Gibson assembly (Gibson, 2011). Homology between the homology arms and the scaffold is required. Use the primers designed in step 5 of section “Primer design to obtain the arms” to incorporate the arms in the HpaI splitting sites of the scaffold donor.

1. Amplify, verify, and purify the homology arms from the mutated TOPO or the synthetic DNA from a company following steps 2–5 from the previous section but using the pair of primers LHAF and LHAR (to obtain the left arm), and the pair of primers RHAF and RHAR (to obtain the right arm) in two separate PCR reactions.
2. Assemble the left homology arm, the right homology arm and the scaffold DNA prepared in step 5 of “Preparation of the donor scaffold DNA” section using Gibson Assembly® Master Mix (NEB) following manufacturer's protocol. We recommend the usage of online calculators for establishing the stoichiometry of the ligation pieces⁹ or designing the ligation with assembly tools¹⁰. Having the right proportion of the different elements during the assembly drastically increases the efficiency.
3. Transform NEB® 10-beta Competent E. coli (NEB) with 2 µL of the assembly following manufacturer's protocol.
4. Grow in 1 LB agar plate (supplemented with ampicillin) per sample using 450 µL of the culture tube.
5. Follow the steps 1 and 2 in section “Picking of colonies and sequencing” for picking and plasmid glycerol stock. In this case pick 10 colonies per donor.
6. Screen by restriction mapping and sequence those that present an expected map.

⁹<http://nebiocalculator.neb.com/#/ligation>

¹⁰<http://nebuilder.neb.com/>

7. Perform a maxiprep (e.g., HiSpeed Plasmid Maxi Kit, Qiagen) of sequence confirmed donors.
8. Measure plasmid concentration and take it to a 500 ng/ul working concentration.

Preparation of Transposase mRNA

Removal of the selection cassette after can be performed by the transfection of mRNA. This mRNA can be either in vitro generated in the lab using the template described in Yusa et al., 2011; Yusa, 2013 (construct pCMV-HAhyPBse); or Li et al., 2013; or it can be commercially acquired¹¹.

1. PCR amplify the coding sequence of codon optimized hyperactive transposase piggybac from *Trichoplusia ni* (Yusa et al., 2011) or the excision-only mutant (Li et al., 2013) to incorporate the T7 promoter:
T7-transposase_F
5'-GAAATTAATACGACTCACTATAGGGCCGCCACCA
TGGGCAGCAGCCTGGAC-3'
T7-transposase_R
5'-GGCAAACAACAGATGGCTGG-3'
2. Column purify PCR product.
3. Use the PCR product as a template for in vitro transcription with a mMESSAGE mMACHINE™ T7 Transcription Kit (Thermo Fisher Scientific) following manufacturer's protocol.
4. Polyadenylate the transcript with a Poly(A) tailing kit (Thermo Fisher Scientific) and purified with a MEGAclean™ Transcription Clean-Up Kit (Thermo Fisher Scientific).

In vitro Testing of sgRNA Efficiency

In silico efficiency of the guides can be further tested with in vitro assays for EGFP reconstitution as described in Mashiko et al. (2013); Mashiko et al. (2014).

CELL CULTURE WORK

Edition of cells under this protocol is normally performed in hiPSCs cultured in Matrigel (Corning) coated plates with daily changes of Essential 8 media (Thermo Fisher Scientific) supplemented with 1% Penicillin/Streptomycin. Cells are normally passaged and handled as single cells (Figure 1C). This is performed by using Accutase (Thermo Fisher Scientific) and supplementing the Essential 8 for 24 h after passaging with ROCK inhibitor to prevent apoptosis (Y-27632, Merck Millipore).

Nucleofection of Parental Cells and Selection

Nucleofection is performed using a 4D-Nucleofector™ X Unit (Lonza) and the P3 Primary Cell 4D-Nucleofector™ X Kit L (Lonza). Expansion of the cells pre-nucleofection can be performed in flasks or 10 cm dishes. Seeding of the cells after nucleofection should be done in a Matrigel pre-coated 1 well

¹¹<https://www.transposagenbio.com>

plate Nunc™ OmniTray™ (Thermo Fisher Scientific) to allow the possibility of doing fluorescence guided picking. Ideally, do 2 different batches of electroporation each with a different sgRNA-Cas, and make five electroporations per sgRNA-Cas.

1. Expand parental cells to have a million cells per electroporation. Cells have to be freshly plated (no more than 4 days after the previous passage).
2. Detach cells using Accutase, neutralize using DMEM/F12 (Thermo Fisher Scientific), centrifuge ($200 \times g$ 3 min), resuspend in 5 ml of media, and count.
3. Transfer 1 million cells into a 1.5 ml Eppendorf tube and centrifuge ($200 \times g$, 3 min). Resuspend with 100 μ l of supplemented nucleofector solution (see manufacturer's protocol) and transfer to electroporation cuvette.
4. Electroporate using the "human stem cells h9" protocol that performs the CB150 pulse in the 4D-nucleofector (Lonza). Use 1.5 μ g of each donor and 2.5 μ g of sgRNA-Cas per electroporation.
5. After pulse, add 1 ml of Essential 8 media supplemented with ROCK inhibitor (10 μ M) to the nucleofector cuvette and then transfer the content of one cuvette to one Matrigel pre-coated 1 well plate.
6. Culture on media supplemented with ROCK inhibitor for 24 h.
7. Culture until small to medium size colonies have grouped and supplement media with Puromycin (0.25 μ g/ml). This concentration can be adjusted in the next 3 days depending on the survival increasing it to 0.5 μ g/ml.
8. Two days after antibiotic treatment small colonies should emerge and continue to populate the plate. Selection with Puromycin should be maintained through the next stages of culturing until treatment with transposase.

FACS or Fluorescence Guided Picking

As explained in the introduction, the presence of repetitive elements in the homology arms increases the chances of having random integration events. Even though the usage of a BFP in the backbone of the plasmid is used to detect these events, there is still a chance of having integration in an unspecific site that does not include the BFP. For this reason we recommend if the presence of repetitive elements in the homology arm cannot be avoided during the design phase to perform fluorescence guided picking of EGFP+/Dtomato+/BFP- colonies rather than generating a panclone through FACS (Figure 1C). It is important to notice that one of the advantages of using fluorescence reporters with strong promoters as in these donors is that some cells not harboring the PSM (and hence not having resistance to Puromycin) can still survive the Puromycin treatment if in the context of a colony containing resistant cells. These WT cells could be easily carried over, reducing the specificity of a drug based approaches (Figure 4A).

Fluorescence Guided Picking

1. Colonies having a size of around 0.25 mm are suitable for fluorescence guided picking. This time will depend on the dividing rate of the iPSCs you are working with.

2. Approximately one week after selection screen the plate for fluorescent colonies. Colonies having an EGFP+/Dtomato+/BFP- fluorescence should be picked and plated in one well of a Matrigel pre-coated 96 well plate (Figure 4B). Picking should be performed using a pipette tip in an angle to gently detach the colony from the plate. It is recommended to have Essential 8 media supplemented with ROCK inhibitor (10 μ M) in the source and receiver plate. The media with ROCK inhibitor should be added to the source plate one hour prior picking. It is recommended to pick all the colonies present in the plate (on average 1–5 colonies/plate are expected). We recommend picking all colonies resending the EGFP+/Dtomato+/BFP- pattern. We discourage picking colonies with mixed fluorescent patterns, where some cells might be EGFP+ or Dtomato+ since the validation by PCR (see step 5) might give conflicting results. It is fine to pick EGFP+/Dtomato+/BFP- that present some WT cells since they can be removed in the first sort (see next section).
3. Depending on the initial size of the colony picked, an average waiting time of 3 days is enough to obtain sufficient material for expanding this one well of a 96 well into 4 wells. Two of these wells are going to be used for genomic DNA extraction.
4. Genomic DNA extraction is done using QuickExtract™ DNA Extraction Solution (Epicenter) and following manufacturer's indications.
5. Perform a PCR for detecting events of random integration without BFP using the plasmid designed in section "Oligonucleotide design for validating knock-in." Both pairs of primers should be used for detecting random integration by assessing the presence of elements of the donor backbone on either side of the homology arms. For assessing this use primers pair VKI Primer 5 and VKI Primer 2, and primers pair VKI Primer 3 and VKI Primer 6. At this point it would be also appropriate to control the junctions between the homology arms and the genomic region. For assessing this, use the primers pair VKI Primer 1 and VKI Primer 2 to detect the left junction; and primers pair VKI Primer 3 and VKI Primer 4 to detect the right junction.
6. Clones presenting backbone of the plasmid integrated in the genome or lacking one of the junctions should be discarded.
7. Expansion of the clones having the correct profile needs to be done before discarding the presence of unedited cells by FACS.
8. For details of the sorting procedure please see the next section

FACS

The starting point for the cell sorting can be from the puromycin selected culture after one/two weeks of nucleofection or from the isolated clone obtained from fluorescence guided picking. In case of the former, it is expected to see around 2% of cells presenting an EGFP+/Dtomato+/BFP- pattern (Arias-Fuenzalida et al., 2017). In the case of the latter, positive cells can range between 50

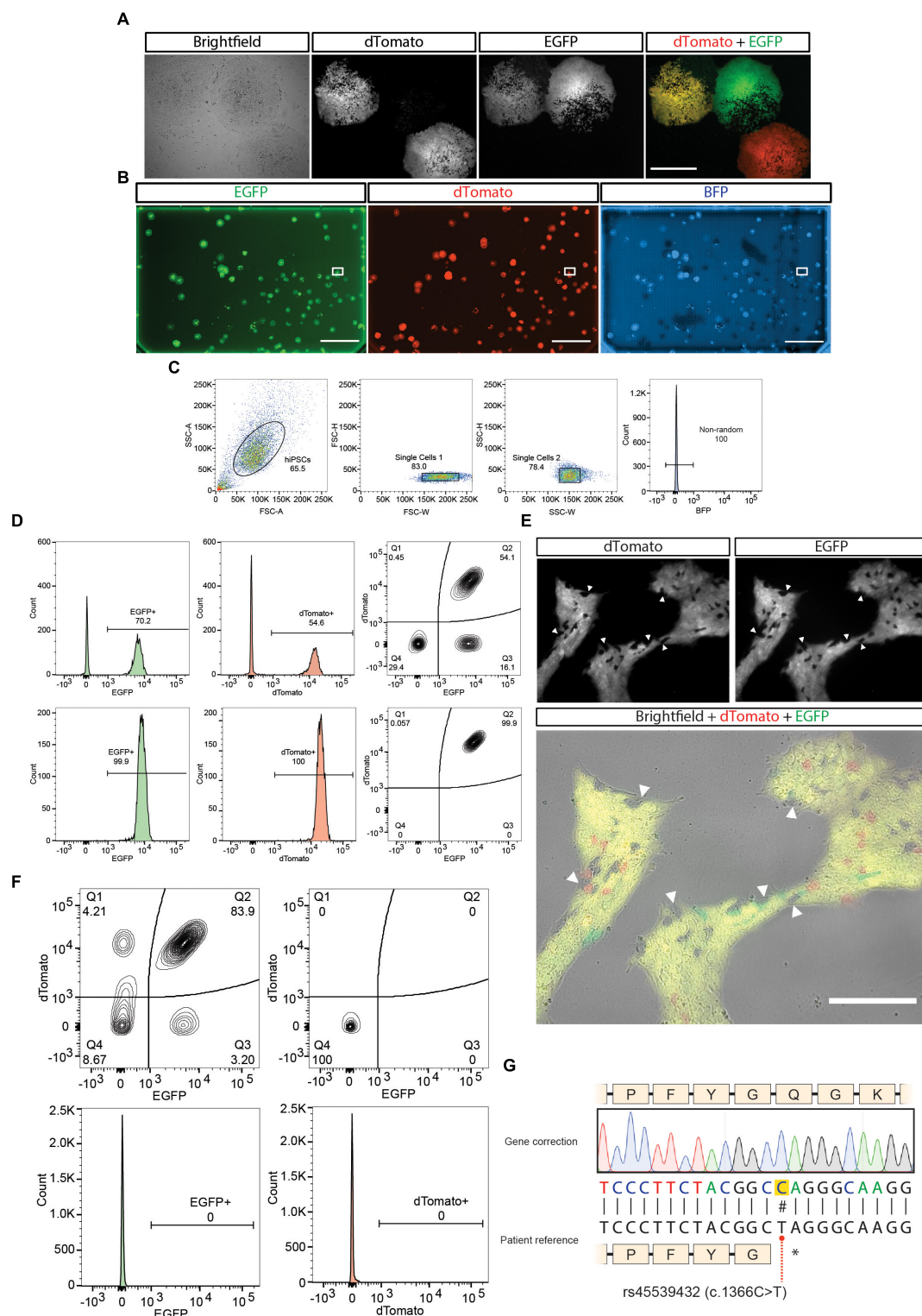


FIGURE 4 | Cell culture work of the gene correction example. **(A)** Representative images of hiPSC colonies expressing the different possible outcomes of gene modification. Scale bar = 500 μ m. **(B)** A one-well plate screened for detecting correct biallelic targeting (dTomato+/EGFP+/BFP-). Bounding box shows selected isolated region. Scale bar = 2 cm. **(C)** Single cell isolation gating strategy performed, plus exclusion of BFP+ cells from an expanded culture derived from the selected region in **(B)**. **(D)** First (upper panel) and second (lower panel) purity sort for purifying dTomato+/EGFP+/BFP- cells. **(E)** Removal of the positive selection module after treatment with transposase. **(F)** First sort after transposase induction (top left panel). Last purity sort before sequencing ($n = 3$) (top right panel and bottom panels). **(G)** Sequencing results of the gene correction of the patient line.

and 100% depending on the number of WT cells that could be forming the colony (**Figures 4C,D**). If using a BD FACS Aria II for performing the cell sorting, we recommend using an 85 μm nozzle, the 2.0 neutral density filter and the refrigeration system set at 4°C.

1. Treat for 5 min the iPSCs culture with Accutase to generate a single cell suspension. Some iPSCs lines might need different times for detachment. In that case it is better to increase the time of incubation with Accutase rather than exerting mechanical force by pipetting.
2. Accutase activity should be stop and the sample centrifuged as explained in section “Nucleofection of parental cells and selection”
3. Cells are resuspend in sorting solution (2% bovine serum albumin, 1% Penicillin/streptomycin in PBS) supplemented with ROCK inhibitor (30 μM). Dilution of the cells can be afterwards adjusted based on the rate of events observed.
4. Filter the cell suspension using a 20 μm pre-separation filter (Miltenyi) to remove doublets.
5. Cells should be sorted into sorting solution supplemented with ROCK inhibitor (30 μM). We recommend performing a first round of sorting using a “yield” mask to maximize the recovery of EGFP+/Dtomato+/BFP- fluorescent cells. In successive runs, change instead to a “4-way purity mask” (**Figure 4D**).
6. To determine the last run of sorting pre-transposase transfection we recommend taking a sample of the cell suspension and staining it with a blue dead cell dye (e.g., SYTOXTM Blue Dead Cell Stain, Thermo Fisher Scientific) to discard the presence of large debris that might be interpreted as non-edited cells still present in the culture (**Figure 4D**).
7. Only after the culture presents ~100% of EGFP+/Dtomato+/BFP- fluorescent cells it is possible to continue to the transposase transfection. After having a dead cell stained sample which presents ~100% of EGFP+/Dtomato+/BFP- you can continue to the next section.

Removal of Positive Selection Module With Transposase

1. The selected polyclone or clone can be plated on two wells of a 6 well plate at a density of have a million cells per well one day before transfection.
2. Transfect the transposase mRNA generated in section “Preparation of transposase mRNA” with StemfectTM RNA transfection kit (Stemgent) according to manufacturer's protocol.
3. The next day repeat the transfection.
4. Allow the culture to recover and proliferate. Once the culture is ready to passage, expand until having 2 10 cm dishes at 80% confluency for sorting (**Figure 4E**). Notice that either the EGFP or the dTomato PSM can be removed first.

FACS of Cells That Underwent PSM Removal

Follow the same instructions performed in section “FACS” to obtain a single cell suspension ready for sorting. We recommend performing a first round of sorting using a “yield” mask to maximize the recovery of EGFP-/Dtomato-/BFP- fluorescent cells. An efficiency between 5 and 15% should be expected of the removal of cells (**Figure 4F**). In successive runs, change instead to a “4-way purity mask.” Only after the culture presents 100% of EGFP-/Dtomato-/BFP- fluorescent cells, a final sequencing should be performed (**Figure 4F**).

Confirming by DNA Sequencing

1. Use the primers designed for extracting the amplicon region to clone in the TOPO vector named SEQPRA to generate an amplicon expanding from genomic regions outside the homology arms.
2. Sequence the amplicon with the primer SEQPR designed in the “Oligonucleotide design for final sequencing” section and confirm the presence of the change in the BTE (**Figure 4G**) as well as the silent mutations for the PAMs and TTAA generation.

CONCLUSION

Gene editing technologies, specifically CRISPR/Cas9 gene editing, are starting to revolutionize biological sciences to a similar extent as the invention of PCR or hiPSCs did (Ledford, 2015). The real and potential applications of gene editing range from increasing crops and livestock yields to disease diagnostics and gene drives (Doudna and Barrangou, 2016). In the context of medicine, one of the applications of this technique is disease modeling by targeted modifications of the genome. Here we covered a procedure for the generation of isogenic lines for doing disease modeling, which allows to evaluate the influence of a specific point mutation or the effect of the genetic background of the patient in the onset and progression of a disease (Bolognin et al., 2018). Due to the high number of clones that needed to be screened to obtain a positive one (Paquet et al., 2016), alternatives techniques were needed by the research community.

We provided a detailed guideline for our previous work (Arias-Fuenzalida et al., 2017), adding the alternative path of doing fluorescence guided picking when the design of the homology arms cannot avoid the inclusion of repetitive elements. We consider that the strength of our protocol resides in the certainty obtained after reaching the different milestones of the pipeline, and the simultaneous biallelic targeting to generate isogenic lines for disease modeling.

AUTHOR CONTRIBUTIONS

JJ, XQ, and JS contributed conception and design of the protocol. JJ wrote the first draft of the manuscript. All authors

contributed to manuscript revision, and read and approved the submitted version.

FUNDING

This project was supported by the LCSB Pluripotent Stem Cell Core Facility. JJ and XQ were supported by fellowships from the FNR (AFR, Aides à la Formation-Recherche). JJ was supported by a Pelican award from the Fondation du Pelican de Mie et Pierre Hippert-Faber. This is an EU Joint Programme-Neurodegenerative Disease Research (JPND) project (INTER/JPND/14/02; INTER/JPND/15/11092422). Acquisition of flow cytometry data was supported by the flow cytometry core of the LCSB bio-imaging platform. Further support comes from the SysMedPD project, which has received

funding from the European Union's Horizon 2020 Research and Innovation Program under grant agreement no. 668738.

ACKNOWLEDGMENTS

We would like to thank Prof. F. Zhang from the McGovern Institute for Brain Research for providing the Cas9 vector.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00190/full#supplementary-material>

REFERENCES

- Arias-Fuenzalida, J., Jarazo, J., Qing, X., Walter, J., Gomez-Giro, G., Nickels, S., et al. (2017). FACS-assisted CRISPR-Cas9 genome editing facilitates parkinson's disease modeling. *Stem Cell Rep.* 9, 1423–1431. doi: 10.1016/j.stemcr.2017.08.026
- Bolognin, S., Fossépré, M., Qing, X., Jarazo, J., Ščančar, J., Moreno, E. L., et al. (2018). 3D cultures of parkinson's disease-specific dopaminergic neurons for high content phenotyping and drug testing. *Adv. Sci.* 6:1800927. doi: 10.1002/adv.201800927
- Chen, Z., and Cole, P. A. (2015). Synthetic approaches to protein phosphorylation. *Curr. Opin. Chem. Biol.* 28, 115–122. doi: 10.1016/j.cbpa.2015.07.001
- Clarke, R., Heler, R., Macdougall, M. S., Yeo, N., Chavez, A., Regan, M., et al. (2018). Enhanced bacterial immunity and mammalian genome editing via RNA-polymerase-mediated dislodging of Cas9 from double-strand DNA breaks. *Mol. Cell* 71, 42.e8–55.e8. doi: 10.1016/j.molcel.2018.06.005
- de Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A., and Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genetics* 7:e1002384. doi: 10.1371/journal.pgen.1002384
- Doudna, J. A., and Barrangou, R. (2016). Applications of CRISPR technologies in research and beyond. *Nat. Biotechnol.* 34:933. doi: 10.1038/nbt.3659
- Gibson, D. G. (2011). Methods in enzymology. *Methods Enzymol.* 498, 349–361. doi: 10.1016/B978-0-12-385120-8.00015-2
- Hedrich, K., Hagenah, J., Djarmati, A., Hiller, A., Lohnau, T., Lasek, K., et al. (2006). Clinical spectrum of homozygous and heterozygous PINK1 mutations in a large german family with parkinson disease: role of a single hit? *Arch. Neurol.* 63, 833–838. doi: 10.1001/archneur.63.6.833
- Hockemeyer, D., and Jaenisch, R. (2016). Induced pluripotent stem cells meet genome editing. *Cell Stem Cell* 18, 573–586. doi: 10.1016/j.stem.2016.04.013
- Ishii, A., Kurosawa, A., Saito, S., and Adachi, N. (2014). Analysis of the role of homology arms in gene-targeting vectors in human cells. *PLoS One* 9:e108236. doi: 10.1371/journal.pone.0108236
- Jasin, M., and Haber, J. E. (2016). The democratization of gene editing: insights from site-specific cleavage and double-strand break repair. *DNA Repair* 44, 6–16. doi: 10.1016/j.dnarep.2016.05.001
- Jehuda, R. B., Shemer, Y., and Binah, O. (2018). Genome editing in induced pluripotent stem cells using CRISPR/Cas9. *Stem Cell Rev. Rep.* 14, 323–336. doi: 10.1007/s12015-018-9811-3
- Ledford, H. (2015). CRISPR, the disruptor. *Nat. News* 522:20. doi: 10.1038/522020a
- Li, X., Burnight, E. R., Cooney, A. L., Malani, N., Brady, T., Sander, J. D., et al. (2013). piggyBac transposase tools for genome engineering. *Proc. Natl. Acad. Sci.* 110, E2279–E2287. doi: 10.1073/pnas.1305987110
- Mashiko, D., Fujihara, Y., Satouh, Y., Miyata, H., Isotani, A., and Ikawa, M. (2013). Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA. *Sci. Rep.* 3:3355. doi: 10.1038/srep03355
- Mashiko, D., Young, S. A., Muto, M., Kato, H., Nozawa, K., Ogawa, M., et al. (2014). Feasibility for a large scale mouse mutagenesis by injecting CRISPR/Cas plasmid into zygotes. *Dev. Growth Diff.* 56, 122–129. doi: 10.1111/dgd.12113
- Nakade, S., Tsubota, T., Sakane, Y., Kume, S., Sakamoto, N., Obara, M., et al. (2014). Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat. Commun.* 5:5560. doi: 10.1038/ncomms6560
- Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., et al. (2016). Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* 533, 125–129. doi: 10.1038/nature17664
- Ran, A. F., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308. doi: 10.1038/nprot.2013.143
- Saito, S., and Adachi, N. (2016). Advances in the development of gene-targeting vectors to increase the efficiency of genetic modification. *Biol. Pharm. Bull.* 39, 25–32. doi: 10.1248/bpb.b15-00701
- Yusa, K. (2013). Seamless genome editing in human pluripotent stem cells using custom endonuclease-based gene targeting and the piggyBac transposon. *Nat. Protoc.* 8, 2061–2078. doi: 10.1038/nprot.2013.126
- Yusa, K., Zhou, L., Li, M., Bradley, A., and Craig, N. L. (2011). A hyperactive piggyBac transposase for mammalian applications. *Proc. Natl. Acad. Sci.* 108, 1531–1536. doi: 10.1073/pnas.1008322108

Conflict of Interest Statement: The authors are inventors in patent PCT/EP2017/051889.

Copyright © 2019 Jarazo, Qing and Schwamborn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Programmable Base Editing of the Sheep Genome Revealed No Genome-Wide Off-Target Mutations

Shiwei Zhou^{1†}, Bei Cai^{1†}, Chong He², Ying Wang¹, Qiang Ding¹, Jiao Liu¹, Yao Liu¹, Yige Ding¹, Xiaoe Zhao³, Guanwei Li¹, Chao Li¹, Honghao Yu⁴, Qifang Kou⁵, Wenzhi Niu⁵, Bjoern Petersen⁶, Tad Sonstegard⁷, Baohua Ma^{3*}, Yulin Chen^{1*} and Xiaolong Wang^{1*}

¹ College of Animal Science and Technology, Northwest A&F University, Yangling, China, ² College of Information Engineering, Northwest A&F University, Yangling, China, ³ College of Veterinary Medicine, Northwest A&F University, Yangling, China, ⁴ Guilin Medical University, Guilin, China, ⁵ Ningxia Tianyuan Tan Sheep Farm, Hongsibu, China, ⁶ Institute of Farm Animal Genetics, Friedrich-Loeffler-Institut, Neustadt, Germany, ⁷ Recombinetics, Saint Paul, MN, United States

OPEN ACCESS

Edited by:

David Jay Segal,
University of California, Davis,
United States

Reviewed by:

Pavel Georgiev,
Institute of Gene Biology (RAS),
Russia
Nathan Ellis,
The University of Arizona,
United States

*Correspondence:

Baohua Ma
mabh@nwfau.edu.cn
Yulin Chen
chenyulin@nwfau.edu.cn
Xiaolong Wang
xiaolongwang@nwfau.edu.cn

[†]Co-first authors

Specialty section:

This article was submitted to
Genomic Assay Technology,
a section of the journal
Frontiers in Genetics

Received: 16 November 2018

Accepted: 27 February 2019

Published: 15 March 2019

Citation:

Zhou S, Cai B, He C, Wang Y,
Ding Q, Liu J, Liu Y, Ding Y, Zhao X,
Li G, Li C, Yu H, Kou Q, Niu W,
Petersen B, Sonstegard T, Ma B,
Chen Y and Wang X (2019)
Programmable Base Editing of the
Sheep Genome Revealed No
Genome-Wide Off-Target Mutations.
Front. Genet. 10:215.
doi: 10.3389/fgene.2019.00215

Since its emergence, CRISPR/Cas9-mediated base editors (BEs) with cytosine deaminase activity have been used to precisely and efficiently introduce single-base mutations in genomes, including those of human cells, mice, and crop species. Most production traits in livestock are induced by point mutations, and genome editing using BEs without homology-directed repair of double-strand breaks can directly alter single nucleotides. The p.96R > C variant of Suppressor cytokine signaling 2 (SOCS2) has profound effects on body weight, body size, and milk production in sheep. In the present study, we successfully obtained lambs with defined point mutations resulting in a p.96R > C substitution in SOCS2 by the co-injection of BE3 mRNA and a single guide RNA (sgRNA) into sheep zygotes. The observed efficiency of the single nucleotide exchange in newborn animals was as high as 25%. Observations of body size and body weight in the edited group showed that gene modification contributes to enhanced growth traits in sheep. Moreover, targeted deep sequencing and unbiased family trio-based whole genome sequencing revealed undetectable off-target mutations in the edited animals. This study demonstrates the potential for the application of BE-mediated point mutations in large animals for the improvement of production traits in livestock species.

Keywords: base editing, genome editing, point mutation, whole genome sequencing, off-target mutation

INTRODUCTION

Clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated (Cas) 9 is widely used to establish site-specific genome-edited cell lines and animal models (Sander and Joung, 2014). The Cas9 protein, under the guidance of single guide (sg) RNA, cleaves DNA at sequence-specific sites in the genome and produces a double-strand break (DSB). To response DSB, cellular DNA repair pathways generated more abundant insertion and deletions (indels) by non-homologous end-joining (NHEJ) than that of homology-directed repair (HDR)-mediated gene correction (Chu et al., 2015; Paquet et al., 2016; Sakuma et al., 2016). Therefore, developing alternative approaches to correct point mutations that do not need DSBs is highly expected. Rat cytidine deaminase (rAPOBEC1) linked to nCas9 (Cas9 nickase) was reported to efficiently convert C→T at target sites without introducing DSBs (Komor et al., 2016). After several generations of modification, base editor 3 (BE3) including rAPOBEC1, nCas9 (A840H),

and uracil DNA glycosylase inhibitor (UGI) was developed; the mutation efficiency was up to 74.9% in mammalian cells (Komor et al., 2016). To further optimize BE3, previous studies have been conducted to improve target specificity (Kim D. et al., 2017), editing efficiency and product purity (Komor et al., 2017), expand the genome-targeting scope (Kim K. et al., 2017), and reduce off-target effects (Kim D. et al., 2017). To prove that BE3 has a high efficiency for converting C:G to T:A base pairs, several groups have used BE3 to silence genes by introducing nonsense mutations (Billon et al., 2017; Kusec et al., 2017).

Sheep are a phenotypically diverse livestock species that are raised globally for meat, milk, and fiber production. Suppressor of cytokine signaling 2 (SOCS2), a member of the SOCS protein family, is a negative regulator of biological processes mediated by various cytokines, such as metabolism, skeletal muscle development, and the response to infection (Inagaki-Ohara et al., 2014; Letellier and Haan, 2016). The most important of these processes is the regulation of GH signaling during growth and development (Yang et al., 2012; Dobie et al., 2018). SOCS2 is the major gene involved in the promotion of bone development in mice, and it plays a vital role in the control of bone mass and body weight (Metcalf et al., 2000; Dobie et al., 2018). A point mutation g.C1901T (p.R96C) in SOCS2 that completely abrogates SOCS2 binding affinity for the phosphopeptide of growth hormone receptor (GHR) is highly associated with an increased body weight and size in sheep (Rupp et al., 2015). We recently reported the usage of the BE3 system to induce nonsense mutations in the goat *FGF5* gene, to generate animals with longer hair fibers (Li G. et al., 2018). It was the first base editing study in large animals and further inspired us to examine the feasibility of induce amino acid exchanges in sheep. In the present study, we obtained BE3-mediated lambs by co-injection of a BE3 mRNA and guide RNA target the p.R96C variant in SOCS2. In addition, we used a parent-progeny whole genome sequencing (WGS) approach to show that no off-target mutations were detected and the mutation frequency in edited animals is equivalent to that in control groups.

MATERIALS AND METHODS

Animals

Tan sheep were maintained at the Ningxia Tianyuan Sheep Farm, Hongsibu, Ningxia Autonomous Region, China. All experimental animals were provided water and standard feed *ad libitum*, consistent with normal sheep, and were treated according to Guidelines for the Care and Use of Laboratory Animals formulated by the College of Animal Science and Technology, Northwest A&F University. The experimental study was approved by the Northwest A&F University Animal Care and Use Committee (Approval ID: 2016NXTS001).

Design of sgRNA

The sequences target the g.C1901T (p.R96C) variant in the ovine SOCS2 gene is listed in **Supplementary Table S1**. Two oligonucleotides (**Supplementary Table S2**) used for the transcription of sgRNA *in vitro* were precisely synthesized and

annealed to form double-stranded oligos. These double-stranded oligos were subcloned into the pUC57-T7-gRNA vector as described previously (Shen et al., 2013). The clones containing the desired sequence were selected, expanded by cultivation, and the plasmid was extracted using a plasmid extraction kit (AP-MN-P-250G; Axygen, Union City, CA, United States), sgRNA was transcribed *in vitro* using the MEGAshortscript Kit (AM1354; Ambion, Foster City, CA, United States) and purified using the MEGAclear Kit (AM1908; Ambion). Subsequently, the BE3 mRNA *in vitro* transcription vector (No. 44758; Addgene, Cambridge, MA, United States) was used as a template to produce BE3 mRNAs following a previously published protocol (Shen et al., 2013).

Production of Single-Nucleotide Mutation Sheep

Healthy ewes (3–5 years old) with regular estrous cycles were selected as donors for zygote collection. The superovulation treatment of donors and the procedures for zygote collection were as described previously (Wang et al., 2015). Briefly, an EAZI-BREED controlled internal drug release (CIDR) Sheep and Goat Device (containing 300 mg of progesterone) was inserted into the vagina of the donor sheep for 12 days and superovulation was performed 60 h before CIDR Device removal. Zygotes at the 1-cell stage were surgically collected and immediately transferred to TCM-199 medium (Gibco, Gaithersburg, MD, United States). A mixture of BE3 mRNA (25 ng μL^{-1}) and sgRNA (10 ng μL^{-1}) was co-injected into the cytoplasm of 1-cell stage zygotes using an Eppendorf FemtoJet system. The injection pressure, injection time, and compensatory pressure were 45 kPa, 0.1 s, and 7 kPa, respectively. Microinjections were performed on the heated stage of the Olympus ON3 micromanipulation system. Injected embryos were cultured in Quinn's Advantage Cleavage Medium and Blastocyst Medium (Sage BioPharma, Toronto, ON, Canada) for ~24 h and were then transferred into surrogates, as reported previously (Wang et al., 2016). Pregnancy was determined by observed estrous behaviors of surrogates at every ovulation cycle. After 150 days of pregnancy, newborn lambs were delivered and genotyped.

Genotyping of Delivered Animals

Peripheral venous blood samples were collected from newborn lambs at day 15 after birth for genomic DNA extraction. Polymerase chain reaction (PCR) amplification-based Sanger sequencing was conducted using KOD-NEO-Plus enzyme (DR010A; TOYOKAI, Osaka, Japan) and primers are listed in **Supplementary Table S3**.

Prediction of Off-Target Sites

Potential off-target sites with up to three mismatches were predicted using the openly available tool SeqMap (Jiang and Wong, 2008). The process for searching for off-target sites was implemented as previously described (Wang et al., 2015; Niu et al., 2017). The primers used for amplifying off-target sites by captured deep sequencing are given in **Supplementary Table S4**.

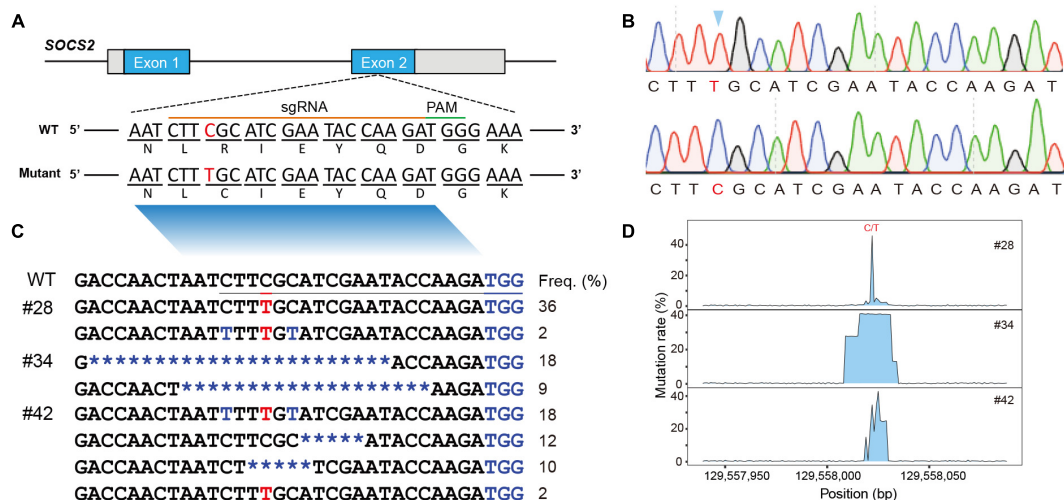


FIGURE 1 | Animals and mutational spectra in the edited goat genome. **(A)** Scheme showing the target site in the sheep SOCS2 gene. sgRNA sequences are presented in yellow. PAM sequences are highlighted in green. The BE3-mediated nucleotide substitutions (g.C1901T, p.R96C) are highlighted. **(B)** Sanger sequencing chromatogram of intended mutations directed by the BE3 system. **(C)** Genotypes of target sites derived from deep sequencing in the three founder animals. **(D)** Mutation rate at the targeted region.

Captured Deep-Sequencing

On-target and potential off-target mutations were amplified using a KAPA HiFi HotStart PCR Kit (#KK2501; KAPA Biosystems, Wilmington, MA, United States) for deep sequencing library generation. Pooled PCR amplicons were sequenced using the MiSeq with TruSeq HT Dual Index system (Illumina, San Diego, CA, United States).

Whole Genome Sequencing

Genomic DNA of nine animals from three edited families were used for WGS. Nine DNA libraries with insert sizes of approximately 300 bp were constructed following the manufacturer's instructions, and 150-bp paired-end reads were generated using the Illumina HiSeq XTen PE150 platform. The qualified reads were mapped to the sheep reference genome (Jiang et al., 2014) using the BWA (v0.7.13) tool (Li and Durbin, 2009). Local realignment and base quality recalibration were assessed with the Genome Analysis Toolkit (GATK) (McKenna et al., 2010). Single-nucleotide polymorphisms (SNPs) and small indels (<50 bp) were called using GATK (McKenna et al., 2010) and SAMtools (Li et al., 2009).

Identification of Off-Target Mutations

The called SNPs were filtered according to the following criteria: (1) SNPs that were identified by both GATK and SAMtools; (2) excluding SNPs that exist in NCBI sheep SNP database (>59 million SNPs); (3) excluding SNPs that exist in our sheep SNP database ($n = 294$, >79 million SNPs¹); (4) within the remaining SNPs, SNPs with C and G converted to other base types were selected. The potential off-target sites were predicted using Cas-OFFinder (Bae et al., 2014) by consider allowing up to

five mismatches. SNPs within the predicted off-target sites were identified as off-target mutations.

Identification of *de novo* Mutations

Putative *de novo* SNPs and indels were identified according to our recent report (Wang et al., 2018). Briefly, the SNPs/indels were identified by both GATK and SAMtools, and SNPs/indels that were found in the NCBI and our own sheep SNP databases were removed. Next, the SNPs/indels inherited from parents were excluded. Additional SNPs/indels were filtered based on parameters including read depth and Phred-scaled likelihood (PL) scores (Wang et al., 2018). Finally, the mis-aligned or miscalled SNPs/indels were removed manually. Genome-wide structure variations (SVs) were called using BreakDancer (Chen et al., 2009), then the SVs specific in the edited animals were remained. To identify the *de novo* SVs, common SVs in every two founders, and the read depth <50%, as well as the scaffolds were removed.

TABLE 1 | Summary of the sheep generated with the defined point mutation via BE3.

Donor sheep	5
Collected embryos	54
BE3-sgRNA	
Injected embryos	53
Developing embryos	20
Recipient sheep	8
Gestation recipient	3
Newborns	4
Expected single base substitution	1
Indel	2
WT	1

¹<http://animal.nwsuaf.edu.cn>

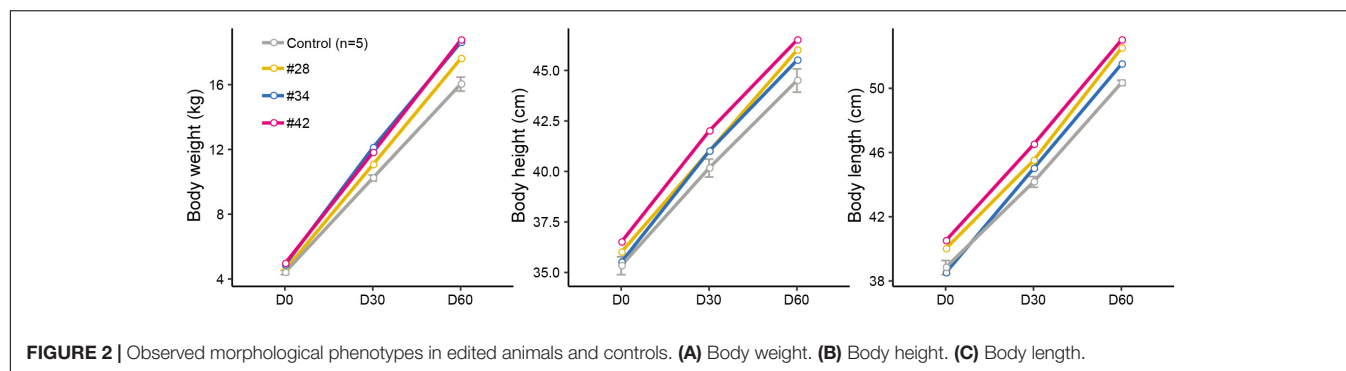


FIGURE 2 | Observed morphological phenotypes in edited animals and controls. (A) Body weight. (B) Body height. (C) Body length.

RESULTS AND DISCUSSION

Generation of Edited Animals

To obtain lambs comprising the precise g.1901C > T mutation in *SOCS2*, we micro-injected the BE3 mRNA and sgRNA into the cytoplasm of 1-cell-stage embryos. The sgRNA was designed to encompass the target point mutation p.R96C in *SOCS2* (Figure 1A). Five mated Tan sheep that were treated for superovulation received 54 one-cell fertilized oocytes; after 53 embryos were subjected to micro-injection, 20 developing embryos were transplanted into eight recipients. Three recipient sheep were confirmed with pregnancy. After ~150 days of gestation, four lambs (#28, #34, #41, and #42) were obtained (Table 1).

Genomic DNA was isolated from the blood samples of the four lambs (#28, #34, #41, and #42) and the targeted region was evaluated by PCR-based Sanger sequencing; this analysis confirmed that three animals (#28, #34, and #42) were edited at the target site (Supplementary Figure S1A). We then used TA cloning and sequencing to further validate the genotypes of the three edited animals, and a nucleotide substitution at the p.R96C mutation site was found in #28 and #42 (Supplementary Figure S1B). TA-cloned sequencing further revealed short indels in the edited animals, for example, the founder animal #34 had 19- and 23-bp deletions and #42 was mosaic with the defined point mutation and a 5-bp deletion (Supplementary Figure S1B). To fully screen the genotypes in the edited animals, these three edited animals were subject to targeted deep sequencing, which confirmed the TA-sequencing results and identified additional low incidence of C-T genotypes within the editing window (Figures 1B,C; Gehrke et al., 2018). We demonstrated that BE3-mediated modification in sheep led to a gene knockout animal (#34), apparent mosaics, and a low incidence of short indels in edited animals (Figure 1D). To further investigate the mosaicism in the BE-edited animals, we sequenced the modified loci in additionally biopsied tissues (tail, muscle, and skin) of the three animals (#28, #34, and #42). We identified same heterozygous genotypes in these tissues as observed in whole blood in #28 and #42 (Supplementary Figure S1A), indicating the genetic modification occurred during early embryogenesis. The non-specificity of the programmable deaminase BE3 often results in short indels and mosaicism

(Kim D. et al., 2017; Park et al., 2017; Sasaguri et al., 2018). Efforts have been made to optimize DNA specificity and minimize bystander effects of BEs (Rees et al., 2017; Gehrke et al., 2018) or to develop advanced cytidine and adenine BEs with high efficiency (e.g., BE4max, AncBE4max, and ABEmax; Koblan et al., 2018).

Although the editing efficiency in this study was as high as 75.0% (3/4), we only generated one animal with the precise point substitution (#28, 25%, 1/4) (Table 1). The efficiency of precise single-base substitution was equivalent to that observed in goats (24%) (Niu et al., 2018) and sheep (Zhou et al., 2018), and was significantly higher than that in zebrafish (4%) (Armstrong et al., 2016). We expect to use a variety of newly invented BEs to improve the DNA specificity and diminish bystander effects in the editing window.

