bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Optimized Cas9 expression improves performance of large-scale CRISPR screening

João M. Fernandes Neto[1,3], Katarzyna Jastrzebski[1,3], Cor Lieftink[1,2], Lenno Krenning[3,4], Matheus Dias[1,3], Ben Morris[1,2], Daimy van der Ven[1,3], Hester Heimans[1,3], René H. Medema[3,4], René Bernards[1,3] and Roderick L. Beijersbergen[1,2#]

[1]Division of Molecular Carcinogenesis, [2]The NKI Robotics and Screening Center, [3]Oncode Institute, [4]Division of Cell Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.

[#]Corresponding author:
Roderick Beijersbergen
Division of Molecular Carcinogenesis
The Netherlands Cancer Institute
Plesmanlaan 121
1066 CX Amsterdam
The Netherlands
Email: r.beijersbergen@nki.nl
Phone +31205121960

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

CRISPR technology is an invaluable tool for large-scale functional genomic screening. Genome editing efficiency and timing are important parameters impacting the performance of pooled CRISPR screens. Here we show that by optimizing Cas9 expression levels, the time necessary for gene editing can be reduced contributing to improved performance of CRISPR based screening.

INTRODUCTION

The simplicity, speed and low-cost of CRISPR technology has led to its widespread application in biomedical research in recent years. In particular, large scale pooled CRISPR screening technologies have yielded many new discoveries in a variety of research fields (1–4). Since its establishment, the technology has undergone many improvements, and the CRISPR toolbox has been significantly expanded (5,6). sgRNA design has improved to a level where efficient screening is possible with whole-genome (WG) libraries containing only two independent sgRNAs per gene (7). Nevertheless, one aspect that remains rather ambiguous is the optimal method of sgRNA and Cas9 delivery. Most publicly available sgRNA libraries can be purchased in a “1 vector system” (e.g. lentiCRISPRv2 backbone) which delivers both the sgRNA library and Cas9 expression at the same time, or in a “2 vector system” (e.g. lentiGuide-Puro backbone) which delivers only the sgRNA library and requires a Cas9-expressing cell line to be generated first. To date, no comprehensive comparison has been made to address differences in screening efficiencies between the two systems. Here, we demonstrate that the level of Cas9 expression is a crucial determinant of editing speed. Therefore, a system that allows for control of Cas9 expression can significantly reduce the time necessary for editing, thereby increasing screen performance.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RESULTS AND DISCUSSION

Pooled CRISPR screens generally use a low multiplicity of infection (MOI<1) to prevent the presence of more than one sgRNA per cell. When using a "1 vector system", most infected cells will integrate a single copy of a sgRNA and Cas9 in a single step (1-step). Depending on the position of the integration, which is unpredictable in pooled screening, this could result in varying levels of sgRNA and Cas9 expression across a population of cells. When using a "2 vector system", first Cas9-expressing cells are established and subsequently a sgRNA library is introduced at low MOI (2-step) into these cells. Although the expression of sgRNAs can still vary depending on the position of the integration, the expression of Cas9 can be optimized independently to increase screening efficiency. We show experimentally that using a 2-step approach results in higher Cas9 expression levels in MCF10A cells, as compared to a 1-step approach (Fig. 1A).

To test whether Cas9 expression affects the performance of CRISPR screens, we performed two screens using either a 1-step or a 2-step system (Sup. Fig. 1A). For each screening system the same collection of sgRNAs (Brunello library) was used. To evaluate the performance of each screen, we analyzed the depletion of essential genes over non-essential genes, as previously described (8,9). By using the 2-step system we observe significantly more depletion of essential genes, in comparison to the 1-step system, indicating that increasing the level of Cas9 expression improved screening efficiency (Fig. 1B, Sup. Figs. 1B, C and Sup. Table 1).

To further study the relationship between Cas9 expression and editing timing and efficiency we expressed Cas9 in 3 different cell lines (SW480, A375 and HEK293T), using either constitutively expressed Cas9 (lentiCas9-P2A-EGFP) or an inducible Cas9 expression (lenti-iCas9-P2A-EGFP). With both vectors, the Cas9 expression should be directly proportional to the expression of EGFP. For the constitutively expressed Cas9, we used fluorescence activated cell sorting to select cell populations with low, medium and high levels of Cas9 expression (Sup. Figs. 2A-D). For the inducible system, a range of Cas9 levels was achieved by increasing doses of doxycycline (10ng/mL to 1 ug/mL). We then used these cell lines expressing different levels of CAS9 to measure editing efficiency both indirectly (using pXPR011 – Sup. Fig. 2E) and directly (using TIDE analysis (10) – Fig. 1C and Sup. Figs. 3A-D) in a time-course experiment. These results show that higher Cas9 levels reduce the time required for editing, meaning that higher Cas9 expression resulted in faster

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

editing kinetics. Importantly, this correlation was observed in all cell lines examined. Of note, a low level of CAS9 expression, due to leakiness of the inducible system, led to gene editing for sgRNAs 1 and 2, even in the absence of doxycycline (Figure 1C). To confirm whether the correlation between Cas9 levels and editing efficiency was independent of sgRNA sequence, we generated a sgRNA library (consisting of 486 sgRNAs targeting essential genes and 210 targeting safe-haven regions) and performed screens in all of our cell lines (Sup. Fig. 4A). As expected, we found that the amount of depletion of sgRNAs targeting essential genes was directly proportional to the level of Cas9 expression in the cells (Fig. 1D and Sup. Figs. 4B-D and Sup. Table 2).

There is a general preconception in the CRISPR field that high Cas9 expression can result in toxicity and increasing off-target effects. Indeed, several reports in different model organisms have shown that high Cas9 expression is toxic (11). We also observed that mammalian cell lines expressing very high levels of Cas9 tend to downregulate Cas9 expression over time (data not shown). To address whether high Cas9 expression increases off-target effects, we analyzed the behavior of sgRNAs targeting safe-havens in our screens. We did not observe any increase in the number of outliers in the conditions with high Cas9 (Fig. 1D and Sup. Figs. 4B-D). Instead, the same outliers were found across the different levels of Cas9 expression and, as expected, became more pronounced in the high Cas9 conditions due to the faster editing speed. This indicates that off-targets are predominantly caused by poor sgRNA design and not by high Cas9 expression. However, because high Cas9 expression exacerbates sgRNA effects, off-target effects affecting cell fitness, do become more apparent.

