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

Ping Xu
Peking University https://orcid.org/0000-0003-4331-8839
Zhiheng Liu
Peking University https://orcid.org/0000-0001-6711-4801
Ying Liu
Peking University
Huazheng Ma
Peking University
Yiyuan Xu
Peking University
Ying Bao
Peking University
Shiyou Zhu
Peking University
Zhongzheng Cao
Peking University
Zhuo Zhou
Peking University
Wensheng Wei (wswei@pku.edu.cn)
Peking University https://orcid.org/0000-0002-8053-2423

Article

Keywords: gene functions, BARBEKO screening, base editing, gene codons, splice sites

Posted Date: August 17th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-57831/v1

Version of Record: A version of this preprint was published at Nature Biotechnology on June 21st, 2021.
See the published version at https://doi.org/10.1038/s41587-021-00944-1.

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

Ping Xu[1,4], Zhiheng Liu[1,2,4], Ying Liu[1,4], Huazheng Ma[1,3,4], Yiyuan Xu[1], Ying Bao[1], Shiyou Zhu[1], Zhongzheng Cao[1,2], Zhuo Zhou[1], Wensheng Wei[1,5,*]

[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 100871, China

[2]Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China

[3]Peking University–Tsinghua University–National Institute of Biological Sciences Joint Graduate Program, Peking University, Beijing 100871, China

[4]These authors contributed equally

[5]Lead Contact

*Correspondence: wswei@pku.edu.cn

Abstract

Canonical CRISPR screens rely on Cas9-induced DNA double-strand breaks (DSBs) to generate targeted gene knockouts. These DSB-dependent methodologies may yield false-positive results by mistakenly assuming targeted loci as essential for cell viability, especially when high-copy-number sites are targeted. Here, we use CRISPR cytosine base editors for genome-scale knockout screens by perturbing gene start codons or splice sites, or by introducing premature termination codons (PTCs). Combining with iBAR strategy we have previously established, we realized an iBARed cytosine Base Editing-mediated gene KnockOut (BARBEKO) screening strategy at a genome-scale (targeting 17,501 genes) in multiple human cell lines. By constructing such a cell library through lentiviral infection at a high MOI (up to 10), we significantly reduced starting cells while producing screening results with improved efficiency and accuracy. More importantly, in comparison with Cas9-mediated cell fitness screens, BARBEKO screens are no longer affected by DNA-cleavage induced cytotoxicity in HeLa, K562, or DSB-sensitive RPE1 cells. We anticipate that BARBEKO offers a valuable tool to complement the current CRISPR screens in various settings.

Introduction

The simplicity of programming CRISPR/Cas9 system to modify specific genomic loci offers an unprecedented opportunity to interrogate gene function in eukaryotes[1–5]. This system has further been employed to develop powerful genetic screening methods for the exploration and functional annotation of genetic elements, which have been broadly applied in various biomedical settings including cancer research and drug discovery[6–9]. Despite its success and broad applications, Cas9-induced double-strand breaks (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[10–14]. 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[15]. Screens of genetic dependency by Cas9 may incur bias in DNA damage response. Despite a computational method for the estimation of the gene-dependency level while taking account of the copy-number effect, it is experimentally demanding to determine copy numbers for cell types of interest, while it is still difficult to resolve mismatch-tolerance effect[11,16]. What's more, it has recently been reported that Cas9-induced DSBs posed obstacles to high-throughput screens in human non-transformed cells via p53-dependent cell growth arrest[17–20]. High-efficiency Cas9 editing could cause cell death in human pluripotent stem cells (hPSCs)[20] and G1 cell cycle arrest in hTERT RPE1 cells[18]. Parallel screens in p53-proficient and -deficient RPE1 cells revealed that Cas9 editing triggered a p53-dependent DNA damage response which negatively impacted the sensitivity of guide-specific effects[18]. However, this is still a subject under heated debate as some groups argued that adequate sgRNA representation in carefully selected cells or clones expressing high-efficiency Cas9 would ensure successful CRISPR-Cas9 screens[17,21].

Nevertheless, to effectively minimize sgRNA mis-association to reduce the false discovery rate, it is a 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[6–9,22]. We have recently established a new screening strategy using re-designed sgRNA harboring internal barcodes (iBARs) that enables high-throughput CRISPR screening at high MOIs, resulting in significant efficiency boost[23]. Although CRISPR^{iBAR} outperformed the conventional methods in positive selection screens, the cytotoxicity of Cas9-induced DSBs[10–14] prevents its broader application from studying gene fitness in negative screens especially with high MOIs[23].

We aim to re-establish CRISPR loss-of-function screening strategy with the following beneficial features: allowing high MOI screening to improve efficiency and economy, ideal for both positive and negative screens, and applicable for screening in non-transformed cells such as primary cells, hPSCs, and RPE1 that usually carry p53 and are sensitive to DSB damage. The simple solution could be the combination of iBAR strategy and CRISPR base editor-mediated gene knockouts. CRISPR-STOP and iSTOP approaches have been proposed to utilize the CRISPR-based cytosine base editor 3 (BE3, C•G to T•A) to introduce nonsense mutations for gene silencing[24,25]. It is foreseeable to achieve broader coverage of genes using CBEs to include additional sites for sgRNA design, splice acceptor sites (Gapinske et al., 2018), splice donor sites and translation initiation sites.

Here, we established a genome-wide iBARed cytosine Base Editing-mediated gene KnockOut (BARBEKO) screening strategy, in which cytosine base editors perturb genes by disrupting splicing sites or translation initiation sites, or introducing premature termination codons (PTCs), and all sgRNAs were re-designed to carry iBARs[23]. BARBEKO approach in genome-scale has been applied in multiple cell lines, HeLa, K562 and RPE1 cells, all at high MOIs for screens of cell fitness. We envision that BARBEKO strategy might be particularly useful for CRISPR screens in complex models such as primary cells, organoids and *in vivo* studies, where the source of cells is usually limited and sensitive to DNA damage, and it is hard, if not impossible, to control transduction efficiency in making libraries.

Results

Design of CBE-based genome-scale sgRNA library for gene knockout screens

In addition to generating effective gene knockouts by utilizing CBEs to introduce PTCs by targeting codons of Gln (5'-CAA, 5'-CAG), Arg (5'-CGA) or Trp (5'-TGG)[24,25], it is foreseeable to achieve gene knockouts by disrupting splice sites (5'-GT, 5'-AG) or start codons (5'-ATG) (Fig. 1a). To examine the effectiveness of CBEs in generating gene knockouts, we designed multiple sgRNAs along the genomic loci of an anthrax toxin receptor gene *ANTXR1* and a diphtheria toxin receptor gene *HBEGF*[9], followed by the transduction of these sgRNAs individually into CBE-expressing HeLa cells. To achieve desirable editing efficiency, AncBE4max, one of the most effective cytosine base editors[28], was employed. All groups (10/10) with sgRNAs targeting *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)[29,30]. Sanger sequencing results confirmed the targeted base transitions (Supplementary Fig. 1). Consistently, all groups (7/7) with sgRNAs targeting *HBEGF* locus obtained resistance to diphtheria toxin (Supplementary Fig. 2). In targeting pan-cancer core fitness genes *RPL11* and *RPL23A*, we also observed robust gene knockouts through AncBE4max in chronic myeloid leukemia K562 cells (Supplementary Fig. 3).

We have previously established an iBAR method that enables high-throughput gene knockout screening on CRISPR library made from high-MOI lentiviral infection[23]. Four verified iBARs from our published iBAR approach were attached to each sgRNA serving as internal replicates in screens (Supplementary Fig. 4a). 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 (Supplementary Table 1). 210,012 sgRNAs covering 17,501 genes (3 sgRNAs per gene) were *in silico* designed, of which 41.8% were newly designed targeting start codons or splice sites, while 58.2% CRISPR-STOP sgRNAs were selected from Kuscu *et al.* (Fig. 1b, and Supplementary Fig. 4b).

BARBEKO strategy achieved high-quality gene fitness screen in HeLa cells at a high MOI

We first applied BARBEKO strategy 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 as ZFC^{iBAR} (Fig. 2b). In

short, we used a Z score to normalize the distribution of log fold change (ZLFC) of each sgRNA^{iBAR}, and combined Robust Rank Aggregation (RRA) analysis to calculate gene Fitness Score (FS), which comprehensively reflected the significance and consistency of the abundance change of 12 sgRNAs^{iBAR} per gene through screening (Supplementary Fig. 5). Using this ZFC^{iBAR} algorithm, both depleted and enriched genes in HeLa cells were revealed under rational cutoffs of gene FS (Fig. 2c). With the help of iBARs serving as internal replicates, ZFC^{iBAR} analysis further improved signal-to-noise ratio of screens, as indicated by the Pearson correlation coefficients of two biological replicates, which increased from 0.75 with sgRNA^{iBAR} ZLFC analysis to 0.96 with gene FS analysis (Fig. 2d-2e). Besides, 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 ZFC analysis without the aid from iBARs (Fig. 2f).

To further evaluate the quality of BARBEKO screening, we compared our results by the area under curve (AUC) of receiver operating characteristic (ROC) curves based on gold-standard reference sets of essential and non-essential genes with a fitness screen utilizing CRISPR^{iBAR} library at an MOI of 3[23], and with a reported Cas9-based screen at a low MOI (~ 0.3)[31]. Gene fitness screens at a high MOI by BARBEKO strategy outperformed both CRISPR^{iBAR} and TKOv1 library screens (Fig. 2g). Similarly, dAUC (delta AUC, difference between sgRNAs targeting essential and non-essential genes) of BARBEKO was much higher than that of TKOv1[32], demonstrating the enhanced specificity of BARBEKO sgRNA library even at high MOIs (Fig. 2h). Taken together, BARBEKO strategy applies well to cell fitness screens and exhibits the potential of high-quality outcomes with much-improved cost- and labor-effectivity.

We went on to compare the results of BARBEKO screens between early and late time points during the fitness screen. The correlation coefficient of two results from Day 15 and Day 21 was high (0.98); however, the number of depleted genes on Day 21 was larger than Day 15 (2,121 vs 1,795) under the same threshold (Supplementary Fig. 6a-6c). These results suggested that a longer duration improve 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 (Supplementary Fig. 6d) mainly belonged to the same GO terms of commonly selected genes of both timepoints (Supplementary Fig. 6e), demonstrating the consistency in the process of screening by BARBEKO strategy to reveal gene functions.

137

138 **Efficiency comparison among different types of sgRNAs**

139 As sgRNAs targeting the gold-standard essential genes are supposed to be depleted in the screen, we
140 categorized these sgRNAs according to the targeting types for efficiency comparison. sgRNA$^{SD/SA}$
141 showed similar ZLFC distribution to sgRNA^{Stop}, while sgRNA^{Start} performed a bit less effectively,
142 assumably due to the presence of alternative translation initiation sites for many targeted genes
143 (Supplementary Fig. 7a). In addition, the efficiency of sgRNA^{SD} was statistically lower than sgRNA^{SA}
144 (Supplementary Fig. 7b), likely due to the context preference of the deaminase domain of rat
145 APOBEC1[34]. Indeed, we found that the 5' guanine (G) adjacent to the targeting cytosine (C)
146 substantially compromised the editing efficiency (Supplementary Fig. 7c). As expected, sgRNA
147 efficiency was influenced by the location of targeted "C" in the editing window as well (Supplementary
148 Fig. 7d). sgRNA^{Stop} targeting different codons also showed distinct ZLFC distributions
149 (Supplementary Fig. 7e), in which targeting the codon of "TGG" had the highest gene knockout
150 efficiency. We infer that the anticodon sequence "CCA" of the DNA strand is more likely edited by
151 the CBE. In conclusion, the above summarized rules might help select better sgRNAs for effective
152 gene knockouts by CBEs.

153

154 **Copy number effect could be diminished in BARBEKO screens**

155 There are a number of reports that Cas9-mediated DNA cleavage in amplified genomic regions induced
156 gene-independent anti-proliferation effect and consequently introduced false positive hits into gene
157 essentiality screens[10,13,16]. To verify if BARBEKO could avert such a problem, we compared sgRNA
158 ZLFC distribution across gene copy numbers of BARBEKO and CRISPR-Cas9 screens in HeLa cells.
159 After normalization by subtracting the ZLFC median of non-targeting sgRNAs, ZLFC of sgRNAs
160 descending in targeting genomic sites correlated with the increased copy numbers in CRISPR-Cas9
161 screens, evidently resulting from DSB-induced cytotoxicity. BARBEKO screen was indeed not
162 affected by copy-number amplification (Fig. 3a). To further confirm this, we selected two genes that
163 are located in amplified genomic regions in HeLa cells, *SDHA* and *TRIP13*[35]. Four *SDHA*-targeting
164 sgRNAs (Fig. 3b) were tested individually in both AncBE4max- and Cas9-expressing cells. No
165 noticeable phenotypic changes were observed in AncBE4max-edited cells, while cell viability was

166 significantly decreased when these loci were perturbed by Cas9 with all four sgRNAs (Fig. 3c). Sanger
167 sequencing and immunoblot analysis further confirmed two CBE-specific sgRNAs were effective in
168 generating *SDHA* knockouts with AncBE4max (Fig. 3d-e), indicating that the decreased cell viability
169 in Cas9 cells was not due to the gene knockouts but the occurrence of multiple DSBs. The same results
170 were obtained for *TRIP13* gene targeting, 3 out of 4 sgRNAs led to decreased cell viability only in
171 Cas9-expressing cells (Supplementary Fig 8a-b).
172
173 **BARBEKO empowers gene fitness screen in K562 cells at ultra-high MOIs**
174 Since library construction with high MOI could significantly reduce the starting cells, we then push
175 the MOI to ~ 10 and test it in K562 cells. K562 contains a Philadelphia chromosome susceptible to
176 single sgRNA-mediated Cas9 cutting; thus, it enables us to examine the potential cytotoxic effect of
177 multiple sgRNAs in BARBEKO screen with ultra-high MOIs. K562 libraries were then made with
178 lentiviral infection at MOIs of ~ 3 and ~ 10 in parallel (Fig. 4a-b). A scatter plot of gene FS showed
179 accordant hits in both depletion and enrichment after screening (Fig. 4c), and the ROC analysis showed
180 comparable AUC scores according to the gold-standard gene sets (Fig. 4d). These results demonstrated
181 that BARBEKO is a robust screening strategy that produces highly consistent results even on cell
182 libraries constructed with lentiviral infection at extremely high MOIs, resulting in much-improved
183 cost- and labor-effectiveness for both positive and negative selection screens. Specifically, to reach
184 1,000-fold coverage per sgRNA, the minimal requirement for a conventional CRISPR library
185 construction at a MOI of ~ 0.3 for 2 experimental repeats is 3.6×10^8 cells; while the number drops to
186 5.4×10^6 for BARBEKO library (4 iBARs/sgRNA serving as internal repeats) at a MOI of ~ 10, over
187 60-fold reduction. Putting the economy aside, this astonishing reduction in cell numbers could be
188 pivotal in large-scale screens when either the source of agents is limited, such as emerging viruses or
189 uncommon toxins, or the screening material is scarce, such as patient-derived cells.
190 To further confirm that BARBEKO approach is immune from Cas9-cleavage-induced
191 cytotoxicity, we chose to test *BCR-ABL* oncogene because this locus suffers from a high-copy tandem
192 amplification during Philadelphia translocation in K562 cells[36]. Cas9 cleavage in this repeated region
193 has been reported to cause false positive selection of essential genes[14]. We plotted the ZLFC of genes
194 located surrounding the fusion gene and compared them with the data from Wang *et al.* (Fig. 4e).

Indeed, the sgRNAs targeting contiguous genes within the amplicons on 22q11.2 and 9q34.1 were significantly dropped out when comparing with the flanking non-amplified regions, indicating Cas9-cleavage-induced cytotoxicity (Fig. 4e, top). These positional effects on non-essential genes were almost completely diminished in two high-MOI screens of BARBEKO approach, while the true essential oncogenic fusion gene *BCR-ABL1* could still be correctly identified (Fig. 4e, middle and bottom). We anticipate that these advantages of BARBEKO strategy are worth to be exploited to the full for critical applications that are sensitive to copy number effect.

BARBEKO enables bidirectional screens of non-transformed cell line RPE1

To understand gene function in more physiological settings, one often need to conduct CRISPR screens in primary cells or other untransformed cells that carry intact and normal cellular machinery, such as the p53 pathway. However, it is currently under heated debate whether these cells are feasible for conventional CRISPR screens because Cas9 cutting-induced DNA damage response could trigger the activation of the p53 pathway, which arrests cell growth and distorts screen outcomes[17–21,37,38]. We then tested if our base-editing screen strategy could offer a solution by assessing the performance of BARBEKO in RPE1, a near-diploid cell line recently come under the spotlight as a model system for non-transformed cell screen[39–41]. To do this, we constructed a control library containing 1,000 non-targeting sgRNAs and an experimental library containing 869 sgRNAs targeting non-essential genes[9]. These two libraries were separately delivered into wild-type, AncBE4max- and Cas9-expressing RPE1 cells at MOIs from low to high (MOI ~ 0.3, 1, 2, 3, 10). Clonogenic survival assays were performed to monitor cell viability (Fi. 5a and Supplementary Fig. 9). Comparing to wild-type RPE1 cells, AncBE4max-expressing cells infected by experimental and control libraries kept similar survival fraction at all levels of MOIs up to 10, while a significant decrease of colony formation ratio was observed especially with high MOIs in the experimental library of Cas9-expressing RPE1 cells (Fig. 5a and Supplementary Fig. 9c). These data clearly showed that BARBEKO is much less toxic to RPE1 cells than conventional CRISPR-Cas9 screens, most likely via preventing p53 activation-associated effects. Prompted by these results, we applied BARBEKO to gene fitness screen in RPE1 cells with a library made of lentiviral transduction at an MOI of ~ 3 (Supplementary Fig. 10a). A volcano plot

displays overall results, and fitness genes were selected under the thresholds of gene FS >4 and <-3 (Fig. 5b).