Phenotypes of Edited Animals

Subsequently, we analyzed the growth curve and body size of mutant and control lambs to assess whether the p.R96C mutation impaired the function of the *SOCS2* protein associated with morphology. The body weight of three edited sheep (#28, #34, and #42) was higher than that in the control group on D0, D30, and D60; body length and height in modified sheep were higher than those in the control group (Figure 2). We did

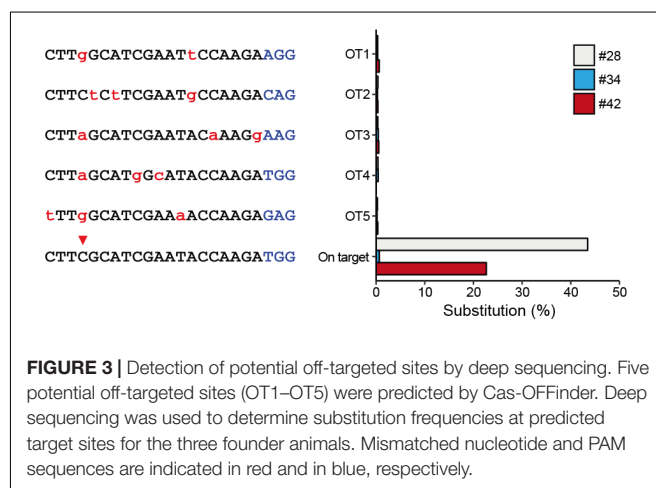


FIGURE 3 | Detection of potential off-targeted sites by deep sequencing. Five potential off-targeted sites (OT1–OT5) were predicted by Cas-OFFinder. Deep sequencing was used to determine substitution frequencies at predicted target sites for the three founder animals. Mismatched nucleotide and PAM sequences are indicated in red and in blue, respectively.

not observe clear phenotypic differences in the three edited animals that were related to their genotypes (substitutions, deletions, or both), even the body parameters of #42 was higher than in the other two animals (Figure 2). Considering the mosaicism in the edited animals, more phenotypic data from a long period is needed to address the correlation of mutation types to phenotypes. Nevertheless, these results are consistent with a spontaneous mutation in *SOCS2* causing a 30–50% increase in the postnatal growth of mice (Horvat and Medrano, 2001), and the *SOCS2* p.R96C mutation in sheep led to an increase body weight, body size, and milk yield (Rupp et al., 2015). Moreover, *SOCS2* deletion protects bone health in inflammatory bowel disease and causes a high-growth

phenotype in mice (Horvat and Medrano, 2001; Dobie et al., 2018). We found that two animals (#34 and #42) with *SOCS2* indels were as healthy as normal sheep, and we did not observe any health issues.

Off-Target Mutations in Edited Animals

To characterize off-target effects induced by the BE system, a deep sequencing assay was used to amplify predicted off-target sites in all the three edited animals. Five off-target sites (OT1–OT5) were predicted using the SeqMap tool (Jiang and Wong, 2008) (Supplementary Table S5). Targeted deep sequencing revealed that the frequency of BE3-induced point mutations is low at all predicted sites

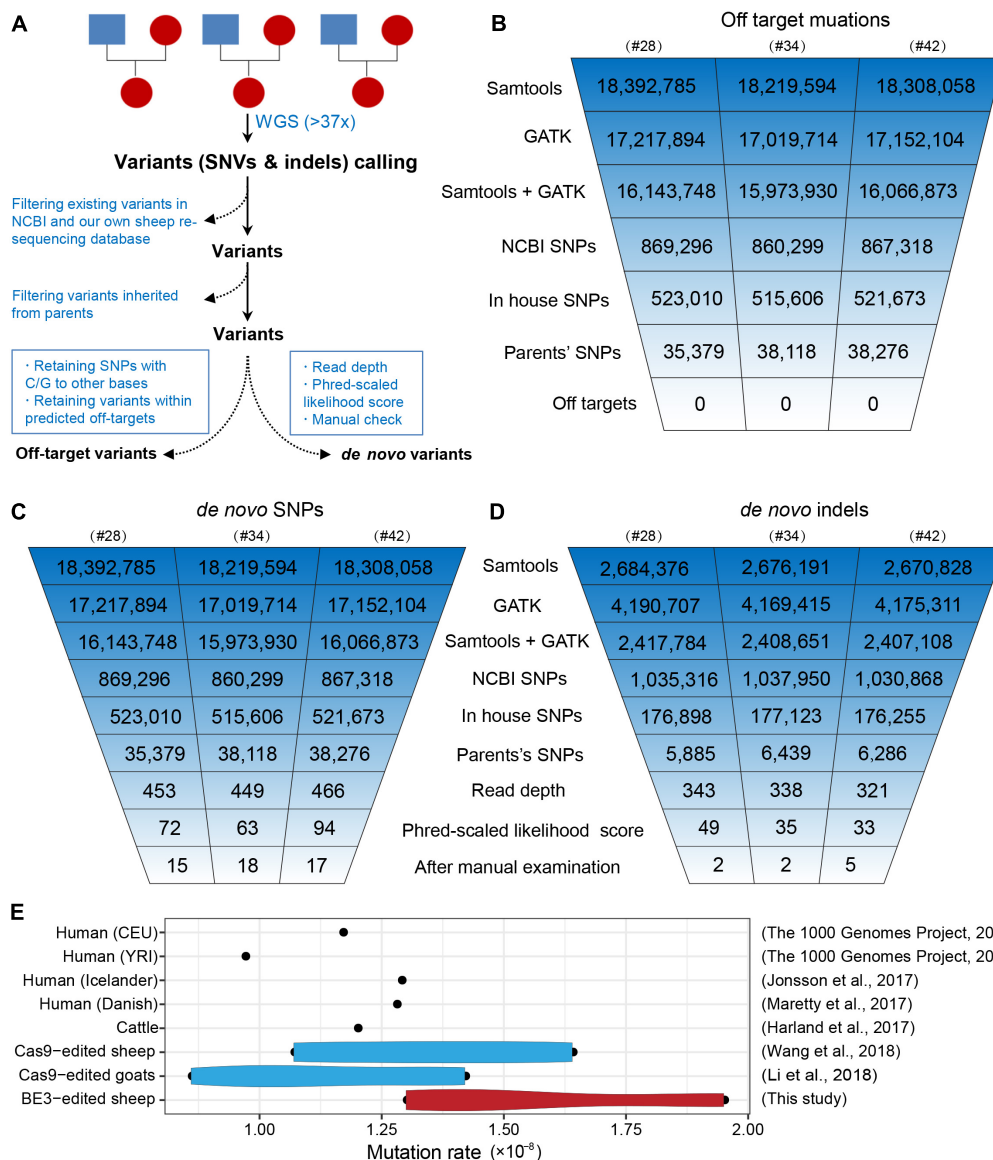


FIGURE 4 | Identification of the genome-wide off-targets and *de novo* mutations by trio-based WGS. **(A)** Schematic representation of the study design for the identification of genome-wide off-target variants and *de novo* mutations. Summary of the filtering process of off-target mutations **(B)**, *de novo* SNPs **(C)**, and *de novo* indels **(D)**. **(E)** Mutation rates in BE3-edited sheep and other populations.

in all three founder animals (#28, #34, and #42) (**Figure 3**), indicating that the incidence of BE3-induced off-target mutations is rare.

To further characterize off-target mutations at the whole-genome scale, we conducted family trio-based WGS to assess off-targets and *de novo* mutations in the three edited animals (#28, #34, and #42) (**Figure 4A**). We calculated the kinship coefficient for pair-wised animals to guarantee the pedigree information (**Supplementary Table S6**). The WGS yielded an average sequence coverage of $37.3\times$ per individual, within a range of 34- to 41-fold, and generated 12–13 million SNPs for each animal (**Supplementary Table S7**). SNPs were first called by both GATK and SAMtools, and an average of 16 million SNPs were identified for each founder. Of the SNPs we were able to map in this study, we next removed naturally occurring variants in the NCBI sheep SNP database (>59 million SNPs) and in our own sheep SNP database (>79 million SNPs from 294 individuals) and filtered out SNPs that were inherited from parents, resulting in $\sim 37,000$ remaining SNPs for each founder animal. We then excluded base substitutions including SNP types C to T/A/G and their antisense type G to A/T/C according to a recent study (Kim D. et al., 2017). Subsequently, we assessed the remaining SNPs that were within the predicted off-target sites (tolerant to five mismatches) (**Supplementary Table S8**) using Cas-OFFinder (Bae et al., 2014), and no single variants were identified (**Figure 4B**), indicating that no off-target mutations were induced by BE3 in the present study. The detailed filtering procedure is summarized in **Supplementary Table S9**.

To identify *de novo* mutations (SNPs and indels) in the edited animals, we used a stringent pipeline for variant filtering, as previously described (Li C. et al., 2018; Wang et al., 2018). Briefly, we selected SNPs that were identified by both GATK and SAMtools and removed existing SNPs in both the NCBI SNP database and our own sheep SNP database as well as the SNPs found in their parents. Next, we filtered out SNPs according to sequence read depth, PL scores, and manual examination of the FASTQ files (Wang et al., 2018). The remaining 15, 18, and 17 SNPs in individuals #28, #34, and #42 were identified as *de novo* SNPs for each progeny (**Figure 4C**). We next validated these *de novo* SNPs with Sanger sequencing. Of the 46 successfully amplified and sequenced SNPs, 44 of them were determined as true variants (**Supplementary Figure S2** and **Supplementary Table S10**), indicating the pipeline for identification of *de novo* SNPs was robust. Similarly, we were able to identify 5, 5, and 10 *de novo* indels in individuals #28, #34, and #42, respectively (**Figure 4D** and **Supplementary Table S11**). To further characterize the large-scale genomic alterations induced by base editing, we called the SVs by BreakDancer (Chen et al., 2009), and identified a total of ten *de novo* SVs in the three BE-mediated animals (**Supplementary Tables S12, S13**), none of these variants were adjacent to the *SOSC2* site.

Additionally, we estimated the mutation rates per base pair per generation (Li C. et al., 2018), and found that no

apparent differences between the BE3 and control animals in term of frequency of *de novo* SNPs in the present study and former studies (**Figure 4E**). Albeit only three trios were analyzed in this study, the mutation rate in base-edited sheep was equivalent to that in human populations (1000 Genomes Project Consortium et al., 2010; Maretty et al., 2017), cattle (Harland et al., 2017), as well as our previously generated CRISPR/Cas9-edited sheep and goat populations (Li C. et al., 2018; Wang et al., 2018). Along with our previous studies reporting the *de novo* mutations in edited animals and their offspring (Li C. et al., 2018; Li G. et al., 2018; Wang et al., 2018) and recently two trio-based studies in mice (Iyer et al., 2018; Willi et al., 2018), we demonstrate that the mutation frequency does not differ in Cas9-mediated or BE-mediated animals, thereby providing evidence to support the reliability of genome editing in large animals for biomedicine and agriculture.

CONCLUSION

In summary, a single sheep carrying the *SOSC2* p.R96C mutation was successfully generated using programmable deaminases BE3. We confirmed that BE3 did not induce unintended off-target mutations at the genome-wide scale, and the mutation frequency in BE-mediated animals was equivalent to those in Cas9-edited animals and in natural populations. This study facilitates gene correction and genetic improvement of large animals caused by single base mutations.

DATA AVAILABILITY

All relevant results are within the paper and its **Supplementary Data Files**. The raw WGS was available at NCBI SRA database under BioProject ID: PRJNA505205.

AUTHOR CONTRIBUTIONS

SZ, XZ, XW, BM, and YC conceived the study. BC, YW, QD, XZ, JL, YL, YD, GL, HY, and BM performed the experiments. CH, CL, and XW analyzed the dataset. QK and WN provided samples. XW, TS, and BP wrote the article.

FUNDING

This work was supported by the National Natural Science Foundation of China (31772571), and Local Grants (NXTS20118-001, 2017NY-072, and 2018KJXX-009). XW is a Tang Scholar at Northwest A&F University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00215/full#supplementary-material>

REFERENCES

- 1000 Genomes Project Consortium, Durbin, R. M., Altshuler, D., Durbin, R. M., Abecasis, G. R., Bentley, D. R., et al. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073. doi: 10.1038/nature09534
- Armstrong, G. A. B., Liao, M., You, Z., Lissouba, A., Chen, B. E., and Drapeau, P. (2016). Homology directed knockin of point mutations in the zebrafish *tardbp* and *fus* genes in ALS using the CRISPR/Cas9 System. *PLoS One* 11:e0150188. doi: 10.1371/journal.pone.0150188
- Bae, S., Park, J., and Kim, J.-S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 30, 1473–1475. doi: 10.1093/bioinformatics/btu048
- Billon, P., Bryant, E. E., Joseph, S. A., Nambiar, T. S., Hayward, S. B., Rothstein, R., et al. (2017). CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of STOP codons. *Mol. Cell* 67, 1068–1079. doi: 10.1016/j.molcel.2017.08.008
- Chen, K., Wallis, J. W., McLellan, M. D., Larson, D. E., Kalicki, J. M., Pohl, C. S., et al. (2009). Breakdancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat. Methods* 6, 677–681. doi: 10.1038/nmeth.1363
- Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., et al. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* 33, 543–548. doi: 10.1038/nbt.3198
- Dobie, R., MacRae, V. E., Pass, C., Milne, E. M., Ahmed, S. F., and Farquharson, C. (2018). Suppressor of cytokine signaling 2 (Socs2) deletion protects bone health of mice with DSS-induced inflammatory bowel disease. *Dis. Model. Mech.* 11:dmm028456. doi: 10.1242/dmm.028456
- Gehrke, J. M., Cervantes, O., Clement, M. K., Wu, Y., Jing, Z., Bauer, D. E., et al. (2018). An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* 36, 977–982. doi: 10.1038/nbt.4199
- Harland, C., Charlier, C., Karim, L., Cambisano, N., Deckers, M., Mni, M., et al. (2017). Frequency of mosaicism points towards mutation-prone early cleavage cell divisions. *bioRxiv* [Preprint]. doi: 10.1101/079863
- Horvat, S., and Medrano, J. F. (2001). Lack of Socs2 expression causes the high-growth phenotype in mice. *Genomics* 72, 209–212. doi: 10.1006/geno.2000.6441
- Inagaki-Ohara, K., Mayuzumi, H., Kato, S., Minokoshi, Y., Otsubo, T., Kawamura, Y. I., et al. (2014). Enhancement of leptin receptor signaling by SOCS3 deficiency induces development of gastric tumors in mice. *Oncogene* 33, 74–84. doi: 10.1038/onc.2012.540
- Iyer, V., Boroviak, K., Thomas, M., Doe, B., Riva, L., Ryder, E., et al. (2018). No unexpected CRISPR-Cas9 off-target activity revealed by trio sequencing of gene-edited mice. *PLoS Genet.* 14:e1007503. doi: 10.1371/journal.pgen.1007503
- Jiang, H., and Wong, W. H. (2008). SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics* 24, 2395–2396. doi: 10.1093/bioinformatics/btn429
- Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J. F., Faraut, T., et al. (2014). The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 344, 1168–1173. doi: 10.1126/science.1252806
- Kim, D., Lim, K., Kim, S. T., Yoon, S. H., Kim, K., Ryu, S. M., et al. (2017). Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat. Biotechnol.* 35, 475–480. doi: 10.1038/nbt.3852
- Kim, K., Ryu, S. M., Kim, S. T., Baek, G., Kim, D., Lim, K., et al. (2017). Highly efficient RNA-guided base editing in mouse embryos. *Nat. Biotechnol.* 35, 435–437. doi: 10.1038/nbt.3816
- Koblan, L. W., Doman, J. L., Wilson, C., Levy, J. M., Tay, T., Newby, G. A., et al. (2018). Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* 36, 843–846. doi: 10.1038/nbt.4172
- Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424. doi: 10.1038/nature17946
- Komor, A. C., Zhao, K. T., Packer, M. S., Gaudelli, N. M., Waterbury, A. L., Koblan, L. W., et al. (2017). Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T: A base editors with higher efficiency and product purity. *Sci. Adv.* 3:eao4774. doi: 10.1126/sciadv.aao4774
- Kuscu, C., Parlak, M., Tufan, T., Yang, J. K., Szlachta, K., Wei, X. L., et al. (2017). CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nat. Methods* 14, 710–712. doi: 10.1038/nmeth.4327
- Letellier, E., and Haan, S. (2016). SOCS2: physiological and pathological functions. *Front. Biosci.* 8, 189–204. doi: 10.2741/e760
- Li, C., Zhou, S., Li, Y., Li, G., Ding, Y., Li, L., et al. (2018). Trio-based deep sequencing reveals a low incidence of off-target mutations in the offspring of genetically edited goats. *Front. Genet.* 9:449. doi: 10.3389/fgene.2018.00449
- Li, G., Zhou, S., Li, C., Cai, B., Yu, H., Ma, B., et al. (2018). Base pair editing of goat embryos: nonsense codon introgression into FGF5 to improve cashmere yield. *bioRxiv* [Preprint]. doi: 10.1101/348441
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Marett, L., Jensen, J. M., Petersen, B., Sibbesen, J. A., Liu, S., Villesen, P., et al. (2017). Sequencing and de novo assembly of 150 genomes from Denmark as a population reference. *Nature* 548, 87–91. doi: 10.1038/nature23264
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Metcalf, D., Greenhalgh, C. J., Viney, E., Willson, T. A., Starr, R., Nicola, N. A., et al. (2000). Gigantism in mice lacking suppressor of cytokine signalling-2. *Nature* 405, 1069–1073. doi: 10.1038/35016611
- Niu, Y., Jin, M., Li, Y., Li, P., Zhou, J., Wang, X., et al. (2017). Biallelic beta-carotene oxygenase 2 knockout results in yellow fat in sheep via CRISPR/Cas9. *Anim. Genet.* 48, 242–244. doi: 10.1111/age.12515
- Niu, Y., Zhao, X., Zhou, J., Li, Y., Huang, Y., Cai, B., et al. (2018). Efficient generation of goats with defined point mutation (I397V) in GDF9 through CRISPR/Cas9. *Reprod. Fertil. Dev.* 30, 307–312. doi: 10.1071/rd17068
- Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., et al. (2016). Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* 533, 125–129. doi: 10.1038/nature17664
- Park, D. S., Yoon, M., Kweon, J., Jang, A. H., Kim, Y., and Choi, S. C. (2017). Targeted base editing via RNA-guided cytidine deaminases in *xenopus laevis* embryos. *Mol. Cells* 40, 823–827. doi: 10.14348/molcells.2017.0262
- Rees, H. A., Komor, A. C., Yeh, W.-H., Caetano-Lopes, J., Warman, M., Edge, A. S. B., et al. (2017). Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* 8:15790. doi: 10.1038/ncomms15790
- Rupp, R., Senin, P., Sarry, J., Allain, C., Tasca, C., Ligat, L., et al. (2015). A point mutation in suppressor of cytokine signalling 2 (Socs2) increases the susceptibility to inflammation of the mammary gland while associated with higher body weight and size and higher milk production in a sheep model. *PLoS Genet.* 11:e1005629. doi: 10.1371/journal.pgen.1005629
- Sakuma, T., Nakade, S., Sakane, Y., Suzuki, K. T., and Yamamoto, T. (2016). MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCH systems. *Nat. Protoc.* 11, 118–133. doi: 10.1038/nprot.2015.140
- Sander, J. D., and Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32, 347–355. doi: 10.1038/nbt.2842
- Sasaguri, H., Nagata, K., Sekiguchi, M., Fujioka, R., Matsuba, Y., Hashimoto, S., et al. (2018). Introduction of pathogenic mutations into the mouse *Psen1* gene by Base Editor and Target-AID. *Nat. Commun.* 9:2892. doi: 10.1038/s41467-018-05262-w
- Shen, B., Zhang, J., Wu, H., Wang, J., Ma, K., Li, Z., et al. (2013). Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Res.* 23, 720–723. doi: 10.1038/cr.2013.46

- Wang, X., Liu, J., Niu, Y., Li, Y., Zhou, S., Li, C., et al. (2018). Low incidence of SNVs and indels in trio genomes of Cas9-mediated multiplex edited sheep. *BMC Genome* 19:397. doi: 10.1186/s12864-018-4712-z
- Wang, X., Niu, Y., Zhou, J., Yu, H., Kou, Q., Lei, A., et al. (2016). Multiplex gene editing via CRISPR/Cas9 exhibits desirable muscle hypertrophy without detectable off-target effects in sheep. *Sci. Rep.* 6:32271. doi: 10.1038/srep32271
- Wang, X., Yu, H., Lei, A., Zhou, J., Zeng, W., Zhu, H., et al. (2015). Generation of gene-modified goats targeting MSTN and FGF5 via zygote injection of CRISPR/Cas9 system. *Sci. Rep.* 5:13878. doi: 10.1038/srep13878
- Willi, M., Smith, H. E., Wang, C., Liu, C., and Hennighausen, L. (2018). Mutation frequency is not increased in CRISPR-Cas9-edited mice. *Nat. Methods* 15, 756–758. doi: 10.1038/s41592-018-0148-2
- Yang, H. L., Sun, C., Sun, C., and Qi, R. L. (2012). Effect of suppressor of cytokine signaling 2 (SOCS2) on fat metabolism induced by growth hormone (GH) in porcine primary adipocyte. *Mol. Biol. Rep.* 39, 9113–9122. doi: 10.1007/s11033-012-1783-9
- Zhou, S., Yu, H., Zhao, X., Cai, B., Ding, Q., Huang, Y., et al. (2018). Generation of gene-edited sheep with a defined Booroola fecundity gene (FecBB) mutation in bone morphogenetic protein receptor type 1B (BMPRI1B) via clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated (Cas) 9. *Reprod. Fertil. Dev.* 30, 1616–1621. doi: 10.1071/RD18086

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Zhou, Cai, He, Wang, Ding, Liu, Liu, Ding, Zhao, Li, Li, Yu, Kou, Niu, Petersen, Sonstegard, Ma, Chen and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



sgRNA-shRNA Structure Mediated SNP Site Editing on Porcine *IGF2* Gene by CRISPR/StCas9

Yongsun Sun[†], Nana Yan[†], Lu Mu, Bing Sun, Jingrong Deng, Yuanyuan Fang, Simin Shao, Qiang Yan, Furong Han, Zhiying Zhang* and Kun Xu*

Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, China

OPEN ACCESS

Edited by:

Youri I. Pavlov,
University of Nebraska Medical
Center, United States

Reviewed by:

Claudio Mussolino,
Institut für Zell- und Gentherapie
(IZG), Germany
Magdy Mahfouz,
King Abdullah University of Science
and Technology, Saudi Arabia

*Correspondence:

Kun Xu
xukunas@nwfau.edu.cn
Zhiying Zhang
zhangzhy@nwfau.edu.cn

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Genomic Assay Technology,
a section of the journal
Frontiers in Genetics

Received: 29 November 2018

Accepted: 01 April 2019

Published: 18 April 2019

Citation:

Sun Y, Yan N, Mu L, Sun B,
Deng J, Fang Y, Shao S, Yan Q,
Han F, Zhang Z and Xu K (2019)
sgRNA-shRNA Structure Mediated
SNP Site Editing on Porcine *IGF2*
Gene by CRISPR/StCas9.
Front. Genet. 10:347.
doi: 10.3389/fgene.2019.00347

The SNP within intron 3 of the porcine *IGF2* gene (G3072A) plays an important role for muscle growth and fat deposition in pigs. In this study, the StCas9 derived from *Streptococcus thermophilus* together with the Drosha-mediated sgRNA-shRNA structure were combined to boost the G to A base editing on the *IGF2* SNP site, which we called “SNP editing.” The codon-humanized StCas9 as we previously reported was firstly compared with the prevalently used SpCas9 derived from *Streptococcus pyogenes* using our idiomatic surrogate report assay, and the StCas9 demonstrated a comparable targeting activity. On the other hand, by combining shRNA with sgRNA, simultaneous gene silencing and genome targeting can be achieved. Thus, the novel IGF2.sgRNA-LIG4.shRNA-IGF2.sgRNA structure was constructed to enhance the sgRNA/Cas9-mediated HDR-based *IGF2* SNP editing by silencing the *LIG4* gene, which is a key molecule of the HDR's competitive NHEJ pathway. The sgRNA-shRNA/StCas9 all-in-one expression vector and the IGF2.sgRNA/StCas9 as control were separately used to transfect porcine PK15 cells together with an ssODNs donor for the *IGF2* SNP editing. The editing events were detected by the RFLP assay, Sanger sequencing as well as Deep-sequencing, and the Deep-sequencing results finally demonstrated a significant higher HDR-based editing efficiency (16.38%) for our sgRNA-shRNA/StCas9 strategy. In short, we achieved effective *IGF2* SNP editing by using the combined sgRNA-shRNA/StCas9 strategy, which will facilitate the further production of base-edited animals and perhaps extend for the gene therapy for the base correction of some genetic diseases.

Keywords: *IGF2* gene, SNP, base editing, CRISPR, StCas9, sgRNA-shRNA

INTRODUCTION

The CRISPR/Cas9 technology (Jinek et al., 2012; Mali et al., 2013) has been widely used for genome editing in various cell types and organisms since its advent. So far, the *S. pyogenes*-derived SpCas9 is the most prevalently applied Cas9 enzyme (Cong et al., 2013). Nevertheless, Cas9 variants from different microbial species can also contribute efficient genome editing, such

as *S. thermophilus* (StCas9) (Xu et al., 2015a) and *Staphylococcus aureus* (SaCas9) (Kleinstiver et al., 2015; Ran et al., 2015).

The Cas9 endonuclease is directed by an artificial single-guide RNA (sgRNA or gRNA) to recognize specifically the target DNA sequence with given protospacer adjacent motif (PAM) by base pairing (Jinek et al., 2012). In mammalian cells, the DNA double-strand breaks (DSBs) induced by the endonucleases can be repaired by two main mechanisms, the error-prone non-homologous end joining (NHEJ) and the donor-dependent homology-directed repair (HDR). The NHEJ pathway generates stochastic nucleotide insertions and deletions (Indels) at the target locus resulting in open reading frame (ORF) shift and loss-of-function of target genes. Alternatively, the HDR pathway can result in desired genome editing events by targeted recombination of designed homologous DNA template donors (Salsman and Delliare, 2017).

In animal breeding researches, researchers usually knock-out/in genes of interest (GOI) to study their function and relationship in the network of signaling pathways (Komor et al., 2017). However, it's always confronted with the concern of genetic safety, when creating gene knock-out/in animals. On the other hand, precise genome editing takes the advantage of the HDR pathway to make point mutations for gene function study, as well as gene therapy and animal breeding researches (Komor et al., 2017; Salsman and Delliare, 2017). However, the HDR efficiency is extremely lower than NHEJ in mammalian cells. Hence, different approaches have been reported to enhance the efficiency of the HDR-based precise genome editing. Firstly, inhibiting the key molecules of the competitive NHEJ pathway, such as DNA ligase IV (LIG4) and KU70 (Maruyama et al., 2015; Hu et al., 2018); Secondly, optimizing the DNA template donors (ssDNA or dsDNA); Thirdly, several molecules or small compounds have been also reported to improve the HDR efficiency significantly, such as RAD51 and RAD52 (homologous recombination related proteins), Nocodazole and CCND1 (synchronizes cell cycle at specific phase) (Lin et al., 2014; Chu et al., 2015; Salsman and Delliare, 2017; Shao et al., 2017).

We have previously developed the novel Droscha-mediated sgRNA-shRNA structure for transcribing multiple sgRNAs to promote the CRISPR/Cas9-based multiplex genome targeting (Yan et al., 2016). Interestingly, we noticed that the by-product shRNA could be used for silencing the *LIG4* gene to aid the HDR-based genome editing. Insulin like growth factor-2 (*IGF2*) is an important gene involved in pig muscle growth and fat deposition, and influencing the heart size. It has been reported that the NO.3072 nucleotide substitution from guanine (G) to adenine (A) within the intron 3 of the *IGF2* gene would obstruct the binding of the transcriptional inhibiting factor ZBED6, resulting in increased *IGF2* expression and muscle yield (Van Laere et al., 2003; Markljung et al., 2009). Our previous research has confirmed that it's the wild-type G at the *IGF2* SNP site in the local pig species (Shao et al., 2017).

In this study, the codon-humanized StCas9 derived from *S. thermophilus* (Xu et al., 2015a) and the novel sgRNA-shRNA

structure (Yan et al., 2016), both as we previously reported, were combined for enhancing the SNP editing of the porcine *IGF2* gene. Taking the advantage of the sgRNA-shRNA structure, simultaneous *IGF2* genome targeting by sgRNA/Cas9 and transient *LIG4* gene silencing by shRNA could be achieved. On the other hand, the CRISPR/StCas9 system with a stricter PAM requirement of NGGNG could reduce the off-target events compared with the CRISPR/SpCas9 system with the PAM pattern of NGG (Muller et al., 2016). Besides, the SNP editing also provided a good idea to avoid the genetic safety problem for the animal breeding study. Therefore, our novel sgRNA-shRNA/StCas9 strategy is of clear significance for gene editing or base correction, which will facilitate the further animal breeding research and the gene therapy study for the correction of genetic mutations.

MATERIALS AND METHODS

Construction of sgRNA/Cas9 Expression Vectors and Surrogate Reporters

The IGF2.sgRNA/StCas9 expression vector (pII3.7-mU6-IGF2.sgRNA-CMV-hStCas9) was constructed with a further modified sgRNA scaffold as shown in **Supplementary Figure S1**, which was designed by referencing our (Xu et al., 2015a) and another (Chen et al., 2013) previous studies for optimizing the sgRNA structure. The IGF2.sgRNA/SpCas9 expression vector was subsequently constructed by replacing the CMV-StCas9 cassette of the IGF2.sgRNA/StCas9 vector with the CBh-SpCas9 cassette amplified from the plasmid pX330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene, #42230).

A series of single strand annealing (SSA)-based surrogate reporters have been developed in our previous studies (Ren et al., 2019). The DsRed-eGFP (RG) and eGFP surrogate reporters (Xu et al., 2015a) were firstly designed and used for sgRNA/Cas9 activity verification in mammalian cells. The DsRed-Puro^R-eGFP (RPG) surrogate reporter with dual-reporter genes was further constructed to assist the enrichment and screening of genetically modified cells by either puromycin selection or fluorescence-activated cell sorting (FACS) (Ren et al., 2015). The IGF2.RG and IGF2.eGFP surrogate reporters were constructed in this study as we previously did (Xu et al., 2015a), while the IGF2.RPG surrogate reporter was constructed in our previous study (Shao et al., 2017).

Cell Culture and Transfection

The human embryonic kidney 293T (HEK293T) and porcine kidney epithelial (PK15) cells were cultured routinely in DMEM supplemented with FBS, penicillin and streptomycin as we previously did (Shao et al., 2017), with an additional 250 ng/ml antimycotic amphotericin B supplemented. The transfection assays were conducted within six-well plates using LipofectamineTM 2000 reagent (Invitrogen) following the manufacturer's protocol, with a total of 3 µg plasmid DNA per well. At least three independent wells were used for parallel transfections for each experiment group.

Surrogate Report Assay for Comparing StCas9 and SpCas9

The human embryonic kidney 293T cells were co-transfected with the IGF2.sgRNA/StCas9 or IGF2.sgRNA/SpCas9 expression vector and the IGF2.RG surrogate reporter within six-well plates. At least three wells were used for parallel transfections for each group, and the molecular ratio for the sgRNA/Cas9 vector and the RG reporter was 1:1. After transfected for 2 days, the cells from each parallel well were harvested independently for flow cytometric analysis to count the DsRed⁺ single and DsRed⁺eGFP⁺ dual positive cells. The flow cytometric data was analyzed by the flowJo v10 software. The percentage of dual-fluorescence positive cells as $\text{DsRed}^+\text{eGFP}^+ / (\text{DsRed}^+\text{eGFP}^+ + \text{DsRed}^+)$ was calculated to evaluate the Cas9 activity indirectly.

shRNA Design and Verification for Porcine LIG4 Gene Interference

Before designing the shRNAs against the porcine *LIG4* gene, three overlapping fragments (**Supplementary Figure S2**) from its complete CDS were amplified by RT-PCR with the PK15 cDNA as the template. The primers used are shown in **Supplementary Table S1**. The three PCR fragments were then sequenced and matched for verifying the sequence information of the *LIG4* gene. Afterward, three shRNA candidates were forecasted accordingly through the Invitrogen BLOCK-iTTM RNAi Designer and the corresponding oligonucleotides as shown in **Supplementary Table S2** were synthesized (Invitrogen). The three shRNA-1/2/3 cassettes were then generated by oligonucleotides-annealing (Xu et al., 2015b) and were cloned into the pLenti-H1 expression vector respectively. The non-specific shRNA vector pLenti-H1-SC (shRNA control) had been constructed previously in our lab. The H1-shRNA-CMV-eGFP expression cassettes from these four pLenti-H1 vectors were further amplified by PCR and cloned into the pB-CBh-puro vector. In addition to the fluorescent *eGFP* marker gene, the upgraded shRNA vectors (pB-CBh-Puro-H1-shRNA-CMV-eGFP) contained the puromycin resistant gene (*Puro*^R) (**Figure 2A**), which was intent for the enrichment of the transfected cells by puromycin selection. The PK15 cells were transfected with these pB-based shRNA vectors. 24 h after the transfection, the cells were selected by puromycin (3 µg/ml) for about another 2 days, and then were harvested for the total RNA preparation and quantitative RT-PCR analysis.

Quantitative RT-PCR Assay for LIG4 Gene Expression

Quantitative RT-PCR (qRT-PCR) assays were conducted as we previously performed (Yan et al., 2016) for detecting the relative transcript level of the *LIG4* gene. Generally, the parallel wells of transfected PK15 cells for each experiment or control group were collected independently for the total RNA isolation, the first-strand cDNA preparation and the further quantitative PCR analysis. The porcine β -actin gene was used as the internal control. The primers used for the qRT-PCR assays were listed in **Supplementary Table S1**.

sgRNA-shRNA Structure Design and Activity Verification

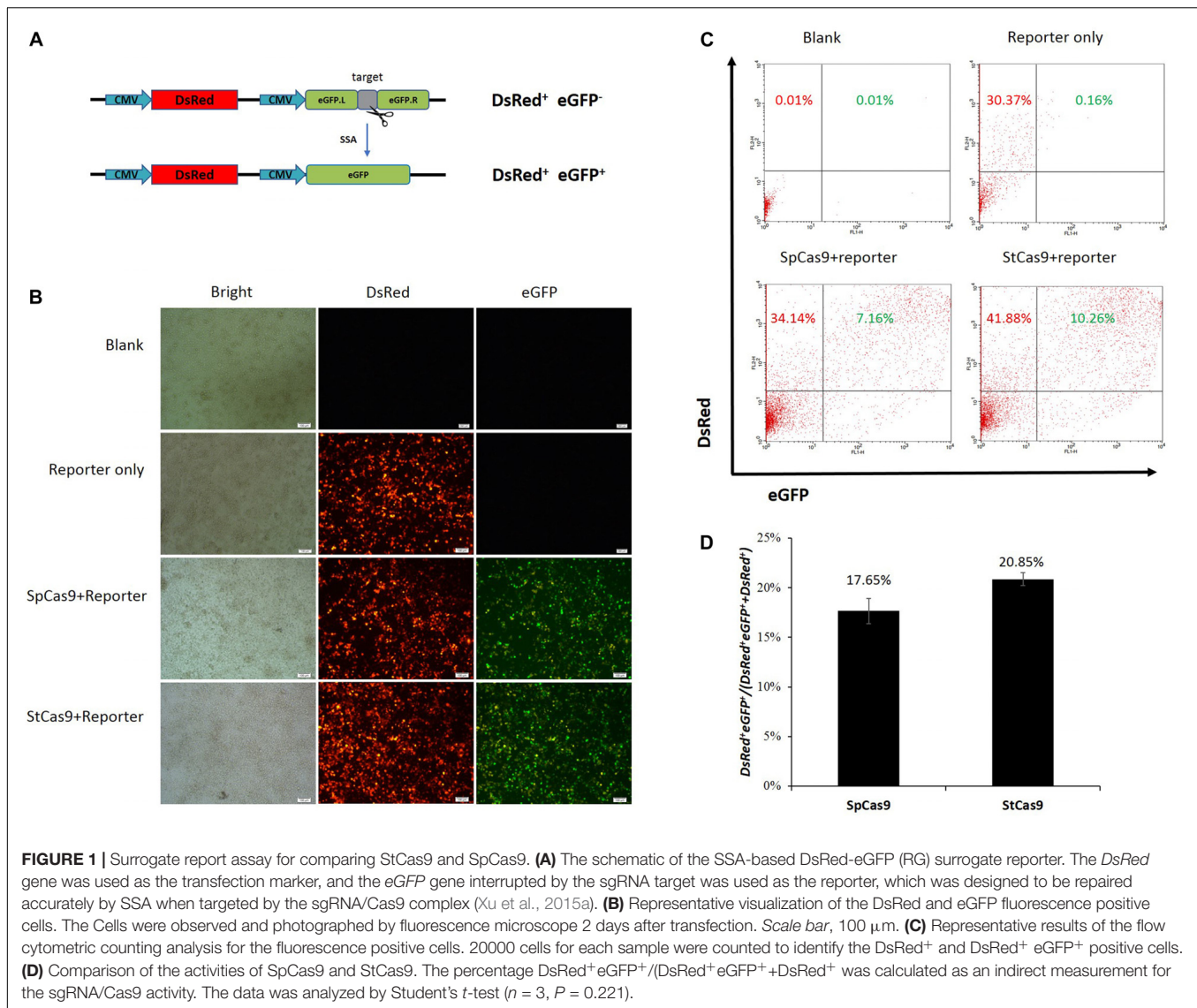
We have reported the novel Drosha-mediated sgRNA-shRNA structure for multiplex genome targeting in previous study (Yan et al., 2016). Here, the LIG4.shRNA-1, which showed the highest activity for silencing the porcine *LIG4* gene in the preceding experiment, was used for the sgRNA-shRNA structure design. For further improvement, a pair of more efficient Drosha-processing sequences from miR-30 (Zeng and Cullen, 2005) were used to replace the former Drosha recognition sites. As designed in **Figure 3A**, the LIG4.shRNA sequence flanked by the Drosha-processing sequences (as shown in **Supplementary Figure S3**) was synthesized directly and inserted into the middle of two identical IGF2.sgRNA sequences. And then the IGF2.sgRNA-LIG4.shRNA-IGF2.sgRNA cassette was cloned into the pLL3.7-mU6-CMV-hStCas9 vector (Xu et al., 2015a), generating the sgRNA-shRNA/StCas9 all-in-one expression vector, which was supposed capable for the simultaneous *IGF2* gene targeting and *LIG4* gene silencing.

Surrogate reporter assay was conducted to verify the sgRNA activity as described above. In brief, HEK293T cells were co-transfected with the IGF2.eGFP surrogate reporter and the sgRNA-shRNA/StCas9, the IGF2.sgRNA/StCas9, or the single StCas9 (pLL3.7-mU6-CMV-hStCas9 with no sgRNA as the negative control) expression vector. The linearized IGF2.eGFP surrogate reporter, which was supposed to repair spontaneously in cells after the transfection, was used as the positive control to co-transfect the cells with the single StCas9 expression vector as we previously did (Xu et al., 2015a). After transfected for 2 days, the cells were photographed and cells from each parallel well were harvested independently for flow cytometric analysis. The percentage of eGFP positive cells was used as an indirect measurement for the IGF2.sgRNA activity.

In addition to the IGF2.sgRNA-LIG4.shRNA-IGF2.sgRNA (Sg-Sh) cassette, an IGF2.sgRNA-SC-IGF2.sgRNA (Sg-SC) cassette was generated by replacing the LIG4.shRNA-1 with the non-specific shRNA control. To verify the shRNA activity driven by the sgRNA-shRNA structure, the Sg-SC cassette, as well as Sg-Sh, was further cloned into the pLenti-H1 vector. Then, the pLenti-H1 based shRNA control (SC), LIG4.shRNA-1(Sh-1), Sg-SC and Sg-Sh expression vectors were used to transfect the PK15 cells, respectively, along with the pB-CBh-Puro vector. The transfected cells were enriched by puromycin selection for about 2 days as above, and then were harvested for detecting the relative expression of the *LIG4* gene by qRT-PCR analysis.

Genome Editing of the Porcine IGF2 Gene

To conduct the HDR-based *IGF2* SNP editing, an 110 nt single-stranded oligodeoxynucleotides (ssODNs) with the desired G > A substitution (**Supplementary Table S2**) was synthesized by GenScript (Nanjing, China). The PAM motif of the IGF2.sgRNA target site within the ssODNs donor was mutated to the *NheI* restriction endonuclease (RE) site for the subsequent restrictive fragment length polymorphism (RFLP) assay.



Porcine kidney epithelial cells were co-transfected with the sgRNA-shRNA/StCas9 or IGF2.sgRNA/ StCas9 expression vector, the IGF2.RPG surrogate reporter and the ssODNs HDR donor. The transfections were conducted within six-well plates and the molecular ratio for RNA/Cas9:RPG reporter:ssODNs donor was 2:1:1. At least three wells were used for parallel transfections for each group. 2 days after the transfection, the cells were transferred into 60 mm dishes and were maintained continuously with puromycin treatment for another 5 days. After the puromycin selection, the resistant cell clones from the parallel dishes for each experiment group were collected as a pool as we previously did (Shao et al., 2017). The genomic DNAs for different pools were extracted, respectively, and the target locus was amplified by PCR for the subsequent RFLP and sequencing detections. The primers used were shown in **Supplementary Table S1**.

For the RFLP assay, the PCR products of the *IGF2* locus were 1319 bp in length. When the *IGF2* gene was edited

successfully as designed, the PCR product would be cut into two fragments (1081 and 238 bp) by the *NheI* RE induced. On the other hand, the PCR products from the two experimental groups were further cloned into the pMD19-T "T-A" cloning vector, respectively, and a total of 40~50 clones were picked for each group for Sanger sequencing. Simultaneously, the *IGF2* locus was also amplified for the Deep-sequencing analysis.

Deep Sequencing Analysis

PCR amplicons of the two experiment groups were amplified, respectively, using different barcode-primer pairs (**Supplementary Table S1**), purified using a gel extraction kit (OMEGA Bio-Tek, China), and sequenced on an Illumina HiSeq (GENEWIZ, China). Among the files provided by GENEWIZ, reads with the sequence CTCaCAGCGCGctAGC (harboring both the desired G > A mutation and the PAM mutations) were considered as the HDR-based editing, and the percentage

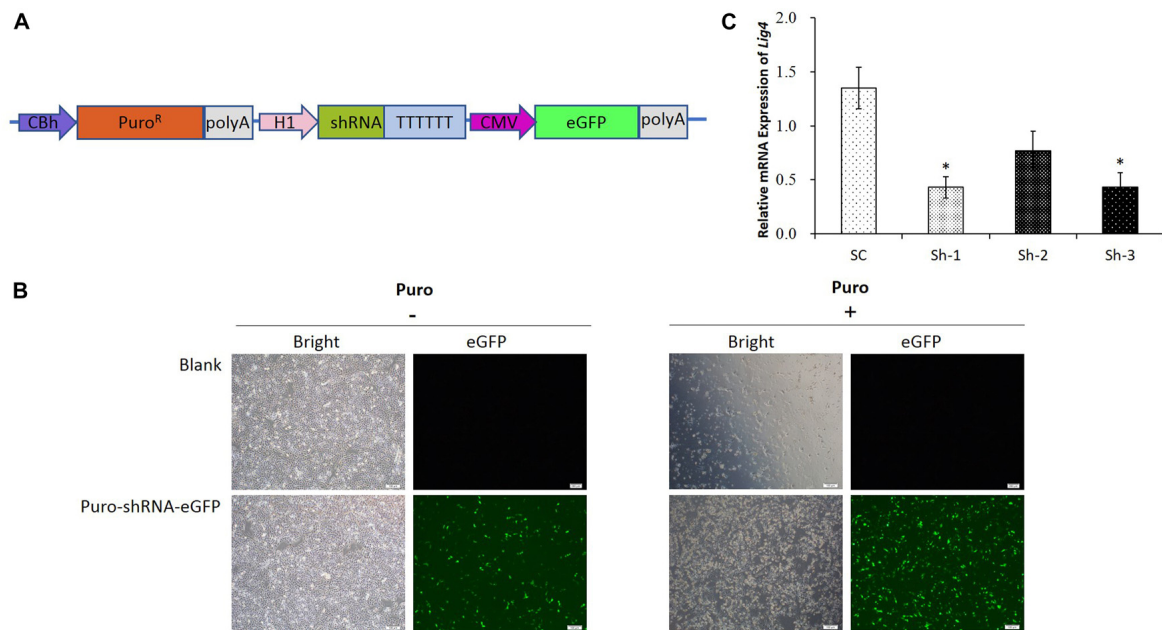


FIGURE 2 | shRNA verification for porcine *LIG4* gene interference. **(A)** The diagrammatic drawing of the *Puro^R*-shRNA-eGFP cassette, which contained the *Puro^R* gene for the puromycin selection of transfected cells, as well as the fluorescent *eGFP* gene for the visualization. **(B)** Representative visualization of the un-transfected and transfected PK15 cells before or after the puromycin selection. The cells transfected with the shRNA vector were enriched significantly after the selection. **(C)** The relative expression of *LIG4* gene down-regulated by different shRNAs. The data was analyzed by Student's *t*-test ($n = 3$, * $P < 0.05$ compared with SC). SC, the non-specific shRNA control.

of HDR reads relative to all reads was calculated as the HDR efficiency for each group. The deep sequencing data are available under the BioProject ID: PRJNA526113.