In conclusion, our findings suggest that in a heterogeneous population of cells, gene editing via CRISPR/Cas9 behaves similarly to an enzymatic reaction, i.e. the percentage of edited cells increases over time until it reaches a plateau. Although the editing rate depends on factors such as sgRNA design (specificity of target site selection), target region (open versus closed chromatin) and DNA repair capacity (NHEJ versus HDR) (12), our data show that editing speed is highly dependent on the level of Cas9 expression. The maximum achievable editing efficiency does not seem to be influenced by Cas9 expression, as it accumulates to the same level (Fig. 1C). However, in CRISPR screens increasing editing speed is desirable and therefore also higher levels of Cas9. In addition, due to variation in the

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

characteristics of single cell derived clonal lines, the use of polyclonal populations is preferred (13). The introduction of a construct expressing Cas9 together with GFP, for example, can be used for the polyclonal selection of high Cas9 expressing cells by using GFP as a “selection” marker. This strategy is favored over the generation and subsequent validation of single cell derived clonal lines (14). Our data highlight that Cas9 expression is a crucial parameter that influences the timing of gene editing, and the careful optimization per model system can significantly impact the outcome of large-scale pooled CRISPR screens.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

METHODS

Cell culture

MCF10A, SW480, A375 and HEK293T cell lines were obtained from ATCC. MCF10A cells were cultured in DMEM/F-12 medium containing 2.5 mM L-glutamine and 15 mM HEPES, supplemented with 5% horse serum, 10 µg/mL insulin, 0.5 µg/mL hydrocortisone and 0.1 µg/mL cholera toxin. SW480 cells were cultured in RPMI medium; A375 and HEK293T cells were cultured in DMEM medium. All the media were supplemented with 10% FBS, 1% penicillin/streptomycin and 2 mM L-glutamine. All cell lines were cultured at 37°C and with 5% CO_2. All cell lines were validated by STR profiling and mycoplasma tests were performed every 2-3 months.

Western blots

After the indicated culture period, cells were washed with chilled PBS, then lysed with RIPA buffer (25mM Tris-HCl, pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS) containing Complete Protease Inhibitor cocktail (Roche) and phosphatase inhibitor cocktails II and III (Sigma). Samples were then centrifuged for 10 min at 15,000 x g at 4°C and supernatant was collected. Protein concentration of the samples was normalized after performing a Bicinchoninic Acid (BCA) assay (Pierce BCA, Thermo Scientific), according to the manufacturer's instructions. Protein samples (denatured with DTT followed by 5 min heating at 95°C) were then loaded in a 4-12% polyacrylamide gel. Gels were run (SDS-PAGE) for approximately 45 min at 175 volts. Proteins were transferred from the gel to a polyvinylidene fluoride (PVDF) membrane at 330 mA for 90 min. After the transfer, membranes were incubated in blocking solution (5% bovine serum albumin (BSA) in PBS with 0.1% Tween-20 (PBS-T)). Subsequently, membranes were probed with primary antibody in blocking solution (1:1000) overnight at 4°C. Membranes were then washed 3 times for 10 min with PBS-T, followed by 1 h incubation at room temperature with the secondary antibody (HRP conjugated, 1:10,000) in blocking solution. Membranes were again washed 3 times for 10 min in PBS-T. Finally, a chemiluminescence substrate (ECL, Bio-Rad) was added to the membranes and signal imaged using the ChemiDoc-Touch (Bio-Rad).

Generation of Cas9-expressing cancer cell lines

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MCF10A cells were transduced with a lentivirus containing Edit-R Inducible Cas9 (Horizon CAS11229) at approximately 40% confluence in the presence of polybrene (4 µg/mL). Cells were incubated overnight, followed by replacement of the lentivirus-containing medium with fresh medium containing Blasticidin (10 µg/mL). After selection, several single cell clones were generated and Cas9 expression was assessed by Western blot. A clone with high Cas9 expression upon doxycycline treatment, and undetectable Cas9 expression in the absence of doxycycline, (named "MCF10A_iCas9") was used for subsequent experiments.

SW480 cells were transduced with Lenti-iCas9-neo (Addgene 85400) at approximately 60% confluence in the presence of polybrene (8 µg/mL). Cells were incubated overnight, followed by replacement of the lentivirus-containing medium with fresh medium containing G418 (100 µg/mL). After selection was completed, a titration of doxycycline (1 to 1000 ng/mL) was performed and the induction of Cas9 expression was assessed by flow cytometry. We determined that 10, 40 and 1000 ng/mL of doxycycline induced low, medium and high levels of Cas9 expression, respectively. Cas9 expression levels were confirmed by Western blot and flow cytometry one week later. The Cas9-expressing cell line was named "SW480_iCas9".

SW480, A375 and HEK293T cells were transduced with lentiCas9-EGFP (Addgene 63592) at approximately 40-60% confluence in the presence of polybrene (8 µg/mL). Cells were incubated overnight, followed by replacement of the lentivirus-containing medium with fresh medium. After 1 week in culture cells were sorted on low, medium and high GFP levels (BD FACSAria™ Fusion Cell Sorter). Cas9 expression levels were confirmed by Western blot and flow cytometry one week later. The Cas9-expressing cells were named according to their cell line name and Cas9 expression level, i.e. "*name*_Cas9$^{expression\ level}$".

Editing efficiency assessment by flow cytometry and TIDE analysis

Parental and Cas9-expressing cell lines, as indicated, were transduced with pXPR_011 (Addgene, 59702) at approximately 40-60% confluence in the presence of polybrene (8 µg/mL). Cells were incubated overnight, followed by replacement of the lentivirus-containing medium with fresh medium containing puromycin (2 µg/mL). Cells were harvested 20 h, 3 days (SW480), 6 days and 10 days (HEK293T and A375) after transduction with pXPR_011, GFP levels were assessed by flow

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cytometry (BD LSRFortessa) and analyzed using FlowJo 10. pXPR_011 results in the expression of both GFP and a sgRNA targeting GFP in the transduced cells. Therefore, editing efficiency can be (indirectly) assessed by analyzing the reduction in GFP signal over time.

SW480_iCas9 cells were transduced with 3 different sgRNAs cloned into pU6-sgRNA-EF1-Puro-T2A-GFP (see sgRNA cloning section below) at approximately 60% confluence in the presence of polybrene (8 µg/mL). Cells were incubated overnight, followed by replacement of the lentivirus-containing medium with fresh medium containing puromycin (2 µg/mL). Cells were kept in puromycin for 7 days. At day 7, a fraction of the cells was harvested, another fraction was analyzed by flow cytometry to confirm equal infection efficiencies indicating similar sgRNA expression levels, and the rest of the cells were placed back in culture and treated with either 10 ng/mL or 1 µg/mL doxycycline. Cells were harvested from these two induction arms after 2, 4, 8, 12, 16, (20, 24 and 32 – only for sgRNA 3) days in continuous culture with doxycycline. DNA was isolated from all samples, Sanger sequencing was performed and editing efficiency was analyzed using TIDE (https://tide.nki.nl/). At day 13, cells were also harvested for western blot and flow cytometry analysis, to assess Cas9 expression levels.

sgRNA cloning

To generate OR9Q2 (non-essential gene) sgRNA-expressing MCF10A cells, we cloned OR9Q2 sgRNA (5'-ATAACCGAGAAGGCCCGCTG-3') sequence into lentiCRISPRv2 (Addgene, #52961) and lentiGuide-Puro (Addgene, #52963). Backbones were digested with BsmBI and cloned using Gibson assembly. sgRNAs targeting 3 different locations in the genome were cloned into a modified version of pU6-sgRNA EF1Alpha-puro-T2A-BFP (Addgene, #60955), where BFP was replaced by superfolder GFP (sfGFP) – named "pU6-sgRNA-EF1-Puro-T2A-GFP". Puro-T2A-BFP was removed using NheI and EcoRI sites. To introduce Puro-T2A-sfGFP, we amplified Puro-T2A as well as sfGFP, adding homology arms to both PCR products. Puromycin-T2A was amplified using the following oligos: FW: 5'-GTTTTTTTCTTCCATTTCAGGTGTCGTGAGCTAGCCCACCATGACCGAGTACAAGCCCAC-3', RV: 5'-AACTCCAGTGAAAAGTTCTTCTCCTTTGCTGGTGGCGACCGGTGGGCCAGGATTCTCCTC-3'

