nature biotechnology

ARTICLES

https://doi.org/10.1038/s41587-021-00944-1

Check for updates

Genome-wide interrogation of gene functions through base editor screens empowered by barcoded sgRNAs

Ping Xu[1,3], Zhiheng Liu[1,3], Ying Liu[1,3], Huazheng Ma[1,2,3], Yiyuan Xu[1], Ying Bao[1], Shiyou Zhu[1], Zhongzheng Cao[1], Zeguang Wu[1], Zhuo Zhou[1] and Wensheng Wei[1]✉

Canonical CRISPR-knockout (KO) screens rely on Cas9-induced DNA double-strand breaks (DSBs) to generate targeted gene KOs. These methodologies may yield distorted results because DSB-associated effects are often falsely assumed to be consequences of gene perturbation itself, especially when high copy-number sites are targeted. In the present study, we report a DSB-independent, genome-wide CRISPR screening method, termed iBARed cytosine base editing-mediated gene KO (BARBEKO). This method leverages CRISPR cytosine base editors for genome-scale KO screens by perturbing gene start codons or splice sites, or by introducing premature termination codons. Furthermore, it is integrated with iBAR, a strategy we devised for improving screening quality and efficiency. By constructing such a cell library through lentiviral infection at a high multiplicity of infection (up to 10), we achieved efficient and accurate screening results with substantially reduced starting cells. More importantly, in comparison with Cas9-mediated fitness screens, BARBEKO screens are no longer affected by DNA cleavage-induced cytotoxicity in HeLa-, K562- or DSB-sensitive retinal pigmented epithelial 1 cells. We anticipate that BARBEKO offers a valuable tool to complement the current CRISPR–KO screens in various settings.

The simplicity of programming a CRISPR–Cas9 system to modify specific genomic loci offers an unprecedented opportunity to interrogate gene function in eukaryotes[1–6]. This system has been further employed to develop powerful genetic screening methods for the functional annotation of genetic elements in various biomedical settings, including cancer research and drug discovery[7–11]. Despite its success and broad applications, Cas9-induced DSBs could have gene-independent anti-proliferation effects, especially in high copy-number and mismatch-tolerance regions, leading to false-positive results in high-throughput screens[12–16]. DSB is one of the most critical lesions that can result in a wide variety of genetic alterations including large- or small-scale deletions, loss of heterozygosity and translocations[17]. Screens of genetic dependency by Cas9 may incur bias in DNA-damage response (DDR). It has recently been reported that Cas9-induced DSBs posed obstacles to high-throughput screens in human nontransformed cells via p53-dependent cell growth arrest[18–21]. High-efficiency Cas9 editing could cause cell death in human pluripotent stem cells (hPSCs)[21] and G1 cell cycle arrest in human telomerase transcriptase subunit, retinal pigmented epithelial 1 cells (hTERT RPE1 cells)[19]. Parallel screens in p53-proficient and -deficient RPE1 cells revealed that Cas9 editing triggered a p53-dependent DDR, which compromised the sensitivity of guide-specific effects[19]. However, some groups argued that adequate single guide (sg)RNA representation in carefully selected cells or clones expressing high-efficiency Cas9 would ensure successful CRISPR–Cas9 screens[18,22].

Nevertheless, to reduce the sgRNA misassociation-associated false discovery rate (FDR), it is common practice to maintain a low multiplicity of infection (MOI) for the lentiviral transduction of the sgRNA library, to ensure that most of the transduced cells harbor only one sgRNA per cell[7–10,23]. We have recently established a new screening strategy using redesigned sgRNA harboring internal barcodes (iBARs) that enables high-throughput CRISPR screening (CRISPR^{iBAR}) at high MOIs, resulting in significant efficiency boost[24]. Although CRISPR^{iBAR} outperformed the conventional methods in positive selection screens, the cytotoxicity of Cas9-induced DSBs[12–16] constrained its application in broader settings such as negative selection screens, especially with high MOIs[24].

We aim to re-establish a CRISPR loss-of-function screening strategy with the following beneficial feature: allowing high-MOI screening to improve efficiency and economy, ideal for both positive and negative selection screens, and applicable to screening in nontransformed cell types such as hPSCs. The simple solution could be the combination of iBAR strategy and CRISPR base editor-mediated gene KOs. CRISPR–STOP and iSTOP approaches have been proposed to utilize the CRISPR-based cytosine base editor 3 to introduce nonsense mutations for gene silencing[25,26]. It is foreseeable that broader coverage of genes using cytosine base editors (CBEs) will be achieved to include additional sites for sgRNA design, splice acceptor sites, splice donor sites and translation initiation sites.

In the present study, we established a genome-wide BARBEKO screening strategy, in which CBEs perturb genes by disrupting splicing sites or translation initiation sites, or introducing premature termination codons (PTCs), and all sgRNAs were redesigned to carry iBARs[24]. The BARBEKO approach to the genome scale has been applied in multiple cell lines—HeLa, K562 and RPE1 cells—all at high MOIs for screens of cell fitness. With proper techniques for

[1]Biomedical Pioneering Innovation Center, 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, China. [2]Peking University-Tsinghua University-National Institute of Biological Sciences Joint Graduate Program, Peking University, Beijing, China. [3]These authors contributed equally: Ping Xu, Zhiheng Liu, Ying Liu, Huazheng Ma. ✉e-mail: wswei@pku.edu.cn

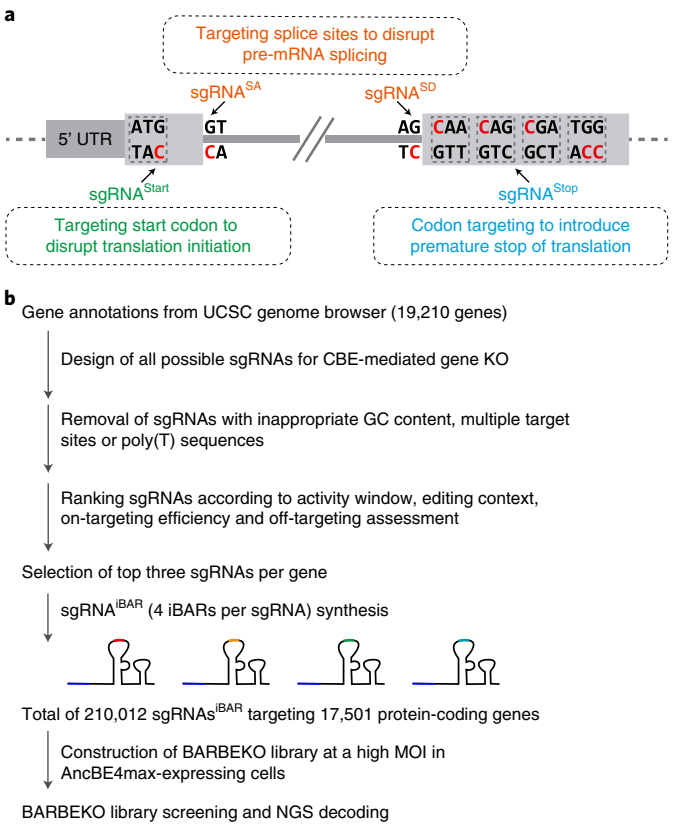


Fig. 1 | Design of CBE-based genome-scale sgRNA library for gene KO screens. a, CBE with sgRNAs targeting start codons (sgRNA^{Start}), splice acceptor sites (sgRNA^{SA}), splice donor sites (sgRNA^{SD}) and codons of gllutamine, arginine or tryptophan (sgRNA^{Stop}) disrupting gene functions. **b**, Selection and filtration of sgRNAs^{iBAR} for the BARBEKO library.

delivery, the BARBEKO strategy could be particularly useful for loss-of-function screens in complex models such as primary cells, organoids and in vivo studies, in which the source of cells is usually limited and sensitive to DNA damage, and when it is hard, if not impossible, to control transduction efficiency in making libraries.

Results

CBE-based genome-wide sgRNA library for KO screens. In addition to generating effective gene KOs by utilizing CBEs to introduce PTCs by targeting codons of glutamine (5′-CAA, 5′-CAG), arginine (5′-CGA) or tryptophan (5′-TGG)[25,26], it is foreseeable to achieve gene KOs by disrupting splice sites (5′-GT, 5′-AG) or start codons (5′-ATG) (Fig. 1a). To examine the effectiveness of CBEs in generating gene KO, we designed multiple sgRNAs along the genomic loci of an anthrax toxin receptor gene *ANTXR1* and a diphtheria toxin receptor gene *HBEGF*[10] (Supplementary Table), followed by the transduction of these sgRNAs individually into CBE-expressing HeLa cells (Extended Data Fig. 1a and Supplementary Fig. 1a). To achieve desirable editing efficiency, AncBE4max, one of the most effective CBEs[27], was employed. By testing the editing kinetics of AncBE4max (Extended Data Fig. 1b and Supplementary Fig. 1b), we chose to treat sgRNA-expressing cells with toxin on day 5 post-transduction. All groups (10/10) with sgRNAs targeting the *ANTXR1* locus obtained resistance to chimeric anthrax toxins (PA/LFnDTA, protective antigen (PA)/N-terminal domain of lethal factor (LF) fused to the catalytic subunit of diphtheria toxin)[28,29] (Extended Data Fig. 1c). Sanger sequencing of resistant cells further confirmed the targeted base transitions (Extended Data Fig. 1d). Consistently, all groups (7/7) with sgRNAs targeting the *HBEGF* locus obtained resistance to diphtheria toxin (Supplementary Fig. 1c,d).

To test the effectiveness of AncBE4max in negative selection screens, we compared the efficiency of gene KOs between AncBE4max and Cas9. By targeting core essential genes *RPL11* and *RPL23A* (Supplementary Table), both Cas9- and AncBE4max-mediated gene KOs efficiently inhibited chronic myeloid leukemia K562 proliferation (Extended Data Fig. 2 and Supplementary Fig. 2). Sanger sequencing analysis demonstrated that CBEs achieved mutagenesis levels comparable with those of Cas9 for gene KOs (Extended Data Fig. 2). Taken together, AncBE4max is competent for both positive and negative selection screens.

We have previously established an iBAR method that enables high-throughput gene KO screening using a CRISPR^{iBAR} library made from high-MOI lentiviral infection[24]. Four verified iBARs were attached to each sgRNA in the BARBEKO library serving as internal replicates in screens (Extended Data Fig. 3a). For the design of BARBEKO at the genome scale, we followed a reasonable scoring scheme considering the AncBE4max activity window, editing context, sgRNA on-targeting efficiency and off-targeting assessment (Fig. 1b and Source Data Fig. 2). Some 210,012 sgRNAs covering 17,501 genes (3 sgRNAs per gene) were designed in silico, of which 41.8% were newly designed, targeting start codons or splice sites, whereas 58.2% CRISPR–STOP sgRNAs were adopted from Kuscu et al.[26] (Extended Data Fig. 3b).

BARBEKO achieved a high-MOI fitness screen in HeLa cells. We first applied the BARBEKO approach to fitness screens at an MOI of 3 in HeLa cells (Fig. 2a). To tailor iBARs to fitness screens, we developed an analysis algorithm termed ZFC^{iBAR} (Fig. 2b). In short, we used a *z*-score to normalize the distribution of $\log_2$(fold-change) (*z*LFC) of each sgRNA^{iBAR} (Supplementary Fig. 3), and combined robust rank aggregation (RRA) analysis[30] to calculate the gene fitness score (FS), which comprehensively reflected the significance and consistency of the abundance change of 12 sgRNAs^{iBAR} per gene. Using ZFC^{iBAR}, both depleted and enriched genes in HeLa cells were revealed under rational cutoffs of gene FS (Fig. 2c and Source Data Fig. 3). With the help of iBARs serving as internal replicates, ZFC^{iBAR} analysis further increased the signal-to-noise ratio of screens, as indicated by Pearson's correlation coefficients of two biological replicates, which increased from 0.75 in sgRNA^{iBAR} *z*LFC analysis to 0.96 in gene FS analysis (Fig. 2d,e). In addition, the F_1 score (harmonic mean of precision and recall, based on gold-standard reference sets[31]) was higher when using ZFC^{iBAR} analysis than the non-iBAR ZFC analysis (Fig. 2f).

Using the area under the curve (AUC) of the receiver operating characteristic (ROC) curves, based on the gold-standard reference sets of essential and nonessential genes, we compared our results with data from a fitness screen utilizing the CRISPR^{iBAR} library[24] at an MOI of 3 and a conventional Cas9 screen at an MOI of 0.3[31]. Fitness screens at a high MOI using the BARBEKO approach outperformed both screens (Fig. 2g, Extended Data Fig. 4a,b and Source Data Figs. 4 and 5). Furthermore, the BARBEKO screen exhibited the maximal extent of depletion in essential genes and a better separation between the distribution of essential and nonessential genes by boxplots, indicating the efficient gene KO and a better-controlled false-positive rate (Extended Data Fig. 4c). Similarly, dAUC (ΔAUC, difference between sgRNAs targeting essential and nonessential genes) of BARBEKO was evidently higher than that of the first-generation CRISPR–KO library[32], demonstrating the enhanced specificity of the BARBEKO library even at high MOIs (Fig. 2h). Taken together, the BARBEKO approach exhibits the potential of high-quality outcomes with much-improved cost and labor effectiveness in fitness screens.

We went on to compare the results of BARBEKO screens between early and late timepoints during the fitness screen. The correlation