Statistical Analysis

For the histograms except **Figure 4H**, the data were collected from three independent experiments, and were analyzed by an unpaired and two-tailed *t*-test. Differences were considered statistically significant (*) when $P < 0.05$. Error bars represented the standard error.

RESULTS

StCas9 Showed Similar Activity With SpCas9

We have developed the CRISPR/StCas9 system and the SSA-based DsRed-eGFP (RG) and eGFP surrogate reporters in our previous study (Xu et al., 2015a). Here, our codon-humanized StCas9 was firstly compared with SpCas9 using the dual-fluorescent RG surrogate report (**Figure 1A**). The *DsRed* gene was used as the transfection marker, and the interrupted *eGFP* gene was used as the reporter, which was designed to be repaired accurately by SSA when targeted by the sgRNA/Cas9 complex (Xu et al., 2015a). The HEK293T cells were observed under a fluorescent microscope 2 days after the transfection. Robust red fluorescence and obvious green fluorescence were evidenced within the cells from

both SpCas9 and StCas9 experiment groups (**Figure 1B**). Then the cells were harvested and the DsRed⁺ single and DsRed⁺eGFP⁺ dual positive cells were counted by flow cytometric analysis (**Figure 1C**). The percentage of the DsRed⁺eGFP⁺ cells was calculated to evaluate the Cas9 activity, and our StCas9 demonstrated a comparable activity (20.85%) with the prevalently used SpCas9 (17.65%) (**Figure 1D**, $P = 0.221$).

shRNA Verification for Porcine LIG4 Gene Interference

Since the transfection efficiency for PK15 cells is limited, we constructed the upgraded shRNA vectors (**Figure 2A**), which contained the *Puro^R* gene for the puromycin selection of the transfected cells, as well as the fluorescent *eGFP* gene for the visualization. The representative pictures for the un-transfected and transfected cells were shown in **Figure 2B**, which demonstrated that the PK15 cells transfected with the shRNA vector were enriched significantly after the puromycin selection. The results of the qRT-PCR analysis further confirmed that the relative expression of *LIG4* gene was declined about 60% by two of the shRNAs (Sh-1/3, **Figure 2C**, * $P < 0.05$ compared with SC).

Functional Assay of the sgRNA-shRNA Structure

The sgRNA-shRNA structure was constructed with two identical sgRNAs targeting the *IGF2* gene and one shRNA

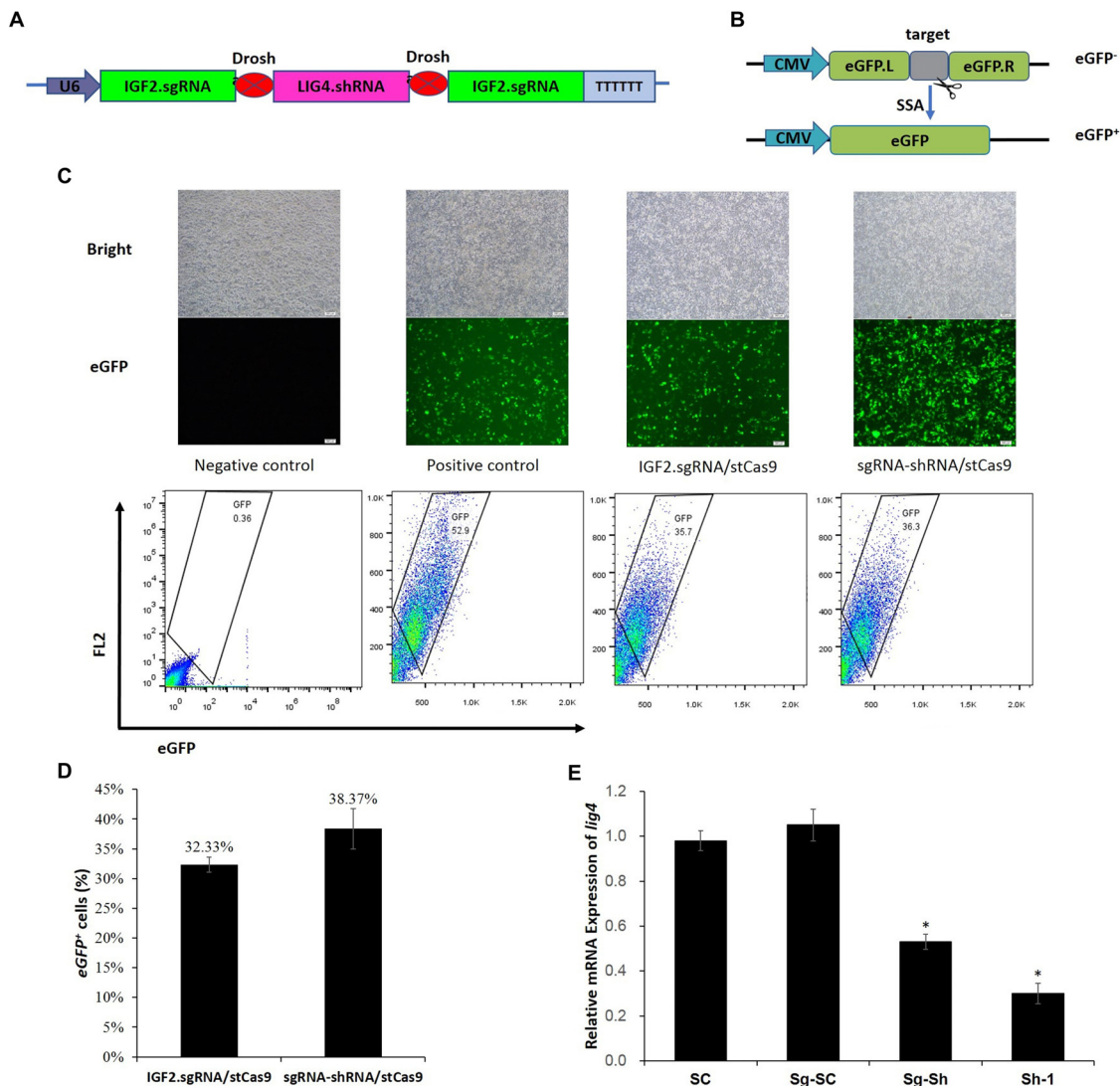


FIGURE 3 | sgRNA-shRNA structure design and activity verification. **(A)** The sgRNA-shRNA structure was designed with two identical sgRNAs targeting the *IGF2* gene and one shRNA against the *LIG4* gene as we previously did (Yan et al., 2016). **(B)** The schematic of the SSA-based single-fluorescent eGFP surrogate reporter. The *DsRed* expression cassette was removed from the dual-fluorescent *DsRed*-eGFP (RG) surrogate reporter to avoid the interference of the robust red fluorescence on the green fluorescence reporter. **(C)** Representative pictures and flow cytometric counting results of eGFP⁺ cells. Negative control: cells transfected with IGF2.eGFP reporter and single StCas9 expression vector; Positive control: cells transfected with linearized IGF2.eGFP reporter and single StCas9 expression vector. **(D)** Comparison of the sgRNA activities driven by IGF2.sgRNA/StCas9 and sgRNA-shRNA/StCas9 ($n = 3$, $P = 0.265$). The percentage of eGFP⁺ cells was used as an indirect measurement for the IGF2.sgRNA activity. **(E)** The relative expression of *LIG4* gene down-regulated by different shRNAs or structures ($n = 3$, $*P < 0.05$ compared with SC). SC, the non-specific shRNA control; Sg-SC, sgRNA-shRNA structure with non-specific shRNA control; Sg-Sh, sgRNA-shRNA structure with *LIG4*.shRNA-1; Sh-1, *LIG4*.shRNA-1.

against the *LIG4* gene (Figure 3A) as we previously did (Yan et al., 2016). The adjacent sgRNA and shRNA was linked by the optimized Drosha cutting sequences (Supplementary Figure S3).

Alternatively, the surrogate reporter assay was conducted using the single-fluorescent IGF2.eGFP reporter (Figure 3B) for verifying the sgRNA activity. The linearized IGF2.eGFP surrogate reporter, which was supposed to repair spontaneously in cells after the transfection, was used as the positive control to co-transfect the cells with the single StCas9 expression vector.

The representative pictures of fluorescent cells and the flow cytometric counting results for different experiment groups were shown in Figure 3C. The percentage of eGFP⁺ cells was used as an indirect measurement for the IGF2.sgRNA activity. As shown in Figure 3D, the IGF2.sgRNA driven by the sgRNA-shRNA/StCas9 all-in-one expression vector (38.37%) demonstrated similar activity with that driven by the IGF2.sgRNA/StCas9 vector (32.33%, $P = 0.265$). To further verify the *LIG4*.shRNA activity driven by the sgRNA-shRNA structure, the qRT-PCR analysis was

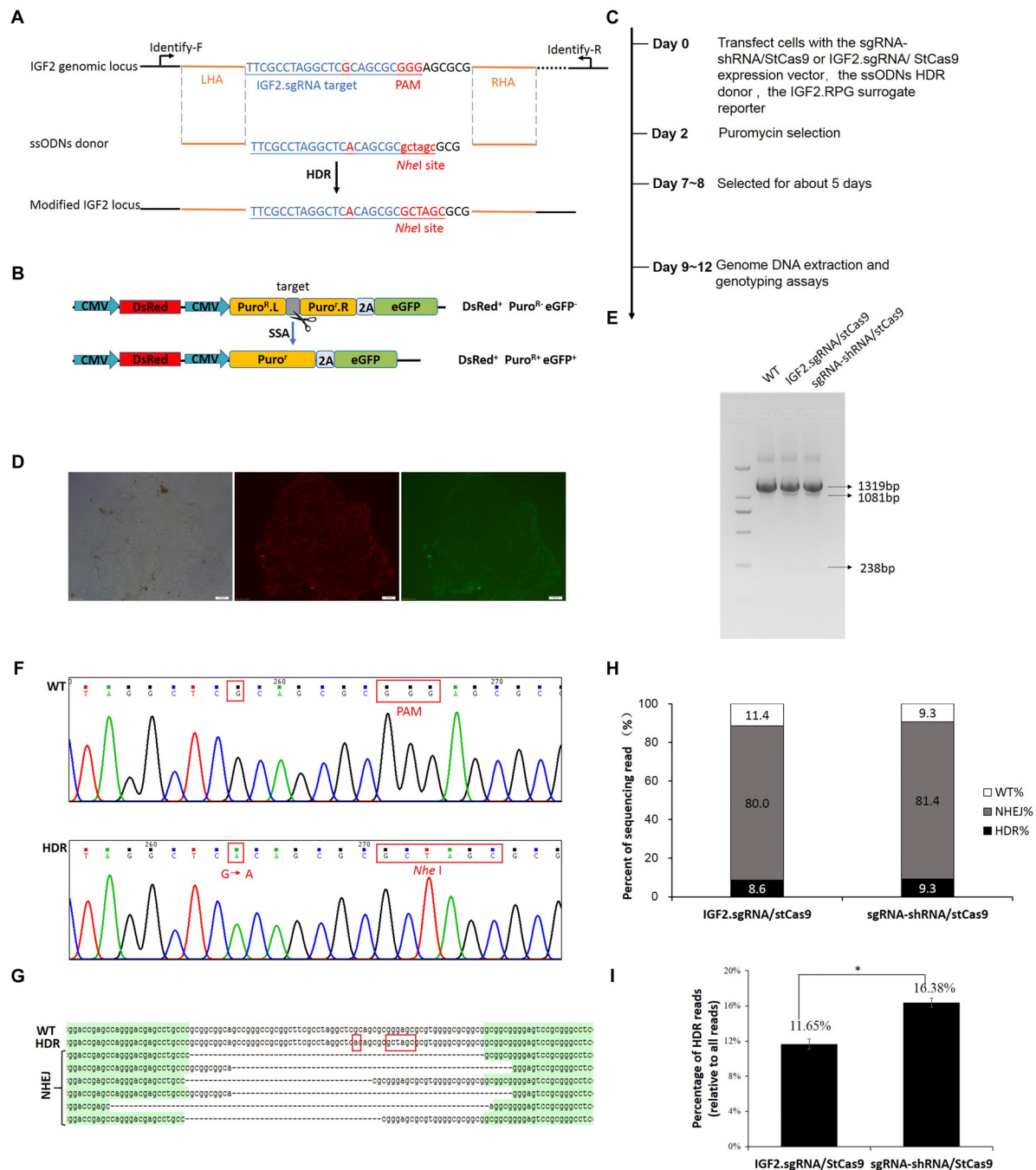


FIGURE 4 | Efficient *IGF2* gene editing by sgRNA-shRNA/StCas9. **(A)** Schematic diagram for the HDR-based *IGF2* gene editing. The ssODNs donor was designed with the desired G > A substitution, as well as the mutations within the PAM motif. The *NheI* RE site was introduced for the subsequent RFLP assay. **(B)** The schematic of the SSA-based DsRed-Puro^R-eGFP (RPG) surrogate reporter, which was used for the enrichment and screening of the targeted positive cells by puromycin selection (Ren et al., 2015). **(C)** Overview of the procedure for *IGF2* gene editing. **(D)** Representative porcine PK15 cell clone screened by puromycin selection. The green and red fluorescence were generated by the expression of the *DsRed* marker gene and the restored eGFP reporter gene within the RPG surrogate reporter. **(E)** RFLP assay for confirming the editing events at the *IGF2* locus. For the RFLP assay, the PCR products (1319 bp) of the *IGF2* locus would be cut into two fragments (1081 and 238 bp) by *NheI* when the *IGF2* gene was edited successfully as designed. **(F)** Representative chromatograms from the Sanger sequencing analysis for confirming the HDR-based editing events. The designed G > A substitution and PAM mutation (Red border) were introduced successfully. **(G)** Representative sequences of the *IGF2* locus with the HDR-based or NHEJ-based edits from the Sanger sequencing analysis. **(H)** Statistical analysis of the HDR-based and NHEJ-based edits from the Sanger sequencing analysis. **(I)** Statistical analysis of the reads with the HDR-based edits from the Deep-sequencing results. The data was analyzed by Student's *t*-test ($n = 3$, $*P < 0.05$).

performed with both the SC (non-specific shRNA control) and Sg-SC (sgRNA-shRNA with non-specific shRNA) negative controls. The results suggested that the relative expression of *LIG4* gene was declined about 50% by the sgRNA-shRNA structure with *LIG4.shRNA* (Sg-Sh, **Figure 3E**, $*P < 0.05$ compared with SC).

Efficient IGF2 Gene Editing by sgRNA-shRNA/StCas9

The IGF2.RPG surrogate reporter was constructed and used for the selection of the genetically modified positive cells as we previously reported (Ren et al., 2015; Shao et al., 2017). PK15 cells were co-transfected with the sgRNA-shRNA/StCas9 or the IGF2.sgRNA/StCas9 expression vector and the ssODNs HDR donor (**Figure 4A**), along with the IGF2.RPG surrogate reporter (**Figure 4B**). After the puromycin selection (**Figure 4C**), the resistant positive cell clones (**Figure 4D**) were pooled and the genomic DNA was extracted for the subsequent RFLP and Sanger sequencing analyses. For the RFLP assay, the PCR products (1319 bp) of the *IGF2* locus would be cut into two fragments (1081 and 238 bp) by *NheI* when the *IGF2* gene was edited successfully as designed (**Figures 4A,E**). The editing events were firstly confirmed by the results of both the RFLP assay (**Figure 4E**) and Sanger sequencing (**Figure 4F**). The Sanger sequencing results of the “T-A clones” for the *IGF2* target locus (**Figure 4G**) further demonstrated 9.3% (4/43) HDR-based and 81.4% (35/43) NHEJ-based repair efficiencies driven by sgRNA-shRNA/StCas9, while the HDR-based and NHEJ-based repair efficiencies for the IGF2.sgRNA/StCas9 group were 8.6% (3/35) and 80.0% (28/35), respectively (**Figure 4H**). As limited clones were sequenced by Sanger sequencing, which may not reveal the real difference between the two experiment groups, we further conducted the Deep-sequencing analysis and the results finally demonstrated a significantly higher HDR-based editing efficiency (16.38%) for the sgRNA-shRNA/StCas9 group (**Figure 4I**, $*P < 0.05$).

DISCUSSION

The HDR-based genome editing holds great promise to the development of safe and highly precise approaches for gene therapy and animal breeding researches (Steyer et al., 2018). Recent years, massive efforts have been made to enhance the CRISPR/Cas9-mediated HDR efficiency. Here, we combined the StCas9 and the novel sgRNA-shRNA structure for boosting the HDR-based “SNP editing” of the porcine *IGF2* gene.

We have developed the *S. thermophilus*-derived CRISPR/StCas9 system for eukaryotic genome editing in our previous study (Xu et al., 2015a). Although the CRISPR/StCas9 system could share the same NGG PAM with the *S. pyogenes*-derived CRISPR/SpCas9 system, it required a stricter NGGNG PAM pattern for full activity (Xu et al., 2015a), which may contribute to reduce the off-target effect compared with the CRISPR/SpCas9 system

(Muller et al., 2016). We found in our study that the key elements of our optimized sgRNA scaffold (Xu et al., 2015a) were almost the same with that of the CRISPR/SpCas9 system (Mali et al., 2013). A series of our subsequent applications also suggested that the two systems may share the same sgRNA structure. However, our StCas9 remained to be compared with the prevalently used SpCas9. To compare the targeting activities of the two Cas9 variants, we used the uniform and further modified sgRNA scaffold (**Supplementary Figure S1**), and the surrogate report assay was conducted using our idiomatic SSA-based RG surrogate reporter. We are glad to see that our StCas9 demonstrated comparable activity with the SpCas9 (**Figure 1D**).

We developed the novel Drosha-mediated sgRNA-shRNA structure mainly for the multiplex genome targeting at the beginning (Yan et al., 2016). However, taking the advantage of the combined sgRNA-shRNA structure, simultaneous genome targeting by sgRNA/Cas9 and transient gene silencing by shRNA could be achieved. Interestingly, we noticed that the by-product shRNA could be used for interfering *LIG4* gene to enhance the HDR-based genome editing. During the multiplex genome targeting assays, we found that the sgRNAs driven by the sgRNA1-shRNA-sgRNA2 structure showed lower activity than the independent sgRNA controls driven by routine sgRNA/Cas9 expression vectors. Hence, we used the sgRNA-LIG4.shRNA-sgRNA structure with identical sgRNAs flanking the *LIG4.shRNA* to guarantee the sgRNA activity for the enhanced HDR-based genome editing (Yan et al., 2016). In this study, we used the same strategy and further applied a pair of more efficient Drosha-processing sequences (Zeng and Cullen, 2005) for the processing of the short RNAs. The surrogate reporter and qRT-PCR assays demonstrated effective sgRNA and shRNA activities driven by sgRNA-shRNA/StCas9 (**Figures 3D,E**). Another point to explain, since we have compared the sgRNA-shRNA structures with *LIG4.shRNA* and non-specific shRNA for enhancing the HDR-based genome editing, this study was designed mainly focusing on the practical question whether our sgRNA-shRNA strategy (with *LIG4.shRNA*) is better than the routine single sgRNA strategy.

It is reported that numerous human genetic diseases were caused by single nucleotide mutation, such as the 878 G > A (*AVPR2* W293X) in X-linked Nephrogenic diabetes insipidus and the 1517 G > A (*FANCC* W506X) in Fanconi anemia (Cox et al., 2017). Moreover, there have been so many SNP sites found in livestock, and most of them are related to animal diseases or the production traits, such as the 3072 G > A in the porcine *IGF2* gene in this study. Thus, the SNP editing provides a good idea for the production of base-edited animals, as well as the gene therapy for the base correction of some genetic diseases.

In view of the unavoidable concern of off-target effect during the genome editing manipulation, our sgRNA-shRNA/StCas9 system still has room for improvement. It has been reported that paired Cas9-nickase (Cas9n) can be used for efficient genome editing with significant decreased off-target events

(Ran et al., 2013). Hence, an improved sgRNA-shRNA/StCas9n system may be a good idea in the further study.

AUTHOR CONTRIBUTIONS

KX and ZZ conceived the research plans. YS, NY, LM, BS, JD, and YF performed the experiments. SS, QY, and FH provided related plasmids. YS, NY, and KX wrote the article. KX provided financial support.

FUNDING

This work was supported by grants from the Shaanxi Natural Science Foundation of China (2017JQ3007), the China Postdoctoral Science Foundation (2018T111111 and

2015M580887), the National Natural Science Foundation of China (NSFC, 31702099) and the National Transgenic Major Project of China (2018ZX08010-09B).

ACKNOWLEDGMENTS

The authors would like to thank Associate Professor Zehui Wei for his help for the statistical analysis of the data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00347/full#supplementary-material>

REFERENCES

- Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G. W., et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479–1491. doi: 10.1016/j.cell.2013.12.001
- Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., et al. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* 33, 543–548. doi: 10.1038/nbt.3198
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi: 10.1126/science.1231143
- Cox, D. B. T., Gootenberg, J. S., Abudayyeh, O. O., Franklin, B., Kellner, M. J., Joung, J., et al. (2017). RNA editing with CRISPR-Cas13. *Science* 358, 1019–1027. doi: 10.1126/science.aag0180
- Hu, Z., Shi, Z., Guo, X., Jiang, B., Wang, G., Luo, D., et al. (2018). Ligase IV inhibitor SCR7 enhances gene editing directed by CRISPR-Cas9 and ssODN in human cancer cells. *Cell Biosci.* 8:12. doi: 10.1186/s13578-018-0200-z
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829
- Kleistiver, B. P., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Topkar, V. V., Zheng, Z., et al. (2015). Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* 33, 1293–1298. doi: 10.1038/nbt.3404
- Komor, A. C., Badran, A. H., and Liu, D. R. (2017). CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* 168, 20–36. doi: 10.1016/j.cell.2016.10.044
- Lin, S., Staahl, B. T., Alla, R. K., and Doudna, J. A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* 3:e04766. doi: 10.7554/eLife.04766
- Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., et al. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826. doi: 10.1126/science.1232033
- Markljung, E., Jiang, L., Jaffe, J. D., Mikkelsen, T. S., Wallerman, O., Larhammar, M., et al. (2009). ZBED6, a novel transcription factor derived from a domesticated DNA transposon regulates IGF2 expression and muscle growth. *PLoS Biol.* 7:e1000256. doi: 10.1371/journal.pbio.1000256
- Maruyama, T., Dougan, S. K., Truttmann, M. C., Bilate, A. M., Ingram, J. R., and Ploegh, H. L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* 33, 538–542. doi: 10.1038/nbt.3190
- Muller, M., Lee, C. M., Gasiunas, G., Davis, T. H., Cradick, T. J., Siksnys, V., et al. (2016). *Streptococcus thermophilus* CRISPR-Cas9 systems enable specific editing of the human genome. *Mol. Ther.* 24, 636–644. doi: 10.1038/mt.2015.218
- Ran, F. A., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., et al. (2015). In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* 520, 186–191. doi: 10.1038/nature14299
- Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., et al. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154, 1380–1389. doi: 10.1016/j.cell.2013.08.021
- Ren, C., Xu, K., Liu, Z., Shen, J., Han, F., Chen, Z., et al. (2015). Dual-reporter surrogate systems for efficient enrichment of genetically modified cells. *Cell Mol. Life Sci.* 72, 2763–2772. doi: 10.1007/s00018-015-1874-6
- Ren, C., Xu, K., Segal, D. J., and Zhang, Z. (2019). Strategies for the enrichment and selection of genetically modified cells. *Trends Biotechnol.* 37, 56–71. doi: 10.1016/j.tibtech.2018.07.017
- Salsman, J., and Delliare, G. (2017). Precision genome editing in the CRISPR era. *Biochem. Cell Biol.* 95, 187–201. doi: 10.1139/bcb-2016-0137
- Shao, S., Ren, C., Liu, Z., Bai, Y., Chen, Z., Wei, Z., et al. (2017). Enhancing CRISPR/Cas9-mediated homology-directed repair in mammalian cells by expressing *Saccharomyces cerevisiae* Rad52. *Int. J. Biochem. Cell Biol.* 92, 43–52. doi: 10.1016/j.biocel.2017.09.012
- Steyer, B., Bu, Q., Cory, E., Jiang, K., Duong, S., Sinha, D., et al. (2018). Scarless genome editing of human pluripotent stem cells via transient puromycin selection. *Stem Cell Rep.* 10, 642–654. doi: 10.1016/j.stemcr.2017.12.004
- Van Laere, A. S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., et al. (2003). A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* 425, 832–836. doi: 10.1038/nature02064
- Xu, K., Ren, C., Liu, Z., Zhang, T., Li, D., Wang, L., et al. (2015a). Efficient genome engineering in eukaryotes using Cas9 from *Streptococcus thermophilus*. *Cell Mol. Life Sci.* 72, 383–399. doi: 10.1007/s00018-014-1679-z
- Xu, K., Zhang, T. T., Guo, L. J., Xin, Y., Zhang, L., and Zhang, Z. Y. (2015b). Alternative methods for the efficient construction of short hairpin RNA expression vectors. *Anal. Biochem.* 478, 23–25. doi: 10.1016/j.ab.2015.03.006
- Yan, Q., Xu, K., Xing, J., Zhang, T., Wang, X., Wei, Z., et al. (2016). Multiplex CRISPR/Cas9-based genome engineering enhanced by Drosha-mediated sgRNA-shRNA structure. *Sci. Rep.* 6:38970. doi: 10.1038/srep38970
- Zeng, Y., and Cullen, B. R. (2005). Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. *J. Biol. Chem.* 280, 27595–27603. doi: 10.1074/jbc.M504714200

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Sun, Yan, Mu, Sun, Deng, Fang, Shao, Yan, Han, Zhang and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Increasing Cytosine Base Editing Scope and Efficiency With Engineered Cas9-PmCDA1 Fusions and the Modified sgRNA in Rice

Ying Wu, Wen Xu, Feipeng Wang, Si Zhao, Feng Feng, Jinling Song, Chengwei Zhang* and Jinxiao Yang*

Beijing Key Laboratory of Maize DNA Fingerprinting and Molecular Breeding, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China

OPEN ACCESS

Edited by:

Kun Xu,
Northwest A&F University, China

Reviewed by:

Sachin Rustgi,
Clemson University, United States
Cem Kuscü,
The University of Tennessee Health
Science Center (UTHSC),
United States

*Correspondence:

Chengwei Zhang
zhangchengwei2017@126.com
Jinxiao Yang
yangjinxiao@maizedna.org

Specialty section:

This article was submitted to
Genomic Assay Technology,
a section of the journal
Frontiers in Genetics

Received: 16 November 2018

Accepted: 09 April 2019

Published: 26 April 2019

Citation:

Wu Y, Xu W, Wang F, Zhao S,
Feng F, Song J, Zhang C and Yang J
(2019) Increasing Cytosine Base
Editing Scope and Efficiency With
Engineered Cas9-PmCDA1 Fusions
and the Modified sgRNA in Rice.
Front. Genet. 10:379.
doi: 10.3389/fgene.2019.00379

Base editors that do not require double-stranded DNA cleavage or homology-directed repair enable higher efficiency and cleaner substitution of targeted single nucleotides in genomic DNA than conventional approaches. However, their broad applications are limited within the editing window of several base pairs from the canonical NGG protospacer adjacent motif (PAM) sequence. In this study, we fused the D10A nickase of several *Streptococcus pyogenes* Cas9 (SpCas9) variants with *Petromyzon marinus* cytidine deaminase 1 (PmCDA1) and uracil DNA glycosylase inhibitor (UGI) and developed two new effective PmCDA1-based cytosine base editors (pBEs), SpCas9 nickase (SpCas9n)-pBE and VQR nickase (VQRn)-pBE, which expanded the scope of genome targeting for cytosine-to-thymine (C-to-T) substitutions in rice. Four of six and 12 of 18 target sites selected randomly in SpCas9n-pBE and VQRn-pBE, respectively were base edited with frequencies of 4–90% in T_0 plants. The effective deaminase window typically spanned positions 1–7 within the protospacer and the single target C showed the maximum C-to-T frequency at or near position 3, counting the end distal to PAM as position 1. In addition, the modified single guide RNA (sgRNA) improved the base editing efficiencies of VQRn-pBE with 1.3- to 7.6-fold increases compared with the native sgRNA, and targets that could not be mutated using the native sgRNA were edited successfully using the modified sgRNA. These newly developed base editors can be used to realize C-to-T substitutions and may become powerful tools for both basic scientific research and crop breeding in rice.

Keywords: base editing, cytosine base editor, SpCas9, VQR, the modified sgRNA, rice

Abbreviations: Cas9, CRISPR-associated protein 9; CBE, cytosine base editor; CRISPR, clustered regularly interspaced short palindromic repeats; C-to-T, cytosine-to-thymine; EQR, engineered SpCas9 variant with residue substitutions D1135E/R1335Q/T1337R; Indel, insertion and deletion; PAM, protospacer adjacent motif; pBE, PmCDA1-based cytosine base editor; PmCDA1, *Petromyzon marinus* cytidine deaminase 1; rAPOBEC1, rat cytidine deaminase APOBEC1; SaCas9, *Staphylococcus aureus* Cas9; SaCas9n, SaCas9 nickase; SaKKH, engineered SaCas9 variant with residue substitutions E782K/N968K/R105H; SaKKHn, SaKKH nickase; sgRNA, single guide RNA; SpCas9, *Streptococcus pyogenes* Cas9; SpCas9n, SpCas9 nickase; SpCas9-NG, engineered SpCas9 variant with residue substitutions R1335V/L1111R/D1135V/G1218R/E1219F/A1322R/T1337R; SpCas9n-NG, SpCas9-NG nickase; tRNA, transfer RNA; UGI, uracil DNA glycosylase inhibitor; VQR, engineered SpCas9 variant with residue substitutions D1135V/R1335Q/T1337R; VQRn, VQR nickase; VRER, engineered SpCas9 variant with residue substitutions D1135V/G1218R/R1335E/T1337R; VRERn, VRER nickase; xCas9, engineered SpCas9 variant with residue substitutions E480K/E543D/E1219V.

INTRODUCTION

Genome-wide association studies have shown that point mutations create elite trait variations in crop plants, and point mutagenesis is one of the main strategies for crop improvement (Henikoff and Comai, 2003; Zhao et al., 2011; Yin et al., 2017). The discovery and development of the CRISPR – Cas9 system (Doudna and Charpentier, 2014; Hsu et al., 2014; Shalem et al., 2015; Wang et al., 2016; Komor et al., 2017a) has provided a powerful genome engineering tool for generating point mutations in plants through precise irreversible base conversion (base editing) without the need for double-stranded DNA backbone cleavages or donor DNA templates (Komor et al., 2016; Nishida et al., 2016). Base editing is much cleaner and more efficient than current methods used in plants [e.g., targeting induced local lesions in genomes (TILLING) and conventional nuclease-mediated, homology-directed repair (HDR)-dependent genome editing] (Henikoff et al., 2004; Slade et al., 2005; Hess et al., 2017; Yang et al., 2017; Kim, 2018).

The first reported CBEs that mediate C-to-T conversion were developed in a wide variety of organisms by fusion of Cas9n with rat cytidine deaminase rAPOBEC1 or activation-induced cytidine deaminase ortholog PmCDA1 (Lu and Zhu, 2017; Li et al., 2017; Ren et al., 2017; Shimatani et al., 2017; Zong et al., 2017). Although highly efficient and useful, these CBEs were restricted to edit sites that contained NGG PAM sequences because of the common SpCas9n that was used (Anders et al., 2014; Nishimasu et al., 2014). This characteristic limited the base editing to a narrow window of several base pairs from the PAM distal region. To circumvent this limitation, several studies have reported new CBEs that use SpCas9 variants or Cas9 homologs that recognize expanded or altered PAMs to increase the targets suitable for base editing. In human cells, several engineered SpCas9 variants that accept NGA (VQR), NGCG (VRER), NGAG (EQR), or NG (xCas9 and SpCas9-NG) PAM sequences have been employed with rAPOBEC1 or activation-induced cytidine deaminase to generate new CBEs (Kim Y.B. et al., 2017; Hu et al., 2018; Nishimasu et al., 2018). In addition, the SaCas9, which recognizes the NNGRRT PAM, and its engineered variant SaKKH, which recognizes the NNNRRT PAM sequence, also have been used to create base editors that expand the editing capability of CBEs (Kim Y.B. et al., 2017).

Most of the Cas9s described above have been used to create new CBEs for plants; the exceptions are VRER and EQR (Hua et al., 2018; Qin et al., 2018; Endo et al., 2019). In addition, wild type SpCas9 was used to broaden the base editing targets for non-canonical NAG PAMs in plants (Hua et al., 2018). Among them, SpCas9n-NG, SaCas9n, and SaKKHn CBEs, both rAPOBEC1-based and activation-induced cytidine deaminase-based or PmCDA1-based, were successfully developed in rice (Qin et al., 2018; Endo et al., 2019). However, only rAPOBEC1-based CBEs were created with SpCas9n and VQRn, and only one editable target site was reported for each CBE (Hua et al., 2018). In this study, to better utilize SpCas9 and VQR to enlarge the base editing

scope in rice, we developed two new effective PmCDA1-based CBEs (pBEs), SpCas9n-pBE, and VQRn-pBE. These two pBEs substantially broaden the target sites from those with NGG PAMs to those with NAG and NGA PAMs. Additionally, the editing efficiency of VQRn-pBE was further increased using the modified sgRNA.

MATERIALS AND METHODS

Plasmid Construction

We modified the pCambia2300 plasmid to construct a vector called 2300-Spe. A schematic illustration of 2300-Spe vector construction is given in **Supplementary Figure S1**. Four fragments were digested at each end by restriction endonucleases to construct the SpCas9n-pBE-basic vector (**Supplementary Figure S2**). Then the four digested fragments together with the KpnI and SbfI digested 2300-Spe backbone were ligated using T4 ligase (NEB, Cat# M0202L) to generate SpCas9n-pBE-basic. Based on the SpCas9n-pBE-basic vector, specific point mutations described by Kleinstiver et al. (2015) were introduced into SpCas9n (D10A) using a Fast MultiSite Mutagenesis System (TransGen Biotech, Beijing, China) to generate VQRn-pBE-basic and VRERn-pBE-basic vectors. Target sequences were cloned before the sgRNA using BsaI according to Xie et al. (2015) to generate pBE constructions. The modified sgRNA linked with tRNA and the *Oryza sativa* U3 (OsU3) terminator was synthesized and digested with BamHI and HindIII, and used to replace the native sgRNA in the SpCas9n-pBE and VQRn-pBE constructions to obtain the corresponding pBEs with the modified sgRNA. Target sites in the same constructs are shown in **Supplementary Table S1**. The primers used in this study are listed in **Supplementary Table S2**.

Rice Transformation

The wild type *Agrobacterium tumefaciens* strain LBA4404 (Weidi Biotech, Shanghai, China) was transformed by the resultant pBE constructs using a freeze/thaw method. Embryogenic calli induced from mature seeds of rice variety Nipponbare (*O. sativa* L. japonica. cv. Nipponbare) were used for the transformation, which was conducted as previously described (Hiei and Komari, 2008). After incubation with *Agrobacterium* for 10 min, the calli were recovered for 3 days and selected on 50 µg/ml hygromycin for 4 weeks to obtain resistant calli. Then, the resistant calli were transferred to regeneration medium (not containing hygromycin) to induce shoot regeneration for 1 month. When the shoots were 4–5 cm long, they were transferred to rooting medium for root induction for about 2 weeks to obtain T₀ plants.

DNA Extraction and Identification of Transgenic Resistant Calli and T₀ Plants

Resistant calli and T₀ plants were harvested for genomic DNA extraction using a DNA-quick Plant System kit (Tiagen Biotech, Beijing, China). The target locus was amplified by PCR with Cas9 specific primers (**Supplementary Table S2**) and samples with a 1150-bp nucleic acid band in agarose gel electrophoresis were identified as transgenic resistant calli or T₀ plants.

Mutant Identification

Several transgenic resistant calli and T_0 plants in a single experiment were used to detect C-to-T conversions and indels. Target loci were amplified by specific primers and the PCR products were purified using an EasyPure PCR Purification Kit (TransGen Biotech). The PCR products were sent for Sanger sequencing (Tsingke Biological Technology, Beijing, China) to detect mutations. C-to-T frequency in calli or T_0 plants was defined as the percentage of mutants with any target C-to-T substitution among all the transgenic samples. Indel frequency was defined as the percentage of mutants with any indels among the resulting C-to-T mutants. Single C-to-T frequency was defined as the percentage of mutants with C-to-T substitution at a specific single position among all the transgenic samples. Homozygous mutants were designated when all the mutations were homozygous. Frequency of mutant genotype was defined as the percentage of mutants with the same genotype among all the mutants.

Detection of Off-Target Mutations

Five to eight single T_0 plants, including base mutated lines and wild type lines, were selected for each off-target site detection. Potential off-target sites were searched on Cas-OFFinder (Bae et al., 2014) and amplified using the primers listed in **Supplementary Table S2**. The PCR products were purified using an EasyPure PCR Purification Kit (TransGen Biotech) and sent for Sanger sequencing (Tsingke Biological Technology) to detect off-target mutations.

RESULTS

SpCas9n-pBE Enables Base Editing at NAG PAM Target Sites in Rice

Previous studies revealed that the most widely used wild type SpCas9 enables efficient genome editing at target sites bearing both the canonical NGG PAM and the non-canonical NAG PAM in rice (Meng et al., 2018). Because the combination of PmCDA1 with SpCas9n leads to C-to-T substitutions at targets with the NGG PAM (Shimatani et al., 2017), we hypothesized that a SpCas9n base editor also could function at targets with the NAG PAM. We fused the D10A nickase of SpCas9 with PmCDA1 and UGI. The fusion protein was driven by the *O. sativa* ubiquitin (*OsUbiq*) promoter and the corresponding cassette was introduced into our tRNA-sgRNA editing system to generate a pBE designated as SpCas9n-pBE (**Figure 1A**).

We first used a resistant rice calli system to determine the feasibility of C-to-T base editing using SpCas9n-pBE. Six targets with NAG PAMs from the *OsWaxy* gene, which encodes an enzyme essential in the biosynthesis of granule-bound starch, were selected (**Supplementary Table S3**). In three of the six targets, C-to-T base editing was detected with frequencies of 13.3–60% (**Figure 1B**). No indels were detected at any of the six on-target loci (**Supplementary Table S3**). Moreover, by analyzing the C-to-T frequency at each single C

in the three edited targets, we found that the editing window spanned bases at positions 1 to 6 upstream of the PAM sequence (**Figure 1C**).

To further assess the use of SpCas9n-pBE in rice plants, the resistant calli were transferred to a regeneration culture to generate stable transgenic T_0 plants. In T_0 plants, four of the six target sites had C-to-T substitutions with frequencies of 7.7–53.8% (**Figure 1B**). Three of the target sites (W-T1, W-T3, and W-T6) were edited in both T_0 plants and calli, whereas W-T2 was edited only in T_0 plants with a frequency of 7.7% (**Figure 1B**). Among the four edited sites, indel was detected only at the W-T1 site in T_0 plants (**Supplementary Table S4**). Except for the edited positions 13 and 12 in targets W-T3 and W-T6, respectively, the deamination window in T_0 plants was consistent with that in the resistant calli (**Figures 1C,D**). Single and double-base substitutions were predominant in the edited targets in T_0 plants. Triple or quadruple-base substitutions also were obtained for targets W-T6 and W-T1 (**Supplementary Table S4**). Furthermore, SpCas9n-pBE was able to be used for multiplex genome editing because two or three target sites were edited simultaneously in the same T_0 plant line (**Supplementary Table S5**). Taken together, our results indicated that SpCas9n-pBE could broaden PAM recognition from NGG to NAG in rice.

VQRn-pBE Enables Base Editing at NGA PAM Target Sites in Rice

With different PAM specificities compared with SpCas9, VRER for NGCG PAMs, and VQR for NGA PAMs were reported to enable efficient genome editing of endogenous genes in zebrafish, human cells, and rice (Kleinstiver et al., 2015; Hu et al., 2016). To further enlarge the scope of C-to-T base editing in rice, we engineered two SpCas9 variants, VRER and VQR, then individually fused them with PmCDA1 and UGI to generate VRERn-pBE and VQRn-pBE (**Figure 2A**).

Eleven targets with NGCG PAMs were selected for VRERn-pBE editing, but the sequencing results showed none of them had C-to-T mutations in the resistant calli (**Supplementary Table S6**), implying VRERn-pBE had poor base editing activity in rice.

Because VQR mediated knockout mutations with a preference for NGAG > NGAT = NGAA in human cells (Kleinstiver et al., 2015), we tested the editing efficiency of VQRn-pBE for targets with NGAG PAMs. Four targets with NGAG PAMs from the *OsWaxy* gene were selected (**Supplementary Table S7**), and three of them were mutated successfully with frequencies of 5.7–77.1% in resistant calli and 10–90% in T_0 plants (**Figure 2B**). The deamination windows spanned positions 1 to 5 of the protospacers, counting from the 5' end of the target (**Figure 2C**). Indels were detected in two targets with relatively high substitution frequencies, and the corresponding indel frequencies were higher in T_0 plants than in calli (**Figure 2D**). Single or double C conversions were the most common genotypes (**Figure 2E**). Among all the T_0 base edited mutants, one homozygous mutant for W-T8 and three for W-T9 were obtained (**Figure 2E**). These results indicate VQRn-pBE could be used to enlarge the scope of base editing in rice.

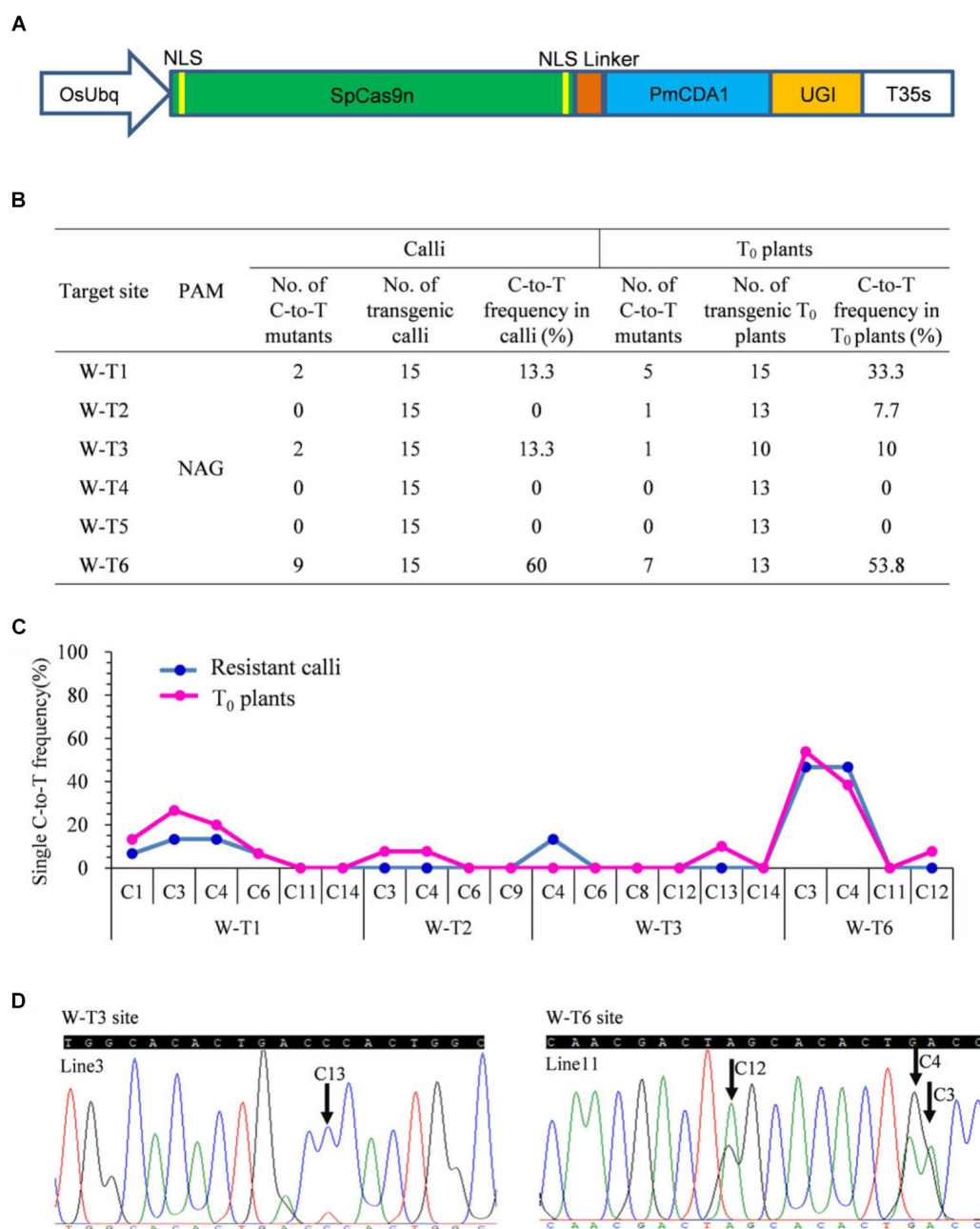


FIGURE 1 | Base editing ability of SpCas9n-pBE in rice. **(A)** Schematic illustration of the SpCas9n-pBE construct. NSL, nuclear localization signal. **(B)** Frequencies of mutations induced by SpCas9n-pBE in resistant calli and T₀ plants. **(C)** Frequencies of targeted single C-to-T substitutions in the targets edited by SpCas9n-pBE in resistant calli and T₀ plants. **(D)** Sequencing chromatograms of T₀ plants at the W-T3 and W-T6 sites of T₀ transgenic lines 3 and 11, with the editing window extending positions 13 and 12 within the protospacer, respectively. Arrows indicate the edited bases. G-to-A conversions in the opposite strand are shown for the W-T6 target of line 11.