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sfGFP was amplified using the following oligos: FW: 5'-GAGGAGAATCCTGGCCCACCGGTCGCCACCAGCAAAGGAGAAGAACTTTTCACTGGAGTT-3' RV: 5'-ATGTATGCTATACGAAGTTATTAGGTCCCTCGACGAATTCTTATTTGTAGAGCTCATCCA-3'

The resulting PCR products were inserted into the open sgRNA vector backbone through Gibson Assembly. To introduce the custom designed sgRNA sequences into the pU6-sgRNA-EF1-Puro-T2A-GFP vector, the vector was digested using BstXI and BamHI. The sgRNAs were PCR-amplified using sgRNA-specific forward primers and a universal reverse primer: FW_1: 5'-TTGGAGAACCACCTTGTTGGAATATGTTTAAGCCTAGAGAGTTTAAGAGCTAAGCTGGAA, FW_2: 5'-TTGGAGAACCACCTTGTTGGTATAGGATAATAGCTGGAAGGTTTAAGAGCTAAGCTGGAA, FW_3: 5'-TTGGAGAACCACCTTGTTGGAGAGGTCTAATTCTAGGGCCGTTTAAGAGCTAAGCTGGAA, RV: 5'-GTAATACGGTTATCCACGCGGCCGCCTAATGGATCCTAGTACTCGAGA.

The resulting PCR products were isolated and used for Gibson Assembly.

Generation of custom sgRNA library

For the design of the custom sgRNA library targeting essential genes and safe-havens we used the Broad GPP sgRNA design portal and the safe-havens as designed previously (15). The sgRNA sequences (Supplemental Table 2) were ordered as a pool of oligonucleotides (Agilent) with flanking sequences to enable PCR amplification and Gibson assembly into pLentiGuide-Puro (pLG, addgene #52963). The pooled oligo library was amplified using pLG_U6_foward 5'-GGCTTTATATATCTTGTGGAAAGGACGAAACACCG-3' and pLG-TRACR_Reverse 5'-GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC-3'. The fragments were purified and cloned into pLG as described by Morgens (16). The representation of the custom sgRNA library was validated by next generation sequencing.

sgRNA libraries and screens

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Two different versions of the Brunello library were used – a "1 vector system" (backbone expresses both Cas9 and the library – Addgene, 73179) and a "2 vector system" (backbone expresses only the library – Addgene, 73178). In this study we also used our Essential/Safe-havens library described above.

The appropriate number of cells to achieve 250-fold representation of the library, multiplied by five to account for 20% transduction efficiency, were transduced at approximately 40-60% confluence in the presence of polybrene (4-8 µg/mL) with the appropriate volume of the lentiviral-packaged sgRNA library. Cells were incubated overnight, followed by replacement of the lentivirus-containing medium with fresh medium containing puromycin (2-4 µg/mL). The lentivirus volume to achieve a MOI of 0.2, as well as the puromycin concentration to achieve a complete selection in 3 days was previously determined for each cell line. Transductions were performed in triplicate (technical for negative selection screens and biological for positive selection screens). After puromycin selection, cells were split into the indicated arms (for each arm, the appropriate number of cells to keep a 250-fold representation of the library was plated at approximately 10-20% confluence) and a T_0 (reference) time point was harvested. Cells were maintained as indicated. In case a passage was required, cells were reseeded at the appropriate number to keep at least a 500-fold representation of the library. Cells (enough to keep at least a 500-fold representation of the library, to account for losses during DNA extraction) were collected when indicated, washed with PBS, pelleted and stored at -80°C until DNA extraction.

DNA extraction, PCR amplification and Illumina sequencing

Genomic DNA (gDNA) was extracted (Zymo Research, D3024) from cell pellets according to the manufacturer's instructions. For every sample, gDNA was quantified and the necessary DNA to maintain a 250-fold representation of the library was used for subsequent procedures (for this we assumed that each cell contains 6.6 pg genomic DNA). Each sample was divided over 50 µl PCR reactions (using a maximum of 1 µg gDNA per reaction) using barcoded forward primers to be able to deconvolute multiplexed samples after next generation sequencing (for primers and barcodes used, see Supplementary Table 3). PCR mixture per reaction: 10 µl 5x HF Buffer, 1 µl 10 µM forward primer, 1 µl 10 µM reverse primer, 0.5 µl Phusion polymerase (Thermo Fisher, F-530XL), 1 µl 10mM dNTPs, adding H_2O and template to 50 µl. Cycling conditions: 30 sec at 98°C, 20× (30 sec at 98°C, 30 sec at 60°C, 1

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

min at 72°C), 5 min at 72 °C. The products of all reactions from the same sample were pooled and 2 µl of this pool was used in a subsequent PCR reaction using primers containing adapters for next generation sequencing (Supplementary Table 2). The same cycling protocol was used, this time for 15 cycles. Next, PCR products were purified using the ISOLATE II PCR and Gel Kit (Bioline, BIO-52060) according to the manufacturer's instructions. DNA concentrations were measured and, based on this, samples were equimolarly pooled and subjected to Illumina next generation sequencing (HiSeq 2500 High Output Mode, Single-Read, 65 bp). Mapped read-counts were subsequently used as input for the further analyses.

Bioinformatics Analysis

For each CRISPR screen the sgRNA count data for each sample was normalized for sequence depth using the method described by DESeq2[3] with the difference that the total value instead of the median of a sample was used. Because of the composition of the sgRNA library with a large fraction of sgRNAs targeting essential genes, the T1 samples were corrected by dividing with the median of T1/T0 ratios for the population of non-essential sgRNAs. For the genome-wide CRISPR screen comparing the efficiency of the 1-step and 2-step systems, a differential analysis was performed using DESeq2 (17). The output was sorted on the DESeq2 test statistic with the most depleted sgRNA at the top. We then used MAGeCK Robust Rank Algorithm to determine enrichment of sgRNAs targeting each gene (18). For the ROC curves in supplemental Fig. 1B and 1C the output of these two analyses were filtered for 50 positive and 50 negative controls genes as described by Evers and colleagues (8). The Comparisons of the distribution of different groups of sgRNAs were performed using the Wilcoxon test.

Reagents

Primary antibodies: Tubulin (Sigma, T9026) and Cas9 (Cell Signaling, 14697). Secondary antibody: Goat Anti-Mouse IgG (H + L)-HRP Conjugate (BioRad, 1706516).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

We thank the NKI's Flow Cytometry and Sequencing facilities. This work was supported by a grant from the Dutch Cancer Society.

Authors' Contributions

R.L.B., R.B. and R.M. supervised all research. J.N. and R.L.B. wrote the manuscript. J.N., K.J., C.L., L.K., M.D., B.M., D.V. and H.H. designed, performed and analyzed the experiments. All authors commented on the manuscript.

Conflicts of interest

The authors declare no conflict of interests.