We noticed several genes were highly enriched at the endpoint of the screen, such as the top-ranked gene neurofibromin 2 (*NF2*), whose sgRNAs^{iBAR} were averagely enriched over ten folds. The distribution of sgRNA counts in the RPE1 screen was distinct from that in the HeLa screen (Supplementary Fig. 10b-c). Gini index, a metrics used to measure the evenness of sgRNA counts in genetic screens[42], increased from 0.263 on Day 0 to 0.595 on Day 24 in RPE1 screen while remained constant in HeLa cells (Supplementary Fig. 10d-e), indicating the uneven distribution of sgRNAs in RPE1 screen and disparate weighting of genes in screens of untransformed cells and cancer cells. GO enrichment analysis showed the Hippo signaling pathway as one of the major terms of these enriched genes in RPE1 (Supplementary Fig. 11a). We listed key components and regulators of the Hippo signaling pathway[43–45], and found that genes directly (*LATS2*, *PTPN14*) or indirectly (*NF2*, *RRMD6*, *SAV1*, *MAP4K4* and *TAOK2*) activating the Hippo signaling pathway were negative regulators of cell proliferation, while YAP/TAZ, the key effectors of Hippo, were essential for cell viability. Perturbations in a number of regulators in the Hippo signaling pathway could effectively unleash cellular proliferation in RPE1 cells (Supplementary Fig. 12a).

In comparison between HeLa and RPE1 screens under the same threshold of FS > 4, 88 and 80 genes were identified as inhibitors of persistent cell proliferation in RPE1 and HeLa cells, respectively. 6 were common hits, including Hippo regulators *NF2* and *SAV1* (Supplementary Fig. 11b). GO enrichment analysis showed that those sgRNAs targeting Hippo regulators were also enriched in the HeLa screen (Supplementary Fig. 11c and 12b). Further comparing Hippo-related genes that were positively or negatively identified in screens from RPE1 and HeLa cells (Supplementary Fig. 12c-d), we found that RPE1 cells, as a non-transformed cell line, were susceptible to perturbations in core Hippo-YAP/TAZ signaling, in agreement with previous studies that aberrant activation of YAP is sufficient to drive uncontrolled cell growth of normal cell types[46]. However, in HeLa - the cancer cell line, in which Hippo pathway is believed to be dysregulated[47–50], these selected genes were distributed in both central and peripheral YAP/TAZ signaling pathways[51].

In sum, BARBEKO screens in HeLa and RPE1 cells identified regulators in Hippo signaling pathway that have strong impact on cell proliferation, suggesting its important implications in cancer research for novel anti-cancer drug targets[52–54].

BARBEKO strategy is immune from generic penalty of DNA damage induced by editing tools

We then scrutinized if BARBEKO is affected by p53-mediated toxic response. It has been reported that Cas9-induced DNA cleavage triggered a p53-dependent cell cycle arrest, resulting in reduced sensitivity of guide RNAs in *TP53* wild-type RPE1 cells, manifested by the failure of depletion of sgRNAs against essential ribosomal genes[18]. From BARBEKO screen in RPE1 cells (Fig. 5b), we listed the rankings of ribosomal genes, essential genes from gold-standard reference sets, non-targeting "genes" and AAVS1 "genes" for the direct comparison (Fig. 5c). We found that both sgRNAs targeting ribosomal genes and gold-standard essential genes were significantly depleted after screening comparing with non-targeting and AAVS1 sgRNAs, indicating a well-functioning screen in RPE1 cells. We also re-analyzed data of gene fitness screens in p53-proficient and -deficient RPE1 cells from Brown *et al.* by ZFC algorithm for direct comparison[21]. The density distribution of gene fitness scores of non-targeting controls coincides well with that of non-essential genes' in BARBEKO screen. In all TKO screens, however, the distribution of non-targeting controls deviated, and the median FS of these controls were higher than that of non-essential genes' in both p53-proficient and -deficient cells (Fig. 5d and Supplementary Fig. 13a). These results indicated that Cas9-mediated DNA cleavage posed long-term impairment on cell proliferation in a manner not fully dependent on p53, thus the cells containing non-targeting sgRNAs obtained growth advantages over those carrying lesions in non-essential genes. In contrast, no such phenotypes were observed in CBE-mediated knockout screens, and this will give rise to accurate outcomes directly associated with the function of sgRNA-targeted loci.

Disruptions in the C-terminus of CDKN1A caused cell death

To study the relationship between p53 and reduced cell viability after Cas9 editing, we compared p53 signaling pathway-related genes by scatter plotting of gene FS or ZLFC/LFC in BARBEKO and Cas9-based screens (Fig. 5e-f and Supplementary 13b-f)[21,39]. In BARBEKO screen, perturbations in p53

signaling pathway showed strong impact on cell proliferation, in agreement with prior reports, including p53 negative regulators MDM2, MDM4 and UBE2N[55–57] being essential for cell viability, positive regulator USP28[58] and p53's downstream effector retinoblastoma protein (RB1)[59] restricting persistent cell proliferation. Unexpectedly, we noticed one sgRNA^{Stop} targeting the C-terminus of cyclin-dependent kinase inhibitor 1A (*CDKN1A*, encoding p21) was drastically depleted (Supplementary Fig. 14a-b). p21, transcriptionally controlled by p53, is a cyclin-dependent kinase inhibitor, whose loss-of-function is supposed to be beneficial to cell proliferation. We therefore presumed that a truncated p21 variant caused by Gln138-targeting sgRNA^{Stop} might aggregate in the nucleus and inhibit cyclin-dependent kinases to induce cell cycle arrest, as reported that loss of phosphorylation at Thr145 constructed arrested cells in G_0 phase[60,61]. Other than acting as the cyclin-dependent kinase inhibitor, p21 has been reported to play versatile roles in multiple cellular processes, such as cell differentiation, migration, apoptosis, and DNA damage repair[62]. As cellular context, subcellular localization and post-translational modifications could all change p21 activities and functions, and we ought to pay special attention to cases like *CDKN1A* perturbation in screens. This is apparently not unique for BARBEKO screens (Supplementary Fig. 14c).

Discussion

We developed a novel approach called BARBEKO that combines cytosine base editor and iBARed sgRNAs for high-throughput genetic screens. In comparison, BARBEKO surpasses conventional CRISPR screening as follows: 1) cell number required for library construction could be dropped up to over 60-fold to reach the same level of coverage (Fig. 4); 2) iBARs serving as internal replicates improved screening quality (Fig. 2f-g); 3) such loss-of-function screens are immune of copy number effect on cytotoxicity (Fig. 3 and Fig. 4e) and immune of gene-independent cytotoxicity, and therefore, BARBEKO is better suited for p53 wild-type cells (Fig. 5d and Supplementary Fig. 13a).

BARBEKO strategy has been applied to gene fitness screens of HeLa, K562 and RPE1 cells, all at high MOIs, and yielded comprehensive list of genes affecting, either positively or negatively, cell proliferation. Among them, a number of genes from Hippo signaling pathway were isolated from both RPE1 and HeLa cells. Interestingly, two MAP4K family kinases, MAP4K3 and MAP4K5, were identified as inhibitors of persistent cell proliferation of K562 cells (Fig. 4a-b). It has been reported that six MAP4K family kinases, including MAP4K1/2/3/4, MINK1 and TNIK, but not MAP4K5, could activate LATS1/2 in the Hippo signaling in HEK293A cells[63]. MAP4K4 and TNIK were identified as inhibitors of persistent cell proliferation of RPE1 cells (Supplementary Fig. 12c), and disruptions of the MINK1 function tended to promote HeLa cell viability. It seems that there exists cell type specificity of MAP4K family kinase-activated Hippo signaling, which controlled steady cell proliferation in both non-transformed and cancer cells. Therefore, we speculated that MAP4K5, not functional in adherent cell types but important for cell proliferation control of K562 cells, might be a novel regulator of the Hippo signaling pathway in suspension cells.

As a matter of fact, negative screening is usually more technically challenging to obtain signal-to-noise ratio with quality and demands much bigger size of library than positive screening[64]. In addition, gene-independent cytotoxicity triggered by Cas9-mediated cleavage often muddles the results of negative screens related to cell fitness, as the depletion level triggered by gene loss-of-function is generally modest[65]. It is an alarming issue that DSB-activated p53 signaling impacts the precision of gene fitness screens from recent reports[17–19,21]. We did notice such gene-independent cytotoxicity of Cas9 cleavage in wild-type RPE1 cells from conventional screening, but not from BARBEKO screening even at high-MOI settings (Fig. 5a and 5d).

High-throughput genetic screens in primary cells or *in vivo* is pivotal in preclinical research. In addition, there is a high demand for more efficient genetic screen approaches for SARS-CoV-2 studies. However, different from the simple manipulation of common materials, many aspects restrict Cas9-mediated approaches from these studies: the limited resource of primary cells derived from patients makes it hard for satisfactory screening conducted in duplicates at a low MOI; low transduction efficiency in making libraries makes it even worse for *in vivo* screening; handling large number of cells increases risk of SARS-CoV-2 exposure. Inspired by the aforementioned analysis of gene fitness screens by BARBEKO strategy, we envision it could be the choice for such screens. Benefited from iBAR approach and the DSB-free editing tool, screens could be conducted without the need to monitor MOI for library construction or duplicate screening experiments replicates.

During the process of our screens, several articles reported some optimized versions of CBEs with extended targeting scope by flexible PAM or expanded activity window[66–69], which could be helpful to CBE-based library design with improved sgRNA quality and coverage. There are about 1,700 genes are missing in current version of 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 DNA damage response based on dCas9[70–75] could also be employed dependent on research needs.

Acknowledgments

This work was supported by funds from the National Science Foundation of China (NSFC31430025), the Beijing Advanced Innovation Center for Genomics at Peking University, the Peking-Tsinghua Center for Life Sciences (to W.W.) and the National Major Science & Technology Project for Control and Prevention of Major Infectious Diseases in China (2018ZX10301401, to Z.Z.). The authors acknowledge the staff of the BIOPIC High-throughput Sequencing Center (Peking University), the National Center for Protein Sciences (Beijing) at Peking University for assistance with Dr H.L., Ms. L.D. and Ms. H.Y. for technical assistance, and the High-Performance Computing Platform at Peking University for providing platforms of NGS data analysis. The authors thank Y. Sun and D. Xu (Peking University) for providing hTERT RPE1 and RPE1-TP53KO-Cas9 cell line.

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, which revised by W.W., Z.Z., H.M., Y.L. and Z.L.

Competing interests

The authors have filed provisional patent via Peking University related to this work.

Methods

Cell culture and reagents

The HeLa CCL2 and HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Gibco), K562 cells were maintained in RPMI1640 medium (Gibco) and hTERT RPE1 cell was cultured in DMEM/F12 medium (Gibco). All cell lines were supplemented with 10% FBS (Biological Industries) and 1% penicillin/streptomycin, cultured with 5% CO2 at 37 degree centigrade. All cells were checked to ensure they are free of mycoplasma contamination.

Cloning

The sequence of AncBE4max was obtained from supplementary information of Koblan *et al.* [28] and synthesized by Synbio Technologies. The AncBE4max construct was cloned into pLenti-P2A-mCherry vector and pLVX-TRE3G vector (TaKaRa, 631187) through double restriction enzyme digestion (NEB) and T4 ligase ligation (NEB, M0202). Individual sgRNA oligos were synthesized by Ruibotech and cloned into pCG-2.0 sgRNA-expressing vector through Golden-Gate assembly.

Phenotypes of toxin-receptor-gene knockouts by AncBE4max

sgRNAs targeting *ANTXR1* and *HBEGF* were lentiviral infected into HeLa cells. GFP^+ cells were FACS sorted and treated with PA/LFnDTA (70 ng/ml PA plus 50 ng/ml LFnDTA) for 48 hours or 7.5 ng/ml Diphtheria toxin (DT, List Biological Laboratories Inc.) for 60 hours and conducted in triplicates 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 by PrimerSTAR HS DNA Polymerase (TaKaRa, R010Q). Then PCR products of *HBEGF* and *ANTXR1* were purified using DNA clean & concentrator-5 (ZYMO research, D4013).

Cell proliferation assay

All sgRNAs were cloned into a lentiviral backbone carrying CMV promoter driven EGFP and packaged into lentiviruses in HEK293T cells. Then sgRNA lentiviruses were delivered into AncBE4max-, Cas9- or KRAB-dCas9-expressing cells at efficiency within 40% - 60%. The percentage of $EGFP^+$ cells was quantified through flow cytometry (LSRFortessa, Becton Dickinson Inc.). The first analysis started from two days post infection, labeled as Day 0, serving as baseline for normalization. Then the percentage of $EGFP^+$ cells were analysis every three days, until Day 15 or Day 24.

Design of genome-scale gene knockout 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 containing targeted cytosine in the 4th to 8th positions were considered (the distal position from PAM is defined as the 1st position, the same below). In consideration of sgRNA on-targeting efficiency, the above sgRNAs 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 smaller than 0.2 or larger than 0.8.

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

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

$sgRNA^{SD/SA}$: Annotations of exon start positions and end positions were extracted from NCBI RefSeq of hg38 assembly to get the genomic sequences around 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 activity window.

$sgRNA^{Stop}$: $sgRNAs^{Stop}$ were introduced from CRISPR-STOP library [25]. We mapped the sgRNA sequences to the human reference genome of hg38 assembly, because the sgRNAs were designed based on hg19 version.

The total number of $sgRNA^{Start}$, $sgRNA^{SD/SA}$ and $sgRNA^{Stop}$ is 512,914. Then we selected three sgRNAs for each gene based on a reasonable scoring scheme (Table S3) for efficient and specific editing. The following situations were considered in selection:

1) sgRNAs with NGG PAM are better than NAG.
2) Distances between sgRNA targeting sites and translational initiation sites. The shortest transcripts of individual genes were considered as reference, then sgRNAs targeting beyond the shortest transcripts were defined as sgRNAs targeting UTR regions.
3) $sgRNA^{SA}$-targeted exons contain multiples of three nucleotides. Here, we considered that skipping of an in-frame exon probably decreases the gene knockout efficiency.
4) A guanine locates in 5' of the targeted cytosine while the targeted cytosine locates in the 4th, 5th or 8th position of sgRNA sequence. The editing efficiency is affected by sequence context of 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 adjacent Cs in the genome, we preferred sgRNAs with the C located in 6th or 7th locations. When high score sgRNAs were more than 3 for one gene, we preferred to select sgRNAs targeting different locations.

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

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

The source code for sgRNA library design can be accessed from https://bitbucket.org/WeiLab/barbeko_sgrna_design/src/master/.

sgRNA plasmid library construction

sgRNA oligonucleotides were array synthesized by Synbio Technologies. Primers (oligo-F and oligo-R) targeting the flanking sequences of oligos were used for PCR amplification of sgRNA sequence form chip. 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, 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).

High-throughput screens via BARBEKO strategy

A total of 7.1×10^7 HeLa cells, 7.1×10^7 (MOI ~3) and 4.3×10^7 (MOI ~10) K562 cells were respectively seeded onto 15-cm plates or T-175 flasks for library construction in duplicates. Then library cells were subjected to puromycin treatment (1 μg/mL for HeLa, 3 μg/mL for K562) for selection. 5 days post infection, a library size of cells was collected as reference group for genome extraction and denoted as Day 0. Cells were passaged every 3 days, and experimental groups were collected on Day 15 and Day 21 of HeLa screen and Day 30 for K562 screens. Genomic DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen) and the sgRNA regions were PCR amplified by Q5 Hot Start High-Fidelity DNA Polymerase (NEB, M0492) or KAPA HiFi HotStart ReadyMixPCR Kit (KAPABIOSYSTEMS, KK2602) with 26-28 cycles of reaction using several pairs of primers (iBAR-

F and iBAR-R). The PCR products were pooled together and purified with DNA Clean & Concentranter-5 (Zymo Research Corporation, D4013), followed by NGS analysis.

For screens in RPE1 cells, a total of 1.8×10^7 RPE1 cells were plated onto 15-cm plates and infected by lentiviral sgRNA library at MOI of 3. The library cells were subjected to puromycin treatment (15 μg/mL) for selection. 200 ng/mL doxycycline (Dox) was added into culture medium to turn on the expression of AncBE4max for 3 days. Then a library size of cells was harvest for genome extraction as reference and denoted as Day 0. The rest of cells were divided into 3 libraries for another 3-day co-culture with Dox. One library for fitness screen was sub-cultured every 3 days and the other two libraries were treated by sub-lethal cisplatin (8 μmol/mL, Selleck) and ionizing radiation (1.9 Gy, Rad Source Technologies Inc. X-ray cabinet RS 2000) every 6 days. Experimental cells were harvested on Day 15 after twice treatment and Day 24 after 3-times treatment.

Computational analysis algorithm for screens

To analyze NGS data of screens by BARBEKO strategy, we developed a new algorithm named ZFC^{iBAR}, which adopted z-score of log_2 fold change to evaluate change of $sgRNA^{iBAR}$ abundance between reference group and experimental group.

Firstly, raw counts of $sgRNA^{iBAR}$ were adjusted by total-count normalization (fitness screens) or median-ratio normalization (drug screens) to correct batch effects. We defined those $sgRNAs^{iBAR}$ of count less than 0.05 quantile in the distribution of reference and experimental groups as small count $sgRNA^{iBARs}$. 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 reference group.