We also detected the off-target effects of VQRn-pBE. Potential off-target sites that contained three to five mismatches with targets W-T7, W-T8, and W-T9 were chosen for the analysis (**Supplementary Table S8**). All these sites in T₀ plants were sequenced and no mutations were detected among any of the selected off-target sites (data not shown).

To avoid gene restriction and to confirm the capability of VQRn-pBE for base editing in rice, we used the *O. sativa* acetolactate synthase gene (*OsALS*), which encodes an essential enzyme in the biosynthesis of branched-chain amino acids. Six target sites with NGAG PAMs were selected (**Supplementary Table S9**). Among all the regenerated T₀ events, five of the six sites were base edited with frequencies of 10–80%; the exception

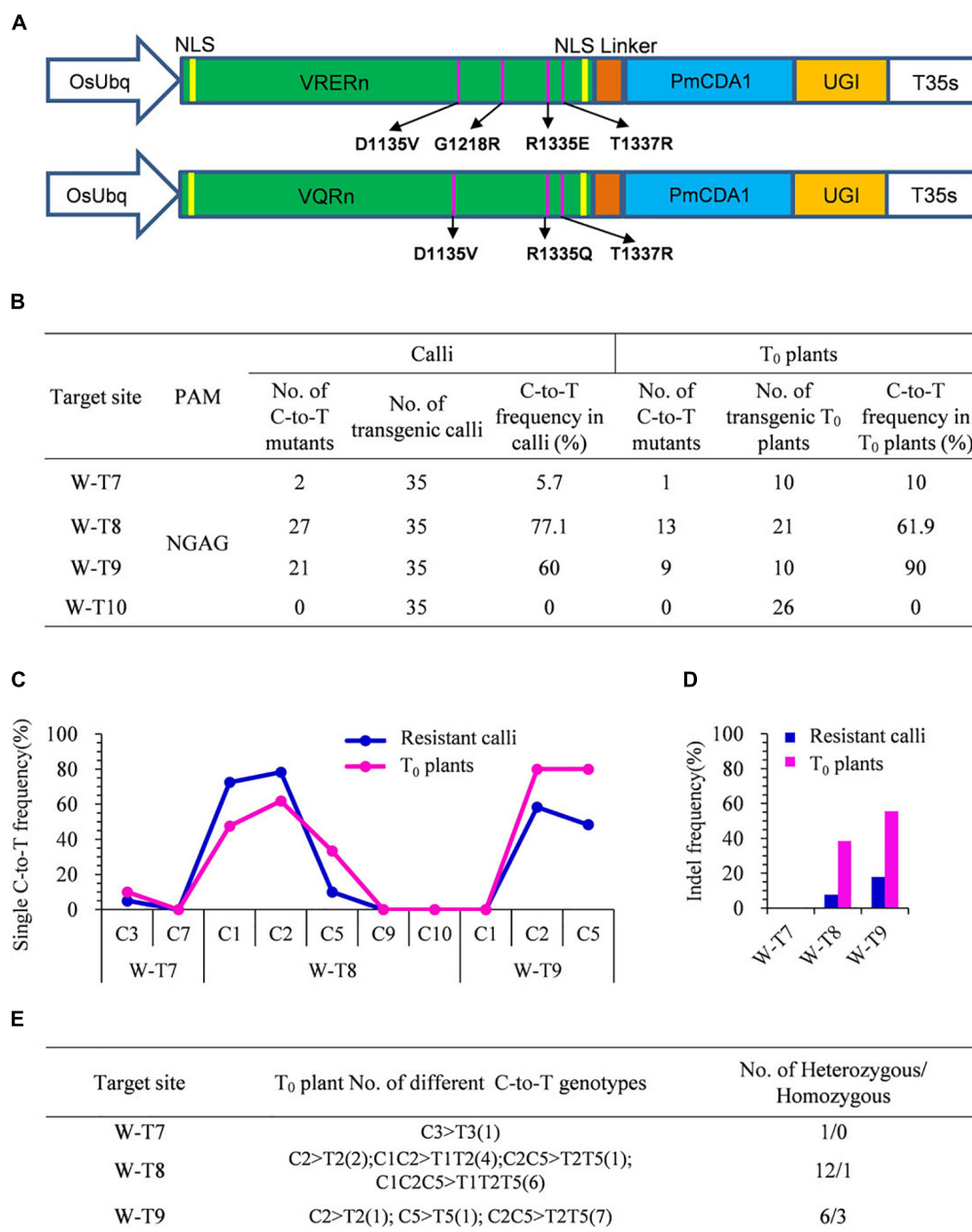


FIGURE 2 | Base editing ability of VQRn-pBE at target sites on the *OsWaxy* gene in rice. **(A)** Schematic illustration of the VRERn-pBE (above) and VQRn-pBE (below) constructs. VQRn and VRERn were generated from SpCas9n with residue substitutions D1135V/R1335Q/T1337R and D1135V/G1218R/R1335E/T1337R, respectively. **(B)** Frequencies of mutations induced by VQRn-pBE at NGAG PAM target sites in resistant calli and T₀ plants. **(C)** Frequencies of targeted single C-to-T substitutions in the targets edited by VQRn-pBE at NGAG PAM sites in resistant calli and T₀ plants. **(D)** Indel frequencies in the targets edited by VQRn-pBE at NGAG PAM target sites in resistant calli and T₀ plants. **(E)** Mutations induced by VQRn-pBE at NGAG PAM sites in T₀ plants.

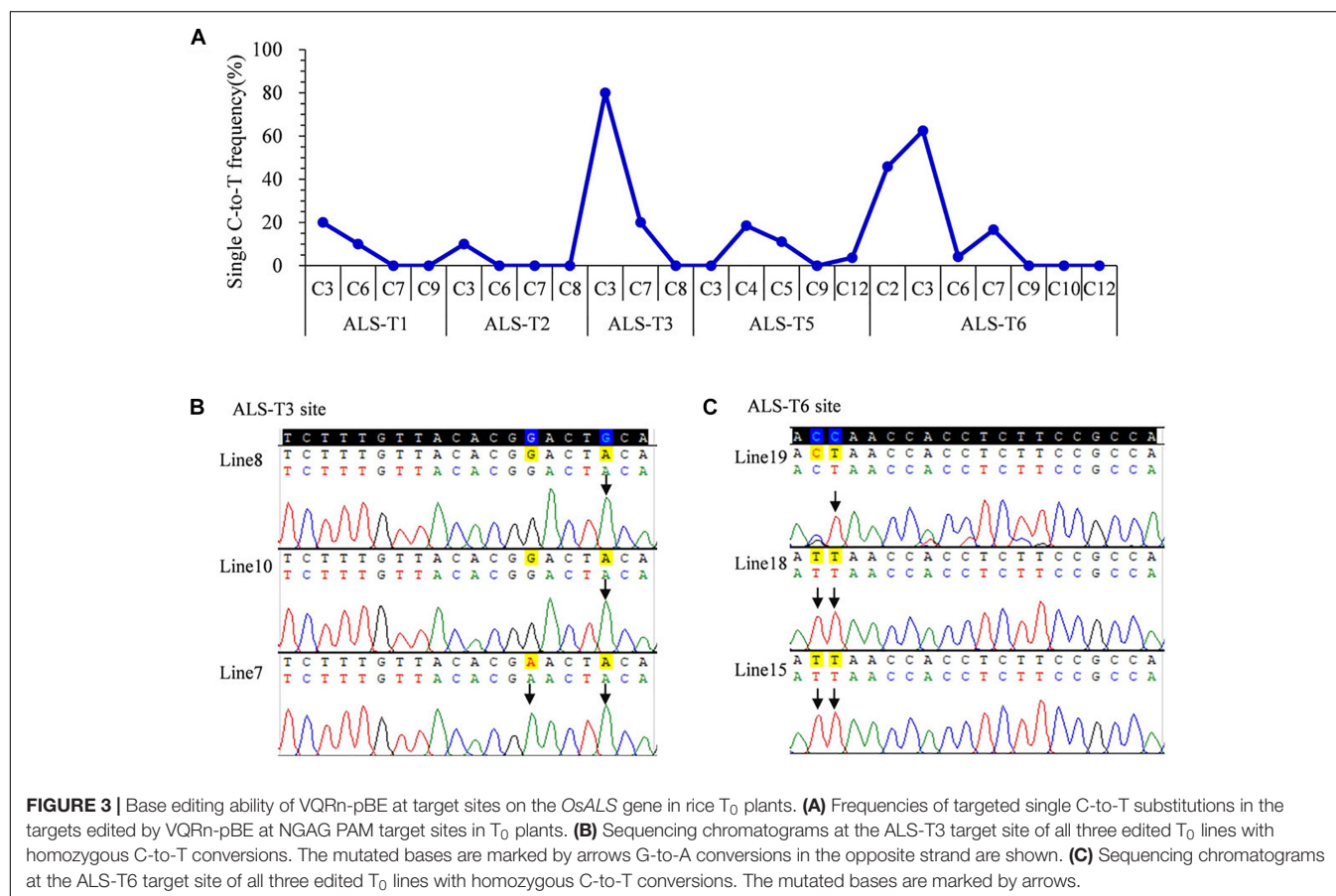
was the ALS-T4 site, which was not edited (Table 1). Indels were identified only in two targets with relatively high editing efficiencies (80 and 70.8%) (Table 1), similar to the results for the *OsWaxy* gene target sites. The effective deamination window typically spanned positions 2–7 within the protospacer, and the frequency of single C-to-T conversion was highest at or near position 3 (Figure 3A). Single C-to-T mutants were detected in

all edited sites, and double C-to-T mutants were more than triple or quadruple mutants (Table 1). Additionally, three mutant lines contained homozygous substitutions were obtained in the ALS-T3 (C3 > T3 and C3C7 > T3T7) and ALS-T6 (C3 > T3 and C2C3 > T2T3) sites (Figures 3B,C and Table 1).

Because VQRn-pBE mediated efficient base editing at sites that contained NGAG PAMs in rice, we determined whether it

TABLE 1 | Frequencies of mutations induced by VQRn-pBE at target sites on the OsALS gene in rice T₀ plants.

Target site	No. of C-to-T mutants	No. of transgenic T ₀ plants	C-to-T frequency (%)	Indel frequency (%)	Genotype of C-to-T mutations	Heterozygous/homozygous
ALS-T1	2	10	20	0	C3 > T3(1); C3C6 > T3T6(1)	2/0
ALS-T2	1	10	10	0	C3 > T3(1)	1/0
ALS-T3	8	10	80	37.5	C3 > T3(6); C3C7 > T3T7(2)	5/3
ALS-T4	0	27	0	0		
ALS-T5	7	27	25.9	0	C4 > T4(3); C5 > T5(1); C12 > T12(1); C4C5 > T4T5(2)	7/0
ALS-T6	17	24	70.8	23.5	C2 > T2(1); C3 > T3(5); C7 > T7(1); C2C3 > T2T3(7); C2C3C7 > T2T3T7(2); C2C3C6C7 > T2T3T6T7(1)	14/3



worked on targets with NGAT, NGAC, or NGAA PAMs. We selected 12 endogenous genomic target sites in the *OsWaxy* gene (Supplementary Table S10). We found that VQRn-pBE produced more C-to-T editing at target sites with the NGAT PAM than at target sites with the NGAC PAM in T₀ plants, although the editing efficiencies were much lower in the former (Table 2). Similar to the results for the NGAG PAM target sites, VQRn-pBE produced indels only at the NGAC PAM W-T15 site, which had a relatively high base editing frequency (41.2% in T₀ plants) (Supplementary Table S11). Single and double C-to-T mutants accounted for most of the detected mutants

(Supplementary Table S11). VQRn-pBE produced no C-to-T editing at the four sites with the NGAA PAM in both calli and T₀ plants (data not shown).

The Modified sgRNA Increases the Base Editing Efficiency of VQRn-pBE

Several studies have reported that modified sgRNAs with a mutation in the streak of T and an extended duplex increased editing efficiency in mammalian cells and rice (Chen et al., 2013; Dang et al., 2015; Hu et al., 2017). To try to enhance

TABLE 2 | Frequencies of mutations induced by VQRn-pBE at target sites on the *OsWaxy* gene with NGAT and NGAC PAMs in rice resistant calli and T₀ plants.

Target site	PAM	Calli			T ₀ plants		
		No. of C-to-T mutants	No. of transgenic calli	C-to-T frequency in calli (%)	No. of C-to-T mutants	No. of transgenic T ₀ plants	C-to-T frequency in T ₀ plants (%)
W-T11	NGAT	2	15	13.3	1	25	4
W-T12		2	15	13.3	1	25	4
W-T13		0	15	0	1	17	5.9
W-T14		0	15	0	0	14	0
W-T15	NGAC	10	15	66.7	7	17	41.2
W-T16		0	15	0	0	17	0
W-T17		0	15	0	0	17	0
W-T18		0	15	0	0	16	0

the C-to-T substitution frequency, we modified the sgRNA as described previously in rice (Hu et al., 2017). We replaced the fourth T in the streak of T with C and extended the duplex by 5 bp (**Supplementary Figure S3**). Then, we used all 22 target sites in the *OsWaxy* gene with SpCas9n-pBE and VQRn-pBE with the modified sgRNA (**Supplementary Tables S3, S7, S10**). With SpCas9n-pBE, the modified sgRNAs showed equal or slightly higher editing frequencies (1.2 and 1.5 folds) than the native sgRNAs at the W-T2, W-T3, and W-T6 sites, efficiencies of 0–6.3% at the W-T4 site, and sharply decreased efficiencies (from 33.3% with the native sgRNA to 5.3% with the modified sgRNA) at the W-T1 site (**Figure 4A**). These results indicate that the modified sgRNA had an unsubstantial or small enhancement effect compared with native sgRNA for SpCas9n-pBE.

With VQRn-pBE, the modified sgRNA showed no base mutations at all four target sites with NGAA PAMs, which is similar to the results obtained with the native sgRNA (data not shown). However, base editing efficiencies were significantly enhanced by 1.3- to 7.6-fold for target sites with the other three PAMs using VQRn-pBE with the modified sgRNA (**Figure 4A**). Moreover, using the modified sgRNA produced C-to-T editing events with efficiencies of 0–12.5% and 5.9% in the W-T10 and W-T17 sites, but equal or slightly decreased efficiencies in the W-T9, W-T11, and W-T13 sites compared with using the native sgRNA (**Figure 4A**). These results indicate that the modified sgRNA was more effective in promoting the base editing efficiency of VQRn-pBE than that of SpCas9n-pBE.

Interestingly, except for the four targets with NGAG PAMs, no indels were identified in any of the targets with the other three PAMs using the modified sgRNA (**Supplementary Figure S4**). We compared the mutant genotypes of four sites with the highest editing efficiencies using VQRn-pBE with the native or modified sgRNAs. We found that the frequencies of single C-to-T mutations decreased and the frequencies of double or multiple C-to-T mutations increased when the modified sgRNA was used (**Figure 4B**). New genotypes also were produced in some target sites, such as C3C13 > T3T13 at the W-T7 site and C1C2C5 > T1T2T5 at the W-T9 site (**Figures 4B,C**). Moreover, the editing window was enlarged to position 13 and 10 within the

protospacer at W-T7 and W-T8 sites when the modified sgRNA was used (**Figures 4B,C**).

Because the base editing efficiency of VQRn-pBE was increased with the modified sgRNA, we also detected its off-target effects. The potential off-target sites were the same as those analyzed using the native sgRNA (**Supplementary Table S8**). The sequences were amplified in T₀ plants for Sanger sequencing and no mutations were found at any of the off-target loci tested (data not shown).

DISCUSSION

Cytosine base editors are powerful new tools for targeted base editing in cells and organisms (Hess et al., 2017; Kim, 2018). The NGG PAM requirement of canonical SpCas9 greatly limits the targeting scope of CBEs. In this study, by fusing SpCas9n and its variants VRERn and VQRn with PmCDA1, we obtained two effective base editors, SpCas9n-pBE, and VQRn-pBE. Consistent results from both calli and T₀ plants confirmed the editing ability of both these base editors. About 66.7% of the selected target sites were base edited using SpCas9n-pBE (4/6) and VQRn-pBE (12/18) with frequencies of 4–90% in T₀ plants (**Figures 1B, 2B** and **Tables 1, 2**). Therefore, the pBEs enlarged the range of cytidine base editing from NGG PAM to NAG and NGA PAMs in rice.

Although VRERn fused with rAPOBEC1 was active in human cells (Kim Y.B. et al., 2017), no mutations were detected among the selected targets in rice with VRERn-pBE. This may be explained by the different genome environments in plant and human cells and the different deaminase base editing systems used. This result suggests that not all Cas9 variants that perform well in human cells will perform well in rice, and other Cas9 variants need to be tested before they are applied in plant.

In human cells, VQR produced different cleaving efficiency at sites that contained NGAN PAMs as follows: NGAG > NGAT = NGAA > NGAC (Kleinstiver et al., 2015). Therefore, we designed four NGAN PAMs for rice and tested the base editing activity of VQRn-pBE for the different target sites. Mutations were detected in 80% (8/10) of the target sites with the NGAG PAM, and in 75% (3/4) and 25% (1/4) at the target

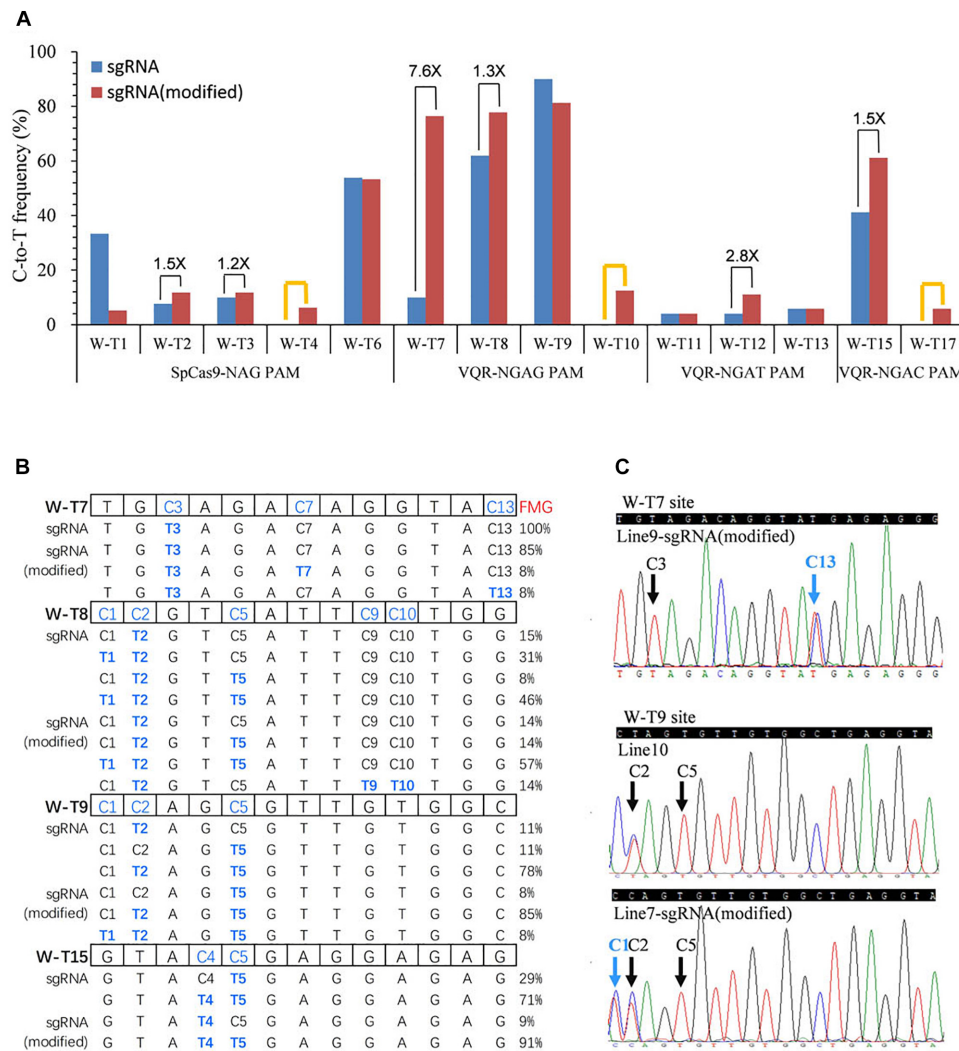


FIGURE 4 | Base editing ability of SpCas9n-pBE and VQRn-pBE with the modified sgRNA in rice T_0 plants. **(A)** C-to-T substitution efficiency of SpCas9n-pBE and VQRn-pBE at target sites with the native and modified sgRNAs. The PAM sequence of each target site is shown below the x axis. Black lines indicate targets where the editing frequency increased using the modified sgRNA compared with using the native sgRNA. Orange lines indicate targets that were not edited using the native sgRNA but were edited successfully using the modified sgRNA. **(B)** Comparison of mutant genotypes and frequency of mutant genotypes (FMGs) using VQRn-pBE with the native or modified sgRNA at four sites with the highest editing efficiencies. No mutations were detected in any cytidine residue located after position 13 in the targets, so the data from positions 14–20 upstream of the PAM sequence are omitted. **(C)** Sequencing chromatograms of T_0 plants at the W-T7 target of line9 and W-T9 target of line 7, with additional C13 and C1 base substitutions obtained using the modified sgRNA. Blue arrows indicate the additional edited bases obtained using the modified sgRNA; black arrows indicate the edited bases obtained using the native sgRNA.

sites with the NGAT and NGAC PAMs. No mutations were found in target sites with the NGAA PAM. These results imply that VQRn-pBE may have a strong preference for targets with NGAN PAMs as follows: NGAG > NGAT > NGAC > NGAA. However, the editing efficiencies were very low (4–7.1%) in the 75% edited targets with a NGAT PAM and much higher (41.2%) in the 25% edited targets with a NGAC PAM. Hence, we can not conclude the editing efficiency of VQRn-pBE was higher for targets with the NGAT PAM than for targets with the NGAC PAM. Recently, Hua and coworkers fused VQRn with rAPOBEC1 to achieve C-to-T substitutions in rice and reported 71.4% editing efficiency at one target site

with a NGAG PAM (Hua et al., 2018). This is similar with our results with VQRn-pBE, which was more efficient at sites harboring NGAG PAMs.

For both SpCas9n-pBE and VQRn-pBE, the editing window typically spanned positions 1 to 7 within the protospacer, and the target C at or near position 3 showed the highest C-to-T substitution frequency. This result is a little different from the reported Cas9n-plant base editor composed of rAPOBEC1, Cas9n, and UGI in rice, in which the deamination window spanned positions 3–9, and the target C at or near position 7 showed the highest editing efficiency (Zong et al., 2017). Moreover, indels seem to be produced at the targets with

relatively high editing efficiencies. Introduction of the Gam protein of bacteriophage Mu (Komor et al., 2017b) can be tested to reduce the indel frequency and improve product purity in the future.

Modified sgRNAs were reported to improve knock out activity in mammalian cells and in rice when combined with wild type SpCas9 or its VQR variant (Chen et al., 2013; Hu et al., 2017). In our study, targets that were not mutated with the native sgRNA were successfully base edited with the modified sgRNA. Moreover, the enhanced efficiency using the modified sgRNA was more predominant for VQRn-pBE than for SpCas9n-pBE. Analysis of the mutant genotypes in targets with high editing frequencies revealed changes that occurred when the modified sgRNA was used: (i) the frequencies of single C-to-T conversions decreased and double or multiple C-to-T substitutions increased, and (ii) mutants with novel single C-to-T conversions in different positions or with new multiple C-to-T mutations at various positions were obtained. Together, these results suggest that the modified sgRNA could be used to increase the editing frequency of VQRn-pBE in rice under applicable circumstances.

Other natural or evolved CRISPR nucleases with different PAM requirements that can broaden the editable targets include *Lachnospiraceae* bacterium Cpf1, *Neisseria meningitidis* Cas9, *Campylobacter jejuni* Cas9, and *Streptococcus thermophilus* Cas9 (Hou et al., 2013; Glemzaite et al., 2015; Kim E. et al., 2017; Tang et al., 2017; Li et al., 2018; Zhong et al., 2018). Recently, a human cytidine deaminase APOBEC3A was used to generate an effective base editor in mammalian cells and plants (Gehrke et al., 2018; Zong et al., 2018). To further expand the scope of base editing in rice, all the

above CRISPR nucleases and human APOBEC3A should be tested in the future.

CONCLUSION

In this study, we described two efficient PmCDA1-based CBE systems, SpCas9n-pBE and VQRn-pBE, that will help to expand the scope of cytosine base editing in rice. The effective deamination window typically spanned positions 1–7 of the protospacer and the target single C showed the highest editing frequency at or near position 3. The mutant genotypes were mainly single or double C-to-T substitutions. Furthermore, the editing efficiency of VQRn-pBE was increased by the modified sgRNA. These base editors will be useful tools for scientific research and crop breeding in rice.

AUTHOR CONTRIBUTIONS

JY, CZ, and YW designed the experiments and wrote the manuscript. YW, FW, SZ, FF, and JS performed all the experiments. WX and FW analyzed the results. JY supervised the project. All authors read and approved the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00379/full#supplementary-material>

REFERENCES

- Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513, 569–573. doi: 10.1038/nature13579
- Bae, S., Park, J., and Kim, J. S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 30, 1473–1475. doi: 10.1093/bioinformatics/btu048
- Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G. W., et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479–1491. doi: 10.1016/j.cell.2013.12.001
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi: 10.1126/science.1231143
- Dang, Y., Jia, G., Choi, J., Ma, H., Anaya, E., Ye, C., et al. (2015). Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biol.* 16:280. doi: 10.1186/s13059-015-0846-3
- Doudna, J. A., and Charpentier, E. (2014). Genome editing: the new frontier of genome engineering with CRISPR-Cas9. *Science* 346:1258096. doi: 10.1126/science.1258096
- Endo, M., Mikami, M., Endo, A., Kaya, H., Itoh, T., Nishimasu, H., et al. (2019). Genome editing in plants by engineered CRISPR–Cas9 recognizing NG PAM. *Nat. Plants* 5, 14–17. doi: 10.1038/s41477-018-0321-8
- Gehrke, J. M., Cervantes, O., Clement, M. K., Wu, Y., Zeng, J., Bauer, D. E., et al. (2018). An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* 36, 977–982. doi: 10.1038/nbt.4199
- Glemzaite, M., Balciunaite, E., Karvelis, T., Gasiunas, G., Grusyte, M. M., Alzbutas, G., et al. (2015). Targeted gene editing by transfection of in vitro reconstituted *Streptococcus thermophilus* Cas9 nuclease complex. *RNA Biol.* 12, 1–4. doi: 10.1080/15476286.2015.1017209
- Henikoff, S., and Comai, L. (2003). Single-nucleotide mutations for plant functional genomics. *Annu. Rev. Plant Biol.* 54, 375–401. doi: 10.1146/annurev.arplant.54.031902.135009
- Henikoff, S., Till, B. J., and Comai, L. (2004). TILLING. Traditional mutagenesis meets functional genomics. *Plant Physiol.* 135, 630–636. doi: 10.1104/pp.104.041061
- Hess, G. T., Tycko, J., Yao, D., and Bassik, M. C. (2017). Methods and applications of CRISPR-mediated base editing in eukaryotic genomes. *Mol. Cell* 68, 26–43. doi: 10.1016/j.molcel.2017.09.029
- Hiei, Y., and Komari, T. (2008). Agrobacterium-mediated transformation of rice using immature embryos or calli induced from mature seed. *Nat. Protoc.* 3, 824–834. doi: 10.1038/nprot.2008.46
- Hou, Z., Zhang, Y., Propson, N. E., Howden, S. E., Chu, L. F., Sontheimer, E. J., et al. (2013). Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15644–15649. doi: 10.1073/pnas.1313587110
- Hsu, P. D., Lander, E. S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262–1278. doi: 10.1016/j.cell.2014.05.010
- Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., et al. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 556, 57–63. doi: 10.1038/nature26155

- Hu, X., Meng, X., Liu, Q., Li, J., and Wang, K. (2017). Increasing the efficiency of CRISPR-Cas9-VQR precise genome editing in rice. *Plant Biotechnol. J.* 16, 292–297. doi: 10.1111/pbi.12771
- Hu, X., Wang, C., Fu, Y., Liu, Q., Jiao, X., and Wang, K. (2016). Expanding the range of CRISPR/Cas9 genome editing in rice. *Mol. Plant* 9, 943–945. doi: 10.1016/j.molp.2016.03.003
- Hua, K., Tao, X., and Zhu, J. K. (2018). Expanding the base editing scope in rice by using Cas9 variants. *Plant Biotechnol. J.* 17, 499–504. doi: 10.1111/pbi.12993
- Kim, E., Koo, T., Park, S. W., Kim, D., Kim, K., Cho, H. Y., et al. (2017). In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat. Commun.* 8:14500. doi: 10.1038/ncomms14500
- Kim, Y. B., Komor, A. C., Levy, J. M., Packer, M. S., Zhao, K. T., and Liu, D. R. (2017). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* 35, 371–376. doi: 10.1038/nbt.3803
- Kim, J. S. (2018). Precision genome engineering through adenine and cytosine base editing. *Nat. Plants* 4, 148–151. doi: 10.1038/s41477-018-0115-z
- Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V., Nguyen, N. T., Zheng, Z., et al. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523, 481–485. doi: 10.1038/nature14592
- Komor, A. C., Badran, A. H., and Liu, D. R. (2017a). CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* 169:559. doi: 10.1016/j.cell.2017.04.005
- Komor, A. C., Zhao, K. T., Packer, M. S., Gaudelli, N. M., Waterbury, A. L., Koblan, L. W., et al. (2017b). Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci. Adv.* 3:eao4774. doi: 10.1126/sciadv.aao4774
- Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424. doi: 10.1038/nature17946
- Li, J., Sun, Y., Du, J., Zhao, Y., and Xia, L. (2017). Generation of targeted point mutations in rice by a modified CRISPR/Cas9 system. *Mol. Plant* 10, 526–529. doi: 10.1016/j.molp.2016.12.001
- Li, S., Zhang, X., Wang, W., Guo, X., Wu, Z., Du, W., et al. (2018). Expanding the scope of CRISPR/Cpf1-mediated genome editing in rice. *Mol. Plant* 11, 995–998. doi: 10.1016/j.molp.2018.03.009
- Lu, Y., and Zhu, J. K. (2017). Precise editing of a target base in the rice genome using a modified CRISPR/Cas9 system. *Mol. Plant* 10, 523–525. doi: 10.1016/j.molp.2016.11.013
- Meng, X., Hu, X., Liu, Q., Song, X., Gao, C., Li, J., et al. (2018). Robust genome editing of CRISPR-Cas9 at NAG PAMs in rice. *Sci. China Life Sci.* 61, 122–125. doi: 10.1007/s11427-017-9247-9
- Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353:aaf8729. doi: 10.1126/science.aaf8729
- Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S., Dohmae, N., et al. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156, 935–949. doi: 10.1016/j.cell.2014.02.001
- Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., et al. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* 361, 1259–1262. doi: 10.1126/science.aas9129
- Qin, R., Li, J., Li, H., Zhang, Y., Liu, X., Miao, Y., et al. (2018). Developing a highly efficient and widely adaptive CRISPR-SaCas9 toolset for plant genome editing. *Plant Biotechnol. J.* 17, 706–708. doi: 10.1111/pbi.13047
- Ren, B., Yan, F., Kuang, Y., Li, N., Zhang, D., Lin, H., et al. (2017). A CRISPR/Cas9 toolkit for efficient targeted base editing to induce genetic variations in rice. *Sci. China Life Sci.* 60, 516–519. doi: 10.1007/s11427-016-0406-x
- Shalem, O., Sanjana, N. E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet.* 16, 299–311. doi: 10.1038/nrg3899
- Shimatani, Z., Kashojiya, S., Takayama, M., Terada, R., Arazoe, T., Ishii, H., et al. (2017). Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat. Biotechnol.* 35, 441–443. doi: 10.1038/nbt.3833
- Slade, A. J., Fuerstenberg, S. I., Loeffler, D., Steine, M. N., and Facciotti, D. (2005). A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nat. Biotechnol.* 23, 75–81. doi: 10.1038/nbt1043
- Tang, X., Lowder, L. G., Zhang, T., Malzahn, A. A., Zheng, X., Voytas, D. F., et al. (2017). A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat. Plants* 3:17018. doi: 10.1038/nplants.2017.103
- Wang, H., La, R. M., and Qi, L. S. (2016). CRISPR/Cas9 in genome editing and beyond. *Annu. Rev. Biochem.* 85, 227–264. doi: 10.1146/annurev-biochem-060815-014607
- Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3570–3575. doi: 10.1073/pnas.1420294112
- Yang, B., Li, X., Lei, L., and Chen, J. (2017). APOBEC: from mutator to editor. *J. Genet. Genomics* 44, 423–437. doi: 10.1016/j.jgg.2017.04.009
- Yin, K., Gao, C., and Qiu, J. L. (2017). Progress and prospects in plant genome editing. *Nat. Plants* 3:17107. doi: 10.1038/nplants.2017.107
- Zhao, K., Tung, C. W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., et al. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* 2:467. doi: 10.1038/ncomms1467
- Zhong, Z., Zhang, Y., You, Q., Tang, X., Ren, Q., Liu, S., et al. (2018). Plant genome editing using FnCpf1 and LbCpf1 nucleases at redefined and altered PAM sites. *Mol. Plant* 11, 999–1002. doi: 10.1016/j.molp.2018.03.008
- Zong, Y., Song, Q., Li, C., Jin, S., Zhang, D., Wang, Y., et al. (2018). Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat. Biotechnol.* 36, 950–953. doi: 10.1038/nbt.4261
- Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., et al. (2017). Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* 35, 438–440. doi: 10.1038/nbt.3811

Conflict of Interest Statement: The authors submitted a patent application based on the results reported in this paper.

Copyright © 2019 Wu, Xu, Wang, Zhao, Feng, Song, Zhang and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Biological Characteristics of Severe Combined Immunodeficient Mice Produced by CRISPR/Cas9-Mediated *Rag2* and *IL2rg* Mutation

Yong Zhao¹, Peijuan Liu¹, Zhiqian Xin¹, Changhong Shi¹, Yinlan Bai², Xiuxuan Sun³, Ya Zhao¹, Xiaoya Wang^{1,4}, Li Liu^{1,5}, Xuan Zhao^{1,4}, Zhinan Chen^{3*} and Hai Zhang^{1,6*}

¹ Laboratory Animal Center, Air Force Medical University, Xi'an, China, ² Department of Microbiology, Air Force Medical University, Xi'an, China, ³ Department of Cell Biology, National Translational Science Center for Molecular Medicine, Air Force Medical University, Xi'an, China, ⁴ College of Veterinary Medicine, Northwest A&F University, Yangling, China, ⁵ Key Laboratory for Space Bioscience and Biotechnology, School of Life Sciences, Northwestern Polytechnical University, Xi'an, China, ⁶ National Translational Science Center for Molecular Medicine, Air Force Medical University, Xi'an, China

OPEN ACCESS

Edited by:

David Jay Segal,
University of California, Davis,
United States

Reviewed by:

Serap Yalin,
Mersin University, Turkey
Wei Xu,
Texas A&M University–Corpus Christi,
United States

*Correspondence:

Zhinan Chen
znchen@fmmu.edu.cn
Hai Zhang
hzhang@fmmu.edu.cn

Specialty section:

This article was submitted to
Genomic Assay Technology,
a section of the journal
Frontiers in Genetics

Received: 27 November 2018

Accepted: 12 April 2019

Published: 30 April 2019

Citation:

Zhao Y, Liu P, Xin Z, Shi C, Bai Y,
Sun X, Zhao Y, Wang X, Liu L, Zhao X,
Chen Z and Zhang H (2019)
Biological Characteristics of Severe
Combined Immunodeficient Mice
Produced by CRISPR/Cas9-Mediated
Rag2 and *IL2rg* Mutation.
Front. Genet. 10:401.
doi: 10.3389/fgene.2019.00401

Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas)9 is a novel and convenient gene editing system that can be used to construct genetically modified animals. Recombination activating gene 2 (*Rag2*) is a core component that is involved in the initiation of V(D)J recombination during T- and B-cells maturation. Separately, the interleukin-2 receptor gamma chain gene (*IL2rg*) encoded the protein-regulated activity of natural killer (NK) cells and shared common receptors of some cytokines. *Rag2* and *IL2rg* mutations cause immune system disorders associated with T-, B-, and NK cell function and some cytokine activities. In the present study, 2 single-guide RNAs (sgRNAs) targeted on *Rag2* and *IL2rg* genes were microinjected into the zygotes of BALB/c mice with Cas9 messenger RNA (mRNA) to create *Rag2/IL2rg*^{-/-} double knockout mice, and the biological characteristics of the mutated mice were subsequently analyzed. The results showed that CRISPR/Cas9-induced indel mutation displaced the frameshift of *Rag2* and *IL2rg* genes, resulting in a decrease in the number of T-, B-, and NK cells and the destruction of immune-related tissues like the thymus and spleen. *Mycobacterium tuberculosis* 85B antigen could not induce cellular and humoral immune response in mice. However, this aberrant immune activity compromised the growth of several tumor heterogenous grafts in the mutated mice, including orthotopic and subcutaneous transplantation tumors. Thus, *Rag2/IL2rg*^{-/-} knockout mice possessed features of severe combined immunodeficiency (SCID), which is an ideal model for human xenograft.

Keywords: *Rag2*, *IL2rg*, CRISPR/Cas9, severe combined immunodeficient mice, biological characteristic

INTRODUCTION

The construction of chimeras of a rodent animal model that harbors human tissues has provided valuable *in vivo* assay systems in biomedical research. To do this, aberrant immune-related genes make it possible to construct chimeric rodent animals. The nude mouse (or athymic nude mouse) was first described by Flanagan (1966), which involved a spontaneous mutation in the

Foxn1 gene, resulting in a lack of fur development and impaired T-cell function (Schorpp et al., 1997). Thereafter, CBA/N and Beige mice, which boasted mutations in the *xid* and *beige* genes, respectively, leading to B-cell- and natural killer (NK)-cell-mediated immune-response failure, were also discovered (Clark et al., 1981; Klaus et al., 1997). After that, *prkdc* gene and *Rag2* mutation mouse, which showed T- and B-cell dysregulation, were defined as a severe combined immunodeficiency (SCID) mouse and used widely in biomedical research (Shinkai et al., 1992; Greiner et al., 1998). Subsequently, SCID mice were greatly improved by the development of non-obese diabetic (NOD) mice, and a new strain of NOD/SCID mice was created by backcrossing SCID mice with NOD mice (Shultz et al., 1995). In these mice, the mature, function lymphocytes were absent, and lower levels of NK cells and cytokine production were present. Further studies were carried out by mating NOD/shi-SCID mice or *Rag2* mutation mice with interleukin-2 receptor gamma chain gene (*IL2rg*) mutation mice, which generated T, B, and NK cells combined deficiency mice, like NOG, NSG, and *Rag2/IL2^{-/-}* double knockout mice (Shultz et al., 2005; Belizário, 2009). These mouse have higher immunocompromised symptoms than the previously mentioned mice did due to the simultaneous absence of mature T-cells, B-cells, and NK cells as well as defective macrophage activity and reduced dendritic cell function (Ito et al., 2002; Shultz et al., 2007; McDaniel and Grisham, 2018). However, all of above-mentioned rodents were known as immunodeficient mice due to one or more immune response being impaired. As a result, these immunodeficient mice are advantageous because of their engraftment, infection control, and tumor control, and thus are a useful tool in biomedical research.

Clustered regularly interspaced short palindromic repeats (CRISPR) was a DNA loci in that it contained multiple, short, and direct repetitions of base sequences and could be found in bacteria and archaea (Jansen et al., 2002; Ishino et al., 2018). In the adjacent region of a CRISPR sequence, a conserved coding protein sequence was also identified, which was labeled as a CRISPR-associated (Cas) gene, and, correspondingly, its encoded protein was referred to as the Cas protein (Barrangou, 2015). Cas proteins form a large family that includes many subtypes; among these, the Cas9 protein originating from the bacterial type II CRISPR/Cas system is a programmable RNA-guided endonuclease that is capable of binding and cutting site-specific cleavage of double-stranded DNA (Mojica and Montolieu, 2016). The Cas9 enzyme recognizes the protospacer adjacent motif (PAM) sequence 5'-NGG-3' and cleaves the DNA at 3 bp to 4 bp upstream of PAM guiding by tracrRNA and crRNA, the damage DNA is subsequently repaired using 2 main pathways: non-homologous end joining (NHEJ) and homology-directed repair (HDR). NHEJ always generates indel (insertion or deletion) mutations, while HDR occurs when the repaired template is presented (Hsu et al., 2014). The original biological function of CRISPR/Cas9 was an adaptive immune defense mechanism against phage for bacteria; the invaded DNA was recognized by CRISPR, after which Cas9 cleaved the exogenous DNA, leading the invading phage to become inactive (Sampson and Weiss, 2014). Since it represents a more convenient, rapid, and efficient way to introduce a mutation in a genome sequence, CRISPR/Cas9

is now known as a key novel genome engineering tool for replacing, deleting, or inserting base pairs into a DNA sequence, which could be used to construct genetically modified animals (Sander and Joung, 2014; Tschaharganeh et al., 2016).

Recombination activating gene 2 (*Rag2*), expressed in adult thymus (Wilson et al., 1994), is an immune-related molecule that is involved in the initiation of immunoglobulin V(D)J gene rearrangement and T-cell-receptor gene recombination during T- and B-cell development (Notarangelo et al., 2016). *Rag2* is essential to the generation of mature T- and B-lymphocytes; importantly, mutations of this gene in humans retards T- and B-cell development, resulting in SCID associated with autoimmune-like Omenn symptom occurrence (Corneo et al., 2001; Notarangelo et al., 2016). Separately, *IL2rg*, expressed in thymus and spleen (Cao et al., 1993), is known as an immune regulator in cytokine secretion; in the growth and differentiation of T-cells, B-cells, and NK cells; and in maintaining the homeostasis of the immune system (Aliyari et al., 2015). Mutation of the *IL2rg* gene prompted a deficiency in functional NK cell and cytokine secretion reduction, including IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21 (Puck et al., 1997). In the present study, we postulated the construct of SCID mice through a mutation in the *Rag2* and *IL2rg* genes using the CRISPR/Cas9 gene editing tool, sought to determine the biological characteristics of the mutated mice by investigation the immune response against the 85B antigen of *Mycobacterium tuberculosis*, and establishment a human tumor xenograft model *in vivo*.

MATERIALS AND METHODS

Animals

BALB/c mice were obtained from the Laboratory Animal Center of Air Force Medical University. ICR mice, which used as recipient animal for transplanting microinjected zygotes, were purchased from Vita River Laboratory Animal Technology Co., Ltd., (Beijing, China). The mice were housed in a temperature- and climate-controlled specific pathogen free facility with a 12-h light/dark schedule. Body weight and the intake of food and water were calculated per week. All mouse experiments were approved by the Institutional Animal Care and Use Committee of Air Force Medical University.

Reagents and Plasmid

A MEGashortscriptTM T7 high-yield transcription kit and MEGAcleanTM kit were provided by Thermo Fisher Scientific (Waltham, MA, United States). Cas9 messenger RNA (mRNA) and protein were purchased from Biomics biotechnologies (Nantong, China) and New England Biolabs (Ipswich, MA, United States), respectively. Pregnant mare's serum gonadotropin (PMSG) and human chorionic gonadotropin (hCG) were purchased from the Ningbo Second Hormone Factory (Ningbo, China). M2 medium was provided by Sigma-Aldrich (St. Louis, MO, United States), while a KSOM powdered media kit (cat: MR-020P-SF) was obtained from Millipore (Burlington, MA, United States). Mouse FITC-CD3, PE-NKp46, and APC-B220 antibodies were purchased from BioLegend (San Diego, CA,

United States). LongAmp Taq DNA polymerase and *Bbs I* restriction enzyme was provided by New England Biolabs (Ipswich, MA, United States). A mouse tail genome extraction kit was sourced from Foregene Biological Technology Co., Ltd., (Chengdu, China). pX330 plasmid was purchased from Addgene. Interferon (IFN) γ , IL-2, and IL-10 cytokine enzyme-linked immunoassay (ELISA) detection kits were purchased from eBioscience (San Diego, CA, United States).

Cell Culture

The brain glioma cell line U87 was purchased from the Type Culture Collection of the Chinese Academy of Sciences (Shanghai, China). Human primary gastric, renal, and bladder carcinoma cell-luciferase and Passage Burkitt's lymphoma cell line Raji-luciferase were obtained from the Laboratory Animal Center of Air Force Medical University. Cells were incubated in high-glucose Dulbecco's modified Eagle medium or Roswell Park Memorial Institute 1640 supplemented with 10% fetal bovine serum under a humidified atmosphere of 5% CO₂ at 37°C.

Preparation of Single-Guide RNA and Microinjection

For the purpose of single-guide RNA (sgRNA) transcription *in vitro*, 220 bp sgRNA sequences targeting *Rag2* exon3 (gene ID: 19374) and *IL2rg* exon1 (gene ID: 16186) were screened on the website of <http://crispr.mit.edu> and synthesized by TsingKe Biological Technology (Xi'an, China). After annealing, double-strand DNA was digested with *Bbs I* restriction enzyme and cloned into pX330 plasmid. Polymerase chain reaction (PCR) was performed to obtain a sgRNA sequence carrying T7 promoter and the 121 bp PCR product then was transcribed with the MEGAshortscript™ T7 high-yield transcription kit according to the manufacturer's protocol and purified. Mice superovulation and microinjection were carried out according to a previous report (Esmail et al., 2016). Briefly, 20 μ g of *Rag2*, 20 μ g of *IL2rg* sgRNA mixture, and 10 μ g of Cas9 mRNA were microinjected into the cytoplasm of collected fertilized eggs. After incubation for 24 h at 37°C, the 2-cell forms of the eggs were then transplanted to the ampulla of recipient pseudopregnancy ICR female mice.

Single-Guide RNA *in vitro* Cleavage Efficiency Assay

PCR reaction was performed with *Rag2* and *IL2rg* specific primers to obtain substrate DNA. After purification, 1 μ g substrate DNA was digested with 2 μ g Cas9 protein, 200 ng sgRNA, and 2 μ L of 10 \times Cas9 buffer at 37°C for 1 h in 20 μ L of reaction volume. Reaction products were run on 1.5% agarose gel to examine cleavage efficiency.

Flow Cytometry

50 μ L of peripheral blood was collected from the tail veins of homozygous mice. Samples were lysed with erythrocyte lysing solution and incubated for 30 min with 1:1,000-diluted FITC-CD3, PE-NKp46, and APC-220 antibodies in a dark place. Then, samples were analyzed by flow cytometry (Becton, Dickinson and Company, Franklin Lakes, NJ, United States) and data

were analyzed with the FlowJo softwares (FlowJo LLC, Ashland, OR, United States).

Real-Time Quantitative RT-PCR

Total RNA was extracted from spleen and/or thymus of homozygotes mice with TRIzol reagent (Invitrogen, Carlsbad, CA, United States) according to the manufacturer's instructions. 500 ng total RNA was reverse-transcribed to cDNA and qPCR was performed using a SYBR Green PCR kit (TakaRa, Dalian, China). Each sample was run in triplicate in a final volume 25 μ L reaction mix, which contained 1 μ L cDNA template, 10 pmol of *Rag2* and *IL2rg* specific primers (Table 1), and 12.5 μ L of SYBR Green solution. Assays were run using following procedures: 1 cycle of 95°C 30 s, followed by 40 cycles of 95°C for 20 s and 60°C for 30 s. Data was analyzed with the $2^{-\Delta\Delta CT}$ method.

Tumor Xenograft Model

15 *Rag2/IL2rg*^{-/-} mice were divided into three group: (1) human primary tumor cells inoculation group; (2) Raji cells inoculation group; and (3) U87 cells inoculation group. The logarithmic growth phase of human primary gastric, renal, and bladder carcinoma cells were collected and 1 \times 10⁷ cells/mouse were implanted subcutaneously in the flank site and bred for 3 weeks. Meanwhile, 1 \times 10⁷ Raji cells were inoculated intravenously to replicate a hematopoietic model. For the glioma xenograft model, 1 \times 10⁷ U87 cells were stereotactically injected into the precuneus, while the other mice were implanted subcutaneously in the flank site. 3 weeks later, all mice were euthanized and tumor formation was observed through skull anatomy in glioma xenograft mice, while the luciferase-labeled cell xenograft mice were visualized using the IVIS Lumina II imaging system (PerkinElmer, Waltham, MA, United States).

Hematoxylin and Eosin Staining

Once the mice were euthanized, the thymus, spleen, and the xenograft tumor tissue samples were sectioned and fixed in 4% formaldehyde for 24 h, followed by dehydration with a series of ethanol solutions and subsequent embedding in paraffin. Then, 5- μ m-thick sections were cut and stained with hematoxylin and eosin (H&E) according to protocol. The histopathological changes were examined under a light microscope (BX43; Olympus, Tokyo, Japan).

Genotype Analysis

Genome DNA was extracted from the tail tip of 1-week-old mice and PCR reaction was performed with *Rag2* and *IL2rg* specific forward and reverse primers, respectively. After purification, PCR products were sequenced with Sanger sequencing and the results

TABLE 1 | Primer sets used for qPCR.

Name of genes	Forward primer	Reverse primer
<i>Rag2</i>	ATTCAACCAGGCTTCTCACTT	TGCAGGCTTCAGTTTGAGATG
<i>IL2rg</i>	AGAGCAAGCACCATGTTGAA	CATTTCGACTGGACATGAGG
Actin	GGAAATCGTGCCTGACATCA	AATAGTGATGACCTGGCCGT

were analyzed with the SnapGene 3.1.1 software (GSL Biotech, Chicago, IL, United States).

Immunization, Lymphocyte Proliferation, Antibody and Cytokine Assay

10 *Rag2/IL2rg*^{-/-} mice were divided into 2 group: (1) 85B antigen treated group; and (2) control group. Recombinant 85B antigen of *Mycobacterium tuberculosis* (MTB) was mixed with aluminum hydroxide adjuvant at a 1:1 ratio. Then, the mutated mice (experimental group) and WT BALB/c mice (control group) were inoculated 3 times intramuscularly with a 2-week interval by 50 µg of 85B antigen/adjuvant mixture at the hind leg. Following immunization, antibody, and lymphocyte proliferation assay were performed as done in a previous report (Zhang et al., 2010; Wang et al., 2014). Briefly, recombinant 85B antigen was coated and the titer of anti-85B specific antibody was carried out by ELISA assay. Meanwhile, lymphocytes were isolated from the spleen of immunized mice and simulated with recombinant 85B antigen (experimental well) or PPD (positive control), post which the supernatants were collected for cytokine measurement. Next, 20 µL MTS (Promega, Madison, MI, United States) was added in each well and incubated for another 4 h; optical density was measured under 490 nm; and the stimulation index (SI) was used to evaluate lymphocyte proliferation, as follows: SI = (A₄₉₀ of stimulated wells - A₄₉₀ of blank cells) / (A₄₉₀ of negative wells - A₄₉₀ of blank wells).

Statistical Analysis

Data are expressed in the form of mean ± standard deviation. A Student's *t* test and one-way analysis of

variance were used for assessing significant differences among experimental groups using the Statistical Package for the Social Sciences software, version 17 (IBM Corp., Armonk, NY, United States). *p*-values of <0.05 or <0.01 were considered to be statistically significant.

RESULTS

Construction of *Rag2/IL2rg*^{-/-} Gene Double-Knockout Mice With CRISPR/Cas9 System From BALB/c Strain

Rag2 and *IL2rg* were involved in the development of T-, B-, and NK cells and the production of cytokines. The mutation of these genes retarded their development in the immature stage and contributed to a lack of both innate and adaptive immune response. Based on this, we planned to construct a SCID BALB/c mouse model targeting the *Rag2* and *IL2rg* genes simultaneously with a CRISPR/Cas9 gene editing tool. After screening on the <http://crispr.mit.edu> website, a pair of 20 bp oligonucleotides was selected as sgRNA targeting sequences from the *Rag2* exon3 sense strand and *IL2rg* exon1 anti-sense strand, respectively (Figure 1A). Subsequently, *in vitro* cleavage efficiency of the sgRNA assay demonstrated that *Rag2* and *IL2rg* sgRNA were endowed with stronger cutting activity for the target sequence, resulting in producing two obvious bands on 1.5% agarose gel (Figure 1B). Similarly, Sanger sequencing also

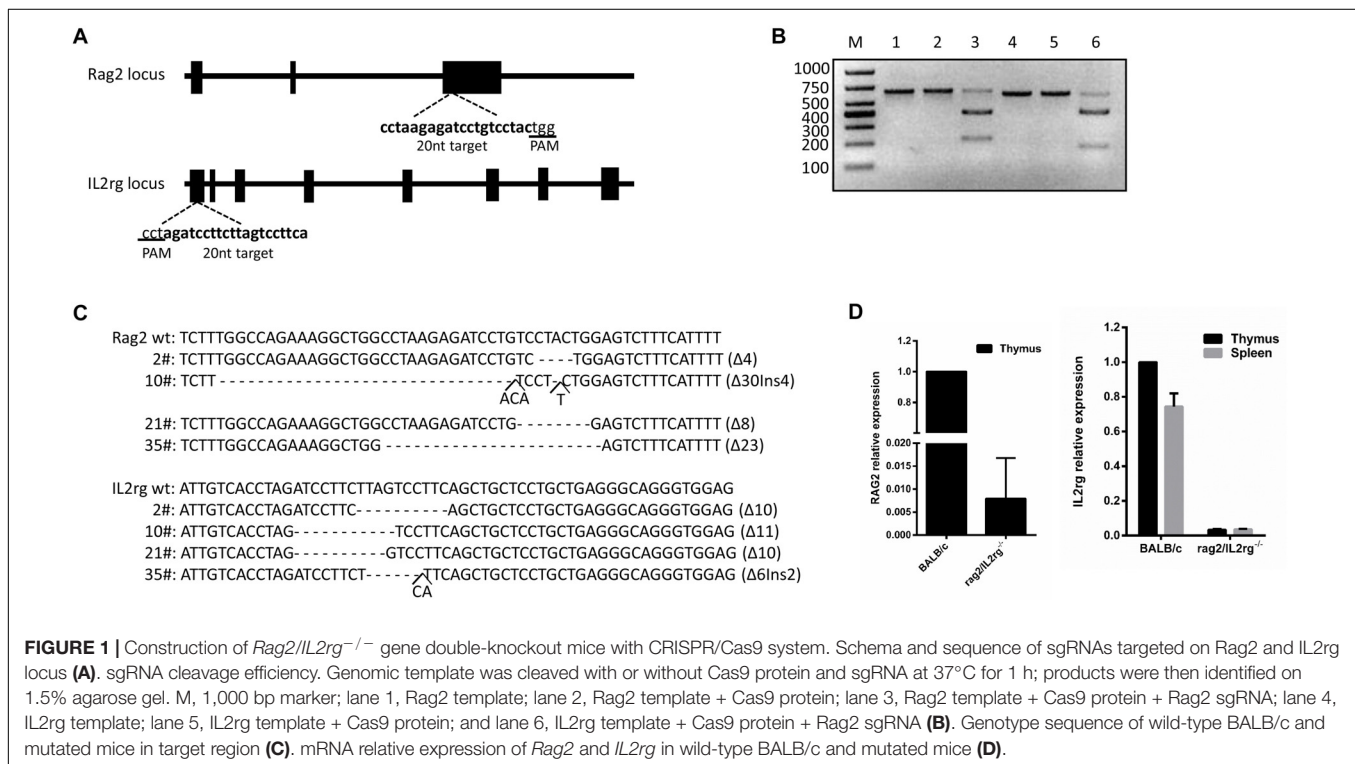


FIGURE 1 | Construction of *Rag2/IL2rg*^{-/-} gene double-knockout mice with CRISPR/Cas9 system. Schema and sequence of sgRNAs targeted on *Rag2* and *IL2rg* locus (A). sgRNA cleavage efficiency. Genomic template was cleaved with or without Cas9 protein and sgRNA at 37°C for 1 h; products were then identified on 1.5% agarose gel. M, 1,000 bp marker; lane 1, *Rag2* template; lane 2, *Rag2* template + Cas9 protein; lane 3, *Rag2* template + Cas9 protein + *Rag2* sgRNA; lane 4, *IL2rg* template; lane 5, *IL2rg* template + Cas9 protein; and lane 6, *IL2rg* template + Cas9 protein + *Rag2* sgRNA (B). Genotype sequence of wild-type BALB/c and mutated mice in target region (C). mRNA relative expression of *Rag2* and *IL2rg* in wild-type BALB/c and mutated mice (D).

suggested that the CRISPR/Cas9 system possessed higher gene editing efficiency *in vivo*. There were 40 pups born after transplantation, of which 20 pups (50%) and 18 pups (45%) showed induced indel mutation on the *Rag2* and *IL2rg* target sequences, respectively. Among these, 4 pups showed mutations simultaneously on both of these sequences, leading to small-fragment deletion or insertion (Figure 1C). Thus, 10# mouse was mated with wild-type female BALB/c mice to examine germline transmission. Three generations later, a similar genotype was observed in homozygote *Rag2/IL2rg*^{-/-} mice, suggesting indel mutations were stably inherited by offspring. *Rag2* and *IL2rg* expression was detected in adult thymus and/or spleen with specific primers (Table 1), data showed *Rag2* and *IL2rg* mRNA transcriptional level were decreased significantly (Figure 1D). Thus *Rag2/IL2rg*^{-/-} gene double-knockout mice were constructed based on BALB/c background.

Rag2/IL2rg^{-/-} Double Knockout Alter the Number of Granulocytes, but Not Physiological Behavior of Mutated Mice

Gene mutation might alter mouse phenotype, so we speculated as to whether *Rag2/IL2rg*^{-/-} double knockout influenced mutated mice' normal behavior or not. Hematological parameters in routine blood test were assayed by biochemical analyzer. Data indicated that granulocyte counts for *Rag2/IL2rg*^{-/-} mice was decreased significantly, which resulting in increasing of percentage of neutrophils, monocytes, eosinophils and basophiles, while other hematological parameters unchanged (Table 2). There were no significant differences regarding body weight or food and water intake between mutated mice and wild-type mice (Figures 2A–C). Thus, *Rag2/IL2rg* mutation did not influence the physiological behaviors of mice.

TABLE 2 | Comparison of Hematological parameters between *Rag2/IL2rg*^{-/-} and BALB/c mice.

Parameters	<i>Rag2/IL2rg</i> ^{-/-}	BALB/c	P values
HGB (g/dL)	16.76 ± 0.5	16.78 ± 0.4	0.956
RBC (× 10 ⁶ cells/μl)	10.54 ± 0.31	10.69 ± 0.29	0.336
WBC (× 10 ³ cells/μl)	1.61 ± 0.4	3.34 ± 0.37	0.000
PLT (× 10 ³ cells/μl)	857.13 ± 73.21	730.38 ± 149.31	0.049
NEU (× 10 ³ cells/μl)	0.99 ± 0.41	1.59 ± 0.34	0.007
LYM (× 10 ³ cells/μl)	0.19 ± 0.07	1.41 ± 0.44	0.000
MONO (× 10 ³ cells/μl)	0.09 ± 0.03	0.04 ± 0.02	0.000
EOS (× 10 ³ cells/μl)	0.22 ± 0.14	0.11 ± 0.11	0.094
BASO (× 10 ³ cells/μl)	0.01 ± 0.01	0 ± 0.01	0.159
NEU (%)	67.18 ± 7.3	47.83 ± 8.59	0.000
LYM (%)	11.71 ± 3.2	46.58 ± 7.28	0.000
MONO (%)	5.75 ± 0.93	1.06 ± 0.51	0.000
EOS (%)	14.65 ± 9.52	3.44 ± 3.09	0.012
BASO (%)	0.59 ± 0.61	0.14 ± 0.05	0.074

p < 0.05 or <0.01 was considered statistically significant, compare with BALB/c mice.

Rag2/IL2rg^{-/-} Double Knockout Retarded Thymus and Spleen Development and Reduced the Number of Lymphocytes

Since *Rag2* and *IL2rg* genes were involved in immune-related tissue development, we attempted to investigate histopathological structure changes in thymus and spleen tissue. Notably, volume and organ weight ratio of thymus and spleen tissue were decreased significantly after mutation (Figures 3A,B). Thymus atrophy, spleen dysplasia, and lymphocyte reduction could be observed with H&E staining; thymus cells decreased and stromal cells increased in thymus tissue, cortical staining became lighter than the medulla, and the boundary between the cortex and medulla was blurred. Additionally, white pulp shrunk and red pulp expanded in the spleen tissue; the numbers of lymphocytes and hematopoietic and monocyte cells in white pulp were reduced; and the boundary between white pulp and red pulp was more unclear. However, the histopathologic structure of the spleen and thymus was normal in wild-type BALB/c mice (Figure 3C). Affected by this, the numbers of T-cells, B-cells, and NK cells in the peripheral blood were decreased significantly (Figure 3D).

MTB Antigen 85B Could Not Stimulate Immune Response in Rag2/IL2rg^{-/-} Double-Knockout Mice

Mycobacterium tuberculosis 85B protein was a prominent antigen designed to stimulate stronger cellular and humoral immune responses in inbred mice (Lu et al., 2018). To investigate the immune response induced by 85B antigen in *Rag2/IL2rg*^{-/-} double-knockout mice, recombinant 85B protein was immunized 3 times and the titer of anti-85B specific antibody was detected by ELISA. Higher-titer antibody could be induced in wild-type BALB/c mice, but not in *Rag2/IL2rg*^{-/-} double-knockout mice, even with an extension of immunization time (Figures 4A,B), and lymphocyte proliferation was inhibited because of lower SI (Figure 4C). Meanwhile, the expressions of IL-2, IL-10, and IFN-γ were decreased significantly (Figures 4D–F). Considering the above data, *Rag2/IL2rg* double knockout not only destructed the histopathological structure of thymus and spleen tissues but also attenuated the cellular and humoral immune responses in mutated mice, suggesting these mice presented the features of an immunocompromised animal.

Rag2/IL2rg^{-/-} Double-Knockout Mice Were More Suitable for the Construction of Human Tumor Xenograft Model

Rag2/IL2rg^{-/-} double knockout impelled SCID in mice; next, we attempted to transplant various human-tissue-derived primary and passage tumor cells into these mutated mice. To establish the orthotopic transplantation of glioma, U87 cells were inoculated intracerebrally into the mutated mice. 3 weeks later, tumor growth could be observed on parenchyma of the brain significantly. H&E staining showed tumor cell arrangement of dense, spindle cells. Notably, nuclear hyperchromatism and

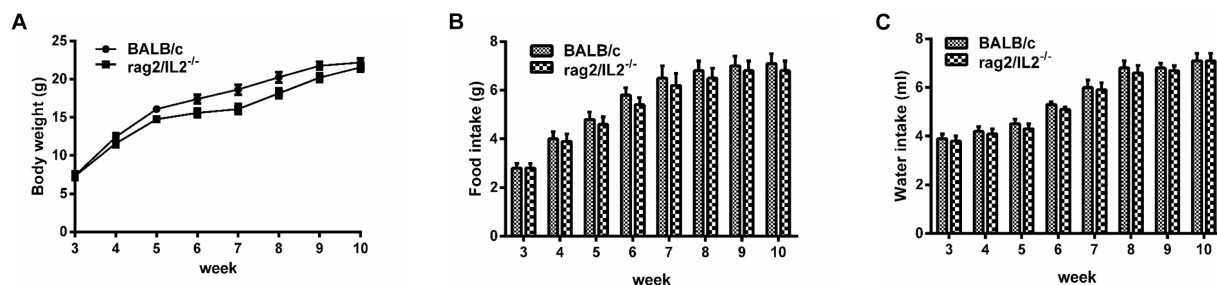


FIGURE 2 | Physiological behavior of *Rag2/IL2rg*^{-/-} gene double-knockout mice. Average body weight of mutated mice measured per week (A). Average food consumption of mutated mice measured per week (B). Average water consumption of mutated mice measured per week (C). Data are presented in the format of mean \pm standard deviation of 3 independent experiments performed in triplicate (* $p \leq 0.05$, ** $p \leq 0.01$, compared with controls).

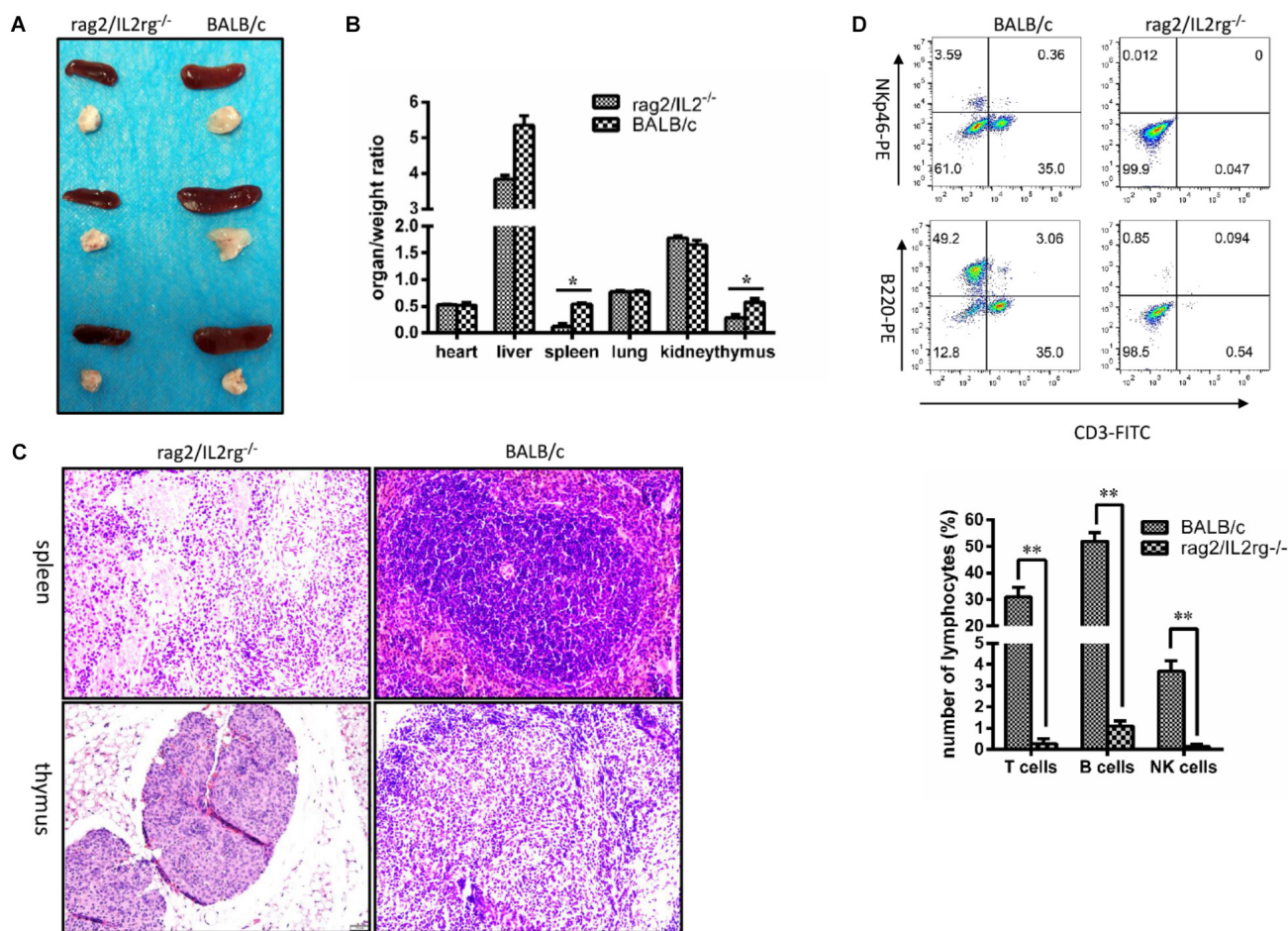
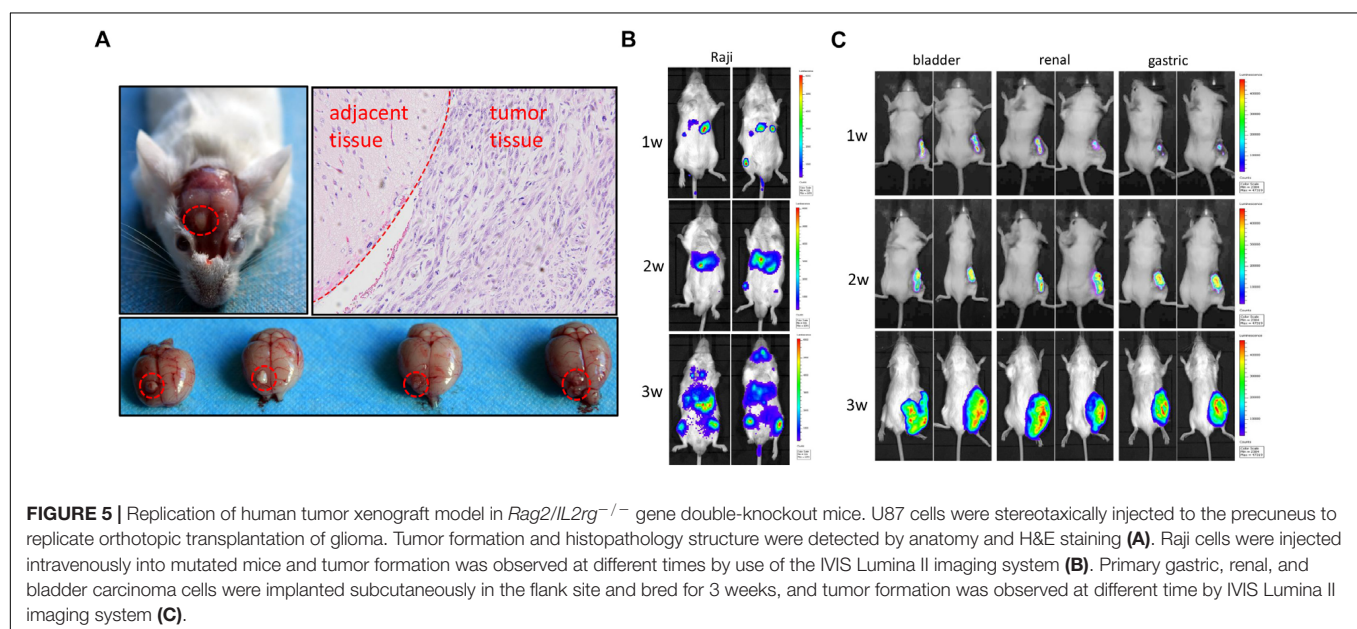
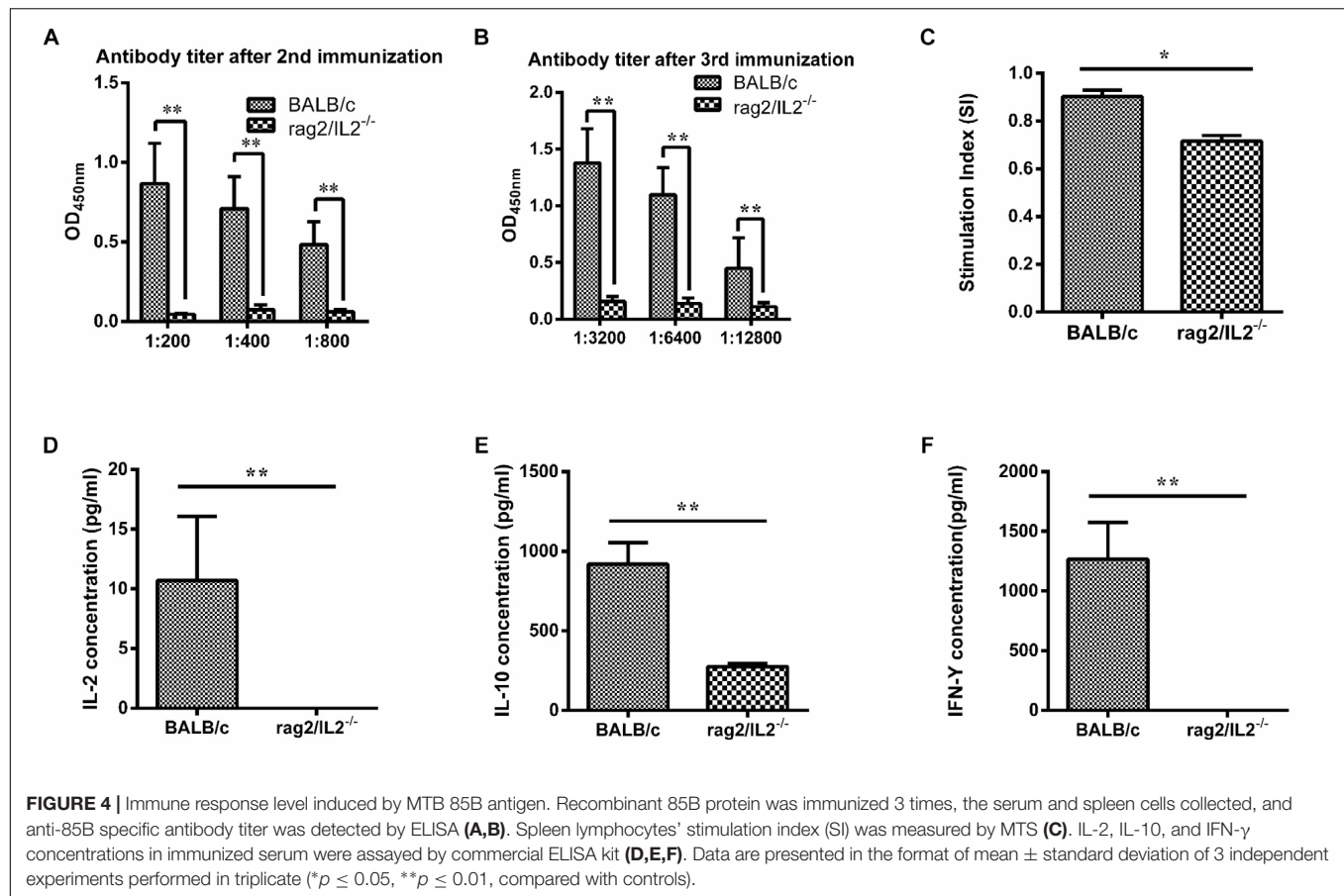


FIGURE 3 | Changes of thymus, spleen tissue structure, and number of lymphocytes in *Rag2/IL2rg*^{-/-} gene double-knockout mice. Six-week mutated and wild-type BALB/c mice were sacrificed. Various organs were extracted, and the organ/weight ratio calculated (A,B). Spleen and thymus tissues were sectioned and stained with the H&E method (C). Peripheral blood was collected from tail veins and, after dilution, sample were stained with FITC-CD3, PE-NKp46, and APC-220 antibodies and detected by flow cytometry FCM (D). Data are presented in the format of mean \pm standard deviation of 3 independent experiments performed in triplicate (* $p \leq 0.05$, ** $p \leq 0.01$, compared with controls).

pathological mitosis were common in carcinoma, while abundant hemorrhage and necrosis were also observed in this region. Although tumor cell infiltration was displayed in the junction

region, there was a clear boundary between carcinoma and paracancerous tissue, which was consistent with the pathological characteristics of glioma (Figure 5A). We also used lymphoma



Raji cells to replicate a hematopoietic tumor model: as shown in Figure 5B, Raji cells penetrated the blood-brain barrier and distributed all over the body after intravenous injection, including in the brain (Figure 5B). Primary cultured cells of

bladder cancer, renal cancer, and gastric cancer from clinical patients were transplanted subcutaneously, and the volume of xenotransplanted tumors was increased along with the time (Figure 5C). Taken together, our findings suggest *Rag2/IL2rg*^{-/-}

double-knockout mice represent an ideal xenograft tumor model, as this mouse type showed compromised growth of various tissue-derived cancer cells and different inoculation methods.

DISCUSSION

In present study, severe combined immunodeficient mice were prepared by CRISPR/Cas9-mediated *Rag2* and *IL2rg* mutation. This mice was produced from clear genetic background of BALB/c inbred strain, with homozygosity as the main characteristic. Although severe combined immunodeficient mice could also be obtained by mating *Rag2* with *IL2rg* mutated mice (Belizário, 2009), but this mating might increase heterozygosity since *Rag2* and *IL2rg* mutated mice were from different strains. In addition, the biological characteristics of this immunodeficient mice were studied, the mice not only developed abnormal lymphatic organs, leading to number of immune cells decreased, but also could not induce immune response even stimulated with recombinant antigen, resulting in immune response defects. Interestingly, this attenuant immune response was more susceptible to compromising tumor xenotransplantation, which made this mice was more adapted to tumor xenograft model.

Gene mutation led to the production of an immunodeficiency animal. *Foxn1* gene mutation brought about a T-lymphocyte-mediated cellular immune response defect of hairless nude mice (Vaidya et al., 2016), *Xid* gene mutations produced a B-lymphocyte-mediated humoral immune response obstacle of CBA/N mice (Szymczak et al., 2013), and *beige* gene mutations also produced NK cell dysfunction of beige mice (Pflumio et al., 1989). At this point, these immunodeficiency animals showed only a single immune cell disorder. Gene mutations could additionally cause a variety of immune dysfunctions, with *prkdc* and *Rag2* spontaneous gene mutations having resulted in SCID mice. Although there were some similarities between these genes, *Rag2* was a vital kinase in the V(D)J recombination. T- and B-cell development was retarded when the *Rag2* gene was mutated because of V(D)J recombination (Carmona et al., 2016). At this moment, T-cell development was blocked at the immature CD3⁺CD4⁺CD8⁺CD25⁺ stage and B-cells were blocked at the B220⁺CD43⁺ IgM⁺ progenitor B-cell stage (Ricetto et al., 2014; von Muenchow et al., 2017). The difference was that *prkdc* mutation led to immune leakiness in 20% of animals at 12 weeks of age, where immunoglobulin could be detected in the serum of *prkdc*-mutated mice (Danska et al., 1996; Katano et al., 2011). This interfered with the experimental results, but there was no immune leakiness in the *Rag2* gene mutation; that was the reason for why we selected the *Rag2* gene in this study.

Immunodeficient mice have been used to construct humanized animal models after the transplantation of human cells or tissues, so we attempted to explore the possibility of building a humanized animal after transplantation model involving human CD34⁺ hematopoietic stem cells and NK cells on *Rag2/IL2rg*^{-/-} mice. Unfortunately, transplanted CD34⁺ hematopoietic stem cells and NK cells abolished the

ability of their survival and proliferation in mutated mice (data not shown), suggesting that mice with *Rag2/IL2rg*^{-/-} gene mutation were not suitable for the construction of humanized mice. Several reasons might be considered as answers for this phenomenon: (1) sgRNAs of *Rag2/IL2rg* produced indel mutation in both of these genes, which induced small-fragment deletion, rather than the absence of large fragments. Although indel also could prompt frameshift mutation in *Rag2/IL2rg* genes, as compared with large-fragment absence, small-fragment mutation of indel seems less impactful on T- or B-cells. The utility of dual sgRNAs in follow-up studies to achieve the loss of large fragments of *Rag2/IL2rg* gene might be a better solution (Song et al., 2016). Additionally, (2) *prkdc*-mutated NOG or NSG mice were currently the best used humanized mice, although there was immune leakiness after *Prkdc* mutation. From the point of view of humanized mice construction, *prkdc* mutation was more suitable than *Rag2*, and (3) except for T-cells, B-cells, NK cells, neutrophils, macrophages, and cytokines could also cause a reaction of graft-versus-host disease, though the *Rag2/IL2rg* mutation reduced the number of neutrophils, lymphocytes in this study, residual granulocytes, macrophages, and cytokines might make it difficult for grafts to survive in mice. Although this immunodeficient mice were not suitable for humanized animal model, but it was a better tool for human tumor xenotransplantation, no matter orthotopic, hematopoietic, and xenotransplantatic tumor. Therefore, this immunodeficient mice was more propitious to application in tumor xenograft.

Except for applications in tumor research, immunodeficient mice could also be used in infection and immunity. Previous study demonstrated innate lymphoid cells were a critical role against *Clostridium difficile* infection based on data from *Rag1*^{-/-} single gene and *Rag2/IL2rg*^{-/-} double gene mutated mice (Abt et al., 2015). T, B cells were absent in *Rag1*^{-/-} and *Rag2/IL2rg*^{-/-} mice, however, innate lymphoid cells, like NK cells, Th17, and Th22 cells were normal in *Rag1*^{-/-} mice, but not in *Rag2/IL2rg*^{-/-} mice. Compared with *Rag1*^{-/-} mice, *Rag2/IL2rg*^{-/-} mice was succumbed to death after *C. difficile* infection owing to absence of NK cells, Th17, and Th22 cells. Thus innate lymphoid cells plays a protective role against *C. difficile* infection. Survival rate and cytokines expression, like IL-22, IL-17, and IFN- γ were assayed after challenge with *C. difficile* virulence strain. In present study, we evaluated the level of immune response, like antibodies titer, lymphocytes proliferation index, Th1 and Th2 cytokines after immunization with recombinant MTB 85B. Similarity, both mice abolished abilities of immune response, no matter immunization with virulence strain or recombinant MTB antigen. The difference was detection indexes, the immune properties of innate lymphoid cells was not investigated in this manuscript. Although we didn't compare the difference of this two types mice, the immune properties should be similarity because both mice was mutated on same genes.

In summary, we have constructed a *Rag2/IL2* gene mutant mouse model using the CRISPR/Cas9 gene editing technology. The *Rag2/IL2* gene mutation did not affect the normal physiological behavior of mice, but the mutated mice displayed

the typical characteristics of immunodeficiency. This mouse model could be used as a good animal model option in tumor research and other related fields.

AUTHOR CONTRIBUTIONS

HZ and ZC conceived the project. HZ designed the experiments. YoZ, PL, ZX, CS, YB, XS, YaZ, XW, LL, and XZ performed the experiments and prepared the manuscript. HZ and ZC

supervised the study and contributed reagents and materials. All authors contributed to data analysis.

FUNDING

This work was supported by Scientific and technological resources coordination project of Shaanxi Province (2018PT-03) and Special fund for military laboratory animals (SYDW(2016)001).