Data and code availability

Data generated during the study will be available in a public repository. Code generated during the study will be available upon request to the authors.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE LEGENDS

Fig. 1: The importance of time and Cas9 expression for efficient gene editing.

A, Comparison of 1-step versus 2-step system. MCF10A and MCF10A_iCas9 cells were transduced with a sgRNA targeting a non-essential gene (OR9Q2), cloned in the lentiCRISPRv2 vector (1-step system) and in the lentiGuide-Puro vector (2-step system), respectively. The level of Cas9 expression was measured by examining Cas9 levels in the western blot. Tubulin was used as loading control. The quantification from 3 biological replicates, as well as a representative blot are shown. **B,** Comparison of depletion of sgRNAs targeting essential genes for 1-step and 2-step systems. MCF10A cells were screened with the same whole-genome sgRNA library (Brunello), using either a 1-step or a 2-step system. The analysis of the depletion (log2 Fold Change) of the sgRNAs targeting essential genes over non-essential genes is displayed. Box plot shows the median (horizontal line), interquartile range (hinges), and the smallest and largest values no more than 1.5 times the interquartile range (whiskers). Comparisons were made using the Wilcoxon test. **C,** Analysis of editing speed in relation to Cas9 expression levels. SW480_iCas9 cells were transduced with 3 different sgRNAs. Cells were kept in puromycin for 7 days. At day 7, a fraction of the cells was harvested as a reference sample, while the rest of the cells were placed back in culture and treated with either 10 ng/mL or 1 µg/mL doxycycline. Cells were kept in continuous culture with doxycycline and harvested from these two arms at the indicated time-points. Gene editing efficiency was analyzed using TIDE. **D,** Analysis of editing speed in relation to Cas9 expression levels for a sgRNA library targeting essential genes and safe-havens. SW480_iCas9 cells were screened with a library of essential and safe-haven sgRNAs. To induce different levels of Cas9 expression, cells were treated with 10, 40 or 1000 ng/mL of doxycycline and cultured for 8 population doublings. The analysis of the depletion (log2 Fold Change) of the sgRNAs targeting essential genes and safe-havens is displayed. Box plot shows the median, interquartile range and the smallest and largest values no more than 1.5 times the interquartile range. Comparisons were made using the Wilcoxon test.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1: Genome-wide CRISPR screen to compare the efficiency of the 1-step and 2-step systems.

A, Schematic of the screen layout. 1-step system: ~120x10^6 MCF10A cells were transduced with the lentiCRISPRv2 Brunello library. After 3 days of puromycin selection, a reference sample was harvested and 1.6x10^7 cells were plated (in triplicate). At day 6 and 11 cells were split and 1.6x10^7 cells were reseeded per replicate. At day 15, all cells were harvested. Abundance of sgRNA for each sample was analyzed by sequencing. 2-steps system: ~150x10^6 MCF10A_iCas9 cells were transduced with the lentiGuide-Puro Brunello library. After 3 days of puromycin selection, a reference sample was harvested and 2x10^7 cells were plated (in 6 replicates). Three replicates were cultured in the absence of doxycycline, while the other 3 replicates were cultured in the presence of 1 µg/mL doxycycline. At day 6 and 11 cells were split and 4x10^7 cells were reseeded per replicate. At day 15, all cells were harvested. Abundance of sgRNAs for each sample was analyzed by sequencing. **B-C**, ROC curves for sgRNAs based on the DESeq2 results sorted on the DESeq2 statistic (B) and for genes based on the rank column in the RRA output, both in increasing order (C) p-values for essential genes. FPR, false-positive rate; TPR, true-positive rate.

Supplementary Figure 2: Editing speed is proportional to Cas9 expression level.

A, Schematic of the generation of Cas9 expressing cell lines. Cell lines were transduced using either lentiCas9-P2A-EGFP or lenti-iCas9-P2A-EGFP. Cells transduced with lenti-iCas9-P2A-EGFP were selected with G418 and then treated with increasing concentrations of doxycycline. Then, using flow cytometry, we selected cell populations with different levels of Cas9 expression (low, medium or high), based on EGFP expression. **B,** Validation of the Cas9 expressing lines using western blot. The level of Cas9 expression was measured by examining Cas9 levels by Western blotting. Tubulin was used as loading control. Representative blots from 2 biological replicates are shown. **C,** Gating strategy used for all flow cytometry experiments. All events were gated on live cells, these were then gated on single

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cells, and EGFP levels were finally plotted on the single cells. **D,** Validation of the Cas9-expressing lines using flow cytometry. The level of Cas9 expression was measured (indirectly) by examining EGFP levels (x-axis) by flow cytometry. **E,** Editing speed is proportional to Cas9 expression level. Parental and Cas9-expressing cell lines were transduced with pXPR_011. Cells were harvested at the indicated time-points after transduction with pXPR_011 and GFP levels (x-axis) were assessed by flow cytometry and analysed using FlowJo 10.

Supplementary Figure 3: Editing speed is proportional to Cas9 expression level.

A, Schematic of editing efficiency assessment by TIDE analysis. SW480_iCas9 cells were transduced with 3 different sgRNAs. Cells were incubated overnight, followed by replacement of the lentivirus-containing medium with fresh medium containing puromycin (2 µg/mL). Cells were kept in puromycin for 7 days. At day 7, a fraction of the cells was harvested as reference sample, (**B**) another fraction was analyzed by flow cytometry to confirm equal infection efficiency indicating similar sgRNA expression levels, while the rest of the cells were placed back in culture and treated with either 10ng/mL or 1µM/mL doxycycline. Cells were harvested from these two arms at different time-points over time. DNA was isolated from all samples, Sanger sequencing was performed and editing efficiency was analyzed using TIDE. At day 13, cells were also harvested for Western blotting (**C**) and flow cytometry analysis (**D**) to assess Cas9 expression levels.

Supplementary Figure 4: Editing speed is proportional to Cas9 expression, independent of sgRNA sequence.

A, Schematic of the screen layout. For each Cas9 expressing subline $2.5x10^6$ cells were transduced with the essential/safe-haven library using a MOI of 0.2. Cells were incubated overnight, followed by replacement of the lentivirus-containing medium with fresh medium containing puromycin (2 µg/mL). After 3 days of puromycin selection, a reference sample was harvested and $2x10^8$ cells were plated (in triplicate). Cells were cultured for 8 population doublings (~10 days), after which $5x10^8$ cells per replicate were harvested, DNA was extracted, samples were barcoded using PCR and sent for sequencing. **B-D,** Editing speed is proportional to Cas9 expression, independent of sgRNA sequence. Cells screened as outlined

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

above were analyzed for the depletion (log2 Fold Change) of the sgRNAs targeting essential genes and safe-havens . Log2FC depicts the log 2 fold change sgRNA counts T1/T0. Box plot shows the median (horizontal line), interquartile range (hinges), and the smallest and largest values no more than 1.5 times the interquartile range (whiskers). Comparisons were made using the Wilcoxon test.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

REFERENCES

1. Wang C, Jin H, Gao D, Wang L, Evers B, Xue Z, et al. A CRISPR screen identifies CDK7 as a therapeutic target in hepatocellular carcinoma. Vol. 28, Cell research. 2018. p. 690–2.
2. Tsui CK, Barfield RM, Fischer CR, Morgens DW, Li A, Smith BAH, et al. CRISPR-Cas9 screens identify regulators of antibody–drug conjugate toxicity. Nat Chem Biol [Internet]. 2019;15(10):949–58. Available from: https://doi.org/10.1038/s41589-019-0342-2
3. Park RJ, Wang T, Koundakjian D, Hultquist JF, Lamothe-Molina P, Monel B, et al. A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors. Nat Genet [Internet]. 2017;49(2):193–203. Available from: https://doi.org/10.1038/ng.3741
4. Flint M, Chatterjee P, Lin DL, McMullan LK, Shrivastava-Ranjan P, Bergeron É, et al. A genome-wide CRISPR screen identifies N-acetylglucosamine-1-phosphate transferase as a potential antiviral target for Ebola virus. Nat Commun [Internet]. 2019;10(1):285. Available from: https://doi.org/10.1038/s41467-018-08135-4
5. Arnold C. What's new in clinical CRISPR? Nat Med [Internet]. 2021;27(2):184–5. Available from: https://doi.org/10.1038/s41591-020-01222-4
6. Yang Y, Xu J, Ge S, Lai L. CRISPR/Cas: Advances, Limitations, and Applications for Precision Cancer Research [Internet]. Vol. 8, Frontiers in Medicine . 2021. p. 115. Available from: https://www.frontiersin.org/article/10.3389/fmed.2021.649896
7. DeWeirdt PC, Sangree AK, Hanna RE, Sanson KR, Hegde M, Strand C, et al. Genetic screens in isogenic mammalian cell lines without single cell cloning. Nat Commun. 2020 Feb;11(1):752.
8. Evers B, Jastrzebski K, Heijmans JPM, Grernrum W, Beijersbergen RL, Bernards R. CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. Nat Biotechnol. 2016 Jun;34(6):631–3.
9. Sanson KR, Hanna RE, Hegde M, Donovan KF, Strand C, Sullender ME, et al. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. Nat Commun. 2018 Dec;9(1):5416.
10. Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic

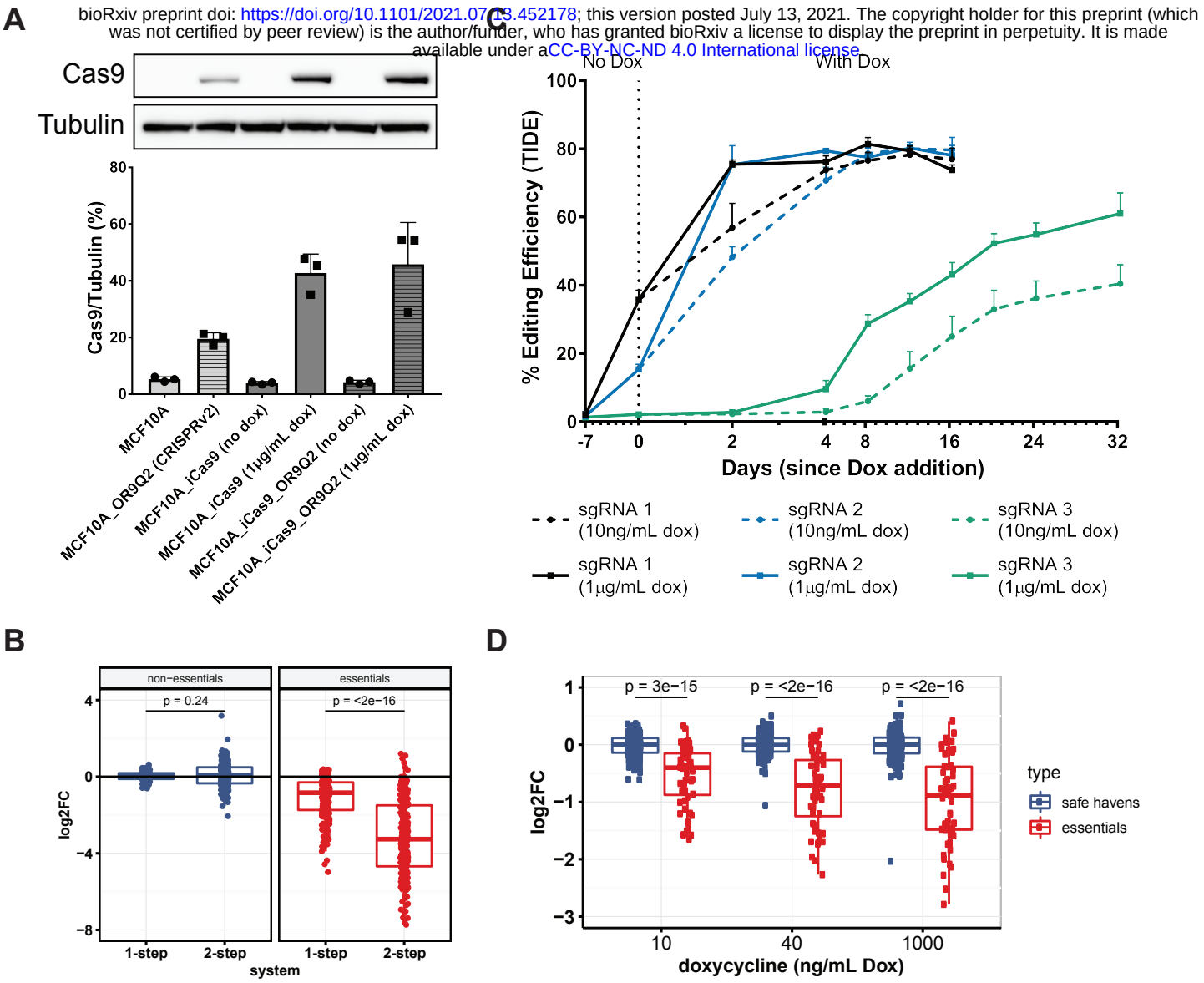
bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acids Res. 2014 Dec;42(22):e168.

11. Ye S, Enghiad B, Zhao H, Takano E. Fine-tuning the regulation of Cas9 expression levels for efficient CRISPR-Cas9 mediated recombination in Streptomyces. J Ind Microbiol Biotechnol. 2020 May;47(4–5):413–23.
12. Schep R, Brinkman EK, Leemans C, Vergara X, van der Weide RH, Morris B, et al. Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. Mol Cell. 2021 Apr;
13. Najm FJ, Strand C, Donovan KF, Hegde M, Sanson KR, Vaimberg EW, et al. Orthologous CRISPR–Cas9 enzymes for combinatorial genetic screens. Nat Biotechnol [Internet]. 2018;36(2):179–89. Available from: https://doi.org/10.1038/nbt.4048
14. Henkel L, Rauscher B, Schmitt B, Winter J, Boutros M. Genome-scale CRISPR screening at high sensitivity with an empirically designed sgRNA library. BMC Biol [Internet]. 2020;18(1):174. Available from: https://doi.org/10.1186/s12915-020-00905-1
15. Joung J, Konermann S, Gootenberg JS, Abudayyeh OO, Platt RJ, Brigham MD, et al. Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. Nat Protoc [Internet]. 2017;12(4):828–63. Available from: https://doi.org/10.1038/nprot.2017.016
16. Morgens DW, Wainberg M, Boyle EA, Ursu O, Araya CL, Tsui CK, et al. Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. Nat Commun [Internet]. 2017;8(1):15178. Available from: https://doi.org/10.1038/ncomms15178
17. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol [Internet]. 2014;15(12):550. Available from: https://doi.org/10.1186/s13059-014-0550-8
18. Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol. 2014 Dec;15(12):554.

Figure 1: The importance of time and Cas9 expression for efficient gene editing

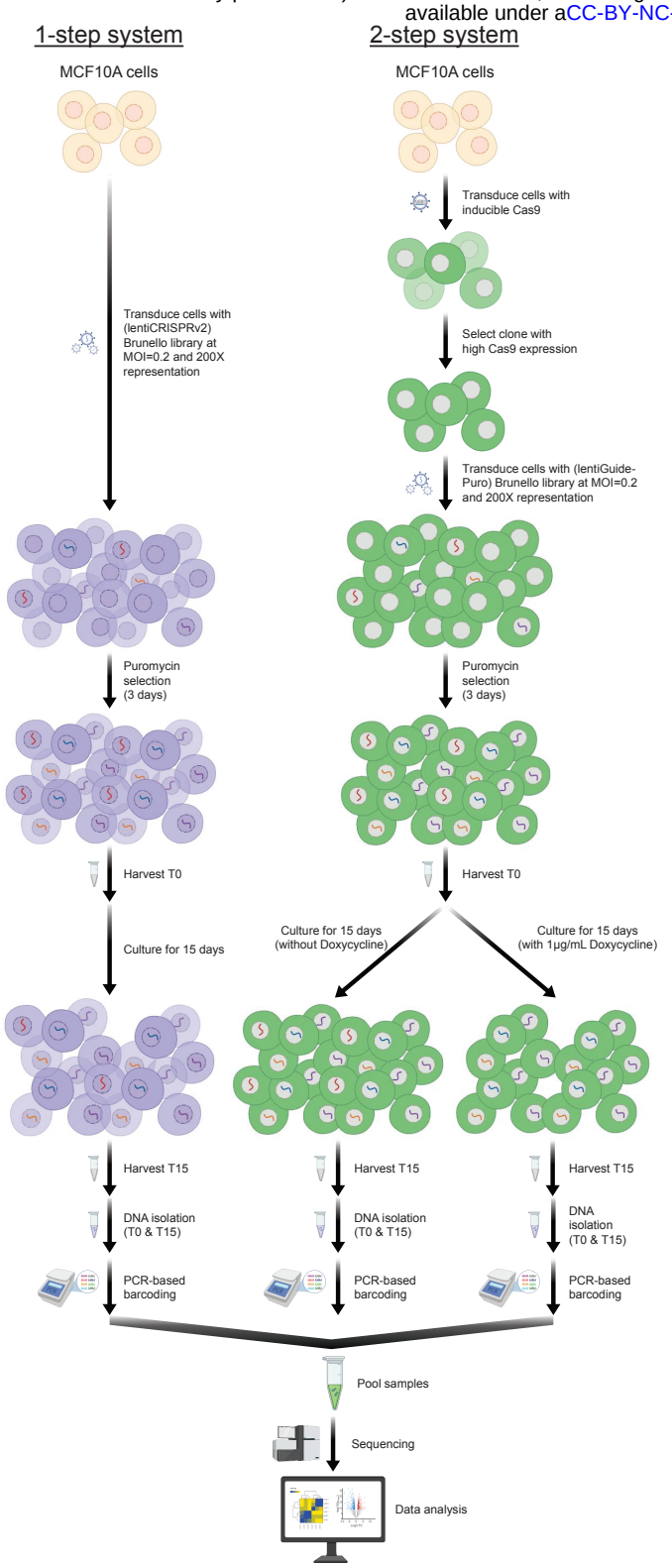
bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



