



# CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets

## Citation

Tsai, Shengdar Q., Nhu T. Nguyen, Jose Malagon-Lopez, Ved V. Topkar, Martin J. Aryee, and J. Keith Joung. 2017. "CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets." *Nature methods* 14 (6): 607-614. doi:10.1038/nmeth.4278. <http://dx.doi.org/10.1038/nmeth.4278>.

## Published Version

doi:10.1038/nmeth.4278

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:37068196>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)



Published in final edited form as:

*Nat Methods*. 2017 June ; 14(6): 607–614. doi:10.1038/nmeth.4278.

## CIRCLE-seq: a highly sensitive *in vitro* screen for genome-wide CRISPR-Cas9 nuclease off-targets

Shengdar Q. Tsai<sup>1,2,3,4,6</sup>, Nhu T. Nguyen<sup>1,2,3</sup>, Jose Malagon-Lopez<sup>1,2,3,4,5</sup>, Ved V. Topkar<sup>1,2,3</sup>, Martin J. Aryee<sup>1,2,3,4,5</sup>, and J. Keith Joung<sup>1,2,3,4</sup>

<sup>1</sup>Molecular Pathology Unit, Massachusetts General Hospital, Charlestown, MA 02129 USA

<sup>2</sup>Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA 02129 USA

<sup>3</sup>Center for Computational and Integrative Biology, Massachusetts General Hospital, Charlestown, MA 02129 USA

<sup>4</sup>Department of Pathology, Harvard Medical School, Boston, MA 02115 USA

<sup>5</sup>Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115 USA

### Abstract

Sensitive detection of off-target effects is important for translating CRISPR-Cas9 nucleases into human therapeutics. *In vitro* biochemical methods for finding off-targets offer potential advantages of greater reproducibility and scalability while avoiding limitations associated with strategies that require the culture and manipulation of living cells. Here we describe CIRCLE-seq (Circularization for *In vitro* Reporting of Cleavage Effects by sequencing), a highly sensitive, sequencing-efficient *in vitro* screening strategy that outperforms existing cell-based or biochemical approaches for identifying CRISPR-Cas9 genome-wide off-target mutations. In contrast to previously described *in vitro* methods, we show that CIRCLE-seq can be practiced using widely accessible next-generation sequencing technology and does not require reference genome sequence. Importantly, CIRCLE-seq can be used to identify off-target mutations associated with

---

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

Correspondence should be addressed to: [jjoung@mgh.harvard.edu](mailto:jjoung@mgh.harvard.edu) or [shengdar.tsai@stjude.org](mailto:shengdar.tsai@stjude.org).

<sup>6</sup>Present address: Department of Hematology, St. Jude Children's Research Hospital, Memphis, TN 38105 USA

### Accession Codes

High-throughput sequencing information associated with this study will be made available through NCBI SRA accession number XXXXXX (note: SRA number requested and pending).

### Software availability

Freely available, open-source software for analysis of CIRCLE-seq data can be obtained at: <https://github.com/tsailabSJ/circleseq>

### Data Availability Statement

Data supporting this work are available in supporting figure tables, supplementary information, or deposited in NCBI SRA.

### Author Contributions

S.Q.T., and J.K.J. conceived of and designed experiments. S.Q.T. and N.T.N. performed all experiments. S.Q.T., J.M.L., V.V.T., and M.J.A. wrote the CIRCLE-seq analysis pipeline and analyzed CIRCLE-seq data. S.Q.T. and J.K.J. wrote the manuscript with input from all authors.

### Competing Financial Interests Statement

J.K.J. is a consultant for Horizon Discovery. J.K.J. has financial interests in Beacon Genomics, Editas Medicine, Poseida Therapeutics, and Transposagen Biopharmaceuticals. J.K.J.'s interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies. S.Q.T., M.J.A., and J.K.J. are scientific co-founders of Beacon Genomics.

cell-type-specific SNPs, demonstrating the feasibility and importance of generating personalized specificity profiles. CIRCLE-seq provides the most accessible, rapid and comprehensive method for identifying genome-wide off-target mutations of CRISPR-Cas9 described to date.

## INTRODUCTION

CRISPR-Cas9 nucleases can be easily programmed to create targeted double-stranded breaks (**DSBs**)<sup>1–5</sup> and this simplicity has driven widespread adoption of this genome editing technology<sup>6–12</sup>. Cas9-induced DSBs can be repaired by cellular DNA repair pathways, resulting in targeted sequence alterations in the genomes of living cells and organisms<sup>13–17</sup>. Efficient cleavage by the commonly used *Streptococcus pyogenes* Cas9 (**SpCas9**) requires 17–20 nts of complementarity between a Cas9-associated guide RNA (**gRNA**) and a target site (protospacer)<sup>5,18–21</sup> adjacent to a 5'-NGG protospacer adjacent motif (**PAM**)<sup>5,22,23</sup>.

For clinical translation of CRISPR-Cas9, defining the frequencies and locations of unintended nuclease-induced off-target mutations is important<sup>10,24–27</sup>. Although cell-based methods for genome-wide off-target identification have been described<sup>28–32</sup>, these can miss off-target mutations that occur with frequencies below ~0.1% in nuclease-treated cell populations<sup>25</sup>. Furthermore, requirements for efficient cellular transfection limit the feasibility, scalability, and reproducibility of these methods, particularly with non-transformed cell types that are most therapeutically relevant.

By contrast, *in vitro* strategies for defining nuclease-induced off-target DSBs have potential advantages over cell-based approaches. Assays using purified components improve reproducibility, bypass the need for efficient cellular transduction or transfection, and avoid cell fitness effects. Importantly, concentrations of active nuclease can be raised to high levels *in vitro*, potentially enabling identification of sequences that may be rarely cleaved in cells. An *in vitro* method for characterizing Cas9 cleavage specificity using partially randomized DNA libraries biased towards specific target DNA sites has been previously described, but a limitation of this approach is that many sites identified do not actually occur in the human genome<sup>33</sup>.

To our knowledge, only a single *in vitro* genome-wide off-target identification method, Digenome-seq<sup>34</sup>, has been described in the literature. This approach relies on nuclease cleavage of genomic DNA, sequencing adapter ligation to all free ends (nuclease- and non-nuclease-induced), high-throughput sequencing, and bioinformatic identification of nuclease-cleaved sites exhibiting signature uniform mapping end positions. However, Digenome-seq analysis requires large number of reads (~400 million) and the high background of random genomic DNA reads makes it challenging to identify low-frequency nuclease-induced cleavage events.

Here we describe CIRCLE-seq (Circularization for In vitro Reporting of Cleavage Effects by sequencing), an *in vitro* screen for identifying genome-wide off-target cleavage sites of CRISPR-Cas9 that virtually eliminates the high background of random reads observed with Digenome-seq. This improvement enables not only substantially more sensitive off-target site detection, but also the ability to easily deploy CIRCLE-seq using widely accessible

benchtop next-generation sequencing platforms (e.g., MiSeq). For most Cas9-guide RNA complexes tested, CIRCLE-seq can identify all off-target sites in human genomic DNA found by GUIDE-seq and HTGTS (high-throughput gene translocation sequencing), two of the most sensitive previously described cell-based methods. Importantly, CIRCLE-seq also identifies many new *bona fide* off-target sites that occur in human cells. We also show that CIRCLE-seq can identify off-target sites in the absence of a reference genome, opening the door to off-target profiling in organisms lacking full genomic sequence or outbred populations with considerable sequence heterogeneity. Lastly, we demonstrate how CIRCLE-seq can be used to identify off-target cleavage sites that are enhanced or diminished by cell-typespecific SNPs, demonstrating the feasibility and importance of defining personalized off-target profiles.

## RESULTS

### Overview and optimization of the CIRCLE-seq method

We reasoned that reducing background genomic DNA reads that occur with Digenome-seq would substantially enhance detection of desired Cas9 nuclease-cleaved genomic DNA fragments. To accomplish this, we envisioned strategies to selectively sequence Cas9-cleaved genomic DNA. We designed restriction enzyme-independent strategies to generate and enzymatically select for the conversion of randomly sheared DNA into one of two different types of covalently closed DNA structures: attachment of stem-loops to linear DNA ends (Supplementary Fig. 1) or circularization of linear fragments (Fig. 1, Supplementary Figs. 2–3). Subsequent nuclease-induced cleavage of either population of covalently closed DNA molecules at on- and off-target sites would release free DNA ends required for subsequent next-generation sequencing adapter-ligation and sequencing. Comparison of these two approaches demonstrated that circularization was orders of magnitude more effective in enriching for Cas9-nuclease cleaved genomic DNA fragments (Supplementary Fig. 4a). Importantly, nearly all sites identified starting from linear DNA fragments with hairpin ends were also identified from circularized DNA and read counts between both methods were strongly correlated, suggesting that circularization does not bias the range or frequency of identified off-target sites (Supplementary Fig. 4b). We named the circularization method CIRCLE-seq and optimization and characterization of its technical reproducibility are described in the Supplementary Note (Supplementary Figs. 3 & 5). In contrast to other genome-wide nuclease off-target discovery methods, CIRCLE-seq uniquely enables sequencing of both sides of a single cleavage site in one DNA molecule using paired-end sequencing.

### CIRCLE-seq enables highly sensitive *in vitro* detection of CRISPR-Cas9 genome-wide off-target cleavage sites

To test the sensitivity of CIRCLE-seq, we used it to identify off-target cleavage sites of Cas9 directed by the single published gRNA profiled with the most recent and accurate version of Digenome-seq<sup>35</sup>. CIRCLE-seq evaluation of SpCas9 with this gRNA (targeted to the human *HBB* gene) on human K562 cell genomic DNA identified not only 26 of 29 off-target sites previously identified by Digenome-seq but also 156 new sites (Supplementary Fig. 6a). For the three sites found by Digenome-seq but not by CIRCLE-seq, we observed supporting

reads in the CIRCLE-seq data (Supplementary Fig. 6b), demonstrating that these sites were simply undersampled in these particular experiments. Of 156 new sites called by CIRCLE-seq but not Digenome-seq, we found 29 also showed evidence of cleavage in the original Digenome-seq data<sup>34</sup> (Supplementary Fig. 6c); Digenome-seq likely failed to call these sites due to stringent informatics scoring criteria required to contend with the abundant genome-wide background reads generated by this method. By contrast, we found that such background reads were rare with CIRCLE-seq (Supplementary Fig. 6d). Indeed, we estimate the enrichment factor of CIRCLE-seq for nuclease-cleaved sequence reads to random background reads is ~180,000-fold better than Digenome-seq based on examination of an on-target site with the two methods and adjusting for sequencing depth (Supplementary Fig. 6d). Start mapping positions of bidirectional CIRCLE-seq reads are consistent with the expected cleavage site of SpCas9 (3 bp before the PAM sequence) (Supplementary Fig. 6e), demonstrating mapping of cleavage positions with nucleotide-level precision. Taken together, these results demonstrate that CIRCLE-seq possesses higher signal-to-noise relative to Digenome-seq using approximately 100-fold fewer sequencing reads, likely accounting for its greater sensitivity for identifying genome-wide off-target sites.

### Direct comparisons of CIRCLE-seq with cell-based off-target determination methods

We next compared CIRCLE-seq with GUIDE-seq, a sensitive cell-based approach for genome-wide off-target site identification previously developed by our group<sup>30</sup>. Initially, we used CIRCLE-seq to assess SpCas9 with six different gRNAs targeted to non-repetitive sequences that had been previously characterized by GUIDE-seq across two different human cell lines. CIRCLE-seq identified variable numbers of genome-wide off-target cleavage sites for these six different gRNAs, ranging in number from as few as 21 to as many as 124 (Fig. 2a, Supplementary Data, and Supplementary Table 2), with up to 6 mismatches relative to the on-target site (Supplementary Fig. 7). For four of the six gRNAs, CIRCLE-seq detected all off-target sites identified by GUIDE-seq (Supplementary Fig. 8) and for two other gRNAs, it detected all but one off-target for each (Supplementary Fig. 8). Closer examination of the CIRCLE-seq data for these experiments revealed supporting reads for these two sites but not of a sufficient number to exceed the statistical threshold for detection. In addition, these two undetected sites had been at the lower boundary of detection in our GUIDE-seq experiments. Taken together, these findings again suggest that these two off-target sites would be detected with modestly increased CIRCLE-seq sequencing depth. Importantly, for all six gRNAs, CIRCLE-seq identified many more off-target sites than previously found by GUIDE-seq, including for a gRNA targeted to the *RNF2* gene for which we had previously been unable to identify off-target sites. We obtained similar results using CIRCLE-seq to profile four additional gRNAs targeted to repetitive sequences that we had previously characterized by GUIDE-seq (Supplementary Note; Supplementary Data, Supplementary Figs. 7 and 8).

We next compared CIRCLE-seq profiles of SpCas9 and two gRNAs targeted to *EMX1* and *VEGFA* that had previously been characterized by the cell-based HTGTS method. These CIRCLE-seq experiments identified 50 of the 53 off-target sites (94%) previously identified by HTGTS (Supplementary Fig. 9). Among the three HTGTS sites not detected by CIRCLE-seq, two were found when additional experimental replicates were performed and

the third had a low HTGTS score (Supplementary Table 3), suggesting that these sites would be detected with greater sequencing depth. Importantly, CIRCLE-seq also found a much greater number of off-target sites than previously identified by HTGTS (Fig. 2c).

### Off-target sites identified by CIRCLE-seq are mutated in human cells

An important question is whether novel off-target cleavage sites identified *in vitro* by CIRCLE-seq (and not by GUIDE-seq or HTGTS) are mutated in cells by Cas9/gRNA complexes. Many off-target sites detected by both CIRCLE-seq and GUIDE-seq have high numbers of mapping CIRCLE-seq sequencing read counts (Fig. 3a), strongly suggesting that GUIDE-seq primarily detects off-target sites that are among the most efficiently cleaved *in vitro*. By contrast, many off-target sites found only by CIRCLE-seq have lower CIRCLE-seq read counts (Fig. 3a), suggesting that these might be missed by GUIDE-seq (Supplementary Fig. 10). In this case, we would anticipate difficulty in validating these sites in cells using standard targeted amplicon sequencing because the error rate of next-generation sequencing places a floor for indel mutation detection of approximately 0.1%. Thus, to determine whether novel off-target sites detected only by CIRCLE-seq (but not in our original GUIDE-seq experiments) might be cleaved in human cells, we reasoned that we could instead perform high depth targeted amplicon sequencing using genomic DNA obtained from cell-based GUIDE-seq experiments and look for tag integration as evidence of off-target cleavage (Fig. 3b). This strategy sidesteps the problem of the indel error rate associated with deep sequencing because tag integration occurs with a negligible background frequency, thereby permitting detection of lower-frequency sites.

Using targeted tag integration sequencing, we examined 98 off-target sites found by CIRCLE-seq (but not by GUIDE-seq) for SpCas9 and gRNAs targeted to *EMX1* and *VEGFA* (site 1). We chose sites that exhibited a range of CIRCLE-seq read counts and numbers of mismatches relative to the on-target site (Supplementary Table 4). As positive controls, we also selected a smaller set of off-target sites with variable numbers of CIRCLE-seq read counts that were found by both CIRCLE-seq and GUIDE-seq (Supplementary Table 4). Targeted amplicon sequencing revealed detection of the dsODN at all of the control off-target sites, with frequencies that correlated well with GUIDE-seq read counts (Figs. 3c and 3d). Notably, we also detected dsODN integration at 24 of the 98 novel off-target sites identified only by CIRCLE-seq (Figs. 3c–e), with frequencies in the low range (0.003 – 0.2%) as expected. The locations of all tag integrations map to the expected cleavage positions 3 bps away from a PAM sequence, consistent with these sites representing *bona fide* off-target cleavage sites (Fig. 3f).

### Reference genome-independent off-target site discovery by CIRCLE-seq

Because each pair of CIRCLE-seq reads yields sequences from both sides of a single CRISPR-Cas9 nuclease cleavage site, we reasoned that our method could identify off-target sites even without a reference genome sequence. We developed a mapping-independent off-target site discovery algorithm that merges CIRCLE-seq paired-end reads and directly searches for off-target cleavage sites resembling the on-target site (**Online Methods**). Using this algorithm, we identified on average ~99.5% of CIRCLE-seq sites detected by our standard reference-based mapping algorithm and with more than 10 CIRCLE-seq reads



(Supplementary Fig. 11). Thus, CIRCLE-seq might potentially be used in a reference-independent fashion to identify off-target cleavage sites for organisms whose genome sequences are less well-characterized and/or show high genetic variability (e.g., non-inbred species in the wild).

### Association of CIRCLE-seq off-target sites with SNPs

With its higher throughput, an *in vitro* method such as CIRCLE-seq provides the opportunity to define patient-specific off-target profiles for any given Cas9/gRNA nuclease. A previously published study described an interesting example of a single SNP influencing off-target cleavage<sup>36</sup>. To more broadly test whether genetic differences can influence nuclease-induced off-target cleavage, we performed CIRCLE-seq experiments on human K562 genomic DNA with six gRNAs we had already assessed on human HEK293 and U2OS genomic DNAs (three gRNAs on HEK293s and three on U2OS). Although many off-target sites for these gRNAs showed well-correlated CIRCLE-seq read counts on DNA from both cell types tested, we also observed 55 sites that were preferentially cleaved in one cell type or the other (Fig. 4a). Further examination revealed that eight of these off-target cleavage sites harbored non-reference single-nucleotide polymorphisms (SNPs) that might account for these cell-type specific differences in cleavage efficiencies (Fig. 4b, Supplementary Table 5). Interestingly, these SNPs were located in regions of protospacer complementarity as well as the PAM (Fig. 4b).

Having identified additional examples where SNPs appear to influence cleavage efficiencies at off-target sites, we next sought to estimate how frequently SNPs might impact off-target cleavage efficiency. To do this, we examined the genotypes of 2504 individuals from the 1000 Genomes Project<sup>37</sup> at all 1247 off-target sites we detected by CIRCLE-seq for the six gRNAs (targeted to standard non-repetitive sequences). We found, on average, genetic variation in ~2.5% of these off-target sites (Fig. 4c). At a population level, we found that superpopulations contained genetic variation in an average of ~20% of these off-target sites (Fig. 4c). In addition, 50% of these off-target sites contained genetic variation for at least one individual sequenced in the 1000 Genomes Project (Fig. 4c). These frequencies are consistent with the expectation that, given existence of ~100 million validated human SNPs in the most recent version of dbSNP<sup>38</sup>, one might expect to find a SNP in ~69% of SpCas9 off-target sites in the human genome. As expected, the range of mismatches observed at the off-target sites we examined is increased when considering diverse individual genotypes from the 1000 genomes project (Fig. 4d). Interestingly, approximately 9% of variant off-target site haplotypes are more closely matched to the intended gRNA target sequence than the corresponding off-target site in the reference genome, suggesting that individuals with these particular genetic variations would have a higher risk of off-target cleavage at those sites (Fig. 4e). Taken together, these results highlight the importance of considering individual genetic variation and illustrate how CIRCLE-seq might be used in future studies to produce personalized genome-wide off-target profiles.

## DISCUSSION

Our results show that CIRCLE-seq method is highly sensitive and the most sequencing-efficient *in vitro* approach described to date for determining genome-wide off-target cleavage sites of CRISPR-Cas9 nucleases. CIRCLE-seq has a substantially reduced rate of observed background reads, enabling it to sensitively identify off-target sites using a small fraction (~1.7%) of the total sequencing reads used with existing *in vitro* methods. Requiring only 4–5 million reads, the method is accessible to most labs and might be amenable to automation and scaling, particularly if the relatively large amount of genomic DNA required for each CIRCLE-seq reaction (25 micrograms; Supplementary Protocol) can be reduced.

CIRCLE-seq might enable the production of larger datasets for training of more accurate predictive algorithms for Cas9 off-target determination. This will require that the method be cost-effectively automated and scaled to identify off-target cleavage sites and relative *in vitro* cleavage efficiencies for thousands of gRNAs. Furthermore, coupled with large-scale cell-based off-target profiles obtained in ENCODE-characterized cell lines<sup>39</sup> with methods such as GUIDE-seq, it may be possible to better understand the impact of chromatin accessibility and epigenetic modifications on the ability of nucleases to induce cellular DSBs. For example, an initial analysis we performed using publicly available DNase-seq datasets suggests that CIRCLE-seq sites were significantly more likely to be detected by GUIDE-seq or targeted tag sequencing if located in DNaseI hypersensitive regions (**data not shown**).

For both routine and therapeutic applications of genome editing, because CIRCLE-seq is more sensitive than cell-based genome-wide off-target detection methods, we envision that the method might be used as an initial screen to identify potential off-target sites that can then be verified with an orthogonal approach in nuclease-modified cells in culture or *in vivo*. In this study, we used targeted sequencing to search for GUIDE-seq dsODN tags to validate low-frequency sites (<0.1%) that would be challenging to identify by standard amplicon sequencing due to the indel error rate associated with next-generation sequencing (typically ~0.1%). However, this approach is limited to cells that can be transfected with the GUIDE-seq dsODN tag. Thus, an important area for future work will be the development of alternate methods to more sensitively measure off-target mutagenesis in cells below the error rate of current high-throughput sequencing technologies. Currently, the true false positive rate for CIRCLE-seq may be challenging to estimate due to limitations in technologies for orthogonal validation.

Our CIRCLE-seq results provide greater support for the concept that human genetic variation can affect off-target cleavage<sup>36</sup>. Our findings illustrate the importance of considering individual genotypes when evaluating off-target risk and suggest that safety assessments of nucleases might ultimately include patient-specific genome-wide activity profiling. Alternatively, CIRCLE-seq performed on a genomic DNA isolated from a large number of genetically diverse individuals may provide an effective strategy to define the vast majority of common and SNP-specific off-target effects for any given nuclease.

Finally, CIRCLE-seq could play an important role in further improvement of CRISPR-Cas nucleases. We recently described an engineered high-fidelity variant, SpCas9-HF1, which



for gRNAs targeted to non-repetitive sites routinely failed to show evidence of off-target effects as judged by GUIDE-seq. A key question for future studies is whether CIRCLE-seq can identify low frequency SpCas9-HF1 off-target mutations that might fall below the detection limit of GUIDE-seq. If such off-target sites were found, we envision that beyond providing an important potential assay for therapeutic applications, CIRCLE-seq would also play a major role in enabling the continued refinement of approaches to improve CRISPR-Cas9 genome-wide specificities.

## ONLINE METHODS

### Cell culture and transfection

Cell culture experiments were performed on human U2OS (gift from T. Cathomen), HEK293 (Thermo-Fisher), K562, and PGP1 fibroblast cells (gift from G. Church). U2OS and HEK293 cells were cultured in Advanced DMEM (Life Technologies) supplemented with 10% FBS, 2 mM GlutaMax (Life Technologies) and penicillin/streptomycin at 37°C with 5% CO<sub>2</sub>. K562 cells were cultured in RPMI 1640 (Life Technologies) supplemented with 10% FBS, 2 mM GlutaMax and penicillin/streptomycin at 37°C with 5% CO<sub>2</sub>. Human PGP1 fibroblasts were cultured in Eagle's DMEM (ATCC) with 10% FBS, 2 mM GlutaMax and penicillin/streptomycin at 37°C with 5% CO<sub>2</sub>. For CIRCLE-seq experiments, genomic DNA was isolated using Genra Puregene Tissue Kit (Qiagen) and quantified by Qubit (Thermo Fisher). For targeted tag-integration deep-sequencing experiments, U2OS cells (program DN-100), HEK293 cells (program CM-137), and K562 cells (program FF-120) were transfected in 20 µl Solution SE (Lonza) on a Lonza Nucleofector 4-D, according to the manufacturer's instructions. In U2OS cells, 500 ng of pCAG-Cas9 (pSQT817), 250 ng of gRNA encoding plasmids, and 100 pmol of GUIDE-seq end-protected dsODN were cotransfected. Genomic DNA for targeted tag integration sequencing was harvested approximately 72 hours post transfection using the Agencourt DNAdvanced Genomic DNA Isolation Kit (Beckman Coulter Genomics).

### *In vitro* transcription of gRNAs

Annealed oligonucleotides containing gRNA target sites were cloned into plasmid NW59 containing a T7 RNA polymerase promoter site. The gRNA expression plasmid was linearized with HindIII restriction enzyme (NEB) and purified with MinElute PCR Purification Kit (Qiagen). The linearized plasmid was used as DNA template for *in vitro* transcription of the gRNA using MEGAshortscript Kit, according to the manufacturer's instructions (Thermo-Fisher).

### CIRCLE-seq library preparation

For the experiments with gRNAs previously evaluated by GUIDE-seq, CIRCLE-seq experiments were performed on genomic DNA from the same cells in which they were evaluated by GUIDE-seq (either U2OS or HEK293 cells). Purified genomic DNA was sheared with a Covaris S200 instrument to an average length of 300 bp, end-repaired, A-tailed, and ligated to uracil-containing stem-loop adapter oSQT1288 5'-P-CGGTGGACCGATGATCUATCGGTCCACCG\*T-3', where \* indicates phosphorothioate linkage. Adapter-ligated DNA was treated with a mixture of Lambda Exonuclease (NEB) and *E. coli*

Exonuclease I (NEB), then with USER enzyme (NEB) and T4 polynucleotide kinase (NEB). DNA was circularized at 5 ng/ul concentration with T4 DNA ligase, and treated with Plasmid-Safe ATP-dependent DNase (Epicentre) to degrade remaining linear DNA molecules. *In vitro* cleavage reactions were performed in a 100 µl volume, with Cas9 nuclease buffer (NEB), 90 nM SpCas9 protein (NEB), 90 nM *in vitro* transcribed gRNA, and 250 ng of Plasmid-Safe-treated circularized DNA. Digested products were A-tailed, ligated with a hairpin adapter, treated with USER enzyme (NEB), and amplified by PCR using Kapa HiFi polymerase (Kapa Biosystems). Completed libraries were quantified by droplet digital PCR (Bio-Rad) and sequenced with 150 bp paired end reads on an Illumina MiSeq instrument. A detailed user protocol for CIRCLE-seq library construction is provided (Supplementary Protocol).

### Targeted deep-sequencing

U2OS cells were transfected with Cas9 and gRNA expression plasmids, in addition to the GUIDE-seq dsODN as described above. Off-targets sites identified by CIRCLE-seq were amplified from the isolated U2OS genomic DNA using Phusion Hot Start Flex DNA polymerase (New England Biolabs) with primers listed in Supplementary Table 6. Triplicates of PCR products were generated from each transfection condition with 100 ng of genomic DNA as the input for each PCR. PCR products were normalized in concentration, pooled into different libraries corresponding to different transfection conditions, and purified with Ampure XP magnetic beads (Agencourt). Illumina Tru-seq deep-sequencing libraries were constructed using 500 ng of each pooled samples (KAPA Biosystems), quantified by real-time PCR (KAPA Biosystems), and sequenced on an Illumina MiSeq instrument.

### CIRCLE-seq data analysis

Paired-end reads were merged and then mapped using *bwa*<sup>40</sup> *mem* with default parameters. The start mapping positions of reads that map in the expected orientation with mapping quality  $\geq 50$  were tabulated and genomic intervals that are enriched in nuclease-treated samples were identified. The interval and 20-bp of flanking reference sequence on either side was searched for potential nuclease-induced off-target sites with an edit distance of less than or equal to 6, allowing for gaps.

*samtools*<sup>41,42</sup> *mpileup* was used to non-reference genetic variation in identified off-target sites. Positions with average quality score greater than 20 were considered as possible variants and confirmed by visual inspection (Supplementary Table 3).

Reference-independent discovery of off-target cleavage sites was performed by reverse complementing the sequence of one read of a pair and concatenating it with the other. An interval of starting 20-bp on either side of the junction was directly searched for potential off-target cleavage sites with edit distance of  $\leq 6$  allowing for gaps and read counts corresponding to identified sites were tabulated.

### CIRCLE-seq open-source analysis software

To enable the broad use of CIRCLE-seq for genome-wide detection of nuclease off-target sites, we developed a freely available, open-source Python package *circleseq* for the analysis

of CIRCLE-seq experimental data. Provided with a simple sample manifest, the *circleseq* software performs full end-to-end analysis of CIRCLE-seq sequencing data with a single command, and returns tables of candidate off-target cleavage site positions, as well as visual alignments of off-target sequences. Source code and running instructions will be made freely available online (<https://github.com/tsailabSJ/circleseq>).

### Digenome-seq data analysis

Read counts of mapping positions in a narrow window ( $\pm 3$  bp) around cleavage sites identified by CIRCLE-seq were tabulated from original Digenome-seq sequencing alignments.

### Chromatin Accessibility analysis

To determine if sites identified by both CIRCLE-Seq and GUIDE-seq/targeted tag sequencing were associated with chromatin accessibility, we used DNase-seq data for closely related HEK293T and U2OS (GEO samples GSM1008573 and GSM2341641, respectively). We ran a Cochran-Mantel-Haenzel test for each cell stratified by the mismatch number (3:5), where the response is whether or not the CIRCLE-Seq site was also called by GUIDE-Seq and the predictor is the categorical variable for chromatin accessibility as inferred from DNaseI hypersensitivity ('open' or 'closed'). We considered all the sites found by CIRCLE-Seq at HEK293 and U2OS cells whose mismatch with the on-target site ranges from 3 to 5 (as we do not have enough sites for other class of mismatches).

### Statistics

An empirical read count distribution was used to determine statistical enrichment of CIRCLE-seq read counts.

For analysis of Digenome-seq data, significant evidence of cleavage at a 0.01 significance level was evaluated by fitting a negative binomial distribution, and statistically significant sites by this criteria were included in Fig. 2b.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

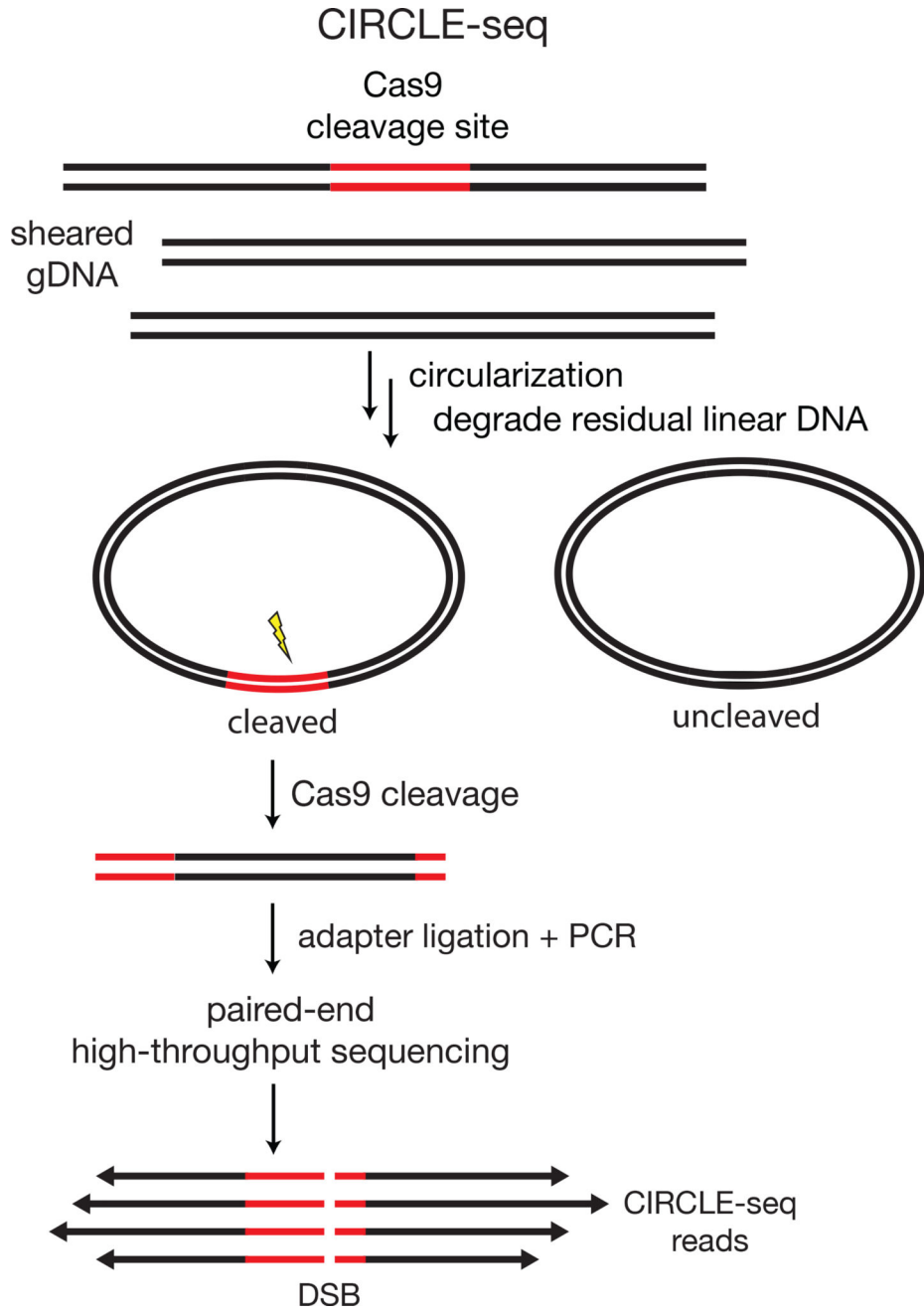
This work was supported by a National Institutes of Health (NIH) Director's Pioneer Award (DP1 GM105378), NIH R35 GM118158, and NIH R01 GM107427 (to J.K.J.), and the Jim and Ann Orr Research Scholar Award (to J.K.J.).

### References

1. Cong L, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013; 339:819–823. [PubMed: 23287718]
2. Hwang WY, et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol*. 2013; 31:227–229. [PubMed: 23360964]
3. Jinek M, et al. RNA-programmed genome editing in human cells | eLife. *Elife*. 2013; 2:e00471. [PubMed: 23386978]

4. Mali P, et al. RNA-guided human genome engineering via Cas9. *Science*. 2013; 339:823–826. [PubMed: 23287722]
5. Jinek M, et al. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*. 2012; 337:816–821. [PubMed: 22745249]
6. Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*. 2014; 346:1258096. [PubMed: 25430774]
7. Bolukbasi MF, Gupta A, Wolfe SA. Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery. *Nat Meth*. 2016; 13:41–50.
8. Mali P, Esvelt KM, Church GM. Cas9 as a versatile tool for engineering biology. *Nat Meth*. 2013; 10:957–963.
9. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*. 2014; 32:347–355. [PubMed: 24584096]
10. Maeder ML, Gersbach CA. Genome-editing Technologies for Gene and Cell Therapy. *Mol Ther*. 2016; 24:430–446. [PubMed: 26755333]
11. Lin J, Musunuru K. Genome Engineering Tools for Building Cellular Models of Disease. *FEBS J*. 2016; doi: 10.1111/febs.13763
12. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*. 2014; 157:1262–1278. [PubMed: 24906146]
13. Brandsma I, Gent DC. Pathway choice in DNA double strand break repair: observations of a balancing act. *Genome Integr*. 2012; 3:9. [PubMed: 23181949]
14. Symington LS, Gautier J. Double-strand break end resection and repair pathway choice. *Annu. Rev. Genet*. 2011; 45:247–271. [PubMed: 21910633]
15. Kass EM, Jasin M. Collaboration and competition between DNA double-strand break repair pathways. *FEBS Letters*. 2010; 584:3703–3708. [PubMed: 20691183]
16. Wyman C, Kanaar R. DNA double-strand break repair: all's well that ends well. *Annu. Rev. Genet*. 2006; 40:363–383. [PubMed: 16895466]
17. Rouet P, Smih F, Jasin M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol*. 1994; 14:8096–8106. [PubMed: 7969147]
18. Sternberg SH, LaFrance B, Kaplan M, Doudna JA. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature*. 2015; 527:110–113. [PubMed: 26524520]
19. Kiani S, et al. Cas9 gRNA engineering for genome editing, activation and repression. *Nat Meth*. 2015; :1–6. DOI: 10.1038/nmeth.3580
20. Dahlman JE, et al. Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat Biotechnol*. 2015; :1–4. DOI: 10.1038/nbt.3390 [PubMed: 25574611]
21. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol*. 2014; 32:279–284. [PubMed: 24463574]
22. Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*. 2014; doi: 10.1038/nature13579
23. Shah SA, Erdmann S, Mojica FJM, Garrett RA. Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol*. 2013; 10:891–899. [PubMed: 23403393]
24. Tsai SQ, Joung JK. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nature*. 2016; 17:300–312.
25. Bolukbasi MF, Gupta A, Wolfe SA. Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery. *Nat Meth*. 2015; 13:41–50.
26. Gori JL, et al. Delivery and Specificity of CRISPR-Cas9 Genome Editing Technologies for Human Gene Therapy. *Hum Gene Ther*. 2015; 26:443–451. [PubMed: 26068008]
27. Cox DBT, Platt RJ, Zhang F. Therapeutic genome editing: prospects and challenges. *Nat Med*. 2015; 21:121–131. [PubMed: 25654603]
28. Gabriel R, et al. An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol*. 2011; 29:816–823. [PubMed: 21822255]
29. Ran FA, et al. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature*. 2015; 520:186–191. [PubMed: 25830891]

30. Tsai SQ, et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol.* 2015; 33:187–197. [PubMed: 25513782]
31. Frock RL, et al. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat Biotechnol.* 2015; 33:179–186. [PubMed: 25503383]
32. Crosetto N, et al. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat Meth.* 2013; 10:361–365.
33. Pattanayak V, et al. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol.* 2013; 31:839–843. [PubMed: 23934178]
34. Kim D, et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Meth.* 2015; doi: 10.1038/nmeth.3284
35. Kim D, Kim S, Kim S, Park J, Kim J-S. Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. *Genome Res.* 2016; gr.199588.115. doi: 10.1101/gr.199588.115
36. Yang L, et al. Targeted and genome-wide sequencing reveal single nucleotide variations impacting specificity of Cas9 in human stem cells. *Nature Communications.* 2014; 5:5507.
37. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature.* 2015; 526:68–74. [PubMed: 26432245]
38. Sherry ST, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001; 29:308–311. [PubMed: 11125122]
39. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012; 489:57–74. [PubMed: 22955616]
40. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2010; 26:589–595. [PubMed: 20080505]
41. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics.* 2011; 27:2987–2993. [PubMed: 21903627]
42. Li H, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009; 25:2078–2079. [PubMed: 19505943]



**Figure 1. Overview of CIRCLE-seq methods for detection of genome-wide CRISPR-Cas9 nuclease cleavage**  
 Schematic overview of the circularization for *in vitro* reporting of cleavage effects by sequencing (CIRCLE-seq) method. Genomic DNA is sheared and circularized by ligation of stem-loop adapters, nicking of stem-loop regions to expose 4 nt palindromic overhangs, and intramolecular ligation. Undesired linear DNA molecules are degraded away by exonuclease treatment. Circular DNA molecules containing a Cas9 cleavage site (red) can be subsequently linearized with Cas9, releasing newly cleaved DNA ends for adapter ligation, PCR amplification, and paired-end high-throughput sequencing. Each pair of reads

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



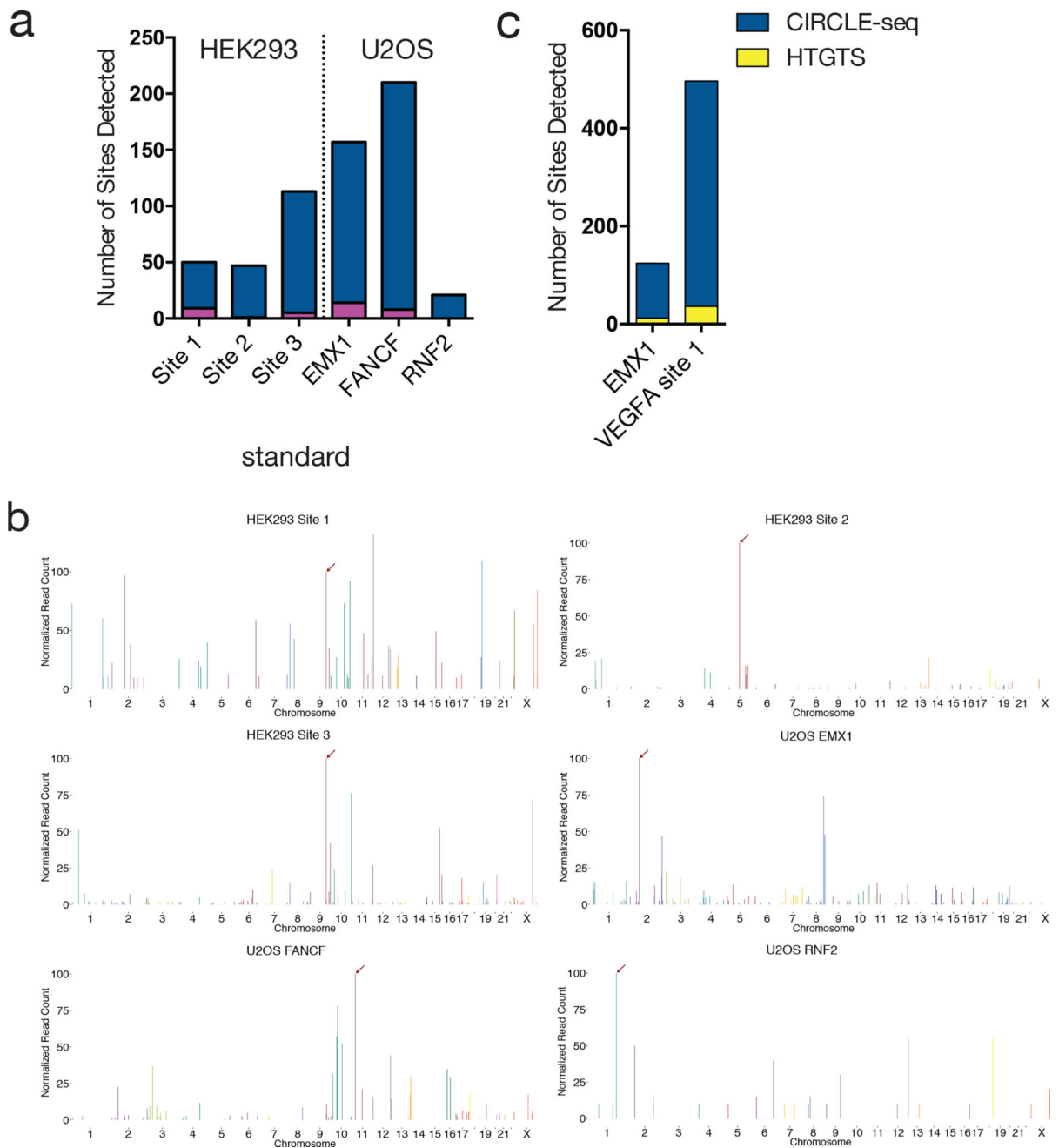
generated by Cas9 cleavage contains complete sequence information for a single off-target site.

Author Manuscript

Author Manuscript

Author Manuscript

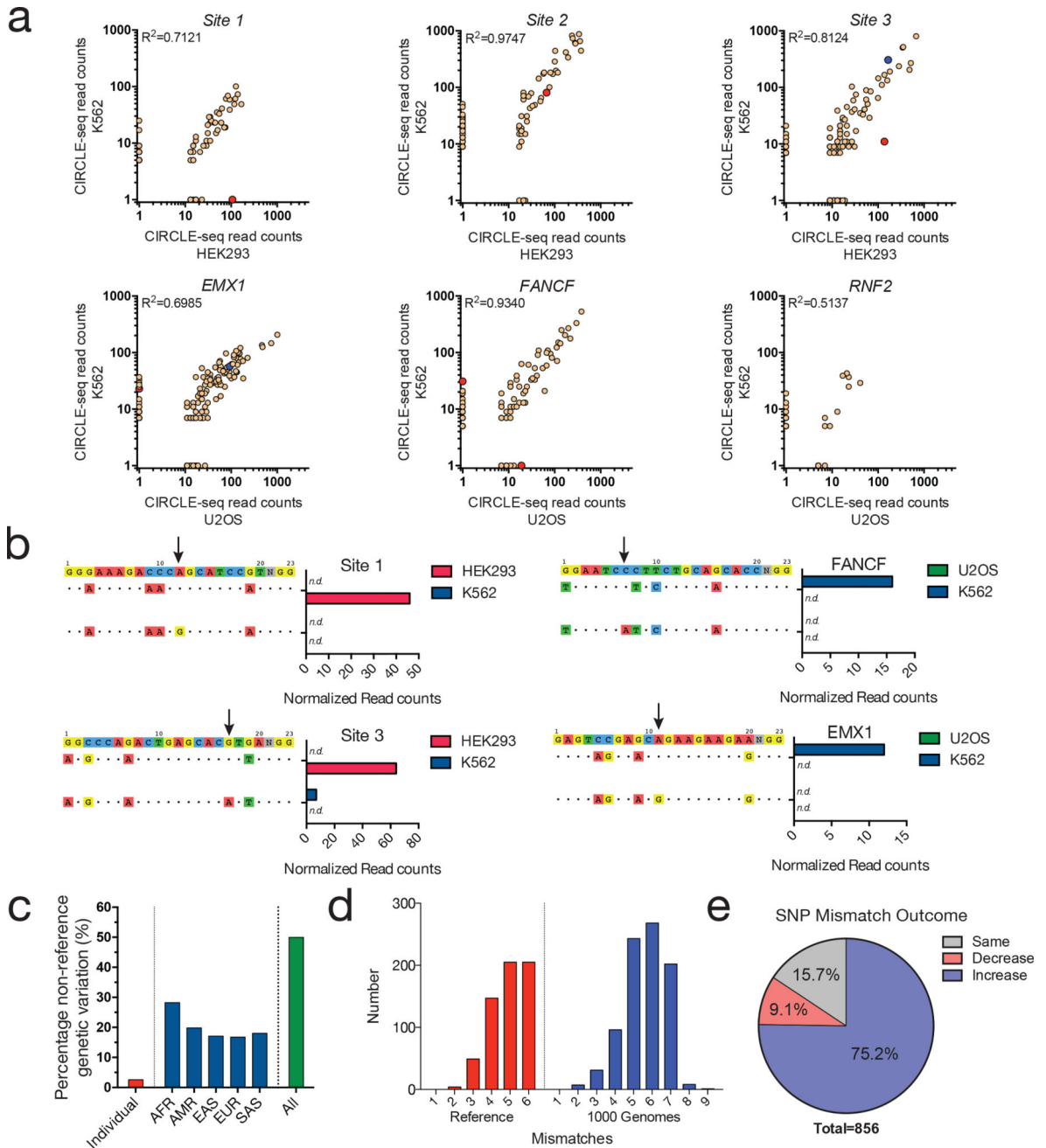
Author Manuscript



**Figure 2. Comparisons of CIRCLE-seq with cell-based GUIDE-seq and HTGTS methods**  
 (a) Histogram showing the number of sites identified exclusively by CIRCLE-seq (blue) and by both CIRCLE-seq and GUIDE-seq (magenta) for gRNAs designs toward standard and more challenging repetitive targets (b) Manhattan plots of CIRCLE-seq detected off-target sites, with bar heights representing CIRCLE-seq read count (normalized to site with highest read count) and organized by chromosomal position. (c) Histogram showing the number of sites detected exclusively by CIRCLE-seq (blue) or by both CIRCLE-seq and HTGTS (yellow).



target sites are ordered top to bottom by CIRCLE-seq read count with mismatches to the intended target sequence indicated by colored nucleotides. Observed tag integration frequencies observed for control (blue) and nuclease-treated (red) cells are plotted on a log scale. (e) Pie charts showing fractions of CIRCLE-seq sites analyzed that are also detected by targeted tag sequencing. (f) Plots of integration positions observed by targeted tag sequencing. PAM bases are the last three nucleotides from the right. Integrations occur at positions proximal to the location of the predicted DSB (three base pairs away the PAM).



**Figure 4. Using CIRCLe-seq to assess the impacts of personalized SNPs on off-target site analysis**

(a) Scatterplots of CIRCLe-seq read counts from experiments performed on genomic DNA from two different cell types. Sites with non-reference genetic variation in only one cell type are highlighted in red, while those with non-reference variation in both cell types are highlighted in blue. (b) Examples of allele-specific CIRCLe-seq read counts at off-target sites with non-reference genetic variation. Mismatches to the intended target sequence are indicated with colored nucleotides, while matching bases are indicated with a dot. The base position harboring the differential genetic change between cell types is indicated with a small arrow. (c) Proportion of CIRCLe-seq off-target sites where non-reference genetic

variation was identified in genotyped individuals from the 1000 Genomes Project: African (AFR), Ad Mixed American (AMR), East Asian (EAS), European (EUR), and Southeast Asian (SAS) superpopulations, and a combined population average. (d) Histogram showing distribution of CIRCLE-seq off-target sites by numbers of mismatches in reference human genome sequence (red) and in 1000 Genomes Project data-derived off-target site haplotypes (blue). (e) Proportion of 1000 Genomes Project-derived haplotypes with increased (blue), decreased (red), or same (grey) numbers of mismatches in off-target sites identified by CIRCLE-seq.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript