

Genome-wide identification of CRISPR/Cas9 off-targets in human genome

Cell Research (2014) 24:1009-1012. doi:10.1038/cr.2014.87; published online 1 July 2014

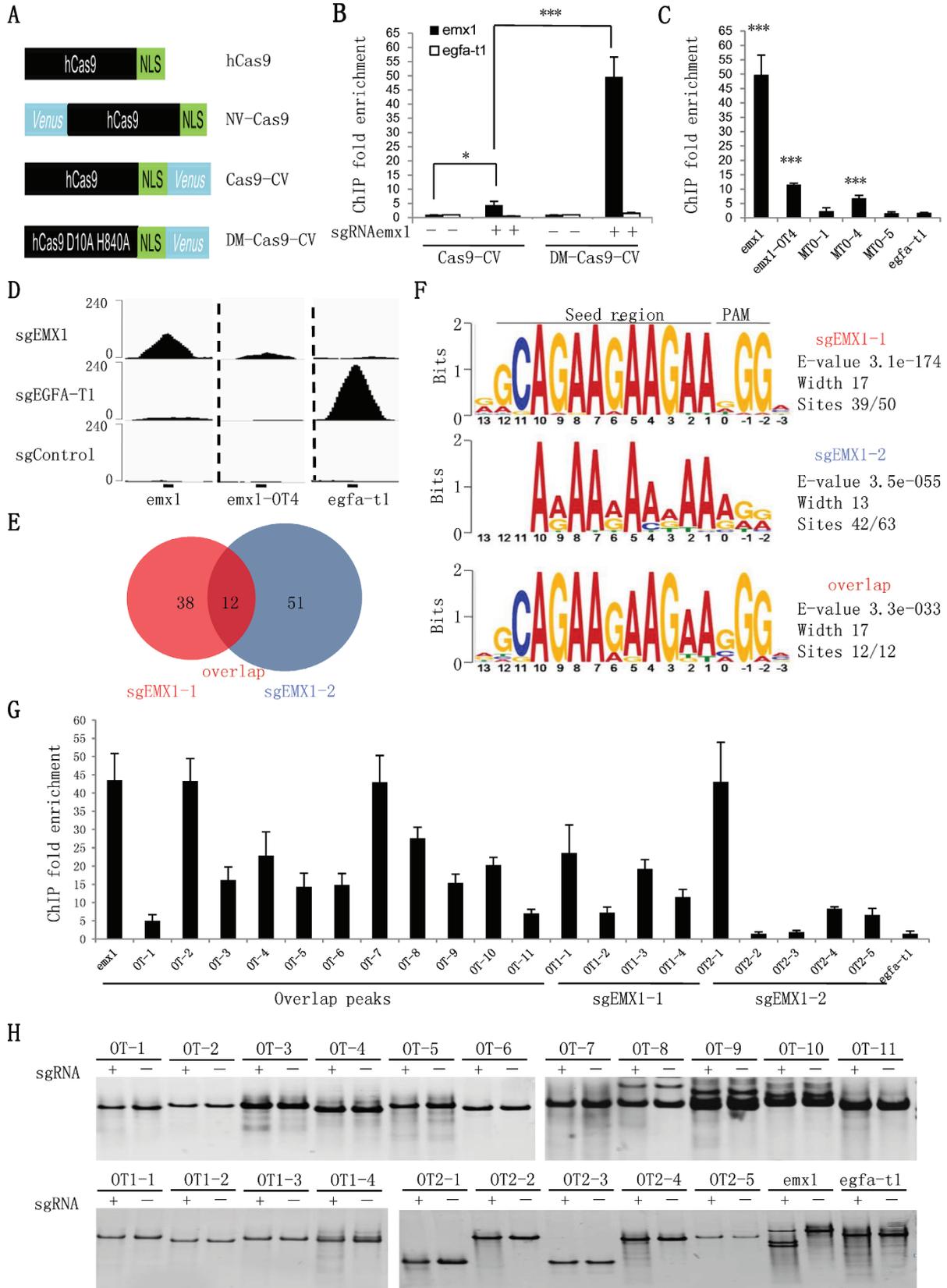
Dear Editor,

Genome editing techniques have been rapidly developing in recent decades [1]. Among them, site-specific cleavage of genomic loci in various organisms by homing endonucleases (HEases) [2], Zinc finger nucleases (ZFNs) [3], transcription activator-like effector nucleases (TALENs) [4], and most recently the CRISPR (clustered regularly interspersed short palindromic repeats)/Cas9 system [5], has been utilized widely not only in laboratories but also for translational studies. The central issue of genome editing is how to achieve specific and robust recognition of particular genomic sequences. In the case of HEases, ZFNs, and TALENs, this is achieved by specific intermolecular interactions between nucleotides and protein motifs, while for CRISPR/Cas9, the specificity is due to Watson-Crick base pairing between CRISPR RNA (crRNA) and its recognition sequences. The crRNA targets a 20-bp complementary target DNA sequence, which is flanked by a proto-spacer adjacent motif (PAM). Recent crystal structure studies and single-molecule DNA curtain experiments suggested that PAM site is essential for the initiation of Cas9 binding while the seed sequence corresponding to 3' end of the crRNA complementary recognition sequence which is also directly adjacent to PAM is critical for subsequent Cas9 binding, R-loop formation, and activation of nuclease activities in Cas9 [6-8]. While extensive efforts have been focused on optimizing the efficiency of targeting and cleavage by the CRISPR/Cas9 system in various organisms, relatively few studies investigated the mistargeting, so called off-targeting, activities [9-12]. In those studies, limited numbers of potential target DNA sequences with point or combined mismatches in comparison with the authentic targeting sites were tested for *in vitro* and *in vivo* cleavage activities. While these studies suggest the off-targeting activities of Cas9 on those examined sites are limited, for a comprehensive view an unbiased genome-wide analysis is required. Here, by a robust unbiased approach, we demonstrated that CRISPR/Cas9 had crRNA-specific

off-target binding activities in human genome. However, most of those binding off-targets could not be efficiently cleaved both *in vivo* and *in vitro*, suggesting that the off-target cleavage activity of CRISPR/Cas9 in human genome is very limited.

To unbiasedly determine the off-targets of CRISPR/Cas9 *in vivo*, we hypothesized that the Cas9/crRNA complex must first bind significantly to those off-targets, which could be revealed by chromatin immunoprecipitation and high throughput genome sequencing (ChIP-seq). We tagged *Venus* protein to either N- or C-terminal of humanized Cas9 (hCas9) protein (Figure 1A), which was subsequently co-transfected together with different single guide RNAs (sgRNAs) into human HEK-293T cells. Cas9-*Venus* C-terminal fusion protein (Cas9-CV) showed similar sgRNA-dependent *in vivo* cleavage activity on its genomic target as untagged wild type protein, while N-terminal *Venus*-Cas9 (NV-Cas9) fusion showed no activity (Supplementary information, Figure S1A). In addition, Cas9-CV also showed similar cleavage activity on a previously identified *emx1* off-target (*emx1*-OT4) [9, 11] but did not cleave the control locus (Supplementary information, Figure S1B-S1D). Mutations of both HNH nuclease and RuvC catalytic domains (DM-Cas9-CV) abolished the cleavage activity (Supplementary information, Figure S1A, lane5). We performed chromatin immunoprecipitation using high affinity nanobody for the *Venus* protein (GBP) [13]. Cas9-CV was significantly enriched in the *emx1* locus but not control *efga-t1* locus in an sgRNA-dependent manner, while DM-Cas9-CV showed a greatly enhanced binding in comparison with Cas9-CV (Figure 1B and Supplementary information, Figure S1E). Importantly, *emx1*-OT4 could also be significantly enriched by the ChIP approach (Figure 1C and Supplementary information, S1F-S1I). Therefore, we used DM-Cas9-CV in all subsequent ChIP experiments.

We performed ChIP-seq analysis in HEK-293T cells co-transfected which DM-Cas9-CV and no sgRNA or sgRNAs targeting either the *emx1* or *efga-t1* locus. Biological repeats were performed to reduce potential noises in the assay. In pooled ChIP-seq libraries, the



original targeting sites of *emx1*, *egfa-t1*, as well as the known off-target of *emx1*, *emx1-OT4*, showed significant specific sgRNA-dependent enrichment (Figure 1D). To achieve more stringent identification of the off-targets, during MACS peak calling, we set the threshold as $FDR < 0.5\%$. In ChIP-seq libraries generated from 293T cells without transfected sgRNA, no peak was identified, while in libraries generated from cells transfected with *egfa-t1* sgRNA, only the original target site was identified in biological repeats (Supplementary information, Table S1A). For *emx1* sgRNA, 50 and 63 peaks were identified in each biological repeat, and 12 overlapped peaks were finally obtained (Figure 1E). Interestingly, most of the 50 (39/50) and 63 (42/63) peaks contain conserved motifs which correlate well with PAM and its 5' 10-12 bp seed region, while all 12 peaks that appeared in both biological repeats contain such conserved motifs (Figure 1F and Supplementary information, Figure S1J-S1L and Table S1C-S1E). We further confirmed these identified peaks by quantitative PCR. Most of them showed significant sgRNA-dependent specific enrichment, with some showing comparable enrichment as the original *emx1* locus (Figure 1G and Supplementary information, Figure S1M). Finally, we checked whether these sites corresponding to peaks could indeed be cut by Cas9. Surprisingly, in both *in vitro* and *in vivo* cleavage assays, most of these binding off-targets could not be significantly cleaved while the *emx1* original site and its known off-target (*emx1-OT4*) were almost completely cleaved by Cas9/sgRNA*emx1* (Figure 1H and Supplementary information, Figure S1I, S1N). Only two binding off-targets, OT2-1 and OT2-4, reproducibly showed weak cleavage (Figure 1H and Supplementary information, Figure S1N). These results suggest that substrate binding could be uncoupled from the cleavage step in the CRISPR/Cas9 system.

One of the major concerns about genome editing is the potential off-target effect of editing enzymes which

may lead to unexpected genomic instabilities such as mutations and chromosomal translocations. By an unbiased genome-wide ChIP-seq approach, we analyzed binding off-targets of CRISPR/Cas9 in human genome. Surprisingly, while Cas9 could bind to various genomic sequences containing PAM and conserved seed sequences in an sgRNA-specific manner, its cleavage off-targets are very limited in comparison with other genome-editing enzymes, such as HEases, ZFNs, and TALENs. This might be largely due to additional involvement of the target sequence annealing step in activating the cleavage activities of CRISPR/Cas9 complex on its targets. On the other hand, the sgRNA-specific off-target binding activities may significantly affect other recently developed approaches which combine the nucleotide sequence binding specificity of CRISPR/Cas9 with other non-cleavage associated functions such as transcription regulation [14] and fluorescent labeling [15].

For the sgRNA targeting *emx1*, there are many more genomic loci which contain the PAM and conserved seed (10+3 base pairs) region in the human genome (Figure 1F and data not shown). It could be speculated that binding of Cas9/sgRNA*emx1* to those loci might be blocked by multiple factors including cell type- and/or development-specific local chromatin structure and modifications. Our preliminary results from HeLaS3 cells (Supplementary information, Figure S1P, Table S1B and S1F) have identified 19 potential binding off-targets for sgRNA *emx1*, and most of them did not overlap with the binding off-targets identified in HEK293T cells. Nevertheless, most of those HeLa cell-specific off-targets also contain similar conserved seed and PAM regions (Supplementary information, Figure S1Q-S1R). This suggests that off-targets might be cell type dependent and determined by various complicated factors in addition to primary DNA sequences. In addition, we hypothesize that different sgRNAs might have greatly variable levels of off-target binding activities which might correlate with

Figure 1 GFP nanobody-based ChIP-seq analysis to unbiasedly identify genome-wide off-targets of CRISPR/Cas9 in human genome. **(A)** Schematic view of Cas9 constructs. *Venus* was fused to wild type hCas9 or hCas9 double mutant (D10A and H840A). **(B)** The sgRNA-specific *in vivo* binding of Cas9-CV and DM-Cas9-CV to endogenous *emx1* and *egfa-t1* loci. 293T cells were transfected with sgRNA for the *emx1* locus or without sgRNA as control. The ChIP signals are divided by the no-sgRNA signals. **(C)** ChIP enrichment on *emx1*, previously identified *emx1* off-target (*emx1-OT4*), control *egfa-t1*, as well as other genomic loci containing sequences with 2-3 nucleotide mismatches in comparison to the original *emx1* target sequence, was examined in 293T cells transfected with sgRNA for *emx1*. **(D)** ChIP signals (normalized read counts) around target and off-target sites in ChIP-seq libraries. **(E)** ChIP-seq peaks identified in sgEMX1 libraries. **(F)** Consensus binding motif analysis of peaks in sgEMX1 libraries. Sequence motifs were identified within 101 base pairs around peak summits using MEME. **(G)** Confirmation of ChIP-seq-identified peaks in sgEMX1 libraries by quantitative PCR. All 12 overlapped peaks as well as randomly selected peaks from the rest of individual sgEMX1 libraries were checked. **(H)** *In vitro* cleavage assay of identified binding off-target sites. Error bars indicate the standard deviation (STDEV). Student's *t*-test was performed (* $P < 0.05$, *** $P < 0.0005$). All experiments were replicated at least twice. See also Supplementary information, Figure S1 and Table S1.

the kinetics of target DNA duplex disruption, formation of DNA-RNA heteroduplex, and R-loop expansion. The contribution of nucleotide sequence and composition of both the seed region and its 5' surrounding region needs further detailed studies. Our unbiased approach provides a valuable tool to further investigate the molecular mechanism of CRISPR/Cas9 and to optimize its *in vivo* applications.

Notes

During the preparation of this letter, two *Nat Biotechnol* papers appeared online (<http://dx.doi.org/10.1038/nbt.2889> and <http://dx.doi.org/10.1038/nbt.2916>) with similar conclusions.

Acknowledgments

We thank Drs L-L Du, Z-R Shen, and Mr D-P Ju (National Institute of Biological Sciences, Beijing (NIBS)) for reagents and members of YZ lab for discussion and support. This work was funded by the "Program for Excellent Talents" by Beijing municipal government (2013D008013000002) and "National Thousand Young Talents Program" of China to YZ. We thank the municipal government of Beijing and the Ministry of Science and Technology of China for funds allocated to NIBS.

Jinzhi Duan^{1,2,*}, Guangqing Lu^{1,2,*}, Zhan Xie²,
Mingliang Lou², Jiao Luo², Lei Guo², Yu Zhang^{1,2}

¹Graduate Program, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China; ²National Institute of Biological Sciences, Beijing 102206, China

*These two authors contributed equally to this work.

Correspondence: Yu Zhang

E-mail: zhangyu@nibs.ac.cn

References

- 1 Li M, Suzuki K, Kim NY, *et al. J Biol Chem* 2014; **289**:4594-4599.
- 2 Paques F, Duchateau P. *Curr Gene Ther* 2007; **7**:49-66.
- 3 Urnov FD, Rebar EJ, Holmes MC, *et al. Nat Rev Genet* 2010; **11**:636-646.
- 4 Joung JK, Sander JD. *Nat Rev Mol Cell Biol* 2013; **14**:49-55.
- 5 Mali P, Esvelt KM, Church GM. *Nat Methods* 2013; **10**:957-963.
- 6 Nishimasu H, Ran FA, Hsu PD, *et al. Cell* 2014; **156**:935-949.
- 7 Jinek M, Jiang F, Taylor DW, *et al. Science* 2014; **343**:1247997.
- 8 Sternberg SH, Redding S, Jinek M, *et al. Nature* 2014; **507**:62-67.
- 9 Fu Y, Foden JA, Khayter C, *et al. Nat Biotechnol* 2013; **31**:822-826.
- 10 Pattanayak V, Lin S, Guilinger JP, *et al. Nat Biotechnol* 2013; **31**:839-843.
- 11 Hsu PD, Scott DA, Weinstein JA, *et al. Nat Biotechnol* 2013; **31**:827-832.
- 12 Cho SW, Kim S, Kim Y, *et al. Genome Res* 2014; **24**:132-141.
- 13 Zhou ZX, Zhang MJ, Peng X, *et al. Genome Res* 2013; **23**:705-715.
- 14 Gilbert LA, Larson MH, Morsut L, *et al. Cell* 2013; **154**:442-451.
- 15 Chen B, Gilbert LA, Cimini BA, *et al. Cell* 2013; **155**:1479-1491.

(Supplementary information is linked to the online version of the paper on the *Cell Research* website.)