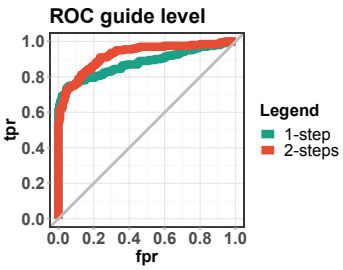
Supplemental Figure 1

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

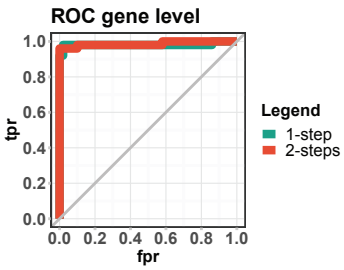
A



B



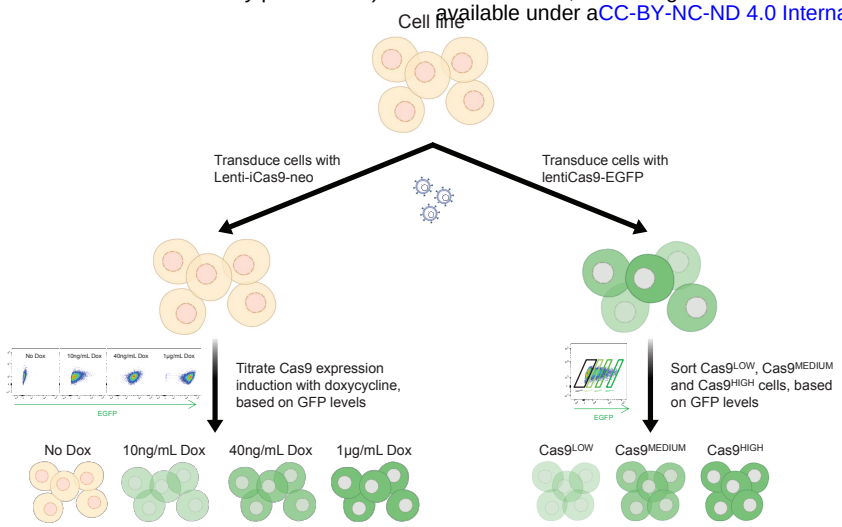
C



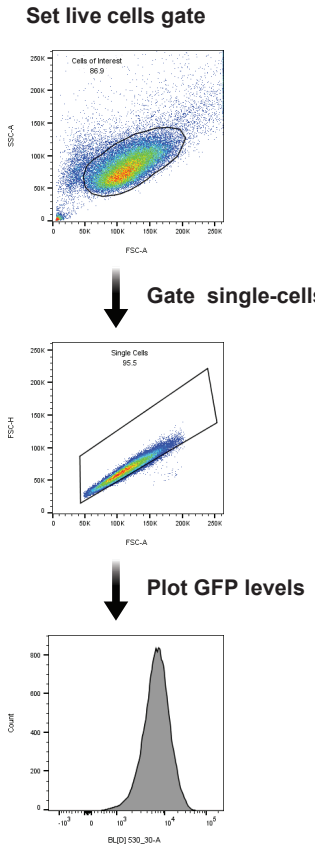
Supplemental Figure 2

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

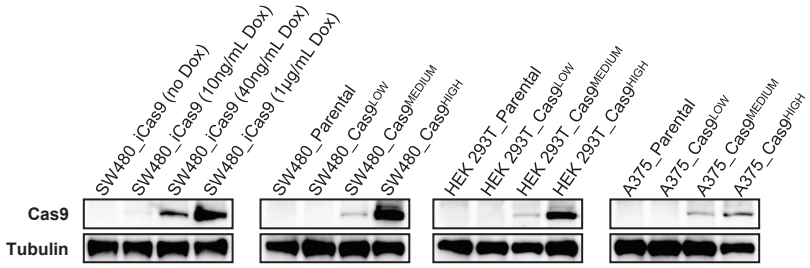
A



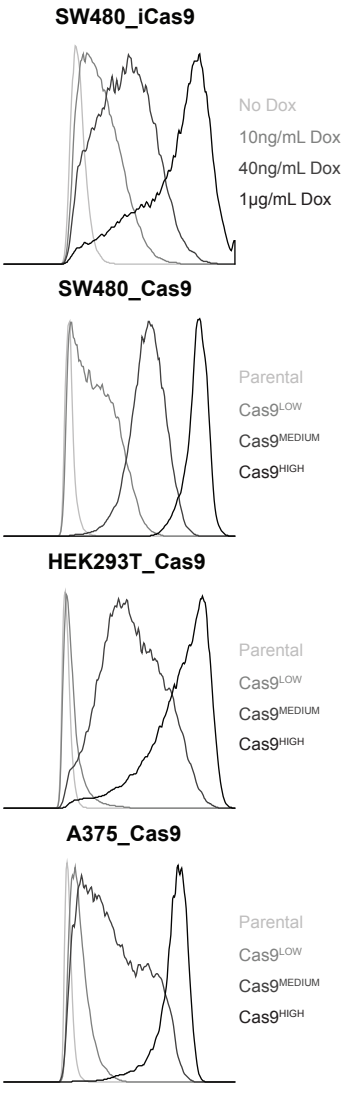
C



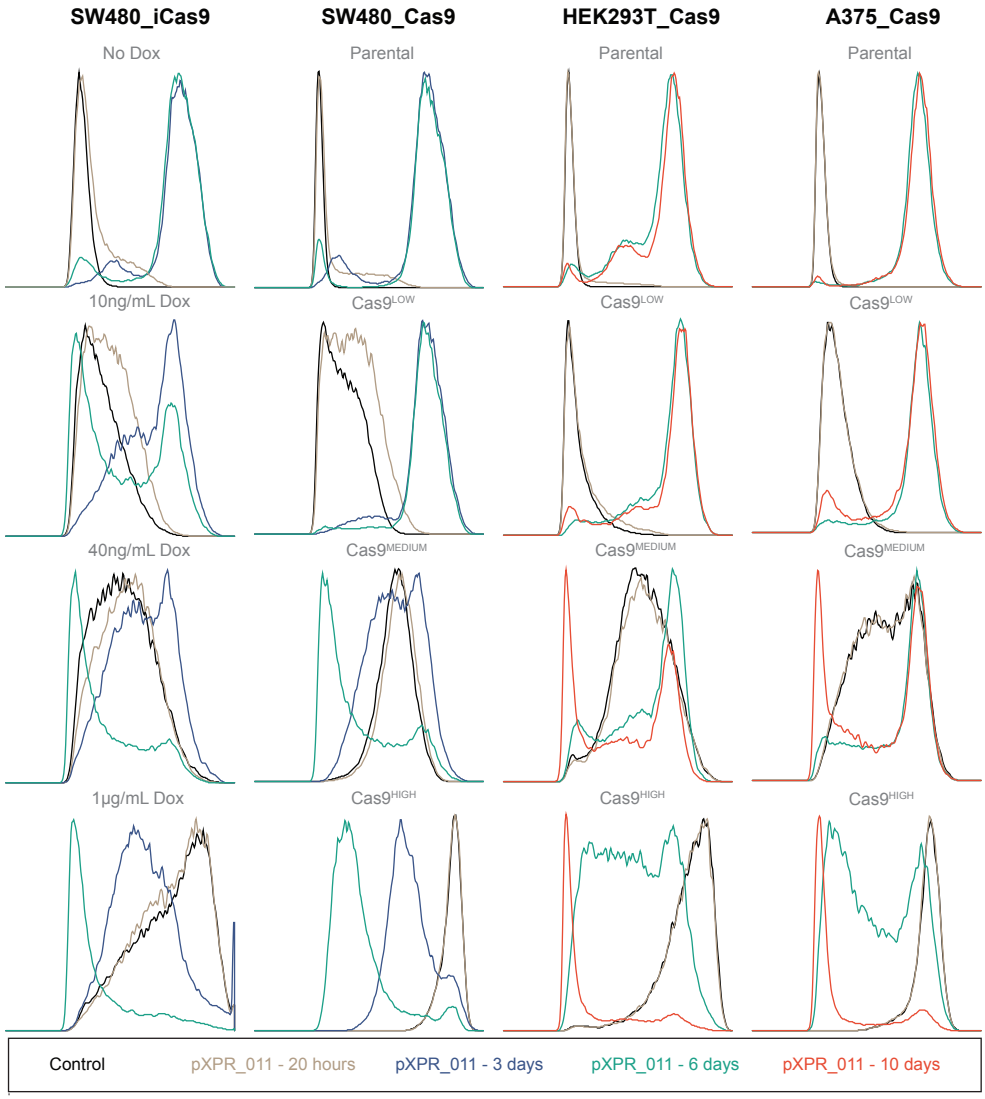
B



D



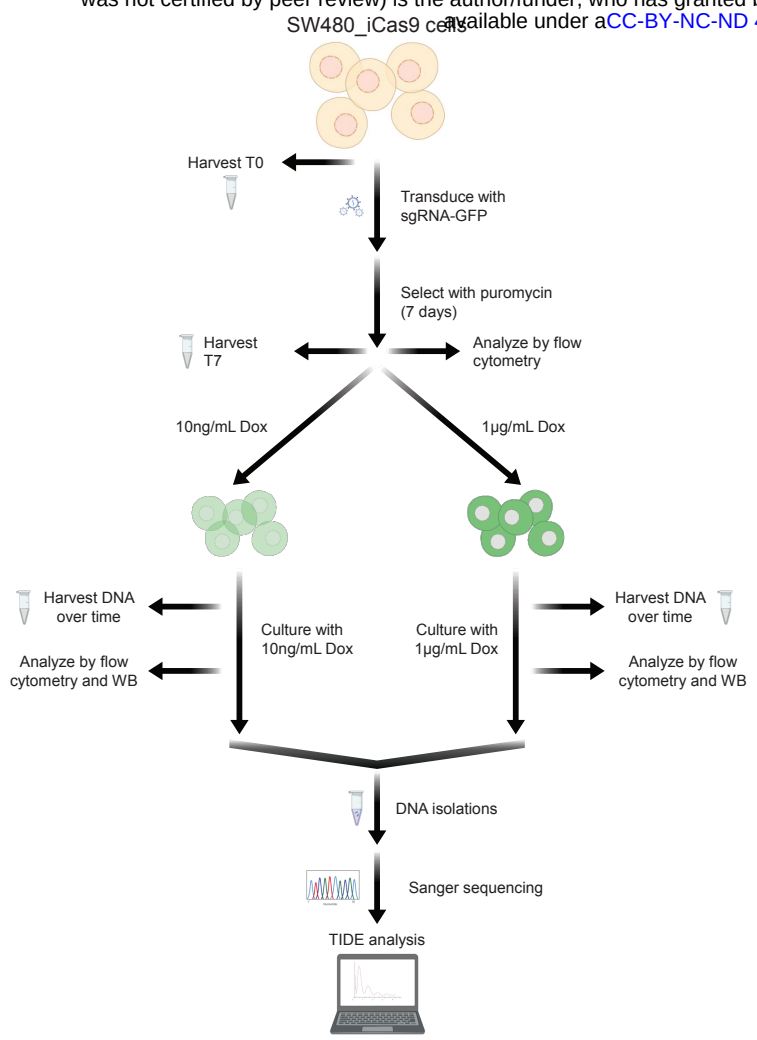
E



