


METHOD

Open Access



# Guide RNAs with embedded barcodes boost CRISPR-pooled screens

Shiyou Zhu<sup>1†</sup>, Zhongzheng Cao<sup>1,2†</sup>, Zhiheng Liu<sup>1,2†</sup>, Yuan He<sup>1†</sup>, Yinan Wang<sup>1,2</sup>, Pengfei Yuan<sup>3</sup>, Wei Li<sup>4,5</sup>, Feng Tian<sup>1</sup>, Ying Bao<sup>1</sup> and Wensheng Wei<sup>1\*</sup> 

## Abstract

We report a new method using re-designed guide RNAs with internal barcodes (iBARs) embedded in their loop regions. Our iBAR approach outperforms the conventional method by producing screening results with much lower false-positive and false-negative rates especially with a high multiplicity of infection (MOI). Importantly, the iBAR approach reduces the starting cells at high MOI significantly with greatly improved efficiency and accuracy compared with the canonical CRISPR screens at a low MOI. This new system is particularly useful when the source of cells is limited or when it is difficult to control viral infection for *in vivo* screening.

## Introduction

The CRISPR/Cas9 system enables editing at targeted sites in the genome with high efficiency and specificity [1–3]. One of its extensive applications is to identify the functions of coding genes, non-coding RNAs, and regulatory elements through high-throughput pooled screening in combination with next-generation sequencing (NGS) analysis. By introducing a pooled single-gRNA (sgRNA) or paired-gRNA (pgRNA) library into cells expressing Cas9 or catalytically inactive Cas9 (dCas9) fused with effector domains, investigators can perform multifarious genetic screens by generating diverse mutations, large genomic deletions, transcriptional activation, or transcriptional repression [4–11]. To generate a high-quality cell library of gRNAs for any given CRISPR screen, one must use a low multiplicity of infection (MOI) during cell library construction to ensure that the majority of cells harbors one sgRNA/pgRNA to minimize the false-positive discovery rate [7, 12, 13]. To further reduce the FDR (false discovery rate) and increase data reproducibility, in-depth coverage of gRNAs and multiple biological replicates are often necessary to obtain the hit genes with high statistical

significance [12], resulting in increased library size. Additional difficulties may arise when one performs a large number of genome-wide screens, when the cell materials for library construction are limited, or when one conducts more challenging screens (i.e., *in vivo*) for which it is difficult to arrange the experimental replications or control the MOI. To address these technical difficulties, we aimed to design a novel method that enabled us to use a high MOI to generate CRISPR libraries in target cells and to obtain improved statistics for the screening results. We hypothesized that both the false-positive and false-negative rates of high MOI screens could be drastically reduced if we had more experimental replicates for each of the gRNAs. If we assigned the gRNAs different barcodes that were somehow embedded within the gRNA sequences, we could then trace the performance of each gRNA multiple times within the same experiment by counting both the gRNAs and their internal barcode (iBAR) sequences.

Two papers have recently reported methods to generate random barcodes outside the sgRNA body for pooled CRISPR screening [14, 15]. Assuming each sgRNA would create both desired loss-of-function (LOF) and non-LOF alleles, calculating all reads of any given sgRNA is unable to accurately assess the importance of its targeting gene in negative screening. Much improved statistical results could be achieved by linking one UMI (unique molecular identifier) with one editing outcome of each sgRNA to enable single-cell lineage tracing so as

\* Correspondence: [wswwei@pku.edu.cn](mailto:wswwei@pku.edu.cn)

<sup>†</sup>Shiyou Zhu, Zhongzheng Cao, Zhiheng Liu and Yuan He contributed equally to this work.

<sup>1</sup>Biomedical Pioneering Innovation Center (BIOPIC), Beijing Advanced Innovation Center for Genomics, Peking-Tsinghua Center for Life Sciences, Peking University Genome Editing Research Center, State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, China

Full list of author information is available at the end of the article



to lower the false-negative rate [14], or by counting the decreased number of RSLs (random sequence labels) affiliated with sgRNAs to improve screening quality [15]. Different from these two methods, we aimed to develop a method through embedded iBARs to enable pooled screening with CRISPR library made of viral infection at a high MOI, so as to reduce library size and improve data quality.

## Results

### A design-based CRISPR sgRNA<sup>iBAR</sup> library screening method

Ideally, the embedded iBAR should not affect the efficiency of the gRNA in guiding the Cas9 or dCas9 nuclease to the target site. We decided to place the barcode sequence in the tetra loop of the gRNA scaffold outside of the Cas9-sgRNA ribonucleoprotein complex, which has been subjected to frequent alterations for various purposes maintaining enough activity of its upstream guide sequence [11, 16]. We arbitrarily designed a 6-nt-long iBAR (iBAR<sub>6</sub>) that gave rise to 4096 barcode combinations, providing sufficient variation for our purposes (Fig. 1a). To determine whether the insertions of these extra iBAR sequences affected the gRNA activities, we constructed a library of a pre-determined sgRNA targeting the anthrax toxin receptor gene *ANTXR1* [17] in combination with all 4096 types of iBAR<sub>6</sub>. This special sgRNA<sup>iBAR-ANTXR1</sup> library was constructed in HeLa cells that constantly express Cas9 [7, 8] through lentiviral transduction at a low MOI of 0.3. After three rounds of PA/LFnDTA toxin treatment and enrichment, the sgRNA along with its iBAR<sub>6</sub> sequences from toxin-resistant cells were examined through NGS analysis as previously reported [7]. The majority of sgRNAs<sup>iBAR-ANTXR1</sup> and the sgRNAs<sup>ANTXR1</sup> without barcodes were significantly enriched, whereas almost all the non-targeting control sgRNAs were absent in the resistant cell populations. Importantly, the enrichment levels of sgRNAs<sup>iBAR-ANTXR1</sup> with different iBAR<sub>6</sub>s appeared to be random between two biological replicates (Fig. 1b). After calculating the nucleotide frequency at each position of iBAR<sub>6</sub>, we failed to observe any bias of nucleotides from either of the replicates (Fig. 1c). Additionally, the GC contents in iBAR<sub>6</sub> did not seem to affect the sgRNA cutting efficiency (Additional file 1: Figure S1). However, there was a small number of iBAR<sub>6</sub>s whose affiliated sgRNA<sup>ANTXR1</sup> did not perform well in either screening replicate (Additional file 2: Table S1). To rule out the possibility that these iBAR<sub>6</sub>s had negative effects on sgRNA activity, we selected six different iBARs from the bottom of the sgRNA<sup>iBAR-ANTXR1</sup> ranking for further investigation. Compared to the control sgRNA<sup>ANTXR1</sup> without a barcode, all six of these sgRNAs<sup>iBAR-ANTXR1</sup> showed comparable efficiency in generating both DNA

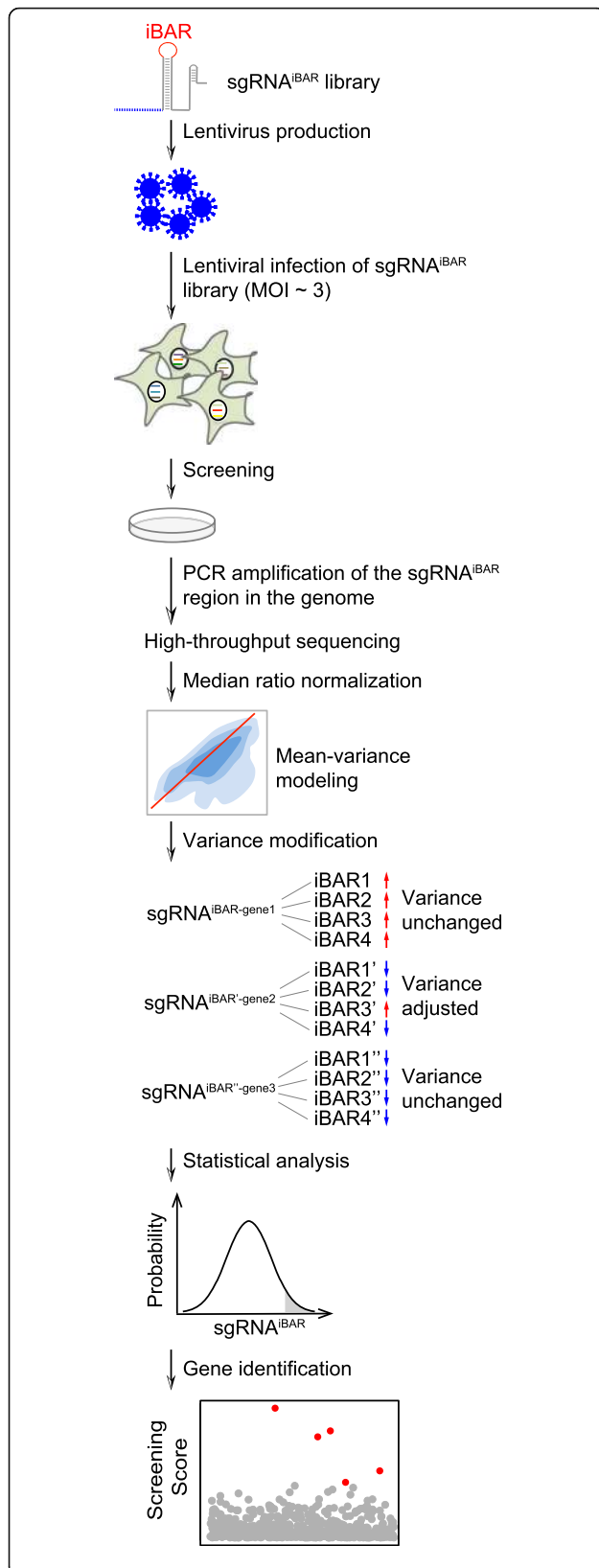
double-stranded breaks (DSBs) at target sites (Fig. 1d) and *ANTXR1* gene disruption leading to the toxin resistance phenotype (Fig. 1e). We further confirmed the negligible effects of iBARs on sgRNA efficiency by four different sgRNAs targeting *CSPG4*, *MLH1*, and *MSH2*, respectively (Additional file 1: Figure S2). Taken together, these results indicate that this re-designed sgRNA<sup>iBAR</sup> retains sufficient activity of sgRNA, making it possible to generally apply this strategy in CRISPR-pooled screens.

Based on the iBAR strategy, we then set out to broaden its application to perform a novel sgRNA<sup>iBAR</sup> library screen at a high MOI. We followed the standard procedure to harvest the library cells, extract their genomic DNA for PCR amplification of sgRNA with iBAR coding regions, and perform NGS analysis [7, 12, 13]. The MAGeCK algorithm could be used to calculate the statistical significance of an sgRNA score through normalization of its raw counts, estimation of its variance using a negative binomial (NB) model, and determination of its ranking using a null model with a uniform distribution [18]. Taking the iBAR into consideration, we assessed the consistency of any sgRNA count change among all the associated iBARs within the same experimental replicate. This process effectively eliminates free riders that were associated with functional sgRNAs due to lentiviral infection at a high MOI in cell library construction. Specifically, for the iBAR system, we purposely adjusted the model-estimated variance for only those sgRNAs whose fold changes with multiple iBARs were in opposite directions, resulting in increased *P* values for these outliers. Finally, we identified hit genes based on sgRNA scores and technical variance between biological replicates (Fig. 2 and “Methods” section). We developed this specific MAGeCK-based algorithm named MAGeCK<sup>iBAR</sup> for the analysis of gRNA<sup>iBAR</sup> library screening that is open source and freely available for download (“Methods” section).

### Comparison of screenings at MOI of 0.3, 3, and 10 for essential genes involved in TcdB toxicity

We then constructed an sgRNA<sup>iBAR</sup> library covering every annotated human gene. For each of the 19,210 human genes, three unique sgRNAs were designed using our newly developed DeepRank method (“Methods” section), each of which was randomly assigned four iBAR<sub>6</sub>s. In addition, 1000 non-targeting sgRNAs, each with four iBAR<sub>6</sub>s, were included as negative controls (Additional file 2: Table S2). For the ease of statistical comparison, every set of three unique non-targeting sgRNAs was artificially named a negative control gene. The 85-nt sgRNA<sup>iBAR</sup> oligos were designed in silico (Additional file 1: Figure S3), synthesized using array synthesis, and cloned as a pooled library into a lentiviral backbone. Cas9-expressing HeLa cells were

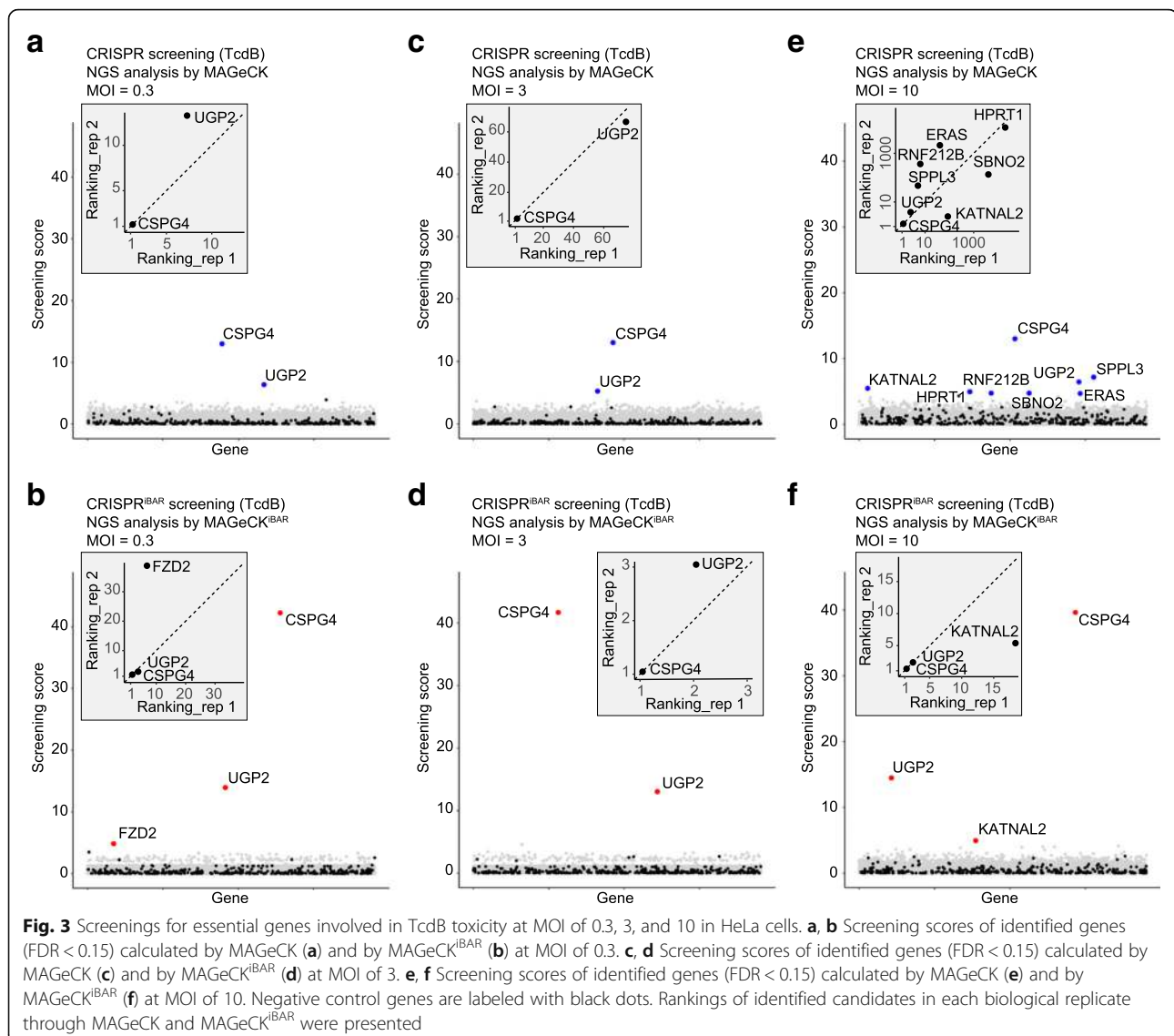




**Fig. 2** Schematic of CRISPR-pooled screening using sgRNA<sup>iBAR</sup>. For a given sgRNA<sup>iBAR</sup> library, four different iBARs were randomly assigned to each sgRNA. The sgRNA<sup>iBAR</sup> library was introduced into target cells through lentiviral infection with a high MOI (i.e., ~ 3). After library screening, sgRNAs with their associated iBARs from enriched cells were determined through NGS. For data analysis, median ratio normalization was applied, followed by mean-variance modeling. The variance of sgRNA<sup>iBAR</sup> was determined based on the fold change consistency of all iBARs assigned to the same sgRNA. The *P* value of each sgRNA<sup>iBAR</sup> was calculated using the mean and modified variance. Robust rank aggregation (RRA) scores of all genes were considered to identify hit genes. A lower RRA score corresponded to a stronger enrichment of the hit genes

transduced with the sgRNA<sup>iBAR</sup> library lentivirus at three different MOIs (0.3, 3, and 10) with 400-fold coverage for sgRNAs to generate cell libraries, in which each sgRNA<sup>iBAR</sup> was covered 100-fold. To evaluate the effect of iBAR design for CRISPR screening at different MOIs, we performed a positive screening to identify genes that mediate the cytotoxicity of *Clostridium difficile* toxin B (TcdB), one of the key virulence factors of this anaerobic bacillus [19]. We have previously reported the first identification of the functional receptor of TcdB, CSPG4 [20], whose coding gene was also identified and ranked at the very top from a genome-scale CRISPR library screening [21]. In this reported CRISPR screening, *UGP2* gene was also top-ranked hit, and *FZD2* was identified and confirmed to encode the secondary receptor that mediates the TcdB's killing effect on host cells. Of note, the role of *FZD2* was significantly dwarfed by CSPG4 so that the *FZD2* gene could only be identified using the truncated TcdB that had CSPG4-interacting region deleted [21]. In our screens on TcdB, we used MAGeCK<sup>iBAR</sup> (Additional file 2: Table S3) and MAGeCK (Additional file 3: Table S4) to analyze data from iBAR and the conventional CRISPR screens, respectively. We consequently obtained top-ranked genes (FDR < 0.15) from both.

For screening at a low MOI of 0.3, *CSPG4* and *UGP2* were identified and ranked at the top (Fig. 3a), consistent with the previous report [21]. When taking iBARs into account, we identified *FZD2* in addition to *CSPG4* and *UGP2* (Fig. 3b). Because *FZD2* is a proven receptor of TcdB which plays much weaker role than CSPG4 in HeLa cells [21], these results demonstrated that iBAR method offered superior quality and sensitivity to conventional CRISPR screening when constructing cell library at a low MOI. In addition, rankings of *CSPG4* and *UGP2* were far more consistent in CRISPR<sup>iBAR</sup> screening between two experimental replicates, again indicating the much higher quality for the new method (Fig. 3a, b). At high MOIs (3 and 10), *CSPG4* and *UGP2* could be isolated from both CRISPR and CRISPR<sup>iBAR</sup> screens, but the data quality was significantly higher with the latter (Fig. 3c–f). In general, the higher the MOI, the worse the signal-to-noise rate for the



traditional method. At a MOI of 10, the number of false-positive hits was drastically increased in the conventional method, but not in CRISPR<sup>iBAR</sup> screening (Fig. 3e, f). Impressively, *CSPG4* and *UGP2* remained top ranked from CRISPR<sup>iBAR</sup> screening even at an MOI of 10, although the data quality slightly declined (Fig. 3f). Noticeably, nearly all *CSPG4*- and *UGP2*-targeting sgRNAs<sup>iBAR</sup> were significantly enriched after TcdB treatment (Additional file 1: Figure S4), strikingly different from other genes identified at an MOI of 10 using conventional method, such as *SPPL3*, a likely false-positive result (Additional file 1: Figure S4). In comparison of the two biological replicates, *CSPG4* and *UGP2* were all ranked at the top in both biological replicates from CRISPR<sup>iBAR</sup> screens with all MOI conditions (Fig. 3b, d, f), but not from the conventional CRISPR screens where *UGP2* was ranked lower than 60th in both replicates at

an MOI of 3 (Fig. 3c) and many false-positive hits appeared in both replicates at an MOI of 10 (Fig. 3e). These results showed that iBAR method maintained the quality of data even at a high MOI as that at a low MOI for conventional CRISPR screening. Additionally, iBAR approach enables highly repeatable results between biological replicates as multiple replications could be conducted within one experiment (Fig. 3).

#### Identification of genes important for 6-TG-mediated cytotoxicity using the CRISPR<sup>iBAR</sup> and conventional CRISPR-pooled screens

To further evaluate the power of iBAR method, we went on conducting a screening to identify genes that modulate cellular susceptibility to 6-TG [22], a cancer drug that could be processed to inhibit DNA synthesis. We decided to construct the genome-scale sgRNA<sup>iBAR</sup> library



at a MOI of 3 to generate a cell library with high coverage (2000-fold) for each sgRNA, in which each sgRNA<sup>iBAR</sup> was covered 500-fold. The overall read distribution of both experimental replicates was shown (Additional file 1: Figure S5a), and the reference cell libraries of both replicates reached 97% coverage of all originally designed sgRNAs (Additional file 1: Figure S5b). Over 95% of the sgRNAs in the original libraries retained three to four iBARs, indicating the good quality of libraries in which most sgRNAs had sufficient barcode variants for screening and data analysis (Additional file 1: Figure S5c). The fold change of all genes correlated well between the two biological replicates (Additional file 1: Figure S6). For the same 6-TG screening of two sgRNA library replicates, we also employed MAGeCK and MAGeCK<sup>iBAR</sup> analyses. For MAGeCK<sup>iBAR</sup>, we consequently obtained adjusted variance and mean distributions for all the sgRNAs<sup>iBAR</sup> that heightened the variance of sgRNAs with enrichment inconsistent among different iBAR repeats (Additional file 1: Figure S7).

From the positively selected sgRNAs with statistical significance, we identified the top-ranked genes (FDR < 0.15) whose corresponding sgRNAs were consistently enriched among different iBARs (Fig. 4a, Additional file 3: Table S5), and we also found these top genes using the MAGeCK algorithm without taking barcodes into account (Fig. 4b, Additional file 3: Table S6). Consistent with a previous report [23], the sgRNAs targeting *HPRT1* gene were top ranked by both methods. Four genes (*MLH1*, *MSH2*, *MSH6*, and *PMS2*) were previously reported to be involved in 6-TG-mediated cell death [6]. We examined and confirmed the cutting activities of all except one of the primary designed sgRNAs targeting these four genes (Additional file 1: Figure S8), indicating that these genes were indeed irrelevant to 6-TG-mediated cell death in HeLa cells we used (Fig. 4c). When analyzing the two biological replicates separately, the top 20 genes of each replicate showed a high level of consistency with CRISPR<sup>iBAR</sup> screening (Spearman correlation coefficient for rankings = 0.74), whereas the two replicates shared much less commonality when using the conventional method (Spearman correlation coefficient for rankings = -0.09) (Fig. 4d, Additional file 3: Table S7).

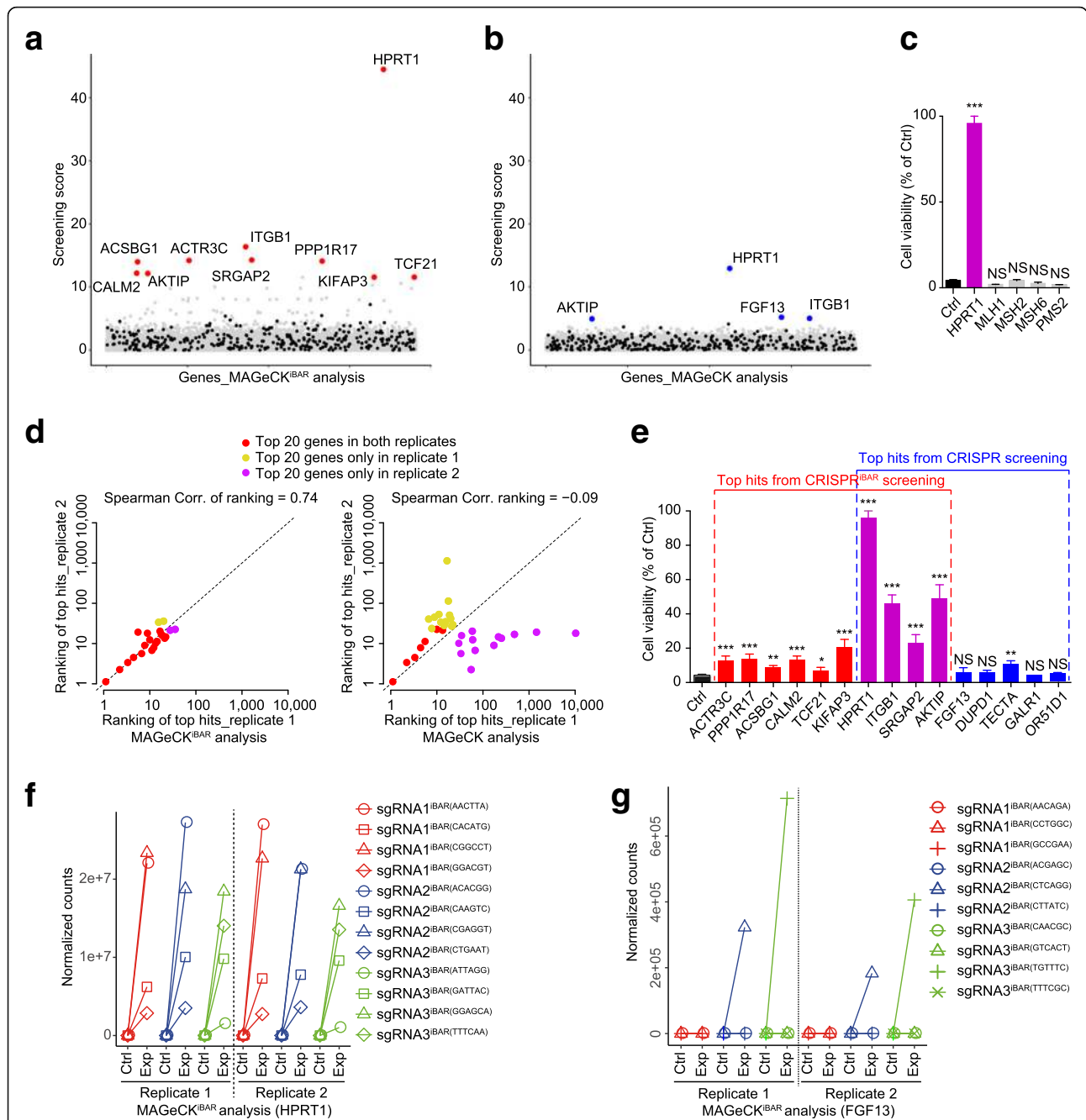
To validate the screening results, we de novo designed and combined two sgRNAs to make a mini-pool to target each candidate gene, and each pool was introduced into HeLa cells through lentiviral infection (Additional file 3: Table S8). The effects of the sgRNA pools on cell viability against 6-TG treatment were quantified by a 3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl-2-H-tetrazolium bromide (MTT) assay. Top 10 genes from CRISPR<sup>iBAR</sup> as well as CRISPR screens were chosen for validation. Noticeably, two non-targeting control genes were identified and ranked in the top 10 candidate list from the conventional

CRISPR screen (Additional file 3: Table S6). These evident false-positive results are predictable because of the high MOI we used to generate the cell library. We successfully confirmed that the top 10 candidate genes from CRISPR<sup>iBAR</sup> of both replicates were all true-positive results; in contrast, only five genes from the top 10 candidate list from the conventional method turned out to be true positives (Fig. 4e). Among them, four genes (*HPRT1*, *ITGB1*, *SRGAP2*, and *AKTIP*) were obtained using both methods, whereas six genes (*ACTR3C*, *PPP1R17*, *ACSBG1*, *CALM2*, *TCF21*, and *KIFAP3*) were only identified and ranked at the top from CRISPR<sup>iBAR</sup>. In summary, iBAR improved accuracy with lower false-positive and false-negative rates for high MOI screens compared with conventional method.

We further assessed the performance of each sgRNA<sup>iBAR</sup> targeting the top four candidate genes (*HPRT1*, *ITGB1*, *SRGAP2*, and *AKTIP*). All the different iBARs of the enriched sgRNAs appeared to have little effect on the enrichment levels of their affiliated sgRNAs, and the order of iBARs associated with any particular sgRNA appeared to be random (Additional file 1: Figure S9), further supporting our prior notion that the iBARs did not affect the efficiency of their affiliated sgRNAs. All four *HPRT1*-targeting sgRNAs<sup>iBAR</sup> were significantly enriched after 6-TG treatment in both replicates (Fig. 4f). Most sgRNAs<sup>iBAR</sup> of other CRISPR<sup>iBAR</sup> identified genes were enriched after 6-TG selection (Additional file 1: Figure S10). In contrast, only a very few of sgRNAs<sup>iBAR</sup> of some top-ranked genes from conventional CRISPR screening were enriched, including *FGF13* (Fig. 4g), *GALRI*, and two negative control genes (Additional file 1: Figure S11), leading to false-positive hits in the MAGeCK but not MAGeCK<sup>iBAR</sup> analysis (Additional file 1: Figure S12).

## Discussion

The library coverage was significantly increased with a high MOI in the transduction with a fixed number of cells for library construction, so the starting cells for library construction could be decreased more than 10-fold (MOI = 3) and 35-fold (MOI = 10) to match and even top the results from conventional screening at an MOI of 0.3 (Additional file 3: Table S10). Four barcodes for each sgRNA, as we designed, appeared to provide sufficient internal repeats to enable the high level of consistency between the two biological replicates using the iBAR method (Fig. 3, Fig. 4d, Additional file 3: Table S7). Therefore, in addition to the significant reduction in cells for library construction, the internal replicates offered by iBARs within the same experiment would lead to more uniform conditions and fairer comparisons versus separate biological replicates, consequently improving statistical scores. The advantage of the iBAR method would become greater when large-scale CRISPR screens



**Fig. 4** Comparison of the CRISPR<sup>iBAR</sup> and conventional CRISPR-pooled screens for the identification of human genes important for 6-TG-mediated cytotoxicity in HeLa cells. **a, b** Screening scores of the top-ranked genes calculated by MAGeCK<sup>iBAR</sup> (**a**) and by MAGeCK (**b**). Identified candidates (FDR < 0.15) were labeled, and only top 10 hits were labeled for MAGeCK<sup>iBAR</sup> screens. Negative control genes were labeled with black dots. **c** Validation of reported genes (*MLH1*, *MSH2*, *MSH6*, and *PMS2*) involved in 6-TG cytotoxicity. **d** Spearman correlation coefficient of the top 20 positively selected genes between two biological replicates using MAGeCK<sup>iBAR</sup> analysis (left) or conventional MAGeCK analysis (right). **e** Validation of top candidate genes isolated by either MAGeCK<sup>iBAR</sup> or MAGeCK analysis. Mini-pooled sgRNAs targeting each gene were delivered to cells through lentiviral infection. Transduced cells were cultured for an additional 10 days before 6-TG treatment. Data are presented as the mean ± S.E.M. (*n* = 5). *P* values were calculated using Student's *t* test. \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001; NS, not significant. The sgRNA sequences for validation are listed in Additional file 3: Table. **S8. f, g** The sgRNA<sup>iBAR</sup> read counts for *HPRT1* targeting (**f**) and *FGF13* targeting (**g**) before (Ctrl) and after (Exp) 6-TG screening in two replicates

in multiple cell lines are in demand or when the cell samples for screening are scarce (i.e., samples from patients or those of primary origin). Especially for in vivo screening in which the lentiviral transduction rate is hard to predict and variable conditions in different animals might greatly impact the screening outcomes, the iBAR method could be an ideal solution to resolve these technical limitations.

Because multiple cuttings decrease cell viability, CRISPR library constructed at a high MOI might have abnormal false discovery rate for negative screening [24, 25]. We therefore performed a genome-scale negative screening at an MOI of 0.3 to assess the iBAR method in calling essential genes. For positive screening using iBAR, we modified the model-estimated variance of sgRNAs with different fold change directions among barcodes to enlarge variance so that the mis-associated sgRNAs were subjected to adequate penalty. For negative screening, however, sgRNA depletion through mis-association had little effect on its consistency of fold change directions as non-functional sgRNAs remained unchanged. Therefore, we treated barcodes only as internal replicates without the penalty procedure. We indeed achieved improved statistics with higher true-positive and lower false-positive rates for negative screening using iBAR method at a low MOI than the conventional approach using gold standard essential genes [26] (Additional file 1: Figure S13).

Notwithstanding the technical advancement of the iBAR method to offer the same benefit of internal replicates, we must be cautious with the MOI during viral transduction to generate the original cell library in negative screens based on measuring cell viability. Although massive integrations have been reported not to affect cell fitness [27], multiple cuttings on DNA caused by higher MOI in cells with active Cas9 have been shown to reduce cell viability [24, 25]. Strategies without cuttings, such as CRISPRi/a [10] or iSTOP systems [28], could be better choices to combine with the iBAR system for negative screening at a high MOI.

Although we had data to support that iBAR<sub>6</sub> had little effect on the activities of sgRNAs, we would not recommend to use barcodes with consecutive T (> 4) so as to avoid any minor effects. Ultimately, 4096 types of iBAR<sub>6</sub> provided sufficient varieties to make CRISPR libraries. In addition, the length of the iBAR is not limited to 6 nt. We have tested different lengths of iBARs and found that their lengths could be up to 50 nt without affecting functions of their affiliated sgRNAs.

(Additional file 1: Figure S14). In addition, it is not necessary to design different barcode sets for different sgRNAs. A fixed set of iBARs assigned to all sgRNAs should work as well as random assignment in library screening. Our iBAR strategy with a streamlined analytic tool MAGeCK<sup>iBAR</sup> would facilitate large-scale CRISPR

screens for broad biomedical discoveries in various settings.

## Methods

### Cells and reagents

HeLa and HEK293T cell lines were maintained in Dulbecco's modified Eagle's medium (DMEM; Gibco C11995500BT) supplemented with 1% penicillin/streptomycin and 10% fetal bovine serum (FBS; CellMax BL102-02) and cultured with 5% CO<sub>2</sub> at 37 °C. All cells were checked for the absence of mycoplasma contamination.

### Plasmid construction

The lentiviral sgRNA<sup>iBAR</sup>-expressing backbone was constructed by changing the position of the BsmBI (Thermo Scientific™, ER0451) site using BstBI (NEB, R0519) and XhoI (NEB, R0146) from Plenti-sgRNA-Lib (Addgene, #53121). sgRNA- and sgRNA<sup>iBAR</sup>-expressing sequences were cloned into the backbone using the BsmBI-mediated Golden Gate cloning strategy [29].

### Design of the genome-scale CRISPR sgRNA<sup>iBAR</sup> library

Gene annotations were retrieved from the UCSC hg38 genome, which contains 19,210 genes. For each gene, three different sgRNAs that had at least one mismatch in the 16-bp seed region in the genome with a high level of predicted targeting efficiency were designed using our newly developed DeepRank algorithm. We then randomly assigned four 6-bp iBARs (iBAR<sub>6</sub>s) to each sgRNA. We designed an additional 1000 non-targeting sgRNAs, each with four iBAR<sub>6</sub>s, to serve as negative controls.

### Construction of the CRISPR sgRNA<sup>iBAR</sup> plasmid library

The 85-nt DNA oligonucleotides were designed and array synthesized. Primers (oligo-F and oligo-R) targeting the flanking sequences of oligos were used for PCR amplification. The PCR products were cloned into the lentiviral vector constructed above using the Golden Gate method [29]. The ligation mixtures were transformed into Trans1-T1 competent cells (Transgene, CD501-03) to obtain library plasmids. Transformed clones were counted to ensure at least 100-fold coverage for the scale of the sgRNA<sup>iBAR</sup> library. The library plasmids were extracted following the standard protocol (QIAGEN 12362) and transfected into HEK293T cells with the two lentivirus package plasmids pVSVG and pR8.74 (Addgene, Inc.) to obtain the library virus. The iBAR library containing all 4096 iBAR<sub>6</sub>s for one *ANTXR1*-targeting sgRNA was constructed using the same protocol.



### Screening of the sgRNA<sup>iBAR-ANTXR1</sup> library containing all 4096 types of iBAR<sub>6</sub>

A total of  $2 \times 10^7$  cells were plated on 150-mm Petri dishes and infected with the library lentivirus at an MOI of 0.3. After 72 h of infection, cells were re-seeded and treated with 1 µg/ml of puromycin (Solarbio P8230) for 48 h. For each replicate,  $5 \times 10^6$  cells were collected for genome extraction. Screening of the sgRNA<sup>iBAR-ANTXR1</sup> library was performed using PA/LFnDTA toxin [30, 31] after library-infected cells were cultured for 15 days [7]. Then, sgRNA with the iBAR coding region in genomic DNA was amplified (TransGen, AP131-13) using Primer-F and Primer-R and then subjected to high-throughput sequencing analysis (Illumina HiSeq2500) using an NEBNext Ultra DNA Library Prep Kit for Illumina (NEB E7370L).

### Screening of the genome-scale CRISPR/Cas9 sgRNA<sup>iBAR</sup> library for genes important for TcdB cytotoxicity and for genes essential for cell viability

A total of  $1.6 \times 10^8$  cells (MOI = 0.3),  $1.53 \times 10^7$  cells (MOI = 3), and  $4.6 \times 10^6$  cells (MOI = 10) were plated on 150-mm Petri dishes respectively for sgRNA library construction for two replicates. Cells were infected with the library lentivirus of different MOIs and treated with 1 µg/ml of puromycin for 72 h post infection. sgRNA<sup>iBAR</sup>-integrated cells were cultured for an additional 15 days to maximize gene knockout. Cells were re-seeded onto 150-mm Petri dishes, treated by TcdB (100 pg/ml) for 10 h, and followed by the removal of the loosely attached round cells through repeated pipetting [20]. For each round of screening, the cells were cultured in fresh medium without TcdB to reach ~50–60% confluence. All resistant cells in one replicate were pooled and subjected to another round of TcdB screening. For the subsequent three rounds of screening, the TcdB concentration was 125 pg/ml, 150 pg/ml, and 175 pg/ml, respectively. After four rounds of treatment, the resistant cells and untreated cells were collected for genomic DNA extraction, amplification of sgRNA, and NGS analysis. Seven pairs of primers were used for PCR amplification (Additional file 3: Table S9), and PCR products were mixed for NGS. For negative screening at an MOI of 0.3, a total of  $4.6 \times 10^7$  (two replicates) sgRNA<sup>iBAR</sup>-integrated cells were cultured for 28 days before NGS decoding.

### Screening of the genome-scale CRISPR/Cas9 sgRNA<sup>iBAR</sup> library for genes important for 6-TG cytotoxicity

A total of  $5 \times 10^7$  cells were plated on 150-mm Petri dishes, and two replicates were obtained. Cells were infected with the library lentivirus at an MOI of 3 and treated with 1 µg/ml puromycin 72 h after infection. sgRNA<sup>iBAR</sup>-integrated cells were cultured for an additional 15 days, re-seeded at a total number of  $5 \times 10^7$ , and then treated with 200 ng/ml 6-TG (Selleck). For the

following two rounds of screening, the 6-TG concentration was 250 ng/ml and 300 ng/ml. For each round of selection, the drug was maintained for 7 days, and the cells were cultured in fresh medium without 6-TG for another 3 days. Then, all the resistant cells in one replicate were grouped together and subjected to another round of 6-TG screening. After three rounds of treatment, the resistant cells and untreated cells were collected for genomic DNA extraction, amplification of sgRNA with iBAR regions, and deep-sequencing analysis.

### Positive screening data analysis

MAGECK<sup>iBAR</sup> is the analysis strategy developed for screens using an sgRNA<sup>iBAR</sup> library based on the MAGECK algorithm [18]. MAGECK<sup>iBAR</sup> takes great advantage of Python, Pandas, NumPy, and SciPy. The analysis algorithm contains three main parts: analysis preparation, statistical tests, and rank aggregation. In the analysis preparation stage, the inputted raw counts of sgRNAs<sup>iBAR</sup> are normalized, and the coefficients of the population mean and variance are then modeled. In the statistical test stage, we use tests to determine the significance of the difference between the treatment and control normalized reads. In the rank aggregation stage, we aggregate the ranks of all the sgRNAs<sup>iBAR</sup> targeting each gene to obtain the final gene ranking.

### Normalization and preparation

We first obtained the raw counts of sgRNAs<sup>iBAR</sup> from sequencing data. Because the sequencing depth and sequencing error might affect the raw counts of the sgRNAs<sup>iBAR</sup>, normalization was needed before the following analysis. A size factor was estimated to normalize the raw counts with different sequencing depths. However, because a few highly enriched sgRNAs might have strong influences on the total read counts, the ratio to total read counts should not be used in the normalization. Thus, we chose the median ratio normalization [32]. Suppose there were  $n$  sgRNAs in the library, with  $i$  ranging from 1 to  $n$ , and  $m$  experiments in total (both control and treatment groups), with  $j$  ranging from 1 to  $m$ . The size factor  $s_j$  can be expressed as follows:

$$s_j = \text{median} \left( \frac{k_{ij}}{\prod_{v=1}^m k_{iv}^{1/m}} \right)$$

Thus, we obtained the normalized counts of sgRNAs<sup>iBAR</sup> in each experiment by calculating the corresponding size factor. In the mean-variance modeling step, the NB distribution was used to estimate the mean and variance of every sgRNA<sup>iBAR</sup> across biological replicates and different treatments [33]:

$$K_{ij} \sim NB(\mu_{ij}, \sigma_{ij}^2)$$

We used the model adopted by MAGeCK to calculate the coefficients of the mean and variance [18]. The mean-variance model satisfied the following relationship:

$$\sigma^2 = \mu + k\mu^b$$

To determine the  $k$  and  $b$  coefficients from all the sgRNAs<sup>iBAR</sup> in the library, the function can be transformed into a linear function:

$$\log_2(\sigma^2 - \mu) = \log_2 k + b \log_2 \mu$$

The means of the treatment and control counts were calculated directly, and the corresponding variance could be calculated from the mean and coefficients. For the CRISPR-iBAR analysis, we evaluated the enrichment of sgRNAs through the performances of different iBARs. We designed four iBARs for each sgRNA to serve as internal replicates. Due to the high MOI during library construction, there must be free riders of false-positive sgRNAs associated with true-positive hits. The free rider here was used to describe the sgRNAs targeting irrelevant genes that were mis-associated with functional sgRNAs to enter the same cells. We modified the variance of sgRNAs<sup>iBAR</sup> based on the enrichment directions of different iBARs for each sgRNA. If all the iBARs of one sgRNA presented the same direction of fold change, i.e., all greater or less than that of the control group, then the variance would be unchanged. However, if one sgRNA with different iBARs revealed inconsistent directions of fold change, then this kind of sgRNA would be penalized by increasing its variance. The final adjusted variance for inconsistent sgRNAs<sup>iBAR</sup> would be the model-estimated variance plus the experimental variance calculated from the Ctrl and Exp samples.

Finally, the score of an sgRNA<sup>iBAR</sup> was calculated by the mean and normalized variance of the treatment compared to those of the control group:

$$\text{score}_i = \frac{t_i - c_i}{v_i}$$

where  $t_i$  is the mean of the treatment counts of the  $i$ th sgRNA and  $c_i$  and  $v_i$  are the mean and variance of control counts of the  $i$ th sgRNA. Because the variance is used as the denominator to calculate the score, the enlarged variance for the inconsistent sgRNAs<sup>iBAR</sup> results in lower score.

#### Statistical test and rank aggregation

The normal distribution was used to test the score<sub>*i*</sub> of the treatment counts. The two sides of scores in a standard normal distribution provided the greater-tail and lesser-tail  $P$  value separately. To obtain the gene

ranks, we used RRA, which is an appropriate method for aggregating rankings [34]. MAGeCK adopted a modified RRA method by limiting the enriched sgRNAs [18]. Suppose for one gene there are  $n$  sgRNAs with different iBARs in the library of  $M$  sgRNAs<sup>iBAR</sup> in total; every sgRNA<sup>iBAR</sup> has a rank in the library of  $R = (R_1, R_2, \dots, R_n)$ . First, the ranks of sgRNAs<sup>iBAR</sup> should be normalized by the total number of sgRNAs<sup>iBAR</sup> in the library. We obtained the normalized rank  $r = (r_1, r_2, \dots, r_n)$  for each  $r_i = R_i/M$ , in which  $1 \leq i \leq n$ . Then, we calculated the sorted normalized ranking  $sr$ , making  $sr_1 \leq sr_2 \leq \dots \leq sr_n$ . The sorted normalized rank follows a uniform distribution between 0 and 1. The probability  $\beta_k, n(sr)$  in which  $sr_i \leq r_i$  follows a  $\beta$  distribution  $\beta(k, n+1-k)$ , making  $\rho = \min(\beta_1, \beta_2, \dots, \beta_n, n)$ . For every gene, the  $\rho$  score can be obtained by RRA and further adjusted by Bonferroni correction [34]. We adopted MAGeCK, which developed  $\alpha$ -RRA, to select the top  $\alpha\%$  sgRNAs from the ranking list. The  $P$  values of sgRNAs lower than a threshold (0.25 for instance) were selected. Only the top sgRNAs of one gene were considered in the RRA calculation, thus making  $\rho = \min(\beta_1, \beta_2, \dots, \beta_j, n)$ , in which  $1 \leq j \leq n$ .

#### Negative screening data analysis

During the analyzing process of positive screening at high MOI based on the iBAR strategy, we modified the model-estimated variance of sgRNAs with different fold change directions among corresponding barcodes. But for negative screening, most of the non-functional sgRNAs would be unchanged. So the variance modification algorithm based on fold change directions of corresponding barcodes becomes not sufficient to justify whether certain sgRNA is a false-positive result. Therefore, we treated barcodes as internal replicates directly. When taking iBAR into consideration, we performed two times robust rank aggregation for the negative screening rather than variance adjustment for the inconsistent sgRNAs<sup>iBAR</sup>. The first round of robust rank aggregation aggregates the sgRNA<sup>iBAR</sup> level to sgRNA level, and the second round aggregates the sgRNA level to gene level.

#### Validation of candidate genes

To validate each gene, we chose two sgRNAs designed in the library and cloned into a lentiviral vector with a puromycin selection marker. We mixed two sgRNA plasmids and co-transfected them into HEK293T cells with two lentiviral package plasmids (pVSVG and pR8.74) using the X-tremeGENE HP DNA transfection reagent (Roche). The HeLa cells stably expressing Cas9 were infected with the lentivirus for 3 days and treated with 1  $\mu$ g/ml puromycin

for 2 days. Then, 5000 cells were added into each well, and five replicates were obtained for each group. After 24 h, the experimental groups were treated with 150 ng/ml 6-TG, and the control groups were treated with normal medium for 7 days. Then, MTT (Amresco) staining and detection were performed following the standard protocol. The experimental wells treated with 6-TG were normalized to the wells without 6-TG treatment.

## Additional files

**Additional file 1: Figure S1.** CRISPR screening of a collection of sgRNAs<sup>iBAR-ANTXR1</sup> containing all 4096 types of iBAR<sub>6</sub> divided by the GC contents of iBARs. Figure S2. Evaluation of the effects of iBARs on sgRNA activity. Figure S3. DNA sequences of the designed oligos. Figure S4. The sgRNA<sup>iBAR</sup> read counts for *CSPG4* targeting (a), *SPPL3* targeting (b), *UGP2* targeting (c), *KATNAL2* targeting (d), *HPRT1* targeting (e), *RNF212B* targeting (f), *SBNQ2* targeting (g) and *ERAS* targeting (h) before (Ctrl) and after (Exp) TcdB screening at MOI of 10 calculated by MAGeCK in two replicates. Figure S5. sgRNA distribution and coverage in different samples. Figure S6. The Pearson Correlation of log<sub>10</sub> (fold change) of all genes between two biological replicates after 6-TG screening at an MOI of 3. Figure S7. Mean-variance model of all the sgRNAs<sup>iBAR</sup> after variance adjustment using MAGeCK<sup>iBAR</sup> analysis. Figure S8. Efficiency of original designed sgRNAs targeting *MLH1*, *MSH2*, *MSH6* and *PMS2*. Figure S9. Fold changes of each sgRNA<sup>iBAR</sup> targeting the indicated top candidate genes (*HPRT1*, *ITGB1*, *SRGAP2* and *AKTIP*) in two experimental replicates. Figure S10. The sgRNA<sup>iBAR</sup> read counts for targeting *ITGB1* (a), *SRGAP2* (b), *AKTIP* (c), *ACTR3C* (d), *PPP1R17* (e), *ACSBG1* (f), *CALM2* (g), *TCF21* (h) and *KIFAP3* (i) in two replicates. Figure S11. The sgRNA<sup>iBAR</sup> read counts for targeting *GALR1* (a), *DUPD1* (b), *TECTA* (c), *OR51D1* (d), *Neg89* (e) and *Neg67* (f) in two replicates. Figure S12. Normalized sgRNA read counts of *HPRT1*, *FGF13*, *GALR1* and *Neg67* via conventional MAGeCK analysis in two experimental replicates. Figure S13. Assessment of screen performance through MAGeCK and MAGeCK<sup>iBAR</sup> analyses by using gold standard essential genes as determined by ROC curves. Figure S14. The effects of different lengths of iBARs on sgRNA activity. (PDF 4136 kb)

**Additional file 2: Table S1.** Results of sgRNA<sup>iBAR-ANTXR1</sup> library screening for the cytotoxicity of PA/LFnDTA. Table S2. Sequences of the human genome-scale sgRNA<sup>iBAR</sup> library. Table S3. Results of the human genome-scale sgRNA<sup>iBAR</sup> library screening at different MOI for the cytotoxicity of TcdB using MAGeCK<sup>iBAR</sup> analysis. (XLSX 15253 kb)

**Additional file 3: Table S4.** Results of the human genome-scale sgRNA<sup>iBAR</sup> library screening at different MOI for the cytotoxicity of TcdB using conventional MAGeCK analysis. Table S5. Results of the human genome-scale sgRNA<sup>iBAR</sup> library screening for the cytotoxicity of 6-TG using MAGeCK<sup>iBAR</sup> analysis. Table S6. Results of the human genome-scale sgRNA<sup>iBAR</sup> library screening for the cytotoxicity of 6-TG using conventional MAGeCK analysis. Table S7. Top 20 gene list of two biological replicates using MAGeCK<sup>iBAR</sup> and MAGeCK analysis. Table S8. sgRNA design for the functional validation of candidate genes from 6-TG screening and sgRNA design for the test of iBAR effects on activity. Table S9. Primers used for PCR amplification of the genomic DNAs and for library construction. Table S10. Comparison of the numbers of cells required for CRISPR library construction for TcdB screenings at different MOIs. (XLSX 12014 kb)

## Acknowledgements

We acknowledge the staff of the BIOPIC sequencing facility (Peking University) for their assistance. The process of data analysis was supported by High-performance Computing Platform of Peking University. We thank the support of National Post-doctoral Innovative Talent Supporting Program.

## Funding

This project was supported by funds from the National Science Foundation of China (NSFC31430025), the Beijing Advanced Innovation Center for

Genomics at Peking University, and the Peking-Tsinghua Center for Life Sciences (to W.W.); the National Major Science & Technology Project for Control and Prevention of Major Infectious Diseases in China (2018ZX10301401, to S.Z.).

## Availability of data and materials

The accession number for the raw sequencing data reported in this paper is NCBI Sequence Read Archive (SRA): SRP164294 (PRJNA494953) [35]. The MAGeCK-based analysis program named as MAGeCK<sup>iBAR</sup> written by python3 for iBAR based CRISPR screening is available at <https://bitbucket.org/WeiLab/mageck-ibar/>. [36]

## Authors' contributions

WW conceived and supervised the project. WW, SZ, ZC, and PY designed the experiments. SZ, ZC, and YH performed the experiments with the help from FT and YB. YW designed the oligos used to construct the iBAR library. ZL performed the data analysis with the help of WL. SZ, ZC, ZL, YH, and WW wrote the manuscript with the help of the other authors. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

Not applicable for this study.

## Consent for publication

Not applicable for this study.

## Competing interests

A patent has been filed relating to the data presented, which would not limit the academic use of the data and the method. W.W. is a founder and scientific advisor for EdiGene. All other authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>Biomedical Pioneering Innovation Center (BIOPIC), Beijing Advanced Innovation Center for Genomics, Peking-Tsinghua Center for Life Sciences, Peking University Genome Editing Research Center, State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, China. <sup>2</sup>Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China. <sup>3</sup>EdiGene Inc, Life Science Park, 22 KeXueYuan Road, Changping District, Beijing 102206, China. <sup>4</sup>Center for Genetic Medicine Research, Children's National Health System, Washington, DC 20010, USA. <sup>5</sup>Department of Genomics and Precision Medicine, The George Washington School of Medicine and Health Sciences, Washington, DC 20010, USA.

Received: 10 December 2018 Accepted: 11 January 2019

Published online: 24 January 2019

## References

- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012;337:816–21.
- Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013;339:819–23.
- Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. *Science*. 2013;339:823–6.
- Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen T, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*. 2014;343:84–7.
- Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic screens in human cells using the CRISPR-Cas9 system. *Science*. 2014;343:80–4.
- Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera Mdel C, Yusa K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol*. 2014;32:267–73.

7. Zhou Y, Zhu S, Cai C, Yuan P, Li C, Huang Y, Wei W. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature*. 2014;509:487–91.
8. Zhu S, Li W, Liu J, Chen CH, Liao Q, Xu P, Xu H, Xiao T, Cao Z, Peng J, et al. Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. *Nat Biotechnol*. 2016;34:1279–86.
9. Liu Y, Cao Z, Wang Y, Guo Y, Xu P, Yuan P, Liu Z, He Y, Wei W. Genome-wide screening for functional long noncoding RNAs in human cells by Cas9 targeting of splice sites. *Nat Biotechnol*. 2018;36:1203–10.
10. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, et al. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell*. 2014;159:647–61.
11. Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*. 2015;517:583–8.
12. Peng J, Zhou Y, Zhu S, Wei W. High-throughput screens in mammalian cells using the CRISPR-Cas9 system. *FEBS J*. 2015;282:2089–96.
13. Zhu S, Zhou Y, Wei W. Genome-wide CRISPR/Cas9 screening for high-throughput functional genomics in human cells. *Methods Mol Biol*. 2017;1656:175–81.
14. Michlits G, Hubmann M, Wu SH, Vainorius G, Budusan E, Zhuk S, Burkard TR, Novatchkova M, Aichinger M, Lu Y, et al. CRISPR-UMI: single-cell lineage tracing of pooled CRISPR-Cas9 screens. *Nat Methods*. 2017;14:1191–7.
15. Schmierer B, Botla SK, Zhang J, Turunen M, Kivioja T, Taipale J. CRISPR/Cas9 screening using unique molecular identifiers. *Mol Syst Biol*. 2017;13:945.
16. Shechner DM, Hacisuleyman E, Younger ST, Rinn JL. Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat Methods*. 2015;12:664–70.
17. Bradley KA, Mogridge J, Mourez M, Collier RJ, Young JA. Identification of the cellular receptor for anthrax toxin. *Nature*. 2001;414:225–9.
18. Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, Irizarry RA, Liu JS, Brown M, Liu XS. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol*. 2014;15:554.
19. Lyras D, O'Connor JR, Howarth PM, Sambol SP, Carter GP, Phumoonna T, Poon R, Adams V, Vedantam G, Johnson S, et al. Toxin B is essential for virulence of *Clostridium difficile*. *Nature*. 2009;458:1176–9.
20. Yuan P, Zhang H, Cai C, Zhu S, Zhou Y, Yang X, He R, Li C, Guo S, Li S, et al. Chondroitin sulfate proteoglycan 4 functions as the cellular receptor for *Clostridium difficile* toxin B. *Cell Res*. 2015;25:157–68.
21. Tao L, Zhang J, Meraner P, Tovaglieri A, Wu X, Gerhard R, Zhang X, Stallcup WB, Miao J, He X, et al. Frizzled proteins are colonic epithelial receptors for *C. difficile* toxin B. *Nature*. 2016;538:350–5.
22. Tan YY, Epstein LB, Armstrong RD. In vitro evaluation of 6-thioguanine and alpha-interferon as a therapeutic combination in HL-60 and natural killer cells. *Cancer Res*. 1989;49:4431–4.
23. Duan J, Nilsson L, Lambert B. Structural and functional analysis of mutations at the human hypoxanthine phosphoribosyl transferase (HPRT1) locus. *Hum Mutat*. 2004;23:599–611.
24. Jackson SP. Sensing and repairing DNA double-strand breaks. *Carcinogenesis*. 2002;23:687–96.
25. Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, Dharia NV, Montgomery PG, Cowley GS, Pantel S, et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet*. 2017;49:1779–84.
26. Hart T, Brown KR, Sircoulomb F, Rottapel R, Moffat J. Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. *Mol Syst Biol*. 2014;10:733.
27. Zhou Y, Wang P, Tian F, Gao G, Huang L, Wei W, Xie XS. Painting a specific chromosome with CRISPR/Cas9 for live-cell imaging. *Cell Res*. 2017;27:298–301.
28. Billon P, Bryant EE, Joseph SA, Nambiar TS, Hayward SB, Rothstein R, Ciccia A. CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of STOP codons. *Mol Cell*. 2017;67:1068–79 e1064.
29. Engler C, Gruetzner R, Kandzia R, Marillonnet S. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One*. 2009;4:e5553.
30. Wei W, Lu Q, Chaudry GJ, Leppla SH, Cohen SN. The LDL receptor-related protein LRP6 mediates internalization and lethality of anthrax toxin. *Cell*. 2006;124:1141–54.
31. Qian L, Cai C, Yuan P, Jeong SY, Yang X, Dealmeida V, Ernst J, Costa M, Cohen SN, Wei W. Bidirectional effect of Wnt signaling antagonist DKK1 on the modulation of anthrax toxin uptake. *Sci China Life Sci*. 2014;57:469–81.
32. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11:R106.
33. Robinson MD, Smyth GK. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics*. 2008;9:321–32.
34. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*. 2012;28:573–80.
35. Zhu S, Cao Z, Liu Z, He Y, Wang Y, Yuan P, Li W, Tian F, Bao Y, Wei W. Guide RNAs with embedded barcodes boost CRISPR-pooled screens. *Sequence Read Archive (SRA)* 2019. Dataset. <https://www.ncbi.nlm.nih.gov/sra/SRP164294>
36. Zhu S, Cao Z, Liu Z, He Y, Wang Y, Yuan P, Li W, Tian F, Bao Y, Wei W. Guide RNAs with embedded barcodes boost CRISPR-pooled screens. *Bitbucket Software*. <https://bitbucket.org/WeiLab/mageck-ibar/>.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