Secondly, log_2 fold change of each $sgRNA^{iBAR}$ was calculated as following:

$$sgRNA^{iBAR}\ LFC = log_2 \frac{normC_{exp} + normC_{small}}{normC_{ref} + normC_{small}}$$

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

Thirdly, to calculate the standard deviation of z-score normalization, $sgRNA^{iBAR}$ LFC was divided into numbers of bins according to corresponding count in reference group and fitted with a linear model, which was applied to calculate the LFC standard deviation for all $sgRNAs^{BAR}$. Inspired by Colic *et al.* [76], the z-score of LFC (ZLFC) was calculated as following:

$$sgRNA^{iBAR}\ ZLFC = \frac{sgRNA^{iBAR}\ LFC}{LFC_{std}}$$

where, LFC_{std} was the standard deviation calculated from the linear model. The empirical P value of sgRNA^{iBAR} ZLFC was calculated.

Fourthly, ZLFC of sgRNAs were calculated as the mean of ZLFC of corresponding sgRNAs^{BAR}, then ZLFC of genes were calculated as the mean of ZLFC of corresponding sgRNAs.

$$sgRNA\ ZLFC = \frac{sgRNA^{iBAR}\ ZLFC}{n}$$

$$gene\ ZLFC = \frac{sgRNA\ ZLFC}{m}$$

where, ***n*** was the number of sgRNAs^{BAR} belonging to certain sgRNA and ***n*** equaled to 4 in BARBEKO strategy here, while ***m*** was the number of sgRNAs belonging to certain gene and ***m*** equaled to 3 here.

Fifthly, Robust Rank Aggregation (RRA) was utilized to calculate the ranking significance for certain sgRNA or gene by the ranking of sgRNAs^{BAR} in the whole library[77]. For bidirectional screens, RRA was calculated twice based on ranking of enrichment and depletion.

Finally, gene fitness score (FS) was calculated based on gene ZLFC and RRA as following:

$$gene\ FS = gene\ ZLFC + [-log_{10}(RRA + 10^{-4})]$$

where, the final RRA value was dependent on the plus or minus sign of gene ZLFC.

ZFC algorithm has been implemented by Python 3 and can be downloaded from https://github.com/wolfsonliu/zfc.

Clonogenic survival assay

RPE1 cells were seeded onto 6-well plates (1×10^5 per well) and treated by lentiviral infection for 24 hours. one-day post treatment, negative control groups without any treatment were counted and sub-cultured into new 6-well plates at the density of 200 cells per well, while experimental groups were seeded of the same volume as control. Cells were cultured for an additional 9 days, then viable colonies were fixed by methanol, stained by 0.1% crystal violet (Solarbio, G1062), and counted manually.

Analysis of copy number effect

Information of absolute copy number was obtained from measurements by Liu *et al.* [35] and average gene copy number of HeLa CCL2 cells was used in our analysis. Relative sgRNAs ZLFC of protein-coding genes was calculated by original sgRNA ZLFC subtracting the median ZLFC of non-targeting

534 sgRNAs.
535
536 **Immunoblot**
537 Cells were lysed by RIPA lysis buffer (CWBIO, CW2333) with protease inhibitor cocktail (CWBIO,
538 CW2200), then samples were concentrated by BCA protein assay (Pierce, 23227) and prepared with
539 SDS-PAGE loading buffer (CWBIO, CW0027). Western blot was performed following standard
540 methods. Primary antibodies used here were anti-β-tubulin (CWBIO, CW0098M) anti-SDHA (Cell
541 Signaling TECHNOLOGY, 11998). And goat anti-rabbit IgG-HRP (Jackson Immunoresearch,
542 111035003) or goat anti-mouse IgG-HRP (Jackson Immunoresearch, 115035003) secondary
543 antibodies were used. The membranes were incubated with Clarity^{TM} Western ECL Substrate kit (Bio-
544 rad, 1705060) and imaged with Chemidoc^{TM} Imaging system (Bio-rad, 1708370). Relative protein
545 level was analyzed by ImageJ software.
546

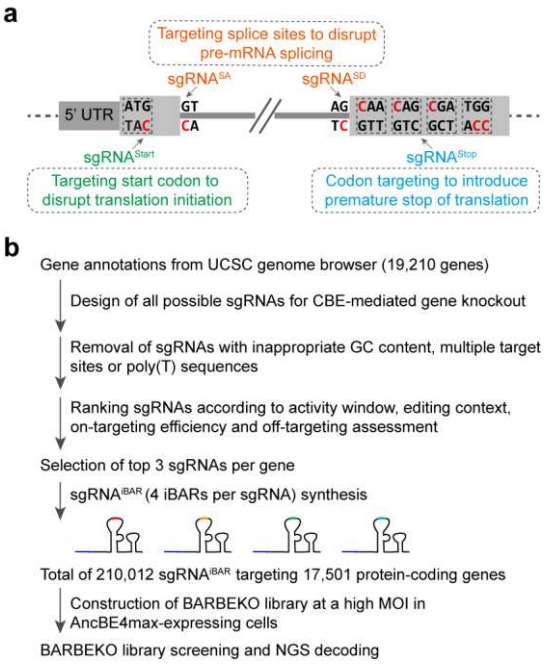
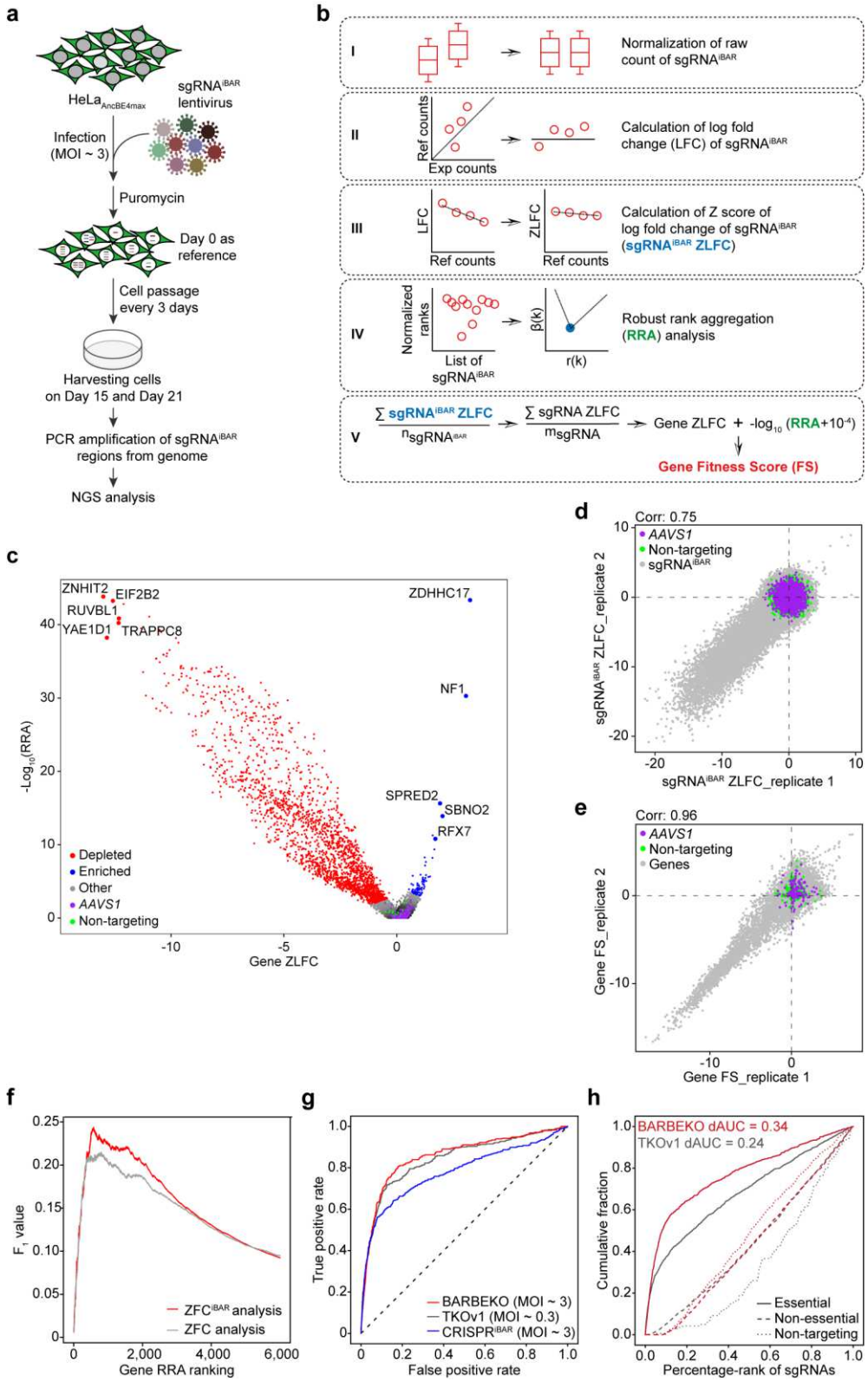


Fig. 1 | Design of CBE-based genome-scale sgRNA library for gene knockout screens. a, CBE with sgRNAs targeting start codons (sgRNA^{Start}), splice acceptor sites (sgRNA^{SA}), splice donor sites (sgRNA^{SD}) and codons of Gln, Arg or Trp (sgRNA^{Stop}) disrupts gene functions. **b**, The flow chart depicts selection and filtration of sgRNAs^{iBAR} for BARBEKO library.



554

555

556 **Fig. 2 | BARBEKO strategy achieved high-throughput gene fitness screens in HeLa cells. a**,

557 Workflow of gene fitness screen in HeLa cells. AncBE4max-expressing HeLa cells were infected by

lentiviruses library of BARBEKO at a MOI of ~ 3, then reference cells denoted by Day 0 were harvested 5 days post infection, and experimental groups were harvest on Day 15 and Day 21. **b**, The schematics of ZFC algorithm describes the analysis processes of NGS data from BARBEKO screens. Gene Fitness Score (FS) is an integrated index of Z score of Log_2(fold change) (ZLFC) and the value of RRA. **c**, Volcano plot shows overall outcome of gene fitness screen in HeLa cells by BARBEKO analyzed by ZFC^{iBAR}. Depleted and enriched genes in screen are plotted in red and blue respectively, and the top5 genes in both directions are labelled individually. Every 3 sgRNAs targeting AAVS1 and non-targeting sgRNAs were randomly grouped as "genes", plotting in purple and green. **d**, Scatter plot of $sgRNA^{iBAR}$ ZLFC of two biological replicates, Pearson correlation coefficient is indicated on the top. $sgRNA^{iBARs}$ targeting AAVS1 locus and non-targeting $sgRNA^{iBARs}$ as negative control are labelled in purple and green. **e**, Scatter plot of gene Fitness Score (FS) of two biological replicates, Pearson correlation coefficient is indicated on the top. **f**, Comparison of F1 value against gene RRA ranking when considering iBARs as internal replicates or ignoring iBARs for ZFC analysis. F_1 value is determined by gene gold-standard sets. **g**, Receiver operating characteristic (ROC) analysis of depleted genes for BARBEKO library at MOI ~3 (considering iBARs as internal replicates in analysis), $CRISPR^{iBAR}$ library at a MOI of ~ 3 (ignoring iBARs in analysis) and TKOv1 library at a MOI of ~ 0.3 [78] screened in HeLa cells. **h**, Comparison of AUCs for essential (solid line), non-essential (dashed line), and non-targeting (dotted line) sgRNAs between BARBEKO a ta MOI of ~ 3 and TKOv1 at a MOI of ~ 0.3 screened in HeLa cells. The dAUCs value from essential and non-essential sgRNAs is indicated in the upper left corner.

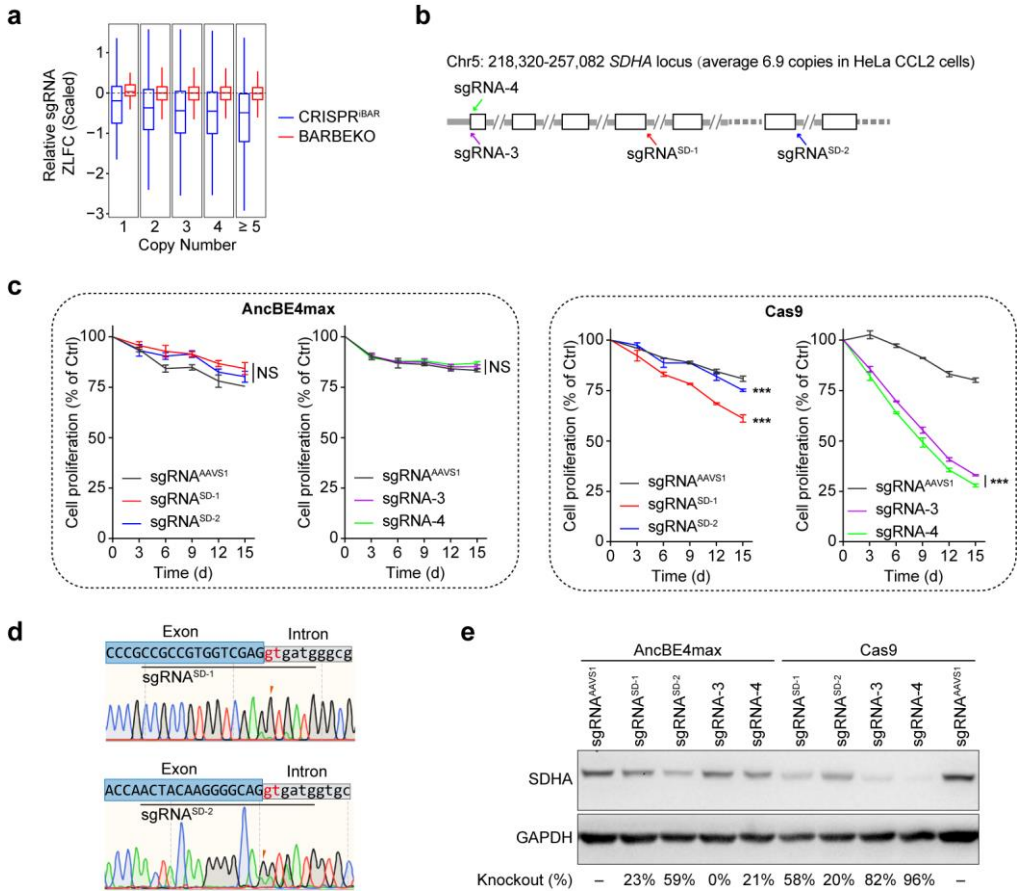


Fig. 3 | Copy number effect could be diminished in BARBEKO screens. a, Boxplot diagram shows relative sgRNA ZLFC of BARBEKO and CRISPR^{iBAR} screens at a MOI of ~ 3 according to gene copy number in HeLa cells. ZLFC of sgRNAs targeting protein-coding genes subtracted the median ZLFC of non-targeting sgRNAs serving as relative sgRNA ZLFC. **b**, Schematic shows 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. 4 sgRNAs were individually delivered into AncBE4max- and Cas9-expressing cells for validation, in which sgRNA^{SD-1} and sgRNA^{SD-2} targeting splice donor sites of *SDHA*. sgRNA^{AAVS1} served as negative control. *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, ***$p < 0.001$. **d**, The sanger sequencing chromatograms of sgRNA^{SD-1} and sgRNA^{SD-2} targeting splice donor sites of *SDHA* genomic region after AncBE4max editing. The orange arrows indicate peaks of targeted "Gs" in splice donor sites. **e**, Immunoblot analysis shows the abundance of SDHA protein of AncBE4max- or

595 Cas9-edited cells with the indicated sgRNAs. KO efficiency was calculated based on the protein
596 level of $sgRNA^{AAVS1}$ group.
597

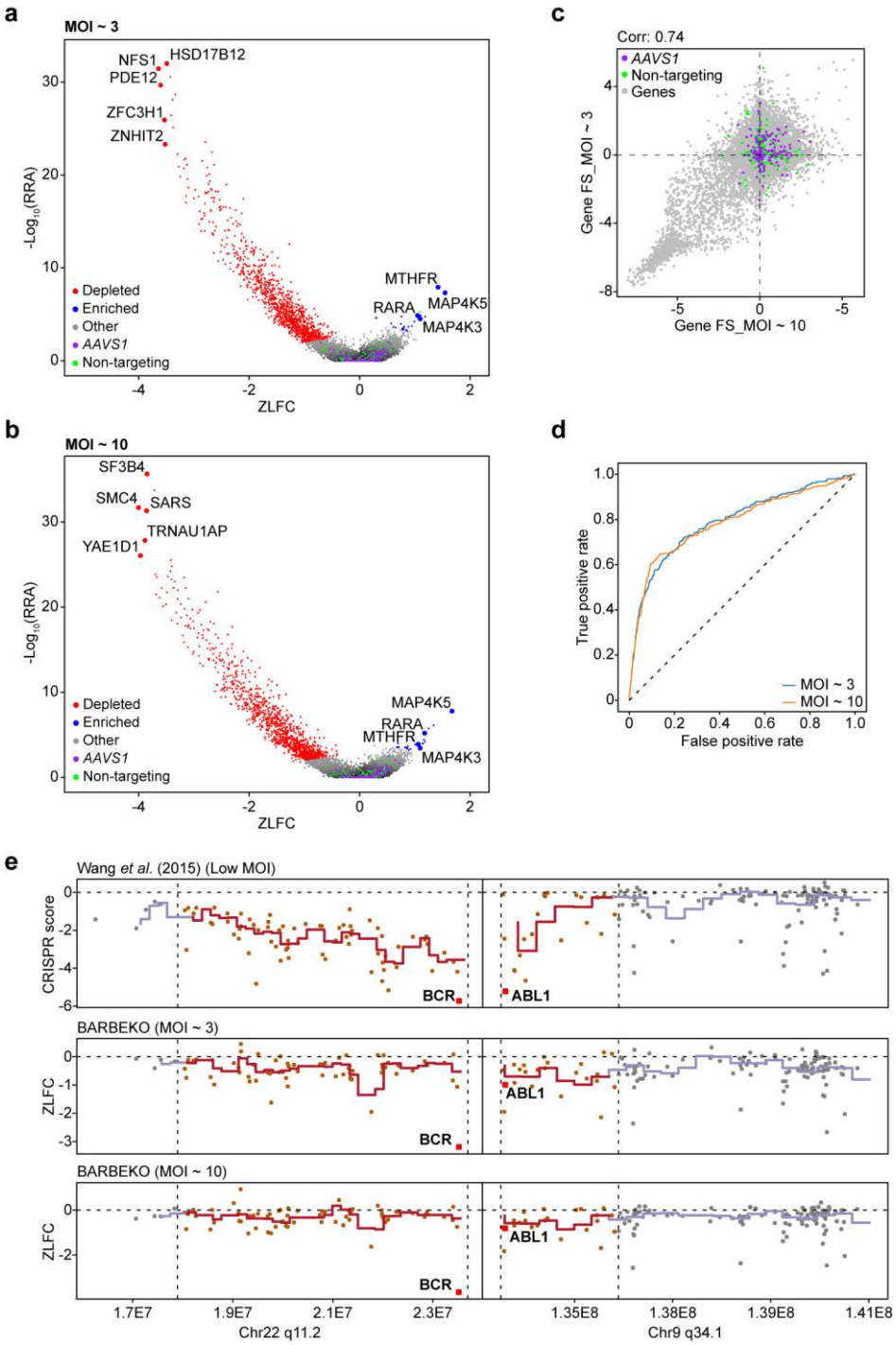


Fig. 4 | BARBEKO Empowers Gene Fitness Screen in K562 Cells at MOI ~3 and MOI ~10. a and **b**, Volcano plot shows overall outcomes of gene fitness screens in K562 cells by BARBEKO at MOIs of ~ 3 (A) and ~ 10 (B). Depleted and enriched genes are plotted in red and blue, respectively. The top5 depleted genes and commonly enriched genes of both screens are labelled individually. Every 3 sgRNAs targeting AAVS1 and non-targeting sgRNAs were randomly grouped as "negative

control genes", plotting in purple and green. **c**, Scatter plot of gene FS in screens at MOIs of ~ 3 and ~ 10, Pearson correlation coefficient is indicated on the top. **d**, Receiver operating characteristic (ROC) analysis of depleted genes in screens at MOIs ~ 3 and ~ 10 according to essential genes and non-essential genes of gold-standard sets. **e**, ZLFC or CRISPR Scores of genes locating around BCR-ABL fusion gene are plotted sequentially. Data from Wang *et al.* is shown in the top lane [8], and the results of BARBEKO at MOIs of ~ 3 and ~ 10 in K562 are shown in the middle and bottom. Genes in this region are separated into 200 bins to calculate mean of ZLFC or CRISPR Score, which are represented by solid lines. The high-copy tandem amplified region is indicated in red, and the flanking regions are in grey.

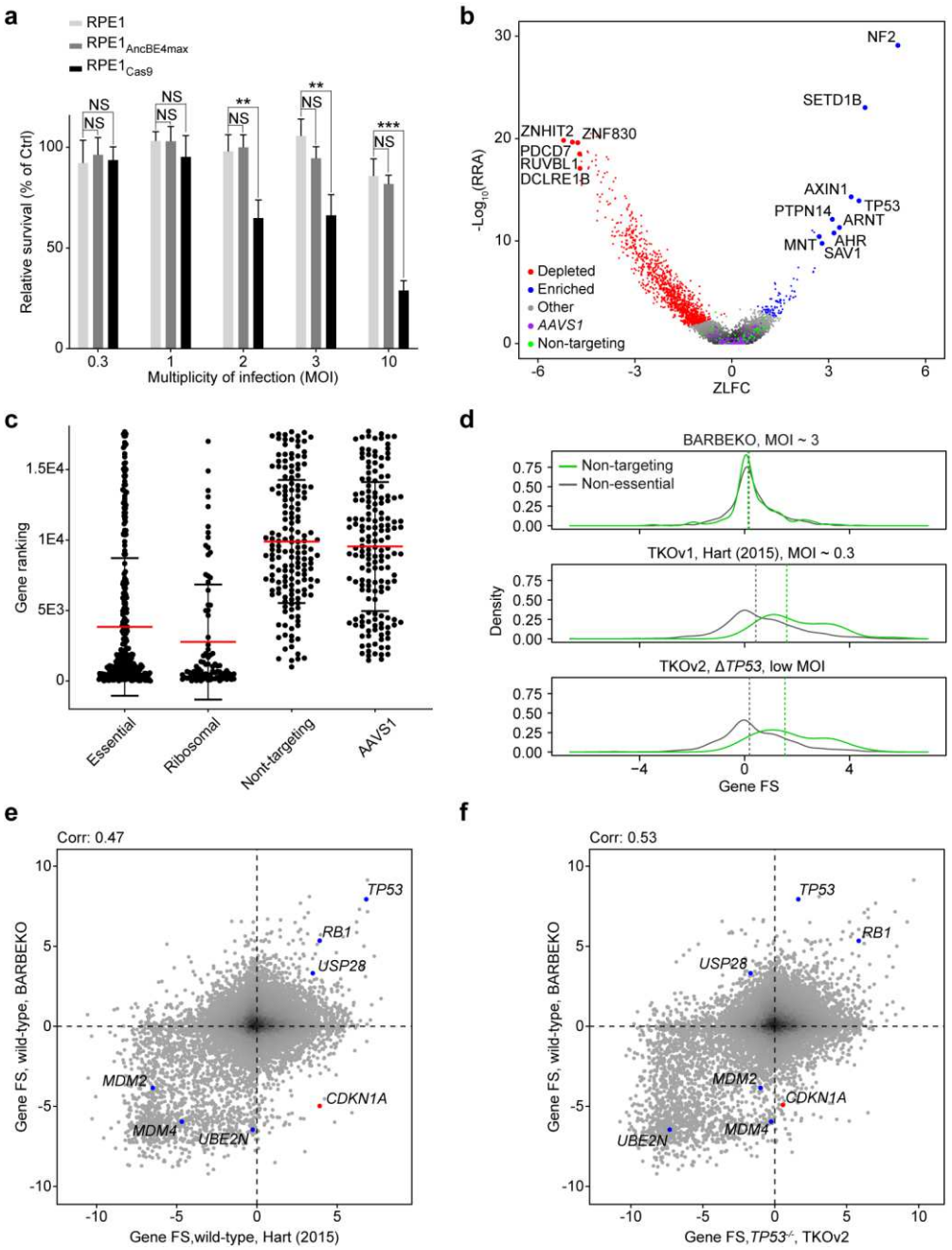


Fig. 5 | BARBEKO enables bidirectional screens of RPE1 cells at a high MOI. a, Clonogenic survival of RPE1 cells in response to sgRNA library transduction at gradient MOIs. Non-targeting control library (1,000 sgRNAs) and non-essential-gene-targeting experimental library (869 sgRNAs) were transduced to wild type, AncBE4max- and Cas9-expressing RPE1 cells at MOIs of 0.3, 1, 2, 3, 10. Clonogenic assay was performed in triplicates 3 days post infection, and the survival fraction (SF) of experimental group was normalized by control SF to calculate the relative percentage. Data are presented as the mean and s.d., $^{**}p < 0.01$; $^{***}p < 0.001$. **b**, Volcano plot shows overall outcome

624 of gene fitness screen in RPE1 cells by BARBEKO at a MOI of ~ 3. The top5 depleted genes and
625 top9 enriched gene were labelled individually. **c**, Scatter plot shows distribution of gene rankings of
626 4 different categories. Essential genes and ribosomal genes were extracted from reference gene sets,
627 while non-targeting and AAVS1 "genes" represented the mean rankings of 3 sgRNAs by randomly
628 sampling. **d**, Comparisons between non-targeting "genes" (green curves) and non-essential genes
629 (grey curves) of the Fitness Score density distribution. Data from TKO screens were re-analyzed by
630 ZFC, and their sgRNAs targeting EGFP, LacZ and luciferase were considered as non-targeting to
631 human genome. **e** and **f**, Scatter plot shows comparisons of gene Fitness Score between BARBEKO
632 and TKOv1 screens in *TP53* wild-type RPE1 cells (**c**); comparisons between BARBEKO in *TP53*
633 wild-type RPE1 and TKOv2 in $TP53^{KO}$ RPE1 cells (**d**). *TP53* and p53 signaling genes are
634 highlighted in blue, while *CDKN1A* is especially highlighted in red. Pearson correlation coefficient is
635 indicated on the top.

636

637 # References

638 1. Chang, N. *et al.* Genome editing with RNA-guided Cas9 nuclease in Zebrafish embryos. *Cell Research* **23**,
639 465–472 (2013).
640 2. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–823 (2013).
641 3. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9–crRNA ribonucleoprotein complex mediates
642 specific DNA cleavage for adaptive immunity in bacteria. *PNAS* **109**, E2579–E2586 (2012).
643 4. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity.
644 *Science* **337**, 816–821 (2012).
645 5. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823–826 (2013).
646 6. Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic
647 screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature Biotechnology* **32**, 267–
648 273 (2014).
649 7. Shalem, O. *et al.* Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **343**, 84–87
650 (2014).
651 8. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic Screens in Human Cells Using the CRISPR-Cas9
652 System. *Science* **343**, 80–84 (2014).
653 9. Zhou, Y. *et al.* High-throughput screening of a CRISPR/Cas9 library for functional genomics in human
654 cells. *Nature* **509**, 487–491 (2014).
655 10. Aguirre, A. J. *et al.* Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9
656 Targeting. *Cancer Discovery* **6**, 914–929 (2016).
657 11. Fortin, J.-P. *et al.* Multiple-gene targeting and mismatch tolerance can confound analysis of genome-
658 wide pooled CRISPR screens. *Genome Biology* **20**, 21 (2019).

659 12. Gonçalves, E. *et al.* Structural rearrangements generate cell-specific, gene-independent CRISPR-Cas9 loss
660 of fitness effects. *Genome Biology* **20**, 27 (2019).
661 13. Munoz, D. M. *et al.* CRISPR Screens Provide a Comprehensive Assessment of Cancer Vulnerabilities but
662 Generate False-Positive Hits for Highly Amplified Genomic Regions. *Cancer Discov* **6**, 900–913 (2016).
663 14. Wang, T. *et al.* Identification and characterization of essential genes in the human genome. *Science* **350**,
664 1096–1101 (2015).
665 15. Shrivastav, M., De Haro, L. P. & Nickoloff, J. A. Regulation of DNA double-strand break repair pathway
666 choice. *Cell Research* **18**, 134–147 (2008).
667 16. Meyers, R. M. *et al.* Computational correction of copy number effect improves specificity of CRISPR–Cas9
668 essentiality screens in cancer cells. *Nature Genetics* **49**, 1779–1784 (2017).
669 17. Bowden, A. R. *et al.* Parallel CRISPR-Cas9 screens clarify impacts of p53 on screen performance. *eLife* **9**,
670 e55325 (2020).
671 18. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR–Cas9 genome editing induces a
672 p53- mediated DNA damage response. *Nature Medicine* **24**, 927–930 (2018).
673 19. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. Reply to "CRISPR screens are feasible in
674 *TP53* wild-type cells". *Mol Syst Biol* **15**, (2019).
675 20. Ihry, R. J. *et al.* p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nature Medicine*
676 **24**, 939–946 (2018).
677 21. Brown, K. R., Mair, B., Soste, M. & Moffat, J. CRISPR screens are feasible in *TP 53* wild-type cells. *Mol*
678 *Syst Biol* **15**, (2019).
679 22. Peng, J., Zhou, Y., Zhu, S. & Wei, W. High-throughput screens in mammalian cells using the CRISPR-Cas9
680 system. *FEBS Journal* **282**, 2089–2096 (2015).

681 23. Zhu, S. *et al.* Guide RNAs with embedded barcodes boost CRISPR-pooled screens. *Genome Biology* **20**,
682 20 (2019).
683 24. Billon, P. *et al.* CRISPR-Mediated Base Editing Enables Efficient Disruption of Eukaryotic Genes through
684 Induction of STOP Codons. *Molecular Cell* **67**, 1068-1079.e4 (2017).
685 25. Kuscu, C. *et al.* CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nature*
686 *Methods* **14**, 710–712 (2017).
687 26. Gapinske, M. *et al.* CRISPR-SKIP: programmable gene splicing with single base editors. *Genome Biology*
688 **19**, (2018).
689 27. Yuan, J. *et al.* Genetic Modulation of RNA Splicing with a CRISPR-Guided Cytidine Deaminase. *Molecular*
690 *Cell* **72**, 380-394.e7 (2018).
691 28. Koblan, L. W. *et al.* Improving cytidine and adenine base editors by expression optimization and ancestral
692 reconstruction. *Nature Biotechnology* **36**, 843–846 (2018).
693 29. Bradley, K. A., Mogridge, J., Mourez, M., Collier, R. J. & Young, J. A. T. Identification of the cellular receptor
694 for anthrax toxin. *Nature* **414**, 225–229 (2001).
695 30. Wei, W., Lu, Q., Chaudry, G. J., Leppla, S. H. & Cohen, S. N. The LDL Receptor-Related Protein LRP6
696 Mediates Internalization and Lethality of Anthrax Toxin. *Cell* **124**, 1141–1154 (2006).
697 31. Hart, T., Brown, K. R., Sircoulomb, F., Rottapel, R. & Moffat, J. Measuring error rates in genomic
698 perturbation screens: gold standards for human functional genomics. *Molecular Systems Biology* **10**, 733–
699 733 (2014).
700 32. Sanson, K. R. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nature*
701 *Communications* **9**, (2018).
702 33. Dempster, J. M. *et al.* Agreement between two large pan-cancer CRISPR-Cas9 gene dependency data

703 sets. *Nat Commun* **10**, 1–14 (2019).

704 34. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in

705 genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).

706 35. Liu, Y. *et al.* Multi-omic measurements of heterogeneity in HeLa cells across laboratories. *Nature*

707 *Biotechnology* 1 (2019) doi:10.1038/s41587-019-0037-y.

708 36. Wu, S. Q. *et al.* Extensive amplification of bcr/abl fusion genes clustered on three marker chromosomes

709 in human leukemic cell line K-562. *Leukemia* **9**, 858–862 (1995).

710 37. Enache, O. M. *et al.* Cas9 activates the p53 pathway and selects for p53-inactivating mutations. *Nat Genet*

711 (2020) doi:10.1038/s41588-020-0623-4.

712 38. Geisinger, J. M. & Stearns, T. CRISPR/Cas9 treatment causes extended TP53-dependent cell cycle arrest

713 in human cells. *Nucleic Acids Res* doi:10.1093/nar/gkaa603.

714 39. Drainas, A. P. *et al.* Genome-wide Screens Implicate Loss of Cullin Ring Ligase 3 in Persistent Proliferation

715 and Genome Instability in TP53-Deficient Cells. *Cell Reports* **31**, 107465 (2020).

716 40. Noordermeer, S. M. *et al.* The shieldin complex mediates 53BP1-dependent DNA repair. *Nature* **560**, 117

717 (2018).

718 41. Olivieri, M. *et al.* A Genetic Map of the Response to DNA Damage in Human Cells. *Cell* (2020)

719 doi:10.1016/j.cell.2020.05.040.

720 42. Wang, B. *et al.* Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. *Nature*

721 *Protocols* 1 (2019) doi:10.1038/s41596-018-0113-7.

722 43. Ma, S., Meng, Z., Chen, R. & Guan, K.-L. The Hippo Pathway: Biology and Pathophysiology. *Annual Review*

723 *of Biochemistry* **88**, 577–604 (2019).

724 44. Park, H. W. *et al.* Alternative Wnt Signaling Activates YAP/TAZ. *Cell* **162**, 780–794 (2015).

725 45. Yu, F.-X., Zhao, B. & Guan, K.-L. Hippo Pathway in Organ Size Control, Tissue Homeostasis, and Cancer.
726 *Cell* **163**, 811–828 (2015).
727 46. Nishimoto, M. *et al.* Transformation of normal cells by aberrant activation of YAP via cMyc with TEAD.
728 *Scientific Reports* **9**, 10933 (2019).
729 47. Calses, P. C., Crawford, J. J., Lill, J. R. & Dey, A. Hippo Pathway in Cancer: Aberrant Regulation and
730 Therapeutic Opportunities. *Trends in Cancer* **5**, 297–307 (2019).
731 48. Han, Y. Analysis of the role of the Hippo pathway in cancer. *Journal of Translational Medicine* **17**, 116
732 (2019).
733 49. Harvey, K. F., Zhang, X. & Thomas, D. M. The Hippo pathway and human cancer. *Nat Rev Cancer* **13**, 246–
734 257 (2013).
735 50. Jho*, M. K. & E. Cross-talk between Wnt/β-catenin and Hippo signaling pathways: a brief review. *BMB*
736 *Reports* **47**, 540–545 (2014).
737 51. van Soldt, B. J. & Cardoso, W. V. Hippo-Yap/Taz signaling: Complex network interactions and impact in
738 epithelial cell behavior. *WIREs Developmental Biology* **n/a**, e371 (2019).
739 52. Bae, J. S., Kim, S. M. & Lee, H. The Hippo signaling pathway provides novel anti-cancer drug targets.
740 *Oncotarget* **8**, (2017).
741 53. Moroishi, T., Hansen, C. G. & Guan, K.-L. The emerging roles of YAP and TAZ in cancer. *Nat Rev Cancer*
742 **15**, 73–79 (2015).
743 54. Nguyen, C. D. K. & Yi, C. YAP/TAZ Signaling and Resistance to Cancer Therapy. *Trends in Cancer* **5**, 283–
744 296 (2019).
745 55. Cheng, J. *et al.* A small-molecule inhibitor of UBE2N induces neuroblastoma cell death via activation of
746 p53 and JNK pathways. *Cell Death & Disease* **5**, e1079–e1079 (2014).

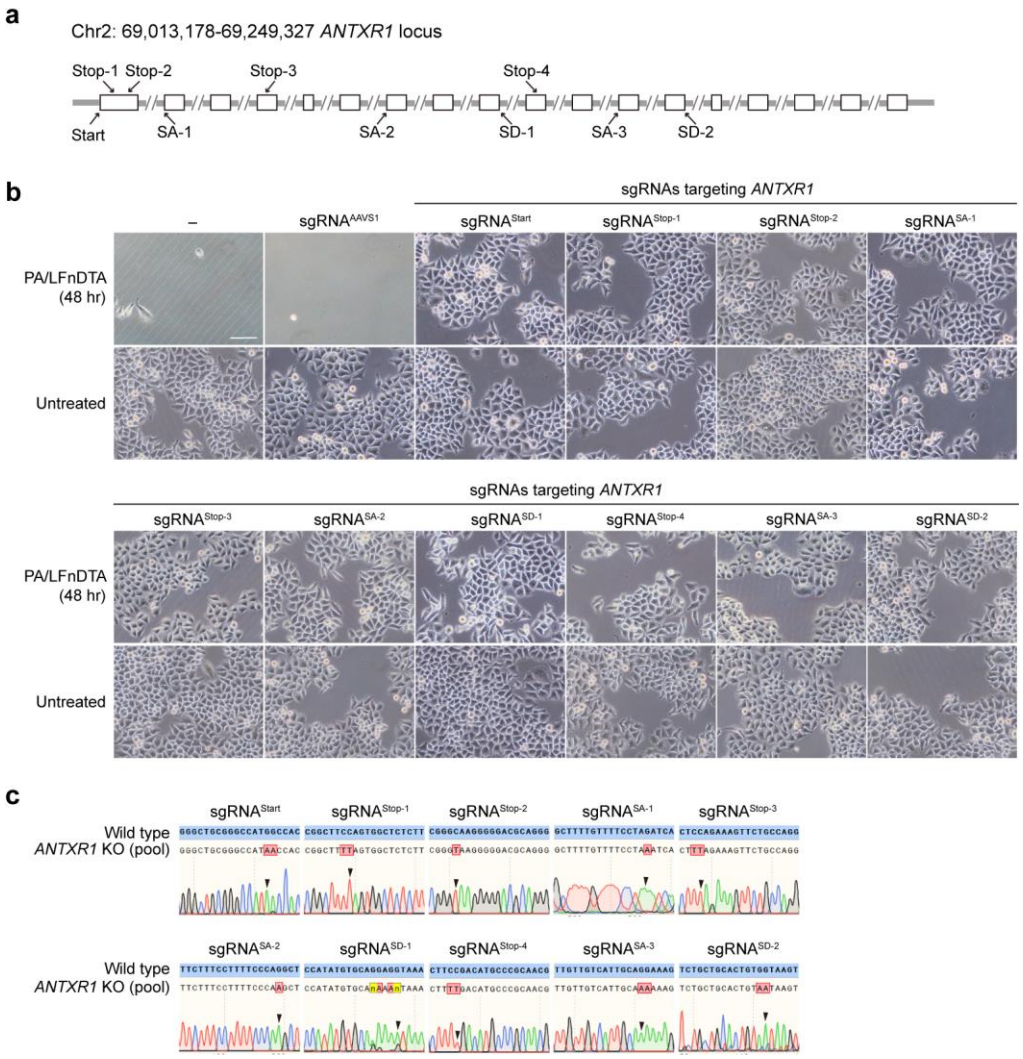
747 56. Francoz, S. *et al.* Mdm4 and Mdm2 cooperate to inhibit p53 activity in proliferating and quiescent cells in
748 vivo. *PNAS* **103**, 3232–3237 (2006).
749 57. Wang, H., Zhi, H., Ma, D. & Li, T. MiR-217 promoted the proliferation and invasion of glioblastoma by
750 repressing YWHAG. *Cytokine* **92**, 93–102 (2017).
751 58. Zhang, D., Zaugg, K., Mak, T. W. & Elledge, S. J. A Role for the Deubiquitinating Enzyme USP28 in Control
752 of the DNA-Damage Response. *Cell* **126**, 529–542 (2006).
753 59. Goodrich, D. W., Wang, N. P., Qian, Y.-W., Lee, E. Y.-H. P. & Lee, W.-H. The retinoblastoma gene product
754 regulates progression through the G1 phase of the cell cycle. *Cell* **67**, 293–302 (1991).
755 60. Rössig, L. *et al.* Akt-Dependent Phosphorylation of p21Cip1 Regulates PCNA Binding and Proliferation of
756 Endothelial Cells. *Molecular and Cellular Biology* **21**, 5644–5657 (2001).
757 61. Zhou, B. P. *et al.* Cytoplasmic localization of p21 Cip1/WAF1 by Akt-induced phosphorylation in HER-
758 2/neu -overexpressing cells. *Nature Cell Biology* **3**, 245–252 (2001).
759 62. Kreis, N.-N., Louwen, F. & Yuan, J. The Multifaceted p21 (Cip1/Waf1/CDKN1A) in Cell Differentiation,
760 Migration and Cancer Therapy. *Cancers* **11**, 1220 (2019).
761 63. Meng, Z. *et al.* MAP4K family kinases act in parallel to MST1/2 to activate LATS1/2 in the Hippo pathway.
762 *Nat Commun* **6**, 8357 (2015).
763 64. Doench, J. G. Am I ready for CRISPR? A user's guide to genetic screens. *Nature Reviews Genetics* **19**, 67–
764 80 (2018).
765 65. Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR–Cas9. *Nature*
766 *Reviews Genetics* **16**, 299–311 (2015).
767 66. Cheng, T.-L. *et al.* Expanding C–T base editing toolkit with diversified cytidine deaminases. *Nat Commun*
768 **10**, 1–10 (2019).

769 67. Huang, T. P. *et al.* Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of
770 base editors. *Nat Biotechnol* (2019) doi:10.1038/s41587-019-0134-y.
771 68. Jiang, W. *et al.* BE-PLUS: a new base editing tool with broadened editing window and enhanced fidelity.
772 *Cell Research* **28**, 855–861 (2018).
773 69. Kim, Y. B. *et al.* Increasing the genome-targeting scope and precision of base editing with engineered
774 Cas9-cytidine deaminase fusions. *Nature Biotechnology* **35**, 371–376 (2017).
775 70. Gehrke, J. M. *et al.* An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities.
776 *Nature Biotechnology* (2018) doi:10.1038/nbt.4199.
777 71. Grünewald, J. *et al.* CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat*
778 *Biotechnol* **37**, 1041–1048 (2019).
779 72. Li, X. *et al.* Base editing with a Cpf1–cytidine deaminase fusion. *Nature Biotechnology* **36**, 324–327 (2018).
780 73. Rees, H. A., Wilson, C., Doman, J. L. & Liu, D. R. Analysis and minimization of cellular RNA editing by DNA
781 adenine base editors. *Science Advances* **5**, eaax5717 (2019).
782 74. Wang, X. *et al.* Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. *Nature*
783 *Biotechnology* (2018) doi:10.1038/nbt.4198.
784 75. Wang, X. *et al.* Cas12a Base Editors Induce Efficient and Specific Editing with Low DNA Damage Response.
785 *Cell Reports* **31**, 107723 (2020).
786 76. Colic, M. *et al.* Identifying chemogenetic interactions from CRISPR screens with drugZ. *Genome Medicine*
787 **11**, 52 (2019).
788 77. Kolde, R., Laur, S., Adler, P. & Vilo, J. Robust rank aggregation for gene list integration and meta-analysis.
789 *Bioinformatics* **28**, 573–580 (2012).
790 78. Hart, T. *et al.* High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer

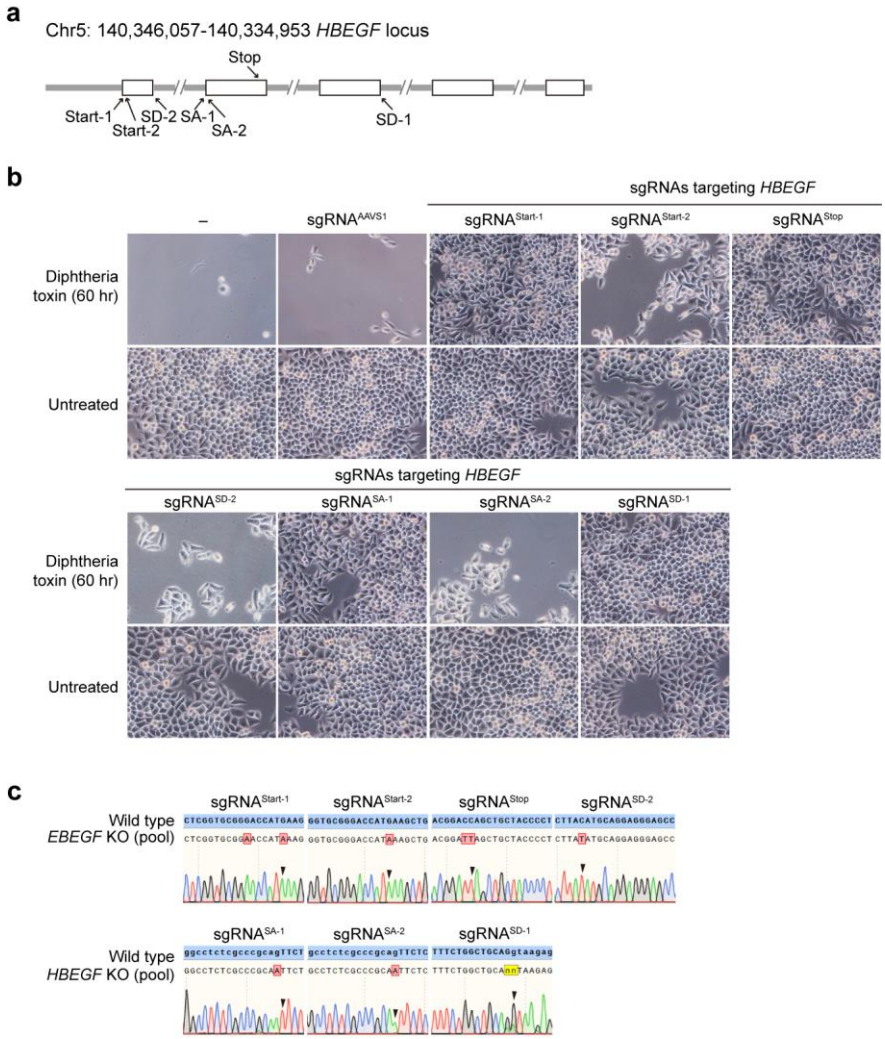
Liabilities. *Cell* **163**, 1515–1526 (2015).

79. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1–10 (2019).

Supplementary Figures

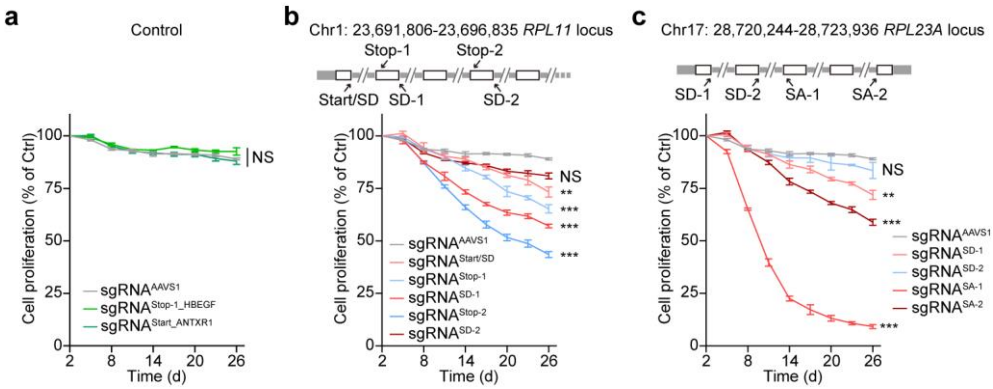


Supplementary Fig. 1 | Effect of *ANTXR1* deficiency by AncBE4max on PA/LFnDTA-triggered cytotoxicity in HeLa cells. a, Schematic indicates sgRNA targeting sites at *ANTXR1* genomic locus. **b**, Images of HeLa cells with or without PA/LFnDTA treatment for 48 hours after AncBE4max editing with indicated sgRNAs. The results shown are from one group of sgRNA transfected HeLa cells and conducted in triplicates with individual PA/LFnDTA toxin treatment. **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.

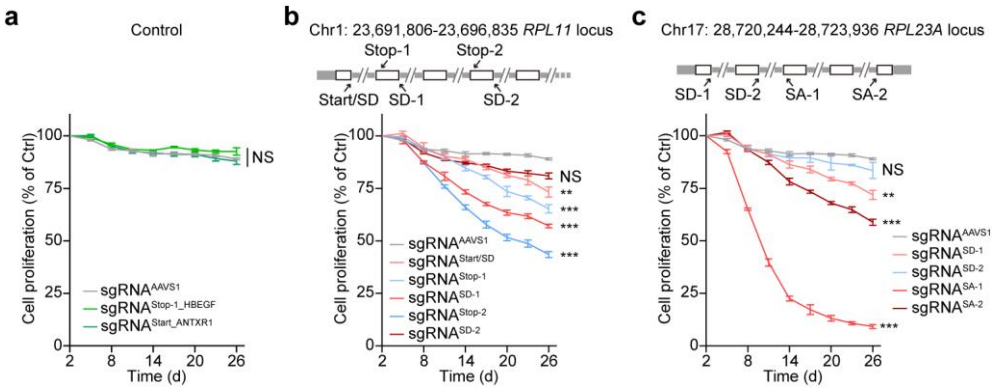


807
808

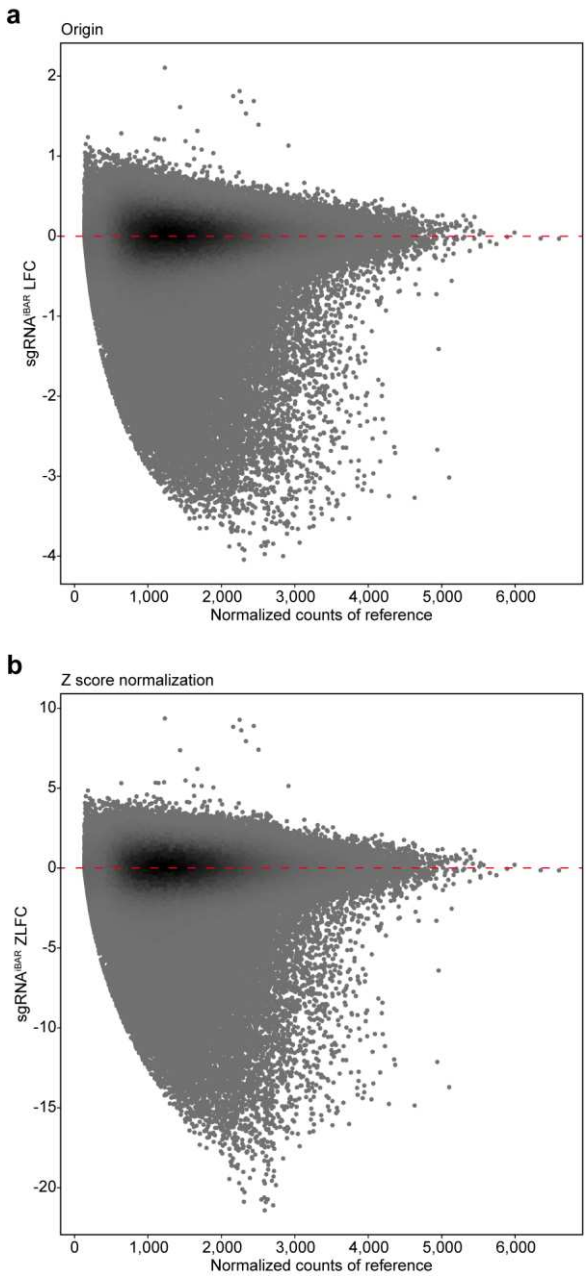
809 **Supplementary Fig. 2 | Effect of *HBEGF* deficiency by AncBE4max on diphtheria-toxin-**
810 **triggered cytotoxicity in HeLa cells. a**, Schematic indicates sgRNA targeting sites at *HBEGF*
811 genomic locus. **b**, Images of HeLa cells with or without DT treatment for 60 hours after AncBE4max
812 editing with indicated sgRNAs. The results shown are from one group of sgRNA transfected HeLa
813 cells and conducted in triplicates with individual DT treatment. **c**, Sanger sequencing chromatograms
814 of sgRNA-targeting *HBEGF* genomic fragments of DT resistant cells, black arrows indicate peaks of
815 targeted cytosines and their editing results.
816



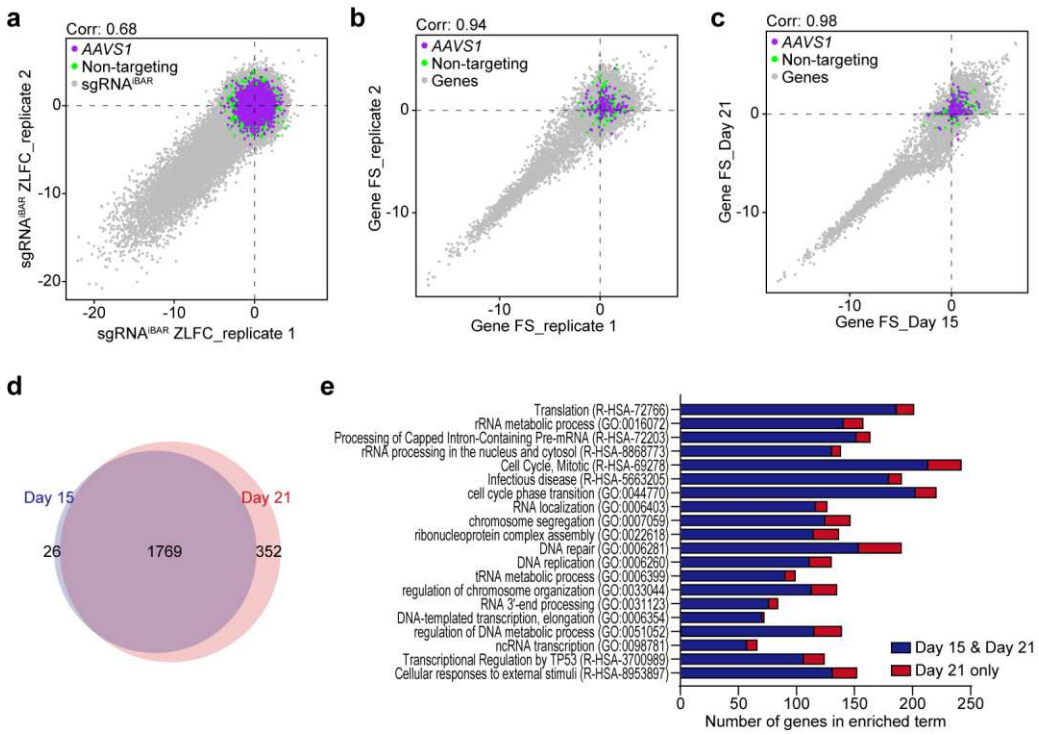
Supplementary Fig. 3 | Effect of indicated sgRNAs targeting ribosomal genes on cell proliferation in K562 cells. a, sgRNA^{Stop} targeting *HBEGF*, sgRNA^{Start} targeting *ANTXR1* and sgRNA^{AAVS1} served as negative control. **b**, Effects of indicated sgRNAs targeting ribosomal gene *RPL11* on cell proliferation in K562 cells. Results shown are from one experiment conducted in technical triplicates, and data are presented as the mean and s.d. of three technically independent experiments. *P*-values represent comparisons with sgRNA^{AAVS1} at the end point (day 26), calculated using a one-tailed Student's t-test and adjusted using the Benjamini–Hochberg method. $^{**}p < 0.01$; $^{***}p < 0.001$. c, Effects of indicated sgRNAs targeting ribosomal gene *RPL23A* on cell proliferation in K562 cells.



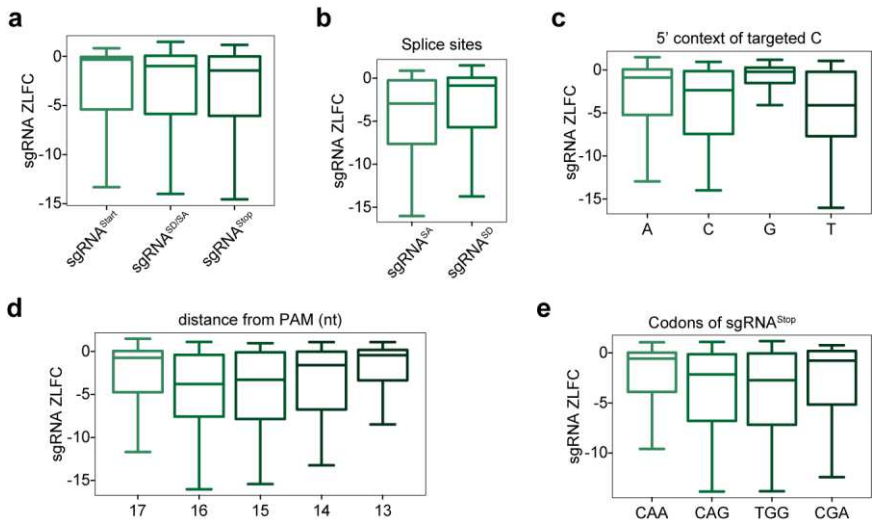
Supplementary Fig. 4 | Information of sgRNA^{iBARs} and BARBEKO library. a, Schematic shows the scaffold sequence of sgRNA^{iBAR}, in which 4 iBARs employed in BARBEKO library are highlighted in red. **b**, Pie chart shows the composition of BARBEKO library that newly designed sgRNA^{Start} and sgRNAs 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%.



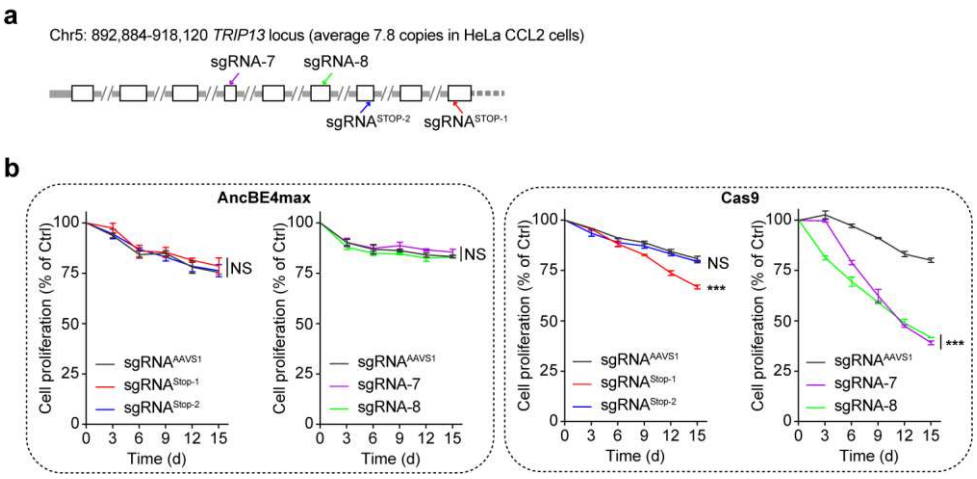
Supplementary Fig. 5 | Z-score normalization of sgRNA^{iBAR} log fold change. a and **b**, Density distribution of sgRNA^{iBAR} log fold change (LFC) (**a**) and Z score of sgRNA^{iBAR} LFC (ZLFC) (**b**). Z-score normalization of LFC helps to normalize varying degrees of fold change at different abundance of sgRNAs^{iBAR} in reference group.



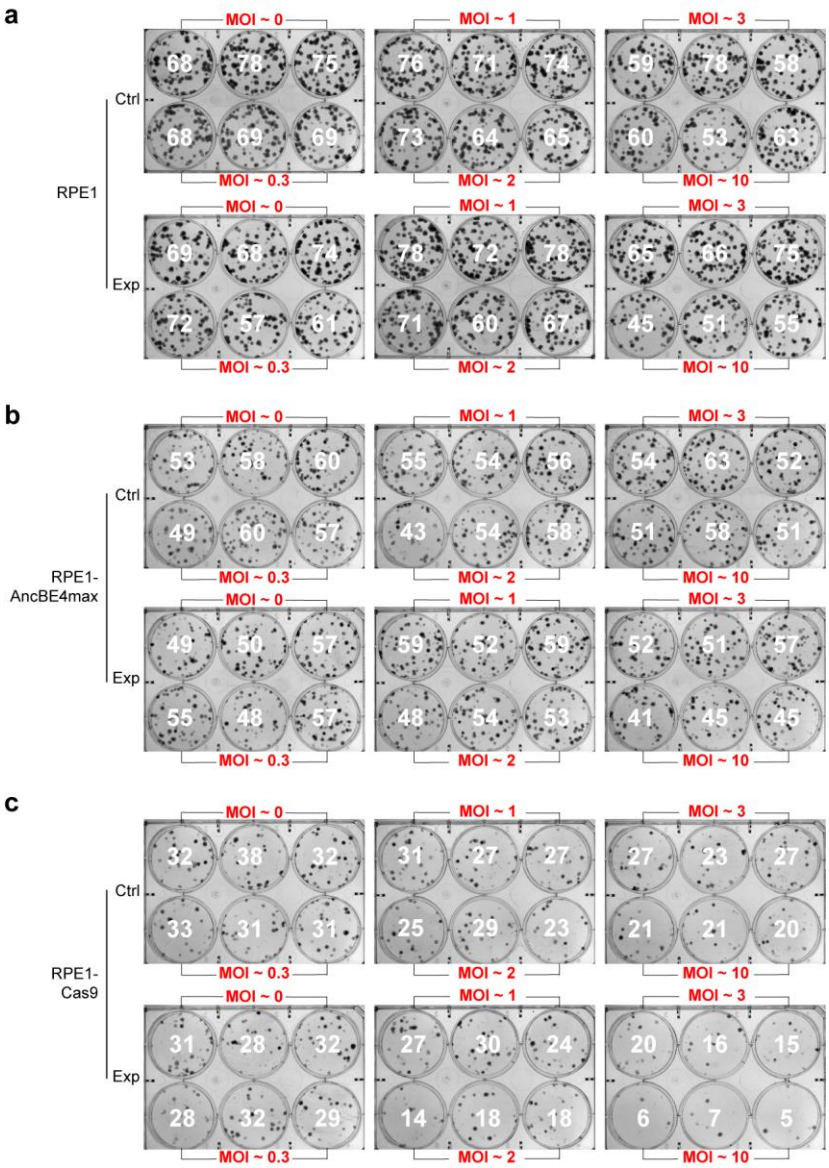
Supplementary Fig. 6 | Comparison of depleted hits between timepoints during fitness screening in HeLa cells. a, Scatter plot of sgRNA^{iBAR} ZLFC of two biological replicates on Day 15, Pearson correlation coefficient is indicated on the top. sgRNAs^{iBAR} targeting AAVS1 locus and non-targeting sgRNAs^{iBAR} as negative control are labelled in purple and green. **b**, Scatter plot of gene Fitness Score (FS) on Day 15 of two biological replicates, Pearson correlation coefficient is indicated on the top. **c**, Scatter plot of gene FS of Day 15 and Day 21, Pearson correlation coefficient is indicated on the top. **d**, Venn diagram shows number of common and different depleted hits of Day 15 and Day 21. **e**, Gene Ontology (GO) analysis of common and Day 21-only selected genes. Analyzed by Metascape[79], GO terms are ranked from top to bottom based on *p*-value of Day 21 results. Blue bars represent number of commonly depleted genes and red bars represent number of Day 21-only selected genes in each GO terms.



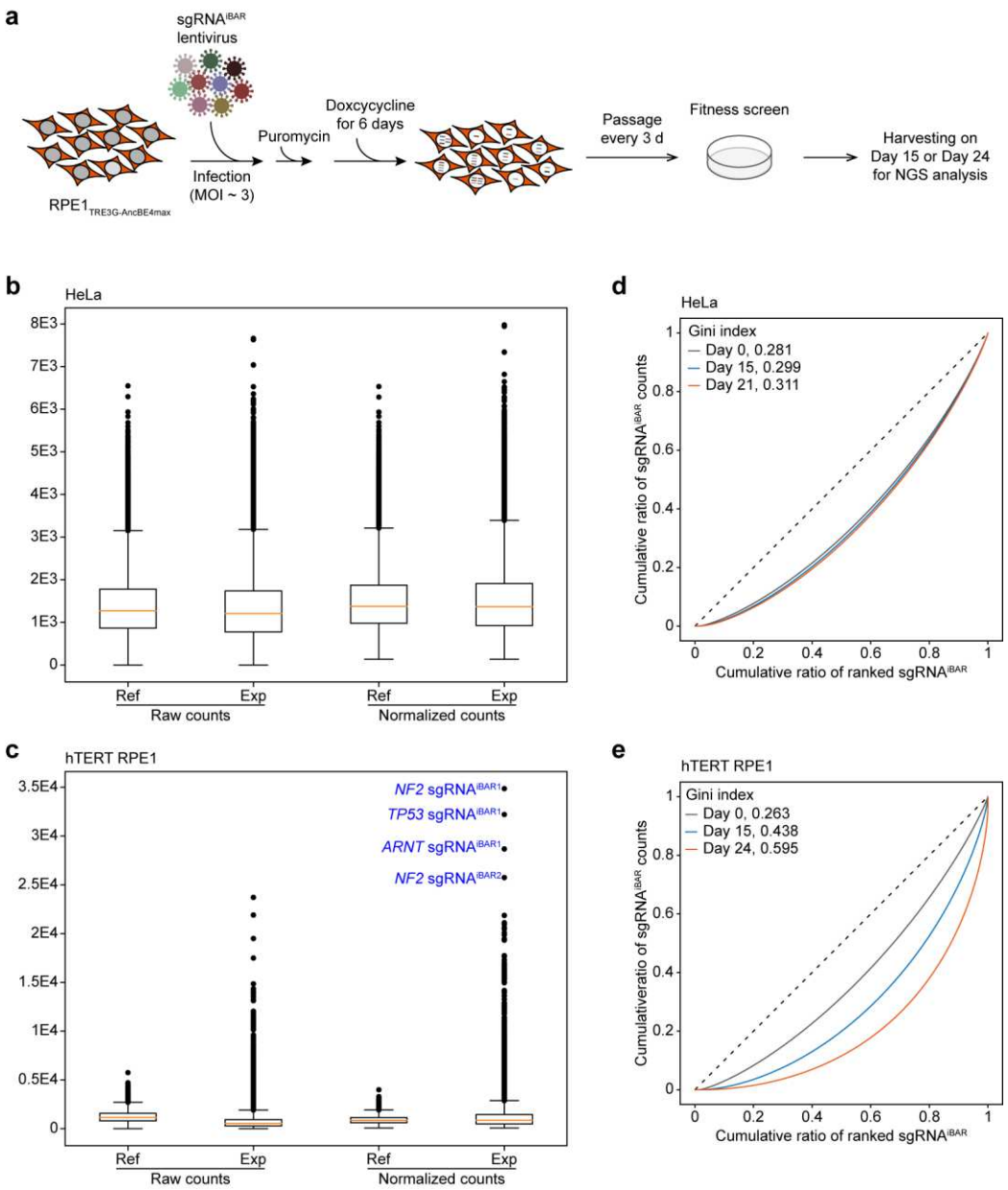
Supplementary Fig. 7 | Efficiency comparison among different types of sgRNAs. a, Efficiency comparison across 3 types of sgRNAs, sgRNA^{Start} targeting start codons, sgRNA$^{SD/SA}$ targeting splice sites and sgRNA^{Stop} targeting codons of Gln (CAA, CAG), Arg (CGA) and Trp (TGG). **b**, Efficiency 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.



Supplementary Fig. 8 | Effect of *TRIP13* deficiency by AncBE4max or Cas9 on cell proliferation in HeLa cells. a, Schematic shows 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. **b**, 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. $sgRNA^{AAVS1}$ served as negative control. *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, ***$p < 0.001$.



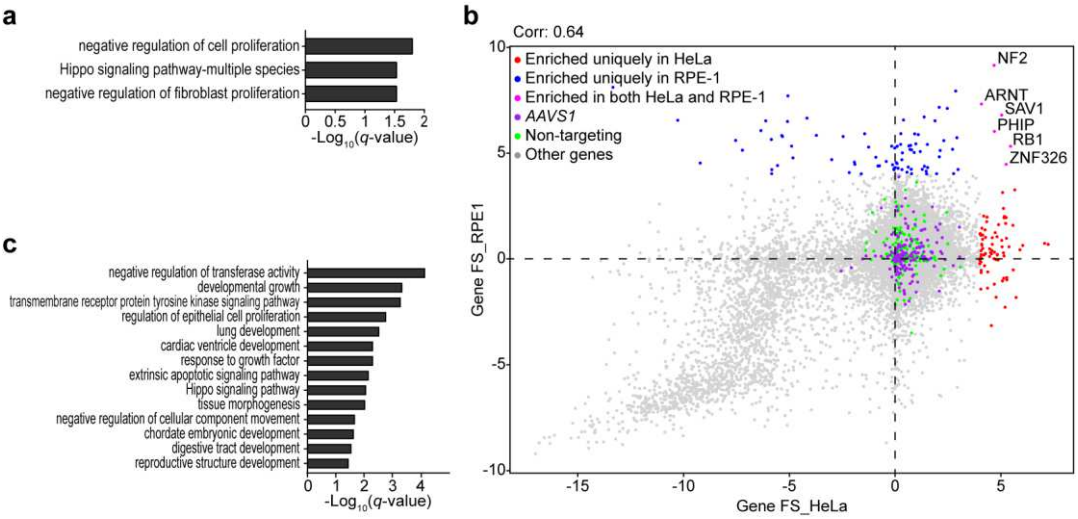
Supplementary Fig. 9 | Clonogenic survival assay of RPE1 cells in response to sgRNA library transduction at gradient MOIs. Images of clonal formation for control and experimental libraries 12 days post transduction in wild-type (**a**), AncBE4max-expressing (**b**) and Cas9-expressing (**c**) RPE1 cells. The multiplicity of infection (MOI) is indicated in red, and the number of clones counted manually is indicated in the middle of wells in white.



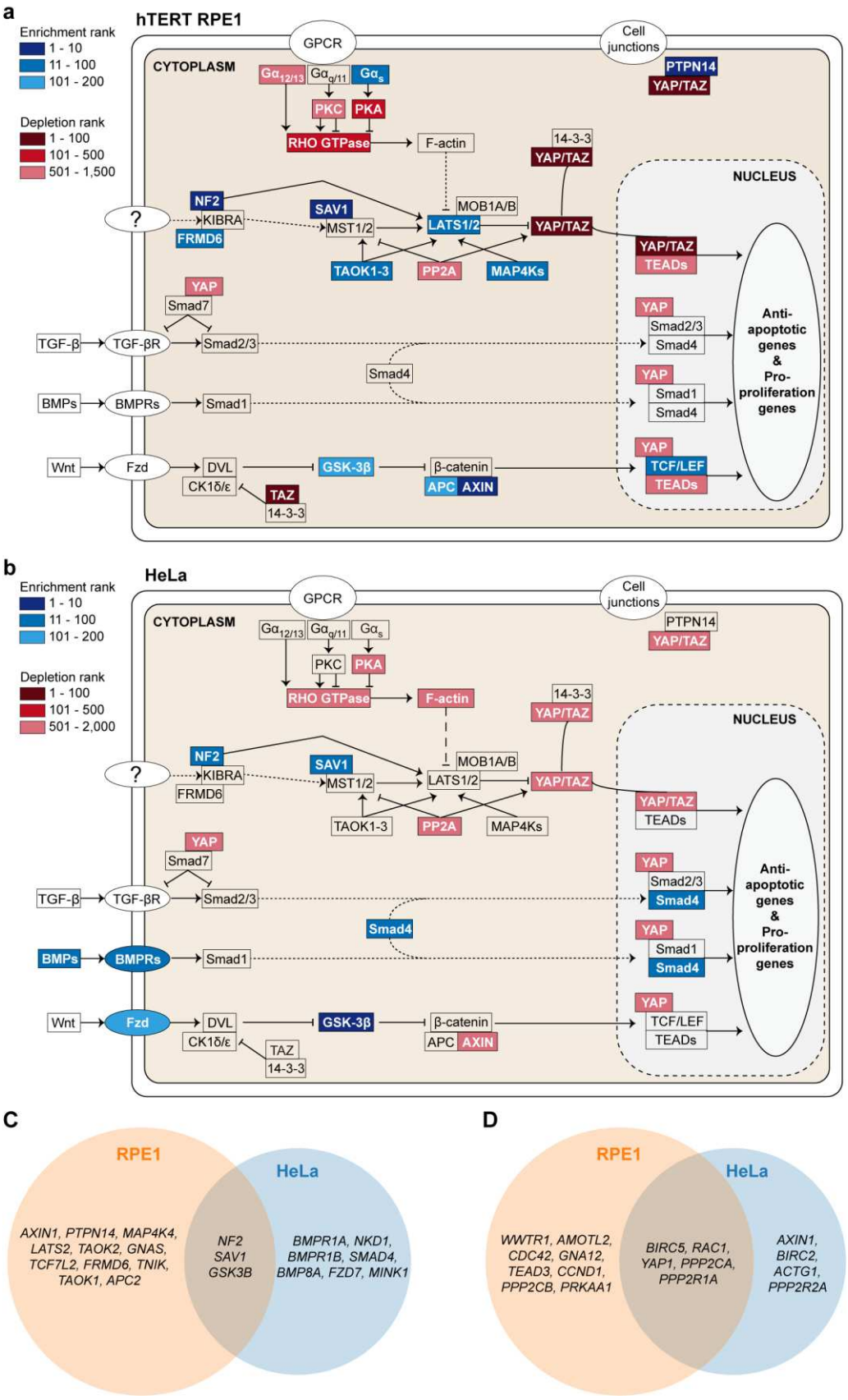
892
893

894 **Supplementary Fig. 10 | Comparisons of distribution of sgRNA^{iBARs} counts between HeLa and**
895 **RPE1 screens. a**, Workflow of gene fitness screens in wild-type RPE1 cells. AncBE4max-
896 expressing RPE1 cells were infected by lentiviruses library of BARBEKO at a MOI of ~ 3, then
897 reference cells denoted by Day 0 were harvested 9 days post infection, and experimental groups were
898 harvest on Day 15 and Day 24. **b**, Boxplot shows distribution of raw and normalized counts of
899 sgRNAs^{iBAR} in HeLa reference and experimental group. **c**, Boxplot shows distribution of raw and
900 normalized counts of sgRNAs^{iBAR} in RPE1 reference and experimental group. The top4 enriched
901 sgRNAs^{iBAR} in experimental group are labelled in blue. **d**, Lorenz curve of sgRNA^{iBAR} counts on

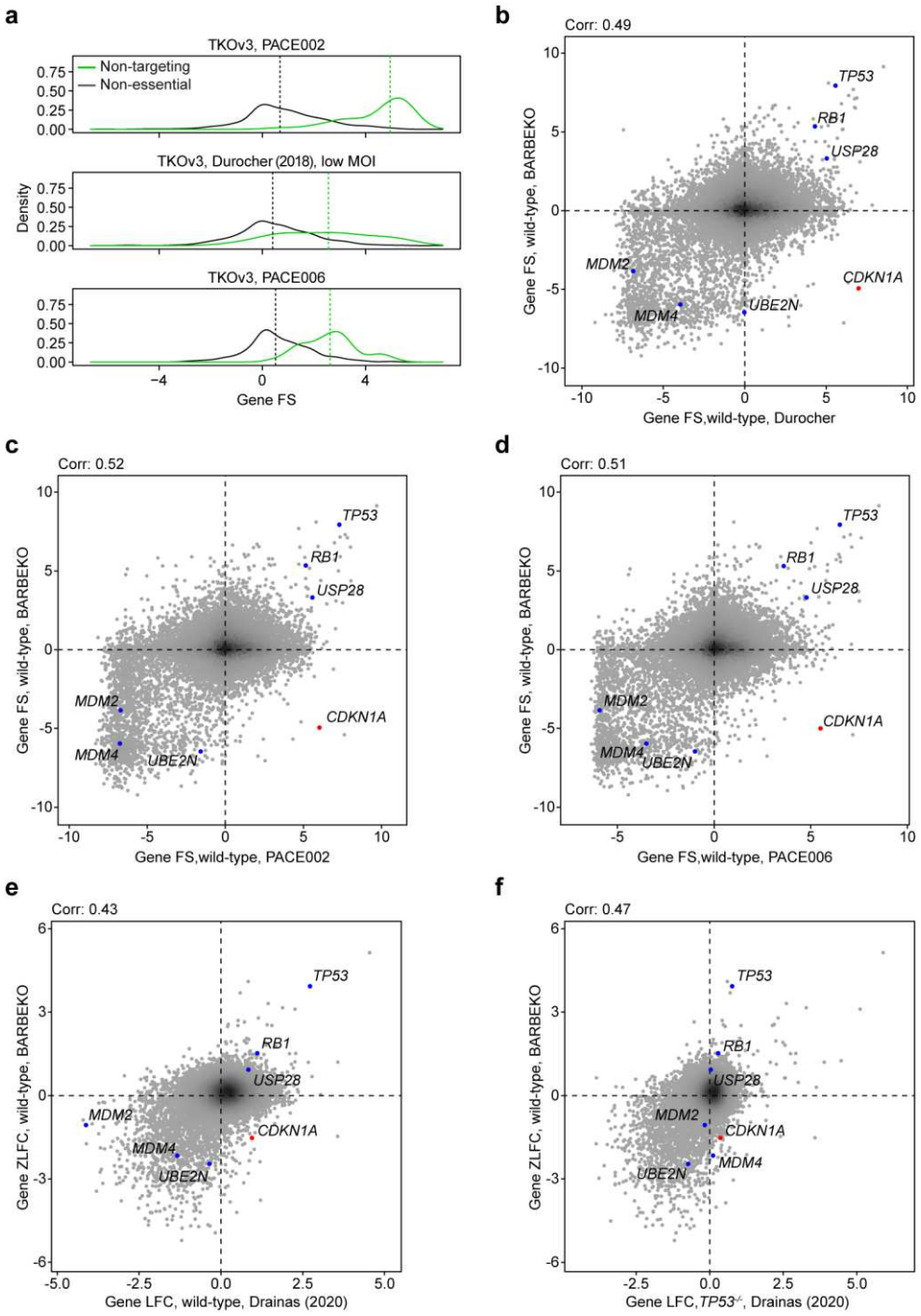
Day 0, Day 15 and Day 21 in HeLa fitness screen, sgRNAs^{iBAR} are ranked by counts from lowest to highest. Gini index is indicated in the top left corner. **e**, Lorenz curve of sgRNA^{iBAR} counts on Day 0, Day 15 and Day 24 in RPE1 fitness screen, sgRNAs^{iBAR} are ranked by counts from lowest to highest. Gini index is indicated in the top left corner.



Supplementary Fig. 11 | Comparisons of enriched genes between HeLa and RPE1 screens. a and **c**, GO enrichment analysis by Metascape of enriched genes in RPE1 (A) and HeLa (C), GO Terms of q-value ≤ 0.05 are exhibited by barplots. **b**, Scatter plot displays gene Fitness Score of HeLa and RPE1. Genes enriched in both screens are highlighted, in which red dots represent genes enriched uniquely in HeLa, blue dots represent genes enriched uniquely in RPE1, while pink dots with symbols represent genes commonly enriched in these two cell lines. Pearson correlation coefficient is indicated on the top.

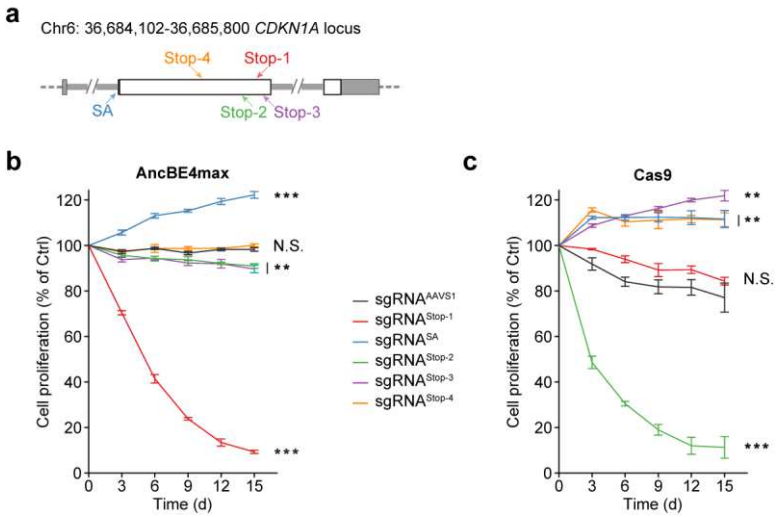

a
hTERT RPE1
Enrichment rank
1 - 10
11 - 100
101 - 200
Depletion rank
1 - 100
101 - 500
501 - 1,500
GPCR
Cell junctions
CYTOPLASM
Gα12/13
Gαq/11
Gαs
PKC
PKA
RHO GTPase
F-actin
PTPN14
YAP/TAZ
14-3-3
YAP/TAZ
NUCLEUS
NF2
SAV1
MOB1A/B
?
KIBRA
MST1/2
LATS1/2
YAP/TAZ
FRMD6
TAOK1-3
PP2A
MAP4Ks
YAP/TAZ
TEADs
YAP
Smad7
YAP
Smad2/3
Smad4
TGF-β
TGF-βR
Smad2/3
Smad4
YAP
Smad1
Smad4
BMPs
BMPRs
Smad1
Anti-apoptotic genes & Pro-proliferation genes
YAP
TCF/LEF
TEADs
Wnt
Fzd
DVL
CK1δ/ε
GSK-3β
β-catenin
APC
AXIN
TAZ
14-3-3
b
HeLa
Enrichment rank
1 - 10
11 - 100
101 - 200
Depletion rank
1 - 100
101 - 500
501 - 2,000
GPCR
Cell junctions
CYTOPLASM
Gα12/13
Gαq/11
Gαs
PKC
PKA
RHO GTPase
F-actin
PTPN14
YAP/TAZ
14-3-3
YAP/TAZ
NUCLEUS
NF2
SAV1
MOB1A/B
?
KIBRA
MST1/2
LATS1/2
YAP/TAZ
FRMD6
TAOK1-3
PP2A
MAP4Ks
YAP/TAZ
TEADs
YAP
Smad7
YAP
Smad2/3
Smad4
TGF-β
TGF-βR
Smad2/3
Smad4
YAP
Smad1
Smad4
BMPs
BMPRs
Smad1
Anti-apoptotic genes & Pro-proliferation genes
YAP
TCF/LEF
TEADs
Wnt
Fzd
DVL
CK1δ/ε
GSK-3β
β-catenin
APC
AXIN
TAZ
14-3-3
C
RPE1
HeLa
AXIN1, PTPN14, MAP4K4, LATS2, TAOK2, GNAS, TCF7L2, FRMD6, TNIK, TAOK1, APC2
NF2 SAV1 GSK3B
BMPR1A, NKD1, BMPR1B, SMAD4, BMP8A, FZD7, MINK1
D
RPE1
HeLa
WWTR1, AMOTL2, CDC42, GNA12, TEAD3, CCND1, PPP2CB, PRKAA1
BIRC5, RAC1, YAP1, PPP2CA, PPP2R1A
AXIN1, BIRC2, ACTG1, PPP2R2A

Supplementary Fig. 12 | Disruptions in Hippo signaling pathway influenced cell viability of RPE1 and HeLa cells. a, Genes of Hippo signal pathway were highly enriched in RPE1 fitness screen. Schematic shows rankings of genes or regulators of Hippo signal pathway, in which enriched genes are represented in blue rectangles while depleted genes are represented in red rectangles. Pathway map is modified from KEGG-Hippo signaling pathway (map04390). **b**, Schematic shows rankings of genes or regulators of Hippo signal pathway in HeLa screening, in which enriched genes are represented in blue rectangles while depleted genes are represented in red rectangles. **c**, Enriched Hippo-related genes in RPE1 and HeLa cell fitness screens. **d**, Depleted Hippo-related genes in RPE1 and HeLa cell fitness screens.



Supplementary Fig. 13 | Comparison of BARBEKO and CRISPR-Cas9 screens in RPE1 cells. **a**, Comparisons between non-targeting "genes" (green curves) and non-essential genes (grey curves) of the density distribution of Fitness Score in TKOv3 screens. Dotted lines represent the median FS of each categories. **b-d**, Scatter plot shows gene Fitness Score comparisons between BARBEKO and three TKOv3 screens in *TP53* wild-type RPE1 cells. *TP53* and p53 signaling genes are highlighted in blue, while *CDKN1A* is especially highlighted in red. Pearson correlation coefficient is indicated on

940 the top. **e-f**, Scatter plot shows comparisons between gene ZLFC of BARBEKO and gene LFC of
941 screens from Drainas *et al.*[39] in p53-proficient (**e**) and -deficient RPE1 cells (**f**).
942



Supplementary Fig. 14 | 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. 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, $^{**}p < 0.01$, $^{***}p < 0.001$.

953 **Supplementary Tables**

954 **Supplementary Table 1. List of primers used in this manuscript**

955 **Supplementary Table 2. List of sgRNA sequences for individual experiments**

956 **Supplementary Table 3. sgRNA scoring table for library design**

957 **Supplementary Table 4. Information of BARBEKO sgRNA library**

958 **Supplementary Table 5. Genetic screen results of HeLa, K562 and RPE1 by BARBEKO**

959 **strategy**

960 **Supplementary Table 6. Genetic screen results of HeLa by CRIPSR^{iBAR} Strategy**

961 **Supplementary Table 7. List of 1,000 non-targeting sgRNAs**

Figures

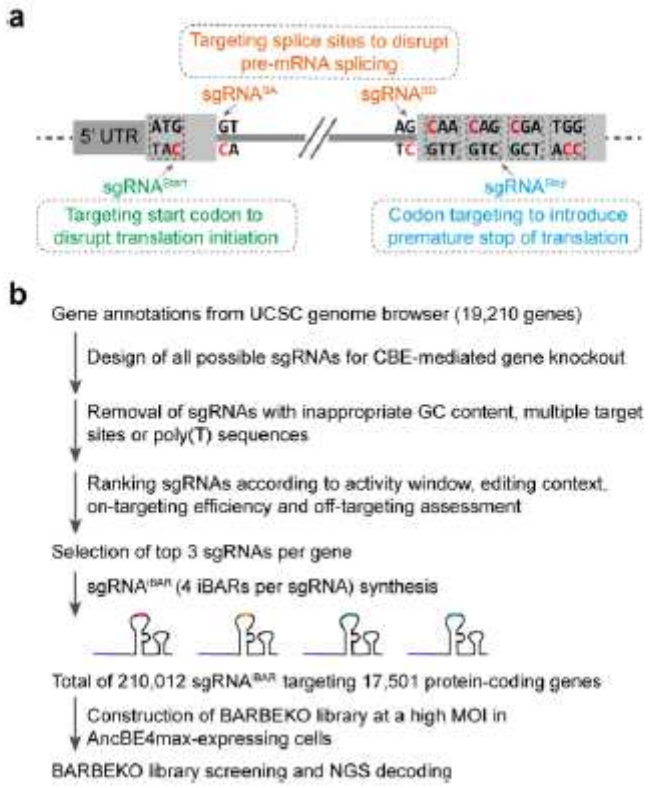


Figure 1

Design of CBE-based genome-scale sgRNA library for gene knockout screens. a, CBE with sgRNAs targeting start codons (sgRNAStart), splice acceptor sites (sgRNASA), splice donor sites (sgRNASD) and codons of Gln, Arg or Trp (sgRNAStop) disrupts gene functions. b, The flow chart depicts selection and filtration of sgRNAsiBAR for BARBEKO library.

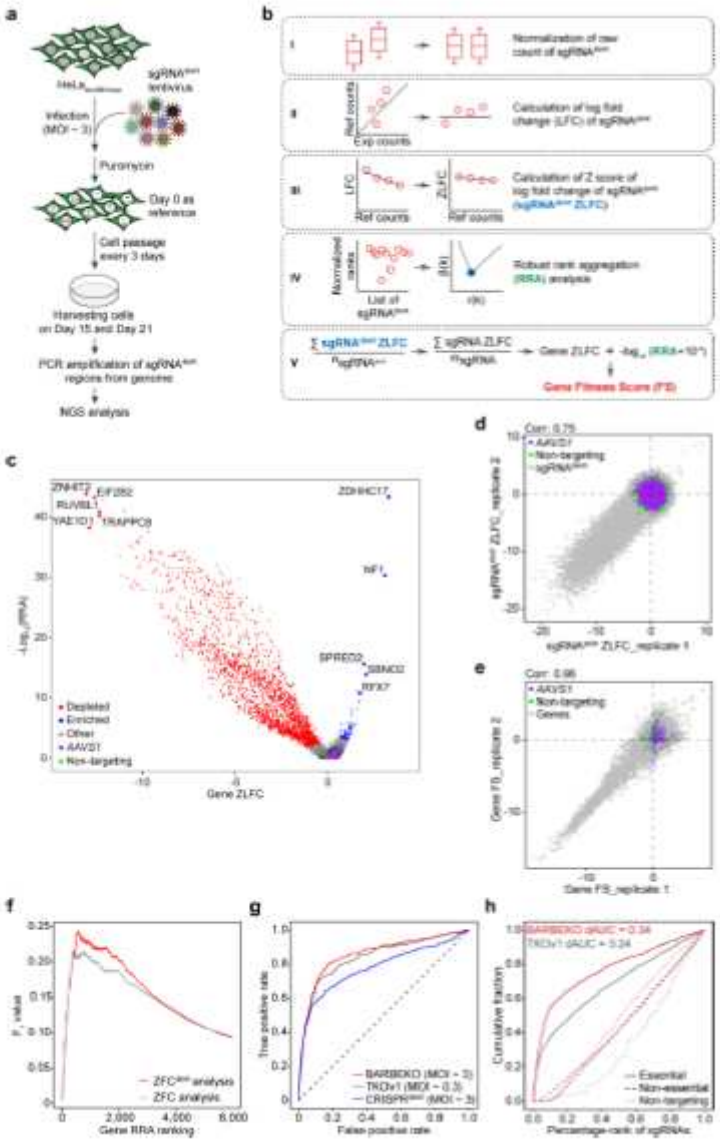


Figure 2

BARBEKO strategy achieved high-throughput gene fitness screens in HeLa cells. a, Workflow of gene fitness screen in HeLa cells. AncBE4max-expressing HeLa cells were infected by lentiviruses library of BARBEKO at a MOI of ~ 3, then reference cells denoted by Day 0 were harvested 5 days post infection, and experimental groups were harvest on Day 15 and Day 21. b, The schematics of ZFC algorithm describes the analysis processes of NGS data from BARBEKO screens. Gene Fitness Score (FS) is an integrated index of Z score of Log2(fold change) (ZLFC) and the value of RRA. c, Volcano plot shows overall outcome of gene fitness screen in HeLa cells by BARBEKOanalyzed by ZFCiBAR. Depleted and enriched genes in screen are plotted in red and blue respectively, and the top5 genes in both directions are labelled individually. Every 3 sgRNAs targeting AAVS1 and non-targeting sgRNAs were randomly grouped as “genes”, plotting in purple and green. d, Scatter plot of sgRNAiBAR ZLFC of two biological replicates, Pearson correlation coefficient is indicated on the top. sgRNAiBARs targeting AAVS1 locus and non-targeting sgRNAiBARs as negativecontrol are labelled in purple and green. e, Scatter plot of gene Fitness

Score (FS) of two biological replicates, Pearson correlation coefficient is indicated on the top. f, Comparison of F1 value against gene RRA ranking when considering iBARs as internal replicates or ignoring iBARs for ZFC analysis. F1 value is determined by gene gold-standard sets. g, Receiver operating characteristic (ROC) analysis of depleted genes for BARBEKO library at MOI ~3 (considering iBARs as internal replicates in analysis), CRISPRiBAR library at a MOI of ~ 3 (ignoring iBARs in analysis) and TKOv1 library at a MOI of ~ 0.3 78 screened in HeLa cells. h, Comparison of AUCs for essential (solid line), non-essential (dashed line), and non-targeting (dotted line) sgRNAs between BARBEKO a ta MOI of ~ 3 and TKOv1 at a MOI of ~ 0.3 screened in HeLa cells. The dAUCs value from essential and non-essential sgRNAs is indicated in the upper left corner.

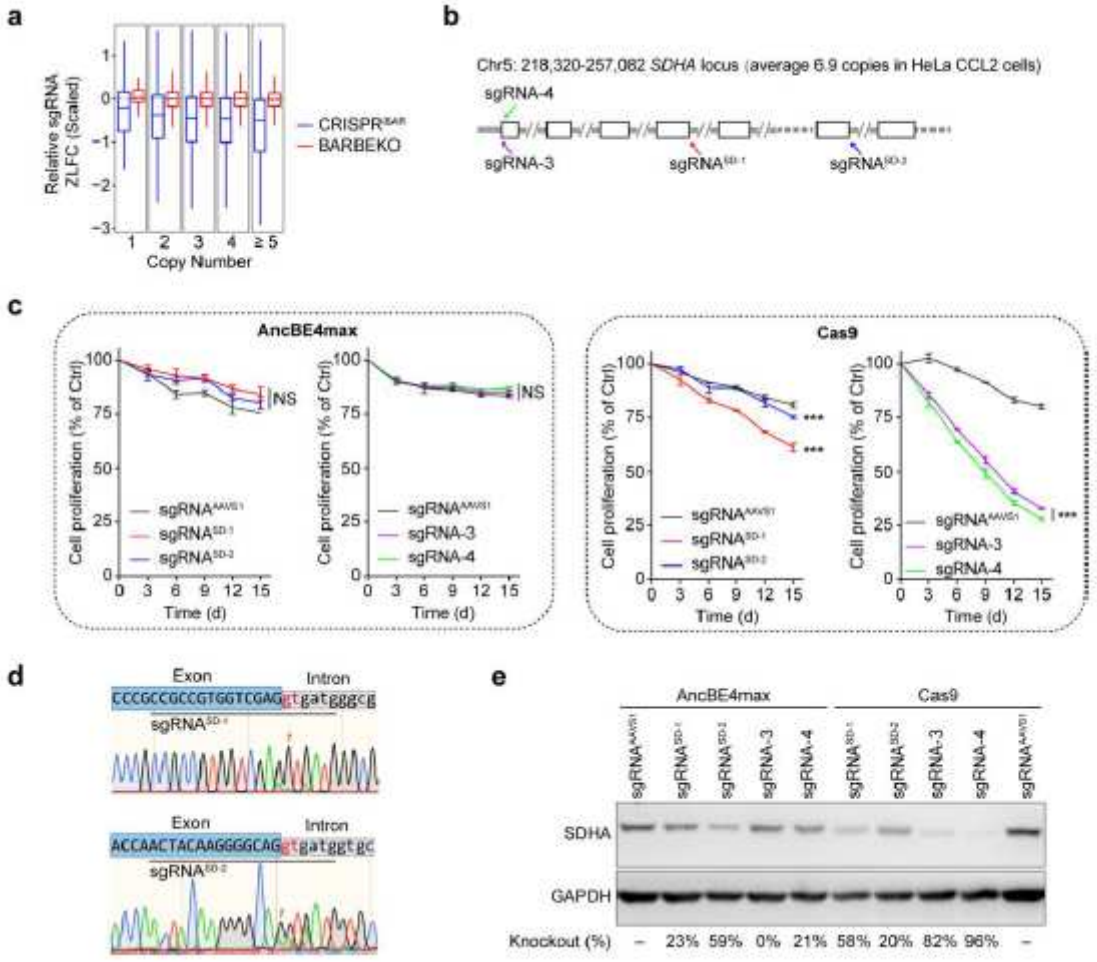


Figure 3

Copy number effect could be diminished in BARBEKO screens. a, Boxplot diagram shows relative sgRNA ZLFC of BARBEKO and CRISPRiBAR screens at a MOI of ~ 3 according to gene copy number in HeLa cells. ZLFC of sgRNAs targeting protein-coding genes subtracted the median ZLFC of non-targeting sgRNAs serving as relative sgRNA ZLFC. b, Schematic shows genomic region of a highly amplified gene SDHA and the targeting sites of sgRNAs selected from BARBEKO (sgRNASD-1 and sgRNASD-2) or TKO (sgRNA-3 and sgRNA-4) libraries. c, Effects of indicated sgRNAs targeting SDHA on cell proliferation in HeLa cells. 4 sgRNAs were individually delivered into AncBE4max- and Cas9-expressing cells for

validation, in which sgRNASD-1 and sgRNASD-2 targeting splice donor sites of SDHA. sgRNAAAVS1 served as negative control. P-values represent comparisons with sgRNAAAVS1 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. d, The sanger sequencing chromatograms of sgRNASD-1 and sgRNASD-2 targeting splice donor sites of SDHA genomic region after AncBE4max editing. The orange arrows indicate peaks of targeted "Gs" in splice donor sites. e, Immunoblot 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 sgRNAAAVS1 group.

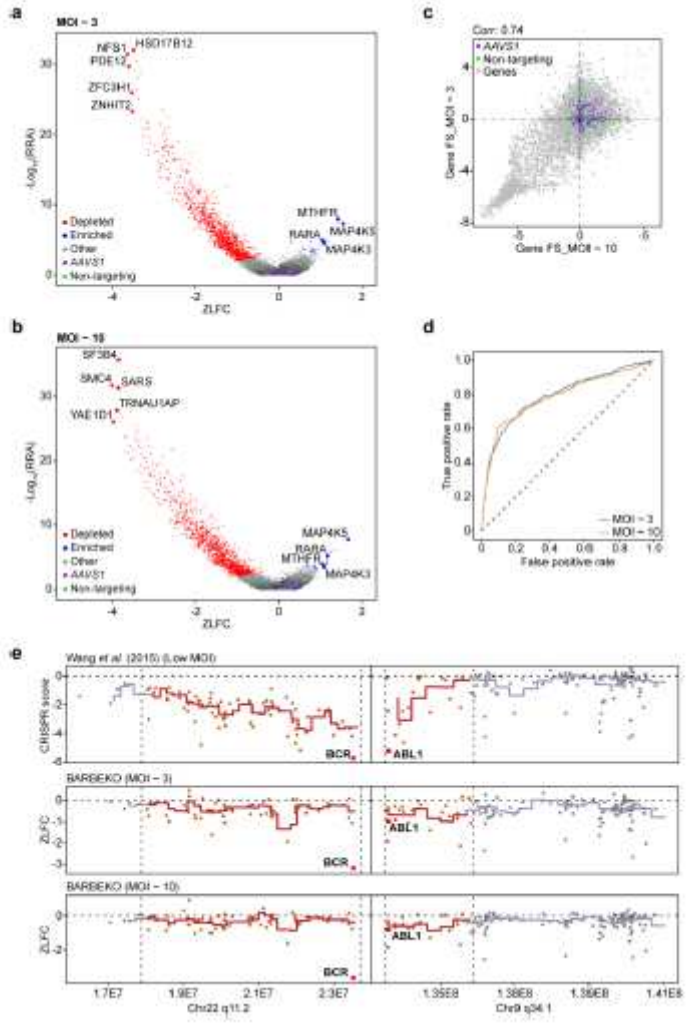


Figure 4

BARBEKO Empowers Gene Fitness Screen in K562 Cells at MOI ~3 and MOI ~10. a and b, Volcano plot shows overall outcomes of gene 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 top5 depleted genes and commonly enriched genes of both screens are labelled individually. Every 3 sgRNAs targeting AAVS1 and non-targeting sgRNAs were randomly grouped as "negative control genes", plotting in purple and green. c, Scatter plot of gene FS in screens at MOIs of ~ 3 and ~ 10, Pearson correlation coefficient is indicated on the top. d, Receiver operating characteristic (ROC) analysis of depleted genes in screens at MOIs ~ 3 and

~ 10 according to essential genes and non-essential genes of gold-standard sets. e, ZLFC or CRISPR Scores of genes locating around BCR-ABL fusion gene are plotted sequentially. Data from Wang et al. is shown in the top lane 8, and the results of BARBEKO at MOIs of ~ 3 and ~ 10 in K562 are shown in the middle and bottom. 610 Genes in this region are separated into bins to calculate mean of ZLFC or CRISPR Score, which are represented by solid lines. The high-copy tandem amplified region is indicated in red, and the flanking regions are in grey.

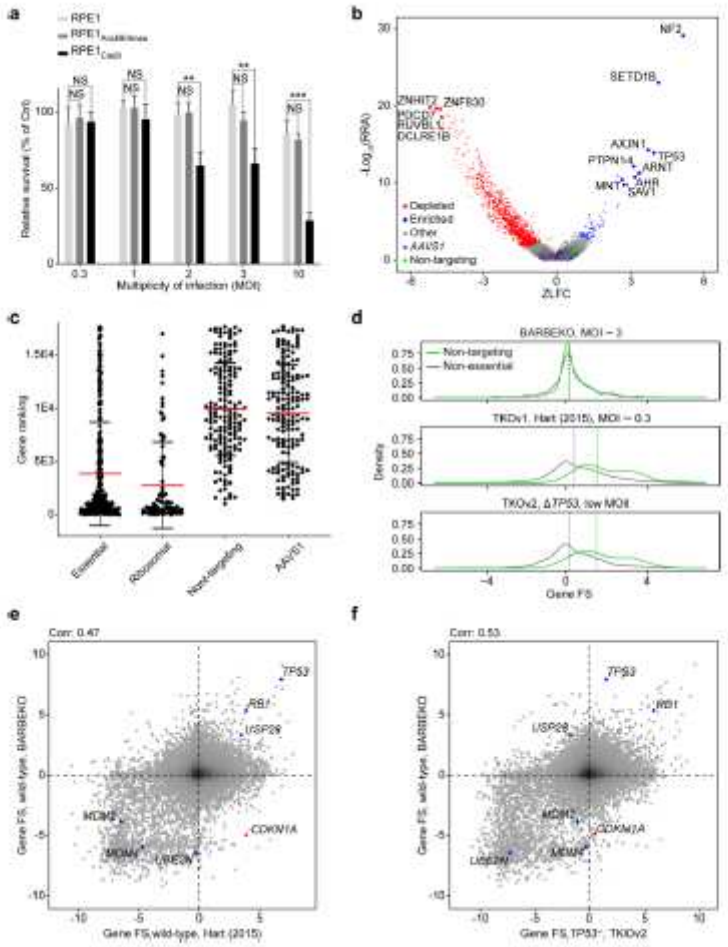


Figure 5

BARBEKO enables bidirectional screens of RPE1 cells at a high MOI. a, Clonogenic survival of RPE1 cells in response to sgRNA library transduction at gradient MOIs. Non-targeting 6 control library (1,000 sgRNAs) and non-essential-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, 10. Clonogenic assay was performed in triplicates 3 days post infection, and the survival fraction (SF) of experimental group was normalized by control SF to calculate the relative percentage. Data are presented as the mean and s.d., **p < 0.01; ***p < 0.001. b, Volcano plot shows overall outcome of gene fitness screen in RPE1 cells by BARBEKO at a MOI of ~ 3. The top5 depleted genes and top9 enriched gene were labelled individually. c, Scatter plot shows distribution of gene rankings of 4 different categories. Essential genes and ribosomal genes were extracted from reference gene sets, while non-targeting and AAVS1 “genes” represented the

mean rankings of 3 sgRNAs by randomly sampling. d, Comparisons between non-targeting “genes” (green curves) and non-essential genes (grey curves) of the Fitness Score density distribution. Data from TKO screens were re-analyzed by ZFC, and their sgRNAs targeting EGFP, LacZ and luciferase were considered as non-targeting to human genome. e and f, Scatter plot shows comparisons of gene Fitness Score between BARBEKO and TKOv1 screens in TP53 wild-type RPE1 cells (c); comparisons between BARBEKO in TP53 wild-type RPE1 and TKOv2 in TP53KO RPE1 cells (d). TP53 and p53 signaling genes are highlighted in blue, while CDKN1A is especially highlighted in red. Pearson correlation coefficient is indicated on the top.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.xlsx
- SupplementaryTable2.xlsx
- SupplementaryTable3.xlsx
- SupplementaryTable4.xlsx
- SupplementaryTable5.xlsx
- SupplementaryTable6.xlsx
- SupplementaryTable7.xlsx
- FigS1.png
- FigS2.png
- FigS3.png
- FigS4.png
- FigS5.png
- FigS6.png
- FigS7.png
- FigS8.png
- FigS9.png
- FigS10.png
- FigS11.png
- FigS12.png
- FigS13.png
- FigS14.png