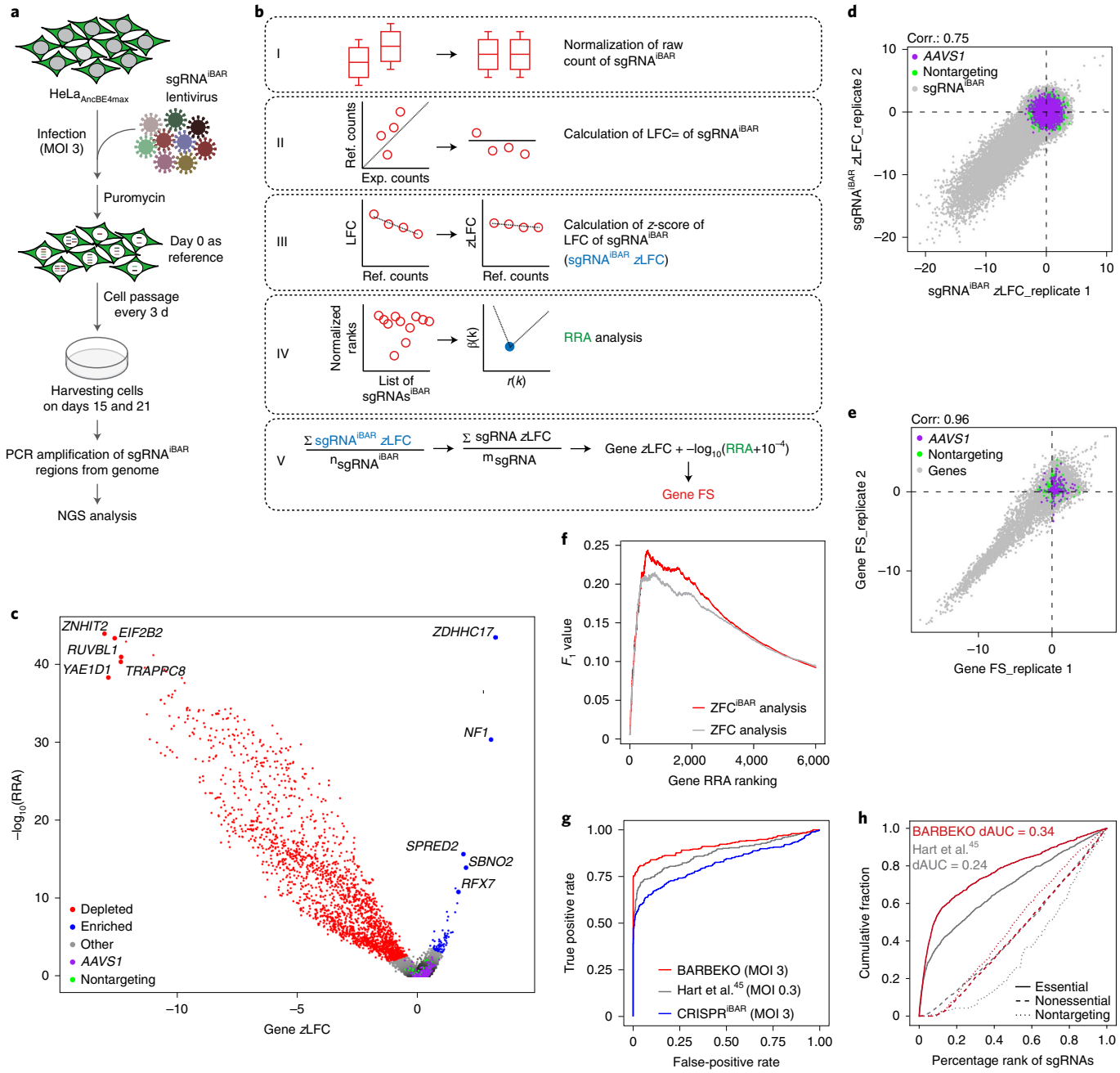


Fig. 2 | BARBEKO achieves high-MOI fitness screens in HeLa cells. a, Workflow of fitness screen in HeLa cells. AncBE4max-expressing HeLa cells were infected by the lentivirus library of BARBEKO at an MOI of ~3, and then reference cells denoted by day 0 were harvested 5 d post-infection, and experimental groups were harvested on days 15 and 21. **b**, The schematics of the ZFC^{iBAR} algorithm describing the analytical processes of NGS data from screening. The gene FS is an integrated index of *z*-score of zLFC and the value of $\log_{10}$(RRA). Ref., reference; Exp., experimental. **c**, A volcano plot showing the overall outcome of the fitness screen in HeLa cells by BARBEKO analyzed by ZFC^{iBAR}. Depleted and enriched genes are plotted in red and blue. respectively, and the top five genes in both directions are labeled individually. Every three sgRNAs targeting the adeno-associated virus integration site 1 (*AAVS1*) and nontargeting sgRNAs were randomly grouped as controls, plotted in purple and green. **d**, Scatter plot of sgRNA^{iBAR} zLFC of two biological replicates. Pearson's correlation coefficient (Corr.) is indicated at the top. The sgRNAs^{iBAR} targeting the *AAVS1* locus and nontargeting sgRNAs^{iBAR} as negative control are labeled in purple and green, respectively. **e**, The scatter plot of gene FS of two biological replicates, Pearson's correlation coefficient is indicated at the top. **f**, Comparison of the F_1 value against gene RRA ranking when considering iBARs as internal replicates (ZFC^{iBAR}) or ignoring iBARs for ZFC analysis (ZFC). The F_1 value is determined by gene gold-standard reference sets[31]. **g**, ROC analysis based on gold-standard reference sets of depleted genes for the BARBEKO library at an MOI of ~3 (considering iBARs as internal replicates in analysis), CRISPR^{iBAR} library[24] at an MOI of ~3 (ignoring iBARs in analysis) and TKOv1 library at an MOI of ~0.3 screened in HeLa cells analyzed by BAGEL from Hart et al.[45]. **h**, Comparison of AUCs for essential (solid line), nonessential (dashed line) and nontargeting (dotted line) sgRNAs between BARBEKO at an MOI of ~3 and TKOv1 at an MOI of ~0.3 screened in HeLa cells. The dAUC values from essential and nonessential sgRNAs are indicated in the upper left corner.

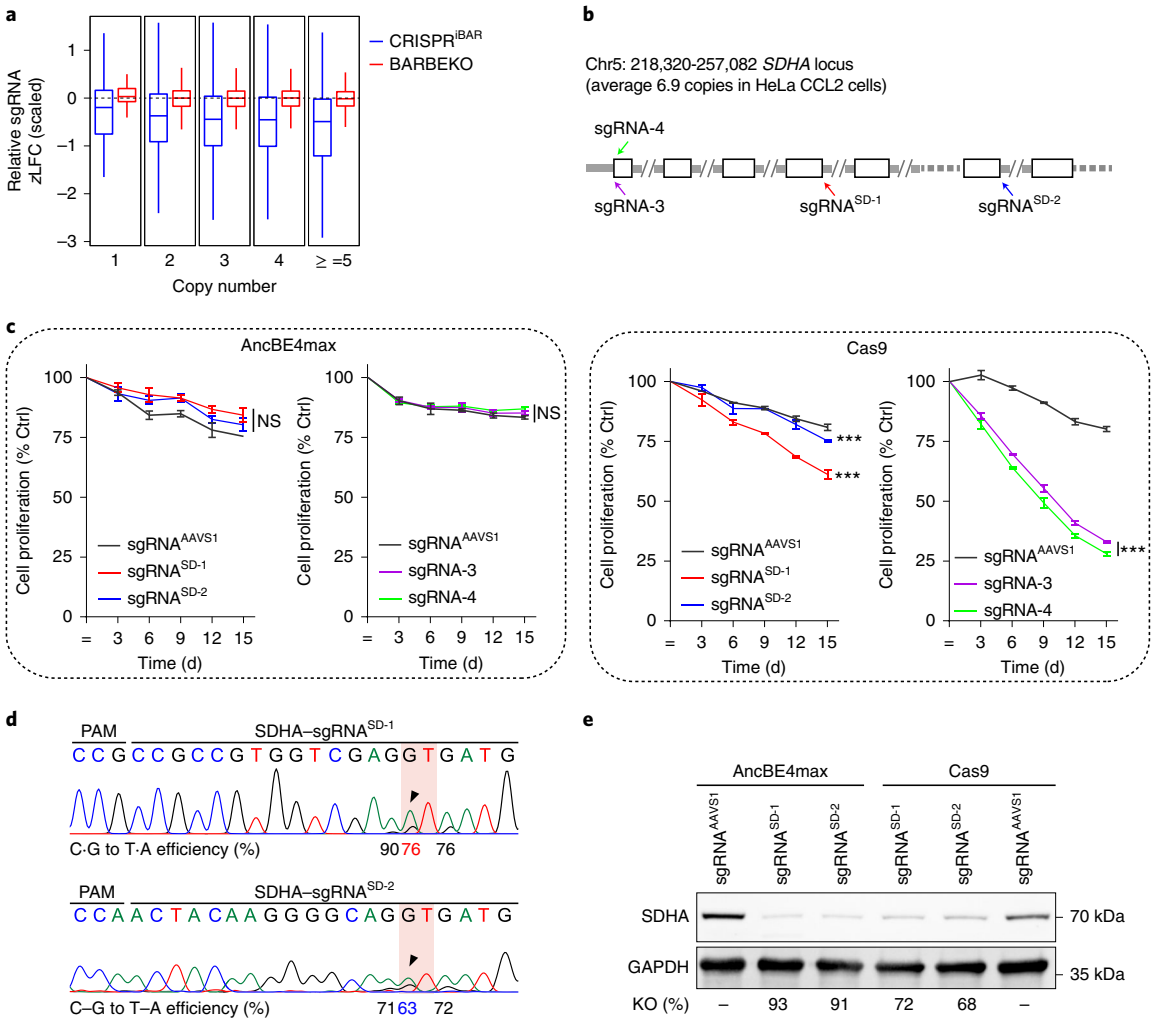


Fig. 3 | Copy-number effect could be diminished in BARBEKO screens. a, Boxplot diagram showing relative sgRNA zLFC of BARBEKO and CRISPRiBAR screens at an MOI of ~3 according to gene copy number in HeLa cells. The zLFC of sgRNAs targeting protein-coding genes subtracted the median zLFC of nontargeting sgRNAs serving as relative sgRNA zLFC. Boxplots are represented as follows: center line indicating the median, box limits indicating the upper and lower quartiles, and whiskers indicating 1.5× the interquartile range. The numbers of sgRNAs targeting loci with a different copy number (CN) separately in the CRISPRiBAR and the BARBEKO library are: CN = 1, 91 and 81; CN = 2, 27,134 and 25,433; CN = 3, 24,245 and 23,065; CN = 4, 1,856 and 1,730; and CN ≥ 5, 433 and 425. **b**, Schematic showing genomic region of a highly amplified gene *SDHA* and the targeting sites of sgRNAs selected from BARBEKO (sgRNA^{SD-1} and sgRNA^{SD-2}) or TKO (sgRNA-3 and sgRNA-4) libraries. **c**, Effects of indicated sgRNAs targeting *SDHA* on cell proliferation in HeLa cells. Four sgRNAs were individually delivered into AncBE4max- and Cas9-expressing cells for validation, in which sgRNA^{SD-1} and sgRNA^{SD-2} targeted splice donor sites of *SDHA*. The sgRNA^{AAVS1} served as a negative control (Ctrl). Data are presented as the mean ± s.d. of three independent experiments. *P* values represent comparisons with sgRNA^{AAVS1} at the end point (day 15), calculated using a one-tailed Student's *t*-test and adjusted using the Benjamini–Hochberg method, ***$P < 0.001$. NS, not significant. **d**, The Sanger sequencing chromatograms of sgRNA^{SD-1} and sgRNA^{SD-2} targeting splice donor sites of the *SDHA* genomic region after AncBE4max editing. The orange arrows indicate peaks of targeted 'Gs' in splice donor sites. **e**, Western blot analysis shows the abundance of SDHA protein of AncBE4max- or Cas9-edited cells with the indicated sgRNAs. KO efficiency was calculated based on the protein level of the sgRNA^{AAVS1} group. These experiments were repeated independently with similar results. GAPDH, glyceraldehyde 3-phosphate dehydrogenase; SDHA, succinate dehydrogenase complex flavoprotein subunit A.

coefficient of two results from days 15 and 21 was high (0.98); however, the number of depleted genes on day 21 was larger than that on day 15 (2,121 versus 1,795) under the same threshold (Extended Data Fig. 4d,f). These results suggested that a longer duration improved the sensitivity of fitness screens, in agreement with a prior report[33]. Gene ontology (GO) enrichment analysis indicated that 352 genes identified only in the later timepoint (Extended Data Fig. 4g) mainly belonged to the same GO terms of commonly selected genes of both timepoints (Extended Data Fig. 4h), demonstrating the consistency in the process of screening using the BARBEKO strategy to reveal gene functions.

Efficiency comparison among different types of sgRNAs. As sgRNAs targeting the gold-standard essential genes are supposed to be depleted in the screen, we categorized these sgRNAs according to the targeting types for efficiency comparison. The sgRNA$^{SD/SA}$ showed similar *z*LFC distribution to sgRNA^{Stop}, whereas sgRNA^{Start} performed a bit less effectively, presumably due to the presence of alternative translation initiation sites for many targeted genes (Extended Data Fig. 5a). In addition, the efficiency of sgRNA^{SD} was statistically lower than that of sgRNA^{SA} (Extended Data Fig. 5b), probably due to the context preference of the deaminase domain of rat APOBEC1 (ref. [34]). Indeed, we found that the 5′-guanine adjacent

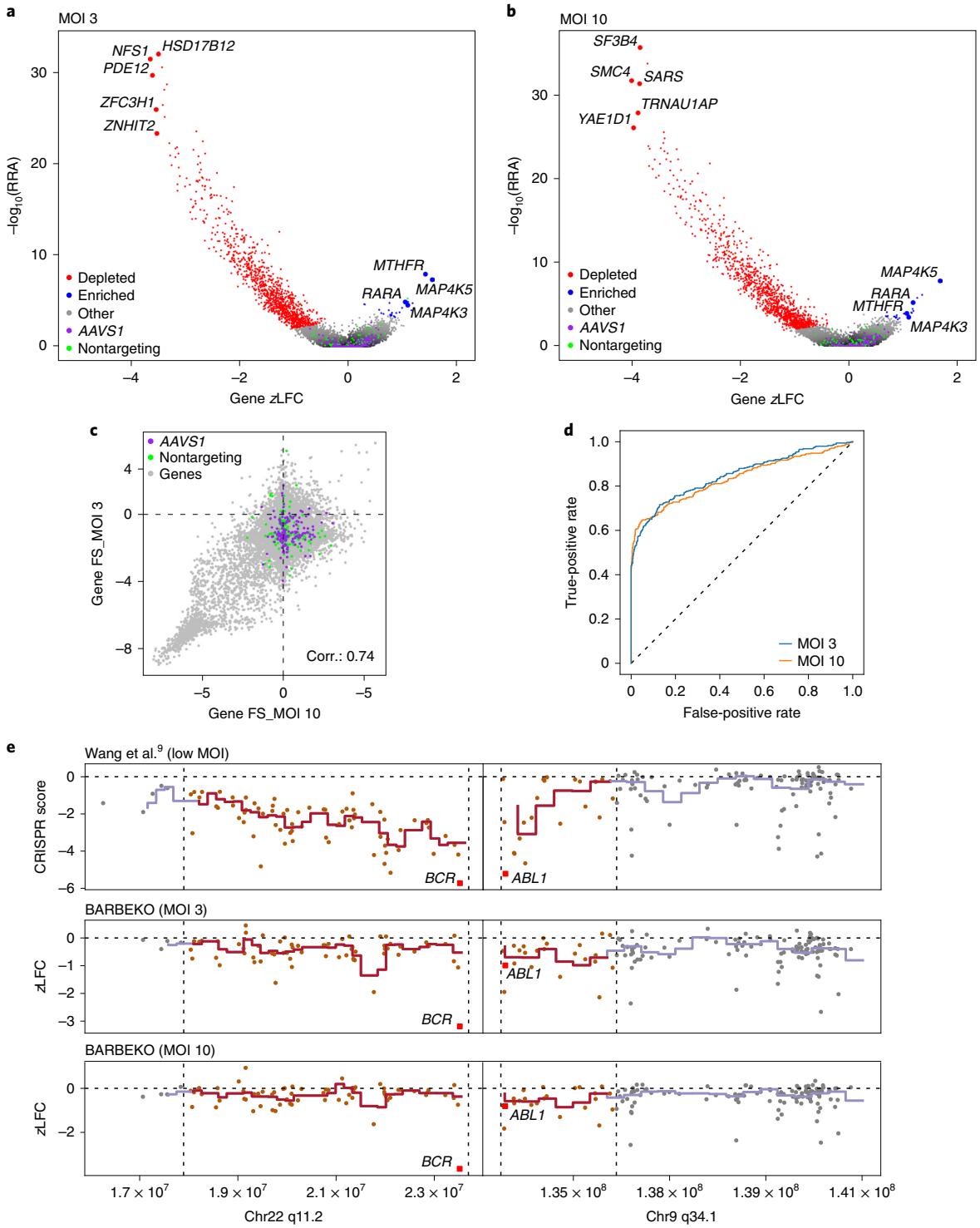


Fig. 4 | BARBEKO empowers fitness screen in K562 cells at extra-high MOIs. a,**b**, Volcano plot showing overall outcomes of fitness screens in K562 cells by BARBEKO at MOIs of ~3 (**a**) and ~10 (**b**). Depleted and enriched genes are plotted in red and blue, respectively. The top five depleted genes and commonly enriched genes of both screens are labeled individually. Every three sgRNAs targeting AAVS1 and nontargeting sgRNAs were randomly grouped as 'negative control genes', plotted in purple and green, respectively. **c**, Scatter plot of gene FS in screens at MOIs of ~3 and ~10; Pearson's correlation coefficient (Corr.) is indicated at the top. **d**, ROC analysis of depleted genes in screens at MOIs of ~3 and ~10 according to essential genes and nonessential genes of gold-standard reference sets. **e**, The zLFC or CRISPR scores of genes locating around the *BCR–ABL* fusion gene are plotted sequentially. Data from Wang et al.[9] are shown in the top lane, and the results of the BARBEKO at MOIs of ~3 and ~10 in K562 are shown in the middle and bottom lanes. Genes in this region are separated into 200 bins to calculate the mean of the zLFC or the CRISPR score, which is represented by a solid line. The high-copy tandem-amplified region is indicated in red and the flanking regions in gray.

to the targeting cytosine substantially compromised the editing efficiency (Extended Data Fig. 5c). As expected, sgRNA efficiency was influenced by the location of targeted 'C' in the editing window as well (Extended Data Fig. 5d). The sgRNA^{Stop} targeting different codons also showed distinct *z*LFC distributions (Extended Data Fig. 5e), in which targeting the codon 'TGG' had the highest gene KO efficiency. We infer that the anticodon sequence 'CCA' of the DNA strand is more likely to be edited by the CBE. In conclusion, the above-summarized rules would help to design sgRNAs for effective gene KOs by CBEs.

Copy-number effect could be diminished in BARBEKO screens. A number of reports suggested that Cas9-mediated DNA cleavage in amplified genomic regions induced a gene-independent, anti-proliferation effect and consequently introduced false positives into gene essentiality screens[12,15,35]. To verify whether BARBEKO could avert such a problem, we compared sgRNA *z*LFC distribution across gene copy numbers of BARBEKO and CRISPR^{iBAR} screens in HeLa cells. The *z*LFC of sgRNAs descending in targeting genomic sites correlated with the increased copy numbers in CRISPR^{iBAR} screens, evidently resulting from DSB-induced cytotoxicity (Fig. 3a). In contrast, the BARBEKO screen was not affected by copy-number amplification. To confirm this, we selected two genes that are located in amplified genomic regions in HeLa cells, *SDHA* and *TRIP13* (ref. [36]). Four *SDHA*-targeting sgRNAs (Fig. 3b) were tested individually in both AncBE4max- and Cas9-expressing cells. No noticeable phenotypic changes were observed in AncBE4max-edited cells, whereas cell viability was significantly decreased when these loci were perturbed by Cas9 with all four sgRNAs (Fig. 3c). Sanger sequencing and western blot analysis further confirmed that two sgRNAs were effective in generating *SDHA* KOs with AncBE4max or Cas9 (Fig. 3d,e, Extended Data Fig. 6a and Source Data Fig. 1), indicating that the decreased cell viability in Cas9 cells was not due to the gene KOs but to the occurrence of multiple DSBs. Similar results were obtained for *TRIP13* gene targeting: three out of four sgRNAs led to decreased cell viability only in Cas9-expressing cells (Extended Data Fig. 6b,c).

BARBEKO empowers screens in K562 cells at ultra-high MOIs. As library construction with a high MOI could significantly reduce the starting cells, we then pushed the MOI to about 10 and tested it in K562 cells. K562 cells contain a Philadelphia chromosome susceptible to single sgRNA-mediated Cas9 cutting; thus, it enables us to examine the potential cytotoxic effect of multiple sgRNAs in the BARBEKO screens with ultra-high MOIs. K562 libraries were then made with lentiviral infection at MOIs of 3 and 10 in parallel (Fig. 4a,b and Source Data Fig. 3). A scatter plot of gene FS showed compatible hits in both depletion and enrichment after screening (Fig. 4c), and the ROC analysis showed comparable AUC scores according to the gold-standard gene reference sets (Fig. 4d). These results demonstrated that BARBEKO is a robust strategy that produces highly consistent results even on cell libraries constructed with lentiviral infection at extremely high MOIs, resulting in much-improved cost and labor effectiveness for both positive and negative selection screens. Specifically, to reach 1,000-fold coverage per sgRNA, the minimal requirement for a conventional CRISPR library construction at an MOI of 0.3 for 2 experimental repeats is 3.6×10^8 cells, whereas the number drops to 5.4×10^6 for the BARBEKO library (4 iBARs per sgRNA serving as internal repeats) at an MOI of 10, a reduction of over 60-fold. Putting economy aside, this astonishing reduction in cell numbers could be pivotal in large-scale screens when either the source of agents is limited, such as emerging viruses or uncommon toxins, or the screening material is scarce, such as patient-derived cells.

To further confirm that the BARBEKO approach is immune to Cas9-cleavage-induced cytotoxicity, we chose to test the *BCR–ABL* oncogene because this locus suffers from a high-copy tandem amplification during Philadelphia translocation in K562 cells[37]. Cas9 cleavage in this repeated region has been reported to cause false positives of essential genes[16]. We plotted the *z*LFC of genes located surrounding the fusion gene and compared them with the data from Wang et al.[9] (Fig. 4e). Indeed, the sgRNAs targeting contiguous genes within the amplicons on 22q11.2 and 9q34.1 were significantly dropped out compared with the flanking nonamplified regions, indicating Cas9-cleavage-induced cytotoxicity (Fig. 4e, top lane). These positional effects on nonessential genes were almost completely diminished in two high-MOI screens of the BARBEKO approach, whereas the true essential oncogenic fusion gene *BCR–ABL1* could still be correctly identified (Fig. 4e, middle and bottom lanes).

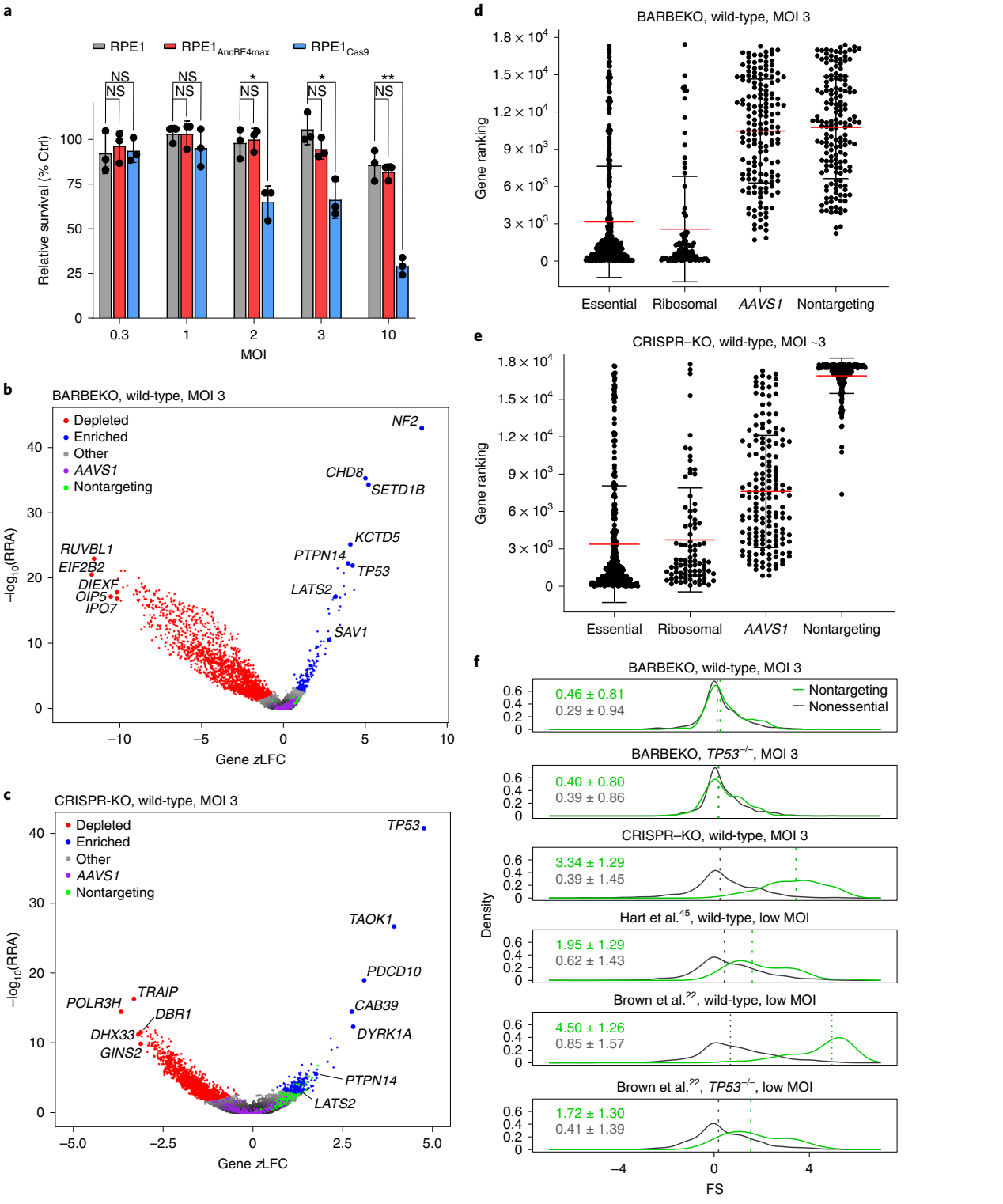
Several computational methods have been developed for conventional CRISPR–KO screens to correct false positives resulting from the copy-number effect. So, we utilized CRISPRCleanR, an unsupervised method[38], to correct the results of the CRISPR–KO screen in K562 cells for comparisons. After data processing, most high-copy-number genes were given near-zero log(fold-change) (LFC) scores, all of which were closer to the value in the BARBEKO screen without correction (Supplementary Fig. 4). These comparisons demonstrated the advantage of the BARBEKO approach in reducing the false-positive rate due to the copy-number effect. Thus, BARBEKO offers a clear advantage without the need for computational correction, which is particularly useful for screens conducted in cells lacking copy-number information. We anticipate that these advantages of BARBEKO are worth being exploited to the full for critical applications that are sensitive to the copy-number effect.

BARBEKO enables precise screens in nontransformed cells. To understand gene function in relative physiological settings, one often needs to conduct CRISPR screens in primary cells or nontransformed cells carrying intact and normal cellular machinery, such as the p53 pathway. However, it is currently under heated

Fig. 5 | Comparing BARBEKO with CRISPR–KO screens in RPE1 cells. a, Clonogenic survival of RPE1 cells in response to sgRNA library transduction at gradient MOIs. Nontargeting control (Ctrl) library (1,000 sgRNAs) and nonessential, gene-targeting experimental library (869 sgRNAs) were transduced to wild-type, AncBE4max- and Cas9-expressing RPE1 cells at MOIs of 0.3, 1, 2, 3 and 10, and three independent samples of each condition were used for clonogenic assay 3 d post-infection. The survival fraction (SF) of the experimental group was normalized by control SF to calculate the relative percentage. Data are presented as the mean ± s.d., and *P* values are calculated using a one-tailed Student's *t*-test and adjusted using the Benjamini–Hochberg method: $^{**}P < 0.01$; $^{***}P < 0.001$. **b**,**c**, Volcano plots showing the overall outcome of fitness screen in wild-type RPE1 cells by the BARBEKO (**b**) and CRISRP-KO (**c**) method at an MOI of ~3. The top five depleted and enriched genes, together with top-ranking Hippo genes, are labeled individually. **d**,**e**, Scatter plots showing the distribution of gene rankings of four different categories. Gene rankings of BARBEKO (**d**) and CRISPR-KO (**e**) screens are calculated according to the gene FS from small to large. Essential genes and ribosomal genes are extracted from reference gene sets, whereas nontargeting and *AAVS1* controls are composed of three corresponding sgRNAs through random sampling. Data are presented as mean ± s.d., and the mean value of gene rankings of each categories is highlighted in red. **f**, Comparisons of density distribution of gene FS between nontargeting controls (green curves) and nonessential genes (gray curves). The mean ± s.d. of each distribution is indicated at the left. The vertical dashed lines represent the median of each distribution. Data from Hart et al.[45] and Brown et al.[22] were reanalyzed by ZFC algorithm, and their sgRNAs targeting EGFP, LacZ and luciferase were considered to be nontargeting to the human genome.

debate whether such cells are feasible for conventional CRISPR–KO screens because Cas9 cutting-induced DDR could trigger the activation of the p53 pathway, which arrests cell growth and thereby confounds screen outcomes[18–22,39,40]. We decided to directly compare BARBEKO and CRISPR–KO screens in RPE1 cells, a model system for normal cell screens[41–43], which were derived from primary RPE cells (RPE-340) and immortalized by expressing hTERT[44].

To test the feasibility of high-MOI transduction in RPE1 cells, we first constructed two sublibraries, a control library containing 1,000 nontargeting sgRNAs (Supplementary Table) and an experimental library containing 869 sgRNAs targeting nonessential genes[10]. With the confirmation of the editing efficiency of AncBE4max in RPE1 cells (Supplementary Fig. 5), we separately delivered these two libraries into wild-type, AncBE4max- and Cas9-expressing RPE1 cells at increasing MOIs. Clonogenic survival assays were performed to monitor cell viability (Supplementary Fig. 6). Comparing with wild-type RPE1 cells, AncBE4max-expressing cells held a similar survival fraction at all levels of the MOI up to 10, whereas a significantly diminished clonal formation ratio was observed in Cas9-expressing cells infected at high MOIs (Fig. 5a). Collectively,



these results indicate that BARBEKO can be applied to fitness screens in RPE1 cells at high MOIs.

Promoted by these results, we performed genome-wide BARBEKO and CRISPR–KO screens in RPE1 cells at an MOI of 3 for a head-to-head comparison. After data processing with the ZFC^{iBAR} algorithm, fitness genes were bidirectionally selected under the same thresholds of gene FS > 4 and <−3 (Fig. 5b,c and Source Data Figs. 3 and 6). We then compared the distribution of rankings of gold-standard essential genes and ribosomal genes together with negative controls composed of *AAVS1* and nontargeting sgRNAs using random sampling. In the BARBEKO screen, most gold-standard essential and ribosomal genes were top ranked and distinct from controls (Fig. 5d). In contrast, the difference between ribosomal/essential genes and *AAVS1* controls was decreased in the CRISPR–KO screen (Fig. 5e). In ROC analysis, the AUC scores of BARBEKO were all evidently higher than the CRISPR–KO screen based on three gold-standard gene reference sets (Extended Data Fig. 7a–c). Further comparisons with additional low-MOI CRISPR–KO screens from publications[22,45,46], using boxplots based on the gold-standard reference sets (Extended Data Fig. 7d) or five essential gene categories from GO datasets (Extended Data Fig. 7e), revealed that BARBEKO screening showed improved signal-to-noise ratios in the identification of true essential genes.

For this head-to-head comparison, another notable difference was the evident enrichment of nontargeting sgRNAs in the CRISPR–KO screen, indicating that cells without sgRNA-mediated DSBs have growth advantages (Fig. 5e). Consistently, such distribution of nontargeting sgRNAs was also observed in conventional CRISPR–KO screens conducted at low MOIs (Fig. 5f). These results indicate that Cas9-mediated DSBs imposed impairment on cell fitness of RPE1 and, consequently, cells containing nontargeting sgRNAs grew out of those carrying lesions by gene-targeting sgRNAs. In sharp contrast, such phenotypes were not observed in BARBEKO screens (Fig. 5d,f).

In addition, the distribution of nonessential genes of the BARBEKO screens was more concentrated than that of the CRISPR–KO screens, which contained evidently larger s.d.s (Fig. 5f and Extended Data Fig. 7d). These results suggest that Cas9-mediated DSBs might randomly trigger a wide variety of genetic alterations, including deletions and translocations[17], which affects neighboring genes and results in guide-independent perturbations in cell fitness. Eventually, these nonspecific perturbations might lead to the increased variance of gene FS in CRISPR–KO screens, but not in BARBEKO ones because CBE editing caused little impact on neighboring genes[47].

BARBEKO outperforms CRISPR–KO screens in positive selection. By analyzing positively selected genes from the CRISPR–KO screens, we found that negative controls composed of nontargeting sgRNAs accounted for about 20% of total hits under the same threshold as the BAREBKO screen (Fig. 6a). These apparent false positives were probably derived from the growth advantages over other cells harboring DNA lesions induced by Cas9 cleavage. As not all sgRNAs in the library are equally functional in any specific cell line due to different cellular contexts, such as different chromatin structure and genetic variants, we speculated that these nonfunctional sgRNAs would perform like nontargeting controls in CRISPR–KO screens and confound the identification of genuine cell fitness suppressors (Fig. 5f). By GO analysis, we found that positively selected genes from the CRISPR–KO screen were enriched in several regulatory pathways of cell fitness with marginal significance indicated by the FDR (Fig. 6b). In contrast, pathways known to modulate cell proliferation missing in the parallel CRISPR–KO screen were significantly more enriched in the BARBEKO screens (Fig. 6c), such as the mitogen-activated protein kinase (MAPK) cascade and the Hippo signaling pathway. By listing key components and regulators of the Hippo pathway[48–50], we found that genes directly (*LATS2*, *PTPN14*) or indirectly (*NF2*, *RRMD6*, *SAV1*, *MAP4K4*, *TNIK*, *TAOK1/3* and *WWC1*) activating the Hippo pathway were negative regulators of cell proliferation (Supplementary Fig. 7a), whereas YAP/TAZ, the key effectors of the Hippo pathway, were essential for cell viability. Actually, perturbations in a number of regulators of the Hippo pathway could effectively unleash cellular proliferation in RPE1 cells (Supplementary Fig. 7b).

BARBEKO is immune to false positives from DDR. Given the critical role of p53 in Cas9-induced DDR which influences the precision of CRISPR–KO screens, we applied BARBEKO to fitness screen in *TP53*$^{-/-}$ RPE1 cells to compare the effect of p53 on these two methods (Extended Data Fig. 8a and Source Data Fig. 3). Most candidates identified from BARBEKO screens in wild-type and *TP53*$^{-/-}$ RPE1 cells were concordant, as indicated by the correlation coefficients (0.78) (Fig. 6d). The ROC analysis indicated that the BARBEKO approach enabled the identification of essential genes with comparable quality in both genetic backgrounds (Extended Data Fig. 7a–c). In addition, the distribution of rankings of essential and ribosomal genes, *AAVS1* and nontargeting controls in *TP53*$^{-/-}$ RPE1 cells were similar to the results of BARBEKO in wild-type cells (Extended Data Fig. 8b). Notably, tight distributions of nontargeting and nonessential sgRNAs were also observed in the BARBEKO screen in *TP53*$^{-/-}$ RPE1 (Fig. 5f and Extended Data Fig. 7d).

By comparing positively selected genes, we found that the screen recaptured key components and regulators of the Hippo pathway in *TP53*$^{-/-}$ RPE1 cells (Supplementary Fig. 7a). We also identified that sgRNAs targeting *TP53* and *USP7* performed differently from screens in wild-type cells (Fig. 6d). As a positive control, sgRNAs targeting *TP53* were enriched only in wild-type cells. Accordingly, sgRNAs targeting *USP7*, a gene that encodes a protein-stabilizing p53 (ref. [51]), were depleted in wild-type cells but enriched in p53-deficient cells.

We further analyzed uniquely selected genes of BARBEKO and CRISRP–KO screens in wild-type RPE1 cells to evaluate the impact of p53 (Extended Data Fig. 9a–d). Unique essential candidates of the CRISPR–KO screen further concentrated on the DSB repair pathway (accession no. GO:0006302) (Extended Data Fig. 9d), suggesting that DSBs sensitized RPE1 cells to loss of genes participating in DDR. In particular, sgRNAs targeting *NHEJ1* and *LIG4*, both of which encode pivotal regulators of the nonhomologous end-joining (NHEJ) pathway[52,53], were depleted in the CRISPR–KO screen. In addition, XRCC3, a homologous recombination repair pathway regulator[54], showed essentiality only in the CRISPR–KO screen (Fig. 6d,e). As wild-type RPE1 cells are sensitive to Cas9-induced DSBs, which rely on NHEJ and homologous recombination pathways for repair, disruption of these genes reduces cell fitness and causes false-positive results of CRISPR–KO screens. Furthermore, by analyzing two pairs of conventional CRISPR–KO screens at low MOIs in wild-type and *TP53*$^{-/-}$ RPE1 cells from published articles[22,41,46] (Supplementary Fig. 8), we found that p53-dependent essentiality of *NHEJ1*, *LIG4* or *XRCC3* was pervasive in these screens.

Disruptions in the C terminus of p21 caused cell death. By comparing differently selected genes among screens in RPE1 cells, we noticed that sgRNAs targeting cyclin-dependent kinase inhibitor 1A (*CDKN1A*, encoding p21) was depleted unexpectedly (Fig. 6d); p21, transcriptionally controlled by p53, is a cyclin-dependent kinase inhibitor, with loss of function that is supposed to benefit cell proliferation. Further analysis identified one sgRNA targeting the C terminus of p21, denoted as sgRNA$^{Stop\text{-}1}$, that was dramatically depleted (Extended Data Fig. 10a,b). Based on previous reports about the effect of p21 on cell fitness, we postulated that a truncated p21 variant caused by Gln138-targeting sgRNA$^{Stop\text{-}1}$ might aggregate in the nucleus, which inhibits cyclin-dependent kinases and induces cell cycle arrest[55,56]. Other than acting as the cyclin-dependent

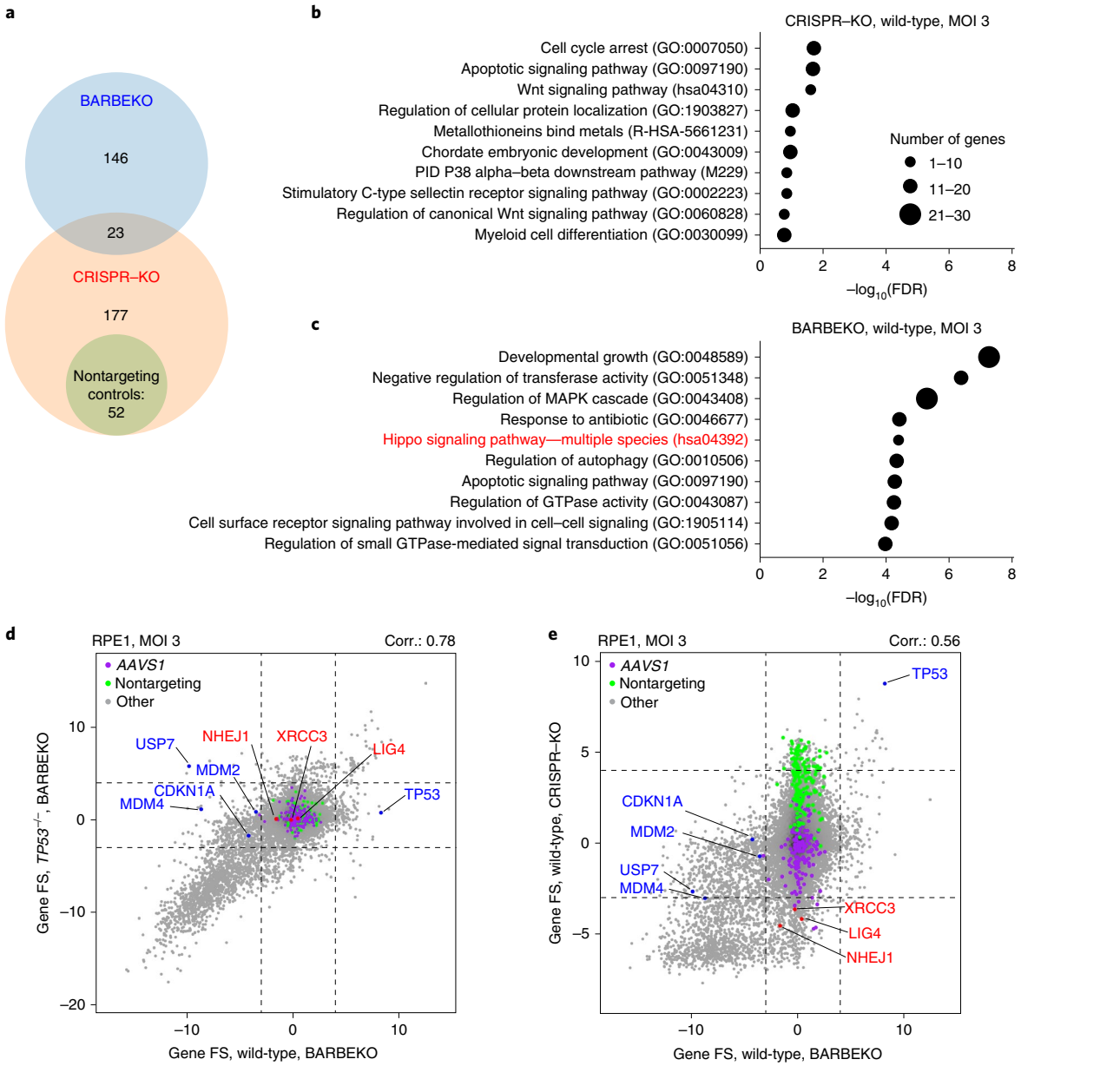


Fig. 6 | Comparisons of differently selected hits among fitness screens of RPE1. a, Venn diagram showing the numbers of commonly and differently enriched hits between BARBEKO and CRISPR-KO screens. Nontargeting controls are labeled in the green circle. **b**,**c**, GO enrichment analysis by Metascape of enriched hits from BARBEKO (**b**) and CRISPR-KO (**c**) screens. GO terms are ranked by the value of the FDR from small to large. Top-ranking terms with FDR < 0.05 are listed. The size of each circle indicates the number of genes belonging to each term. The Hippo signaling pathway is highlighted in red. **d**, Scatter plot of gene FS comparing BARBEKO screens between wild-type and *TP53*$^{-/-}$ RPE1 cells. Control genes grouped by *AAVS1* and nontargeting sgRNAs are highlighted in purple and green. Genes participating in DDR are highlighted in red and those belonging to the p53 signaling are highlighted in blue. Pearson's correlation coefficient (Corr.) is indicated at the top right corner. The dashed lines represent thresholds used in the BARBEKO screens. **e**, Scatter plot of gene FS comparing the BARBEKO screen with the CRISPR-KO screen in wild-type RPE1 cells.

kinase inhibitor, p21 has been reported to play versatile roles in multiple cellular processes, such as cell differentiation, migration, apoptosis and DDR[57]. As cellular context, subcellular localization and post-translational modifications could all change p21 activities and functions[58], we ought to pay special attention to cases such as *CDKN1A* perturbation in screens. This is apparently not unique for BARBEKO screens (Extended Data Fig. 10c).

Discussion

We developed a new approach called BARBEKO that combines CBEs and iBARed sgRNAs for high-throughput genetic screens. In comparison, BARBEKO surpasses conventional CRISPR–KO screening as follows: (1) cell number required for library construction could be significantly dropped to reach the same level of coverage; (2) iBARs serving as internal replicates improved screening quality; and (3) such loss-of-function screens are immune to copy-number effect and gene-independent cytotoxicity induced by editing tools. These make BARBEKO particularly valuable in screens for DSB-sensitive cell types, and when the screening deals with cell fitness.

The BARBEKO strategy has been applied to fitness screens of HeLa, K562 and RPE1 cells, all at high MOIs, and yielded a

comprehensive list of genes affecting, either positively or negatively, cell proliferation. As a matter of fact, negative screening is usually more technically challenging to obtain a satisfactory signal-to-noise ratio and demands a much bigger size of library than positive selection screens[59]. In addition, gene-independent cytotoxicity triggered by Cas9-mediated cleavage often muddles the results of negative selection screens related to cell fitness, because the depletion level triggered by gene loss of function is generally modest[60]. It is an alarming issue that DSB-activated p53 signaling impacts the precision of fitness screens from recent reports[18–20,22]. Besides the copy-number effect, KOs of key regulators of DSB repair pathways, such as NHEJ1, LIG4 and XRCC3, gave rise to false positives in CRISPR–KO screens in wild-type RPE1 cells. In addition, nontargeting and nonfunctional sgRNAs tend to be enriched in CRISPR–KO screens in cells sensitive to DSBs, leading to an elevated rate of false positives. Consequently, true positive hits were compromised in such screens, leading to a high false-negative rate in identifying negative regulators for cell fitness.

Besides Cas9-induced DNA damage, the lentiviral infection may cause cytotoxicity. This effect needs to be taken into consideration for BARBEKO screening with very high MOIs. In addition, the MAGeCK^{iBAR} algorithm[24] is recommended rather than the ZFC^{iBAR} for data processing of positive selection screens using the BARBEKO approach. MAGeCK^{iBAR} was customized to deal with the acute problem of sgRNA misassociation in positive selection screens at high MOIs.

During the process of our screens, several articles reported some optimized versions of CBEs with extended targeting scope via a flexible protospacer adjacent motif (PAM) or an expanded activity window[61–64], which could be helpful to CBE-based library design with improved sgRNA quality and coverage. About 1,700 genes are missing in the current version of the BARBEKO library because of the limited targeting scope of AncBE4max. Other CBE constructs with higher efficiency, fewer off-targeting in DNA and RNA level or lower DDR based on dCas9 (refs. [65–70]) could also be employed dependent on research needs.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-021-00944-1.

Received: 12 August 2020; Accepted: 6 May 2021;
Published online: 21 June 2021

References

1. Chang, N. et al. Genome editing with RNA-guided Cas9 nuclease in Zebrafish embryos. *Cell Research* **23**, 465–472 (2013).
2. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
3. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl Acad. Sci. USA* **109**, E2579–E2586 (2012).
4. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
5. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
6. Zhang, L. & Zhou, Q. CRISPR/Cas technology: a revolutionary approach for genome engineering. *Sci. China Life Sci.* **57**, 639–640 (2014).
7. Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).
8. Shalem, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
9. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR–Cas9 system. *Science* **343**, 80–84 (2014).
10. Zhou, Y. et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509**, 487–491 (2014).
11. Shen, Z. & Ou, G. CRISPR–Cas9 knockout screening for functional genomics. *Sci. China Life Sci.* **57**, 733–734 (2014).
12. Aguirre, A. J. et al. Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discov.* **6**, 914–929 (2016).
13. Fortin, J.-P. et al. Multiple-gene targeting and mismatch tolerance can confound analysis of genome-wide pooled CRISPR screens. *Genome Biol.* **20**, 21 (2019).
14. Gonçalves, E. et al. Structural rearrangements generate cell-specific, gene-independent CRISPR-Cas9 loss of fitness effects. *Genome Biol.* **20**, 27 (2019).
15. Munoz, D. M. et al. CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. *Cancer Discov.* **6**, 900–913 (2016).
16. Wang, T. et al. Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
17. Shrivastav, M., De Haro, L. P. & Nickoloff, J. A. Regulation of DNA double-strand break repair pathway choice. *Cell Res.* **18**, 134–147 (2008).
18. Bowden, A. R. et al. Parallel CRISPR–Cas9 screens clarify impacts of p53 on screen performance. *eLife* **9**, e55325 (2020).
19. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **24**, 927–930 (2018).
20. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. Reply to 'CRISPR screens are feasible in *TP53* wild-type cells'. *Mol. Syst. Biol.* **15**, e8679 (2019).
21. Ihry, R. J. et al. p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nat. Med.* **24**, 939–946 (2018).
22. Brown, K. R., Mair, B., Soste, M. & Moffat, J. CRISPR screens are feasible in *TP 53* wild-type cells. *Mol. Syst. Biol.* **15**, e71 (2019).
23. Peng, J., Zhou, Y., Zhu, S. & Wei, W. High-throughput screens in mammalian cells using the CRISPR–Cas9 system. *FEBS J.* **282**, 2089–2096 (2015).
24. Zhu, S. et al. Guide RNAs with embedded barcodes boost CRISPR-pooled screens. *Genome Biol.* **20**, 20 (2019).
25. Billon, P. et al. CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of STOP codons. *Mol. Cell* **67**, 1068–1079.e4 (2017).
26. Kuscu, C. et al. CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nat. Methods* **14**, 710–712 (2017).
27. Koblan, L. W. et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–846 (2018).
28. Bradley, K. A., Mogridge, J., Mourez, M., Collier, R. J. & Young, J. A. T. Identification of the cellular receptor for anthrax toxin. *Nature* **414**, 225–229 (2001).
29. Wei, W., Lu, Q., Chaudry, G. J., Leppla, S. H. & Cohen, S. N. The LDL receptor-related protein LRP6 mediates internalization and lethality of anthrax toxin. *Cell* **124**, 1141–1154 (2006).
30. Kolde, R., Laur, S., Adler, P. & Vilo, J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* **28**, 573–580 (2012).
31. Hart, T., Brown, K. R., Sircoulomb, F., Rottapel, R. & Moffat, J. Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. *Mol. Syst. Biol.* **10**, 733–733 (2014).
32. Sanson, K. R. et al. Optimized libraries for CRISPR–Cas9 genetic screens with multiple modalities. *Nat. Commun.* **9**, 5416 (2018).
33. Dempster, J. M. et al. Agreement between two large pan-cancer CRISPR–Cas9 gene dependency data sets. *Nat. Commun.* **10**, 1–14 (2019).
34. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
35. Meyers, R. M. et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
36. Liu, Y. et al. Multi-omic measurements of heterogeneity in HeLa cells across laboratories. *Nat. Biotechnol.* **37**, 314–322 (2019).https://doi.org/10.1038/s41587-019-0037-y
37. Wu, S. Q. et al. Extensive amplification of bcr/abl fusion genes clustered on three marker chromosomes in human leukemic cell line K-562. *Leukemia* **9**, 858–862 (1995).
38. Iorio, F. et al. Unsupervised correction of gene-independent cell responses to CRISPR–Cas9 targeting. *BMC Genomics* **19**, 604 (2018).
39. Enache, O. M. et al. Cas9 activates the p53 pathway and selects for p53-inactivating mutations. *Nat. Genet.* **52**, 748–749 (2020).
40. Geisinger, J. M. & Stearns, T. CRISPR/Cas9 treatment causes extended TP53-dependent cell cycle arrest in human cells. *Nucleic Acids Res.* **48**, 9067–9081 (2020).
41. Drainas, A. P. et al. Genome-wide screens implicate loss of cullin ring ligase 3 in persistent proliferation and genome instability in TP53-deficient cells. *Cell Rep.* **31**, 107465 (2020).
42. Noordermeer, S. M. et al. The shieldin complex mediates 53BP1-dependent DNA repair. *Nature* **560**, 117–121 (2018).

43. Olivieri, M. et al. A genetic map of the response to DNA damage in human cells. *Cell* **182**, 481–496.e21 (2020).
44. Bodnar, A. G. et al. Extension of life-span by introduction of telomerase into normal human cells. *Science* **279**, 349–352 (1998).
45. Hart, T. et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526 (2015).
46. Behan, F. M. et al. Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature* **568**, 511–516 (2019).
47. Dang, L. et al. Comparison of gene disruption induced by cytosine base editing-mediated iSTOP with CRISPR/Cas9-mediated frameshift. *Cell Prolif.* **53**, e12820 (2020).
48. Ma, S., Meng, Z., Chen, R. & Guan, K.-L. The Hippo pathway: biology and pathophysiology. *Annu. Rev. Biochem.* **88**, 577–604 (2019).
49. Park, H. W. et al. Alternative Wnt signaling activates YAP/TAZ. *Cell* **162**, 780–794 (2015).
50. Yu, F.-X., Zhao, B. & Guan, K.-L. Hippo pathway in organ size control, tissue homeostasis, and cancer. *Cell* **163**, 811–828 (2015).
51. Sheng, Y. et al. Molecular recognition of p53 and MDM2 by USP7/HAUSP. *Nat. Struct. Mol. Biol.* **13**, 285–291 (2006).
52. Pannunzio, N. R., Watanabe, G. & Lieber, M. R. Nonhomologous DNA end-joining for repair of DNA double-strand breaks. *J. Biol. Chem.* **293**, 10512–10523 (2018).
53. Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* **18**, 495–506 (2017).
54. Brenneman, M. A., Wagener, B. M., Miller, C. A., Allen, C. & Nickoloff, J. A. XRCC3 controls the fidelity of homologous recombination: roles for XRCC3 in late stages of recombination. *Mol. Cell* **10**, 387–395 (2002).
55. Rössig, L. et al. Akt-dependent phosphorylation of p21Cip1 regulates PCNA binding and proliferation of endothelial cells. *Mol. Cell. Biol.* **21**, 5644–5657 (2001).
56. Zhou, B. P. et al. Cytoplasmic localization of p21 Cip1/WAF1 by Akt-induced phosphorylation in HER-2/neu-overexpressing cells. *Nat. Cell Biol.* **3**, 245–252 (2001).
57. Kreis, N.-N., Louwen, F. & Yuan, J. The multifaceted p21 (Cip1/Waf1/CDKN1A) in cell differentiation, migration and cancer therapy. *Cancers* **11**, 1220 (2019).
58. Sikder, R. K. et al. Differential effects of clinically relevant N- versus C-terminal truncating CDKN1A mutations on cisplatin sensitivity in bladder cancer. *Mol. Cancer Res.* https://doi.org/10.1158/1541-7786.MCR-19-1200 (2020).
59. Doench, J. G. Am I ready for CRISPR? A user's guide to genetic screens. *Nat. Rev. Genet.* **19**, 67–80 (2018).
60. Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR–Cas9. *Nat. Rev. Genet.* **16**, 299–311 (2015).
61. Cheng, T.-L. et al. Expanding C–T base editing toolkit with diversified cytidine deaminases. *Nat. Commun.* **10**, 3612 (2019).
62. Huang, T. P. et al. Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-019-0134-y (2019).
63. Jiang, W. et al. BE-PLUS: a new base editing tool with broadened editing window and enhanced fidelity. *Cell Res.* **28**, 855–861 (2018).
64. Kim, Y. B. et al. Increasing the genome-targeting scope and precision of base editing with engineered Cas9–cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).
65. Gehrke, J. M. et al. An APOBEC3A–Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* **36**, 977–982 (2018).
66. Grünewald, J. et al. CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat. Biotechnol.* **37**, 1041–1048 (2019).
67. Li, X. et al. Base editing with a Cpf1–cytidine deaminase fusion. *Nat. Biotechnol.* **36**, 324–327 (2018).
68. Rees, H. A., Wilson, C., Doman, J. L. & Liu, D. R. Analysis and minimization of cellular RNA editing by DNA adenine base editors. *Sci. Adv.* **5**, eaax5717 (2019).
69. Wang, X. et al. Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. *Nat. Biotechnol.* **36**, 946–949 (2018).
70. Wang, X. et al. Cas12a base editors induce efficient and specific editing with low DNA damage response. *Cell Rep.* **31**, 107723 (2020).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

Cells and reagents. The HeLa CCL2 from Z. Jiang's laboratory (Peking University) and HEK293T cells from C. Zhang's laboratory (Peking University) were cultured in Dulbecco's modified Eagle's medium (DMEM; Gibco). K562 cells from H. Wu's laboratory (Peking University) were maintained in RPMI 1640 medium (Gibco). The hTERT RPE1 cells from Y. Sun's laboratory (Peking University) were cultured in DMEM/F12 medium (Gibco). AncBE4max or Cas9 was lentivirally delivered into cells, and single clones for screening were selected based on editing efficiency and expression levels of nCas9/Cas9 proteins. As the editing efficiency is pivotal for the quality of screens, use of freshly established cells expressing AncBE4max or Cas9 is recommended for experiments. All cell lines were supplemented with 10% fetal bovine serum (Biological Industries) and 1% penicillin–streptomycin, cultured with 5% CO_2 at 37 °C. All cells were checked to ensure that they were free of *Mycoplasma* contamination.

Cloning. The sequence of AncBE4max was obtained from the supplementary information of Koblan et al.[27] and synthesized by Synbio Technologies. The AncBE4max construct was cloned into a pLenti-P2A-mCherry vector through double restriction enzyme digestion (New England Biolabs) and T4 ligase ligation (New England Biolabs, catalog no. M0202). Individual sgRNA oligos (Supplementary Table) were synthesized using Ruibotech and cloned into pCG-2.0 sgRNA-expressing vector through Golden-Gate assembly.

Phenotypes of toxin-receptor–gene KOs by AncBE4max. The sgRNAs targeting *ANTXR1* and *HBEGF* were lentivirally infected into HeLa cells. Green fluorescent protein-positive (GFP^+) cells were FACS sorted and treated with PA/LFnDTA (70 ng ml^{-1} of PA + 50 ng ml^{-1} of LFnDTA) for 48 h or 7.5 ng ml^{-1} of diphtheria toxin (List Biological Laboratories Inc.) for 60 h, and conducted in triplicate with individual treatment. Phenotype images were acquired with an inverted wide-field fluorescence microscope (Olympus IX71) equipped with a CCD camera (CoolSnap HQ2, Photometrics). Cells were harvested and subjected to genome extraction using the DNeasy Blood and Tissue Kit (QIAGEN). Targeted fragments were PCR amplified using specific primers (Supplementary Table) by PrimerSTAR HS DNA Polymerase (TaKaRa, catalog no. R010Q). Then the PCR products of *HBEGF* and *ANTXR1* were purified using DNA Clean & Concentrator-5 (ZYMO research, catalog no. D4013).

Cell proliferation assay. Specific sgRNAs (Supplementary Table) were cloned into a lentiviral backbone carrying cytomegalovirus promoter-driven enhanced GFP (EGFP) and packaged into lentiviruses in HEK293T cells. Then sgRNA lentiviruses were delivered into AncBE4max- or Cas9-expressing cells at an efficiency within 40–60%. The percentage of $EGFP^+$ cells was quantified through flow cytometry (LSRFortessa, Becton Dickinson Inc.). The first analysis started from 2 d post-infection, denoted as day 0, serving as a baseline for normalization. Then the percentage of $EGFP^+$ cells was analyzed every 3 d, until day 15 or day 18.

Design of genome-scale gene KO sgRNA library of CBE. Gene annotations of 19,210 genes were retrieved from the UCSC hg38 genome. All possible sgRNAs with 'NGG' or 'NAG' PAMs (where N is any nucleobase) containing targeted cytosine in positions 4–8 were considered (the distal position from PAM is defined as position 1, the same below). In consideration of sgRNA on-targeting efficiency, the above sgRNAs that met one of the following descriptions were removed:

(1) perfectly matching more than one human genomic regions based on bowtie-1.2.1.1 and index 'GCA_000001405.15_GRCh38_no_alt_analysis_set'
(2) containing thymine homopolymers of length ≥4
(3) GC content <0.2 or >0.8

Then, we selected library sgRNAs from the candidate pool as follows:

$SgRNA^{Start}$: annotations of the genomic position of translational start codons were obtained from the CCDS database (CCDS.20160908 release). We selected sgRNAs targeting the cytosine of 'CAT' (the reverse complementary sequence of 'ATG') in the activity window, and ensured that there was no other in-frame 'ATG' in the top 30% of CDSs.

$SgRNA^{SD/SA}$: annotations of exon start positions and end positions were extracted from the National Center for Biotechnology Information (NCBI) RefSeq of hg38 assembly to get the genomic sequences around the splice site. We selected sgRNAs targeting the cytosine of 'CT' (reverse complementary sequence of splicing donor site) and 'AC' (reverse complementary sequence of splicing acceptor site) in the activity window.

$SgRNA^{Stop}$: $sgRNAs^{Stop}$ were introduced from the CRISPR–STOP library[26]. We mapped the sgRNA sequences to the human reference genome of hg38 assembly, because the sgRNAs were designed based on the hg19 version.

The total number of $sgRNA^{Start}$, $sgRNA^{SD/SA}$ and $sgRNA^{Stop}$ was 512,914. Then we selected three sgRNAs for each gene based on a reasonable scoring scheme (Source Data Fig. 2) for efficient and specific editing. The following situations were considered in the selection:

(1) SgRNAs with NGG PAMs are better than those with NAG.
(2) Distances between sgRNA targeting sites and translational initiation sites: the shortest transcripts of individual genes were considered as a reference, and then sgRNAs targeting beyond the shortest transcripts were defined as sgRNAs targeting UTR regions.
(3) $SgRNA^{SA}$-targeted exons contain multiples of three nucleotides. In the present study, we considered that skipping of an in-frame exon probably decreases the gene KO efficiency.
(4) A guanine locates 5′ to the targeted cytosine whereas the targeted cytosine locates in positions 4, 5 and 8 of the sgRNA sequence. The editing efficiency is affected by sequence context of the targeted cytosine[34].
(5) SgRNAs contain adenine, guanine or cytosine homopolymers of length ≥4.
(6) The number of matched positions of sgRNAs mapping to the reference genome with 1-bp mismatch based on bowtie-1.2.1.1 and index 'GCA_000001405.15_GRCh38_no_alt_analysis_set'.

For these sgRNAs targeting the same or the same type of adjacent cytosines in the genome, we preferred sgRNAs with the cytosine located in the sixth or seventh location. When high-score sgRNAs were >3 for one gene, we preferred to select sgRNAs targeting different locations.

After selection, the final sgRNA library contained 52,502 sgRNAs targeting 17,501 protein-coding genes (3 sgRNAs per gene); 500 nontargeting sgRNAs and 499 sgRNAs targeting the *AAVS1* safe harbor locus (chr19: 55113873–55117983 in the human hg38 assembly) were used as negative controls. For sgRNAs targeting the *AAVS1* locus, we designed all possible sgRNAs containing cytosines in an activity window with 'NGG' PAM, then we selected 499 sgRNAs that have more than five mismatching sites to any loci in the human reference genome.

All sgRNAs were combined with four iBARs of 'CTCGCT', 'GATGGT', 'GCACTG' and 'TCCACT', which has been validated by parallel performances in screens.

The sgRNA plasmid library construction. The sgRNA oligonucleotides were synthesized using semiconductor chip synthesis technology (Synbio Technologies). Primers (oligo-F and oligo-R) targeting the flanking sequences of oligos were used for PCR amplification of the sgRNA sequence from the oligo pool. The clean-up PCR products were cloned into the lentiviral $sgRNA^{iBAR}$ backbone using Golden-Gate assembly. Then the Golden-Gate products were electroporated into competent cells (TaKaRa, catalog no. 9028) to obtain library plasmids. The lentivirus library was produced by co-transfection of library plasmids with two viral packaging plasmids pVSVG and pR8.74 (Addgene) into HEK293T cells using the X-tremeGENE HP DNA transfection reagent (Roche).

Titration of sgRNA library lentiviruses. HeLa, K562 or RPE1 cells were seeded at an appropriate density into 6-well plates on day 0, then 0, 2, 4, 8, 16 or 32 µl of viruses was added with 8 µg ml^{-1} of polybrene on day 1. The culture medium was refreshed on day 2, and then cells were cultured for another 24 h. On day 3, cells were detached, counted and replated in duplicate at the same density as day 0. Puromycin was added into one of the 6-well plates on day 4 for a 48-h treatment. The concentration of puromycin was tested in advance to ensure that cells free of sgRNA-expressing vectors would be killed thoroughly within 48 h. Next, the ratio of infected cells was detected through viable cell counting of both puromycin-treated and -untreated groups.

The MOI calculation was based on its definition and the equation of Poisson's distribution. We calculated the volume of viruses required for infection at MOIs of 3 or 10 based on the equation of m and v whereas the cell number and virus titer were fixed.

High-throughput screens via BARBEKO. A total of 7.1×10^7 (MOI ~3) HeLa cells, and 7.1×10^7 (MOI ~3) and 4.3×10^7 (MOI ~10) K562 cells were respectively seeded on to 15-cm plates or T-175 flasks in duplicate for lentiviral transduction of sgRNA library. Then, 3 d after infection, library cells were subjected to puromycin treatment (1 µg ml^{-1} for HeLa, 3 µg ml^{-1} for K562) for 48-h selection. Then, one library size of cells (7.1×10^7 or 4.3×10^7) was harvested as the reference group and denoted as day 0, and one library size of cells was maintained and passaged every 3 d. Next, experimental groups were harvested on days 15 and 21 of HeLa screening, and on day 30 of K562 screening. Genomic DNA was extracted using the DNeasy Blood and Tissue kit (QIAGEN). All extracted genomes were used as PCR templates and the sgRNA regions were PCR amplified using Q5 HotStart High-Fidelity DNA Polymerase (New England Biolabs, catalog no. M0492) or KAPA HiFi HotStart ReadyMixPCR Kit (KAPABIOSYSTEMS, catalog no. KK2602) with 26–28 cycles of reactions using several pairs of primers (Supplementary Table). Up to 2 µg or 6 µg of genomic DNA was used as a template in one 100-µl PCR reaction with Q5 or KAPA polymerase, respectively, and the total number of PCR reactions was determined by the amount of extracted genomic DNA. Then, the PCR products were pooled together and purified with DNA Clean & Concentrator-5 (Zymo Research Corp., catalog no. D4013), followed by next-generation sequencing (NGS) analysis.

For screens in RPE1 cells, a total of 1.8×10^7 RPE1 cells for each screening was plated on to 15-cm plates and infected by lentiviral sgRNA library at an MOI of 3. The library cells were subjected to puromycin treatment (15 µg ml^{-1}) for 48-h selection. Then, 5 d post-transduction, a library size of cells (1.8×10^7) was harvested for genome extraction as reference and denoted as day 0. Another library size of cells was maintained and passaged every 3 d, and then experimental cells were harvested on day 15.

Computational analysis algorithm for screens. To analyze NGS data of screens using the BARBEKO strategy, we developed a new algorithm named ZFC^{iBAR}, which adopted the *z*LFC to evaluate change of sgRNA^{iBAR} abundance between the reference group and the experimental group.

First, raw counts of sgRNA^{iBAR} were adjusted by total-count normalization or median-ratio normalization to correct batch effects. We defined those sgRNAs^{iBAR} of count <0.05th quantile in the distribution of reference and experimental groups as small-count sgRNAs^{iBAR}. The mean count of small-count sgRNAs^{iBAR} is added to all sgRNAs^{iBAR} to deal with the impact on LFC caused by small counts in the reference group.

Second, the LFC of each sgRNA^{iBAR} was calculated as follows:

$$\text{sgRNA}^{\text{iBAR}}\ \text{LFC} = \log_2 \frac{\text{norm}C_{\text{exp}} + \text{norm}C_{\text{small}}}{\text{norm}C_{\text{ref}} + \text{norm}C_{\text{small}}}$$

where normC_{exp} and normC_{ref} were normalized counts of sgRNAs^{iBAR} of experimental and reference groups, respectively, and normC_{small} was the normalized mean count of small-count sgRNAs^{iBAR}.

Third, to calculate the s.d. of *z*-score normalization, the sgRNA^{iBAR} LFC was divided into numbers of bins according to the corresponding count in the reference group and fitted with a linear model, which was applied to calculate the LFC s.d. for all sgRNAs^{iBAR}. Inspired by Colic et al.[71], the *z*LFC was calculated as follows:

$$\text{sgRNA}^{\text{iBAR}}\ z\text{LFC} = \frac{\text{sgRNA}^{\text{iBAR}}\ \text{LFC}}{\text{LFC}_{\text{std}}}$$

where LFC$_{std}$ was the s.d. calculated from the linear model. The empirical *P* value of sgRNA^{iBAR} *z*LFC was calculated.

Fourth, the *z*LFC of sgRNAs was calculated as the mean of the *z*LFC of the corresponding sgRNAs^{iBAR}, and then the *z*LFC of genes was calculated as the mean *z*LFC of the corresponding sgRNAs.

$$\text{sgRNA}\ z\text{LFC} = \frac{\text{sgRNA}^{\text{iBAR}}\ z\text{LFC}}{n}$$

$$\text{Gene}\ z\text{LFC} = \frac{\text{sgRNA}\ z\text{LFC}}{m}$$

where *n* was the number of sgRNAs^{iBAR} belonging to a certain sgRNA and equaled 4 in the BARBEKO strategy in the present study, whereas *m* was the number of sgRNAs belonging to a certain gene and equaled 3 in the present study.

Fifth, RRA was utilized to calculate the ranking significance for a certain sgRNA or gene by ranking sgRNAs^{iBAR} in the whole library[30]. For bidirectional screens, RRA was calculated twice based on ranking of enrichment and depletion.

Finally, the gene FS was calculated based on the gene *z*LFC and RRA as follows:

$$\text{Gene FS} = \text{Gene}\ z\text{LFC} + \left[-\log_{10}\left(\text{RRA} + 10^{-4}\right)\right]$$

where the final RRA value was dependent on the plus or minus sign of gene *z*LFC.

Clonogenic survival assay. RPE1 cells were seeded on to 6-well plates (1×10^5 per well) and treated by lentiviral infection for 24 h. Then, 1 d post-treatment, negative control groups without any treatment were counted and subcultured into new 6-well plates at the density of 200 cells per well, whereas experimental groups were seeded at the same volume. Cells were cultured for an additional 9 d, then viable colonies were fixed by methanol, stained by 0.1% Crystal Violet (Solarbio, catalog no. G1062) and counted manually.

Analysis of copy-number effect. Information of absolute copy number was obtained from measurements by Liu et al.[36] and the average gene copy number of HeLa CCL2 cells was used in our analysis. The relative sgRNA *z*LFC of protein-coding genes was calculated using the original sgRNA *z*LFC, subtracting the median *z*LFC of nontargeting sgRNAs.

Western blotting. Cells were lysed by radioimmunoprecipitation assay lysis buffer (CWBIO, catalog no. CW2333) with protease inhibitor cocktail (CWBIO, catalog no. CW2200), then samples were concentrated by bicinchoninic acid protein assay (Pierce, catalog no. 23227) and prepared with sodium dodecylsulfate–polyacrylamide gel electrophoresis loading buffer (CWBIO, CW0027). Western blotting was performed following standard methods. The primary antibodies used in the present study were anti-β-tubulin (CWBIO, CW0098M), anti-SDHA (Cell Signaling Technology, catalog no. 11998) and anti-CRISPR–Cas9 (Abcam, catalog no. ab204448). In addition, goat anti-rabbit immunoglobulin G–horseradish peroxidase (IgG–HRP) (Jackson Immunoresearch, catalog no. 111035003) or goat anti-mouse IgG–HRP (Jackson Immunoresearch, catalog no. 115035003) secondary antibodies were used. The membranes were incubated with Clarity Western ECL Substrate kit (Bio-Rad, catalog no. 1705060) and imaged with Chemidoc Imaging system (Bio-Rad, catalog no. 1708370). The relative protein level was analyzed using ImageJ software.

Editing efficiency detection by Sanger sequencing. Cas9- or AncBE4max-expressing K562 cells were infected by indicated and *AAVS1* sgRNAs at an MOI of 3. Then, GFP-positive cells were FACS sorted 2 d post-infection, denoted as day 0. About 2×10^5 cells were collected from day 0 to day 6, and targeted fragments were PCR amplified using specific primers by TransTaq DNA Polymerase High Fidelity (Transgen, catalog no. AP131). The editing efficiency was detected by Sanger sequencing, comparing with controls. The Sanger sequencing results of Cas9 and AncBE4max were, respectively, analyzed by Tide (https://tide.nki.nl) or EditR (https://moriaritylab.shinyapps.io/editr_v10/) for quantification.

Statistics and reproducibility. The exact *P* values are listed as follows:

In Fig. 3c: AncBE4max-sgRNA^{SD-1}, 0.9949; AncBE4max-sgRNA^{SD-2}, 0.3150; AncBE4max-sgRNA-3, 0.9761; AncBE4max-sgRNA-4, 0.9976; Cas9-sgRNA^{SD-1}, 6.0119E-5; Cas9-sgRNA^{SD-2}, 6.0119E-5; Cas9-sgRNA-3, 4.5510E-8; and Cas9-sgRNA-4, 4.7793E-8.

In Fig. 5a: from left to right: 0.3193, 0.4285, 0.4898, 0.1522, 0.3785, 0.0047, 0.0683, 0.0035, 0.2594 and 0.0003.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The raw sequencing data of screens are available under NCBI BioProject accession no. PRJNA643641. Source data are available for this paper.

Code availability

The source code for sgRNA library design can be accessed from https://bitbucket.org/WeiLab/barbeko_sgrna_design/src/master. The ZFC algorithm has been implemented by Python 3 and can be downloaded from https://github.com/wolfsonliu/zfc.

References

71. Colic, M. et al. Identifying chemogenetic interactions from CRISPR screens with drugZ. *Genome Med.* **11**, 52 (2019).

Acknowledgements

This work was supported by funds from the National Science Foundation of China (grant nos. NSFC31930016 to W.W. and NSFC31870893 to Z.Z.), Beijing Municipal Science & Technology Commission (grant no. Z181100001318009), and the Beijing Advanced Innovation Center for Genomics at Peking University, the Peking–Tsinghua Center for Life Sciences (to W.W.). We thank the staff of the BIOPIC High-throughput Sequencing Center (Peking University), H. Lyu, L. Du and H. Yang of the National Center for Protein Sciences (Beijing) at Peking University for technical assistance, and the High-Performance Computing Platform at Peking University for NGS data analysis. We thank Y. Sun and D. Xu (Peking University) for providing hTERT RPE1 and RPE1-TP53KO-Cas9 cell lines.

Author contributions

W.W. conceived and supervised the project. W.W., P.X. and Y.L. designed the experiments. P.X., H.M., Y.X., Y.B., S.Z. and Z.C. performed the experiments. P.X., Z.L. and Y.L. analyzed experimental data. P.X. wrote the manuscript and W.W., Z.Z., H.M., Y.L., Z.L. and Z.W. revised it.

Competing interests

The authors declare no competing interests.

Additional information

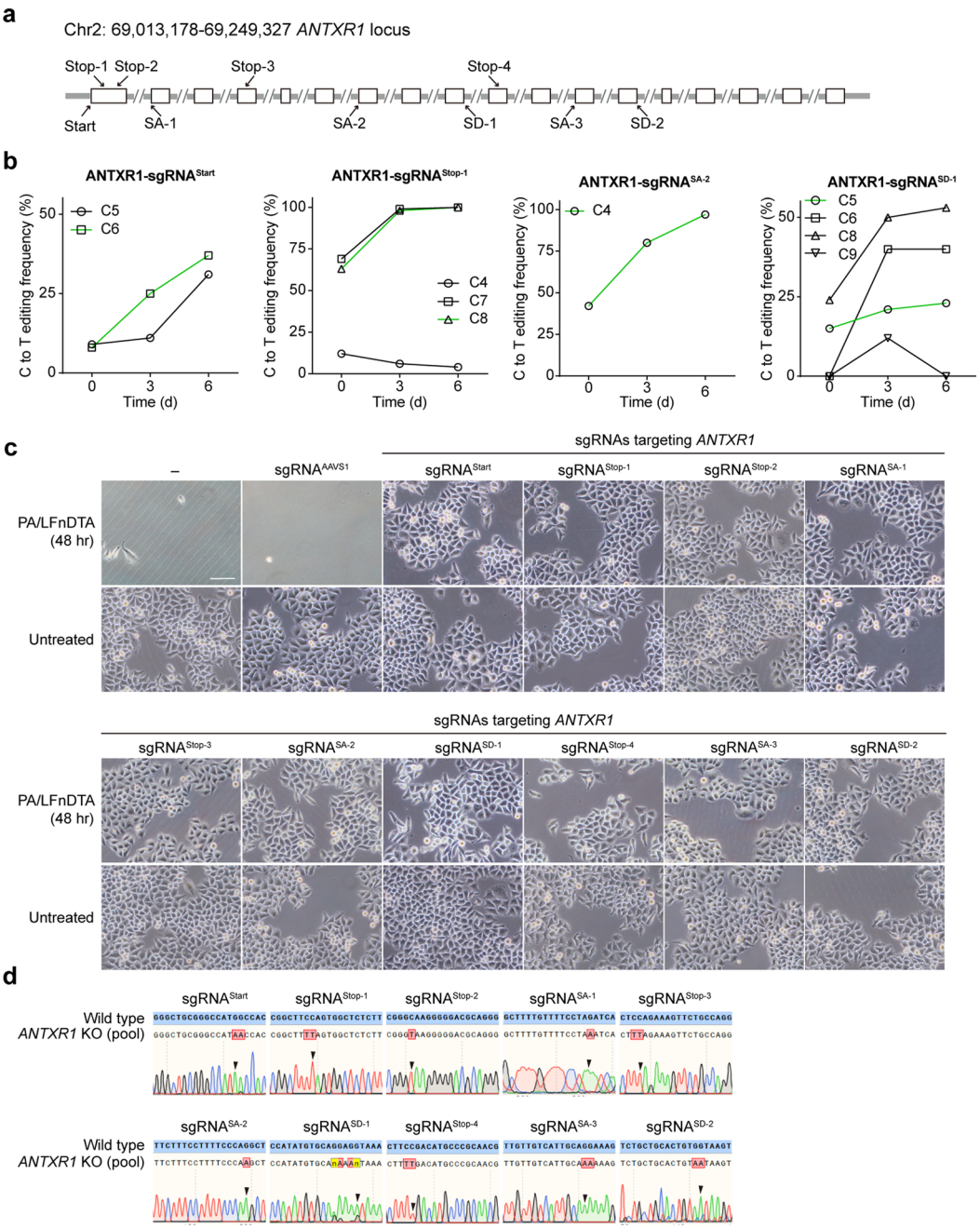
Extended data is available for this paper at https://doi.org/10.1038/s41587-021-00944-1.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-021-00944-1.

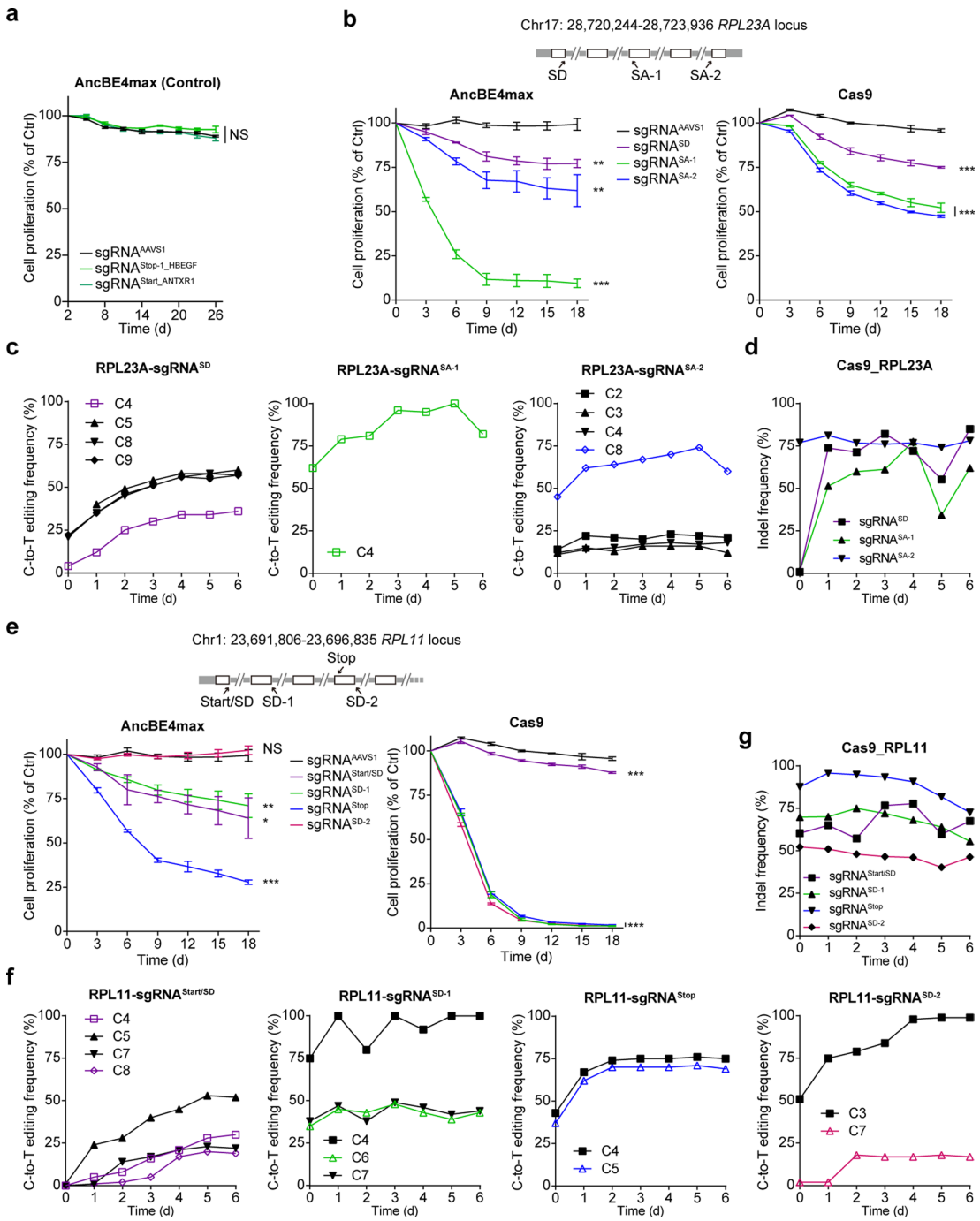
Correspondence and requests for materials should be addressed to W.W.

Peer review information *Nature Biotechnology* thanks the anonymous reviewers for their contribution to the peer review of this work.

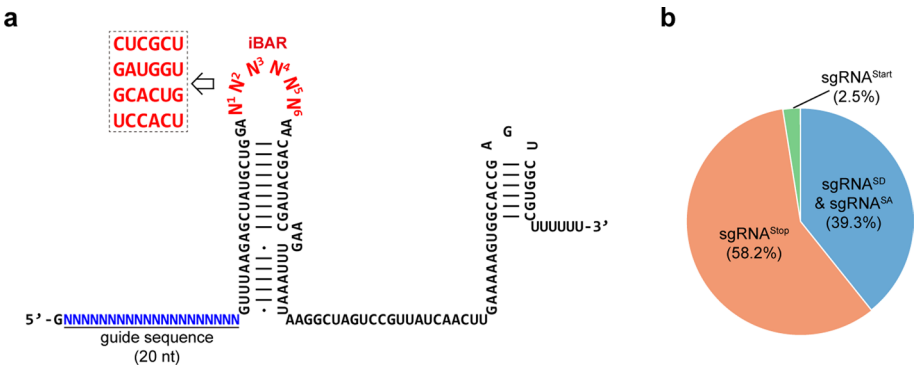
Reprints and permissions information is available at www.nature.com/reprints.



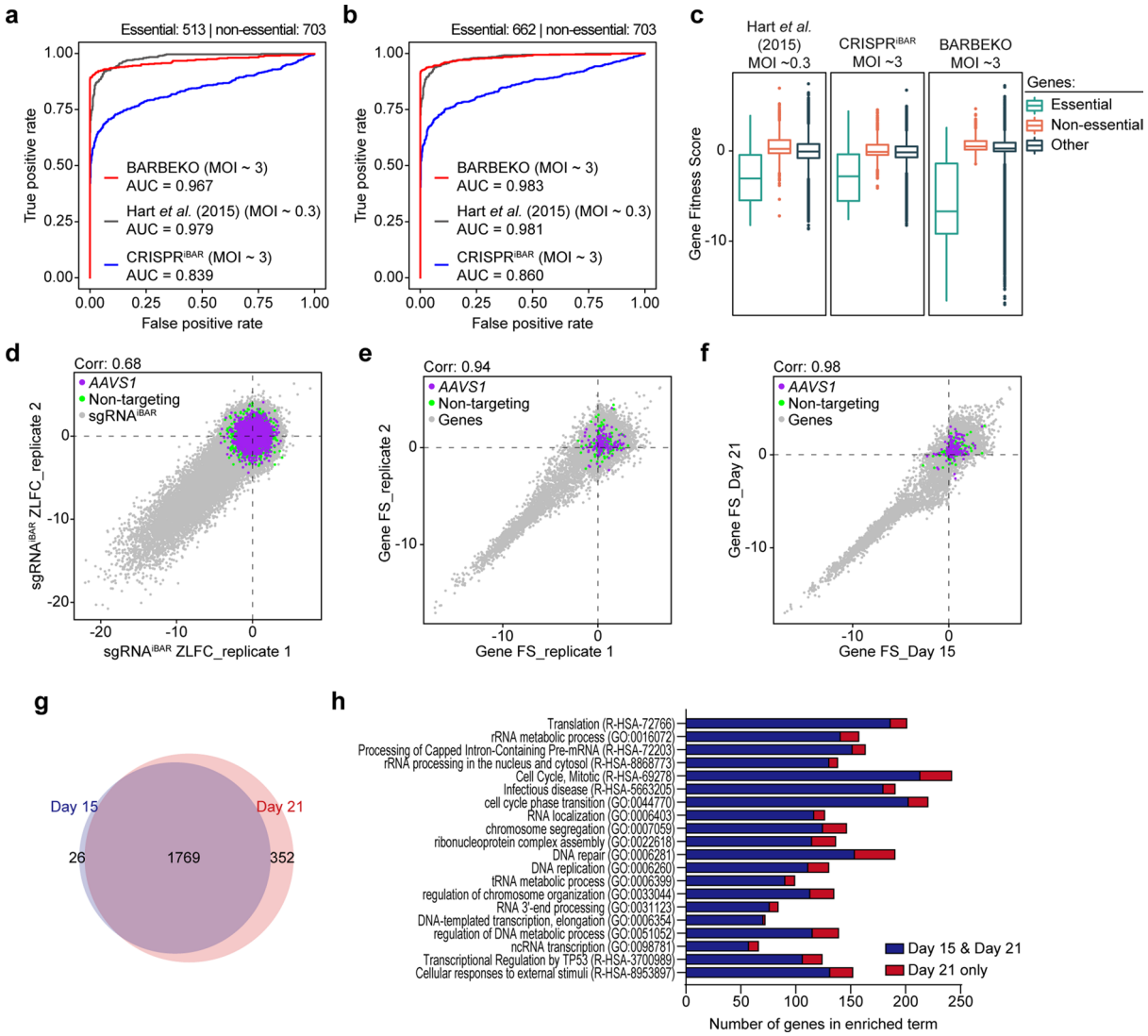
Extended Data Fig. 1 | Effect of *ANTXR1* deficiency by AncBE4max on PA/LFnDTA-triggered cytotoxicity in HeLa cells. a, Schematic indicates sgRNA targeting sites at *ANTXR1* genomic locus. **b**, Images of HeLa cells with or without PA/LFnDTA treatment for 48 hours after AncBE4max editing with indicated sgRNAs. The results shown are from one group of sgRNA transfected HeLa cells and conducted in triplicates with individual PA/LFnDTA toxin treatment. Scale bar: 100 μm. **c**, Sanger sequencing chromatograms of sgRNA-targeting *ANTXR1* genomic fragments of PA/LFnDTA toxin resistant cells, black arrows indicate peaks of targeted cytosines and their editing results. **d**, C-to-T editing frequency of indicated sgRNAs targeting *ANTXR1* in HeLa cells detected by sanger sequencing. Sorting of the sgRNA-expressing cells was conducted 2 days post-transduction (denoted as day 0), and cells were harvested on days 0, 3 and 6. The green lines indicated the editing frequency of targeted cytosine for gene knockouts, and the other blank lines indicated the editing frequency of cytosine locating in the activity windows of AncBE4max.



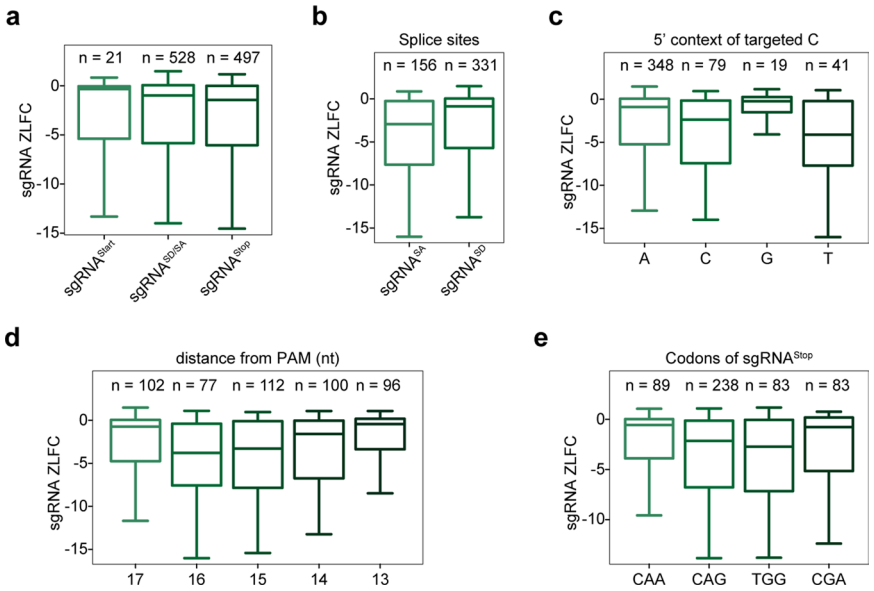
Extended Data Fig. 2 | Comparing knockout efficiency between AncBE4max and Cas9 by targeting ribosomal genes on cell proliferation. a, sgRNA^{Stop} targeting *HBEGF*, sgRNA^{Start} targeting *ANTXR1* and sgRNA^{AAVS1} served as negative controls. **b**, Effects of indicated sgRNAs targeting ribosomal gene *RPL23A* on cell proliferation in K562 cells by AncBE4max (left) and Cas9 (right). Data are presented as the mean ± s.d. of 3 independent experiments. P values represent comparisons with sgRNA^{AAVS1} at the endpoint (day 18) using a one-tailed Student's t-test and adjusted using the Benjamini–Hochberg method. $^{**}p < 0.01$; $^{***}p < 0.001$. **c**, Editing efficiency of AncBE4max with indicated sgRNAs targeting *RPL23A* detected by sanger sequencing. sgRNA-expressing cells were sorted on 2 days post transduction (denoted as day 0) and cells were harvested daily until day 6. The colored lines indicated the conversion efficiency of targeted cytosine for gene knockouts and the other blank lines indicated the conversion efficiency of cytosine locating in the activity windows of AncBE4max (the same with f). **d**, Editing efficiency of Cas9 with indicated sgRNAs targeting *RPL23A* were detected by sanger sequencing. **e**, Effects of indicated sgRNAs targeting ribosomal gene *RPL11* on cell proliferation in K562 cells by AncBE4max (left) and Cas9 (right). **f**, Editing efficiency of AncBE4max with indicated sgRNAs targeting *RPL11* detected by sanger sequencing. **g**, Editing efficiency of Cas9 with indicated sgRNAs targeting *RPL11* were detected by sanger sequencing.



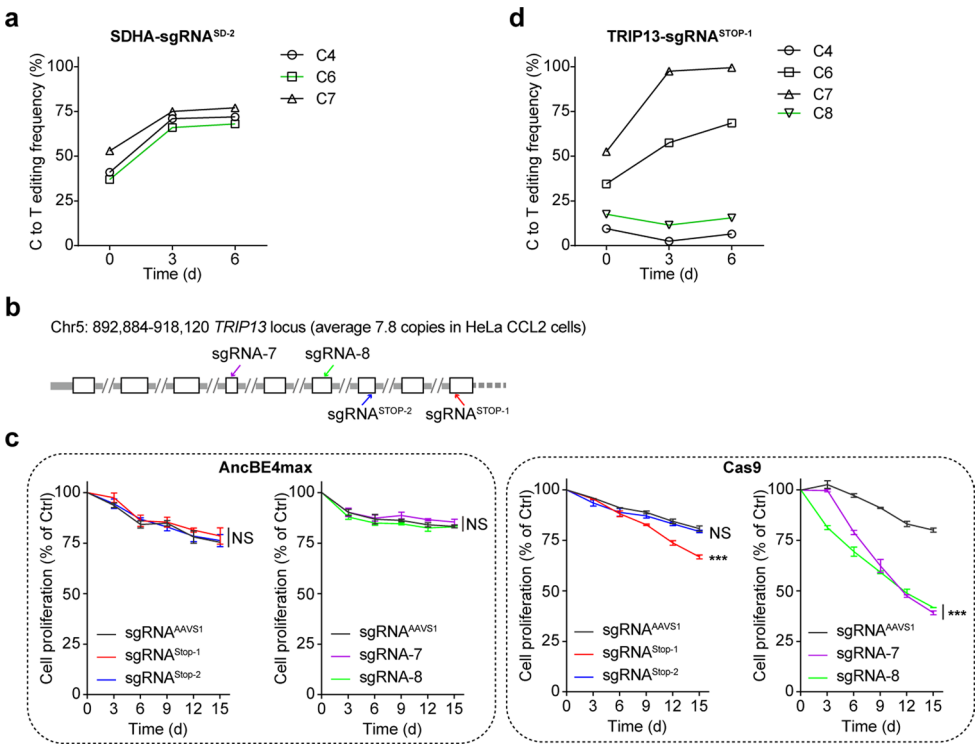
Extended Data Fig. 3 | Information of sgRNAs^{iBAR} and BARBEKO library. a, Schematic shows the scaffold sequence of sgRNA^{iBAR}, in which 4 iBARs employed in BARBEKO library are highlighted in red. **b**, Pie chart shows the composition of BARBEKO library that newly designed sgRNA^{Start} and sgRNAs targeting splice sites (sgRNA^{SD} and sgRNA^{SA}) account for 2.5% and 39.3% respectively, and sgRNA^{Stop} introduced from Kuscu *et al.* account for 58.2%.



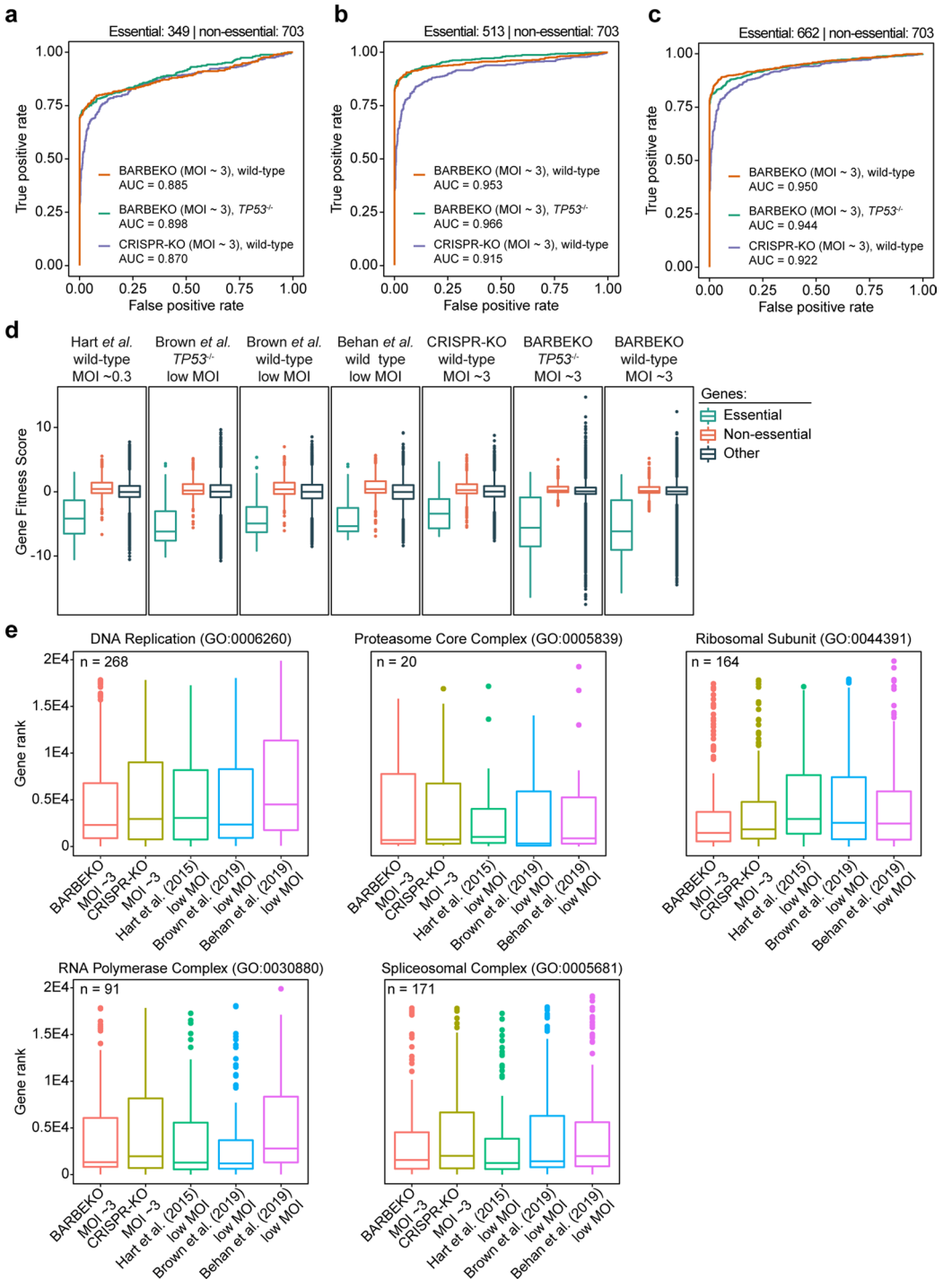
Extended Data Fig. 4 | Comparisons of BARBEKO screening with CRISPR-KO results and comparisons of depleted hits of BARBEKO between timepoints in HeLa cells. a-b, ROC analysis for screens in HeLa cells based on reference gene sets with 513 (**a**) and 662 (**b**) essential genes. **c**, Boxplots showing the distribution of gene FS of 349 essential, 703 non-essential and other genes in conventional CRISPR-KO, CRISPR^{iBAR} and BARBEKO screening. Boxplots are represented as follows: center line indicating the median, box limits indicating the upper and lower quartiles, whiskers indicating the 1.5x interquartile range and all other observed points plotting as outliers. **d**, Scatter plot of sgRNA^{iBAR} ZLFC of two biological replicates on day 15, Pearson correlation coefficient is indicated on the top. sgRNAs^{iBAR} targeting AAVS1 locus and non-targeting sgRNAs^{iBAR} as negative controls are labelled in purple and green. **e**, Scatter plot of gene Fitness Score (FS) on day 15 of two biological replicates, Pearson correlation coefficient is indicated on the top. **f**, Scatter plot of gene FS of day 15 and day 21, Pearson correlation coefficient is indicated on the top. **g**, Venn diagram shows the numbers of common and different depleted hits of day 15 and day 21. **h**, Gene Ontology (GO) analysis of common and day 21-only selected hits. GO terms are ranked from top to bottom based on P value of day 21 results using Metascape. Blue bars represent the numbers of commonly depleted hits and red bars represent the numbers of day 21-only selected hits in each GO terms.



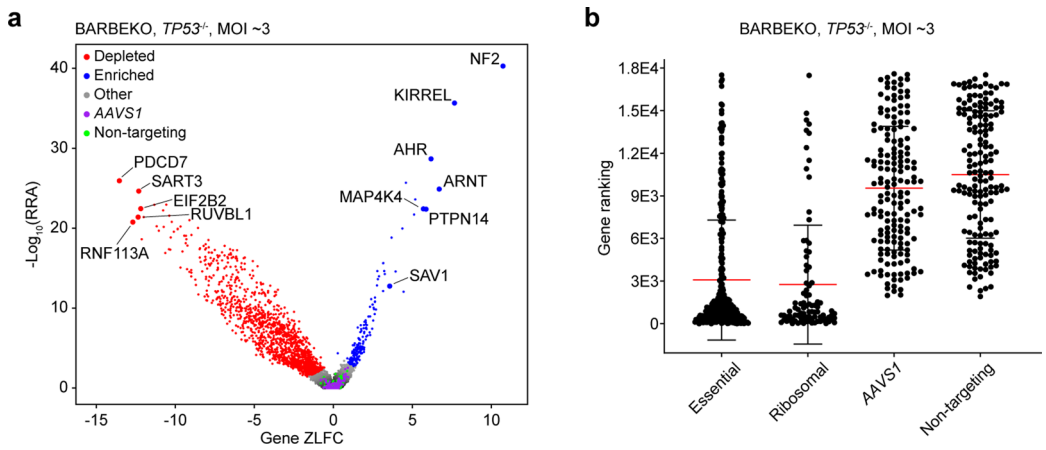
Extended Data Fig. 5 | Efficiency comparison among different types of sgRNAs. a, Efficiency comparison across 3 types of sgRNAs, sgRNA^{Start} targeting start codons, sgRNA$^{SD/SA}$ targeting splice sites and sgRNA^{Stop} targeting codons of Gln (CAA, CAG), Arg (CGA) and Trp (TGG). **b**, Efficiency comparison between sgRNA^{SA} targeting splice acceptor sites and sgRNA^{SD} targeting splice donor sites. **c**, Editing efficiency comparison across 4 types (A, C, G, T) of 5′ context of sgRNA-targeted cytosine. **d**, Editing efficiency comparison across locations of sgRNA-targeted cytosine in AncBE4max editing window. **e**, Efficiency comparison across sgRNA^{Stop} targeting CAA, CAG, TGG and CGA. Boxplots are represented as follows: center line indicating the median, box limits indicating the upper and lower quartiles and whiskers indicating the 1.5x interquartile range. The numbers of sgRNAs in each category are indicated above the corresponding boxplots.



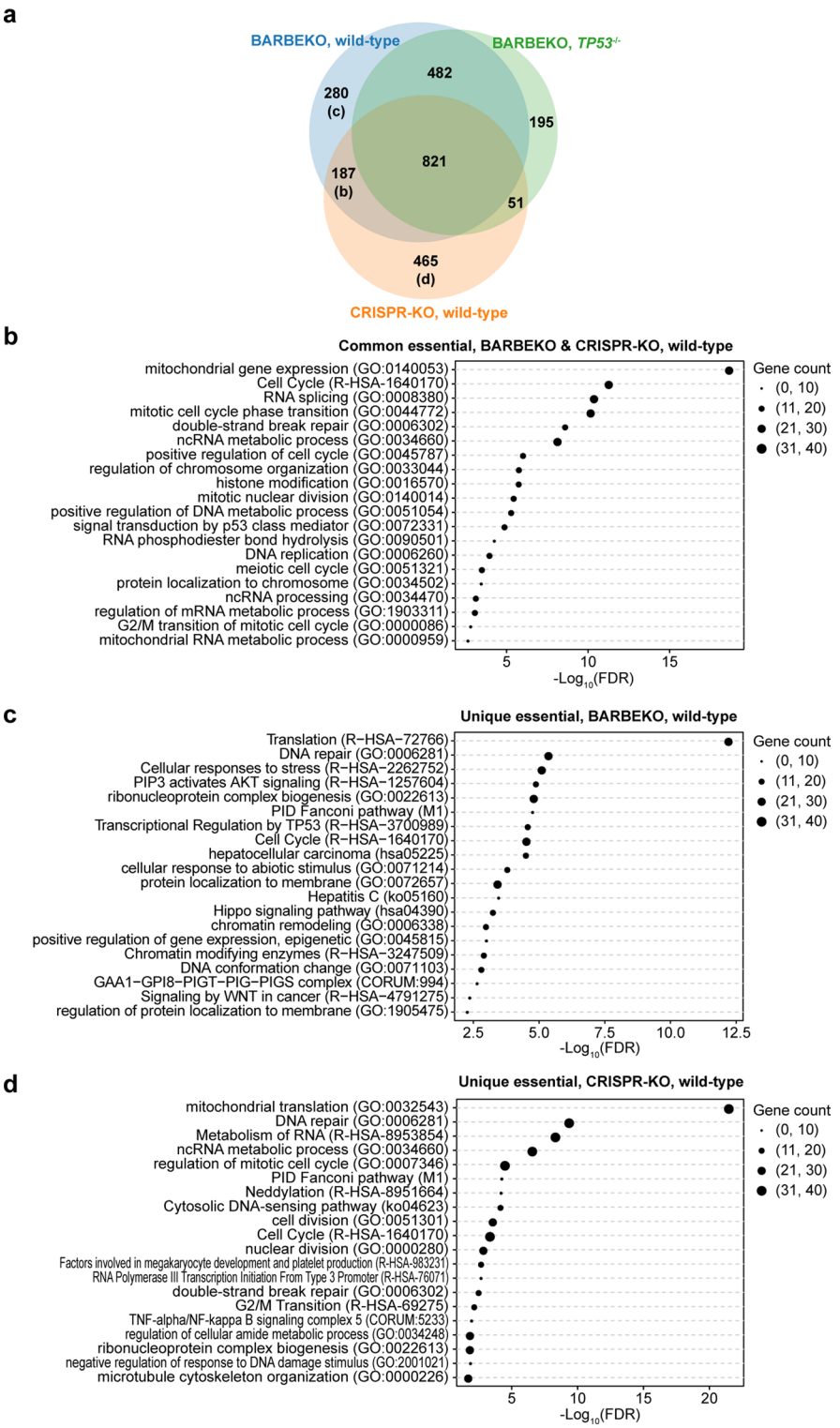
Extended Data Fig. 6 | Editing kinetics and effect on cell proliferation by AncBE4max or Cas9 targeting high-copy loci in HeLa cells. a and **d**, C-to-T editing frequency of sgRNAs targeting high-copy-number *SDHA* (**a**) and *TRIP13* (**d**) loci in HeLa cells detected by sanger sequencing. Sorting of the sgRNA-expressing cells was conducted 2 days post-transduction (denoted as day 0), and cells were harvested on days 0, 3 and 6. The green lines indicated the editing frequency of targeted cytosine for gene knockouts, and the other blank lines indicated the editing frequency of cytosine locating in the activity windows of AncBE4max. **b**, Schematic showing the genomic region of a highly amplified gene *TRIP13* and the targeting sites of sgRNAs selected from BARBEKO (sgRNA$^{Stop\text{-}1}$ and sgRNA$^{Stop\text{-}2}$) or TKO (sgRNA-7 and sgRNA-8) libraries. **c**, Effects of indicated sgRNAs targeting *TRIP13* on cell proliferation in HeLa cells. 4 sgRNAs were individually delivered into AncBE4max- and Cas9-expressing cells for validation. Data are presented as the mean ± s.d. of 3 independent experiments. sgRNA^{AAVS1} served as negative control. P values represent comparisons with sgRNA^{AAVS1} at the end point (day 15), and was calculated using a one-tailed Student's t-test and adjusted using the Benjamini–Hochberg method, $^{***}p < 0.001$.



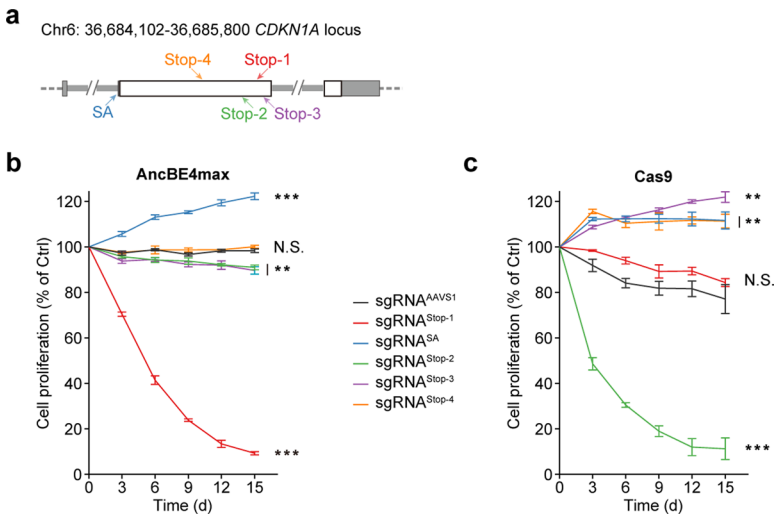
Extended Data Fig. 7 | Comparisons of BARBEKO screening in RPE1 cells with CRISPR-KO based on gold-standard reference gene sets and essential GO terms. a-**c**, ROC analysis for screens based on reference gene sets with 349 (**a**), 513 (**b**) and 662 (**c**) essential genes. **d**, Boxplots showing the distribution of gene FS of 349 essential, 703 non-essential and other genes in conventional CRISPR-KO, high-MOI CRISPR-KO or BARBEKO screening in wild-type or TP53$^{-/-}$ RPE1 cells. Boxplots are represented as follows: center line indicating the median, box limits indicating the upper and lower quartiles, whiskers indicating the 1.5x interquartile range and all other observed points plotting as outliers. **e**, Gene lists were obtained from Gene Ontology, and the numbers of genes are indicated at the top left. Boxplots are represented the same as above. Screening data from publications was re-analyzed by ZFC algorithm for comparisons.



Extended Data Fig. 8 | Fitness screen of *TP53*$^{-/-}$ RPE1 cells by BARBEKO at a high MOI. a, Volcano plot showing the overall outcomes of BARBEKO screen in *TP53*$^{-/-}$ background at a MOI of ~ 3. The top 5 depleted and enriched genes together with top-ranking Hippo genes are labelled. **b**, Scatter plot showing the distribution of gene rankings of 4 categories. Gene rankings of BARBEKO screens are calculated according to the gene FS from small to large. Essential genes and ribosomal genes are extracted from reference gene sets, while non-targeting and AAVS1 controls are composed of 3 corresponding sgRNAs by randomly sampling. The results are presented as the mean ± s.d., and the mean value of gene rankings of each categories is highlighted in red.



Extended Data Fig. 9 | Comparing the depleted hits of BARBEKO and CRISPR-KO screens in RPE1 cells. a, Venn diagram showing the numbers of commonly and differently selected hits of BARBEKO and CRISPR-KO screens. **b-d**, GO enrichment analysis by Metascape of common essential hits in wild-type cells but not in *TP53*$^{-/-}$ cells (**b**), unique essential hits of BARBEKO screen in wild-type cells (**c**) and CRISPR-KO screen in wild-type cells (**d**). GO terms are ranked by the value of FDR from small to large. The size of circle represents the number of genes belonging to each term.



Extended Data Fig. 10 | Perturbations on different sites of the *CDKN1A* locus caused variant phenotypes. a, Schematic shows genomic region of *CDKN1A* and the targeting sites of sgRNAs selected from BARBEKO library (sgRNA$^{Stop\text{-}1}$) and newly designed sgRNAs (sgRNA$^{Stop2\text{-}4}$ and sgRNA^{SD}). **b-c**, Effects of indicated sgRNAs targeting *CDKN1A* on cell proliferation in AncBE4max- (**b**) and Cas9-expressing (**c**) RPE1 cells. sgRNA^{AAVS1} served as negative control. Data are presented as the mean ± s.d. of 3 independent experiments. P values represent comparisons with sgRNA^{AAVS1} at the end point (day 15), calculated using a one-tailed Student's t-test and adjusted using the Benjamini–Hochberg method, $^{**}p < 0.01$, $^{***}p < 0.001$.

nature research

Corresponding author(s): Wensheng Wei

Last updated by author(s): Apr 29, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	R (3.6.0) script (ggplot2, dpyr, tidyr, CRISPRCleanR) and Python (3.7.1) script (biopython, pandas, numpy, matplotlib) were used for calculation and data visualization. Customized bash scripts (with GNU AWK 4.0.2) were used to run analysis and generate raw counts from FASTQ files. GraphPad Prism 6 was used for basic statistical analysis and graph production. ImageJ was used for relative protein level analysis. Metascape (http://metascape.org/gp/index.html#/main/step1) was used for gene GO enrichment analysis. ZFC (0.1.6) developed in this manuscript (https://github.com/wolfsonliu/zfc) was used to analyze gene fitness screens with or without iBARs.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

FASTQ files of raw sgRNA reads by next-generation sequencing are available in NCBI BioProject: PRJNA643641.
Fig. 2, 4, 5 and 6 have associated raw data, all available in NCBI BioProject PRJNA643641. Raw counts sgRNAs of cell fitness screens and results analyzed by ZFC are available in Supplementary Fig. 3, 4, 5 and7.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The library coverage of screens in HeLa, K562 and RPE1 cells was based on the recommended size from Joung et al. (2017).
Data exclusions	No data exclusions.
Replication	The numbers of replications were indicated in the text, methods or figure legends. All attempts at replication were successful.
Randomization	All samples from cultured cells were randomly allocated after mixing for experiments.
Blinding	Blinding was not relevant to cell fitness screening as experimental and control groups were pooled together in sgRNA library.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Primary antibodies used here were anti-β-tubulin (CWBIO, CW0098M), anti-SDHA (Cell Signaling TECHNOLOGY, 11998). And goat anti-rabbit IgG-HRP (Jackson Immunoresearch, 111035003) or goat anti-mouse IgG-HRP (Jackson Immunoresearch, 115035003) secondary antibodies were used.
Validation	All antibodies used in this study were validated by the manufacturer, and the western blot experiments were performed according to the manufacturer's instruction.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	The HeLa CCL2 from Z. Jiang's laboratory (Peking University), HEK293T cells from C. Zhang's laboratory (Peking University), K562 cells from H. Wu's laboratory (Peking University), hTERT RPE1 cells from Y. Sun's laboratory (Peking University).
Authentication	STR detection was applied to K562 cells for authentication.
Mycoplasma contamination	All cells were tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cells were infected by lentivirus containing sgRNAs with polybrene. About 48 to 72 hours later, cells were digested with trypsin and collected for the following FACS according to the fluorescence maker (EGFP).
Instrument	BD LSRFortessa
Software	BD FACSDiva and FlowJo_V10
Cell population abundance	Over 10,000 single cells with normal shape of each sample were analyzed for the percentage of GFP positive in cell proliferation assay.
Gating strategy	FSC-A and SSC-A (P1) were used to gate cells with normal shape, then SSC-W and SSC-H (P2) following FSC-W and FSC-H (P3) were used to gate single cells for further analysis.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.