REFERENCES

- Abt, M. C., Lewis, B. B., Caballero, S., Xiong, H., Carter, R. A., Susac, B., et al. (2015). Innate immune defenses mediated by two ILC subsets are critical for protection against acute clostridium difficile infection. *Cell Host. Microbe* 18, 27–37. doi: 10.1016/j.chom.2015.06.011
- Aliyari, Z., Soleimanirad, S., Sayyah, M. M., Tayefi, N. H., and Nozad, C. H. (2015). IL2rg cytokines enhance umbilical cord blood CD34+ cells differentiation to T cells. *Adv. Pharm. Bull.* 5, 615–619. doi: 10.15171/apb.2015.083
- Barrangou, R. (2015). Diversity of CRISPR-Cas immune systems and molecular machines. *Genome Biol.* 16:247. doi: 10.1186/s13059-015-0816-9
- Belizário, J. E. (2009). Immunodeficient mouse models: an overview. *Open Immunol. J.* 2, 79–85. doi: 10.2174/1874226200902010079
- Cao, X., Kozak, C. A., Liu, Y. J., Noguchi, M., O'Connell, E., and Leonard, W. J. (1993). Characterization of cDNAs encoding the murine interleukin 2 receptor (IL-2R) gamma chain: chromosomal mapping and tissue specificity of IL-2R gamma chain expression. *Proc. Natl. Acad. Sci. U.S.A.* 90, 8464–8468. doi: 10.1073/pnas.90.18.8464
- Carmona, L. M., Fugmann, S. D., and Schatz, D. G. (2016). Collaboration of RAG2 with RAG1-like proteins during the evolution of V(D)J recombination. *Genes Dev.* 30, 909–917. doi: 10.1101/gad.278432.116
- Clark, E. A., Shultz, L. D., and Pollack, S. B. (1981). Mutations in mice that influence natural killer (NK) cell activity. *Immunogenetics* 12, 601–613. doi: 10.1007/bf01561700
- Corneo, B., Moshous, D., Gungor, T., Wulffraat, N., Philippot, P., Le Deist, F. L., et al. (2001). Identical mutations in RAG1 or RAG2 genes leading to defective V(D)J recombinase activity can cause either T-B-severe combined immune deficiency or Omenn syndrome. *Blood* 97, 2772–2776. doi: 10.1182/blood.v97.9.2772
- Danska, J. S., Holland, D. P., Mariathasan, S., Williams, K. M., and Guidos, C. J. (1996). Biochemical and genetic defects in the DNA-dependent protein kinase in murine scid lymphocytes. *Mol. Cell. Biol.* 16, 5507–5517. doi: 10.1128/mcb.16.10.5507
- Esmail, M. Y., Qi, P., Connor, A. B., Fox, J. G., and Garcia, A. (2016). Generating chimeric mice by using embryos from nonsuperovulated BALB/c mice compared with superovulated BALB/c and Albino C57BL/6 mice. *J. Am. Assoc. Lab. Anim. Sci.* 55, 400–405.
- Flanigan, S. P. (1966). 'Nude', a new hairless gene with pleiotropic effects in the mouse. *Genet. Res.* 8, 295–309.
- Greiner, D. L., Hesselton, R. A., and Shultz, L. D. (1998). SCID mouse models of human stem cell engraftment. *Stem Cells* 16, 166–177. doi: 10.1002/stem.160166
- Hsu, P. D., Lander, E. S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262–1278. doi: 10.1016/j.cell.2014.05.010
- Ishino, Y., Krupovic, M., and Forterre, P. (2018). History of CRISPR-cas from encounter with a mysterious repeated sequence to genome editing technology. *J. Bacteriol.* 200:e00580-17. doi: 10.1128/JB.00580-17
- Ito, M., Hiramatsu, H., Kobayashi, K., Suzue, K., Kawahata, M., Hioki, K., et al. (2002). NOD/SCID/gamma(c)(null) mouse: an excellent recipient mouse model for engraftment of human cells. *Blood* 100, 3175–3182. doi: 10.1182/blood-2001-12-0207
- Jansen, R., Embden, J. D., Gaastra, W., and Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* 43, 1565–1575. doi: 10.1046/j.1365-2958.2002.02839.x
- Katano, I., Ito, R., Eto, T., Aiso, S., and Ito, M. (2011). Immunodeficient NOD-scid IL-2Rgamma(null) mice do not display T and B cell leakiness. *Exp. Anim.* 60, 181–186. doi: 10.1538/expanim.60.181
- Klaus, G. G., Holman, M., Johnson-Leger, C., Elgueta-Karstegl, C., and Atkins, C. (1997). A re-evaluation of the effects of X-linked immunodeficiency (xid) mutation on B cell differentiation and function in the mouse. *Eur. J. Immunol.* 27, 2749–2756. doi: 10.1002/eji.1830271102
- Lu, Y., Kang, J., Ning, H., Wang, L., Xu, Y., Xue, Y., et al. (2018). Immunological characteristics of *Mycobacterium tuberculosis* subunit vaccines immunized through different routes. *Microb. Pathog.* 125, 84–92. doi: 10.1016/j.micpath.2018.09.009
- McDaniel, M. B., and Grisham, M. B. (2018). Humanizing the mouse immune system to study splanchnic organ inflammation. *J. Physiol.* 596, 3915–3927. doi: 10.1113/JP275325
- Mojica, F., and Montolieu, L. (2016). On the origin of CRISPR-Cas technology: from prokaryotes to mammals. *Trends Microbiol.* 24, 811–820. doi: 10.1016/j.tim.2016.06.005
- Notarangelo, L. D., Kim, M. S., Walter, J. E., and Lee, Y. N. (2016). Human RAG mutations: biochemistry and clinical implications. *Nat. Rev. Immunol.* 16, 234–246. doi: 10.1038/nri.2016.28
- Pflumio, F., Fonteneau, P., Gaveriaux, C., Cammisuli, S., and Loo, F. (1989). The C57BL/6 nude, beige mouse: a model of combined T cell and NK effector cell immunodeficiency. *Cell Immunol.* 120, 218–229. doi: 10.1016/0008-8749(89)90189-5
- Puck, J. M., Pepper, A. E., Henthorn, P. S., Candotti, F., Isakov, J., Whitwam, T., et al. (1997). Mutation analysis of IL2RG in human X-linked severe combined immunodeficiency. *Blood* 89, 1968–1977.
- Ricetto, A. G., Buzolin, M., Fernandes, J. F., Traina, F., Barjas-de-Castro, M. L., Silva, M. T., et al. (2014). Compound heterozygous RAG2 mutations mimicking hyper IgM syndrome. *J. Clin. Immunol.* 34, 7–9. doi: 10.1007/s10875-013-9956-4
- Sampson, T. R., and Weiss, D. S. (2014). CRISPR-Cas systems: new players in gene regulation and bacterial physiology. *Front. Cell. Infect. Microbiol.* 4:37. doi: 10.3389/fcimb.2014.00037
- Sander, J. D., and Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32, 347–355. doi: 10.1038/nbt.2842
- Schorpp, M., Hofmann, M., Dear, T. N., and Boehm, T. (1997). Characterization of mouse and human nude genes. *Immunogenetics* 46, 509–515. doi: 10.1007/s002510050312
- Shinkai, Y., Rathbun, G., Lam, K. P., Oltz, E. M., Stewart, V., Mendelsohn, M., et al. (1992). RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* 68, 855–867. doi: 10.1016/0092-8674(92)90029-c
- Shultz, L. D., Ishikawa, F., and Greiner, D. L. (2007). Humanized mice in translational biomedical research. *Nat. Rev. Immunol.* 7, 118–130. doi: 10.1038/nri2017
- Shultz, L. D., Lyons, B. L., Burzenski, L. M., Gott, B., Chen, X., Chaleff, S., et al. (2005). Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells. *J. Immunol.* 174, 6477–6489. doi: 10.4049/jimmunol.174.10.6477

- Shultz, L. D., Schweitzer, P. A., Christianson, S. W., Gott, B., Schweitzer, I. B., Tennent, B., et al. (1995). Multiple defects in innate and adaptive immunologic function in NOD/LtSz-scid mice. *J. Immunol.* 154, 180–191.
- Song, Y., Yuan, L., Wang, Y., Chen, M., Deng, J., Lv, Q., et al. (2016). Efficient dual sgRNA-directed large gene deletion in rabbit with CRISPR/Cas9 system. *Cell. Mol. Life Sci.* 73, 2959–2968. doi: 10.1007/s00018-016-2143-z
- Szymczak, W. A., Davis, M. J., Lundy, S. K., Dufaud, C., Olszewski, M., and Pirofski, L. A. (2013). X-linked immunodeficient mice exhibit enhanced susceptibility to *Cryptococcus neoformans* infection. *mBio* 4:e00265-13. doi: 10.1128/mBio.00265-13
- Tschaharganeh, D. F., Lowe, S. W., Garippa, R. J., and Livshits, G. (2016). Using CRISPR/Cas to study gene function and model disease in vivo. *FEBS J.* 283, 3194–3203. doi: 10.1111/febs.13750
- Vaidya, H. J., Briones, L. A., and Blackburn, C. C. (2016). FOXP1 in thymus organogenesis and development. *Eur. J. Immunol.* 46, 1826–1837. doi: 10.1002/eji.201545814
- von Muenchow, L., Tsapogas, P., Alberti-Servera, L., Capoferri, G., Doelz, M., Rolink, H., et al. (2017). Pro-B cells propagated in stromal cell-free cultures reconstitute functional B-cell compartments in immunodeficient mice. *Eur. J. Immunol.* 47, 394–405. doi: 10.1002/eji.201646638
- Wang, P., Wang, L., Zhang, W., Bai, Y., Kang, J., Hao, Y., et al. (2014). Immunotherapeutic efficacy of recombinant *Mycobacterium smegmatis* expressing Ag85B-ESAT6 fusion protein against persistent tuberculosis infection in mice. *Hum. Vaccin. Immunother.* 10, 150–158. doi: 10.4161/hv.26171
- Wilson, A., Held, W., and MacDonald, H. R. (1994). Two waves of recombinase gene expression in developing thymocytes. *J. Exp. Med.* 179, 1355–1360. doi: 10.1084/jem.179.4.1355
- Zhang, H., Peng, P., Miao, S., Zhao, Y., Mao, F., Wang, L., et al. (2010). Recombinant *Mycobacterium smegmatis* expressing an ESAT6-CFP10 fusion protein induces anti-mycobacterial immune responses and protects against *Mycobacterium tuberculosis* challenge in mice. *Scand. J. Immunol.* 72, 349–357. doi: 10.1111/j.1365-3083.2010.02448.x
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Zhao, Liu, Xin, Shi, Bai, Sun, Zhao, Wang, Liu, Zhao, Chen and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Enhancement of Precise Gene Editing by the Association of Cas9 With Homologous Recombination Factors

Ngoc-Tung Tran^{1†}, Sanum Bashir^{1,2†}, Xun Li¹, Jana Rossius^{1,2}, Van Trung Chu^{1,2}, Klaus Rajewsky¹ and Ralf Kühn^{1,2*}

¹ Max-Delbrück-Centrum für Molekulare Medizin, Berlin, Germany, ² Berlin Institute of Health, Berlin, Germany

OPEN ACCESS

Edited by:

David Jay Segal,
University of California, Davis,
United States

Reviewed by:

Robin Ketteler,
University College London,
United Kingdom
Ashok S. Bhagwat,
Wayne State University, United States

*Correspondence:

Ralf Kühn
ralf.kuehn@mdc-berlin.de

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Genomic Assay Technology,
a section of the journal
Frontiers in Genetics

Received: 30 November 2018

Accepted: 05 April 2019

Published: 30 April 2019

Citation:

Tran N-T, Bashir S, Li X,
Rossius J, Chu VT, Rajewsky K and
Kühn R (2019) Enhancement
of Precise Gene Editing by
the Association of Cas9 With
Homologous Recombination Factors.
Front. Genet. 10:365.
doi: 10.3389/fgene.2019.00365

The CRISPR-Cas9 system is used for genome editing in mammalian cells by introducing double-strand breaks (DSBs) which are predominantly repaired via non-homologous end joining (NHEJ) or to lesser extent by homology-directed repair (HDR). To enhance HDR for improving the introduction of precise genetic modifications, we tested fusion proteins of Cas9 nuclease with HDR effectors to enforce their localization at DSBs. Using a traffic-light DSB repair reporter (TLR) system for the quantitative detection of HDR and NHEJ events in human HEK cells we found that Cas9 fusions with CtIP, Rad52, and Mre11, but not Rad51C promote HDR up to twofold in human cells and significantly reduce NHEJ events. We further compared, as an alternative to the direct fusion with Cas9, two components configurations that associate CtIP fusion proteins with a Cas9-SunTag fusion or with guide RNA that includes MS2 binding loops. We found that the Cas9-CtIP fusion and the MS2-CtIP system, but not the SunTag approach increase the ratio of HDR/NHEJ 4.5–6-fold. Optimal results are obtained by the combined use of Cas9-CtIP and MS2-CtIP, shifting the HDR/NHEJ ratio by a factor of 14.9. Thus, our findings provide a simple and effective tool to promote precise gene modifications in mammalian cells.

Keywords: Cas9, gene editing, homologous recombination, CtIP, CRISPR

INTRODUCTION

The RNA guided Cas9 nuclease is used to create targeted double-strand breaks (DSBs) in the genome of mammalian cells and represents a versatile tool for genome editing (Barrangou and Doudna, 2016; Komor et al., 2016). CRISPR-Cas9 mediated DSBs are repaired by either the non-homologous end joining (NHEJ) repair pathway that leads to randomly sized small deletions or insertions (Indels), or by homology-directed repair (HDR) enabling precise sequence modifications that are copied from a repair template. Since HDR requires the presence of a repair template and is restricted to the S and G2 phases of the cell cycle it occurs less frequently than NHEJ. This presents a barrier to applications that rely on precise sequence modifications, such as the correction of mutations in somatic gene therapy or the modeling of disease mutations. To reinforce precise gene editing at Cas9 induced DSBs, tools for shifting the repair pathway choice in favor of HDR must be developed. DSB repair pathway choice is largely determined by the competition between the

53BP1 and BRCA1 regulator proteins, triggering either the protection or resection of DSB ends and the subsequent engagement of the NHEJ or HDR pathway, respectively (**Figure 1A**; Daley and Sung, 2014; Gupta et al., 2014; Zimmermann and de Lange, 2014). 53BP1 is recruited to DSBs by recognition of the Ubiquitin mark at Lysine 15 of histone H2A (Fradet-Turcotte et al., 2013) in chromatin flanking the break sites. 53BP1 blocks CtIP-based end resection (Bunting et al., 2010) and recruits Rif1 and the Shieldin complex (Mirman et al., 2018), which further block end resection and inhibit BRCA1 accumulation (Escribano-Díaz et al., 2013; Zimmermann et al., 2013). In contrast, the HDR pathway requires the exclusion of 53BP1 and the resection of DSB ends in order to be initiated. During the S/G2 phase, BRCA1 excludes Rif1 from DSB repair foci, and recruits CtIP and the MRE11-Rad50-NBS1 (MRN) complex. This complex initiates a cleavage step which is then further respected at the 5' end by Exo1 (Sartori et al., 2007; Symington and Gautier, 2011; Symington, 2016) extending on each side of the DSB (Zakharyevich et al., 2010). The exposed single-stranded DNA (ssDNA) is protected by binding of RPA1 that is subsequently replaced by Rad51 through the action of BRCA2 and Rad52, forming a nucleofilament competent for homology search and strand invasion (Liu et al., 2011).

Previous approaches to enhance HDR include enrichment of cells in the S/G2 phase (Lin et al., 2014; Yang et al., 2016), restriction of Cas9 activity to the S/G2 phase (Gutschner et al., 2016; Howden et al., 2016), inhibition of the NHEJ key molecules DNA ligase IV (Chu et al., 2015; Maruyama et al., 2015) or 53BP1 (Paulsen et al., 2017; Canny et al., 2018) and the fusion of Cas9 with CtIP (**Figure 1B**; Charpentier et al., 2018). Here we extended the Cas9 fusion approach to the HDR key proteins Mre11, CtIP, RPA1, Rad51C, and Rad52. Using CtIP, we compared the direct fusion with Cas9 to alternative molecular configurations that associate CtIP with guide RNA that includes MS2 binding loops (**Figure 1C**) or with a Cas9-SunTag fusion protein (**Figure 1D**), previously developed for binding of transcriptional activators to dCas9 (Tanenbaum et al., 2014; Konermann et al., 2015). In addition, we compared wildtype CtIP, that is activated in the S/G2 cell cycle phases through phosphorylation at the Threonine residue 847, with the phosphomimetic T847E mutant that is also active in the G1 phase (Huertas and Jackson, 2009). Using a traffic-light DSB repair reporter (TLR) system for the quantitative detection of HDR and NHEJ events in human HEK cells we found that both the fusion of CtIP with Cas9 and the MS2, but not the SunTag system strongly shift the balance of DSB repair pathway choice toward HDR. Best results were obtained by the combined use of both systems, increasing the HDR/NHEJ ratio 14.9-fold.

RESULTS

DSB Repair Modification by Cas9 Fusion Proteins

To quantitatively determine CRISPR/Cas9-induced DSB repair by HDR or NHEJ, we used a traffic light reporter (TLR) construct, integrated into the AAVS1 locus of human HEK293 cells (HEK^{TLR}) as previously described (Chu et al., 2015). Briefly,

the reporter cassette includes a CAG promoter for expression of a non-functional coding region for Yellow fluorescent (Venus) protein, disrupted by the replacement of codons 117–152 with a 23 bp gRNA target sequence from the mouse *Rosa26* locus (sgRosa26), followed by a P2A peptide and the coding region for a red fluorescent (TagRFP) protein in a reading frame shifted by 2 bp (**Figure 2A**). If an intact Venus coding sequence is provided as a template and the repair of DSBs occurs via the HDR pathway the reporter cells are detected by the expression of Venus. CRISPR/Cas9-induced DSBs in the target region that are repaired via NHEJ and acquire Indels resulting into the shift of translation into the frame (+2) of P2A-RFP are detectable by the expression of RFP in reporter cells. Analysis by the inDelphi tool (Shen et al., 2018) for the repair of the *Rosa26* target site in HEK293 cells predicts a frequency of 16% of products in the +2 frame. Therefore, Venus positive reporter cells represent all HDR events but the number of RFP positive cells in a sample indicates only a fraction of NHEJ repair events. For the assessment of HDR modifiers we constructed N- or C-terminal fusions of Cas9 with the coding regions of human MRE11A, CtIP (wildtype or the phosphomimetic T847E mutant), RPA1, Rad51C, or Rad52 separated by a flexible linker of 16 residues (**Figure 2B**). To monitor the effects of Cas9 fusions on DSBs repair pathways, we co-transfected HEK^{TLR} cells with plasmids expressing either Cas9 or Cas9 fusions, a vector for expression of sgRosa26 and a Blasticidine resistance gene together with the donor plasmid (pTLR-repair) for repair of the defective Venus reporter gene (**Figure 2C**). The transfected cells were selected with Blasticidine for the enrichment of transfected cells and the frequency of Venus⁺ and RFP⁺ cells was analyzed 4 days later by flow cytometry (**Supplementary Figure 1**) in 4 independent samples. The results were used to calculate mean values and standard deviation. The ratio of Venus⁺ versus RFP⁺ cells is used as a relative index for DSB repair of the reporter by HDR or by NHEJ events resulting into the +2-reading frame. As shown in **Figure 2D**, upon expression of Cas9 we observed 0.95% of Venus⁺ and 7.55% of RFP⁺ cells in the ratio of 0.13 (sample 1). The expression of a N- or C-terminal fusion protein of Cas9 with Rad52 both lead to the increase of Venus⁺ and the decrease of RFP⁺ cells, shifting the Venus/RFP ratio to values of 0.35 and 0.42, respectively. The expression of a N- or C-terminal fusion protein of Cas9 with MRE11A lead to a higher increase of Venus⁺ cells but a lower decrease of RFP⁺ cells (samples 7 and 8), exhibiting Venus/RFP ratios of 0.35 and 0.37, respectively. The expression of a C-terminal fusion of Cas9 with CtIP or its N-terminal fusion with the CtIP (T847E) mutant (samples 5 and 6) increased the level of Venus⁺ cells up to 2.6%, showing Venus/RFP ratios of 0.42 and 0.53, respectively. In contrast to Rad52, MRE11A, and CtIP, the Cas9-RPA1 fusion protein (sample 4) lead to a smaller shift of the Venus/RFP ratio (0.22). As compared to the control (Cas9, sample 1) the increase of Venus⁺ cells and the decrease of RFP⁺ cells in samples 2–8 was significantly different ($p < 0.05$). Only the expression of the Cas9-Rad51C fusion (sample 9) resulted into levels of Venus⁺ cells and RFP⁺ cells that were not significantly different from the control (sample 1). These results show that the use of Cas9 fusions with multiple

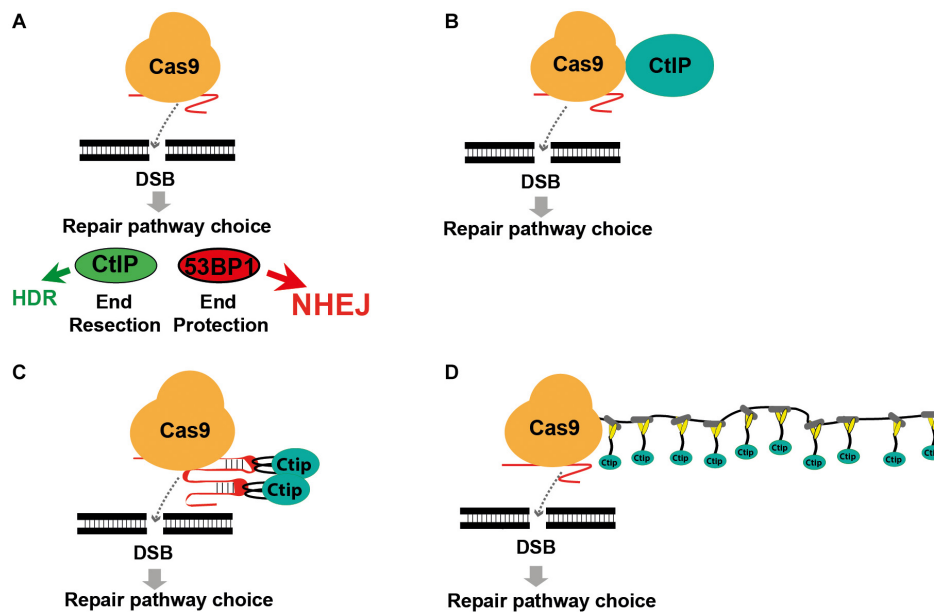


FIGURE 1 | Modifying DSB repair pathway choice by association of Cas9 with CtIP. **(A)** Cas9 induced double-strand breaks are either repaired by HDR or the prevailing NHEJ pathway. DSB repair pathway choice is largely determined by the competition between 53BP1 and BRCA1, triggering either end protection or the CtIP induced resection of DSB ends, followed by the engagement of the NHEJ or HDR pathway. **(B)** Use of a Cas9-CtIP fusion protein for shifting DSB repair choice toward HDR. **(C)** Association of Cas9 and CtIP via sgRNA loops that include MS2 phage aptamer sequences bound by fusion proteins of CtIP and the MS2 coat protein. **(D)** Association of a Cas9-SunTag fusion protein with multiple copies of a SunLigand-CtIP fusion protein.

proteins of the HDR pathway, specifically CtIP, MRE11A and Rad52, can be used to stimulate DSB repair by HDR up to 2.7-fold.

Cas9 Fusion Proteins in 53BP1 Knockout HEK^{TLR} Cells

The 53BP1 key regulator of NHEJ as well as its interaction partners Rif1 and the shieldin complex counteract DSB end resection. In order to elucidate the role of 53BP1 in the TLR assay with Cas9 fusion proteins we generated reporter cells harboring frameshift knockout alleles of the 53BP1 gene (HEK^{TLR/Δ53BP1}) using CRISPR-Cas9 (Supplementary Figure 2A). HEK^{TLR} and HEK^{TLR/Δ53BP1} reporter cells were transfected with expression vectors for Cas9 or Cas9 fusions, the sgRosa26 and a BFP reporter together with pTLR-repair (Figure 3A). The BFP⁺ population representing the transfected cells was analyzed 4 days later by flow cytometry (Supplementary Figure 2B) to determine the frequency of Venus⁺ and RFP⁺ cells. The results were used to calculate mean values shown in bar plots with standard deviation. As shown in Figure 3B, in HEK^{TLR} cells with wildtype alleles of 53BP1 the expression of Cas9-MRE11A, -Rad52, or -CtIP lead to a statistically significant ($p < 0.05$) increase of Venus⁺ cells and a decrease of RFP⁺ cells, as observed previously (Figure 2D). In contrast, the transfection of Cas9 into 53BP1 deficient HEK^{TLR/Δ53BP1} cells lead to twofold higher levels of Venus⁺ cells and a more than threefold reduction of RFP⁺ cells to 1.7% (Figure 3C, sample 1), as compared to 6.73% RFP⁺ in HEK^{TLR} cells (Figure 3B, sample 1). As compared to HEK^{TLR} cells the transfection of HEK^{TLR/Δ53BP1}

cells with Cas9-Mre11, -Rad52, and -CtIP wildtype fusion proteins did not significantly increase the levels of Venus⁺ cells (samples 2–4), whereas the levels of RFP⁺ cells fall below 2% and was not significantly different from the control (sample 1). Only the expression of Cas9-CtIP^{T847E} fusion protein lead to a significant increase of Venus⁺ cells. These results suggest that in wildtype HEK^{TLR} cells Cas9 fusion proteins with MRE11A, Rad52, or CtIP are efficiently counteracting the inhibitory action of 53BP1 on DSB end resection, leading to increased HDR repair. Nevertheless, Cas9 fusion proteins interfere only weakly with the NHEJ promoting activity of 53BP1 since the numbers of RFP⁺ cells in wildtype HEK cells were only moderately suppressed, but strongly reduced in 53BP1 knockout cells. In conclusion, we found that Cas9 fusions with MRE11A, Rad52, and CtIP are useful tools for increasing HDR mediated gene editing but only weak suppressors of NHEJ. Thus, experimental conditions for the concurrent increase of HDR together with NHEJ suppression would require an additional, active inhibition of 53BP1.

Targeting of the Beta-2 Microglobulin Gene in HEK Cells

In addition to the TLR reporter construct we assessed the effect of Cas9 fusion proteins on targeting a red fluorescent mCherry reporter gene into the last exon of the beta-2 microglobulin (B2M) gene of HEK cells. For targeting of the B2M locus we used a sgRNA against B2M (sgB2M) and a repair template vector for HDR (B2M-donor) that leads to the insertion of a 2A peptide and the mCherry coding region upstream of the

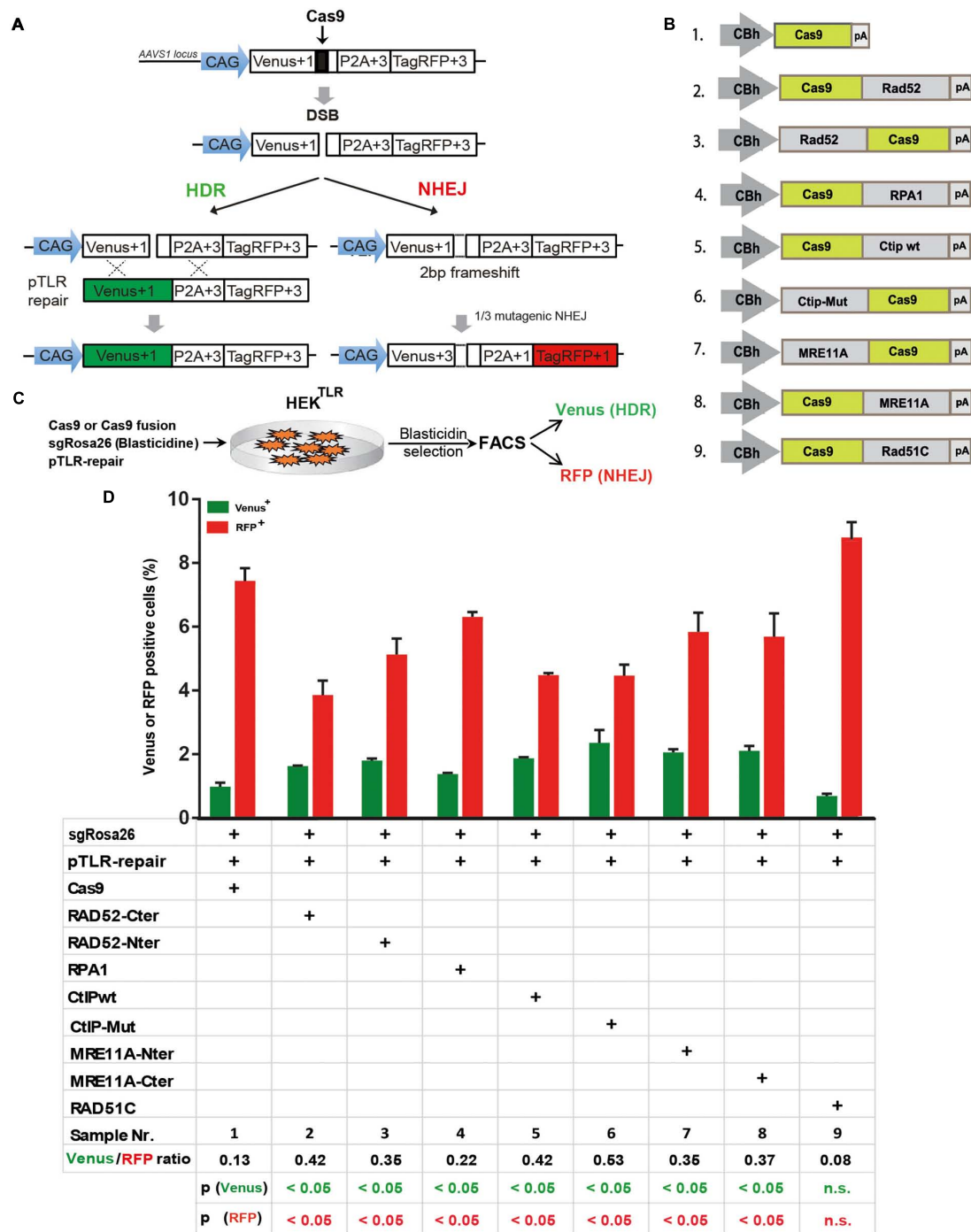


FIGURE 2 | DSB repair modification by Cas9 fusion proteins. **(A)** The traffic light reporter (TLR) construct indicates DSB repair by NHEJ or HDR and was integrated into the AAVS1 locus of HEK cells (HEK^{TLR}). Upon induction of a DSB in the defective Venus coding region, RFP is expressed upon NHEJ repair resulting into deletions that shift translation by 2 bp. Venus expression occurs upon HDR with a repair template vector that includes the intact Venus coding region (pTLR-repair). **(B)** Vectors for the expression of Cas9 or fusion proteins between the N- or C-terminal end of Cas9 and Rad52, RPA1, CtIP wildtype (wt), or the T847E mutant (mut), Mre11A or Rad51C, driven by the CBh promoter. pA – polyadenylation signal. **(C)** For DSB repair assays HEK^{TLR} reporter cells were cotransfected with expression vectors for Cas9 or Cas9 fusion proteins, sgRosa26, Blasticidine and pTLR-repair. After 4 days of Blasticidine selection the samples were analyzed by FACS for the presence of Venus and RFP positive cells. **(D)** Bar graph representation of Venus and RFP positive cells, indicating HDR or NHEJ repair of the TLR reporter, upon transfection with an expression vector for Cas9, or Cas9 in C-terminal (Cter) or N-terminal (Nter) fusion with Rad52, RPA1, CtIP wildtype (wt), CtIP mutant T847E (mut), MRE11A, or Rad51C. Bars show mean values of three independent samples with standard deviation. These values were used to calculate the ratio of Venus⁺ to RFP⁺ cells and *p*-values (*T*-test) to determine the significance in levels of Venus⁺ or RFP⁺ cells between samples 1 and 2–9. n.s., not significant, *p* > 0.05.

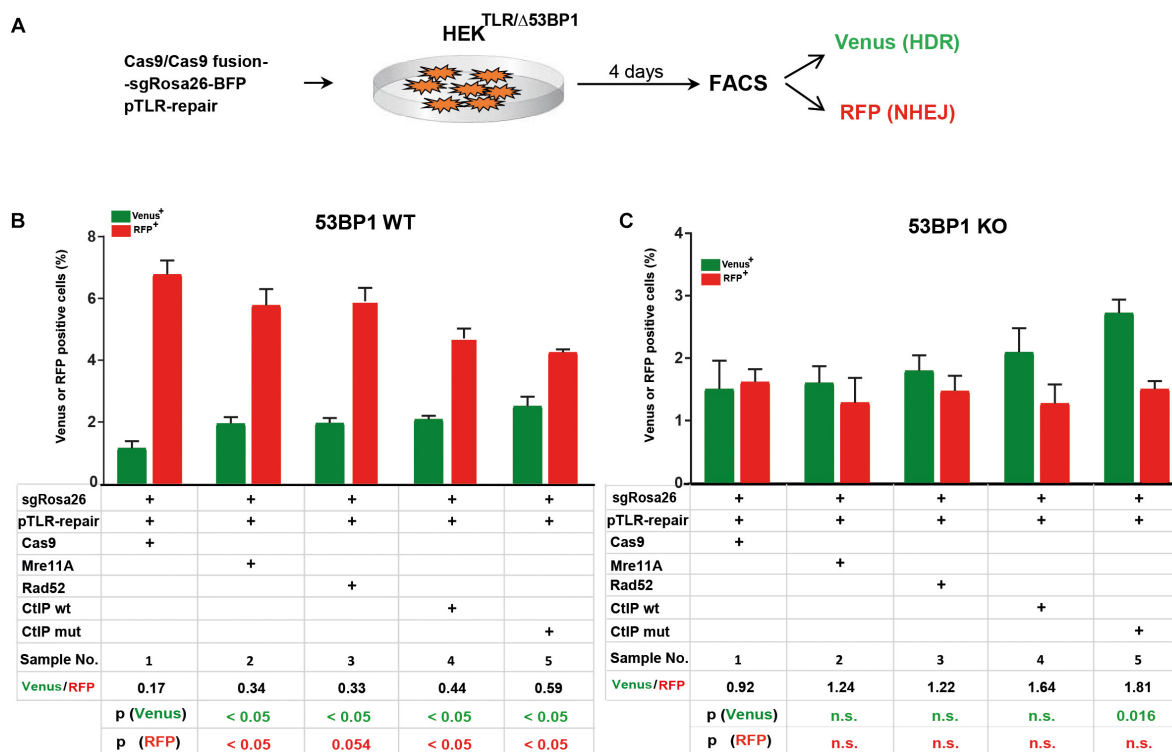


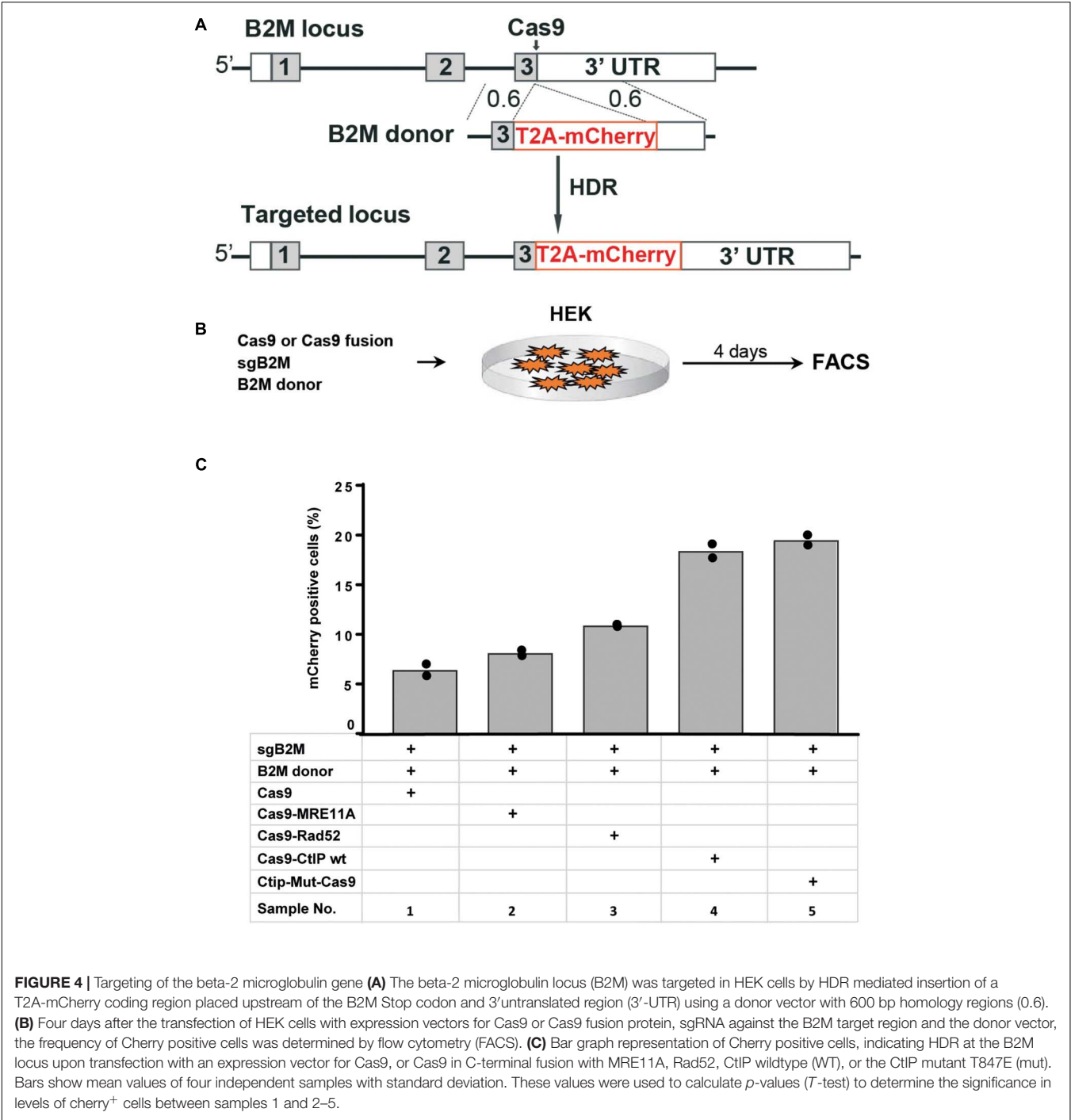
FIGURE 3 | Cas9 fusion proteins in 53BP1 knockout HEK^{TLR} cells. **(A)** For DSB repair assays in HEK^{TLR} and HEK^{TLR} cells harboring 53BP1 knockout alleles (HEK^{TLR} Δ53BP1), reporter cells were cotransfected in duplicate with expression vectors for Cas9 or Cas9 fusion proteins, sgRosa26, BFP and pTLR-repair. After 4 days the samples were analyzed by flow cytometry (FACS) for the presence of Venus and RFP positive cells indicating repair of the TLR construct by HDR or NHEJ. **(B)** Bar graph representation of Venus and RFP positive cells, indicating HDR or NHEJ repair of the reporter in HEK^{TLR} wildtype (53BP1-WT) cells or **(C)** HEK^{TLR} Δ53BP1 (53BP1-KO) cells, upon transfection with an expression vector for Cas9, or Cas9 in C-terminal fusion with MRE11A, Rad52, CtIP wildtype (wt), or the CtIP mutant T847E (mut). Bars show mean values of four (samples 1–3) or three (samples 4, 5) independent samples with standard deviation. These values were used to calculate the ratio of Venus⁺ to RFP⁺ cells and *p*-values (*T*-test) to determine the significance in levels of Venus⁺ or RFP⁺ cells between samples 1 and 2–5. n.s., not significant, *p* > 0.05.

B2M Stop codon (Figure 4A). HEK cells were transfected with expression vectors for Cas9 or Cas9 fusions, sgB2M and the B2M donor vector (Figure 4B). The transfected cell population was analyzed 4 days later by flow cytometry for the presence of Cherry⁺ cells (Supplementary Figure 3). The results were used to calculate mean values of 4 independent samples with standard deviation. As shown in Figure 4C, the transfection with Cas9, sgB2M, and B2M donor resulted into 6.4% Cherry⁺ cells. The use of expression vectors for Cas9-MRE11A or Cas9-Rad52 fusion protein moderately, but significantly (*p* < 0.05) increased the frequency of Cherry⁺ cells to 8.1 or 10.9%, whereas the expression of Cas9-CtIP or Cas9-CtIP(T847E) fusion protein more than doubled the number of Cherry⁺ cells to 18.4 or 19.5%, respectively (Figure 4C). These results show that in particular the Cas9-CtIP fusion proteins are able to support HDR at the endogenous B2M target locus in HEK cells.

Association of CtIP and Cas9 via the MS2 and SunTag Systems

Besides Cas9-CtIP fusion proteins we further assessed the MS2 and the SunTag system as alternative molecular configurations

for the association of CtIP to Cas9 induced DSBs (Figures 1C,D). In a first approach we fused CtIP with a monomer of the MS2 bacteriophage coat protein (MS2-CtIP) which binds as dimer to a MS2 derived 34 nt RNA aptamer sequence motif that can be included in the tetraloop and stem loop of sgRNAs, as described (Konermann et al., 2015; Supplementary Figures 4A,B). Since CtIP must associate via its N-terminus into a tetramer to become functional, the binding of two closely neighbored MS2-CtIP molecules at each MS2-sgRNA loop (Supplementary Figure 4B) may be inefficient or interfere with its tetramerization. Therefore, we also constructed a fusion of CtIP with a single chain MS2 dimer (MS2^{di}-CtIP), that binds to the MS2 aptamer RNA as a monomer (Peabody and Lim, 1996; Supplementary Figure 4C). In a second approach Cas9 was fused with 10 repeats of a GCN4 derived peptide motif (Cas9-SunTag) (Tanenbaum et al., 2014) that is recognized by a high affinity single-chain antibody, designated here as SunLigand (SunL) (Supplementary Figure 4D). For the association of CtIP with Cas9-SunTag we constructed an expression vector for a SunL-CtIP(T847E) fusion protein. For these assays we used HEK reporter cells harboring in the AAVS1 locus a modified TLR reporter construct (HEK^{TLR6}) that includes a Venus coding region disrupted by



the replacement of codons 95–97 with the same *Rosa26* derived target sequence used in the TLR reporter. Since HEK^{TLR6} cells are homozygous for the reporter construct a small population of double positive cells appears upon DSB induction, undergoing HDR repair on one reporter allele and a mutagenic NHEJ event on the other. For DSB repair assays the number of single positive Venus and RFP positive cells was determined, excluding double positive cells. For repair of the TLR6 reporter we used a template vector (pTLR-donor) that includes the

Venus coding sequence, excluding the start codon to prevent background expression. HEK^{TLR6} cells were cotransfected with plasmids for expression of Cas9, Cas9-CtIP or Cas9-SunTag driven by the CAG promoter together with pTLR-donor and the *Rosa26* specific sgRNA including two MS2 aptamer sequences [sgRosa26(MS2)]. Three days after transfection the frequency of Venus⁺ and RFP⁺ cells from triplicate samples was determined by flow cytometry and the mean values and standard deviations were calculated (**Figure 5A**). The ratio of Venus⁺/RFP⁺ cells

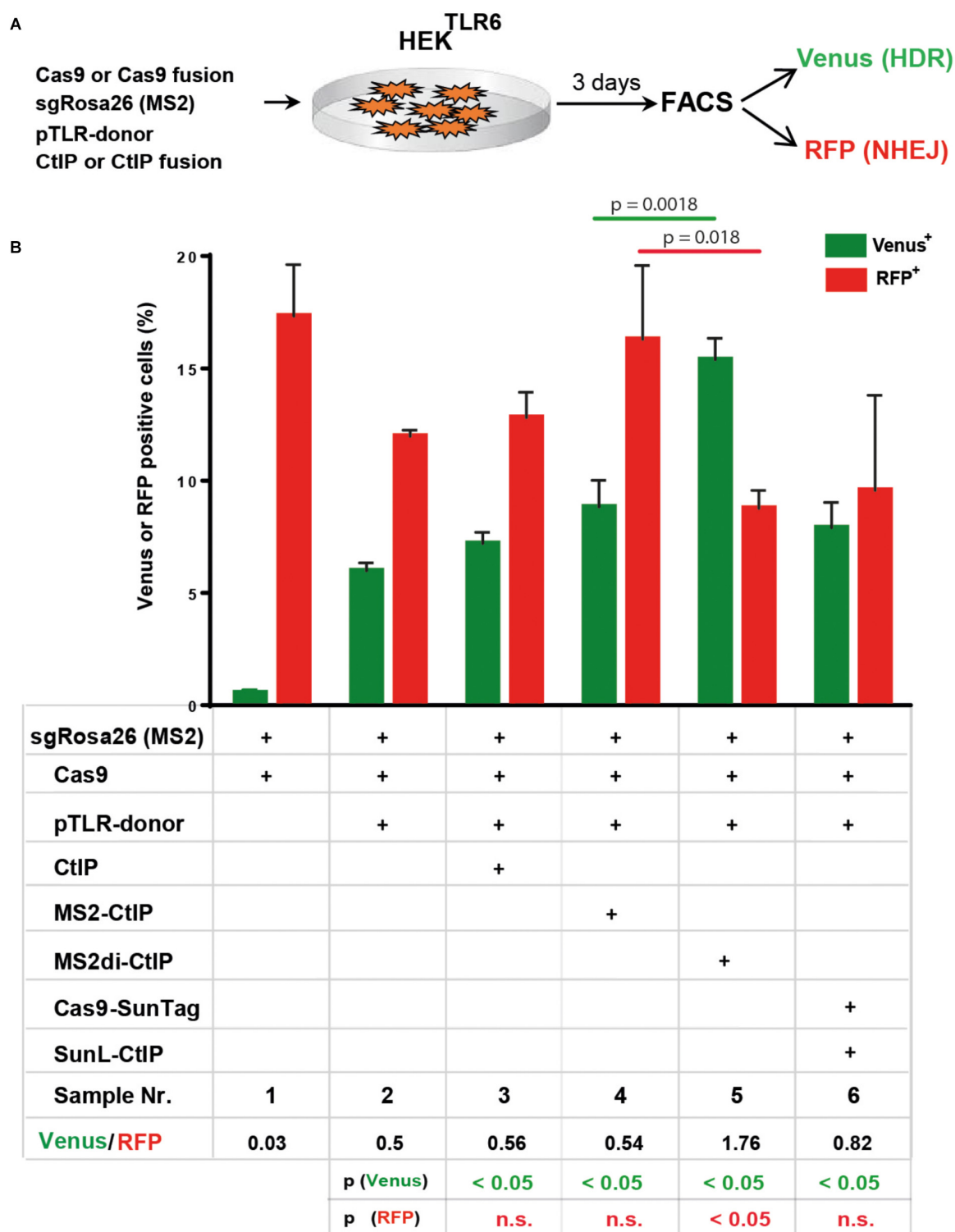


FIGURE 5 | DSB repair modification by MS2-CtIP and SunL-CtIP fusion proteins in HEK^{TLR6} reporter cells. **(A)** Expression vectors for Cas9 or Cas9-SunTag, sgRosa26(MS2), pTLR-donor and CtIP, MS2-CtIP, MS2^{di}-CtIP, or SunL-CtIP were cotransfected into HEK^{TLR6} cells and the frequency of Venus⁺ cells (green columns) and RFP⁺ cells (red columns), indicating DSB repair by HDR or NHEJ, was determined by flow cytometry (FACS) 72 h after transfection. **(B)** Transfection results are shown as bar graphs representing mean values with standard deviation of triplicate samples. Plasmids selected (+) for individual transfection samples are indicated in the table below. Mean values were used to calculate the ratio of Venus⁺ to RFP⁺ cells and *p*-values (*T*-test) to determine the significance in levels of Venus⁺ or RFP⁺ cells between samples 2 and 3–6 (table bottom) and to compare samples 4 and 5 (top horizontal lines). n.s., not significant, *p* > 0.05. Two independent replicates of the assay were performed that confirmed the results.

is used as an index for DSB repair choice by HDR or NHEJ. Of note, this relative value represents the ratio of all reporter HDR events to only a fraction of NHEJ mediated Indels that reconstitute the RFP reading frame. As shown in **Figure 5B**, the transfection of Cas9, sgRosa26(MS2) and pTLR-donor resulted into 5.6% Venus and 11.4% RFP cells (sample 2; Venus/RFP ratio: 0.5), whereas a control without pTLR-donor (sample 1) showed a Venus background of 0.4%. The expression of Cas9 and free CtIP (sample 3) yielded 6.8% Venus and 12.2% RFP cells, showing a similar ratio (0.56) as the sample with Cas9 alone. The expression of MS2-CtIP(T847E) lead to a moderate, but statistically significant increase of Venus⁺ cells with a ratio of 0.54 (sample 4). In contrast to MS2-CtIP(T847E), the expression of MS2^{di}-CtIP(T847E) strongly increased ($p = 0.0018$) the Venus⁺ cells to 14.6% and reduced the RFP⁺ cells ($p = 0.018$) to 8.3%, resulting into a Venus/RFP ratio of 1.76 (sample 5). In contrast to MS2^{di}-CtIP, the expression of Cas9-SunTag and SunL-CtIP did not increase the Venus⁺ cells (sample 6, ratio: 0.82), as compared to the sample with free CtIP. These results show that in fusion with CtIP only the MS2^{di}, but not the MS2 or the SunTag configuration, leads to a strong shift of the HDR/NHEJ balance upon DSB repair of the reporter. Since the MS2 and SunTag systems assemble two or more CtIP molecules in close proximity these configurations may interfere with CtIP oligomerization and the stimulation of end resection.