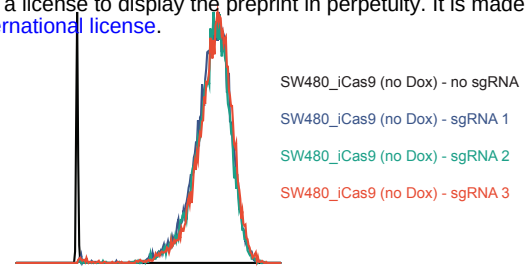
Supplemental Figure 3

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

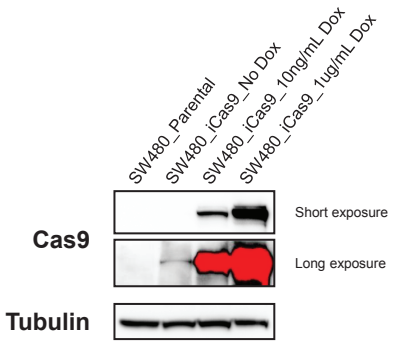
A



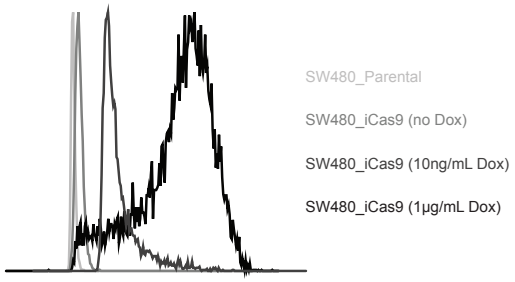
B



C



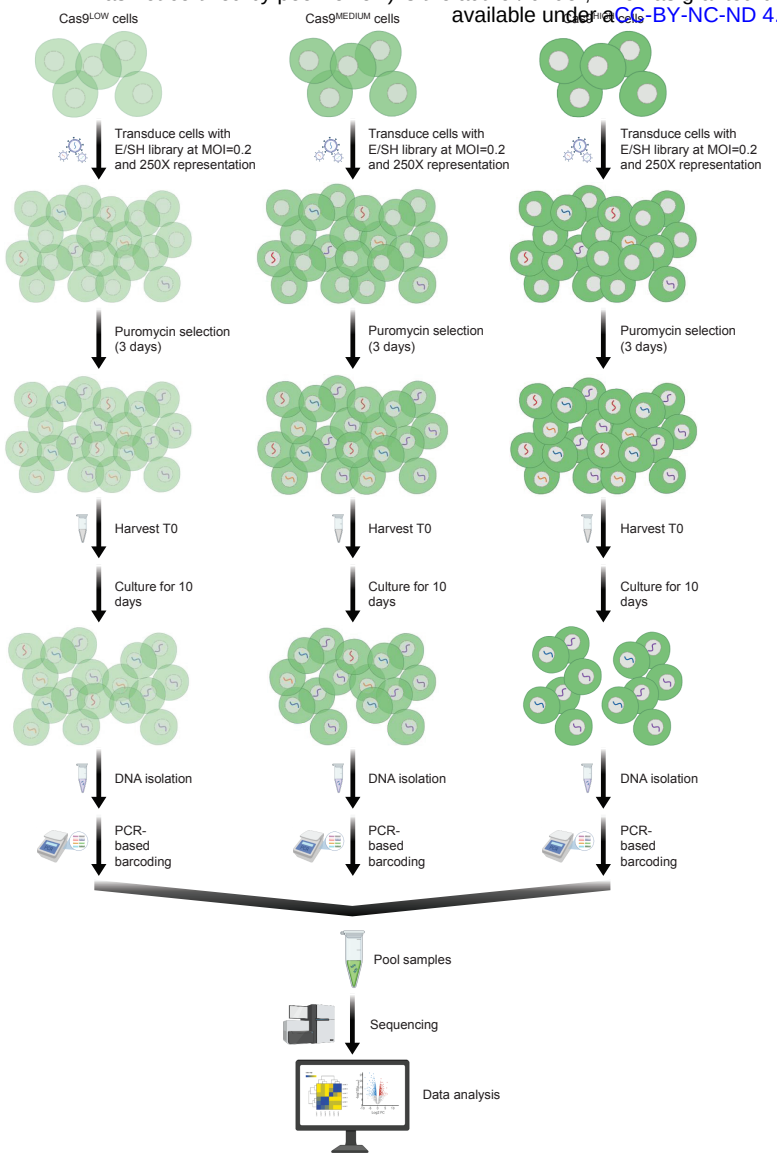
D



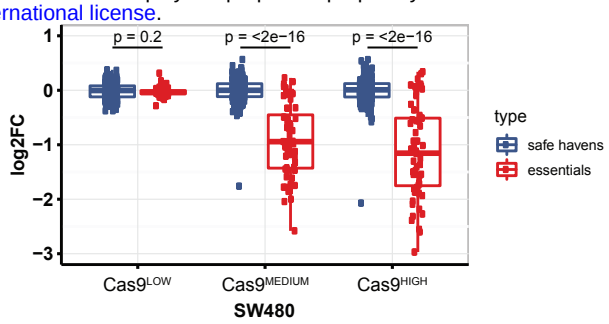
Supplemental Figure 4

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452178; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

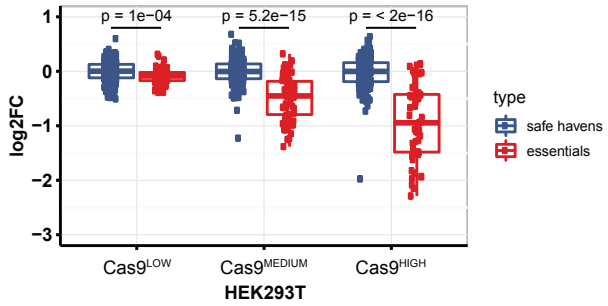
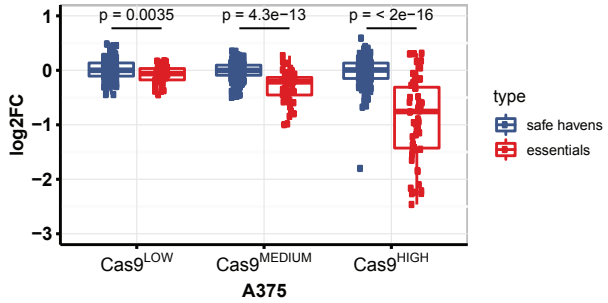
A



B



C



D