Comparison of Cas9-CtIP, MS2^{di}-CtIP, and SunL-CtIP Fusion Proteins

Next, we compared the performance of the Cas9-CtIP fusion protein with the MS2^{di}-CtIP and SunL-CtIP systems in HEK^{TLR6} cells to determine which approach is most effective for HDR stimulation. For a side-by-side comparison we transferred the Cas9-CtIP coding region into the same vector used for Cas9 expression in MS2^{di}-CtIP assays that includes a CAG promoter region and the sgRosa26(MS2) expression cassette (**Figure 6A**). As shown in **Figure 6B**, the transfection of Cas9, sgRosa26(MS2), and pTLR-donor resulted into 3.6% Venus and 9.2% RFP cells (sample 2; ratio: 0.38), whereas a control without pTLR-donor (sample 1) showed a Venus background of 0.7%. The expression of MS2^{di}-CtIP (sample 3) yielded 12.4% Venus and 7.2% RFP cells, showing a Venus/RFP ratio of 1.7. The expression of the Cas-CtIP fusion lead to a further increase of the Venus/RFP ratio to 2.31 although the absolute levels of Venus and RFP cells were decreased (sample 4). Interestingly, the combined expression of Cas-CtIP and MS2^{di}-CtIP lead to a further decrease of RFP⁺ cells to 2.1% (sample 5), resulting into an increased ratio of 5.67. The expression of Cas9-SunTag and SunL-CtIP lead to a Venus/RFP ratio of 1.62, but lower levels of Venus⁺ and RFP⁺ cells as compared to the sample with MS2^{di}-CtIP. These results show that both Cas9-CtIP and MS2^{di}-CtIP are shifting the ratio of HDR/NHEJ repair of the reporter construct. Best results were obtained when both systems were combined, leading to a significant decrease of RFP⁺ cells ($p < 0.05$) as compared to the use of Cas9-CtIP or MS2^{di}-CtIP alone, shifting the HDR/NHEJ balance of DSB repair at the reporter by a factor of 14.9.

DISCUSSION

For the improvement of HDR and precise gene editing at Cas9 induced DSBs we compared three approaches for associating HDR effector proteins with the nuclease or guide RNA. Firstly, we tested the ability of Cas9/HDR effector fusion proteins and found that CtIP but also Rad52 and MRE11A are effective for HDR stimulation, counteracting the inhibition of end resection by 53BP1 without strong suppression of NHEJ. Using CtIP as paradigm we compared the Cas9-CtIP fusion to methods based on anchoring of SunL-CtIP to a Cas9-SunTag protein or MS2-CtIP fusions to gRNA that includes MS2 recognition sequences. Our results show that the performance of the single chain MS2^{di}-CtIP fusion was comparable to the Cas9-CtIP direct fusion and that the combination of both proteins had an additive effect on NHEJ suppression, shifting the HDR/NHEJ repair of the reporter construct by a factor of 14.9. Our findings provide simple and effective tools to promote precise gene modifications in mammalian cells.

Previous studies based on HDR stimulation rarely explored the association of HDR effectors with Cas9/gRNA complexes, except for the work of Charpentier et al. (2018) that described the N-terminal fusion of CtIP with Cas9. In this study it was found that CtIP-Cas9 increased HDR at the AAVS1 locus in human cells twofold without reducing Indel formation and that the N-terminal oligomerization domain of CtIP is sufficient for HDR stimulation. Our results extend the analysis of Cas9-fusion proteins to the HDR effectors MRE11A and Rad52 and confirm the previous finding that the N-terminal but also the C-terminal fusion of CtIP to Cas9 stimulates HDR. In addition, we validated the MS2 aptamer system as a useful tool for HDR stimulation by the association of CtIP with sgRNA loops. In fusion with CtIP the MS2^{di} single chain configuration was more effective than MS2 monomers that require dimer formation for binding to the MS2 RNA aptamer. Whereas the MS2 tagged sgRNA provides two binding sites for MS2 fusion proteins, the SunTag system has previously shown utility for signal amplification by binding of up to 24 SunL-GFP molecules to a Cas9-SunTag fusion protein (Tanenbaum et al., 2014). In fusion with CtIP and Cas9-SunTag providing 10 SunL binding sites, we found the SunTag approach was less effective for HDR stimulation than the MS2 system, which might be caused by an interference with CtIP oligomerization in the presence of multiple localized molecules. A similar effect may account for the improved performance of the MS2^{di} single chain dimer as compared to the use of MS2-CtIP monomers. Nevertheless, it remains to be investigated whether MS2 monomers or the SunTag system provide valid options for the use of other HDR effectors which do not form oligomers.

Remarkably, we observed that the combined use of Cas9-CtIP and MS2^{di}-CtIP had a cumulative effect on NHEJ suppression but not on HDR stimulation, pointing to the enhanced counteraction against the initiation of NHEJ through 53BP1 by localization of three instead of two or a single CtIP molecule at the Cas9/sgRNA complex. This observation may allow further improvements of the toolbox for manipulation of the DSB repair pathway choice by the assembly of two cooperative HDR effector proteins at the DSB site, such as CtIP together with Rad52, MRE11A or other factors.

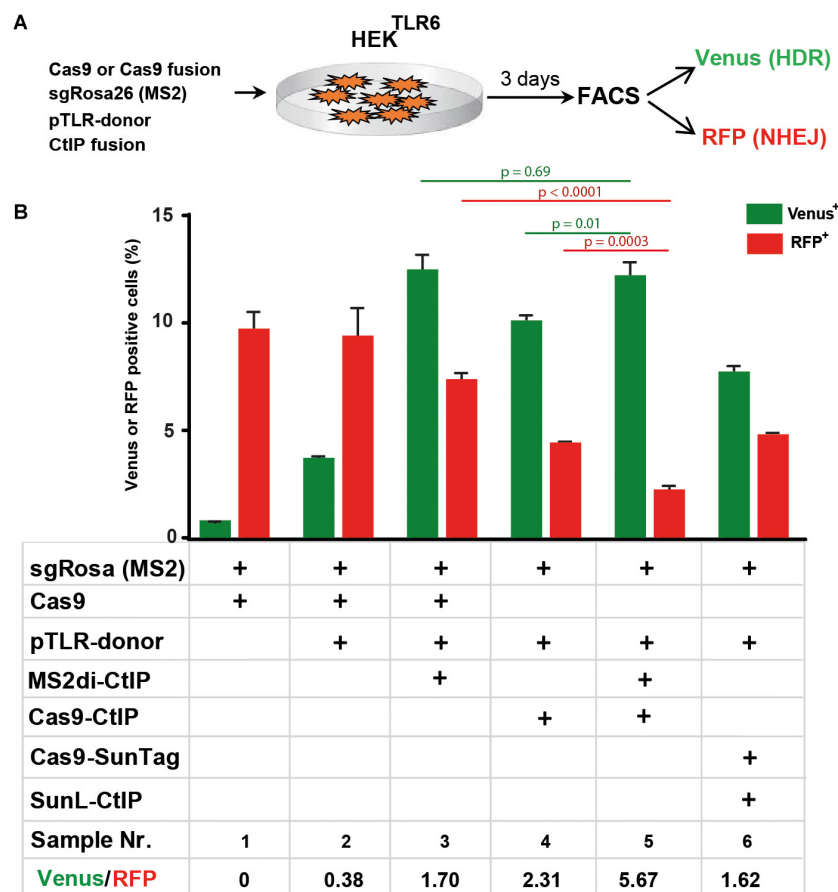


FIGURE 6 | Comparison of Cas9-CtIP, MS2^{di}-CtIP, and SunL-CtIP fusion proteins. **(A)** Expression vectors for Cas9, Cas9-CtIP, or Cas9-SunTag, sgRosa26(MS2), pTLR-donor, and MS2^{di}-CtIP or SunL-CtIP were cotransfected into HEK^{TLR6} cells and the frequency of Venus⁺ cells (green columns) and RFP⁺ cells (red columns), indicating DSB repair by HDR or NHEJ, was determined by FACS analysis 72 h after transfection. **(B)** Transfection results are shown as bar graphs representing mean values \pm standard deviation of triplicate samples. Plasmids selected (+) for individual transfections are indicated as in the table below. Mean values were used to calculate the ratio of Venus⁺ to RFP⁺ cells and *p*-values (*T*-test) to determine the significance in levels of Venus⁺ or RFP⁺ cells between samples 5 and 3 or 4 (top horizontal lines). Two independent replicates of the assay were performed that confirmed the results.

Furthermore, it may be possible to combine the N-terminal CtIP HDR enhancer domain with one or more active domains from other effectors into multifunctional fusions that are able to bypass rate limiting steps of HDR initiation or HDR processing. Besides the DSB repair pathway choice, the availability of donor templates at DSBs represents a limiting factor for the completion of HDR. It has been previously shown that HDR can be enhanced by the co-localization of donor templates and Cas9 via covalent linkage using the SNAP Tag in fusion with Cas9 that couples to oligodeoxynucleotides modified with Benzylguanine (Savic et al., 2018). Therefore, it will be interesting to explore whether the colocalization of donor templates and of HDR effectors at DSBs by using Cas9 fusion proteins together with the MS2 aptamer system has a synergistic effect on HDR enhancement.

In summary, here we employed a TLR reporter system for the assessment of DSB repair by HDR and NHEJ in HEK cells that we had used earlier to validate DNA Ligase IV as a target for NHEJ suppression (Chu et al., 2015). Although we expect that HDR enhancement at the TLR reporter system is

predictive for other genomic targets as previously shown and we indeed confirmed the effect of Cas9-CtIP at the B2M gene, the Cas9 fusion and the MS2 aptamer approach will require further validation using additional target genes and other cell types. For reporting NHEJ activity, the present configuration of the TLR construct has a restriction since it reports only a fraction of NHEJ repair events that restore the RFP reading frame. To increase the utility of TLR reporter constructs in future, modified constructs are required enabling to report NHEJ repair products that occur in all three reading frames. Furthermore, we did not discriminate between DSB repair choice in the G1 and S/G2 cell cycle phases. We anticipate that HDR enhancement occurs primarily in the S/G2 phases in which the HDR pathway usually operates. It has been shown that HDR can be at least partially reactivated in G1 by the combined suppression of 53BP1 and the expression of a degradation resistant Palb2 mutant together with the phosphomimetic CtIP (T847E) mutant (Orthwein et al., 2015). Thus, it will be interesting to determine the effect of Cas9 and MS2 fusions with HDR effectors specifically on DSB repair

in the G1 phase and whether the current protocol or further modifications will enable HDR mediated DSB repair in the G1 phase or even in resting cells.

In the current format of using plasmid-based expression vectors we expect that the CtIP co-localization approach can readily support applications of precise gene editing in human cell lines such as modeling or correcting disease-causing mutations by cotransfection with Cas9/sgRNA vectors. In combination with recombinant proteins its applications may be extended in future to primary human cells such as hematopoietic stem cells, muscle satellite or other stem cells exhibiting sufficient basal levels of HDR to assess its utility for the precise correction of mutations required for somatic gene therapy. Since DSB repair mechanism are conserved in evolution, we expect that the co-localization approach using CtIP or other HDR effectors can also be applied for HDR enhancement in other species.

MATERIALS AND METHODS

Plasmid Constructions

In the first set of experiments (**Figures 2–4**) we used Cas9 fusion vectors that were constructed based on the pX330 vector backbone (Addgene ID 42230), including a CBh promoter for protein expression. First, an oligonucleotide encoding the 16-mer flexible linker (SGSETPGTSESATPES) and multiple cloning sites (PacI and Sall) was in-frame cloned into either the N- or C-terminal end of Cas9. Then, cDNA of DNA repair factors (RAD51, RAD52, CtIP, CtIP_{mutant}, MRE11A, NBN, RPA1, and RAD50) were cloned between the PacI and Sall sites by standard PCR cloning methods. All primers for cloning are listed in **Table 1**. sgRNAs against 53BP1 was purchased as separated oligos, phosphorylated, annealed, and cloned into pX330 by standard cloning technique. pTLR-repair (Addgene 64322) was used as repair template vector for DSB repair assays in HEK^{TLR} reporter cells (Chu et al., 2015). In the second set of experiments (**Figures 5, 6**) we used pU6Rosa-CAG-Cas9 for expression of Rosa26 sgRNA and of Cas9 from the CAG promoter, constructed by ligation of oligonucleotides sgRosa-A/-B into the BbsI sites downstream of a human U6 promoter into plasmid pU6(BbsI). The U6-sgRosa cassette was recovered as AscI fragment and inserted into pCAG-Cas9-bpA-EF1-BFP, upstream of the CAG promoter driving Cas9 expression, followed by a BFP coding region under control of the human EF1 α promoter. For cloning of pCAG-Cas9-CtIP a SphI fragment was isolated from pX330-CBh-Cas9-CtIP and ligated between the SphI sites of pCAG-Cas9-bpA-EF1-BFP. Plasmid pTLR-donor was generated by whole plasmid PCR amplification using 5'-phosphorylated primers TLRtv-1 and TLRtv-2 and pTLR-repair as template, followed by of the PCR fragment. The modification of pTLR-repair removes the Start codon of the Venus coding region, eliminating background fluorescence upon transient transfection. Plasmid pU6Rosa-CAG-CtIP(T847E)-EF1-BFP was generated by cloning of a PCR product, amplified with the primers CtIP-for/-rev using as template pCW-GFP-CtIP-T847E (Addgene ID 71111), into the PacI/NotI sites of pU6Rosa-CAG-Pac-Not-EF1BFP. Plasmid pU6Rosa(MS2) was cloned by ligation

of oligonucleotides sgRosa-A/-B into the BbsI sites downstream of a human U6 promoter into plasmid sgRNA(MS2) (Addgene ID 64124). The U6Rosa(MS2) cassette was recovered by PCR (primers U6MS2-for/-rev) and cloned in between the AscI sites of pU6Rosa-CAG-CtIP(T847E)-EF1-BFP, replacing the U6Rosa cassette, to derive plasmid pU6Rosa(MS2)-CAG-CtIP(T847E)-EF1-BFP. The latter plasmid was digested with PacI upstream of the CtIP coding region for the in frame insertion of the MS2 coat protein coding region, amplified with primers MS2-for/-rev from plasmid MS2-P65-HSF1_GFP (Addgene ID 61423), to complete the pU6Rosa(MS2)-CAG-MS2-CtIP(T847E)-EF1-BFP vector. Alternatively a 875 bp synthetic gene fragment from pMS2-dimer (Thermo Scientific) encoding a single chain dimer of the MS2 protein (Peabody and Lim, 1996) was cloned into the PacI site of pU6Rosa(MS2)-CAG-CtIP(T847E)-EF1-BFP to obtain the pU6Rosa(MS2)-CAG-MS2^{di}-CtIP(T847E)-EF1-BFP vector. The SunLigand single chain antibody coding region was amplified using primers SunLigand-for/-rev from pHRdSV40-scFv-GCN4-sfGFP-VP64-GB1-NLS (Addgene ID 60904) and cloned into the PacI and AfeI site of pMS2-dimer to derive plasmid pSunLigand. Next, the SunLigand coding region was recovered by PCR (primers SunLigand-for, MS2-rev) and cloned into the PacI site of pU6Rosa-CAG-CtIP(T847E)-EF1-BFP in frame with CtIP to derive pU6Rosa-CAG-SunL-CtIP-EF1-BFP. Plasmid pCAG-Cas9-SunTag-bpA was cloned by ligation of a 1483 bp PCR product, amplified from pHRdSV40-dCas9-10xGCN4_v4-P2A-BFP (Addgene ID 60903) using primers Bsm-up and SunTag, into the BsmI/MluI sites of pCAG-Cas9v3a-bpA. pAAVS1-TLR6 was constructed by cloning of a fusion PCR product from a PCR fragment (using primers TLRvenus-1 and TLR6-1) and a PCR fragment (using primers TLR6-2 and TagRFP-2), using pAAVS1-TLR donor (Addgene ID 64215) (Chu et al., 2015) as template, into the backbone of plasmid pCAG-venusTarget+1P2A+3TagRFP (opened with PacI and MluI), resulting into pCAG-TLR6. pCAG-TLR6 was used for isolation of a AscI-AsiSI fragment that was ligated into pAAVS1-TLR donor, resulting into pAAVS1-TLR6, serving as AAVS1 targeting vector with the TLR6 reporter insert.

Generation of HEK Reporter Cells

The generation of HEK^{TLR} cells was previously described (Chu et al., 2015). Cells were cultured in DMEM (Gibco) supplemented with 10% FBS (Gibco), 2 mM sodium pyruvate (Gibco), 2 mM L-glutamine (Gibco), and 1 \times NEAA (Gibco). Cells were maintained in the exponential phase. Transfected cells with Blasticidin-resistant vector were selected with 10 μ g/ml of blasticidin for 4 days before analysis. The HEK^{TLR6} line was maintained in Dulbecco's Modified Eagle's medium with Glutamax (Gibco) supplied with 10% fetal bovine serum (Gibco) and generated by transfection of pAAVS1-TLR6 together with sgRNA plasmid against AAVS1 and a Cas9 plasmid (750 ng each) using Xtreme-gene transfection reagent (Roche). Antibiotic selection was performed using 0.4 μ g/ml Puromycin. Single clones were generated and genotyped using genomic DNA isolated using the Wizard genomic DNA purification kit (Promega #A1125). PCR was performed for knock-in of the TLR6 construct (puro 5') as well as the AAVS1 WT locus

specific as a control. The PCR reaction for TLR Knockin was performed using primers ST_puro_gt_fw and ST_puro_gt_rv. Phusion HF DNA polymerase (NEB # M0530 L) and 200 ng genomic DNA using following conditions; initial denaturation at 98°C for 3 min, followed by 35 cycles of 98°C 30 s, 58°C 30 s and 72°C for 90 s. final extension was done at 72°C for 5 min. AAVS1 WT PCR was performed using primers hAAVS1-For and hAAVS1-Rev by amplifying at 98°C for 5 min, followed by 40 cycles of 98°C 30 s, 60°C 30 s, 72°C for 45 s and final extension at 72°C for 5 min. After confirmation of TLR insertion by genotyping, selected clones were tested for activity of the TLR allele by FACS-based assay

and a single clone was chosen for further assays. All cell lines were confirmed for the absence of mycoplasma using the PCR assay of Uphoff and Drexler (2002).

Targeting of the 53BP1 Gene in HEK^{TLR} Cells

sgRNAs against Exon1 and 2 of the 53BP1 gene were designed using the CrispRGold program¹ and cloned into the pX330 vector using the oligonucleotides sgRNA_53BP1_exon1.1/-1.2 and _exon2.1/-2.2. Plasmid px330-sgRNA_exon1.1 and

¹<https://crisprgold.mdc-berlin.de/>

TABLE 1 | Oligodeoxynucleotides.

Name	Sequence (5'-3')
16mer Linker	AGCGGCAGCGAGACCCCCGGCACCGAGAGCGCCACCCCCGAGAGC
PacI-hCtIP F	ATTTAATTAAAAACATCTCGGGAAGCAGCTGTGGAAGCCC
Sall-hCtIP R	GGC GTCGAC CTATGTCTTCTGCTCCTTGCCCTTTTGAG
PacI-hRAD51 F	AT TTAATTTAA GCAATGCAGATGCAGCTTGAAGCAATGC
Sall-hRAD51 R	GGC GTCGAC TCAGGAAGACAGGGAGAGTCGTAGATTTTGC
PacI-hMRE11A F	AT TTAATTTAA AGTACTGCAGATGCAGCTTGATGATGAAA
Sall-hMRE11A R	GGC GTCGAC TTATCTTCTATTTCTCTTAAAGAACTAGTG
PacI-hNBN F	AT TTAATTTAA CAGAATGGCTTTTCCCGAACTTTGAAGTCGG
Sall-hNBN R	GGC GTCGAC TTATCTTCTCCTTTTAAATAAGGATTGTATCT
PacI-hRAD50 F	AT TTAATTTAA TCCCGGATCGAAAAGATGAGCATTCTGG
Sall-hRAD50 R	GGC GTCGAC TTAATGAACATTGAATCCAGGGAGCT
PacI-hRAD52 F	ATTTAATTAA TCTGGGACTGAGGAAGCAATCTTGGA
Sall-hRAD52 R	GGC GTCGAC TTAAGATGGATCATATTTCTTTCTTCATGCTCCTGGC
sgRNA_53BP1_Exon1.1	AGATTCTCAGCCTGAAAGCC
sgRNA_53BP1_Exon1.2	GCCTGAAAGCCAGGTTCTAG
sgRNA_53BP1_Exon2.1	GCTGGAGAAGAACGAGGAGA
sgRNA_53BP1_Exon2.2	GAACGAGGAGACGGTAATAG
T7_h53BP1_F	GAGGAATGGTGACTTGGTAGGCATGATAGG
T7_h53BP1_R	TGACCTGACTGATGGAACCATG
TLRvenus-1	ATCCTTAATTAAGCCGCCACCATGGTGAGCAAGGGCGAGGAGCTG
TLR6-1	CCATCTTCTAGAAAGACTGGAGTCTGGACGTAGCCTTCGGGCATG
TLR6-2	GACTCCAGTCTTTCTAGAAGATGGATCTTCTTCAAGGACGACGGC
TagRFP-2	ATTTACGCGTGCATCGCTCAATTAAGTTTGTGCCCCAGTTTG
TLRtv-1	GTGAGCAAGGGCGAGGAGCTG
TLRtv2	GGTGGCGGCTTAATCAAGCTTATC
sgRosa-A	CACCGACTCCAGTCTTTCTAGAAGA
sgRosa-B	AAACTCTTCTAGAAAGACTGGAGTC
i53-for	TCCTTAATTAATAACGACTCACTATAGGGGCCACCATGTTGATTTCTGTGAAAACCTTAC
i53-rev	CATTACGCGTTCAACGAAGTCTCAACAGAGGATG
CtIP-for	GATCCTTAATTAATAACGACTCACTATAGGGGCCACCATGAACATCTCGGGAAGCAGCTG
CtIP-rev	TCATTTACGCGTCTATGTCTTCTGCTCCTTGCCCTTT
U6MS2-for	TGAGGGGCGCGCCGAGGGCCTATTTCCCATGATTCC
U6MS2-rev	GAATCCGGCGCGCCAAAAAAGCACCGACTCGGTGCCAC
MS2-for	GATCCTTAATTAAGCCACCATGGCTTCAAACCTTACT
MS2-rev	ATGATCGATCGAGCGGCCGCCACCTTCCTCTT
SunLigand-for	ATTAAGCCACCATGGGCCCCGACATCGTGATG
SunLigand-rev	CCTCCAGCGCTATCGCCAATTGGAGTATTTGTTG
Bsm-up	CCTGTTGAGCTGGAAAACGGC
SunTag	ATTTACGCGTTATCATCCACCTTGCCTTCTTCTTGGGTCCACCTC
ST_puro_gt_fw	TCCCTCTTCCGATGTTGAG
ST_puro_gt_rv	GCGGTCCGGATCGACG

-sgRNA_exon2.2 were co-transfected into HEK^{TLR} cells to generate TP53BP1 knockout subclones. Three days post-transfection, targeted HEK^{TLR} cells were sorted as single cells into 96-well plates, cultured and expanded for genotyping by PCR using primers T7_h53BP1 primers listed in Table 1. Deletion of the 53BP1 gene segment between the Cas9 target sites in the first and second exon was confirmed by PCR by sequencing of the PCR product. Clone 7 that showed homozygous deletions (Supplementary Figure 2A) was selected for further experiments.

DSB Repair Assays

HEK^{TLR} or HEK^{TLR/Δ53BP1} cells were plated into 6-well plates 1 day before transfection at a density of 1×10^5 cells/well. Transfection was carried out using FuGENE HD Reagent (Promega) according to the manufacturer's introduction. Briefly, DNA was diluted in OptiMEM medium (GIBCO), then FuGENE was added in the mixture with the ratio of 3:1 (FuGENE:DNA). The mixture was incubated at room temperature for 10 min and then dropped slowly into pre-plated cells. Next day, medium was changed to the stop reaction. We used the same molar ratio of plasmids for transfection if multiple plasmids were used. HEK^{TLR6} cells were seeded in 24-well plates (50,000 cells per well) 1 day before transfection. Cells were transfected with pU6Rosa-CAG-Cas9, pTLR-donor template and one or two plasmids for the expression of CtIP or Cas9 fusion proteins (375 ng each) using Xtreme gene transfection reagent according to manufacturer's recommendation. All samples were transfected in triplicate unless otherwise stated. FACS analysis for all the experiments was performed upon 72 h after transfection. For preparation of cells for FACS analysis the medium was aspirated from wells and each well was washed with PBS and treated with Trypsin for cell detachment. The cells were collected after adding media by centrifugation at 300 g for 4 min and finally resuspended in 500 μ l PBS. The cells were kept on ice and FACS analysis was performed immediately. Single cells were gated for BFP positive populations and the frequency of Venus (HDR) and RFP (NHEJ) positive cells was determined using a BD LSR Fortessa flow cytometer (BD Biosciences). The results from replicate wells of each sample were used to calculate the mean value and standard deviation (SD). Mean

values and SD were used to calculate *p*-values (*T*-test) using the GraphPad Prism 8 software (GraphPad Software Inc., San Diego, CA, United States).

Synthetic Oligodeoxynucleotides

Synthetic oligonucleotides used in this study are shown in Table 1.

AUTHOR CONTRIBUTIONS

N-TT, SB, KR, and RK conceived and designed the project. N-TT, JR, and SB acquired the data. N-TT, SB, and RK analyzed and interpreted the data. VTC, XL, and JR provided materials. N-TT, SB, KR, and RK wrote the manuscript.

FUNDING

This work was supported by the German Ministry of Education and Research within the VIP program (TAL-CUT 03V0261 to RK).

ACKNOWLEDGMENTS

We thank Hans Peter Rahn (MDC FACS core facility) for excellent service and support, Miguel Rodriguez de los Santos for cloning of the pTLR-donor. Plasmids pCW-GFP-CtIP-T847E (Addgene ID 71111), MS2-P65-HSF1_GFP (Addgene ID 61423), sgRNA(MS2) cloning backbone (Addgene ID 61424), pHRdSV40-dCas9-10xGCN4_v4-P2A-BFP (Addgene ID 60903), pHRdSV40-scFv-GCN4-sfGFP-VP64-GB1-NLS (Addgene ID 60904) were gifts from Daniel Durocher, Feng Zhang, and Ron Vale, obtained via Addgene.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00365/full#supplementary-material>

REFERENCES

- Barrangou, R., and Doudna, J. A. (2016). Applications of CRISPR technologies in research and beyond. *Nat. Biotechnol.* 34, 933–941. doi: 10.1038/nbt.3659
- Bunting, S. F., Callén, E., Wong, N., Chen, H.-T., Polato, F., Gunn, A., et al. (2010). 53BP1 inhibits homologous recombination in *Brca1*-deficient cells by blocking resection of DNA breaks. *Cell* 141, 243–254. doi: 10.1016/j.cell.2010.03.012
- Canny, M. D., Moatti, N., Wan, L. C. K., Fradet-Turcotte, A., Krasner, D., Mateos-Gomez, P. A., et al. (2018). Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR-Cas9 genome-editing efficiency. *Nat. Biotechnol.* 36, 95–102. doi: 10.1038/nbt.4021
- Charpentier, M., Khedher, A. H. Y., Menoret, S., Brion, A., Lamribet, K., Dardillac, E., et al. (2018). CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nat. Commun.* 9:1133. doi: 10.1038/s41467-018-03475-7
- Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., et al. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* 33, 543–548. doi: 10.1038/nbt.3198
- Daley, J. M., and Sung, P. (2014). 53BP1, BRCA1, and the choice between recombination and end joining at DNA double-strand breaks. *Mol. Cell. Biol.* 34, 1380–1388. doi: 10.1128/MCB.01639-13
- Escribano-Díaz, C., Orthwein, A., Fradet-Turcotte, A., Xing, M., Young, J. T. F., Tkáč, J., et al. (2013). A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. *Mol. Cell* 49, 872–883. doi: 10.1016/j.molcel.2013.01.001
- Fradet-Turcotte, A., Canny, M. D., Escribano-Díaz, C., Orthwein, A., Leung, C. C. Y., Huang, H., et al. (2013). 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. *Nature* 499, 50–54. doi: 10.1038/nature12318

- Gupta, A., Hunt, C. R., Chakraborty, S., Pandita, R. K., Yordy, J., Ramnarain, D. B., et al. (2014). Role of 53BP1 in the regulation of DNA double-strand break repair pathway choice. *Radiat. Res.* 181, 1–8. doi: 10.1667/RR13572.1
- Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G. F., and Chin, L. (2016). Post-translational regulation of Cas9 during G1 enhances homology-directed repair. *Cell Rep.* 14, 1555–1566. doi: 10.1016/j.celrep.2016.01.019
- Howden, S. E., McColl, B., Glaser, A., Vadolas, J., Petrou, S., Little, M. H., et al. (2016). A Cas9 variant for efficient generation of indel-free knockin or gene-corrected human pluripotent stem cells. *Stem Cell Rep.* 7, 508–517. doi: 10.1016/j.stemcr.2016.07.001
- Huertas, P., and Jackson, S. P. (2009). Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. *J. Biol. Chem.* 284, 9558–9565. doi: 10.1074/jbc.M808906200
- Komor, A. C., Badran, A. H., and Liu, D. R. (2016). CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* 168, 20–36. doi: 10.1016/j.cell.2016.10.044
- Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517, 583–588. doi: 10.1038/nature14136
- Lin, S., Staahl, B. T., Alla, R. K., and Doudna, J. A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* 3:e04766. doi: 10.7554/eLife.04766
- Liu, J., Ehmsen, K. T., Heyer, W.-D., and Morrical, S. W. (2011). Presynaptic filament dynamics in homologous recombination and DNA repair. *Crit. Rev. Biochem. Mol. Biol.* 46, 240–270. doi: 10.3109/10409238.2011.576007
- Maruyama, T., Dougan, S. K., Truttmann, M. C., Bilate, A. M., Ingram, J. R., and Ploegh, H. L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* 33, 538–542. doi: 10.1038/nbt.3190
- Mirman, Z., Lottersberger, F., Takai, H., Kibe, T., Gong, Y., Takai, K., et al. (2018). 53BP1-RIF1-shieldin counteracts DSB resection through CST- and Pol α -dependent fill-in. *Nature* 560, 112–116. doi: 10.1038/s41586-018-0324-7
- Orthwein, A., Noordermeer, S. M., Wilson, M. D., Landry, S., Enchev, R. I., Sherker, A., et al. (2015). A mechanism for the suppression of homologous recombination in G1 cells. *Nature* 528, 422–426. doi: 10.1038/nature16142
- Paulsen, B. S., Mandal, P. K., Frock, R. L., Boyraz, B., Yadav, R., Upadhyayula, S., et al. (2017). Ectopic expression of RAD52 and dn53BP1 improves homology-directed repair during CRISPR-Cas9 genome editing. *Nat. Biomed. Eng.* 1, 878–888. doi: 10.1038/s41551-017-0145-2
- Peabody, D. S., and Lim, F. (1996). Complementation of RNA binding site mutations in MS2 coat protein heterodimers. *Nucleic Acids Res.* 24, 2352–2359. doi: 10.1093/nar/24.12.2352
- Sartori, A. A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., et al. (2007). Human CtIP promotes DNA end resection. *Nature* 450, 509–514. doi: 10.1038/nature06337
- Savic, N., Ringnalda, F. C., Lindsay, H., Berk, C., Bargsten, K., Li, Y., et al. (2018). Covalent linkage of the DNA repair template to the CRISPR-Cas9 nuclease enhances homology-directed repair. *eLife* 7:e33761. doi: 10.7554/eLife.33761
- Shen, M. W., Arbab, M., Hsu, J. Y., Worstell, D., Culbertson, S. J., Krabbe, O., et al. (2018). Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* 563, 646–651. doi: 10.1038/s41586-018-0686-x
- Symington, L. S. (2016). Mechanism and regulation of DNA end resection in eukaryotes. *Crit. Rev. Biochem. Mol. Biol.* 51, 195–212. doi: 10.3109/10409238.2016.1172552
- Symington, L. S., and Gautier, J. (2011). Double-strand break end resection and repair pathway choice. *Annu. Rev. Genet.* 45, 247–271. doi: 10.1146/annurev-genet-110410-132435
- Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S., and Vale, R. D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159, 635–646. doi: 10.1016/j.cell.2014.09.039
- Uphoff, C. C., and Drexler, H. G. (2002). Comparative PCR analysis for detection of mycoplasma infections in continuous cell lines. *In Vitro Cell. Dev. Biol. Anim.* 38, 79–85.
- Yang, D., Scavuzzo, M. A., Chmielowiec, J., Sharp, R., Bajic, A., and Borowiak, M. (2016). Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with customizable endonucleases. *Sci. Rep.* 6:21264. doi: 10.1038/srep21264
- Zakharyevich, K., Ma, Y., Tang, S., Hwang, P. Y.-H., Boiteux, S., and Hunter, N. (2010). Temporally and biochemically distinct activities of Exo1 during meiosis: double-strand break resection and resolution of double Holliday junctions. *Mol. Cell* 40, 1001–1015. doi: 10.1016/j.molcel.2010.11.032
- Zimmermann, M., and de Lange, T. (2014). 53BP1: pro choice in DNA repair. *Trends Cell Biol.* 24, 108–117. doi: 10.1016/j.tcb.2013.09.003
- Zimmermann, M., Lottersberger, F., Buonomo, S. B., Sfeir, A., and de Lange, T. (2013). 53BP1 regulates DSB repair using Rif1 to control 5' end resection. *Science* 339, 700–704. doi: 10.1126/science.1231573

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Tran, Bashir, Li, Rossius, Chu, Rajewsky and Kühn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Efficient and Precise CRISPR/Cas9-Mediated MECP2 Modifications in Human-Induced Pluripotent Stem Cells

Thi Thanh Huong Le^{1†}, Ngoc Tung Tran^{2†}, Thi Mai Lan Dao¹, Dinh Dung Nguyen¹, Huy Duong Do¹, Thi Lien Ha¹, Ralf Kühn^{2,3}, Thanh Liem Nguyen¹, Klaus Rajewsky² and Van Trung Chu^{2,3*}

¹ Department of Gene Technology, Vinmec Research Institute of Stem Cell and Gene Technology, Hanoi, Vietnam, ² Immune Regulation and Cancer, Max-Delbrück-Center for Molecular Medicine, Berlin, Germany, ³ iPS Cell Based Disease Modeling, Berlin Institute of Health, Berlin, Germany

OPEN ACCESS

Edited by:

Kun Xu,
Northwest A&F University, China

Reviewed by:

Daniela Tropea,
Trinity College Dublin, Ireland
Wei Li,
University of Alabama at Birmingham,
United States
Robin Ketteler,
University College London,
United Kingdom

*Correspondence:

Van Trung Chu
vantrung.chu@mdc-berlin.de

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Genomic Assay Technology,
a section of the journal
Frontiers in Genetics

Received: 21 February 2019

Accepted: 17 June 2019

Published: 02 July 2019

Citation:

Le TTH, Tran NT, Dao TML,
Nguyen DD, Do HD, Ha LT, Kühn R,
Nguyen TL, Rajewsky K and Chu VT
(2019) Efficient and Precise CRISPR/
Cas9-Mediated MECP2 Modifications
in Human-Induced Pluripotent
Stem Cells.
Front. Genet. 10:625.
doi: 10.3389/fgene.2019.00625

Patients with Rett syndrome (RTT) have severe mental and physical disabilities. The majority of RTT patients carry a heterozygous mutation in methyl-CpG binding protein 2 (MECP2), an X-linked gene encoding an epigenetic factor crucial for normal nerve cell function. No curative therapy for RTT syndrome exists, and cellular mechanisms are incompletely understood. Here, we developed a CRISPR/Cas9-mediated system that targets and corrects the disease relevant regions of the MECP2 exon 4 coding sequence. We achieved homologous recombination (HR) efficiencies of 20% to 30% in human cell lines and iPSCs. Furthermore, we successfully introduced a MECP2^{R270X} mutation into the MECP2 gene in human induced pluripotent stem cells (iPSCs). Consequently, using CRISPR/Cas9, we were able to repair such mutations with high efficiency in human mutant iPSCs. In summary, we provide a new strategy for MECP2 gene targeting that can be potentially translated into gene therapy or for iPSCs-based disease modeling of RTT syndrome.

Keywords: MECP2 mutations, CRISPR/Cas9, RETT syndrome, homologous recombination, iPSCs

INTRODUCTION

Rett (RTT) syndrome is a genetic neurodevelopmental disorder found predominantly in female infants. Disease symptoms including cognitive disabilities, decreased coordination and mobility, repetitive movements, and slowed brain growth typically appear after 6 to 18 months of age (Hagberg et al., 1983; Hagberg, 1985; Trevathan and Naidu, 1988). The RTT syndrome is caused by dominant negative mutations in the X-linked transcription factor methyl CpG-binding protein 2 (MECP2). Most mutations occur within one of two functional domains of MECP2: the highly conserved methyl binding domain (MDB) or the transcription repressing domain (TRD) (Rasmussen et al., 1975; Amir et al., 1999). Gene therapy using an adeno-associated viral vector to deliver a functional sequence of the MECP2 gene *in vivo* has been intensively investigated in mouse models (Sinnott and Gray, 2017). However, this approach induces toxicity and side effects due to the supraphysiological expression of exogenous MECP2 (Sinnott and Gray, 2017). The study by Guy et al. using conditional MECP2 alleles has shown that by restoring the physiological expression level of MECP2, the symptoms of RTT syndrome could be reversed in affected adult mice (Guy et al., 2007; Clarke and Abdala Sheikh, 2018). This study suggests that CRISPR/Cas9-mediated correction of MECP2 mutant alleles is a potential gene therapy for RTT.

The type II CRISPR/Cas9 system is a RNA-guided nuclease providing adaptive immunity in *Streptococcus pyogenes*. In mammalian cells, Cas9 nuclease can be used for editing of genomic sequence by the induction of targeted DNA double-strand breaks (DSBs). The induced DSBs are mostly repaired by the non-homologous end-joining (NHEJ) pathway creating micro-deletions/or insertions (INDELs) or, to a lesser extent, by the homologous recombination (HR) pathway allowing precise genetic manipulations if a repair template is provided (Cong et al., 2013; Hsu et al., 2013; Mali et al., 2013; Chu et al., 2015). Many studies have shown that CRISPR/Cas9-mediated mutagenesis can lead to efficient HR in human iPSCs (Byrne et al., 2015; Zhang et al., 2017; Li et al., 2018).

To our knowledge, there is no previous attempt to correct endogenous MECP2 mutant alleles in human cells using the CRISPR/Cas9 system. Here, we provide a system to repair the mutation hotspots of the MECP2 exon 4. We achieved HR efficiency up to 20% to 30% in human cell lines and iPSCs. To show a proof of principle of our system, we precisely inserted the MECP2^{R270X} mutation into the MECP2 gene in wild-type iPSCs. Finally, using the CRISPR/Cas9 system, we subsequently repaired this mutation successfully in mutant iPSCs. Overall, we provide an efficient strategy for MECP2 correction that is crucial for disease modeling and can potentially be translated into gene therapy of the RTT syndrome.

MATERIALS AND METHODS

Cell Culture and Reagents

HEK293 cells were maintained in DMEM (Gibco) supplied with 10% FBS (Gibco). K562 cells were cultured in RPMI 1640 supplied with 10% FBS and 2 mM L-glutamine (Gibco). Human iPSCs (BCRT cell line) were cultured in E8 flex medium (Gibco) according to the manufacturer's manual. Cas9 protein was purchased from NEB (M0386S), synthetic sgRNA was purchased from Synthego, and ssODN was ordered from IDT. To generate RNP complex, the Cas9 protein and synthetic sgRNA were mixed at ratio 1:2 and incubated at 25°C for 10 min.

CRISPR/Cas9 and Donor Vectors Construction

sgRNAs were designed based on CrisprGold software (Chu et al., 2016). Forward and reverse oligos were mixed and phosphorylated individually. Then, annealed oligo duplexes were cloned into the BbsI sites of the CRISPR/Cas9-T2A reporter plasmid (Addgene, 64216). To generate the pMECP2-T2A-mCherry reporter donor vector, the 5' and 3' homology arms (HA) were amplified from genomic DNA using Hercules Phusion polymerase (Agilent). The 5' HA fragment was cloned into XhoI/EcoRI sites of pTV-T2A-mCherry; the 3' HA fragment was cloned into AsiSI/KpnI sites of the pTV-T2A-mCherry vector. To generate the pMECP2 donor vector, 5-kb MECP2 fragment was amplified from genomic DNA, and at the cleavage site, the silent mutations were added to generate a new PstI recognition site.

Transfection, Electroporation, and Flow Cytometry

Human HEK293 cells were plated into six-well plates at 1 day before transfection. On the day of transfection, cells were supplied with fresh complete medium, and the DNA was mixed with FuGENE® HD Reagent (Promega) in Opti-MEM (Invitrogen) according to the manufacturer's introduction. After 15 min of incubation at RT, the mixture was dropped slowly into the well. Next day, the medium was exchanged. The transfected cells were analyzed in different time points. For flow cytometry analysis, HEK293 cells were trypsinized and resuspended in PBS/1% bovine serum albumin (BSA) fluorescence-activated cell sorting (FACS) buffer and analyzed with a Fortessa machine (Becton Dickinson).

For electroporation, human K562 cells were harvested and counted, 2×10^5 cells were resuspended with sgRNA/Cas9 RNP and 100 pmol ssODN in 20 µl electroporation buffer P3 (Lonza) and transferred to a 16-strip cuvette and electroporated using a 4D Nucleofector X unit (Lonza). Then, cells were transferred into the pre-warmed complete medium.

For human BCRT-iPSCs, we used Lipofectamine® 3000 (Life Technologies) according to the manufacturer's manual. Briefly, 1 day before transfection, iPSCs were placed as small clumps in 500 µg/ml of vitronectin (Life Technologies) precoated plate filled with E8 flex completed medium supplemented with Rock inhibitor (10 µM) (Bio Cat). The next day, the medium was replaced without Rock inhibitor 6 to 8 hours before transfection. Diluted plasmids and lipofectamine were prepared as described in manual. In the case of multiple plasmids, pX330 and targeting vector were used for transfection; we used a 1:1 molar ratio. The medium was changed on the next day to stop the reaction. Two days post-transfection, transfected iPSCs were collected by incubation with Accutase (Merck) for 5 min and the cell pellet was then resuspended in a medium containing E8 Flex complete medium, Rock inhibitor (10 µM), RevitaCell Supplement (Life Technologies), and Gentamycin (Lonza) (FACS-PREP medium). GFP⁺ iPSCs were sorted by FACSria (Becton Dickinson) and placed into a Vitronectin-precoated plate filled with FACS-PREP medium. The next day, the medium was replaced by E8 flex completed medium. Cells were maintained in culture for several passages before analysis.

For introducing the MECP2^{R270X} mutation into the MECP2 gene, wild-type iPSCs were transfected with 2 µg of the plasmid expressing Cas9 and sgRNA-5 and 60 pmol of ssODN-R270X donor using Lipofectamine® 3000 (Life Technologies). Two days post-transfection, 10^3 GFP⁺ iPS cells were sorted and plated on a Vitronectin-coated well of a six-well plate in the FACS-PREP medium. The medium was changed the next day. Seven days after sorting, iPSC clones were picked and plated on a new Vitronectin-coated well of a six-well plate. iPSC clones were expanded for 2 weeks. To this end, iPSC clones were harvested for analyzing insertion efficiency by genotyping PCR, RFLP assay, and Sanger sequencing. To repair the MECP2^{R270X} mutation, homozygous mutant iPSC clone (clone 18) was transfected with the plasmid expressing Cas9 and sgRNA-3 together with the donor template plasmid containing silent mutations and a recognition site of PstI restriction enzyme as described above.

Genomic DNA Isolation, PCR, T7EI, and RFLP Assay

Reporter⁺ cells were cultured and harvested at different time points. Single-cell clones were sorted in 96-well plates. Genomic DNA was extracted using the QuickExtract DNA extraction kit (Epicentre) following the manufacturer's instruction. For T7EI assay, PCR was done using Herculase II Fusion DNA Polymerase (Agilent Technology) with PCR gene-specific primers (**Table 1**) using the following conditions: 98°C for 3 min; 39 cycles (95°C for 20 s, 60°C for 20 s, 72°C for 20 s), and 72°C for 3 min. PCR products were run on 2% agarose gels, purified, denatured, annealed, and treated with T7EI (NEB). Cleaved DNA fragments were separated on 2% agarose gels, and the DNA concentration of each band was quantified using the ImageJ software. For genotyping RFLP assay, PCR product was purified and digested with a PstI restriction enzyme for 1 h. The digestion was separated on 2% agarose gels.

DNA sequencing PCR products were directly sequenced by specific primers or cloned into the pSTBlue-1 Blunt vector (Novagen) following the manufacturer's protocol. Plasmid DNAs were isolated using the NucleoSpin Plasmid (Macherey-Nagel). Plasmids were sequenced using T7 forward primer (5'-TAATACGACTCACTATAGGG-3') by the Sanger method (LGCgenomics, Berlin, Germany).

qRT-PCR

Total RNA was extracted from wild-type, mutant, and repaired iPSC clones with RNAeasy Mini Kit (Qiagen) and was reverse-transcribed into cDNA with a SuperScript[™] III kit (Invitrogen). The expression of MECP2 was measured by real-time PCR using SYBR green PCR Master Mix (Thermo Scientific) and StepOnePlus[™] (Applied Biosystems). The relative expression level of MECP2 was normalized with GAPDH housekeeping gene.

Statistical Analysis

Statistical tests were performed using Prism 7.0 (GraphPad) using a paired two-tailed Student's t-test. ****P < 0.0001.

RESULTS

CRISPR/Cas9-Mediated Reporter Insertion in the MECP2 Locus

Our previous study identified the mutation spectrum of the MECP2 gene in Vietnamese patients with RTT syndrome. The recurrent mutations T158M, G269fs, R270X, and R306H are located in the MDB and TRD domains of MECP2 protein encoded by exon 4 of the *MECP2* gene (**Figure 1A**). These mutations are specific for the Vietnamese patients, listed in RettBASE (Le Thi Thanh et al., 2018). We first developed a system to quantitatively determine HR efficiency in the human *MECP2* locus by inserting in-frame the coding sequence of cleavage peptide (T2A) and an mCherry reporter in the last exon of the *MECP2* gene. As a result, correctly targeted cells will express the mCherry reporter (**Figure 1B**). We used the CRISPRGold tool (Chu et al., 2016) to design two gRNAs (sg1 and sg2) targeting sequences proximal to the MECP2's stop codon. T7EI assay indicated that both sgRNAs efficiently targeted the *MECP2* locus upon delivery into HEK293 cells along with Cas9. Sanger sequencing showed a broad range of INDELs at the targeted site (**Figure 1C**). To access HR efficiency, plasmids expressing sg1 and Cas9 were transfected into the HEK293 cells together with donor plasmid carrying the MECP2_T2A_mCherry repair template. Transfected cells were analyzed by flow cytometry at days 14 and 21 post-transfection. We detected about 23% of mCherry⁺ cells in HEK293 cells that received sg1, Cas9, and donor template, but only background signals in control groups transfected with sgRNA/Cas9 or donor template alone (**Figure 1D and E**). We confirmed the corrected integration of the reporter into the targeted *MECP2* locus by using an external forward primer annealing to a genomic sequence outside of the 5' HA and a specific T2A sequence reverse primer for PCR. The expected ~3.6-kb fragment was amplified only in HEK293 cells transfected with sg1, Cas9, and donor template (**Figure 1F**). These data indicate the correct configuration of the HR alleles in mcherry⁺ HEK293 cells. Single mCherry⁺ cells were sorted into individual wells of 96-well plates for genotyping PCR. We found that about 75% of these cells harbor heterozygous

TABLE 1 | Sequences of sgRNAs, ssODNs, and primers were used in this study.

Name	Forward (5'>>3')	Reverse (5'>>3')
sg1	CACCGAGAGTTAGCTGACTTTACA	AAACTGTAAAGTCAGCTAACTCTC
sg2	CACCGTTAGCTGACTTTACACGGAG	AAACCTCCGTGTAAAGTCAGCTAAC
sg3	CACCGCTTTTTTGGCCTCGGCGGCAG	AAACCTGCCGCGCAGGCGCAAAAAGC
sg4	CACCGGGTGGCAGCGCTGCCGCCG	AAACCGCGGCAGCGGCTGCCACCC
sg5	CACCGTCAGGCCATTCCCAAGAAAC	AAACGTTTCTTGGGAATGGCCTGAC
MECP2 T7 set 1	CCAGAGGAGGCTCACTGGAGAGCGAC	GCTTGTCTGGTCAGTAGTATCTGCAGC
MECP2 T7 set 2	ACCACATCCACCCAGGTCATGGTGATC	TGGTGATGGTGGTGGTGCTCCTTCTTG
MECP2 HR	TAGAATAGGTAGGGTGCTCTTCTCCACCGG	CGGGATTCTCCTCCACGTACCCGCATGTT
MECP2 RT-PCR	GGAAATCTGGCCGCTCTGCTGGGA	ATTAGGGTCCAGGGATGTGTGCGCT
Gapdh RT-PCR	ATGCATCTGCAACCACTGCTTAG	GGATGCAGGGATGATGTTCTGGGCAGC
ssODN donor	CATTCCCAAGAAACGGGGCCGAAAGCCGGGAGTGTGGTGGCAGCAGCTGCAGCAGAGGCCAAAAAGAAAGCCGTGAAGGAGTCTTC	TATCCGATCTGTG
ssODN R270X	GGTGATCAACGCCCGGCAGGAAGCGAAAAGCTGAGGCCGACCCCTCAGGCCATTCCCAAAAAGCGCGGCTGAAATCTAGAGAGTGTG	GTGGCAGCCGCTGCCGCCGAGGCCAAAAAGAAAGCCGTGAAGGAGTCTTCTAT

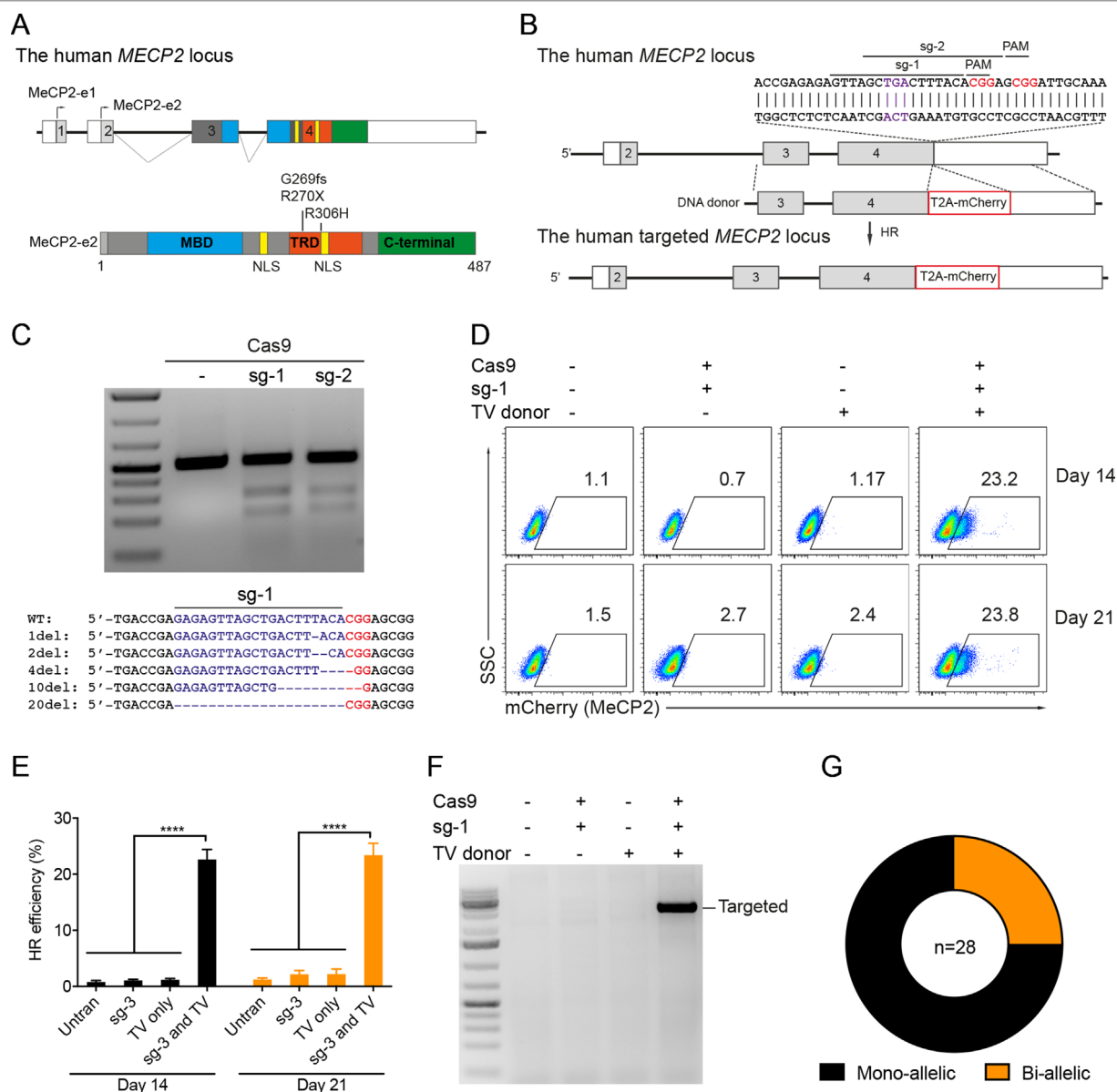


FIGURE 1 | CRISPR/Cas9-mediated reporter insertion in the *MECP2* locus. **(A)** Scheme of human *MECP2* isoforms (above) and *MECP2* protein with hotspot mutations in exon 4 (below). **(B)** Targeting strategy to insert T2A-mCherry into the *MECP2* locus. **(C)** T7E1 assay to show the editing activity of sg1 and sg2 (above) in the targeted *MECP2* sequence. Sequencing data (below) indicates the INDEL spectrum of sg1 in the targeted *MECP2* locus. **(D)** The percentage of mCherry⁺ cells were analyzed by flow cytometry at days 14 and 21 post-transfection. **(E)** The HR efficiency was summarized in the graph from three independent experiments, data show means \pm SD (**** $p < 0.0001$). **(F)** Corrected integration PCR. **(G)** Mono-allelic and bi-allelic knock-in analysis in single mCherry⁺ cell clones. The data represent at least two independent experiments.

integrations of the T2A-mCherry sequence into *MECP2*, whereas homozygous integrations represent about 25% (**Figure 1G**). Taken together, using CRISPR/Cas9 and reporter systems, we achieved efficient HR in the human *MECP2* locus.

CRISPR/Cas9-Mediated Precise Modification in *MECP2* Locus

Next, we developed the CRISPR/Cas9-mediated system to correct the recurrent mutations in the *MECP2* exon 4: G269fs, R270X,

and R306H. We designed two sgRNAs close to these mutations (sg3 and sg4) and a repair template containing HAs of 2.5 kb. To facilitate quantification of the HR efficiency, a PstI restriction site was created by introducing silent mutations into the repair template (**Figure 2A**). High editing activities of sg3 and sg4 were validated in HEK293 cells by T7E1 assays (**Figure 2B**, top panel). Sequencing data of sg3-targeted cells showed a broad range of INDELs (**Figure 2B**, bottom panel). Next, plasmids, carrying Cas9/sg3 and donor template were transfected into HEK293T cells. Thirty days post-transfection, genomic DNA of transfected

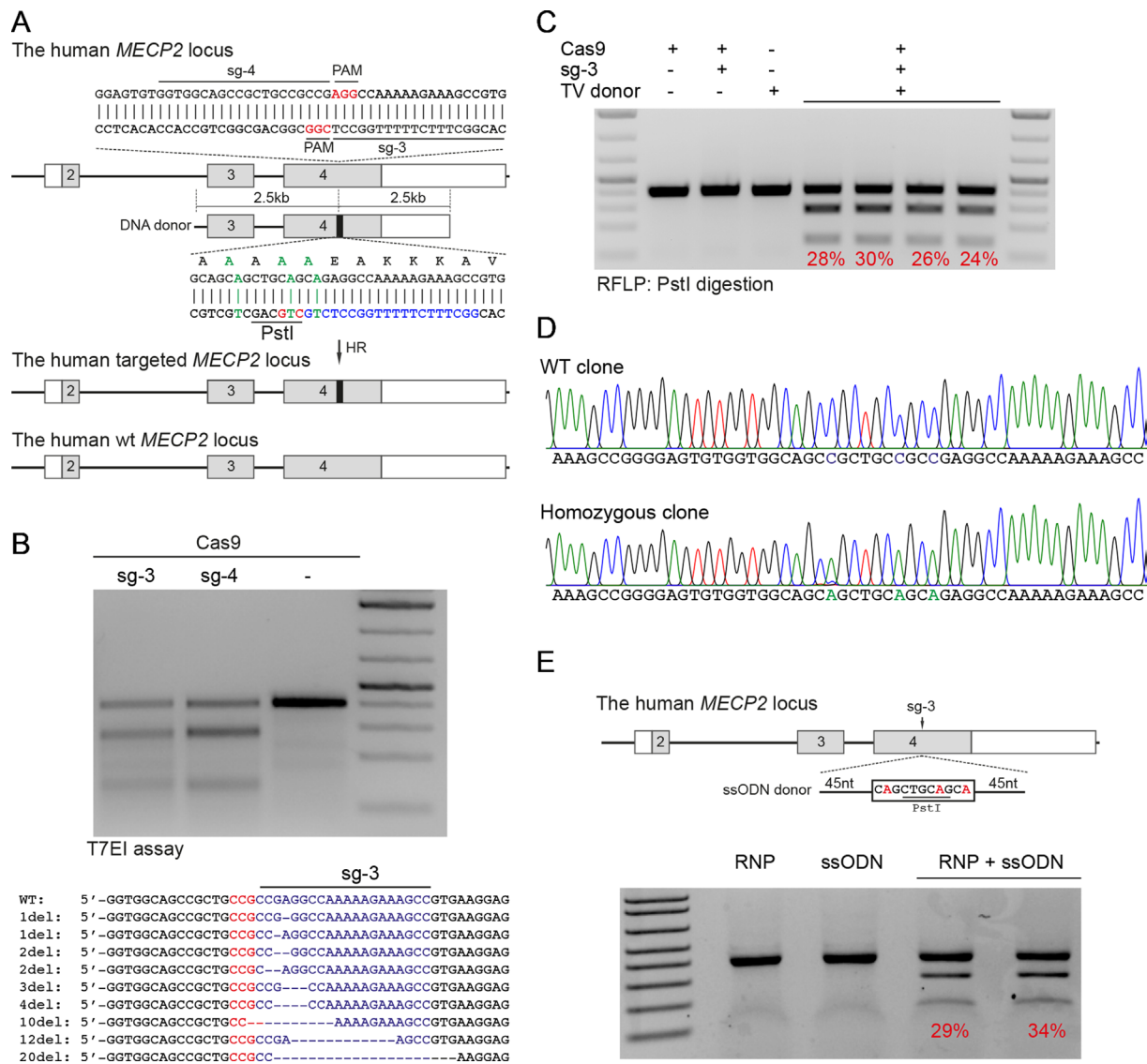


FIGURE 2 | CRISPR/Cas9-mediated efficient precise modification in the *MECP2* locus. **(A)** A strategy to knock-in silent mutations in the *MECP2* exon 4. **(B)** T7EI assay showed the cutting efficiencies of two sgRNAs (above) and sequencing data (sg3) showed the INDELs of the targeted *MECP2* sequence. **(C)** PstI-mediated restriction fragment length polymorphism (RFLP) assay showed the HR efficiency in HEK cells after 30 days of transfection, numbers indicate HR efficiency. **(D)** Sanger sequencing data of single-cell clone: wild-type (WT) and homozygous targeted clone (below). **(E)** A strategy to knock-in silent mutations in K562 cells using CRISPR/Cas9 RNP and ssODN as a repair template (above) and the RFLP assay to access the knock-in efficiency. Numbers in red represent for the band quantification-based knock-in efficiency. The data represent at least two independent experiments.

cells was isolated, and the targeted region was amplified by PCR. PstI-mediated restriction fragment length polymorphism (RFLP) showed that PstI-cleaved bands were only detected in cells co-transfected with sg3, Cas9, and repair template. Band quantification showed HR efficiency of 20% to 30% in the *MECP2* locus (Figure 2C). Sequencing data of the targeted homozygous clone confirmed that the precise modifications were correctly inserted into the *MECP2* gene (Figure 2D).

It is known that HR efficiency is determined by many factors. The type of donor template is considered as the most important one (Song and Stieger, 2017). Double-stranded DNA plasmid, PCR sequences, and ssDNA oligonucleotides (ssODN) are

often used as a donor template for precise insertion of large or small sequence changes at CRISPR/Cas9-induced DSBs. To test whether we can also use ssODN as donor template for *MECP2* precise modifications, we designed a 100-nucleotide ssODN with homology regions of 45 nucleotides each and silent replacements to create a PstI restriction site (Figure 2E). Next, we electroporated the sg3/Cas9 RNP complexes with ssODN into human leukemic K562 cells. Four days post-targeting, the targeted cells were harvested for analyzing HR efficiency. As shown in Figure 2E, we detected PstI-cleaved PCR products only in cells that received both RNP and ssODN. The HR efficiency ranged from 29% to 34% (Figure 2E).

CRISPR/Cas9-Mediated Reporter Insertion in Human iPSCs

Patient-derived iPSCs have been intensively studied for disease modeling, drug discovery, and potential somatic cell therapy. Thus, we next tested whether our system works efficiently in human iPSCs. We exploited the reporter system as described above to evaluate HDR efficiency. The sg1/Cas9 vector and donor template vector were co-transfected into human iPSCs. Two

days later, the transfected iPSCs were enriched and expanded (Figure 3A). The transfection efficiency in human iPSCs was about 25% to 28% (Figure 3B and C). Ten days after expansion, the percentage of mCherry⁺ iPSCs was analyzed by flow cytometry. As shown in Figure 3D and E, the percentage of mCherry⁺ iPSCs was about 20% when transfected with sg1/Cas9 and donor template vectors, whereas there were only background signals in control cells transfected with only sg1/Cas9 or donor

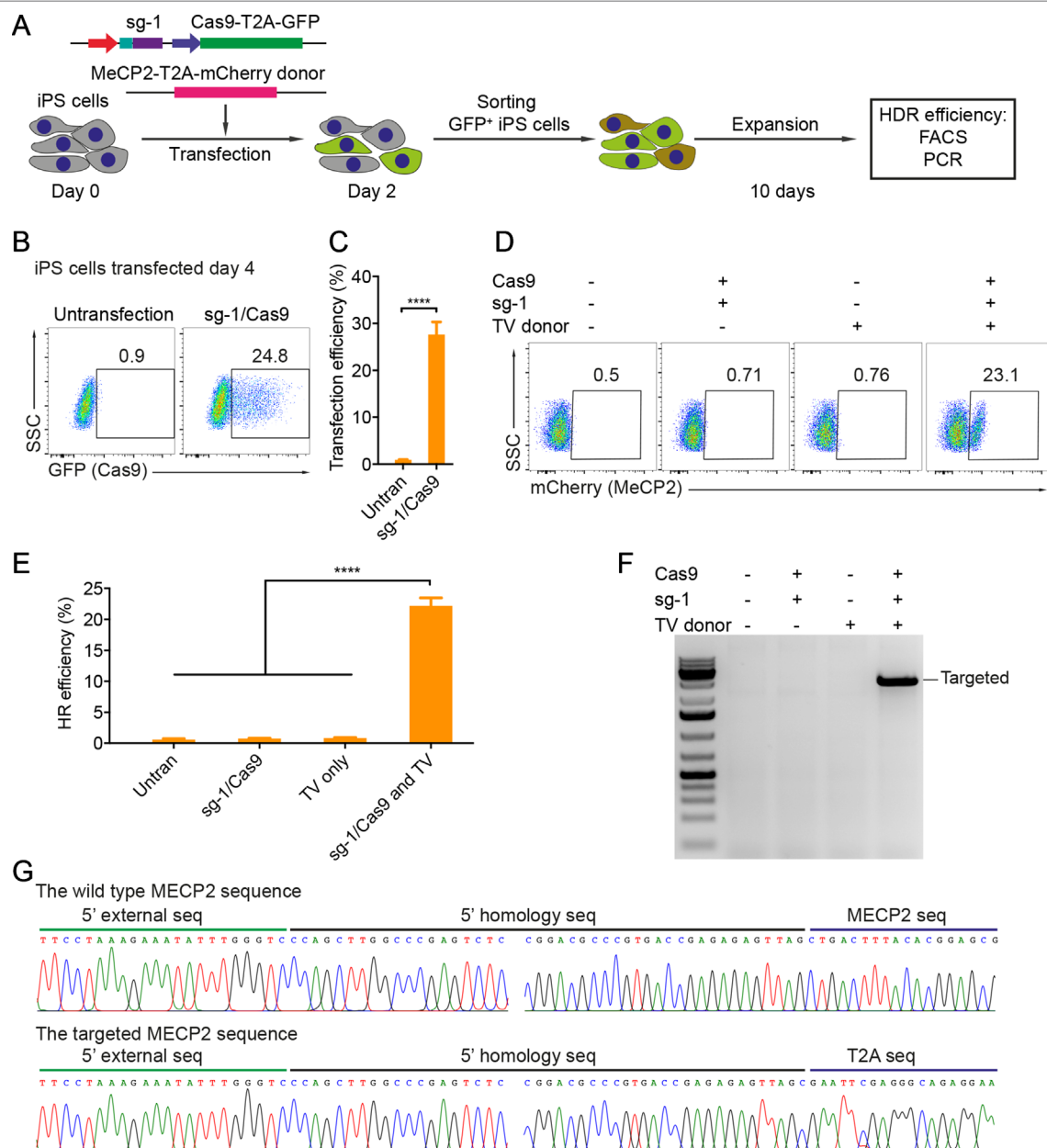


FIGURE 3 | CRISPR/Cas9-mediated reporter insertion in human iPSCs. **(A)** Experimental scheme of gene editing in human iPSCs. **(B)** Transfection efficiency in iPSCs using Lipofectamine 3000 was analyzed by FACS. **(C)** A bar graph indicated the transfection efficiency of human iPSCs, data show means \pm SD (**** p < 0.0001). **(D)** FACS analysis showed the knock-in efficiency in human iPSCs represented by the mCherry⁺ cells. **(E)** A bar graph showed knock-in efficiencies in human iPSCs, data show means \pm SD (**** p < 0.0001). **(F)** Corrected integration PCR and Sanger sequencing data showed the junctions of 5' HA. The data represent at least two independent experiments.

template vectors alone (**Figure 3D and E**). Correct integration PCR proved that the T2A-mCherry reporter was successfully inserted into the *MECP2* locus in human iPSCs (**Figure 3F**). Sequencing data confirmed that the targeted *MECP2* sequences were configured as planned (**Figure 3G**). Overall, we provide a new efficient system to precisely modify the human *MECP2* gene in human iPSCs.

CRISPR/Cas9-Based Modeling MECP2^{R270X} Mutation in Human iPSCs

To precisely introduce a MECP2^{R270X} mutation (c.808 C>T) into the *MECP2* gene in human iPSCs, we designed a new sgRNA (sg5) and ssODN (ssODN-R270X) with homology regions of 60 nucleotides, including the C>T mutation and an XbaI recognition site (**Figure 4A**). Next, we transfected human iPSCs with the sg5/Cas9 vectors and ssODN-R270X donor template. Two days post-transfection, the transfected iPSCs were sorted and single cell-derived clones were expanded (**Figure 4B**). To this end, cell clones were harvested for analyzing the insertion efficiency of the MECP2^{R270X} mutation. Among 22 cell clones, we detected two (9%) homozygous and five (23%) heterozygous clones (**Figure 4C and 4D**). Sequencing data confirmed that we succeeded to insert the R270X mutation into the *MECP2* gene in human iPSCs (**Figure 4E**). To show a proof of principle of our correction system, we used the strategy depicted in **Figure 2A**. The sg3/Cas9 and donor template vectors were transfected into MECP2^{R270X/R270X} iPS cells (clone 18, see **Figure 4C**). Two days later, single cell-derived clones were isolated and expanded for 2 weeks. As shown in **Figure 4F**, the PstI-mediated RFLP assay showed that we were able to correct the MECP2^{R270X} mutation in mutant iPSCs. To assess the expression of the *MECP2* gene, we isolated mRNA from the wild-type, mutant, and repaired iPSC clones. Quantitative RT-PCR data showed that we enabled to restore the expression level of the *MECP2* gene in the repaired iPSCs as comparable as the wild-type iPSCs (**Figure 4G**). Sequencing data of *MECP2* cDNA confirmed that the mutant *MECP2* allele was successfully repaired in the corrected iPSCs (**Figure 4H**). Thus, this system will open new avenues to iPSC-based disease modeling and therapeutic development of RTT syndrome.

DISCUSSION

In this study, we developed an efficient CRISPR/Cas9-mediated system that precisely modifies the human *MECP2* locus. Restoration of the physiological expression level of the *MECP2* gene is critical for the reversion of disease symptoms. Thus, the correction of mutations in the endogenous *MECP2* gene holds promise for gene therapy of RTT syndrome. Using preclinical mouse models of RTT, many laboratories have attempted to deliver intact *MECP2* cDNA *in vivo*, using the AAV vectors (Sinnott and Gray, 2017). Although AAV-mediated *MECP2* delivery leads to positive effects, such as increasing survival and improving weight, this system still causes multiple side effects and toxicity to animals due to the uncontrolled expression level of the *MECP2* transgene

(Collins et al., 2004; Luikenhuis et al., 2004; Gadalla et al., 2013). Thus, the expression level of *MECP2* is essential for restoring the function of defected neurons. Too much *MECP2* protein is as harmful to neural cells as too little. It has been shown that patients with the *MECP2* duplication syndrome have two copies of the *MECP2* gene on their single X-chromosome leading to mental disability and autistic-like behavior (Moretti and Zoghbi, 2006). Furthermore, a reduction of *MECP2* protein due to protein instability is also related to the milder neurological and psychiatric symptoms, such as anxiety and depression (Ramocki et al., 2009; Chao and Zoghbi, 2012). CRISPR/Cas9-mediated mutation correction restores the endogenous expression of the *MECP2* gene that might fully recover RTT symptoms. This system can be potentially applied as gene therapy for RTT patients *in vivo*. However, many challenges remain. For example, the mainly affected cells are neurons that are non-dividing. It is known that the HR pathway occurs in the S-G2 phase of the cell cycle. Thus, non-dividing or low proliferating cells, such as neurons, are not suitable for HR-mediated mutant correction. However, this issue can be potentially addressed by either suppressing the NHEJ pathway or activating HR-related factors (Chu et al., 2015; Charpentier et al., 2018). In addition, the safety of the system needs to be further investigated.

In addition, many studies have shown that *MECP2* is not only essential for the functions of neurons but also important for the functions of glial subtypes in the brain as well as of many cell types in the immune system, including microglia and macrophages (Maezawa and Jin, 2010; Derecki et al., 2012; O'Driscoll et al., 2013; Cronk et al., 2015; Jin et al., 2015; Cronk et al., 2016). Physiological restoration of normal *MECP2* functions in neurons and in all immune cells should be considered as an optimal therapy. It is supported that hematopoietic stem cell transplantation (HSCT) is beneficial in *Mecp2*-deficient mice and that *MECP2* plays important roles for immune cells both in the brain and in the periphery (Maezawa and Jin, 2010; Derecki et al., 2012; O'Driscoll et al., 2013; Cronk et al., 2015; Jin et al., 2015). Thus, correction of the *MECP2* mutations *ex vivo* with the CRISPR/Cas9 system followed by autologous HSCT is a promising future therapeutic option for RTT syndrome.

Our system also works efficiently in human iPSCs opening great avenues for disease modeling, drug screening, and somatic cell therapy. The generation of RTT patient-derived iPSCs (RTT-hiPSCs) has been reported by many groups; however, the natural random X-chromosome inactivation (XCI) status of RTT-hiPSCs is inconsistent. XCI results in cellular mosaicism where some cells express wild-type *MECP2* (isogenic), whereas other cells express mutant *MECP2*. Some studies showed that the maintenance of inactive X-chromosome of the founder cell allows RTT-hiPSCs to express either wild-type or mutant *MECP2* allele (Ananiev et al., 2011; Cheung et al., 2011; Pomp et al., 2011). In contrast, others reported that the inactive X-chromosome of the founder cells is reactivated during the reprogramming process (Kim et al., 2009; Marchetto et al., 2010). Importantly, RTT-hiPSCs could differentiate into neurons that exhibit disease phenotypes of RTT syndrome. Thus, these cells will be valuable for being rescued by CRISPR/Cas9-mediated *MECP2*

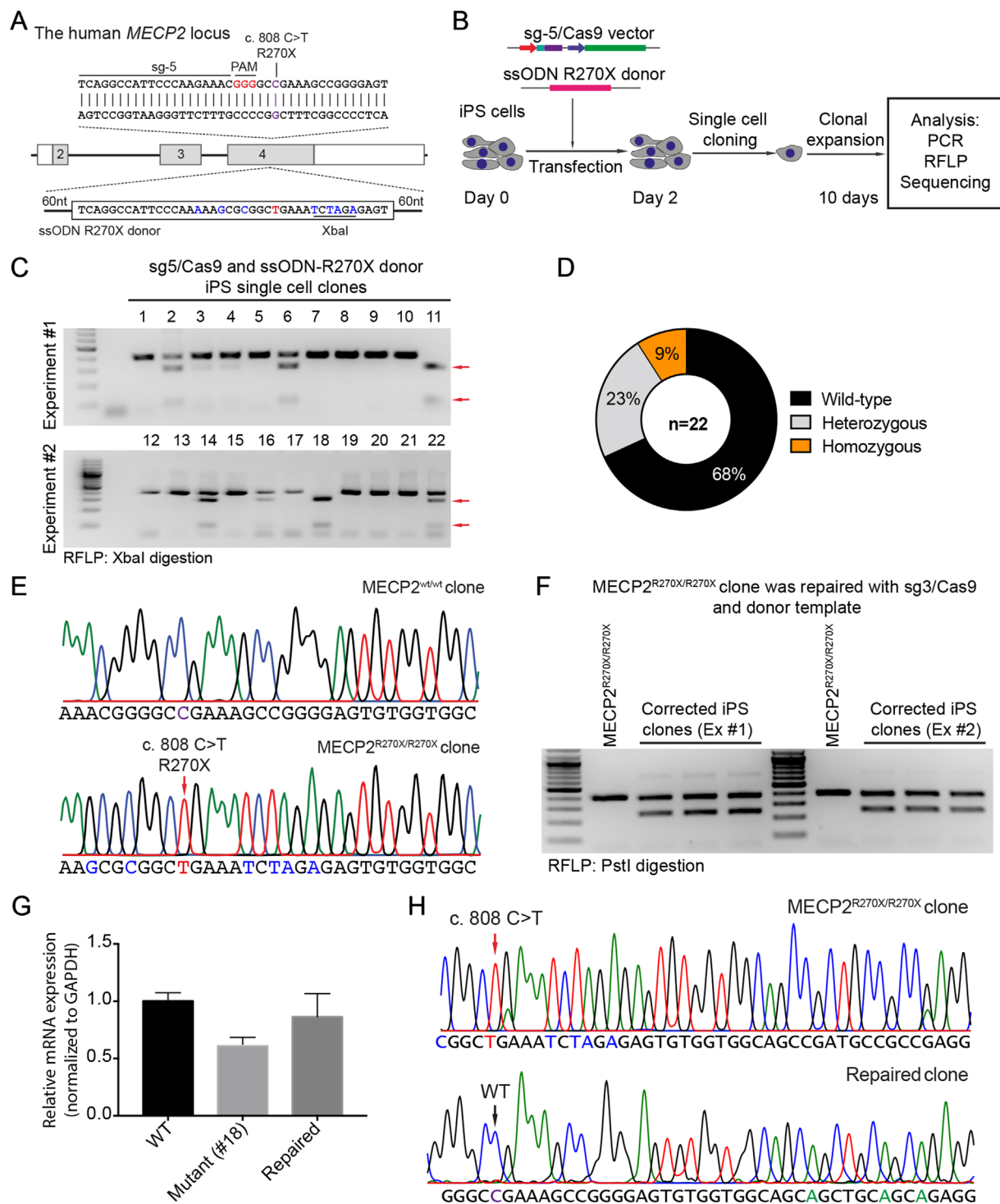


FIGURE 4 | CRISPR/Cas9-mediated insertion of the MECP2^{R270X} mutation in human iPSCs. **(A)** Strategy to insert a c.808 C>T mutation into the *MECP2* gene using CRISPR/Cas9 system. **(B)** Experimental scheme of the MECP2^{R270X} mutation insertion in human iPSCs. **(C)** XbaI-mediated RFLP assay shows heterozygous and homozygous MECP2^{R270X} mutant iPSC clones. **(D)** The donut chart summarizes the data shown in **(C)**. **(E)** Sanger sequencing data confirms insertion of the MECP2^{R270X} mutation in homozygous mutant iPSC clone. **(F)** PstI-mediated RFLP assay shows successful correction of the MECP2^{R270X} mutation in the repaired iPSC clones. **(G)** qRT-PCR data shows the expression level of MECP2 in the wild-type, homozygous mutant, and CRISPR/Cas9-repaired iPSC clones. **(H)** Sanger sequencing data of MECP2 cDNA confirms the MECP2^{R270X} correction in the repaired iPSCs. The data represent at least two independent experiments.

correction. Using CRISPR/Cas9 technology, we have achieved efficient precise MECP2 modifications in human iPSCs. Overall, our work provides an efficient system for repairing MECP2 mutations in RTT-hiPSCs.

DATA AVAILABILITY STATEMENT

All datasets generated/analyzed in this study are included in the manuscript or supplementary files.

AUTHOR CONTRIBUTIONS

TTHL and VTC designed the project. TTHL, NTT, TMLD, DDN, HDD, TLH, and VTC performed experiments and acquired the

data. NTT, VTC, RK, KR, and TLN analyzed and interpreted the data. VTC, NTT, RK, and KR wrote the article.

FUNDING

This work was supported by Vinmec Healthcare System (ISC.17.05 to TTHL and TLN).

ACKNOWLEDGMENTS

The authors thank H.P. Rahn for excellent FACS-related support. The authors thank Dr. Harald Stachelscheid from Berlin-Brandenburg Center for Regenerative Therapies for kindly providing human BCRT-iPSC line. The authors would also like to thank K. Petsch and J. Pempé for technical support.

REFERENCES

- Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., and Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* 23, 185–188. doi: 10.1038/13810
- Ananiev, G., Williams, E. C., Li, H., and Chang, Q. (2011). Isogenic pairs of wild type and mutant induced pluripotent stem cell (iPSC) lines from Rett syndrome patients as *in vitro* disease model. *PLoS One* 6, e25255. doi: 10.1371/journal.pone.0025255
- Byrne, S. M., Ortiz, L., Mali, P., Aach, J., and Church, G. M. (2015). Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. *Nucleic Acids Res.* 43, e21. doi: 10.1093/nar/gku1246
- Chao, H. T., and Zoghbi, H. Y. (2012). MECP2: only 100% will do. *Nat. Neurosci.* 15, 176–177. doi: 10.1038/nn.3027
- Charpentier, M., Khedher, A. H. Y., Menoret, S., Brion, A., Lamribet, K., Dardillac, E., et al. (2018). CtiP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nat. Commun.* 9, 1133. doi: 10.1038/s41467-018-03475-7
- Cheung, A. Y., Horvath, L. M., Grafodatskaya, D., Pasceri, P., Weksberg, R., Hotta, A., et al. (2011). Isolation of MECP2-null Rett syndrome patient hiPS cells and isogenic controls through X-chromosome inactivation. *Hum. Mol. Genet.* 20, 2103–2115. doi: 10.1093/hmg/ddr093
- Chu, V. T., Graf, R., Wirtz, T., Weber, T., Favret, J., Li, X., et al. (2016). Efficient CRISPR-mediated mutagenesis in primary immune cells using CrispRGold and a C57BL/6 Cas9 transgenic mouse line. *Proc. Natl. Acad. Sci. U. S. A.* 113, 12514–12519. doi: 10.1073/pnas.1613884113
- Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., et al. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* 33, 543–548. doi: 10.1038/nbt.3198
- Clarke, A. J., and Abdala Sheikh, A. P. (2018). A perspective on “cure” for Rett syndrome. *Orphanet J. Rare Dis.* 13, 44. doi: 10.1186/s13023-018-0786-6
- Collins, A. L., Levenson, J. M., Vilaythong, A. P., Richman, R., Armstrong, D. L., Noebels, J. L., et al. (2004). Mild overexpression of MECP2 causes a progressive neurological disorder in mice. *Hum. Mol. Genet.* 13, 2679–2689. doi: 10.1093/hmg/ddh282
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi: 10.1126/science.1231143
- Cronk, J. C., Derecki, N. C., Ji, E., Xu, Y., Lampano, A. E., Smirnov, I., et al. (2015). Methyl-CpG binding protein 2 regulates microglia and macrophage gene expression in response to inflammatory stimuli. *Immunity* 42, 679–691. doi: 10.1016/j.immuni.2015.03.013
- Cronk, J. C., Derecki, N. C., Litvak, V., and Kipnis, J. (2016). Unexpected cellular players in Rett syndrome pathology. *Neurobiol. Dis.* 92, 64–71. doi: 10.1016/j.nbd.2015.05.005
- Derecki, N. C., Cronk, J. C., Lu, Z., Xu, E., Abbott, S. B., Guyenet, P. G., et al. (2012). Wild-type microglia arrest pathology in a mouse model of Rett syndrome. *Nature* 484, 105–109. doi: 10.1038/nature10907
- Gadalla, K. K., Bailey, M. E., Spike, R. C., Ross, P. D., Woodard, K. T., Kalburgi, S. N., et al. (2013). Improved survival and reduced phenotypic severity following AAV9/MECP2 gene transfer to neonatal and juvenile male Mecp2 knockout mice. *Mol. Ther.* 21, 18–30. doi: 10.1038/mt.2012.200
- Guy, J., Gan, J., Selfridge, J., Cobb, S., and Bird, A. (2007). Reversal of neurological defects in a mouse model of Rett syndrome. *Science* 315, 1143–1147. doi: 10.1126/science.1138389
- Hagberg, B. (1985). Rett's syndrome: prevalence and impact on progressive severe mental retardation in girls. *Acta Paediatr. Scand.* 74, 405–408. doi: 10.1111/j.1651-2227.1985.tb10993.x
- Hagberg, B., Aicardi, J., Dias, K., and Ramos, O. (1983). A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases. *Ann. Neurol.* 14, 471–479. doi: 10.1002/ana.410140412
- Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31, 827–832. doi: 10.1038/nbt.2647
- Jin, L. W., Horiuchi, M., Wulff, H., Liu, X. B., Cortopassi, G. A., Erickson, J. D., et al. (2015). Dysregulation of glutamine transporter SNAT1 in Rett syndrome microglia: a mechanism for mitochondrial dysfunction and neurotoxicity. *J. Neurosci.* 35, 2516–2529. doi: 10.1523/JNEUROSCI.2778-14.2015
- Kim, D., Kim, C. H., Moon, J. I., Chung, Y. G., Chang, M. Y., Han, B. S., et al. (2009). Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. *Cell Stem Cell* 4, 472–476. doi: 10.1016/j.stem.2009.05.005
- Le Thi Thanh, H., Do Thi Diem, T., Duy, C. V., Thanh, H. L. T., Phuong, H. B. T., and Thanh, L. N. (2018). Spectrum of MECP2 mutations in Vietnamese patients with RETT syndrome. *BMC Med. Genet.* 19, 137. doi: 10.1186/s12881-018-0658-x
- Li, X. L., Li, G. H., Fu, J., Fu, Y. W., Zhang, L., Chen, W., et al. (2018). Highly efficient genome editing via CRISPR-Cas9 in human pluripotent stem cells is achieved by transient BCL-XL overexpression. *Nucleic Acids Res.* 46, 10195–10215. doi: 10.1093/nar/gky804
- Luikenhuis, S., Giacometti, E., Beard, C. F., and Jaenisch, R. (2004). Expression of MECP2 in postmitotic neurons rescues Rett syndrome in mice. *Proc. Natl. Acad. Sci. U. S. A.* 101, 6033–6038. doi: 10.1073/pnas.0401626101
- Maezawa, I., and Jin, L. W. (2010). Rett syndrome microglia damage dendrites and synapses by the elevated release of glutamate. *J. Neurosci.* 30, 5346–5356. doi: 10.1523/JNEUROSCI.5966-09.2010
- Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., et al. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826. doi: 10.1126/science.1232033
- Marchetto, M. C., Carroumeu, C., Acab, A., Yu, D., Yeo, G. W., Mu, Y., et al. (2010). A model for neural development and treatment of Rett syndrome using

- human induced pluripotent stem cells. *Cell* 143, 527–539. doi: 10.1016/j.cell.2010.10.016
- Moretti, P., and Zoghbi, H. Y. (2006). MeCP2 dysfunction in Rett syndrome and related disorders. *Curr. Opin. Genet. Dev.* 16, 276–281. doi: 10.1016/j.gde.2006.04.009
- O'Driscoll, C. M., Kaufmann, W. E., and Bressler, J. P. (2013). MeCP2 deficiency enhances glutamate release through NF-kappaB signaling in myeloid derived cells. *J. Neuroimmunol.* 265, 61–67. doi: 10.1016/j.jneuroim.2013.09.002
- Pomp, O., Dreesen, O., Leong, D. F., Meller-Pomp, O., Tan, T. T., Zhou, F., et al. (2011). Unexpected X chromosome skewing during culture and reprogramming of human somatic cells can be alleviated by exogenous telomerase. *Cell Stem Cell* 9, 156–165. doi: 10.1016/j.stem.2011.06.004
- Ramocki, M. B., Peters, S. U., Tavyev, Y. J., Zhang, F., Carvalho, C. M., Schaaf, C. P., et al. (2009). Autism and other neuropsychiatric symptoms are prevalent in individuals with MeCP2 duplication syndrome. *Ann. Neurol.* 66, 771–782. doi: 10.1002/ana.21715
- Rasmussen, P. J., Sarp, O. P., and Vilmann, H. (1975). Morphology of growing bone surfaces. *Scand. J. Dent. Res.* 83, 254–256. doi: 10.1111/j.1600-0722.1975.tb00434.x
- Sinnett, S. E., and Gray, S. J. (2017). Recent endeavors in MECP2 gene transfer for gene therapy of Rett syndrome. *Discov. Med.* 24, 153–159. <http://www.ncbi.nlm.nih.gov/pubmed/29272692>
- Song, F., and Stieger, K. (2017). Optimizing the DNA donor template for homology-directed repair of double-strand breaks. *Mol. Ther. Nucleic Acids* 7, 53–60. doi: 10.1016/j.omtn.2017.02.006
- Trevathan, E., and Naidu, S. (1988). The clinical recognition and differential diagnosis of Rett syndrome. *J. Child Neurol.* 3 Suppl, S6–16. doi: 10.1177/0883073888003001S03
- Zhang, J. P., Li, X. L., Li, G. H., Chen, W., Arakaki, C., Botimer, G. D., et al. (2017). Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biol.* 18, 35. doi: 10.1186/s13059-017-1164-8

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Le, Tran, Dao, Nguyen, Do, Ha, Kühn, Nguyen, Rajewsky and Chu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership