**Robust and Versatile Arrayed Libraries for Human Genome-Wide CRISPR Activation, Deletion and Silencing**

Jiang-An Yin[1,7], Lukas Frick[1,7], Manuel C. Scheidmann[2], Chiara Trevisan[1], Ashutosh Dhingra[3], Anna Spinelli[1], Yancheng Wu[1], Longping Yao[1], Dalila Laura Vena[1], Elena De Cecco[1], Kathi Ging[1], Tingting Liu[1], Joachim Täger[3], Salvador Rodriguez[3], Jingjing Guo[1], Scott Berry[4,5], Marco Losa[1], Simone Hornemann[1], Martin Kampmann[6], Lucas Pelkmans[4], Dominic Hoepfner[2], Peter Heutink[3], and Adriano Aguzzi[1*]

[1]Institute of Neuropathology, University Hospital Zurich, University of Zurich, 8091 Zurich, Switzerland

[2]Novartis Institutes for BioMedical Research, Novartis Campus, 4056 Basel, Switzerland

[3]German Center for Neurodegenerative Diseases, 72076 Tübingen, Germany

[4]Department of Molecular Life Sciences, University of Zurich, 8057 Zurich, Switzerland

[5]EMBL Australia Node in Single Molecule Science, School of Medical Sciences, University of New South Wales, Sydney, Australia

[6]Institute for Neurodegenerative Diseases, Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94158, USA

[7]These authors contributed equally.

*Correspondence: adriano.aguzzi@usz.ch

## Abstract

Genome-wide CRISPR phenotypic screens are clarifying many fundamental biological phenomena. While pooled screens can be used to study selectable features, arrayed CRISPR libraries extend the screening territory to cell-nonautonomous, biochemical and morphological phenotypes. Using a novel high-fidelity liquid-phase plasmid cloning technology, we generated two human genome-wide arrayed libraries termed T.spiezzo (gene ablation, 19,936 plasmids) and T.gonfio (gene activation and epigenetic silencing, 22,442 plasmids). Each plasmid encodes four non-overlapping single-guide RNAs (sgRNAs), each driven by a unique housekeeping promoter, as well as lentiviral and transposable vector sequences. The sgRNAs were designed to tolerate most DNA polymorphisms identified in 10,000 human genomes, thereby maximizing their versatility. Sequencing confirmed that ~90% of each plasmid population contained ≥3 intact sgRNAs. Deletion, activation and epigenetic silencing experiments showed efficacy of 75-99%, up to 10,000x and 76-92%, respectively; lentiviral titers were ~$10^7$/ml. As a proof of concept, we investigated the effect of individual activation of each human transcription factor (n=1,634) on the expression of the cellular prion protein $PrP^C$. We identified 24 upregulators and 12 downregulators of $PrP^C$ expression. Hence, the T.spiezzo and T.gonfio libraries represent a powerful resource for the individual perturbation of human protein-coding genes.

## Introduction

According to Karl Popper, fundamentally new discoveries cannot be rooted in prior knowledge (Popper, 1992). A powerful strategy to circumvent this limitation is to perform experiments that do not rely on priors. Unbiased genetic screens, whose development Popper did not live to see, fulfill this requirement. In the last decades, RNA interference- or mutagen-mediated screenings have greatly improved our understanding of biology and human health and transformed drug development for diseases (Acevedo-Arozena et al., 2008; Boutros and Ahringer, 2008). CRISPR-mediated techniques (Cong et al., 2013; Jinek et al., 2012) have enormously expanded the toolkits of genetic screening, and now allow for gene ablation (CRISPRko), activation (CRISPRa), interference (CRISPRi) and epigenetic silencing (CRISPRoff) (Amabile et al., 2016; Gilbert et al., 2014; Hsu et al., 2014; Jinek et al., 2012; Kampmann, 2020; Knott and Doudna, 2018; Nunez et al., 2021; Wang et al., 2016). CRISPR-based screenings yield both overlapping and distinct hits compared to RNA-interference-based screenings, and CRISPR-mediated gene perturbations are much more specific than RNA interference-based methods (Evers et al., 2016; Gilbert et al., 2014; Morgens et al., 2016; Smith et al., 2017). Thus, genome-wide CRISPR-based gene perturbation libraries are of essential importance to identify and understand the underlying biology of genes involved in various biological processes and diseases.

Many genome-wide CRISPR-based pooled libraries including CRISPRko, CRISPRa and CRISPRi (Hart et al., 2017; Horlbeck et al., 2016; Konermann et al., 2015; Sanson et al., 2018; Shalem et al., 2014) have been successfully deployed in screens for cellular phenotypes including cell survival, proliferation or sensitivity to insults, and gene expression (Kampmann, 2020). Yet, these pooled libraries cannot be readily applied to investigate cell-non-autonomous phenotypes, in which genetically mutant cells cause other cells to exhibit a mutant phenotype. This limitation becomes obvious when studying, for instance, glia-neuron interactions, which are crucial for brain function in physiological and pathological conditions (Araque and Navarrete, 2010). Furthermore, pooled libraries have limitations for genome-wide high-content optical screens and very limited use for biochemical screens (Feldman et al., 2019; Kampmann, 2020; Kanfer et al., 2021; Yan et al., 2021).

The limitations of pooled libraries can be circumvented by the use of arrayed libraries. However, in contrast to the varied, widely available and rapidly advancing collections of pooled CRISPR libraries, there are still only limited commercial (HorizonDiscovery; Synthego; ThermoFisher) and academic (Erard et al., 2017b; Metzakopian et al., 2017; Schmidt et al., 2015) resources for arrayed CRISPR screens, and their effectiveness is not well-documented. Moreover, these arrayed CRISPR libraries suffer from several limitations. Firstly, arrayed synthetic crRNA libraries are restricted to the usage with easily transfectable cells, and the selection of successfully transfected cells is not possible owing to lack of selection markers. Secondly, plasmid-based libraries featuring one sgRNA per vector

exhibit low and heterogenous gene perturbation efficiency when using one sgRNA per target gene (Chakrabarti et al., 2019), leading to variable editing outcomes. Thirdly, single sgRNAs are under transcriptional control of a single promoter, whose cell-type specificity may diminish its effectiveness. Fourthly, the sgRNA design algorithms employed by most existing libraries are based on the hg38 or earlier versions of human reference genomes (Hanna and Doench, 2020; Sanson et al., 2018). However, human genomes are highly polymorphic, and the genome of cells used for a gene-perturbation screen may significantly diverge from the reference genome, leading to impaired sgRNA function. This issue is particularly important for the study of patient-derived cells, such as cells derived from induced pluripotent stem cells (iPSCs) (Lessard et al., 2017). It would be exceedingly useful to construct a new generation of highly active, robust, generic and versatile CRISPR arrayed libraries with relatively small size to overcome the limitations mentioned above, thereby enabling the study of phenotypes that yet cannot be addressed with the currently existing libraries.

The use of multiple sgRNAs targeting each gene confers improved potency and robustness to gene-perturbation screens (Chavez et al., 2016; Erard et al., 2017a; McCarty et al., 2020) and reaches saturation at four sgRNAs per gene (Sanson et al., 2018). Several methods can be used to assemble multiplexed sgRNAs into one single vector (McCarty et al., 2020), but they require many steps including gel purification, colony picking and plasmid sequencing. These manipulations are not amenable to scalable automation, and this limits their usefulness for cloning high-throughput arrayed libraries cost-effectively and efficiently. To circumvent these limitations, we have developed APPEAL (Automated-liquid-Phase-Plasmid-assEmbly-And-cLoning) which assembles four specific sgRNAs targeting the same gene, driven by four different promoters, into a single vector. APPEAL leverages an antibiotic resistance switch between the precursor plasmid backbone and the final 4sgRNA vector to eliminate the necessity of colony-picking, enabling cost- and time-effective liquid-phase cloning of large numbers of plasmids. We developed a custom sgRNA design algorithm to select highly specific sgRNAs optimized to tolerate most common polymorphisms, and preferentially chose four non-overlapping sgRNAs to maximize their synergistic effect. The T.spiezzo and T.gonfio libraries consist of 19,936 and 22,442 plasmids and target 19,820 and 19,839 human protein-coding genes respectively. On average, ~90% of the plasmid population of each well contains at least three intact sgRNAs. Deletion, activation and epigenetic silencing (CRISPRoff) experiments showed efficacy of 75%-99%, up to 10,000x, and 76%-92% respectively. Lentiviral packaging yielded titers of ~$10^7$/ml. We then investigated the effect of individual activation of each human transcription factor (n=1,634) on the expression of the cellular prion protein $PrP^C$, and identified 24 upregulators and 12 downregulators of $PrP^C$ expression. Thus, the T.spiezzo and T.gonfio libraries represent powerful tools for human genome-wide screens of protein-coding genes.

## Results

*The APPEAL Cloning Method*

Traditional cloning processes require the isolation and verification of single bacterial colonies because the original vector and undesired recombinants may contaminate the desired output. However, colony picking is not easily automatable and reduces the throughput necessary for the simultaneous generation of large numbers of plasmids. Therefore, we have developed APPEAL (Automated-liquid-Phase-Plasmid-assEmbly-And-cLoning), a technology allowing cloning of plasmids, transformation and growing of bacteria in liquid phase, thereby eliminating single colony-picking. Thanks to a twin sequential antibiotic selection in the starting precursor vector (ampicillin) and the final plasmid (trimethoprim), the cloning fidelity of APPEAL reaches levels similar to traditional cloning methods.

We used APPEAL to assemble four sgRNAs, each followed by a distinct variant of tracrRNA and driven by a different ubiquitously active Type-III RNA polymerase promoter (human U6, mouse U6, human H1, and human 7SK) into a single vector (**Figure 1A and Figure S1**) (Adamson et al., 2016; Kabadi et al., 2014; Sanson et al., 2018). The 4sgRNA vector includes a puromycin and TagBFP selection cassette, lentiviral packaging and PiggyBac (PB) transposon elements enabling selection and genomic integration by multiple routes (Metzakopian et al., 2017) (**Figure 1A**).

The four sgRNAs were individually synthesized as 59-meric oligonucleotide primers comprising the 20-nucleotide protospacer sequence and a constant region including amplification primer annealing sites (**Figure S1A**). In three distinct polymerase chain reactions (PCR), the primers were mixed with corresponding constant-fragment templates to produce three individual amplicons. These amplicons and the digested empty vector (pYJA5) contain directionally distinct overlapping ends (approximately 20 nucleotides) enabling the assembly via Gibson cloning (Gibson et al., 2009) (**Figure 1A** and **Figure S1**).

During the cloning process, the antibiotic used for bacterial selection was switched from ampicillin to trimethoprim. In the precursor vector pYJA5, the beta lactamase gene (AmpR) which codes for ampicillin resistance was designed to be flanked by two *BbsI* restriction sites and was removed to minimize the size of final 4sgRNA plasmids for the following Gibson assembly with the three amplicons described above. The trimethoprim dihydrofolate reductase resistance gene (TmpR) was incorporated into the first amplicon and positioned between the sgRNA1 and sgRNA2 cassettes leading to an antibiotic resistance cassette switch of the final construct from ampicillin to trimethoprim **(Figure 1A and Figure S1**) (Adamson et al., 2016; McCarty et al., 2020) avoiding the need of spreading bacterial transformants onto agar plates and subsequent colony picking. These characteristics make APPEAL completely automatable and therefore suitable for cost-effective large-scale liquid-phase generation of arrayed plasmid libraries.

To test the accuracy of APPEAL, we generated the sgRNA-containing amplicons and cloned them

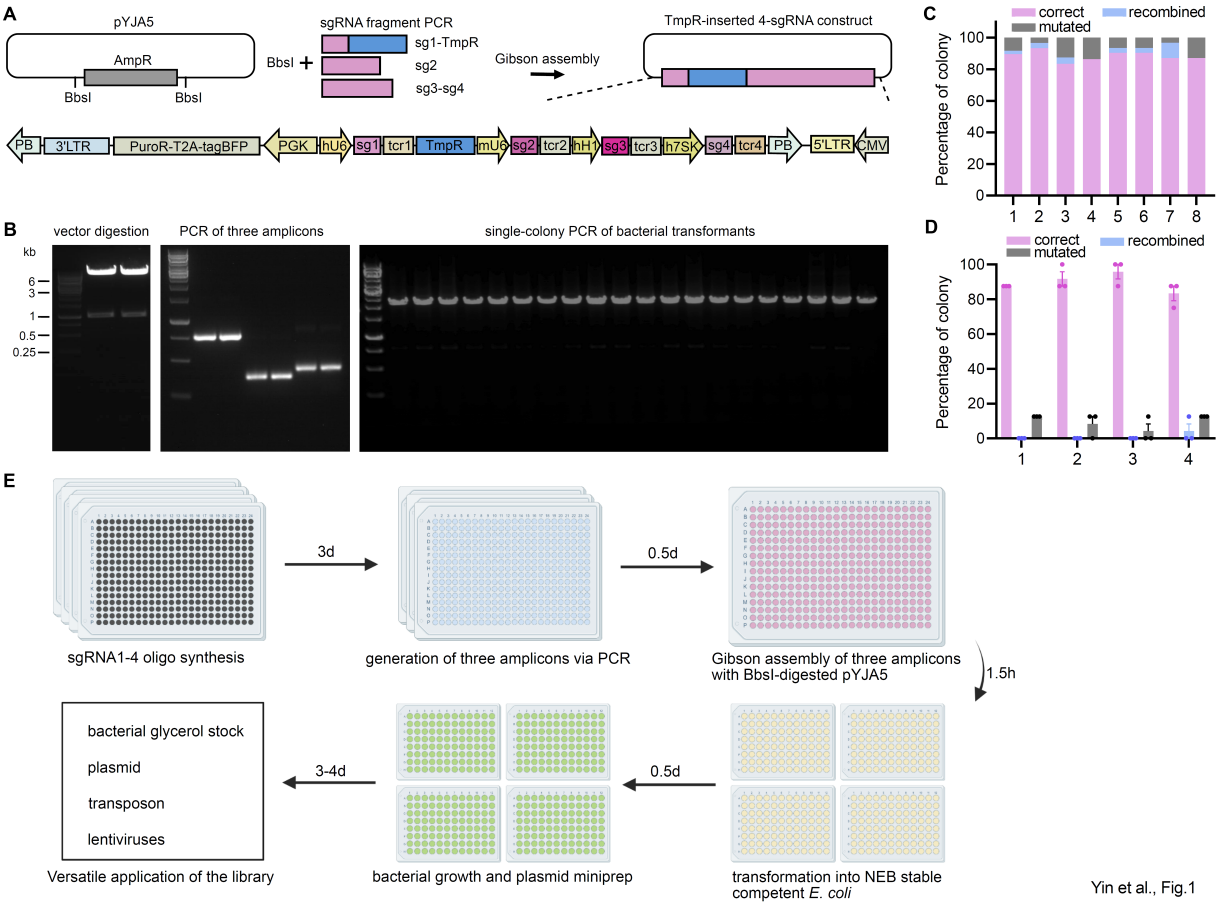



**Figure 1. The APPEAL cloning method. A**, Scheme of APPEAL cloning method and the final 4sgRNA plasmid. The precursor vector pYJA5 was digested with the type-II restriction enzyme *Bbs*I to remove the ampicillin resistance element beta-lactamase (AmpR). sgRNA1-4 were individually incorporated into three amplicons, the first of which includes the trimethoprim dihydrofolate reductase resistance gene (TmpR), by PCR. The digested vector and the three amplicons sharing distinct overlapping ends (~20 base-pairs) were sequentially Gibson-assembled to form the 4sgRNA-pYJA5 plasmid. Only the bacteria harboring the desired 4sgRNA grew in the presence of trimethoprim. hU6, mU6, hH1 and h7SK are ubiquitously expressed RNA polymerase-III promoters. sg, sgRNA; PB, piggyBac transposon element; PuroR, puromycin resistance element; tcr, tracrRNA. **B**, Representative images of pYJA5 restriction fragments, three-fragment PCRs, and single-colony PCR of APPEAL cloning products after transforming into competent *E. coli* and selection with trimethoprim. *Bbs*I digestion of pYJA5 yielded ~1–kilobase (kb) band of the AmpR element and ~7.6-kb band of the linearized vector (left). After PCR with corresponding sgRNA primers, the three amplicons showed the expected size of 761, 360 and 422 bp on agarose gels, respectively (middle). Single-colony PCR with primers flanking the 4sgRNA expression cassettes of APPEAL cloning products in transformed bacteria plate yielded the expected size of 2.2kb in all cases (right). **C**, Percentage of correct 4sgRNA, recombined and mutated 4sgRNA plasmids in 8 independent APPEAL experiments with distinct 4sgRNA sequences. Twenty-two or more colonies were tested in each experiment. **D**, Percentage of correct, recombined and mutated 4sgRNA plasmids in four APPEAL experiments. Each dot represented an independent biological replica consisting of eight colonies (n=24; Mean ± S.E.M.). **E**, Timeline of scaled-up APPEAL cloning in high-throughput format. Approximated time required for generation of one 384-well plate of 4sgRNA plasmids were indicated in hours (h) or days (d). 5-9 plates of 384-well plate of plasmids were generated by 3 full-time employees per week.

into the digested pYJA5 vector using Gibson assembly (**Figure 1B**). After transformation, colony PCR was performed with primers flanking the 4sgRNA insert. On agarose-gel electrophoresis, all PCR products from the tested colonies showed the expected size of 2.2 kilobases, suggesting correct assembly of all three fragments and backbone (**Figure 1B**). We then sequenced single colonies from eight independent cloning procedures (≥22 colonies/procedure). We found that all colonies

showed the desired antibiotic selection switch, and 83%-93% of colonies showed correct fragment assembly in the desired order, resulting in the final 4sgRNA vector (**Figure 1C**).

The presence of repeated sequences may lead to recombination and to the generation of undesired plasmids. To minimize this effect, in the 4sgRNA vectors each sgRNA is driven by a different PolIII promoter and is followed by a unique tracrRNA variant. In the eight cloning trials described above, we found that 0%-10% of tested colonies harbored recombined plasmids (**Figure 1C**). Further, we quantified the colonies with mutation(s) in the 4sgRNA region and found that the frequency of mutated plasmid in the eight cloning trials was 3%-14% (**Figure 1C**). Such mutations may result from errors in oligonucleotide synthesis and the tolerance of DNA mismatches by the Taq DNA ligase during Gibson assembly (Gibson et al., 2009; Lohman et al., 2016). We further tested the robustness of the cloning efficacy of APPEAL by repeating four of the eight cloning trials three times. The percentages of colonies with the correct, recombined, or mutated plasmid were comparable to the previous trials (**Figure 1D**). We thus conclude that the APPEAL cloning method vastly increases the selective pressure for the desired end product and is appropriate for high-quality plasmid generation in liquid phase without resorting to the isolation of single bacterial colonies.

*Scaling up APPEAL to high-throughput Cloning*

In order to perform APPEAL in a high-throughput mode (**Figure 1E**), all steps (sgRNA1-4 containing oligonucleotide synthesis, PCR of three amplicons, and Gibson assembly of the three amplicons with *Bbs*I-digested pYJA5) were performed in 384-well plates. Subsequently, each plate containing the Gibson assembly reaction products was transferred to four deep-well 96-well plates for transformation into recombination-deficient chemically competent *E. coli*. After bacterial cultivation, magnetic bead-based plasmid minipreparation was performed in the same microplates using custom equipment developed in-house. Hence the entire procedure could be performed in liquid phase using automated liquid-handling equipment. This technique allowed us to construct >40,000 individual plasmids with a reduced workforce and within a reasonable production time (~2,000 plasmids/week). The final products can be used as bacterial glycerol stocks, plasmids, transposons and lentiviral stocks (**Figure 1E**).

*Superiority and Robustness of 4sgRNAs in Gene Activation and Ablation*

To test the efficiency and robustness of the 4sgRNA approach in CRISPR-mediated gene perturbation, we tested the efficiency of gene activation (CRISPRa) and ablation (CRISPRko) using the APPEAL-cloned 4sgRNA vector for several target genes in human embryotic kidney cells HEK293 . For CRISPRa, we chose the genes *ASCL1, NEUROD1* and *CXCR4* which show low, moderate and high baseline (Chavez et al., 2016). We found that co-transfection of individual sgRNAs with the

CRISPR activator dCas9-VPR resulted in inefficient or moderate gene activation, whereas the 4sgRNA vector co-transfection with dCas9-VPR significantly increased target gene expression **(Figure 2A**). This is consistent with the previous finding that multiple sgRNAs act more potently than single sgRNAs (Chavez et al., 2016). To further test the robustness and efficacy of the 4sgRNA in gene activation, we tested genes (including the protein-coding genes *HBG1*, *KLF4*, *POU5F1*, *ZFP42*, *IL1R2*, *MYOD1*, *TINCR*, and the long non-coding RNA coding genes *LIN28A*, *LINC00925*, *LINC00514*, and *LINC00028*) that were difficult to activate before the invention of the synergistic activation mediator (SAM) CRISPR activator (Konermann et al., 2013; Mali et al., 2013; Perez-Pinera et al., 2013). With the 4sgRNA approach, we successfully activated the genes **(Figure 2B**). For CRISPRko, we chose to assess sgRNA efficacy by live-cell immunostaining and flow cytometry of the cell-surface proteins CD47, IFNGR1 (also known as CD119) and MCAM (also named as CD146) (Bausch-Fluck et al., 2015) with fluorophore-conjugated primary antibodies. For each gene, 12 single sgRNAs from widely used resources (Hart et al., 2017; Sanjana et al., 2014; Sanson et al., 2018) were tested. Single sgRNAs of all three genes showed variable knockout efficiency (5%-85% for *CD47*, 1%-76% for *IFNGR1*, and 6%-85% for *MCAM*) and 7, 5, and 2 sgRNAs for *CD47, IFNGR1* and *MCAM,* respectively, showed ablation efficiency <60%. In contrast, the respective 4sgRNA plasmids showed reliable knockout efficiency of >80% (**Figure 2C)**.

To further test the efficiency and robustness of the 4sgRNA approach for gene ablation, we examined 9 genes (*ADIPOR1*, *AP2B1*, *CSNK2A1*, *FYN*, *HPRT1*, *TGFBR1*, *APEX1*, *TAZ*, and *PRNP*) for which single sgRNAs were reported to have low or moderate ablation efficiency (de Groot et al., 2018). Through single molecule real-time (SMRT) long-read sequencing of PCR amplicons of the genome-edited cells, we found that 75%-99% of the sequencing reads showed nucleotide deletions for the 9 genes tested (**Figure 2D**). Furthermore, we observed that 4sgRNA-mediated gene knockout resulted in conspicuous deletions of the genomic region between sgRNA cut sites for all nine of the tested genes, as reflected by the size of PCR amplicons in agarose gels (**Figure S2A**). This large genomic DNA ablation increases the likelihood of loss-of-function of the target gene. Together, these results demonstrate the high efficacy and robustness of our 4sgRNA/gene strategy for both gene activation and ablation.

*Lentiviral Packaging and Delivery of the 4sgRNA Vectors*

The 4sgRNA vector is amenable to lentiviral packaging, and the insertion of 4sgRNA expression and trimethoprim selection cassettes (1,800 bases) considerably increases the packing size of lentiviral particles. With a refined transfection ratio of 4sgRNA, packaging and envelope plasmids in HEK293T

cells, we repeatedly obtained titers of $10^7$ transducing units per milliliter (TU/ml) in raw culture-medium supernatants of 24-well plates (**Figure 2E,** see also methods). These viruses greatly increased

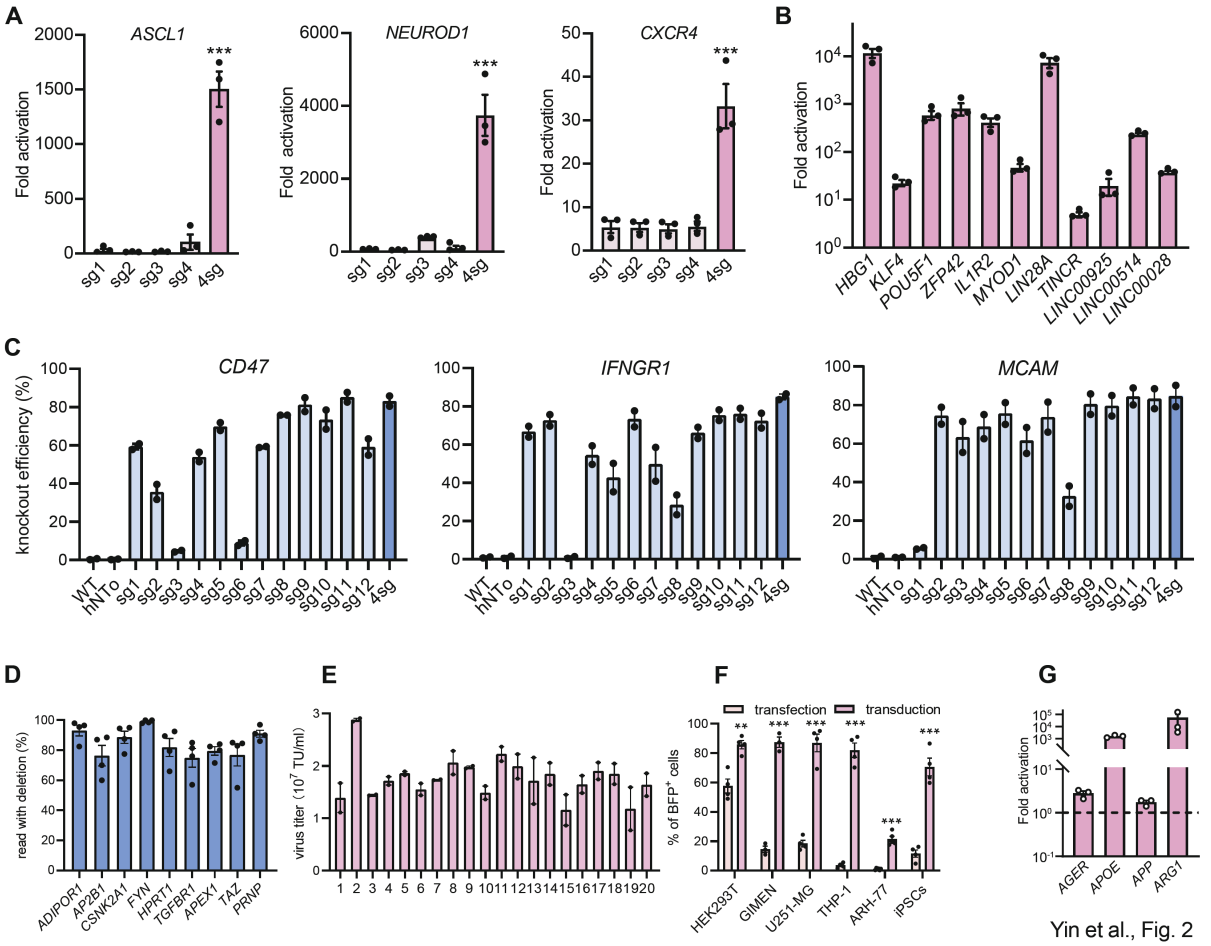



**Figure 2. Robustness of gene activation and ablation with the 4sgRNA approach. A,** Gene activation measured by qRT-PCR of HEK293 cells co-transfected with 4sgRNA plasmids (4sg) generated using the same four single sgRNAs (sg1-4) targeting each gene and dCas9-VPR. Assays were performed 3 days post transfection. Dots: independent experiments (Mean ± S.E.M.). **B**, 4sgRNA plasmids targeting genes poorly activated before the CRISPR-SAM methods (Konermann et al., 2015) were co-transfected with dCas9-VPR into HEK293 cells. Assays were performed 3 days post transfection. Dots (here and henceforth): independent experiments (Mean ± S.E.M.). **C,** Gene disruption efficiency by single guides vs 4sgRNAs in HEK293 cells, measured by flow cytometry of live cells immunostained with fluorophore-conjugated antibodies against CD47, IFNGR1 and MCAM. To compare gene-disruption variability, we tested 12 single sgRNAs (sg1-12) from the Brunello, GeCKOv2 and TKOv3 libraries (light blue) against APPEAL-assembled 4sgRNA plasmids (dark blue) from 4 randomly selected single sgRNAs. HEK293 cells were co-transfected with single sgRNA or 4sgRNA plasmids (encoding puromycin resistance) and the lenti-Cas9-blast plasmid. Cells surviving selection (3 days, 10 µg/ml blasticidin, 1.5 µg/ml puromycin) were maintained without antibiotics for ~1 week. hNTo, non-targeting control plasmid; WT, untransfected wild-type cells. **D.** Robust ablation of genes inadequately disrupted by single sgRNAs (de Groot et al., 2018). Single sgRNAs were assembled into 4sgRNA plasmids and co-transfected with lenti-Cas9-blast into HEK293 cells (as in **C**). Gene disruption was quantified by single-molecule real-time long-read sequencing of the genomic region covering all sgRNAs target sites. **E**, Titers of lentiviral particles packaged from 20 randomly chosen 4sgRNA APPEAL plasmids. Viral particles were produced in 24-well plates with HEK293T cells, and transduced to further HEK293T cells. TagBFP$^+$ cells were quantified by flow cytometry 3 days post transduction. **F**, Gene delivery efficiency of 4sgRNA vector into poorly transfectable cells, measured by flow cytometry of tagBFP$^+$ cells 3 days post transduction. **G**, Gene activation in neurons derived from human induced pluripotent stem cells, measured by qRT-PCR after transduction of 4sgRNA lentiviruses (multiplicity of infection: 1.4). Target neurons expressed stably dCas9-VPR. Assays were performed at day 7 post infection.

sgRNA delivery rates in non-transfectable cells, including the human lymphocyte-related cell lines

THP-1 and ARH-77, the human neuroblastoma cell line GIMEN, the human glioblastoma cell line U251-MG, and patient-derived iPSCs, as indicated by the fraction of tagBFP$^+$ cells measured by flow cytometry (**Figure 2F**). We further examined the efficiency of gene activation in non-transfectable iPSC-derived neurons (iNeurons, which stably express dCas9-VPR) using lentivirus-mediated delivery of the 4sgRNA vector. Gene activation was generally efficient, and its extent depended on the basal expression levels of the target genes (**Figure 2G**).

*Updated Algorithms for Generic, Specific and Synergetic sgRNA Selection*

To enable both gain-of-function and loss-of-function arrayed CRISPR screens, we generated CRISPRa (termed as T.gonfio, meaning swelling up) and CRISPRko (termed as T.spiezzo, meaning sweeping away) arrayed libraries for human protein-coding genes with the high-throughput APPEAL cloning method (**Figure 1E**). Recently developed sgRNA design algorithms have greatly improved the likelihood of obtaining active sgRNAs (Hanna and Doench, 2020). We decided to use the Calabrese and hCRISPRa-v2 sgRNA sequences for T.gonfio, and the Brunello and TKOv3 sequences for T.spiezzo (Hart et al., 2017; Horlbeck et al., 2016; Sanson et al., 2018) as a baseline to generate our arrayed libraries with a novel algorithm to select the optimal combination of four sgRNAs (**Figure 3A**).

Common DNA polymorphisms in human genomes, such as single-nucleotide polymorphisms (SNPs), concern 0.1% of the genome and may reduce the efficacy of sgRNAs (Lessard et al., 2017). Except for the TKOv3 library, all other existing CRISPR libraries did not consider DNA polymorphisms when selecting sgRNA sequences, hampering their usage on primary patient-derived cells. In order to avoid targeting genetically polymorphic regions, we obtained a dataset compiled from over 10'000 complete human genomes (http://db.systemsbiology.net/kaviar/), which provides the coordinates of genetic polymorphisms aligned to the hg38 reference genome. A guide RNA was flagged as unsuitable if the genomic coordinates of the 20-nucleotide protospacer sequence or the two guanine nucleotides of the protospacer adjacent motif (PAM) were affected by a polymorphism with frequency higher 0.1%.

The GuideScan algorithm (http://www.guidescan.com) can predict off-target effects of sgRNAs with high accuracy, showing a strong correlation with the unbiased genome-wide off-target assay GUIDE-Seq (Tsai et al., 2015; Tycko et al., 2019). sgRNAs with GuideScan scores exceeding 0.2 are generally considered specific (Tycko et al., 2019), and we imposed this constraint for sgRNA selection of our libraries.

While choosing sgRNAs for our libraries, two additional points came into consideration. First, previous libraries chose top-ranking sgRNA sequences based on high on-target efficacy scores and low predicted off-target effects, but this could result in the selection of overlapping sgRNAs whose target positions differed only by a few nucleotides. This was especially common in CRISPRa libraries, due to the limited target window for sgRNAs upstream of the transcription start site (TSS). We investigated whether the proximity between the binding sites of the four sgRNAs might affect their activity and if overlapping or spaced sgRNAs should be preferred. We tested six genes and compared 4sgRNA combinations that were spaced by at least 50 nucleotides against combinations that did not meet this criterion. The use of four non-overlapping sgRNAs (spaced at least 50 nucleotides apart) resulted in significantly higher gene activation, suggesting that spatially unconstrained binding of sgRNA-dCas9-VPR complexes is strongly synergistic (**Figure S3A**). Second, since we generated

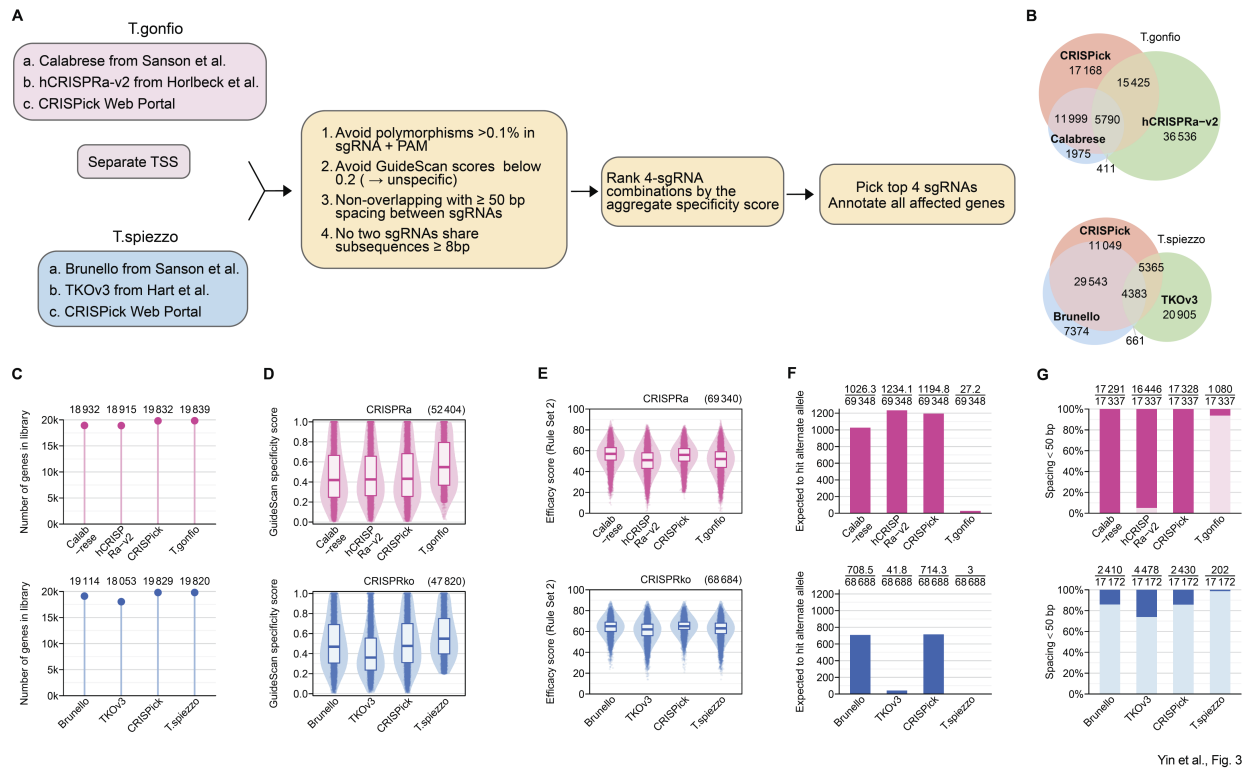



**Figure 3. Algorithms for sgRNA selection and features of the T.spiezzo and T.gonfio libraries. A**, 4sgRNA key elements of the sgRNA selection pipeline. A large pool of potential sgRNAs was provided by combining existing CRISPR libraries and the output of the CRISPick web platform. Each sgRNA was annotated with data on any common genetic polymorphisms affecting the target region, as well as with GuideScan specificity scores as a measure of predicted off-target effects. For the T.gonfio library, a separate 4sgRNA plasmid was designed to target each major TSS (transcription start site). Using a combination of criteria, the best combination of four non-overlapping sgRNAs was chosen (see also Table S1). **B**, For the T.spiezzo and T.gonfio libraries, sgRNAs were selected from existing libraries and resources. The number of sgRNAs derived from each source is shown. **C**, Coverage (number of unique target genes included) of the T.spiezzo and T.gonfio libraries compared with existing libraries and resources. **D**, Cross-library comparison of GuideScan specificity scores for the individual sgRNAs. The top 4 sgRNAs from each library were included, based on their original ranking, for genes that are present in all source libraries. **E**, Cross-library comparison of predicted sgRNA efficacy scores. **F**, Comparison of the number of sgRNAs that are expected to target the alternate allele of a genetically polymorphic region, based on the rate of occurrence in human populations. Only polymorphisms with a frequency >0.1% are considered. **G**, Percentage of 4sgRNA combinations where at least one pair of sgRNAs is spaced fewer than 50 bp apart. Insufficient spacing may lead to steric hindrance between sgRNAs.

our libraries using the Gibson assembly method, if two or more sgRNAs share identical subsequences of 8 nucleotides or more, the prevalence of correct plasmids decreased because of recombination between identical sequences among the four sgRNAs (**Figure S3B and S3C**).

The efficacy of CRISPRa-mediated gene activation relies on sgRNAs targeting a narrow window of 400 base pairs (bp) upstream of TSS of a gene (Gilbert et al., 2014; Sanson et al., 2018). Many genes have more than one TSS that may exhibit different activity in different cell models (Consortium et al., 2014; Sanson et al., 2018). Therefore, the T.gonfio library targets each major TSS with an individual 4sgRNA plasmid (we did not separate TSSs that were spaced fewer than 1000 bp apart). Because some genes or TSSs did not have four sgRNAs that fulfilled these requirements, we supplemented the above-mentioned libraries with sgRNAs from the CRISPick web portal (https://portals.broadinstitute.org/gppx/crispick/public), which designs sgRNAs with the same algorithm that was used for the Calabrese and Brunello libraries for CRISPRa and CRISPRko, respectively.

Finally, after filtering with the above four constraints, all possible combinations of four sgRNAs targeting a gene/TSS were ranked by their aggregate specificity score, enabling the selection of sgRNA sequences with minimized potential off-target effects (**Figure 3A**).

*Features of the T.spiezzo and T.gonfio libraries*

The T.gonfio and the T.spiezzo libraries include 22,442 plasmids and 19,936 plasmids, respectively. Each library contains 116 non-targeting plasmids for control, and is organized into thematic sublibraries (**Table 1**). Transcription factor, secretome and G protein-coupled receptor (GPCR) sublibraries were strictly defined according to current gene catalogs (Lambert et al., 2018; Uhlen et al., 2019). Other sublibraries were based largely on the categories defined by the pooled library hCRISPRa-v2 (Horlbeck et al., 2016). sgRNAs selected via our updated algorithm for our two libraries originated mainly from previously published libraries (**Figure 3B**). We achieved coverages of 19,839 and 19,820 human protein-coding genes with the T.gonfio and T.spiezzo libraries, respectively, which exceed those of the individual pooled libraries from which most sgRNAs were adopted (**Figure 3C**). In the T.gonfio library, 17,528 genes were represented by only one plasmid targeting a single major TSS, whereas for 2,311 genes, multiple plasmids were included to target two or more TSSs (**Figure S3D**). Among the 19,820 genes targeted by the T.spiezzo library, the size of expected deletions in the human genome ranges between tens of base pairs to hundreds of kilobase pairs (**Figure S3D**). By excluding sgRNAs with GuideScan scores <0.2, we enriched for specificity without sacrificing the predicted efficacy **(Figure 3D and 3E**). Importantly, both the T.spiezzo and T.gonfio libraries show a significant improvement in targeting generic regions of the genome by avoiding genetically polymorphic regions **(Figure 3F and S3E**).

Our algorithm for sgRNA selection greatly increased the proportion of sgRNAs spaced at least 50

nucleotides apart (compared with simply selecting the top four sgRNAs from reference libraries) (**Figure 3G**) and thus potentially further increased the activity of our 4sgRNA CRISPRa plasmids, as reflected in Figure S3A. Furthermore, in our libraries, we were able to completely avoid 4sgRNA combinations that shared any identical sub-sequences of 8 base pairs or more in length (**Figure S3F**), thus ensuring minimal spacing for all genes.

In our sgRNA selection pipeline, we avoided unspecific sgRNAs with multiple perfect-match binding sites in the genome, wherever possible. However, when targeting families of closely related, paralogous genes, there were often no specific sgRNAs to choose from. For the sake of simplicity, we nevertheless created a separate 4sgRNA plasmid for each protein-coding gene that possessed its own unique Entrez gene identifier. In the T.spiezzo library such unspecific sgRNAs were mostly excluded, whereas in the T.gonfio library the proportion of sgRNAs with off-site targets (0.8%) was comparable to the reference pooled libraries – most likely owing to the narrow target window around the TSS, which limited potential target sites (**Figure S3G**). Furthermore, CRISPRa activation of unintended genes may also occur if two genes locate on opposite strands of the genome and share a bidirectional promoter region. Such *on-site* targets are an unavoidable consequence of structure of the genome, and were much more common than *off-site* off-target effects. Indeed, we observed that when considering a window of one kilobase surrounding the TSS, around 20% of CRISPRa sgRNAs affected additional protein-coding or non-coding genes. This proportion was almost identical in all libraries we examined, including T.gonfio (**Figure S3H**). All sgRNAs that affect any genes other than the intended gene have been annotated (**Supplemental Dataset 1 and Figure S3I**).

*Sequencing the T.spiezzo and T.gonfio Libraries*

To assess the quality of the T.gonfio and T.spiezzo libraries, we amplified the 4sgRNA-expression cassettes in each well with barcoded primers and then subjected the pools of amplicons to single-molecule real-time (SMRT) long-read (2.2-kilobase) sequencing (Pacific Biosystems) **(Figure S4A**). To quantify technical errors, 74 single-colony-derived (each with distinct 4sgRNA sequences) and sequence-validated plasmids were included in each round of sequencing. The median read count (at CCS7 quality) per plasmid across both libraries was 86; we obtained at least 10 reads for 98.7% of plasmids, and at least one read for 99.9% of plasmids (**Figure S4B**).

Mutations, deletions and recombinations are expected to occur in a minority of plasmids constructed using the Gibson assembly method; because the APPEAL procedure does not rely on colony-picking, these alterations typically affect only a fraction of the plasmid pool in each well. This heterogeneity could be precisely quantified by single-molecule long-read sequencing, which permits a linked analysis of all four promoter, protospacer and tracrRNA sequences. An sgRNA was considered correct only if the 20-nucleotide sgRNA and the following tracrRNA sequence was present and entirely error-

free. Across both libraries, the majority of plasmids contained correct sequences for all four sgRNAs (**Figure 4A and 4B**). When considering the median across all wells in the libraries, the percentage of reads with at least one, two, three or four correct sgRNAs was 98%, 94%, 92% and 78% for the T.gonfio library, and 98%, 92%, 89% and 76% for the T.spiezzo library, respectively (**Figure 4A**). At the 5$^{th}$ percentile of plasmids (i.e., worse than 95% of wells in the library), the fraction of reads with at least three correct sgRNAs remained 77% for the T.gonfio library, and 71% for the T.spiezzo library (**Figure 4B**). Across both libraries, 99.7% of wells passed the minimal quality standard of >50% reads with at least one correct sgRNA (76 wells failed to meet this standard, including the 38 wells with zero CCS7 reads). We thus observed acceptable error rates for the vast majority of wells in both libraries. When considering the four sgRNAs individually, the median percentage of correct reads

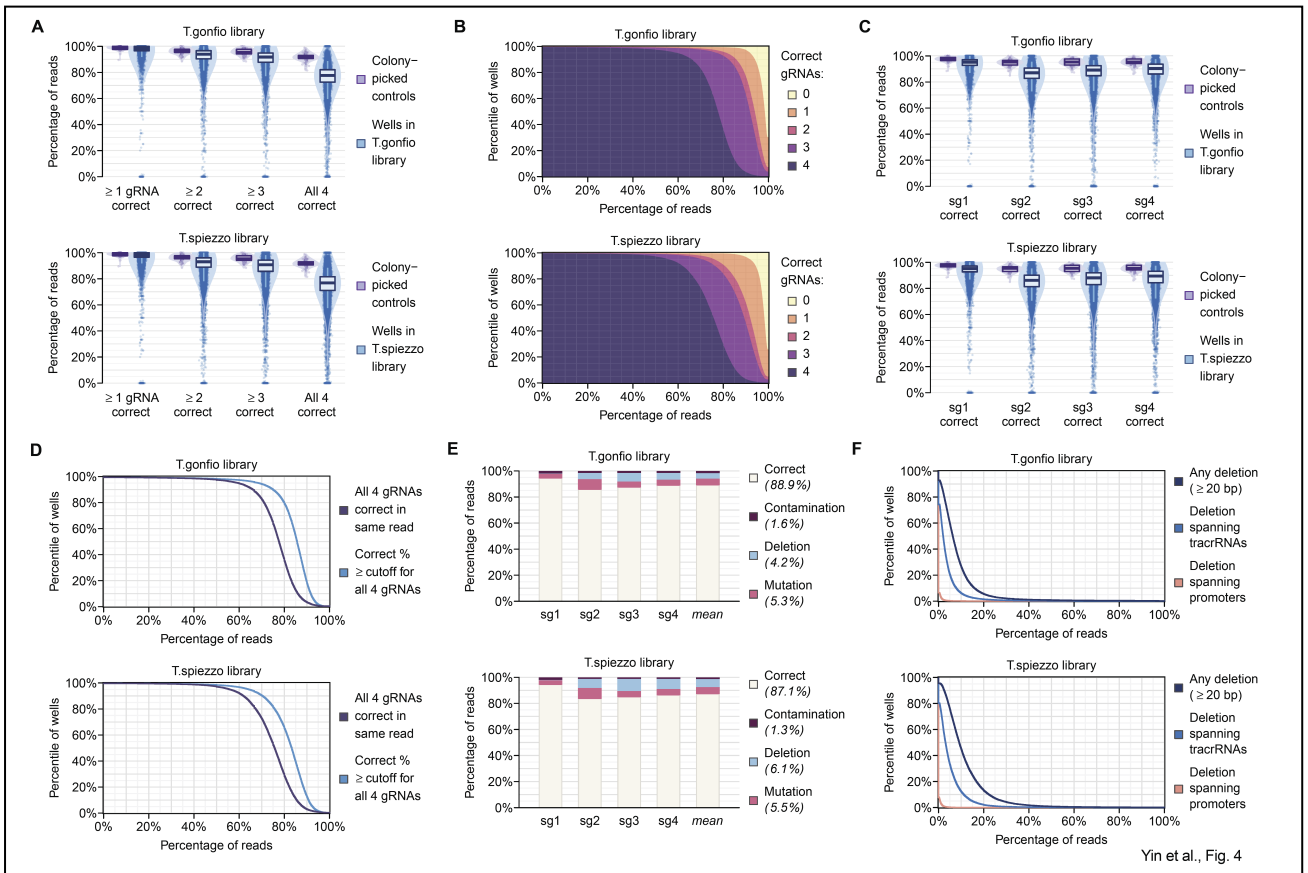



**Figure 4. Genome-wide sequencing of the T.spiezzo and T.gonfio libraries**. **A**, Percentage of reads with 0, 1, 2, 3, or 4 correct sgRNAs for each well in the T.spiezzo and T.gonfio libraries (quantitative SMRT long-read sequencing). To assess technical errors, we added barcoded amplicons of 74 single-colony-derived, sequence-validated 4sgRNA plasmids as internal NGS controls. **B**, Cumulative distribution of the percentage of reads with 0-4 correct sgRNAs in each well of the T.spiezzo and T.gonfio libraries. The box plot denotes the median and interquartile range; the whiskers indicate the 5th and 95th percentile. **C**, Percentage of correct sgRNA-1, sgRNA-2, sgRNA-3, and sgRNA-4 cassettes among the plasmid pools in each well of the T.spiezzo and T.gonfio libraries. **D**, Percentage of reads with four entirely correct sgRNAs in the same vector (black) and minimum percentage threshold passed by each of the four sgRNAs individually (blue) considering the entire pool of plasmids in each well. **E**, Mean percentage of mutations, deletions, and cross-well contaminations in the T.spiezzo and T.gonfio libraries. **F**, Cumulative distribution of plasmids with recombination in each well of the T.spiezzo and T.gonfio libraries.

was ≥85% for all four sgRNAs in both the T.gonfio and T.spiezzo libraries (**Figure 4C**). When performing an arrayed screen, each cell typically receives multiple copies of the 4sgRNA plasmid, so that four correct sgRNAs will still be expressed, even if some copies contain mutations. Indeed, while 65% of wells in the T.gonfio library had ≥75% reads with four entirely correct sgRNAs (within the same read), when each sgRNA is considered separately, 90% of wells were ≥75% correct for each of the four individual sgRNAs (**Figure 4D**). This indicates that mutations may be compensated for by other clones in the same well. We performed an alternative analysis where the promoter sequences were also considered; an sgRNA was considered correct only if the preceding promoter sequence was ≥95% correct. Despite this more stringent criterion, the percentages of correct sgRNAs were very similar to those described above (**Figures S4C and S4D**).

Incorrect sgRNAs were classified as contaminated (matching sgRNAs from other wells), deleted, or mutated (**Figure 4E**). Contaminations were rare, with a mean of 1.6% contaminating sgRNAs in the T.gonfio library, and 1.3% in the T.spiezzo library. Large deletions (involving more than 50% of the sequences of sgRNA and tracrRNA) affected 4.1% of sgRNAs in T.gonfio and 6.1% of sgRNAs in T.spiezzo. Many deletions resulted from recombination between two tracrRNAs (4.1% of reads were affected by deletions spanning from tracrRNA to tracrRNA, whereas only 0.1% of reads contained deletions spanning two promoters) (**Figure 4F**); this is explained by the homology between tracrRNA sequences, which increases their propensity for recombination, whereas the four promoter sequences share less similarity. The mean percentage of plasmids with a deletion affecting at least one sgRNA was 8.1% in the T.gonfio library and 11.4% of sgRNAs in the T.spiezzo library. Finally, mutations affected 5.3% of sgRNAs in the T.gonfio library and 5.5% of sgRNAs in the T.spiezzo library. However, these estimates include sequencing errors and may overestimate the error rate. Importantly, less than 0.1% of sgRNA sequences comprising mutations acquired novel off-target activities (**Figures S4E and S4F**). An entirely correct sequence was observed for 88.9% and 87.1% of sgRNAs for T.gonfio and T.spiezzo, respectively. We conclude that APPEAL cloning resulted in the generation of these libraries with low overall error rates.

*Benchmarking of Individual 4sgRNA Ablation Plasmids with Various Delivery Methods in Multiple Cell Models*

Next, we sought to benchmark the 4sgRNA plasmid approach against commercially available CRISPR reagents (lentiviral packaged sgRNAs and synthetic guide RNAs) in various cell models including immortalized human colon cancer cell line HCT116, iPSCs, and kidney organoids using several delivery methods (transduction, transfection, and electroporation) (**Figure 5A**). Due to the availability of all related CRISPR knockout reagents, we focused on knockout assays and chose

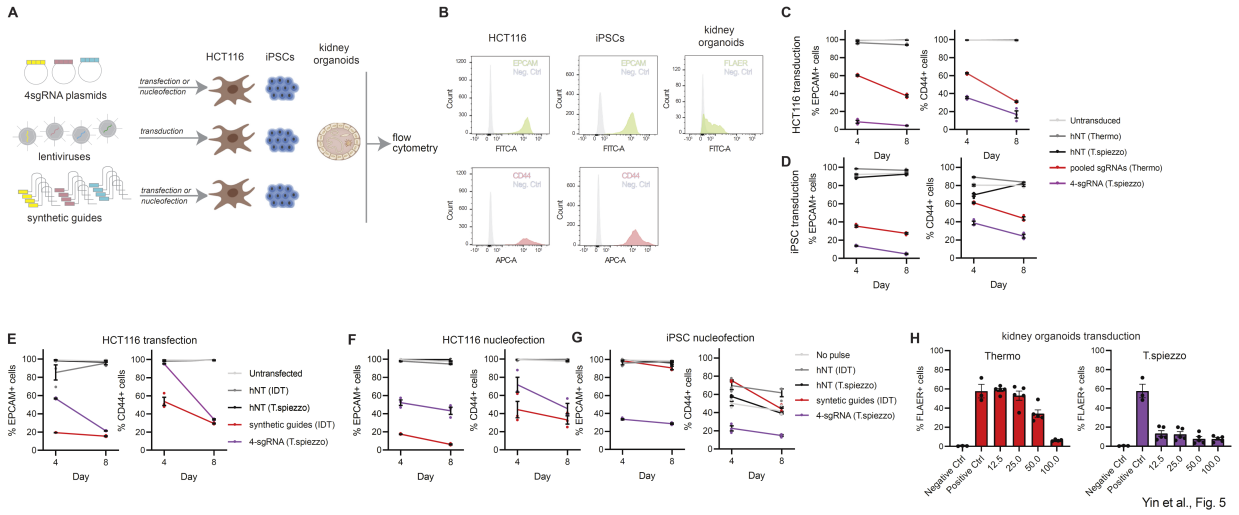



**Figure 5: Benchmarking the 4sgRNA knockout approach in various cell models.**

**A,** Schematic of the experiment. 4sgRNA plasmids, synthetic guide RNAs, or lentivirally packaged sgRNAs were either transfected, nucleofected or transduced in Cas9 expressing HCT116, iPSCs and nephron progenitor cells (NPCs, which were matured to kidney organoids). **B,** Flow-cytometry histograms of Cas9-expressing HCT116 and iPSC cells immunostained with anti-EPCAM-FITC (*green; top left and middle*) or anti-CD44-A647 (*red; bottom left and middle*) antibodies, and single-cell-dissociated kidney organoids stained with fluorescently labelled aerolysin (FLAER) (*green, top right*). **C** and **D,** Comparing the percentages of EPCAM or CD44 positive HCT116-Cas9 (**C**) and iPSC-iCas9 cells (**D**) transduced with lentiviruses carrying the 4sgRNA vector (*T.spiezzo*) or a mixture of four individual pre-packaged lentiviruses (*Thermo*) targeting *EPCAM* or *CD44* to the untransduced (*no virus*) and non-targeting (*hNT*) controls after 4-8 days post transduction (*n*=3; error bars represent S.E.M.). **E,** Comparing the percentages of EPCAM or CD44 positive HCT116-Cas9 cells transfected with 5 µg of the 4sgRNA vector (*T.spiezzo*) or 10 µM of four individual synthetic guide RNAs (*IDT, Integrated DNA Technologies*) targeting *EPCAM* or *CD44* compared to untransfected and non-targeting (*hNT*) control after four and eight days post transfection (*n*=3; error bars represent S.E.M.). **F,** Comparing the percentages of EPCAM or CD44 positive HCT116-Cas9 cells electroporated with 5 µg of the 4sgRNA vector (*T.spiezzo*) or 10 µM of four individual synthetic guide RNAs (*IDT*) targeting *EPCAM* or *CD44* compared to the no pulse and non-targeting (*hNT*) controls after four and eight days post electroporation (*n*=3; error bars represent S.E.M.). **G,** Comparing the percentages of EPCAM or CD44 positive iPSC-iCas9 cells electroporated with 5 µg of the 4sgRNA vector (*T.spiezzo*) or 10 µM of four individual synthetic guide RNAs (*IDT*) targeting *EPCAM* or *CD44* compared to the no pulse and non-targeting (*hNT*) controls after four and eight days post electroporation (*n*=3; error bars represent S.E.M.). **H,** Bar plots showing percentage of fluorescently labelled aerolysin (FLAER) positive cells dissociated from kidney organoids transduced with lentiviruses carrying the 4sgRNA vector (*T.spiezzo*) or four individual pre-packaged lentiviruses (*Thermo*) targeting *PIGA* at increasing viral volumes compared to the unstained (negative ctrl) and un-transduced (positive Ctrl) controls (*n*=4-5; error bars represent S.E.M.).

Epithelial Cell Adhesion Molecule (*EPCAM*), cell-surface glycoprotein *CD44,* and phosphatidylinositol glycan anchor biosynthesis class A (*PIGA*) as targets, based on their expression and possible detection with live-cell immunostaining and flow cytometry quantification in the cell models that are used (**Figure 5B**). First, we transduced Cas9 expressing HCT116 (HCT116-Cas9) or doxycycline-induced Cas9 expressing iPSC (iPSC-iCas9) cells with either the lentiviral packaged 4sgRNA vector

or with a pool of four individually packaged sgRNAs (ThermoFisher) at a multiplicity of infection (MOI) of 5. Transduction of both reagents resulted in a significant reduction of EPCAM and CD44 detection in a time-dependent fashion, however, our 4sgRNA-plasmid-derived lentiviruses resulted in a more pronounced knockout efficiency in both cell models at four and eight days post transduction (**Figure 5C and 5D**). Next, we transfected HCT116-Cas9 cells (iPSCs were not transfected due to their poor transfectability) either with our 4sgRNA plasmids or a pool of four individual synthetic sgRNAs (Integrated DNA Technologies). We found that synthetic sgRNAs showed an earlier knockout effect than our 4sgRNA plasmids at day 4 post transfection but both reagents resulted in a similar reduction of EPCAM and CD44 detection at day 8, indicating differential kinetics of gene knockout between plasmids and synthetic sgRNAs (**Figure 5E**). Next, we electroporated HCT116-Cas9 and iPSC-iCas9 cells with either the 4sgRNA vectors or a pool of four individual synthetic sgRNAs (Integrated DNA Technologies) at increasing concentrations. In HCT116-Cas9 cells, both reagents worked in a concentration dependent manner, while the four synthetic sgRNAs approach resulted in a faster and more efficient reduction of EPCAM and CD44 than our 4sgRNA vectors (**Figure 5F; Figure S5A and S5C**). In contrast, in iPSC-iCas9 cells, electroporation of synthetic sgRNAs showed minimal knockout efficacy whereas the 4sgRNA vector resulted in fast and high editing efficiencies, showed by low detection percentages for both EPCAM and CD44 after four and eight days (**Figure 5G; Figure S5B and S5D**).

To further test whether our 4sgRNA vector approach can efficiently edit target genes in complex cellular models, we used an inducible Cas9 iPSC line (Ungricht et al., 2022) to generate nephron progenitor cells (NPCs) and further differentiated them into kidney organoids following an established protocol (Morizane and Bonventre, 2017). The *PIGA* gene, which is essential for the synthesis of glycosylphosphatidylinositol inositol (GPI) anchors was targeted and its editing efficiency was assessed by staining with a non-toxic fluorescently labelled aerolysin (FLAER assay) (Metzakopian et al., 2017). At NPCs stage, cells were transduced with the lentiviral packaged 4sgRNA vector or with a pool of four individually packaged sgRNAs (ThermoFisher) at increasing volumes of viral supernatant and after 48 days the organoids were dissociated into single cells and subsequently stained with FLAER. Lentiviruses carrying the 4sgRNA vector showed already a high knockout efficiency at low lentiviral volumes whereas four individually packaged sgRNAs required a higher volume to achieve a similar knockout efficiency even with equal viral titers (**Figure 5H; Figure S5E**).

Together, these results demonstrate that our library shows equal or superior gene perturbation performance compared to commercially available CRISPR reagents and furthermore underlines the versatility of our 4sgRNA approach regarding various cell models with different delivery methods.

*Transcription Factors Regulating the Expression of the Cellular Prion Protein $PrP^C$*

Prion diseases are devastating, incurable neurodegenerative diseases (Scheckel and Aguzzi, 2018). The cellular prion protein $PrP^C$ is encoded by the *PRNP* gene and is essential for the development

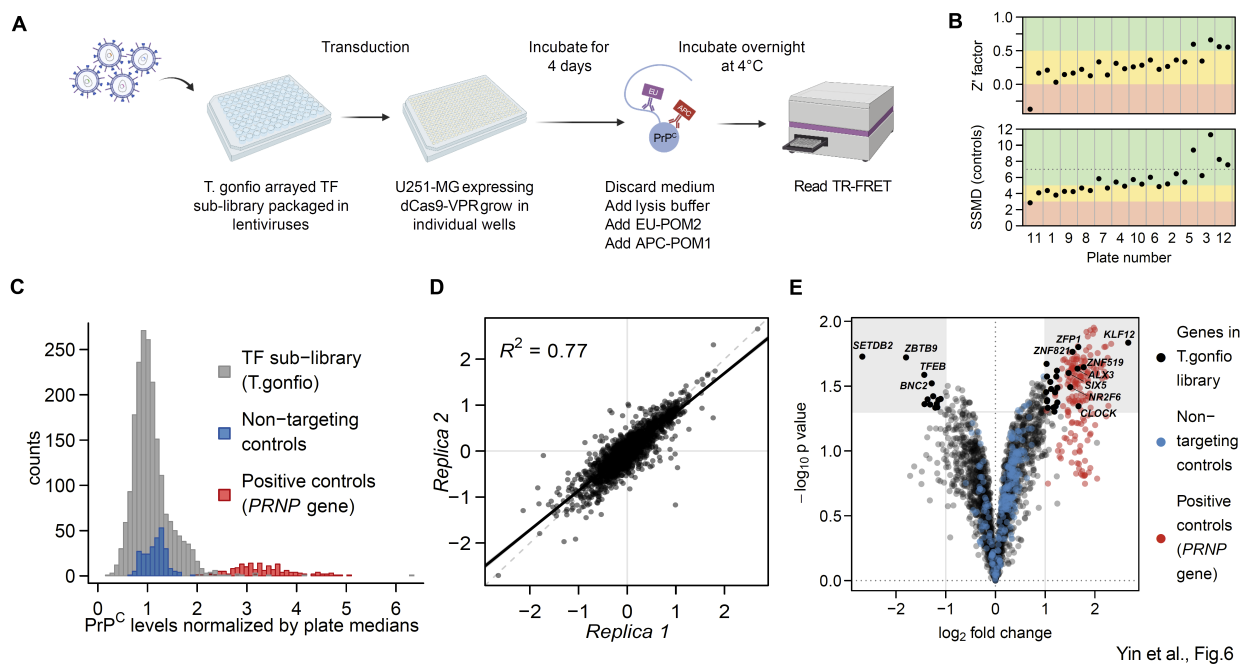



**Figure 6**. **Identifying transcription factors (TFs) regulating $PrP^C$ expression through arrayed T.gonfio TF sublibrary screen. A**, Schematic of the primary $PrP^C$ TF sublibrary screen. **B**, Z' factor and SSMD of each plate from the primary screen. **C**, Distribution of positive controls (4sgRNA targeting *PRNP*), negative controls (4sgRNA non-targeting control) and samples. **D**, Duplicate correlation across samples from the primary screen. **E**, Volcano plot displaying –log10 p values and log2 fold change values across the T.gonfio TF sublibrary. The 36 candidate genes identified are depicted in black.

of prion diseases (Bueler et al., 1993). Previous microRNA and siRNA screens have uncovered a complex pattern of regulated expression of $PrP^C$ (Heinzer et al., 2021; Pease et al., 2019) However, the transcription factor(s) (TFs) controlling $PrP^C$ expression remains unclear. We measured $PrP^C$ expression in a focused arrayed activation screen with the T.gonfio sublibrary encompassing all human TFs (**Figure 6A**). We adopted the previously established biochemical method to detect endogenous $PrP^C$ expression in cell lysate with time-resolved fluorescence resonance energy transfer (TR-FRET) using Europium (EU)-conjugated-POM2 and allophycocyanin (APC)-labelled-POM1, a pair of antibodies binding distinct domains of $PrP^C$, (**Figure 6A**).

We packaged the TF sublibrary of T.gonfio, consisting of 1,634 plasmids, into lentiviruses. We then individually transduced each vector into the human glioblastoma cell line U-251MG stably expressing the CRISPR activator dCas9-VPR with a MOI of 3. Experiments were performed as triplicates in 384-well microplates, each including 14 wells with non-targeting (NT) and 14 *PRNP* targeting controls (**Figure S6A**). Cells were lysed 4 days post transduction; one replica plate was used to determine cell viability with CellTiter-Glo® (Promega), and two replicas were used to assess $PrP^C$ levels by the TR-FRET method (**Figure 6A**). Heatmaps were generated to detect plate gradients and other systematic anomalies (**Figure S6B**). The Z' factor was used to evaluate the discrimination between

positive and negative controls (Zhang, 2011).

Out of the 24 plates, 19 and 4 plates had a Z' factor of 0-0.5 and >0.5, respectively, whereas one plate had a Z'<0 (Zhang, 2011) (**Figure 6B**). The strictly standardized mean difference (SSMD) assessment of the separation between negative and positive controls gave comparable results (**Figure 6B**). This was further confirmed with the finding that distinct levels of $PrP^C$ was detected between non-targeting and *PRNP* targeting controls (**Figure 6C**), indicating that the screen was of sufficient quality to proceed with candidate gene selection. The Pearson correlation coefficient ($R^2$) between duplicates was 0.77 (**Figure 6D**), indicating a satisfactory replicability. Hit calling was based on an absolute $log_2$ fold change of ≥1 and a p-value of ≤ 0.05. Based on these cut-offs, 24 and 12 genes out of 1,634 were found to upregulate or downregulate $PrP^C$ expression, respectively (**Figure 6E**). These results confirm the feasibility and power of these libraries for studying phenotypes of interest in an arrayed format.

*Suitability of the T.gonfio Library to Targeted Epigenetic Silencing (CRISPRoff)*

CRISPR-mediated targeted epigenetic silencing (Amabile et al., 2016; Nunez et al., 2021) is an instrumental loss-of-function perturbation method used as an alternative to knockout for interrogating gene functions, especially for cell models that are sensitive to DNA breakage (e.g. iPSCs) (Ihry et al., 2018). The recently developed CRISPRoff epigenome memory editor has been shown to be very robust and efficient in targeted gene silencing and the memory can persist even in iPSCs-differentiated neurons (Nunez et al., 2021). Interestingly, the sgRNA targeting window of CRISPRoff is quite broad and covers the sgRNA targeting window of both CRISPRa and CRISPRi (Nunez et al., 2021), suggesting that sgRNAs from a CRISPRa library may be able to induce gene silencing with CRISPRoff in combination with a CRISPRoff plasmid. When aligning the sgRNA targeting sequences of our T.gonfio library to those included for CRISPRoff, we found that 96.8% of the T.gonfio sgRNAs target the same targeting window (**Figure 7A**). The residual 3.2% sgRNAs fell within adjacent sequences (<100bp) of this window (**Figure 7A**). This encouraged us to examine the possibility of using T.gonfio for effective CRISPRoff.

We first tested the silencing efficiency of the cell-surface proteins ITGB1, CD81 and CD151 that were assessed in the seminal CRISPRoff study (Nunez et al., 2021). The 4sgRNA plasmids targeting the TSSs of these genes were co-transfected into HEK293T cells along with either CRISPRoff, or with a CRISPRoff mutant carrying a catalytically inactive version of the DNA methyltransferase (Nunez et al., 2021). Transfected cells were cultured for 3 days under puromycin selection and 7

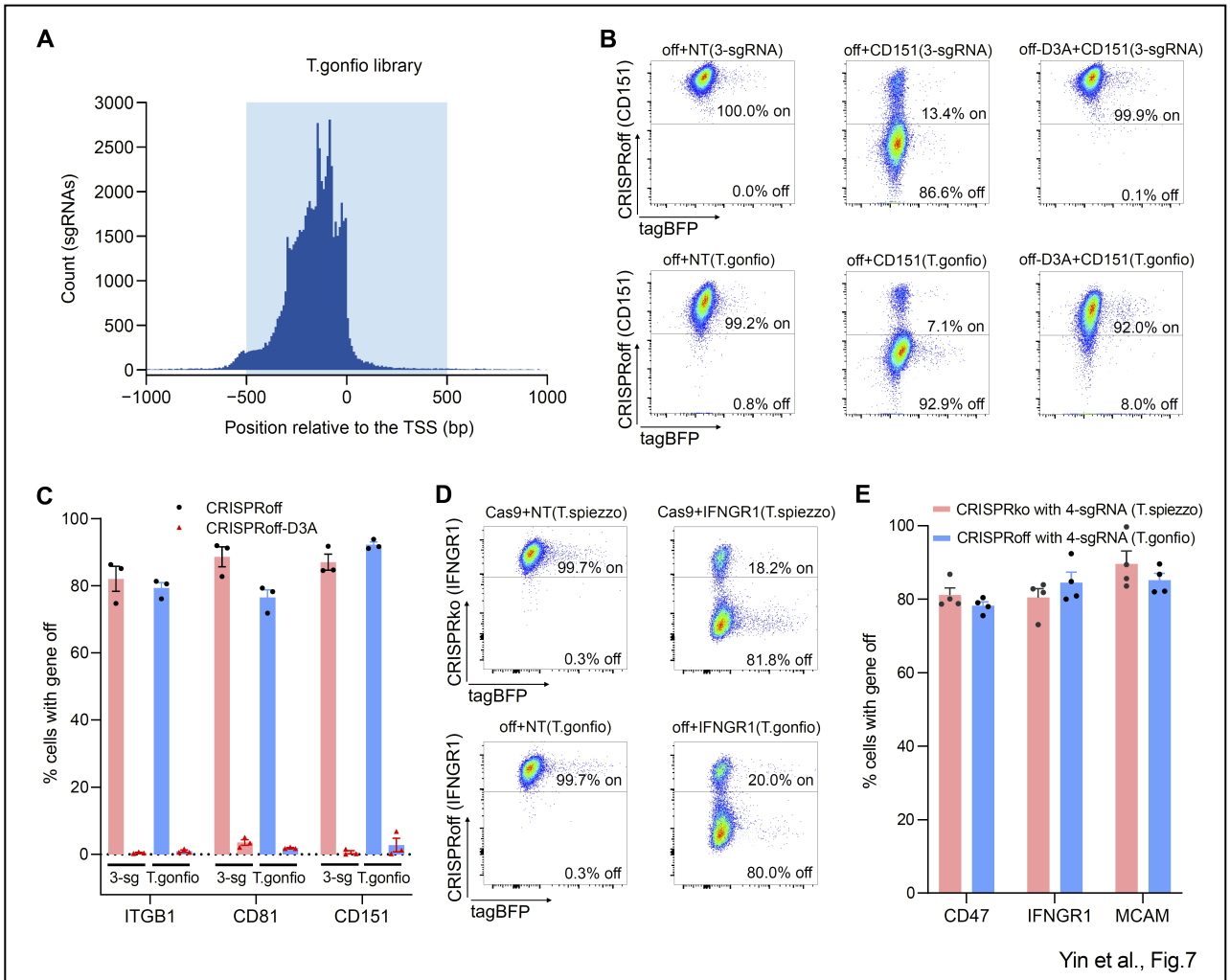



**Figure 7. Targeted epigenetic silencing (CRISPRoff) with the T.gonfio library. A**, Alignment of the T.gonfio sgRNAs sequences to the CRISPRoff targeting window. **B**, An example of flow cytometry measurement of CD151 in HEK293T cells after exposure (10 days) to pools of three separated sgRNAs (3-sgRNA, used in the CRISPRoff study) or the respective T.gonfio 4sgRNA. Off represented the CRISPRoff plasmid (Addgene #167981); off-D3A represented the CRISPRoff mutant carrying a catalytically inactive version of the DNA methyltransferase. **C**, Quantification of percentage of cells with ITGB1, CD81 and CD151 silencing 10 days post CRISPRoff or CRISPRoff-D3A mutant with pools of three single sgRNAs (3-sg) from the published resource or 4sgRNA from the T.gonfio library in HEK293T cells. Each dot represents an independent biological repeat of the assay. Data were presented in Mean ± S.E.M.. **D**, An example of flow cytometry measurement of percentage of cells with IFNGR1 silencing 10 days post CRISPR knockout with 4sgRNA plasmid from the T.spiezzo library or CRISPRoff with 4sgRNA from the T.gonfio library in HEK293T cells. **E**, Quantification of percentage of cells with CD47, IFNGR1 and MCAM silencing 10 days post CRISPR knockout with 4sgRNA plasmids from the T.spiezzo library or CRISPRoff with 4sgRNA plasmids from the T.gonfio library in HEK293T cells. Each dot represents a biological repeat of the assay. Data were presented in Mean ± S.E.M..

days without selection. Then, the cell-surface expression of ITGB1, CD81 and CD151 was determined by live-cell immunostaining with fluorophore-conjugated antibodies and quantified by flow cytometry. To benchmark the efficiency of our 4sgRNA plasmids, a pool of three single sgRNA plasmids used in the CRISPRoff study (Nunez et al., 2021) was used as reference for each gene. Interestingly, we achieved a comparable gene silencing efficiency with our 4sgRNA plasmids compared to the pool of three CRISPRoff individual sgRNA plasmids for all target genes, whereas using the mutant dCas9-DNA methyltransferase complex (CRISPRoff-D3A) resulted in minimal or no reduction of gene expression (**Figure 7B and 7C**).

We then tested three additional cell-surface proteins, CD47, IFNGR1 and MCAM, in the knockout efficiency assay (see **Figure 2C**), to further confirm the feasibility of using the T.gonfio library for CRISPRoff. In addition, we were curious to compare the gene silencing efficacy of T.gonfio by CRISPRoff with that induced from gene knockout via the corresponding T.spiezzo library. While the T.spiezzo plasmids combined with Cas9 induced 80-90% gene ablation as expected, the T.gonfio 4sgRNA plasmids with CRISPRoff also induced a similar extent of gene silencing (**Figure 7D** and **7E**). These data demonstrate that the T.gonfio library can be adopted for efficient epigenetic gene silencing using the CRISPRoff technology.

## Discussion

The endeavor described here addresses the limited availability of arrayed genome-wide libraries, which are essential to the study of complex and non-autonomous phenotypes. Furthermore, we focused on issues that limit the versatility of CRISPR screens, including the variable targeting efficacy of single sgRNAs and the impact of human genomic variability onto editing. While arrayed CRISPR screens may seem exceedingly laborious, we found that they can be performed rapidly by standardizing workflows and deploying inexpensive automation steps such as 384-well pipetting. Crucially, arrayed screens can drastically improve signal/noise ratios and allow for hits to be unequivocally called without any sequencing.

Our aspiration to cover the entire protein-coding genome with ablating, activating and silencing tools entailed the generation of >42,000 individual plasmids. This would have required prohibitive resources, since agar-plating, colony-picking and gel extraction of desired DNA fragments (McCarty et al., 2020) are extremely time-consuming and cannot be easily automated. We have therefore invented APPEAL, a high-fidelity plasmid construction method that does not require bacterial plating and radically simplifies the cloning procedure, enabling the automatable, rapid and cost-effective generation of complex 4sgRNA vectors. What is more, APPEAL is generically applicable to any kind of molecular cloning task. The crucial feature of APPEAL is the placement of a dihydrofolate reductase gene within the clonable insert, which dramatically increases the selective growth of correctly assembled plasmids. When combined with adjustments aimed at minimizing the likelihood of illegitimate recombination events, such as the insertion of unique promoter and tracrRNA sequences for each of the four sgRNA cassettes, APPEAL allowed us to generate ~ 2,000 individual 4sgRNA plasmids per week with high accuracy. Finally, the sgRNA selection algorithm was adapted to ensure that the sgRNAs were non-overlapping and would tolerate to the largest possible extent the DNA polymorphisms found among 10’000 human genomes with minimal off-target effects.

### *Efficacy of CRISPR-Based Gene Perturbation*

Although the algorithms for predicting sgRNA activity are continuously improving (Hanna and Doench, 2020), the efficacy of pooled and arrayed gene-perturbation screens can still be jeopardized by suboptimal guide design and by insufficient numbers of guides for each gene, leading to false-negative hit calls. Conversely, the deployment of multiple individual sgRNAs per gene may prohibitively increase costs due to increased numbers of vectors and cells needed to reach acceptable coverage. By increasing the efficiency and robustness of CRISPR-mediated gene activation and ablation, the integration of 4 sgRNA into each vector reduces dramatically the number of cells needed in for pooled screens. Therefore, the 4sgRNA libraries described here not only enable complex arrayed screens, but can also lower the cost and augment the reliability of pooled screens.

*Versatility of 4sgRNA-Based Libraries*

The design of genetic screens in cell and in vivo has undergone fundamental changes and is continuously improving. To ensure the broadest possible adaptability to multiple experimental protocols, we included two selection markers in our 4sgRNA vector (puromycin resistance and TagBFP) as well as the motifs necessary for lentiviral packaging and transposon-mediated integration. Therefore, users can select the delivery method most appropriate for the experimental model at hand (immortal cell lines, hiPSCs, organoids, or primary cells). Furthermore, each sgRNA is driven by a different housekeeping promoter. Besides ensuring activity in the broadest range of cells and tissues, this design minimizes the risk that the entire construct be transcriptionally silenced by promoter methylation. Moreover, the sgRNA selection algorithm was tuned to identify the least polymorphic regions of each gene, thereby extending the likelihood of perturbation to patient-derived cells that may substantially differ from the human reference genome.

We became interested in exploring whether pooled versions of the arrayed libraries may outperform the existing pooled libraries. We therefore generated pools by mixing the individually purified T.spiezzo and T.gonfio plasmids. Conventional pooled libraries can suffer from inhomogeneous sgRNAs representation (up to 1000-fold) (Gautron et al., 2021; Imkeller et al., 2020) which can reduce sensitivity and signal-to-noise ratio, whereas pooling plasmid arrays allows for strictly controlling the stoichiometry of each component. Moreover, the pooled libraries have a much smaller size compared to the existing libraries that require up to 10 guides per gene. This can not only reduce workload and cost, but enables screens when cell numbers are limiting – which is often a problem with human primary cells.

Finally, we found that the T.gonfio library can be efficiently used for epigenetic silencing (CRISPRoff). This opens the possibility of performing both loss-of-function and gain-of-function screens using the same library in cell lines expressing the appropriate dCas9 proteins, further saving time, cost and labor in the execution of gene-perturbation screens.

*Quality and homogeneity of the 4sgRNA vectors*

By dramatically lowering the tolerance towards incorrect plasmid assemblies, APPEAL eliminates the necessity of isolating clonal bacterial colonies. Consequently, each APPEAL reaction product may potentially represent a polyclonal pool of plasmids. This source of variability was quantitatively assessed by sequencing: in the average well, 90% of the plasmid population contains three or more intact sgRNAs. Some 4sgRNA plasmids showed mutations, mainly in the region of the sgRNA and

tracrRNA sequences, most likely originating from oligonucleotide synthesis and the ligation processes using Taq DNA ligase (the enzyme required for Gibson assembly). Mismatches within the overlapping sequences of the PCR amplicons can be tolerated by Taq DNA ligase during the annealing process, leading to incorrect assemblies during Gibson cloning. The mutation rate found in our sgRNA vector sequences is consistent with the expected error rate occurring during Gibson assembly, leading to mutations in approximately 10% of the plasmids (Gibson et al., 2009).

Despite the use of four different promoters and tracrRNA variants, we observed a recombination between sgRNA expression cassettes in ~10% of reads, resulting in a deletion of the intervening sequence. However, 85% of recombining plasmids retained ≥1 correct sgRNA sequence, and in the median well 99.7% of reads had at least one sgRNA+tracrRNA module that was 100% correct. Thus, the plasmid collections remained functional even in wells affected by recombination events. The average percentage of entirely correct protospacer and tracrRNA sequences for each of the four sgRNAs was ~90%.

Some of the mutations described above may render the construct inactive, or they may lead to off-target effects when a mutated sgRNA binds elsewhere in the genome. However, the latter occurrence was extremely rare, affecting <0.5% of mutated sgRNAs, and targeting additional genes in only ~0.01% of cases (**Figure S4E** and **S4F**). We conclude that the entirety of sequence alterations in the plasmid pools generated by APPEAL had no practical effect apart from reducing the number of active sgRNAs in a minority of plasmids. Notably, the 74 single-colony-derived control plasmids displayed several errors attributable to faulty sequencing. Since it can be plausibly assumed that the APPEAL cloning products show such sequencing errors, the total error rates in T.gonfio and T.spiezzo can be regarded as worst-case limits which most likely overestimate the actual error rates.

*Limitations of the T.spiezzo and T.gonfio libraries*

The delivery of multiple sgRNAs to the same cell may increase the likelihood of off-target effects. We therefore developed an updated sgRNA selection algorithm (Table S1) to adopt the most specific combination of four sgRNAs from existing, well-validated resources. We predict that the T.spiezzo and T.gonfio libraries will enable the identification of hits that may remain unrecognized with existing libraries. In any phenotypic screening approach, subsequent validation of hit genes is required. Many orthogonal approaches exist for a second-round validation of hits identified with our libraries. Thus, a combination of our libraries and orthogonal resources reduces the impact of possible off-target effects and enables powerful and efficient genetic screens.

## Contributorship

J-A.Y. designed, supervised, and coordinated the research, invented the APPEAL cloning method, performed validation experiments of APPEAL with the assistance of A.S. and K.M., developed the homemade Gibson assembly mix for the two libraries with assistance of A.S., produced (~30%) homemade competent cells with Y.W. (~30%), L.Y. (~30%), K.G. (~8%) and A.S. (~2%), transformed the Gibson assembly products (100%) of the two libraries into competent cells together with Y.W. (~50%) and L.Y. (~50%), stored (100%) bacterial glycerol stock of the two libraries into 384-well deep-well plate, developed the 96-well plate deep-well magnetic-beads based plasmid miniprep together with Y.W. and L.Y., setup the lentiviral production of the 4gRNA plasmids together with K.M., performed the 1sgRNA vs 4sgRNA gene activation real-time quantitative PCR with A.S. and K.M., performed 1sgRNA vs 4sgRNA gene knockout efficiency on CD47, IFNGR1 and MCAM together with J.G., analysed the 4sgRNA knockout efficiency obtained from SMRT long-read sequencing, performed transfection and transduction of non-transfectable cells with 4sgRNA vector together with K.M. and M.L., contributed to L.F. for the design of new algorithms for 4sgRNA selection, set up the barcoded 384-well plate library plasmid sequencing with A.S., performed CRISPRoff test experiments on ITGB1, CD81, CD151, CD47, IFNGR1, MCAM together with J.G, analysed data, and wrote the manuscript with A.A. L.F. developed the 4sgRNA selection algorithm and analysed the in-silico features of the libraries, analysed the SMRT long-read sequencing of the two libraries, analysed the arrayed screen data on $PrP^C$, aligned the sgRNA targeting window of T. gonfio library to the targeting window of sgRNAs for efficient targeted epigenetic silencing, wrote the manuscript. M.S. performed the benchmarking experiment on 4sgRNA knockout efficiency with commercially available resources (synthetic sgRNA and lentiviruses) in HCT116, iPSCs and kidney organoids with various vector delivery methods including transfection, transduction and electroporation, analysed the data, and wrote the manuscript. C.T. performed the TF sublibrary arrayed screen TR-FRET experiment and wrote a first draft of the respective section of the manuscript. A.D. performed cell culture and transduction of 4sgRNA lentiviruses into iNeurons, packaged the T.gonfio TF sublibrary plasmids into lentiviruses together with J.T. and S.R. A.S. performed maxiprep, *Bbs*I digestion of the pYJA5 vector, and purification of the digested pYJA5, performed three fragment PCRs of the entire two libraries, Gibson assembly of the three PCR amplicons with digested pYJA5 of the entire two libraries, tested Taq DNA ligase together with J-A.Y., and the above-mentioned experiments together with J-A.Y. Y.W. produced (~30%) homemade competent cells, transformed the Gibson assembly products (~50%) of the two libraries into competent cells, performed ~50% of bacterial glycerol stock of the two libraries into 96-well deep-well plates, and miniprepped ~50% of plasmids of the two libraries. L.Y. produced (~30%) homemade competent cells, transformed the Gibson assembly products (~50%)

of the two libraries into competent cells, performed ~50% of bacterial glycerol stock of the two libraries into 96-well deep-well plates, and miniprepped ~50% of plasmids of the two libraries. D.L.V. performed barcoded PCR of the entire two libraries and pooled down each plate of the PCR products into a single tube and purified the PCR products for SMRT long-read sequencing. E.D.C produced Taq DNA ligase used for Gibson assembly reaction. K.G. performed real-time quantitative PCR of cDNAs from iNeurons, assisted sometimes A.S. for three fragments PCR and Gibson assembly, and the above-mentioned experiments. T.L. prepared samples for 4sgRNA knockout assay with SMRT long-read sequencing of the 9 genes with assistance of K.M.. J.G. performed 1sgRNA vs 4sgRNA knockout efficiency and CRISPRoff assay with flow cytometry analyses. S.B. assisted the method development for comparing activity of 1sgRNA and 4sgRNA for activation. M.L. assisted the flow cytometry analyses of 4sgRNA virus titration and delivery rate with transfection and transduction to non-transfectable cells. S.H. supported the production of Taq DNA ligase and the production of FRET antibodies for $PrP^{C}$ detection. M.K. helped the design of sgRNAs for the 1st trial 384-well plate cloning of 4sgRNA plasmids. L.P. and D.H. supervised research. P.H. supervised research, appropriated the funding, supervised the planning and the execution of the experiments, offered continuous feedback and mentoring at DZNE. A.A. conceived the primary idea of generating arrayed libraries, appropriated the funding, supervised the planning and the execution of the experiments, offered continuous feedback and mentoring, coordinated the activities of the research team, and wrote the paper with input from all authors.

## Acknowledgements

A.A. is the recipient of grants from the Nomis Foundation, the Swiss National Research Foundation, the Swiss Personalized Health Network (SPHN, 2017DRI17), an ERC (European Research Council) Advanced grant, and a donation from the estate of Dr. Hans Salvisberg. J-A.Y. is the recipient of the postdoc grant Forschungskredit from University of Zurich and the Career Development Awards grant of the Synapsis Foundation – Alzheimer Research Switzerland ARS. The work of A.D. J.T., S.R and P.H. was supported by funds from the DZNE. We thank Drs. Patrick Hsu (University of California, Berkeley), John Doench (the Broad Institute of MIT and Harvard) and Jacob Corn (ETH Zurich) for their help and suggestions, Dr. Allan Bradley for sharing the Lenti-PB plasmid, Dr. Luke A. Gilbert and Mr. Greg C. Pommier (University of California, San Francisco) and Dr. Jonathan S. Weissman (Massachusetts Institute of Technology) for providing reagents and advice for experiments on targeted epigenetic silencing (CRISPRoff). We thank Kevin Maggi, Rafaela Ribeiro and Andres Gonzalez Guerra for technical assistance, Drs. Anna Bratus-Neuenschwander and Weihong Qi at the Functional Genomics Center Zurich (FGCZ) for support with SMRT sequencing and the Cytometry Facility of University of Zurich for technical assistance, as well as Dr. Merve Avar and Dr. Daniel

Heinzer for supporting C.T. with the $PrP^C$ FRET assays. Figure 1E, Figure 6A and Figure S4A illustrating the high-throughput APPEAL cloning method, TR-FRET assay and the strategy of barcoded SMRT sequencing of plasmids of our libraries were created with BioRender.

## References


Acevedo-Arozena, A., Wells, S., Potter, P., Kelly, M., Cox, R.D., and Brown, S.D. (2008). ENU mutagenesis, a way forward to understand gene function. Annu Rev Genomics Hum Genet *9*, 49-69.

Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nunez, J.K., Chen, Y., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y.*, et al.* (2016). A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell *167*, 1867-1882 e1821.

Amabile, A., Migliara, A., Capasso, P., Biffi, M., Cittaro, D., Naldini, L., and Lombardo, A. (2016). Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. Cell *167*, 219-232 e214.

Araque, A., and Navarrete, M. (2010). Glial cells in neuronal network function. Philos Trans R Soc Lond B Biol Sci *365*, 2375-2381.

Bausch-Fluck, D., Hofmann, A., Bock, T., Frei, A.P., Cerciello, F., Jacobs, A., Moest, H., Omasits, U., Gundry, R.L., Yoon, C.*, et al.* (2015). A mass spectrometric-derived cell surface protein atlas. PLoS One *10*, e0121314.

Boutros, M., and Ahringer, J. (2008). The art and design of genetic screens: RNA interference. Nat Rev Genet *9*, 554-566.

Bueler, H., Aguzzi, A., Sailer, A., Greiner, R.A., Autenried, P., Aguet, M., and Weissmann, C. (1993). Mice devoid of PrP are resistant to scrapie. Cell *73*, 1339-1347.

Chakrabarti, A.M., Henser-Brownhill, T., Monserrat, J., Poetsch, A.R., Luscombe, N.M., and Scaffidi, P. (2019). Target-Specific Precision of CRISPR-Mediated Genome Editing. Mol Cell *73*, 699-713 e696.

Chavez, A., Tuttle, M., Pruitt, B.W., Ewen-Campen, B., Chari, R., Ter-Ovanesyan, D., Haque, S.J., Cecchi, R.J., Kowal, E.J.K., Buchthal, J.*, et al.* (2016). Comparison of Cas9 activators in multiple species. Nat Methods *13*, 563-567.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A.*, et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Consortium, F., the, R.P., Clst, Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T.*, et al.* (2014). A promoter-level mammalian expression atlas. Nature *507*, 462-470.

de Groot, R., Luthi, J., Lindsay, H., Holtackers, R., and Pelkmans, L. (2018). Large-scale image-based profiling of single-cell phenotypes in arrayed CRISPR-Cas9 gene perturbation screens. Mol Syst Biol *14*, e8064.

Erard, N., Knott, S.R.V., and Hannon, G.J. (2017a). A CRISPR Resource for Individual, Combinatorial, or Multiplexed Gene Knockout. Mol Cell *67*, 1080.

Erard, N., Knott, S.R.V., and Hannon, G.J. (2017b). A CRISPR Resource for Individual, Combinatorial, or Multiplexed Gene Knockout. Mol Cell *67*, 348-354 e344.

Evers, B., Jastrzebski, K., Heijmans, J.P., Grernrum, W., Beijersbergen, R.L., and Bernards, R. (2016). CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. Nat Biotechnol *34*, 631-633.

Feldman, D., Singh, A., Schmid-Burgk, J.L., Carlson, R.J., Mezger, A., Garrity, A.J., Zhang, F., and Blainey, P.C. (2019). Optical Pooled Screens in Human Cells. Cell *179*, 787-799 e717.

Gautron, A., Bachelot, L., Aubry, M., Leclerc, D., Quemener, A.M., Corre, S., Rambow, F., Paris, A., Tardif, N., Leclair, H.M.*, et al.* (2021). CRISPR screens identify tumor-promoting genes conferring melanoma cell plasticity and resistance. EMBO Mol Med *13*, e13466.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009).

Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods *6*, 343-345.
Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C.*, et al.* (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647-661.
Hanna, R.E., and Doench, J.G. (2020). Design and analysis of CRISPR-Cas experiments. Nat Biotechnol *38*, 813-823.
Hart, T., Tong, A.H.Y., Chan, K., Van Leeuwen, J., Seetharaman, A., Aregger, M., Chandrashekhar, M., Hustedt, N., Seth, S., Noonan, A.*, et al.* (2017). Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. G3 (Bethesda) *7*, 2719-2727.
Heinzer, D., Avar, M., Pease, D.P., Dhingra, A., Yin, J.A., Schaper, E., Dogancay, B., Emmenegger, M., Spinelli, A., Maggi, K.*, et al.* (2021). Novel regulators of PrPC biosynthesis revealed by genome-wide RNA interference. PLoS Pathog *17*, e1010013.
HorizonDiscovery.
Horlbeck, M.A., Gilbert, L.A., Villalta, J.E., Adamson, B., Pak, R.A., Chen, Y., Fields, A.P., Park, C.Y., Corn, J.E., Kampmann, M.*, et al.* (2016). Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. Elife *5*.
Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262-1278.
Ihry, R.J., Worringer, K.A., Salick, M.R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C.*, et al.* (2018). p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. Nat Med *24*, 939-946.
Imkeller, K., Ambrosi, G., Boutros, M., and Huber, W. (2020). gscreend: modelling asymmetric count ratios in CRISPR screens to decrease experiment size and improve phenotype detection. Genome Biol *21*, 53.
Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.
Kabadi, A.M., Ousterout, D.G., Hilton, I.B., and Gersbach, C.A. (2014). Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. Nucleic Acids Res *42*, e147.
Kampmann, M. (2020). CRISPR-based functional genomics for neurological disease. Nat Rev Neurol.
Kanfer, G., Sarraf, S.A., Maman, Y., Baldwin, H., Dominguez-Martin, E., Johnson, K.R., Ward, M.E., Kampmann, M., Lippincott-Schwartz, J., and Youle, R.J. (2021). Image-based pooled whole-genome CRISPRi screening for subcellular phenotypes. J Cell Biol *220*.
Knott, G.J., and Doudna, J.A. (2018). CRISPR-Cas guides the future of genetic engineering. Science *361*, 866-869.
Konermann, S., Brigham, M.D., Trevino, A., Hsu, P.D., Heidenreich, M., Cong, L., Platt, R.J., Scott, D.A., Church, G.M., and Zhang, F. (2013). Optical control of mammalian endogenous transcription and epigenetic states. Nature *500*, 472-476.
Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H.*, et al.* (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583-588.
Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650-665.
Lessard, S., Francioli, L., Alfoldi, J., Tardif, J.C., Ellinor, P.T., MacArthur, D.G., Lettre, G., Orkin, S.H., and Canver, M.C. (2017). Human genetic variation alters CRISPR-Cas9 on- and off-targeting specificity at therapeutically implicated loci. Proc Natl Acad Sci U S A *114*, E11257-E11266.
Lohman, G.J., Bauer, R.J., Nichols, N.M., Mazzola, L., Bybee, J., Rivizzigno, D., Cantin, E., and Evans, T.C., Jr. (2016). A high-throughput assay for the comprehensive profiling of DNA ligase fidelity. Nucleic Acids Res *44*, e14.
Mali, P., Aach, J., Lee, J.H., Levner, D., Nip, L., and Church, G.M. (2013). Barcoding cells using cell-surface programmable DNA-binding domains. Nat Methods *10*, 403-406.
McCarty, N.S., Graham, A.E., Studena, L., and Ledesma-Amaro, R. (2020). Multiplexed CRISPR

technologies for gene editing and transcriptional regulation. Nat Commun *11*, 1281.
Metzakopian, E., Strong, A., Iyer, V., Hodgkins, A., Tzelepis, K., Antunes, L., Friedrich, M.J., Kang, Q., Davidson, T., Lamberth, J.*, et al.* (2017). Enhancing the genome editing toolbox: genome wide CRISPR arrayed libraries. Sci Rep *7*, 2244.
Morgens, D.W., Deans, R.M., Li, A., and Bassik, M.C. (2016). Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. Nat Biotechnol *34*, 634-636.
Morizane, R., and Bonventre, J.V. (2017). Generation of nephron progenitor cells and kidney organoids from human pluripotent stem cells. Nat Protoc *12*, 195-207.
Nunez, J.K., Chen, J., Pommier, G.C., Cogan, J.Z., Replogle, J.M., Adriaens, C., Ramadoss, G.N., Shi, Q., Hung, K.L., Samelson, A.J.*, et al.* (2021). Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. Cell *184*, 2503-2519 e2517.
Pease, D., Scheckel, C., Schaper, E., Eckhardt, V., Emmenegger, M., Xenarios, I., and Aguzzi, A. (2019). Genome-wide identification of microRNAs regulating the human prion protein. Brain Pathol *29*, 232-244.
Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W.*, et al.* (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods *10*, 973-976.
Popper, K.R. (1992). The logic of scientific discovery (London ; New York: Routledge).
Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat Methods *11*, 783-784.
Sanson, K.R., Hanna, R.E., Hegde, M., Donovan, K.F., Strand, C., Sullender, M.E., Vaimberg, E.W., Goodale, A., Root, D.E., Piccioni, F.*, et al.* (2018). Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. Nat Commun *9*, 5416.
Scheckel, C., and Aguzzi, A. (2018). Prions, prionoids and protein misfolding disorders. Nat Rev Genet *19*, 405-418.
Schmidt, T., Schmid-Burgk, J.L., and Hornung, V. (2015). Synthesis of an arrayed sgRNA library targeting the human genome. Sci Rep *5*, 14987.
Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G.*, et al.* (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84-87.
Smith, I., Greenside, P.G., Natoli, T., Lahr, D.L., Wadden, D., Tirosh, I., Narayan, R., Root, D.E., Golub, T.R., Subramanian, A.*, et al.* (2017). Evaluation of RNAi and CRISPR technologies by large-scale gene expression profiling in the Connectivity Map. PLoS Biol *15*, e2003213.
Synthego.
ThermoFisher.
Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P.*, et al.* (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol *33*, 187-197.
Tycko, J., Wainberg, M., Marinov, G.K., Ursu, O., Hess, G.T., Ego, B.K., Aradhana, Li, A., Truong, A., Trevino, A.E.*, et al.* (2019). Mitigation of off-target toxicity in CRISPR-Cas9 screens for essential non-coding elements. Nat Commun *10*, 4063.
Uhlen, M., Karlsson, M.J., Hober, A., Svensson, A.S., Scheffel, J., Kotol, D., Zhong, W., Tebani, A., Strandberg, L., Edfors, F.*, et al.* (2019). The human secretome. Sci Signal *12*.
Ungricht, R., Guibbal, L., Lasbennes, M.C., Orsini, V., Beibel, M., Waldt, A., Cuttat, R., Carbone, W., Basler, A., Roma, G.*, et al.* (2022). Genome-wide screening in human kidney organoids identifies developmental and disease-related aspects of nephrogenesis. Cell Stem Cell *29*, 160-175 e167.
Wang, H., La Russa, M., and Qi, L.S. (2016). CRISPR/Cas9 in Genome Editing and Beyond. Annu Rev Biochem *85*, 227-264.
Yan, X., Stuurman, N., Ribeiro, S.A., Tanenbaum, M.E., Horlbeck, M.A., Liem, C.R., Jost, M., Weissman, J.S., and Vale, R.D. (2021). High-content imaging-based pooled CRISPR screens in mammalian cells. J Cell Biol *220*.
Zhang, X.D. (2011). Illustration of SSMD, z score, SSMD*, z* score, and t statistic for hit selection in RNAi high-throughput screens. J Biomol Screen *16*, 775-785.

## Materials and Methods

### DNA constructs

The DNA constructs used in the study, except for the 4sgRNA expression plasmids (whose construction is described separately in the following), include hCas9 (Addgene #41815) and lentiCas9-Blast (Addgene #52962), SP-dCas9-VPR (Addgene #63798) and pXPR_120 (lenti-dCas9-VPR-Blast, addgene #96917), psPAX2 (Addgene #12260), VSV-G (Addgene #8454), and pYJA5.

The pYJA5 construct was created by modifying the lenti-PB vector (Metzakopian et al., 2017) (a gift from Dr. Allan Bradley) in two steps. First, the DNA fragment flanked by the recognition sites for the restriction enzymes *Mlu*I and *Age*I in the lenti-PB vector was replaced by a synthesized DNA fragment that included the human U6 promoter and the fourth variants of tracrRNA, as well as an ampicillin resistance gene (β-lactamase expression cassette). Two *Bbs*I (type II restriction enzyme) recognition sites flanking the β-lactamase expression cassette were introduced into the new fragment, in order to facilitate the removal of the β-lactamase expression cassette. In a second step, the original ampicillin resistance (β-lactamase) expression cassette in the lenti-PB vector was removed between the two *Bsp*HI restriction enzyme recognition sites. After its removal, the insertion of 4sgRNA expression cassettes containing a trimethoprim resistance gene (dihydrofolate reductase) achieves antibiotic-switch-based cloning. Furthermore, all *BsmB*I recognition sites were mutated. Detailed sequences of the pYJA5 and 4sgRNA-pYJA5 constructs are included in the supplementary information.

Single sgRNAs were cloned into the pYJA4 vector individually via the previously established method (Koike-Yusa et al., 2014).

### In silico 4sgRNA libraries design

#### *Pooling existing libraries*

To provide a starting point for guide RNA selection, we collected sgRNAs from previously published and validated libraries and tools, which each employed their own algorithms to select sgRNAs with high predicted on-target efficacy. We included the Calabrese (Sanson et al., 2018) and hCRISPRa v2 (Horlbeck et al., 2016) libraries for CRISPRa, and the TKOv3 (Hart et al., 2017) and Brunello (Doench et al., 2016; Sanson et al., 2018) libraries for CRISPRko. We complemented these source libraries with sgRNAs from the CRISPick tool (formerly GPP sgRNA Designer) (Doench et al., 2016; Hanna and Doench, 2020; Sanson et al., 2018), to ensure optimal coverage of difficult-to-target and newly annotated genes (the website was accessed in April 2020, following the update of the 20th March 2020).

#### *Gene definitions*

Entrez gene identifiers were used to provide common gene definitions for sgRNAs from all sources. If the source library did not provide Entrez identifiers, the official gene symbols were mapped to Entrez IDs, and the genomic location was used to disambiguate gene symbols, when necessary. Genes that were not defined as protein-coding by NCBI or Ensembl were excluded (according to the annotation files ftp://ftp.ncbi.nih.gov/gene/DATA/GENE_INFO/Mammalia/Homo_sapiens.gene_info.gz and ftp://ftp.ensembl.org/pub/release-99/tsv/homo_sapiens/Homo_sapiens.GRCh38.99.entrez.tsv.gz, both downloaded on 25 March 2020). The final libraries included 19839 protein-coding genes for CRISPRa, and 19819 for CRISPRko; the difference in gene counts arises from a small number of genes that are present for only one modality in our source libraries (for example, highly polymorphic genes related to adaptive immunity, such as the T Cell Receptor Alpha Locus (TRA) gene, are available for CRISPRa, but not CRISPRko).

#### *TSS definitions*

To ensure good coverage of alternative transcripts, and broad applicability of the CRISPRa library in multiple cell lines, we adopted the alternative transcription start site (TSS) definitions from the hCRISPRa-v2 library (Horlbeck et al., 2016). The authors of this library used the FANTOM5 CAGE-Seq dataset (FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al., 2014), supplemented by Ensembl (Yates et al., 2020) transcript models, to define TSS positions; additional TSSs were targeted by their own set of sgRNAs if the FANTOM5 scores indicated significant transcriptional activity, and if they were spaced more than one kilobase apart from the primary TSS. We chose a separate set of four sgRNAs for each TSS, treating multiple TSSs as if they were separate genes. To group sgRNAs by TSS, we mapped sgRNAs from all sources (including the top five sgRNAs from the CRISPick sgRNA Designer) to their genomic locations, and iterated through each sgRNA, starting with the lowest genomic coordinate; a new TSS group was defined if the distance from one guide to the next exceeded 1000 base pairs. Additional TSSs were only targeted if a valid combination of four guides was available. Multiple TSSs were included for 2311 genes (using 4803 four-guide combinations), whereas a single TSS was targeted for the remaining 17528 genes.

### *Avoidance of genetic polymorphisms*

For each sgRNA, we checked for overlaps with regions of frequent genetic polymorphism in human populations, in either the 20-nucleotide protospacer sequence, or the two guanosine nucleotides of the protospacer adjacent motif (NGG). We avoided sgRNAs whose target region contained any genetic polymorphisms with frequencies greater than 0.1%. Variant frequencies were derived from the Kaviar database (Glusman et al., 2011), which includes curated genomic data on single nucleotide variants, indels, and complex variants from over 77000 individuals (including over 13000 whole genomes). The dataset (only variants seen more than 3 times, version 160204-hg38) was downloaded on 7 August 2019. The polymorphism frequencies in the Kaviar database were generally similar to those from TOPMED, gnomAD, and the 1000 Genomes Project.

### *Specificity scores*

In order to select a four-guide combination with minimal off-target effects, we computed specificity scores for each sgRNA from our source libraries. We used the approach introduced by the authors of the GuideScan (Perez et al., 2017) tool: For each guide, potential off-target sites were weighted by their CFD (cutting frequency determination) scores (Doench et al., 2016), and CFD scores were aggregated into a single score using the formula: 1 / (1 + sum of CFD scores from all off-target sites (Hsu et al., 2013). Because the pre-computed GuideScan Cas9 database does not contain all sgRNAs (it excludes those with perfect-match or one-mismatch off-target sites in the reference genome), we annotated sgRNAs using both the GuideScan and CRISPOR (Concordet and Haeussler, 2018; Haeussler et al., 2016) tools. Local installations of these tools were used, and the source code was downloaded in December 2020 (GuideScan version 2018-05-16, and CRISPOR version 4.97); the output of the local installations was confirmed to be identical to that of the web-based tools. When available, GuideScan specificity scores were used (considering up to three mismatches); otherwise, CRISPOR specificity scores were used (considering up to four mismatches). CRISPOR three-mismatch (3MM) and four-mismatch (4MM) specificity scores analogous to those from GuideScan were computed, using the detailed output files listing each off-target site. GuideScan and CRISPOR specificity scores were highly correlated, but not identical, due to slight differences in the number of off-target sites identified for the same sequence. When selecting sgRNAs, we avoided low-specificity guides with 3MM scores below 0.2; this cut-off point was recently shown to have good predictive power for identifying sgRNAs with significant off-target activity (Tycko et al., 2019). However, this criterion had to be relaxed in cases where all eligible sgRNAs had specificity scores below 0.2, for example, when targeting genes present in multiple copies in the genome, or those belonging to large gene families with many closely related paralogs and pseudogenes. Finally, in order to choose among all eligible four-guide combinations, we computed an aggregate specificity score, using the formula: 1 / (1 + sum of CFD scores from all four guides), and picked the combination with the highest score, indicating high predicted specificity.

### *Guide RNA spacing*

To allow for unhindered multiple binding for synergistic effect, we aimed to select four sgRNAs whose "cut" locations were spaced at least 50 base pairs apart. However, for CRISPRa, target sequences should be located within a window of about 400 base pairs upstream of the TSS for optimal activity (Gilbert et al., 2014), which is reflected in the selection of sgRNAs in the source libraries; thus, overlaps were unavoidable for some genes. For CRISPRko, on the other hand, overlaps were often inevitable when targeting genes with very short coding sequences. In those cases, we nevertheless aimed to minimize the total number of overlaps between neighbouring guides. Furthermore, all four-guide combinations strictly adhered to another criterion: No two sgRNAs were allowed to share identical sub-sequences of more than seven base pairs. This was done primarily to minimize recombination events between identical regions during Gibson assembly of the plasmid. However, this also enforced minimal spacing of the four selected guides.

### *Selection of four sgRNAs*

After integration and annotation of sgRNAs from the source libraries, we selected the final combination of four sgRNAs for each gene or TSS. First, sgRNAs containing a stretch of four or more T nucleotides were excluded, since this sequence can induce termination of transcription. Next, all possible four-guide combinations for each gene were generated, and combinations that shared identical subsequences greater than seven base pairs in length were excluded. The potential combinations were then ranked, using a list of criteria that were applied in order; if multiple combinations were tied in first place, the decision was made using the next criterion down the list. The criteria were as follows: 1) Maximize the number of sgRNAs (from zero to four) that fulfil certain minimal requirements – the sgRNA can be mapped to a defined genomic location in the reference genome with an N(GG) PAM; there are no overlaps with frequent genetic polymorphisms (>0.1%); the 3MM specificity score is at least 0.2; and for CRISPRko only, the guide conforms to the criteria of Graf et al. ( Graf et al., 2019); 2) maximize the number of sgRNAs with exactly one perfect match location in the reference genome, 3) minimize the number of overlaps between two neighbouring sgRNAs spaced fewer than 50 base pairs apart, 4) minimize the number of sgRNAs derived from the CRISPick sgRNA Designer tool, rather than the previously published libraries, 5) for CRISPRa, minimize the number of sgRNAs derived from the "supplemental 5" rather than "top 5" sgRNAs for the hCRISPRa-v2 library, and for CRISPRko, minimize the number of CRISPick -derived sgRNAs ranked outside the top 10, and 6) maximize the aggregate specificity score from all 4 guides. The highest-ranked four-guide combination was chosen. Since the aggregate specificity score was the only quantitative criterion, it acted as a tiebreaker, and had the greatest impact on the choice of guides.

### *Sublibrary allocation*

To facilitate focussed screens of a subset of the genome, we divided the entire set of protein-coding genes into mutually exclusive sub-libraries. Two of our sub-libraries – Transcription Factors, and Secretome – were based on recent publications that combined bioinformatics analyses with expert curation to arrive at a comprehensive list of genes in those categories (Lambert et al., 2018; Uhlén et al., 2019). These lists were obtained from the publication's supplemental data (for the secretome) or the authors' website (for the transcription factors; humantfs.ccbr.utoronto.ca, database version 1.01). Ensembl gene IDs were translated to Entrez gene IDs, making use of HUGO gene symbols to disambiguate one-to-many mappings for a few genes. A third sub-library was based on a list of G-protein coupled receptors, curated by the HUGO Gene Nomenclature Committee (HGNC) (Braschi et al., 2019) (https://www.genenames.org/cgi-bin/genegroup/download?id=139&type=branch, accessed on 11 Marc 2020). An additional seven thematic sub-libraries were adopted from the hCRISPRa-v2 library (Horlbeck et al., 2016): Membrane Proteins, Kinases/Phosphatases/Drug Targets, Mitochondria/Trafficking/Motility, Stress/Proteostasis, Cancer/Apoptosis, Gene Expression, and Unassigned. The first two of these thematic sub-libraries were updated to incorporate a small number of additional transmembrane receptors, transporters, kinases and phosphates, using Gene Ontology terms (exported from

BioMart (Smedley et al., 2009) on 25 March 2020) and a list of membrane proteins provided by the Human Protein Atlas project (Uhlén et al., 2015) (https://www.proteinatlas.org/search/protein_class:Predicted+membrane+proteins, accessed on 11 March 2020). If a gene belonged to multiple categories, it was assigned to the first sub-library (in the order in which they are listed in this section), and all remaining genes were added to the Unassigned sub-library.

### *Classification of unintended gene perturbations*

Some sgRNAs are expected to perturb additional genes, in addition to the intended target gene. In certain cases, this occurs at a different locus than the intended target site: For example, gene families of very close paralogs can often only be targeted with sgRNAs that have multiple perfect-match binding sites in the genome. However, in most cases, this involves a single locus – the intended binding site – where a sgRNA may perturb more than one gene. In the case of CRISPRa, the same promoter region is often shared by two genes located on opposite strands of the chromosome, so that their transcription start sites (TSSs) lie only a few hundred base pairs apart. In this case, guide RNAs that effectively activate one gene would inevitably also activate the other. As a guide for users of the library, and to aid the interpretation of hit genes, we annotated sgRNAs with a complete list of all genes they target. For the purpose of summarizing this phenomenon across the entire library, we classified each sgRNA as 1) only targeting the intended gene, 2) targeting unintended genes, but in a single location, or 3) targeting unintended genes at other locations. For this analysis, if two perfect-match sgRNA binding sites had any target genes in common, they were considered to target unintended genes at the same location (which is especially relevant for sgRNAs targeting the pseudo-autosomal region of chromosomes X and Y).

### *Annotation of unintended target genes*

To annotate each sgRNA with all its potential target genes, a database of TSS locations was constructed by merging the FANTOM5 dataset (lifted over to the hg38 genome (Abugessaisa et al., 2017), version 3) with data from BioMart (Smedley et al., 2009) (exported on 25 March 2020), using Entrez gene IDs as a common identifier. Similarly, data on coding sequence (CDS) and exon locations were compiled from BioMart, the “TxDb.Hsapiens.UCSC.hg38.knownGene” Bioconductor package (version 3.10.0), and GENCODE (Frankish et al., 2019) annotation data (Release 33), and location data were merged using Entrez gene identifiers (if available) or Ensembl gene identifiers. Genes annotated as pseudogenes, or whose categorization was unclear, were excluded from further analysis. For CRISPRa, perfect-match sgRNA binding sites within a window of 1000 base pairs around TSSs were considered. For CRISPRko, sgRNA cut locations had to lie within the coding sequences (CDSs) of protein-coding genes, or within the exons of non-coding RNAs.

### *Annotation of predicted deletions*

In the case of CRISPRko, when four sgRNAs are active within the same cell, the multiple, closely spaced double-strand breaks commonly lead to the loss of a DNA segment between the sgRNA cut locations. Thus, in addition to annotating individual sgRNAs, we also determined which genes are affected by the predicted deletion – the segment between the first and last cut site. We also took deletions induced by (perfect-match) off-target binding sites into consideration. Because deletions may be less likely to occur if the cut sites are very far apart, we imposed a maximum distance of one megabase between cut sites, so that multiple predicted deletions (or isolated cut positions) on the same chromosome were possible.

## In silico comparison of CRISPR libraries

To compare in silico characteristics of existing libraries and the 4sg library, the top four guides per gene were selected. Whereas the Brunello (Doench et al., 2016; Sanson et al., 2018) and TKOv3 (Hart et al., 2017) libraries were designed to contain four sgRNAs per gene, the Calabrese library (Sanson et al., 2018) was divided by the authors into Set A and Set B, each containing three sgRNAs per gene. To define the top four sgRNAs, the sgRNAs from Set A were supplemented with a randomly selected sgRNA from Set B. For the hCRISPRa v2 (Horlbeck et al., 2016) and CRISPick

libraries, the four highest-ranked sgRNAs were chosen (using the “Pick Order” column in the output from the CRISPick sgRNA designer tool). Since the libraries differed in the genes they covered, and since different genes vary in the availability of potential sgRNAs with high predicted activity and specificity, only genes present in all libraries were used for benchmarking. Furthermore, for genes for which the 4sg and hCRISPRa v2 libraries included more than one transcription start site (TSS), only the sgRNAs targeting the main TSS was included, defined as the TSS with the highest score in the FANTOM5 dataset, or – if data were unavailable for that gene – the most upstream TSS. To compare the expected number of sgRNA binding sites affected by genetic polymorphisms, the frequencies of the most common polymorphisms overlapping each sgRNA were summed up. This is a conservative estimate, since SNPs with frequencies below 0.1% were excluded. Furthermore, in the case of multiple single nucleotide polymorphisms (SNPs) overlapping with a sgRNA, only the most frequent was considered. Because linkage disequilibrium between SNPs affecting the same sgRNA is highly likely, a precise estimation of the total probability of overlaps with polymorphisms would require access to the individual sequencing data underlying the SNP databases.

## Software and code

The annotation and selection of sgRNAs for the library design was performed using the R statistical programming environment (R Core Team, 2020), version 3.6.3, and the Bioconductor suite (Huber et al., 2015), version 3.10. Source code is available at https://github.com/Lukas-1/CRISPR_4sgRNA.

## SMRT long-read next-generation sequencing of libraries

### Barcoding, amplification, and long-read sequencing

To assess the frequency of mutations, recombinations and deletions within the polyclonal population of 4sgRNA plasmids, single-molecule long-read sequencing was performed. Plasmids were amplified using polymerase chain reaction (PCR) with barcoded primers that uniquely identified each targeted gene. In our pilot sequencing run, this was achieved using a combination of 16 different forward primers and 24 different reverse primers (distinguishing the rows and columns of the 384-well plate, respectively). The amplified region was 2225 base pairs in length, encompassing the entire 4sg expression cassette (containing all four promoter, guide RNA and tracrRNA sequences, as well as the trimethoprim resistance element), and was flanked by two 10-bp paired barcode sequences. Amplicons from all wells were pooled and single-molecule real-time (SMRT) sequencing data were generated using the PacBio Sequel instrument.

### Processing of long-read sequencing data

Circular Consensus Reads (CCS) consensus calling was done using the PacBio SmrtLink software using default parameters, and only consensus reads with at least 5 full-pass subreads and an estimated read accuracy $\geqslant$ 99.9% were retained. Barcode demultiplexing was also done using the SmrtLink software, and to minimize incorrect well assignments, reads with a barcode score lower than 60 or a score lead lower than 30 were omitted. Reads were further filtered using a custom script to ensure the following criteria were met for each barcode: Either the barcode sequence must be present in full and entirely correct, or the sequence must be at least 8 bp in length and flanked by an entirely correct 20-bp flanking constant region (representing the constant region of the primers used for PCR amplification). These additional steps proved necessary to ensure that only complete reads were retained (containing the forward and reverse primer sequences), and to exclude truncated reads whose terminal sequences were incorrectly interpreted as truncated barcodes. Finally, consensus reads with an average per-basePhred Quality Score below 85 were excluded (with the highest achievable mean Phred Quality score being 93). In the pilot sequencing run, 78351 consensus reads remained after filtering, with an average of 204 reads per well (ranging from 63 to 1098).

### Analysis of consensus reads

To quantify the percentage of correct guide RNA sequences, and to identify contaminations from other wells, each read was searched for the sgRNA + tracrRNA sequences in the forward and reverse directions, and all perfect matches were counted. To further characterize incorrect

sequences, each consensus read was aligned to the corresponding barcoded reference sequence for that well, with the "pairwiseAlignment" function of the Biostrings R/Bioconductor package, version 2.54.0, using default parameters. The region corresponding to the sgRNA + tracrRNA sequence of the reference was then extracted from the aligned read, and each sequence was classified as a) entirely correct, b) a contamination (if it is a perfect match for a sgRNA sequence from another well), c) a large deletion (if >50% of the aligned sequence was composed of gaps), or d) some other mutation.

## APPEAL high-throughput generation of libraries

### Oligo synthesis

Twenty-nucleotide sgRNA sequences were incorporated into oligonucleotide sequences with appended constant sequences and synthesized in 384-well plates using the high affinity purification (HAP) purification method by Sangon Biotech (China). The sgRNA1 (sgRNA1 sequence, $N_{20sg1}$) oligonucleotide sequence is: 5'- ttgtggaaaggacgaaacaccG$N_{20sg1}$GTTTAAGAGCTAAGCTG-3'; sgRNA2 (sgRNA2 sequence, $N_{20sg2}$) oligo sequence is: 5'-cttggagaaaagccttgtttG$N_{20sg2}$GTTTGAGAGCTAAGCAGA-3'; sgRNA3 (sgRNA3 sequence, $N_{20sg3}$) oligo sequence is: 5'-gtatgagaccactctttcccG$N_{20sg3}$GTTTCAGAGCTAAGCACA-3'; and sgRNA4 (reverse complement sequence of sgRNA4, $N_{20\ crsg4}$) oligo sequence is 5'-ATTTCTGCTGTAGCTCTGAAACN$_{20crsg4}$Cgaggtacccaagcggc-3'. The oligonucleotides were diluted with ultrapure water to a working concentration of 4 µM.

### Three-fragment polymerase chain reactions (*PCRs*)

A total of 10 µL PCR reaction per well was performed in 384-well plates.

The C1 fragment (amplicon size 761 bp) PCR mix was prepared as follows:

| volume/well | Reagent |
|---|---|
| 0.2 µL | C1 fragment 1 ng/µL |
| 0.2 µL | mU6 Rev primer 10 µM<br>(common primer, sequence attached) |
| 2 µL | 5X HF buffer |
| 0.2 µL | dNTPs 10 mM |
| 0.1 µL | Phusion High-Fidelity DNA polymerase |
| 6.8 µL | $ddH_2O$ |

9.5 µL of the mix were aliquoted in each well of the 384-well plate, and 0.5 µl of sgRNA1 primer (at 4 µM concentration) was added to each well and mixed.

The M fragment (amplicon size 360 bp) PCR mix was prepared as follows:

| volume/well | Reagent |
|---|---|
| 0.2 µL | M fragment 1ng/µL |
| 0.2 µL | M Rev primer 10 µM<br>(common primer, sequence attached) |
| 2 µL | 5X HF buffer |
| 0.2 µL | dNTPs 10 mM |
| 0.1 µL | Phusion High-Fidelity DNA polymerase |
| 6.8 µL | $ddH_2O$ |

9.5 µL of the mix were aliquoted in each well of the 384-well plate, and then 0.5 µl of sgRNA2 primer (at 4 µM concentration) was added to each well and mixed.

C2s fragment (amplicon size 422 bp) PCR mix was prepared as follows:

| vol/well | Reagent |
|---|---|
| 0.2 µL | C2s fragment 1ng/µL |
| 2 µL | 5x HF buffer |
| 0.2 µL | dNTPs 10 mM |
| 0.1 µL | Phusion High-Fidelity DNA polymerase |
| 6.5 µL | $ddH_2O$ |

9 µL of the mix were aliquoted in each well of the 384-well plate, and then 0.5 µL of sgRNA3 primer (at 4 µM concentration) and 0.5 µL of sgRNA4 primer (also at 4 µM concentration) were added to each well and mixed.

The Integra ViaFlo 384-well pipetting system was used for all 384-well liquid handling. All the PCR plates were sealed tightly and centrifuged at 2000 rpm for 2 minutes, and placed in thermocyclers with the following program: Preheat the lid at 99 °C; Initial denaturation at 98 °C for 30 seconds, 36 cycles comprising 98 °C for 10 seconds, 60 °C for 30 seconds, and 72 °C for 25 seconds, and final extension at 72 °C for 5 minutes, followed by cooldown to 20 °C. All PCR products were then diluted with 9 µL of ultrapure water for later Gibson assembly. The success of PCR on each plate was confirmed by DNA agarose gel electrophoresis of several random samples on the plate.

**Gibson assembly**

Assembly of the three fragment PCR products into the pYJA5 vector was performed in a 384-well plate by Gibson assembly, with the following reaction mix:

| Volume | Reagent |
|---|---|
| 2 µL | C1 amplified fragment (estimated around 16 ng/µL) |
| 1 µL | M amplified fragment (estimated around 16 ng/µL) |
| 1 µL | C2s amplified fragment (estimated around 20 ng/µL) |
| 1 µL | pYJA5 *Bbs*I digested purified vector, diluted to 120 ng/µL |
| 5 µL | 2x homemade HiFi Gibson master mix |

The mix was incubated in the thermocycler at 50 °C for 1 hour, and then used for transformation of competent cells or stored immediately at -20 °C.

**Transformation and bacterial storage**

Transformation was carried out in 96-well deep-well plates (2.3 mL, Axygene P-DW-20-C) in the cold room. 5 µL (per well) of Gibson mix from the 384-well plate was transferred into four 96-well plates and spun down to the bottom of each well. 50 µL (per well) of homemade competent cells (NEB stable competent cells) were dispensed and mixed twice with the Gibson mix. The plates were then kept immersed in ice for 30 minutes. Heat shock was performed for 30 seconds at 42 °C by placing the plate into a water bath. Plates were placed back on ice for 5 minutes. 300 µL of homemade SOC medium (0.5% Yeast Extract, 2% Tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM Glucose) were then added into the plate and incubated for 1 hour at 37 °C under shaking at 900 rpm using a thermo-shaker. Then, 900 µL (per well) of Terrific Broth (TB) medium (https://openwetware.org/wiki/Terrific_Broth) containing 15 µg/mL trimethoprim and 15 µg/mL tetracycline was added to the transformation mix, and incubated at 30 °C under shaking at 900 rpm

for 40-48 hours.

Bacteria were then stored at a final concentration of 16.7% (v/v) glycerol in both 96-well plates (300 µL final storage volume) and 384-well plates (150 µL final storage volume) at -80 ° C.

**Magnetic-beads-based 96-well plasmids miniprep**

50 µl of the Gibson assembly product transformed bacteria were transferred into 1.2 mL of TB medium (with 15 µg/mL trimethoprim and 15 µg/mL tetracycline in 96-well deep well plate) immediately before the storage of the bacteria, and grown at 30 °C at 900 rpm for 40-48 hours. The bacteria were then subjected to in-house magnetic-beads-based plasmids miniprep procedures, which were adopted from the canonical plasmids miniprep protocols (Birnboim and Doly, 1979). Briefly, the bacteria were pelleted at 4000 rpm for 10 min and resuspended in 200 µl of P1 buffer [50 mM glucose, 10 mM EDTA, 25 mM Tris (pH 8.0)], and subsequently lysed in 200 µl of P2 buffer [0.2 M NaOH, 1% SDS (w/v) ], and the lysis mixture was neutralized in 200 µl of P3 buffer (3 M KOAc, pH 6.0) and subjected to centrifugation at 4000 rpm for 10 min at 4 °C. Then 400 µl of the supernatant were transferred into a new deep-well plate and 1000 µl of cold absolute ethanol were added and mixed, then centrifuged at 4000 rpm for 10 min at 4 °C. The supernatant was discarded and 50 µl of $ddH_2O$ was added to the plasmid pellet and mixed to dissolve the plasmids. Then 75 µl of beads buffer [2.5 M NaCl, 10 mM Tris base, 1mM EDTA, 3.36 mM HCl, 20% (w/v) PEG8000, 0.05% (w/v) Tween 20] and 50 µl of SpeedBeads™ magnetic carboxylate modified particles (GE Healthcare 65152105050250, 1:50 dilution in beads buffer) were added to the plasmids, mixed and incubated for 5 min on a magnetic rack to separate the beads from the supernatant. The beads were then washed twice with 70% ethanol and dried in a water bath (65 °C). Plasmid DNA was then eluted by 150 µl of sterile tris-EDTA buffer [1 mM EDTA, 10 mM Tris-HCl (pH 8.0)] from the beads at 65 °C for 10 min and transferred to a new low-profile 96-well plate. To ensure the full cloning procedure was correct, two wells of plasmids from each 96-well plate were subjected to Sanger sequencing.

## Cell culture, transfection, transduction, and flow cytometry

All cells were cultured at 37 °C with appropriate growth medium with 5% $CO_2$. Transfection was performed via Lipofectamine 3000 (Thermo Fisher Scientific) at a cell density of 80-90% and with 0.25 µg of sgRNA plasmids and 0.25 µg of Cas9 or dCas9-VPR plasmids in 24-well plates. For lentiviral transduction, a multiplicity of infection of ~1-2 was used and 3 days post infection, the cells were subjected to flow cytometry or RNA extraction for real-time quantitative PCR. For validation of gene knockout/silencing efficiency, cells were cultured for 3 days under puromycin selection and then around one-week without selection before subjected to live-cell staining and flow cytometry. Flow cytometry analysis was performed by the BD Canto II or LSRFortessa™ Cell Analyzer at the core facility center of the University of Zurich.

## Real-time quantitative PCR

Total RNA of HEK293 cells or iNeurons were isolated by the TRIzol Reagent (Thermo Fisher Scientific) according to the manual. 600 ng of RNA were reversed transcribed into cDNA via QuantiTect Reverse Transcription Kit (Qiagen). Real-time quantitative PCR was done with SYBR green (Roche) according to the manual with the primer sets for each gene as follows. GAPDH, ACTB and HMBS were used as internal control.

| Gene | Forward primer sequence 5'-3' | Reverse primer sequence 5'-3' |
|---|---|---|
| ACTB | CATGTACGTTGCTATCCAGGC | CTCCTTAATGTCACGCACGAT |
| CXCR4 | ACTACACCGAGGAAATGGGCT | CCCACAATGCCAGTTAAGAAGA |
| NEUROD1 | GGATGACGATCAAAAGCCCAA | GCGTCTTAGAATAGCAAGGCA |
| LINC00925 | AATTGTCCTGTGAAGTGAAG | TTCCTCTGTCTCCATTGTCA |
| TINCR | GGTCTGGGCTCCCAGGTGGA | TGTCAGGGACTGGGGCTCC |

| | | |
|---|---|---|
| POU5F1 | TATTTGGGAAGGTATTCAGC | CTTACACATGTTCTTGAAGC |
| KLF4 | CTGGCGGGAGGAGCTCTCC | CGGCTCCGCCGCTCTCCA |
| LIN28A | GGGATGGATATATGAAGTAAGG | TAGCTACCATGACACTATTAAT |
| IL1R2 | GCTGCCAGAAGCTGCCG | CTCAGGGCTACAGGCTCCC |
| EGFR | GTTTGCCAAGGCACGAGTAA | GAGAAAATGATCTTCAAAAGTGCCC |
| HBG1 | AATGTGGAAGATGCTGGAGG | GCCAAAGCTGTCAAAGAACC |
| MYOD1 | GCTCCAACTGCTCCGAC | TGCTGGACAGGCAGTCTA |
| ZFP42 | CAGGTGTTTGCTGAAGACAG | GTTGCTGGCTCATGTTTTCC |
| LINC00028 | CACTCCCACACCCCAAC | TGCCAACTTTCCACAGCTAG |
| LINC00514 | AGAAGTGGTTTGGGCGC | CCGTTTCCATTGTTGTATCCTG |
| ASCL1 | CGCGGCCAACAAGAAGATG | CGACGAGTAGGATGAGACCG |
| GAPDH | ATGATCTTGAGGCTGTTG | CTCAGACACCATGGGGAA |
| AGER | TGGATGAAGGATGGTGTGCC | CACAGCTGTAGGTTCCCTGG |
| APOE | CTGCTCAGCTCCCAGGTC | TTGTTCCTCCAGTTCCGATT |
| F2R | CCGCAGGCCAGAATCAAAAG | ACAAAGAGTGTCAGCCAGGAG |
| HES7 | CCCCAAGATGCTCAAGCCG | GGTTCCGGAGGTTCTGGTC |
| LPAR4 | GGCGGTATTTCAGCCTCTTT | AGCAGGTGGTGGTTGCATTG |
| ARG1 | TCCCGATGTGCCAGGATTCT | ACGTCTCTCAAGCCAATATA |
| APP | CTCGTCACGTGTTCAATATG | GGGTGTGCTGTCTGTCCTTC |
| HMBS | GAAGGATGGGCAACTGTACC | ATGGTAGCCTGCATGGTCTC |
| PRNP | GTGCACGACTGCGTCAAT | CCTTCCTCATCCCACTATCAGG |

## Quantification of gene editing efficiency via SMRT long-read sequencing

HEK293 cells were seeded in 24-well plates at a density of 4.0 ×$10^5$ cells per well. 24 hours later, cells growing at ~90 % confluency were co-transfected with lentiCas9-Blast (Addgene, 52962, 250 ng per well) and sgRNA plasmids (250 ng per well) using the Lipofectamine 3000 transfection reagent (Thermo Fischer Scientific, L3000015) according to the manufacturer's introductions. 24 hours post transfection, the cells were split to puromycin (1μg/ml) containing medium for 72 hours. Then cells were cultured in medium without selection for around one week and afterwards cells were harvested for genomic DNA isolation using the DNeasy blood & tissue kit (Qiagen, 69506).

Barcoded primers (flanking 4sgRNA targeting region) were synthesized to amplify the genomic edited region of the corresponding genes. Genomic DNA was used as the template for PCR amplification of the targeted region in the genome using Phusion high-fidelity DNA polymerase (New England Biolabs, M0530S). For each PCR reaction of 50 μl volume, 150 ng genomic DNA, 0.5 μl Phusion DNA polymerase, 5 μM forward/reverse primers, 10 mM dNTP, and 10 μl 5× Phusion HF buffer were included, followed by temperature conditions: Initial denaturation at 98 °C for 30 seconds, 37 cycles including 98 °C for 10 seconds, 60 °C for 30 seconds, and 72 °C for 30 seconds per Kb, and final extension at 72 °C for 10 min. Then the PCR products were purified with gel extraction using the NucleoSpin gel and PCR clean-up kit (Macherey Nagel, 40609.250). Purified PCR

amplicons were pooled with roughly equal molar amount (determined by Nanodrop) and subjected to SMRT long-read sequencing.

### Lentiviral packaging

HEK293T cells were grown to 80-90% confluency in DMEM + 10% FBS on poly-D-lysine coated 24-wells plates and transfected with the 3 different plasmids (Transfer plasmid, pAX-2 and VSV-G; ratios: 5:3:2) with lipofectamine 3000 for lentivirus production. After 6 hours, or overnight incubation, the medium is changed to virus harvesting medium (DMEM + 10% FBS + 1% BSA). The supernatant containing the lentiviral particles was then harvested 48-72 hours after the change to virus harvesting medium. Suspended cells or cellular debris was pelleted with centrifugation at 1500 rpm for 5 min. Then clear supernatant was titrated and stored at 80 °C.

For the titration of the lentiviral particles, the same number of HEK293T cells were grown in 24-well plates, and infected by adding small volumes (V) of the above-mentioned viral supernatant (e.g. 3 µL). A representative batch of cells was used to determine the cell count at the time of infection (N). 72 hours after infection, the cells were harvested and analysed by flow cytometry to quantify the fraction of infected cells (BFP positive). The percentage of positive cells (P) is then used to calculate the titre (T) of the virus according to the following formula: $T = \frac{\mathrm{P*N}}{V}$

### Cell culture

HCT116-Cas9 cells were grown in DMEM medium (GIBCO) supplemented with 10% FBS (GIBCO) and 1x penicillin/streptomycin (GIBCO) in a humidified incubator at 37 °C with 5% $CO_2$. For passaging, HCT116-Cas9 cells were washed once with D-PBS (GIBCO) and detached using 0.25% Trypsin (GIBCO). iPSC-iCas9 cells were cultured in mTeSR (Stem Cell Technologies) supplemented with penicillin/streptomycin (GIBCO) and doxycycline (200 ng/ml; Clontech) on laminin-521 (Biolamina) coated plates at 37°C and 5% $CO_2$. For routine maintenance, approx. 70% confluent cultures were dissociated into single cells with TrypLE (GIBCO) and seeded at a seeding density of 10,000-25,000 cells/cm2 in mTeSR supplemented with 2 µM ROCK inhibitor Y27632 (Tocris) onto laminin-521 coated plates. After 24h, the medium was replaced with mTeSR without ROCKi followed by daily medium changes. For kidney organoid differentiation and maintenance please refer to this publication *PMID: 34847364*. HEK293T cells were cultured in DMEM medium (GIBCO) supplemented with 10% FBS (GIBCO) in a humidified incubator at at 37 °C with 5% $CO_2$.

### Lentiviral production

To produce virus, individual 4sgRNA plasmids were co-transfected into HEK293T cells with Ready-to-Use Lentiviral Packaging Plasmid Mix (Cellecta). HEK293T cells were cultured in DMEM medium (as described above) and seeded in 100 $cm^2$ collagen I-coated tissue culture plates in a total volume of 10 ml growth medium. 4sgRNA plasmid was mixed with Ready-to-Use Lentiviral Packaging Plasmid Mix in OptiMEM (GIBCO) to a volume of 250 µl. TransIT transfection reagent (Mirius) was diluted with OptiMEM (GIBCO) to a total volume of 250 µl and incubated for 5 minutes at room temperature (RT). Both solutions were mixed and incubated for 15 minutes at RT. TransIT-plasmid mix was added dropwise to the cells and cultured in a humidified incubator at at 37 °C with 5% $CO_2$. Medium was exchanged 24h post-transfection. Viral particles were harvested after 72 hours by filtering the viral supernatant through a 0.22 µm Steriflip-GP filter (Merck) and immediately snap frozen in liquid nitrogen and stored at −80°C until usage.

### Transduction

Each batch of virus was titrated by transduction of $1.5x10^5$ HCT116-Cas9 or iPSC-iCas9 cells per well in a 6-well plate or $2.5x10^3$ NPC-iCas9 cells per well of an ultra-low attachment (ULA) 384 well plate with different dilutions of the virus in each well. For each viral concentration and no virus control, three to four replicate wells were seeded. After 24 hours, medium containing 2 µg/ml puromycin and 8 µg/ml polybrene (HCT116-Cas9) and 1 µg/ml (iPSC-iCas9) or no (NPCs-iCas9) puromycin (GIBCO) was exchanged. A non-virus control was always included and untransduced control cells did not survive 72 hours of puromycin treatment. After puromycin selection, live and dead cells were

counted. The viral titer of each batch was identified by calculating the percentage of puromycin surviving cells relative to the no virus control. For NPCs, all viral volumes were functionally read out via FACS analysis. For all subsequent transductions, the viral volume was calculated to reach a MOI of 5. Cells were cultured as described above and knockout was measured 4 or 8 days (HCT116-Cas9 or iPSC-iCas9) or >14 days (NPC-iCas9) after transduction.

### Transfection

HCT116-Cas9 or iPSC-iCas9 cells were seeded so that cells reached 60-80% confluency 24 hours post-seeding. On the day of transfection, growth medium was exchanged for medium without penicillin or streptomycin. 4sgRNA plasmids or synthetic guides complexed with the tracrRNA following the manufacturer's protocol (IDT) as described below were diluted at different concentrations, (1, 2, 5 µg of plasmid or 5, 10 µM tracr-complexed synthetic guide RNAs) in OptiMEM (GIBCO). Lipofectamine 2000 (Invitrogen) was diluted in OptiMEM according to manufacturer's protocol and incubated for 5 minutes at RT. DNA or RNA and diluted Lipofectamine 2000 were mixed dropwise and incubated for 20 minutes at RT. The DNA/Lipofectamine 2000 mixture was added gently to the cells. Growth medium was exchanged 24h post-transfection. Cells were cultured as described above and knockout was measured 4 or 8 days after transfection.

### Nucleofection

$2x10^5$ HCT116-Cas9 and iPSC-iCas9 were resuspended in 20 µl SE cell line nucleofection solution (Lonza) (HCT116-Cas9) or P3 primary cell nucleofection solution (Lonza) (iPSC-iCas9). Cells were mixed and incubated at room temperature for 2 min in PCR tubes. Different concentrations of the 4sgRNA plasmid (1, 2, 5 µg) or synthetic guide RNAs (5, 10 µM) were mixed and the cell/reagent/nucleofection mix was transferred to Nucleofection cuvette strips (Lonza). Cells were electroporated using a 4D nucleofector (4D-Nucleofector Core Unit: Lonza, AAF-1002B; 4D-Nucleofector X Unit: AAF-1002X; Lonza). Programs were adapted for the different cell types (HCT116-Cas9: EN-113, iPSC-iCas9: CD-118). After nucleofection, prewarmed cell-specific growth media was used to transfer transfected cells in culture plates containing pre-warmed cell-specific growth media. Cells were cultured as described above and knockout was measured 4 or 8 days post-nucleofection.

### Live immunostaining and FACS analysis

Cells were harvested and resuspended at a concentration of $1x10^6$ cells/100µl in FACS buffer (1x PBS (GIBCO), 0.5M EDTA (Sigma) and 1% FBS (GIBCO)). Afterwards, per $1x10^6$ cells, 1 µl of Alexa488 anti-human EPCAM (Abcam, ab112067) or Alexa647 anti-mouse/human CD44 (Biolegend, 103018) was added. After incubation for 10-20 minutes at RT, cells were washed 2x with 2 ml FACS buffer. Afterwards, cells were resuspended in 250 µl FACS buffer and analyzed with a Fortessa (BD) or Canto (BD) analyzer.

### Preparation of crRNA–tracrRNA duplex and precomplexing of Cas9/RNP

To prepare the duplex, each Alt-R crRNA and Alt-R tracrRNA (IDT) was reconstituted to 200 µM with Nuclease-Free Duplex Buffer (IDT). Oligos were mixed at equimolar concentrations in a sterile PCR tube (e.g., 10 µl Alt-R crRNA and 10 µl Alt-R tracrRNA). Oligos were annealed by heating at 95°C for 5 min in PCR thermocycler and the mix was slowly cooled to room temperature.

### FLAER assay

NPCs were transduced with lentiviruses carrying the 4sgRNA plasmid targeting the gene *PIGA* or with a pool of four individual lentiviruses each carrying a sgRNA targeting the gene *PIGA* as described above. After 46 days post-transduction, organoids were dissociated into single cells and stained with FLAER-488 reagent (Biozol) in 3% BSA (blocking solution) according to the manufacturer's protocol. Subsequently, the percentage of FLAER-negative cells in each condition were analyzed using a Fortessa FACS analyzer (BD).

## Cell culturing for $PrP^C$ screening

U-251 MG human cells (Kerafast, Inc., Boston, MA, USA, AccessionID: CVCL_0021) expressing dCas9-VPR (plasmid #96917; https://www.addgene.org/96917/) were cultured in T150 tissue culture flasks (TPP, Trasadingen, Switzerland) in OptiMEM without Phenol (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% FBS (Takara, Goteborg, Sweden), 1% NEAA (Gibco), 1% GlutaMax (Gibco), 1% Penicillin/Streptomycin (P/S) (Gibco) and blasticidin (Gibco) at a concentration of 10 ug/mL. Once the cells reach a confluency of 80-90%, they were harvested with Accutase (Gibco), washed with PBS (Kantonsapotheke, Zurich, Switzerland) and resuspended in medium, pooled, and counted using TC20 (BioRad) Cell Counter with trypan blue (Gibco).

## $PrP^C$ screening workflow

U-251 MG dCas9-VPR (5'000 cells per well) were seeded in 30ul of medium into white 384-well CulturPlates (Greiner Bio-One, item no.:781080). The plates were incubated in a rotating tower incubator (LiCONiC StoreX STX, Schaanwald, Liechtenstein) for 24 hours. Afterwards, plates were removed from the incubator and cells were transduced with lentiviruses containing the sgRNA against each TFs. At the same time, in each plate, 14 wells were transduced with non-targeting (NT) and other 14 wells with *PRNP* targeting controls. Experiments were performed in triplicates. Plates were incubated in a rotating tower incubator for four days. Subsequently, one replica was used to determine cell viability: plates were removed from the incubator and centrifuged at 1000xg for 1 minute (Eppendorf 5804R, Hamburg, Germany). Medium was removed by inverting the plates and replaced with 25ul of fresh medium and 25ul of CellTiter-Glo® (Promega). The plates were incubated on a plate shaker (Eppendorf ThermoMixer Comfort) for 2 min (room temperature, 400 rpm shaking conditions) and, after 10 minutes of incubation at room temperature without shaking, the luminescence was measured with the EnVision plate reader (Perkin Elmer). The other two replica were used to assess $PrP^C$ levels by the TR-FRET method. Four days post transduction, the medium was removed by inverting the plates, and cells were lysed in 10 µL lysis buffer (0.5% Na-Deoxycholate (Sigma Aldrich St. Louis, MO, USA) 0.5% Triton X (Sigma Aldrich), supplemented with EDTA-free cOmplete Mini Protease Inhibitors (Roche, Basel, Switzerland) and 0.5% BSA (Merck, Darmstadt, Germany). Following lysis, assay plates were incubated on a plate shaker (Eppendorf ThermoMixer Comfort) for 10 min (4˚C, 400 rpm shaking conditions) prior to centrifugation at 1000xg for 1 min and incubated at 4˚C for two additional hours. Following incubation, plates were centrifuged once more under same conditions mentioned above and 5 µL of each FRET antibody pair was added (2.5 nM final concentration for donor and 5 nM for acceptor, diluted in 1x Lance buffer (Perkin Elmer)). For FRET, two distinct anti-PrP antibodies, POM1 (binding to amino acid residue a.a 144–152) and POM2 (binding to a.a 43–92) (Polymenidou et al., 2008), targeting different epitopes of $PrP^C$ were coupled to a FRET donor, Europium (EU) and a FRET acceptor, Allophycocyanin (APC), respectively, following previously reported protocols (Ballmer et al., 2017). Plates were centrifuged once more and incubated overnight at 4˚C. TR-FRET measurements were read out using previously reported parameters (Ballmer et al., 2017) on an EnVision multimode plate reader (Perkin Elmer).

## References

Birnboim, H.C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7, 1513-1523.

Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera Mdel, C., and Yusa, K. (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat Biotechnol 32, 267-273.

Metzakopian, E., Strong, A., Iyer, V., Hodgkins, A., Tzelepis, K., Antunes, L., Friedrich, M.J., Kang, Q., Davidson, T., Lamberth, J., et al. (2017). Enhancing the genome editing toolbox: genome wide CRISPR arrayed libraries. Sci Rep 7, 2244.

Abugessaisa, I., Noguchi, S., Hasegawa, A., Harshbarger, J., Kondo, A., Lizio, M., Severin, J., Carninci, P., Kawaji, H., and Kasukawa, T. (2017). FANTOM5 CAGE profiles of human and mouse reprocessed for GRCh38 and GRCm38 genome assemblies. Sci. Data 4, 170107.

Ballmer BA, Moos R, Liberali P, Pelkmans L, Hornemann S, Aguzzi A. Modifiers of prion protein biogen[1]esis and recycling identified by a highly parallel endocytosis kinetics assay. J Biol Chem. 2017; 292 (20):8356–68.

Braschi, B., Denny, P., Gray, K., Jones, T., Seal, R., Tweedie, S., Yates, B., and Bruford, E. (2019). Genenames.org: the HGNC and VGNC resources in 2019. Nucleic Acids Res. 47, D786–D792.
Concordet, J.-P., and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. 46, W242–W245.
Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. 34, 184–191.
FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest, A.R.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J.L., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., et al. (2014). A promoter-level mammalian expression atlas. Nature 507, 462–470.
Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., et al. (2019). GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. 47, D766–D773.
Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell 159, 647–661.
Glusman, G., Caballero, J., Mauldin, D.E., Hood, L., and Roach, J.C. (2011). Kaviar: an accessible system for testing SNV novelty. Bioinforma. Oxf. Engl. 27, 3216–3217.
Graf, R., Li, X., Chu, V.T., and Rajewsky, K. (2019). sgRNA Sequence Motifs Blocking Efficient CRISPR/Cas9-Mediated Gene Editing. Cell Rep. 26, 1098-1103.e3.
Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol. 17, 148.
Hanna, R.E., and Doench, J.G. (2020). Design and analysis of CRISPR-Cas experiments. Nat. Biotechnol.
Hart, T., Tong, A.H.Y., Chan, K., Van Leeuwen, J., Seetharaman, A., Aregger, M., Chandrashekhar, M., Hustedt, N., Seth, S., Noonan, A., et al. (2017). Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. G3 Bethesda Md 7, 2719–2727.
Horlbeck, M.A., Gilbert, L.A., Villalta, J.E., Adamson, B., Pak, R.A., Chen, Y., Fields, A.P., Park, C.Y., Corn, J.E., Kampmann, M., et al. (2016). Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. ELife 5.
Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. 31, 827–832.
Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods 12, 115–121.
Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell 172, 650–665.
Perez, A.R., Pritykin, Y., Vidigal, J.A., Chhangawala, S., Zamparo, L., Leslie, C.S., and Ventura, A. (2017). GuideScan software for improved single and paired CRISPR guide RNA design. Nat. Biotechnol. 35, 347–349.
Polymenidou M, Moos R, Scott M, Sigurdson C, Shi YZ, Yajima B, et al. The POM monoclonals: a comprehensive set of antibodies to non-overlapping prion protein epitopes. PLoS One. 2008; 3(12):e3872.
R Core Team (2020). R: A Language and Environment for Statistical Computing (Vienna, Austria: R Foundation for Statistical Computing).
Sanson, K.R., Hanna, R.E., Hegde, M., Donovan, K.F., Strand, C., Sullender, M.E., Vaimberg, E.W., Goodale, A., Root, D.E., Piccioni, F., et al. (2018). Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. Nat. Commun. 9, 5416.
Smedley, D., Haider, S., Ballester, B., Holland, R., London, D., Thorisson, G., and Kasprzyk, A. (2009). BioMart--biological queries made easy. BMC Genomics 10, 22.
Tycko, J., Wainberg, M., Marinov, G.K., Ursu, O., Hess, G.T., Ego, B.K., Aradhana, null, Li, A., Truong, A., Trevino, A.E., et al. (2019). Mitigation of off-target toxicity in CRISPR-Cas9 screens for essential non-coding elements. Nat. Commun. 10, 4063.
Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science 347, 1260419.
Uhlén, M., Karlsson, M.J., Hober, A., Svensson, A.-S., Scheffel, J., Kotol, D., Zhong, W., Tebani, A., Strandberg, L., Edfors, F., et al. (2019). The human secretome. Sci. Signal. 12.
Yates, A.D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., et al. (2020). Ensembl 2020. Nucleic Acids Res. 48, D682–D688.

## Supplementary Figures

**Figure S1. Details of APPEAL cloning method. A**, Step-by-step details of APPEAL cloning method. **B**, Zoom-in illustration of homologous ends overlapping among the three amplicons and the digested vector pYJA5.

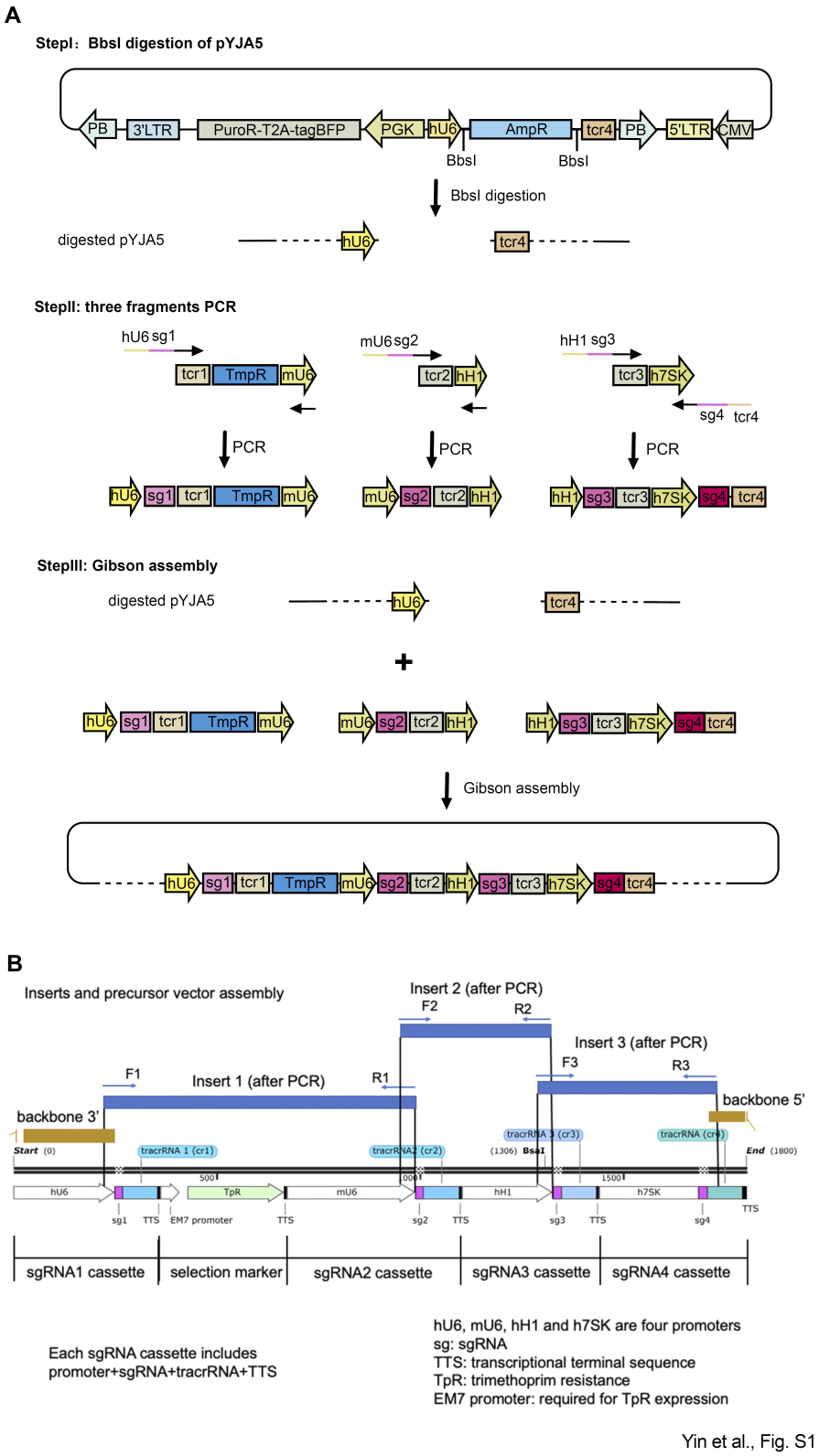
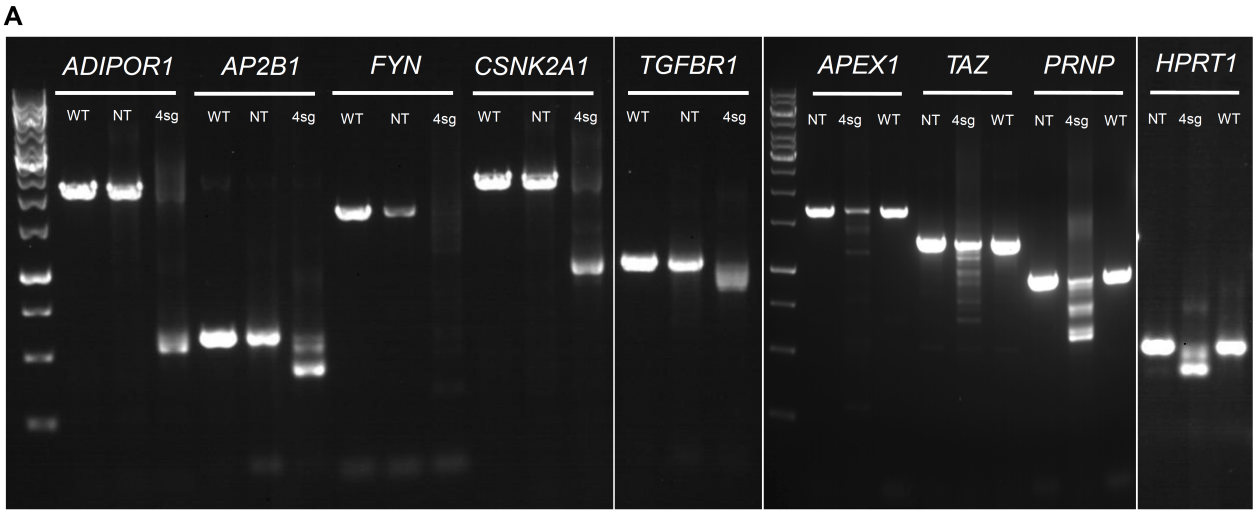


**Figure S2. High efficiency of 4sgRNAs in gene ablation. A**, Gel examination of 4sgRNA knockout plasmids in generating genomic DNA deletions.




Yin et al., Fig. S2

**Figure S3. The effect of sgRNA spacing and homology on 4sgRNA plasmids and other features of T.spiezzo and T.gonfio libraries. A**, Comparison of the effect of overlapping and non-overlapping sgRNAs on gene activation in HEK293 cells. **B**, Correlation between the extent of homology among the 4 sgRNAs and the percentage of correct plasmids. **C**, Correlation between the extent of homology and the frequency of shortened amplicon regions (indicating deletions). **D**, Summary of the number of transcription start sites (TSSs) per gene that are each targeted by a separate plasmid in the T.gonfio library (top), and the estimated size of deletions between the first and last cut sites of each 4 sgRNA plasmid in the T.spiezzo library (bottom). **E**, Percentage of sgRNAs that target genomic site affected by a polymorphism with frequency higher than 0.1% in the T.spiezzo and T.gonfio libraries in comparison with the top 4 sgRNAs from existing resources. **F**, Percentage of sgRNAs that share 8 or more base pairs of homology in the T.spiezzo and T.gonfio libraries in comparison with the top 4 sgRNAs from existing resources. **G** and **H**, Comparison of the percentage of sgRNAs predicted to target unintended genes at off-site locations (**G**) and all locations (**H**) – the latter include mostly sgRNAs with *on-site* unintended targets. **I**, All plasmids in the T.spiezzo and T.gonfio libraries were assigned to mutually exclusive categories, based on whether any of the 4 sgRNAs may target additional, unintended genes.

**Figure S4. Genome-wide sequencing of the T.spiezzo and T.gonfio libraries. A,** PacBio long-read sequencing workflow: polymerase chain reaction (PCR) was performed in each well of a 384-well plate using primers appended with row- and column-specific barcodes. All wells from one plate were pooled and ligated with plate-specific barcodes, and multiple plates were further pooled for sequencing. **B**, High-quality read count for each well in the T.spiezzo and T.gonfio libraries. **C** and **D**, Cumulative distribution of each well of plasmids with 0, 1, 2, 3, 4 entirely correct sgRNA and tracrRNA sequences, as well as an associated promoter sequence that was at least 95% correct, in the T.spiezzo and T.gonfio libraries. **E** and **F**, Predicted off-target effects for mutated sgRNAs in the T.spiezzo and T.gonfio libraries. Guide RNAs were considered to target a gene if they lay within coding sequences or exons (for CRISPR knockout plasmids) or within 1000 base pairs of a transcription start site (for CRISPR activation plasmids).

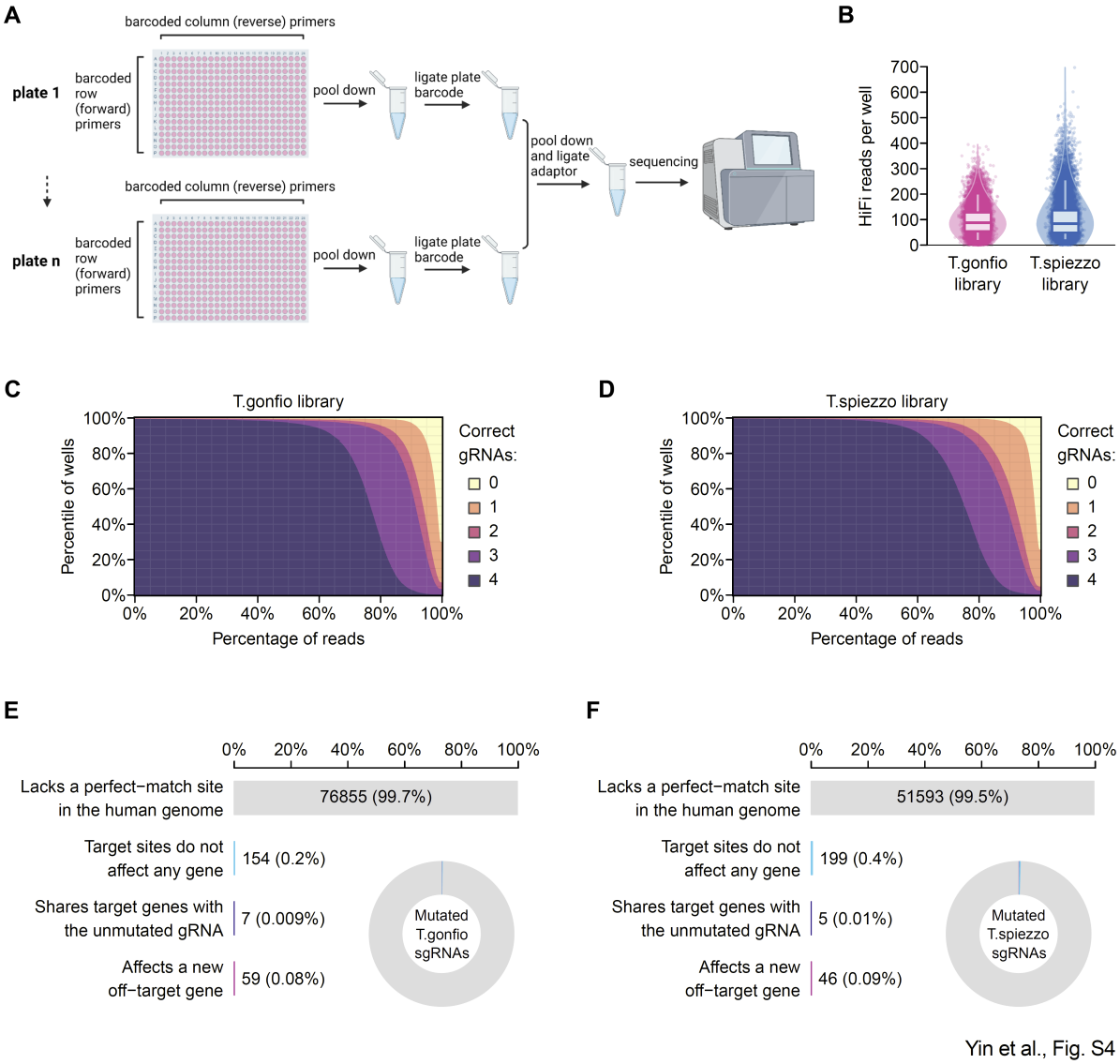
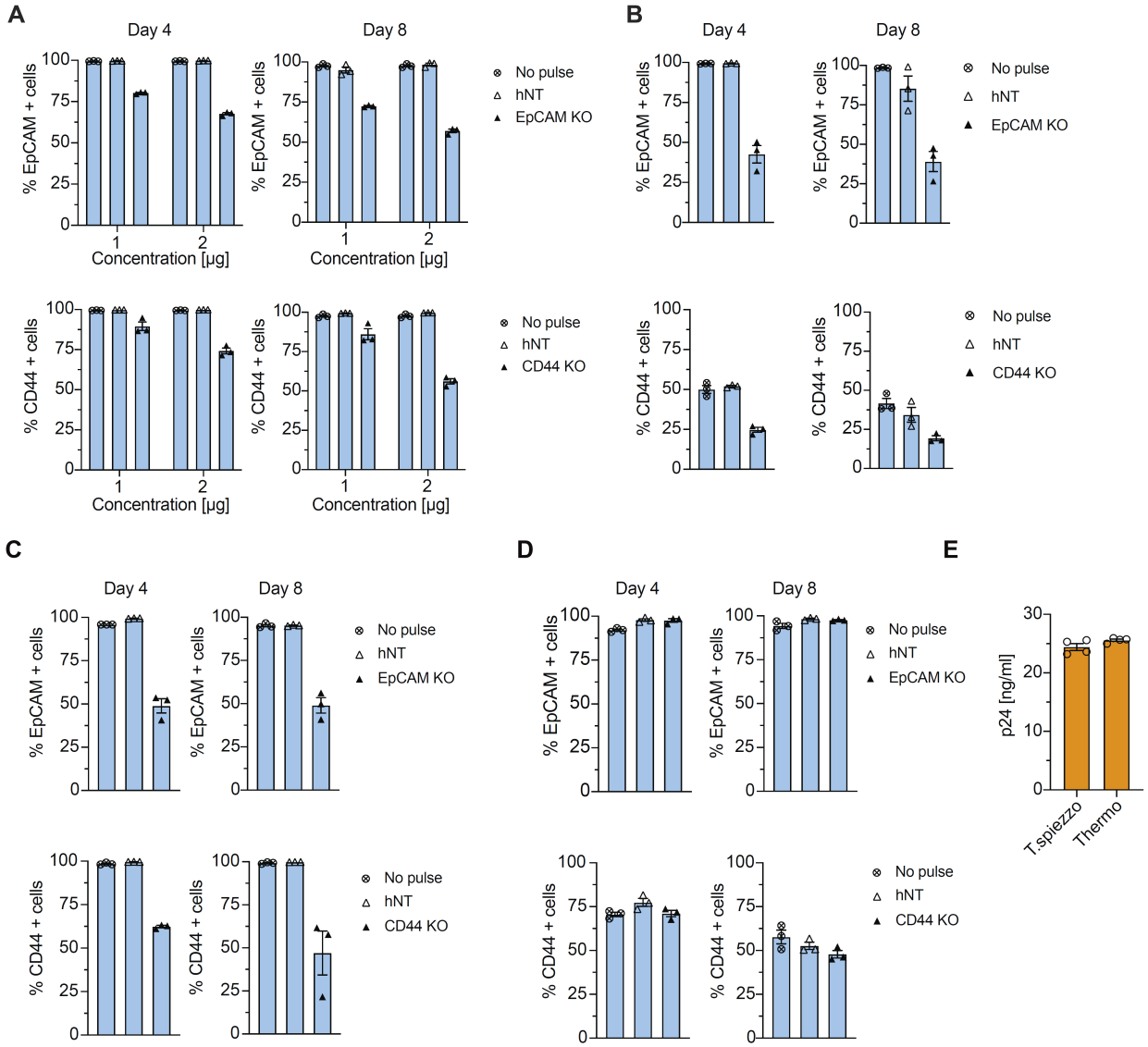


**Figure S5: Benchmarking of the 4sgRNA approach against commercially available lentiviruses and synthetic guide RNAs**

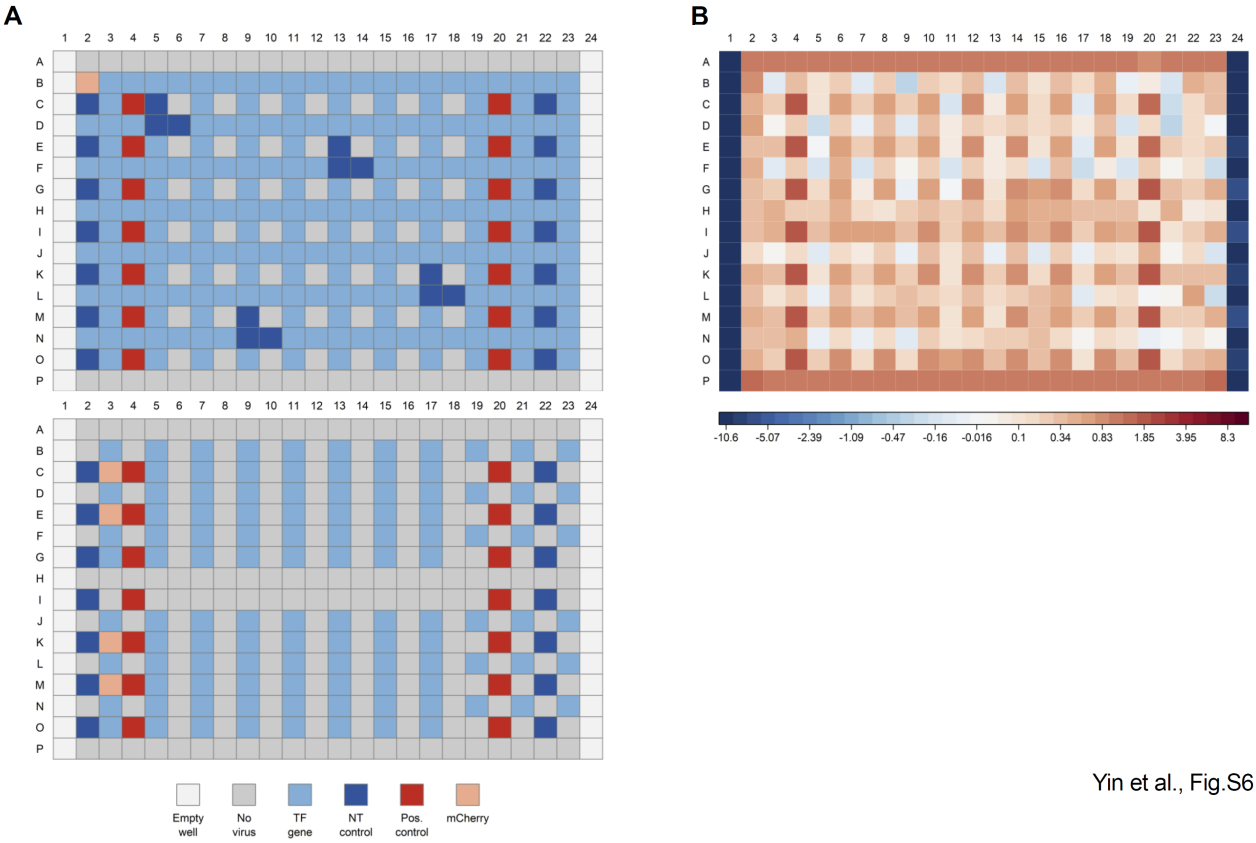
**A**, Bar plots showing percentage of EPCAM (*top*) or CD44 (*bottom*) positive HCT116-Cas9 cells electroporated with the 4sgRNA vector (*T.spiezzo*) targeting *EPCAM* or *CD44* compared to the no pulse and non-targeting (*hNT*) control at 1 or 2 µg after four and eight days post electroporation (*n*=3; error bars represent SEM). **B**, Bar plots showing percentage of EPCAM (*top*) or CD44 (*bottom*) positive iPSC-iCas9 cells electroporated with the 4sgRNA vector (*T.spiezzo*) targeting *EPCAM* or *CD44* compared to the no pulse and non-targeting (*hNT*) control at 5 µg after four and eight days post electroporation (*n*=3; error bars represent SEM). **C**, Bar plots showing percentage of EPCAM (*top*) or CD44 (*bottom*) positive HCT116-Cas9 cells electroporated with four individual synthetic guide RNAs (*IDT*) targeting *EPCAM* or *CD44* compared to the no pulse and non-targeting (*hNT*) control at 5 µM after four and eight days post electroporation (*n*=3; error bars represent SEM). **D**, Bar plots showing percentage of EPCAM (*top*) or CD44 (*bottom*) positive iPSC-iCas9 cells electroporated with four individual synthetic guide RNAs (*IDT*) targeting *EPCAM* or *CD44* compared to the no pulse and non-targeting (*hNT*) control at 5 µM after four and eight days post electroporation (*n*=3; error bars represent SEM). **E**, Bar plot showing ELISA analysis of p24 quantification in supernatant containing lentiviruses carrying the 4sgRNA vector (*T.spiezzo*) or four individual packaged sgRNAs (*Thermo*) targeting *PIGA* (*n*=4; error bars represent SEM).




Yin et al., Fig. S5

**Figure S6: Schematic of plate layout and heat map of TF sublibrary screen on modifiers of $PrP^C$ expression**

**A**, Plate layout of $PrP^C$ TFs sublibrary screen: positive controls (sgRNA targeting *PRNP*) are shown in red, non-targeting (NT) controls in blue, sgRNA targeting the TFs in light blue, not-transduced wells in gray and mCherry control in orange. **B**, Plate heat map plotted to examine temperature-induced gradients or dispensing errors.




Yin et al., Fig.S6

## Supplementary information

1. Sequence of the empty pYJA5 vector

CAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCC
ACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTT
CCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGACGGTA
CAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAG
GCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCT
GGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAAAC
TCATTTGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTT
GGAATCACACGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACT
CCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGATAA
ATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTATATAAAATTATTCATA
ATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTTCTATAGTGAATAGAG
TTAGGCAGGGATATTCACCATTATCGTTTCAGACCCACCTCCCAACCCCGAGGGGACCCGAC
AGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCGATTAGT
GAACGGATCGGCACTGCGTGCGCCAATTCTGCAGACAAATGGCAGTATTCATCCACAATTTTA
AAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACA
GACATACAAACTAAAGAATTACAAAAACAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAG
GGACAGCAGAGATCCAGTTTGGTTAGTACCGGGCCCTACGCGTTACTTAACCCTAGAAAGAT
AATCATATTGTGACGTACGTTAAAGATAATCATGCGTAAAATTGACGCATGTGTTTTATCGGTC
TGTATATCGAGGTTTATTTATTAATTTGAATAGATATTAAGTTTTATTATATTTACACTTACATAC
TAATAATAAATTCAACAAACAATTTATTTATGTTTATTTATTTATTAAAAAAAAACAAAAACTCAAA
ATTTCTTCTATAAAGTAACAAAGCaaaaaaaGCACCGACTCGGTGCCACTTTTTCAAGTTGATAA
CGGACTAGCCTTATTTCAACTTGCTACAGCATTTCTGCTGTAGCTCTGAAACccGTCTTCTTAC
CAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCT
GACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCA
ATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGG
AAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTG
CCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTAC
AGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGAT
CAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCG
ATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAAT
TCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCAT
TCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACC
GCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTC
TCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCT
TCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCA
AAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATT
GAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAA
ACAAATAGGGGTTCCGCGGAAGACcccCggtgtttcgtcctttccacaagatatataaagccaagaaatcgaaatactttca
agttacggtaagcatatgatagtccattttaaaacataattttaaaactgcaaactacccaagaaattattactttctacgtcacgtattttgtact
aatatctttgtgtttacagtcaaattaattctaattatctctctaacagccttgtatcgtatatgcaaatatgaaggaatcatgggaaataggccct
cTTCCTGCCCGACCTTGGgGATCCAATTCTACCGGGTAGGGGAGGCGCTTTTCCCAAGGCAG
TCTGGAGCATGCGCTTTAGCAGCCCCGCTGGGCACTTGGCGCTACACAAGTGGCCTCTGGC
CTCGCACACATTCCACATCCACCGGTAGGCGCCAACCGGCTCCGTTCTTTGGTGGCCCCTTC
GCGCCACCTTCTACTCCTCCCCTAGTCAGGAAGTTCCCCCCCGCCCCGCAGCTCGCGTCGTG
CAGGACGTGACAAATGGAAGTAGCAGTCTCACTAGTCTCGTGCAGATGGACAGCACCGCTGA
GCAATGGAAGCGGGTAGGCCTTTGGGGCAGCGGCCAATAGCAGCTTTGCTCCTTCGCTTTCT
GGGCTCAGAGGCTGGGAAGGGGTGGGTCCGGGGGCGGGCTCAGGGGCGGGCTCAGGGGC
GGGGCGGGCGCCCGAAGGTCCTCCGGAGGCCCGGCATTCTGCACGCTTCAAAAGCGCACG
TCTGCCGCGCTGTTCTCCTCTTCCTCATCTCCGGGCCTTTCGACCTGCATCCATCTAGATCTC
GAGCAGCTGAAGCTTACCATGACCGAGTACAAGCCCACGGTGCGCCTCGCCACCCGCGACG
ACGTCCCCAGGGCCGTACGCACCCTCGCCGCCGCGTTCGCCGACTACCCCGCCACGCGCCA

CACCGTCGATCCGGACCGCCACATCGAGCGGGTCACCGAGCTGCAAGAACTCTTCCTCACG
CGCGTCGGGCTCGACATCGGCAAGGTGTGGGTCGCGGACGACGGCGCCGCGGTGGCGGTC
TGGACCACGCCGGAGAGCGTCGAAGCGGGGGCGGTGTTCGCCGAGATCGGCCCGCGCATG
GCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCCTCCTGGCGCCG
CACCGGCCCAAGGAGCCCGCGTGGTTCCTGGCCACCGTCGGgGTCTCGCCCGACCACCAGG
GCAAGGGTCTGGGCAGCGCCGTCGTGCTCCCCGGAGTGGAGGCGGCCGAGCGCGCCGGG
GTGCCCGCCTTCCTGGAGACCTCCGCGCCCCGCAACCTCCCCTTCTACGAGCGGCTCGGCT
TCACCGTCACCGCCGACGTCGAGGTGCCCGAAGGACCGCGCACCTGGTGCATGACCCGCAA
GCCCGGTGCCGGCGGCGGGTCCGGAGGAGAGGGCAGAGGAAGTCTCCTAACATGCGGTGA
CGTGGAGGAGAATCCTGGCCCAATGAGCGAGCTGATTAAGGAGAACATGCACATGAAGCTGT
ACATGGAGGGCACCGTGGACAACCATCACTTCAAGTGCACATCCGAGGGCGAAGGCAAGCC
CTACGAGGGCACCCAGACCATGAGAATCAAGGTGGTCGAGGGCGGCCCTCTCCCCTTCGCC
TTCGACATCCTGGCTACTAGCTTCCTCTACGGCAGCAAGACCTTCATCAACCACACCCAGGG
CATCCCCGACTTCTTCAAGCAGTCCTTCCCTGAGGGCTTCACATGGGAGAGAGTCACCACAT
ACGAGGACGGGGGCGTGCTGACCGCTACCCAGGACACCAGCCTCCAGGACGGCTGCCTCAT
CTACAACGTCAAGATCAGAGGGGTGAACTTCACATCCAACGGCCCTGTGATGCAGAAGAAAA
CACTCGGCTGGGAGGCCTTCACCGAGACtCTGTACCCCGCTGACGGCGGCCTGGAAGGCAG
AAACGACATGGCCCTGAAGCTCGTGGGCGGGAGCCATCTGATCGCAAACATCAAGACCACAT
ATAGATCCAAGAAACCCGCTAAGAACCTCAAGATGCCTGGCGTCTACTATGTGGACTACAGAC
TGGAAAGAATCAAGGAGGCCAACAACGAGACCTACGTCGAGCAGCACGAGGTGGCAGTGGC
CAGATACTGCGACCTCCCTAGCAAACTGGGGCACAAGCTTAATTGAGCGGCCGCTAGGTACC
TTTAAGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGG
ACTGGAAGGGCTAATTCACTCCCAAAGAAGTCAAGATCTGCTTTTTGCCTGTACTGGGTCTCT
CTGGTTAGACCAGAGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACT
AGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCG
TCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCT
AGCAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTG
CCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAA
TGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGC
AGGACAGCAAGGGGGAGGATTGGGAAGTCAATAGCAGGCATGCTGGGGATGCGGTGGGCTC
TATGGGCGGCCGTTAATGATATCTATAACAAGAAAATATATATATAATAAGTTATCACGTAAGT
AGAACATGAAATAACAATATAATTATCGTATGAGTTAAATCTTAAAAGTCACGTAAAAGATAATC
ATGCGTCATTTTGACTCACGCGGTCGTTATAGTTCAAAATCAGTGACACTTACCGCATTGACAA
GCACGCCTCACGGGAGCTCCAAGCGGCGACTGAGATGTCCTAAATGCACAGCGACGGATTC
GCGCTATTTAGAAAGAGAGAGCAATATTTCAAGAATGCATGCGTCAATTTTACGCAGACTATCT
TTCTAGGGTTAAATTAAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCA
TGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCC
GGgAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTG
CGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCA
ACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCG
CTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGT
TATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCC
AGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCA
TCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGG
CGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATAC
CTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTC
AGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGA
CCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGC
CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGA
GTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGaACAGTATTTGGTATCTGCGCTCT
GCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCG
CTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAG
AAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGA
TTTTGGTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCG

CACATTTCCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAA
ATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCA
AAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAG
AACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGA
ACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAA
AGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGG
GAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGT
AACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCATTCAGG
CTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGA
AAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGT
TGTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGACTAGTTA
TTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAA
CTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAAT
GACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTT
ACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGA
CGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCC
TACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTA
CATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACG
TCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCG
CCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGCGCGTTT
TGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAG
GGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTC
TGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTA
GCAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGACGC
AGGACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACG
CCAAAAATTTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAA
GCGGGGGAGAATTAGATCGCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAAT
ATAAATTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCC
TGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAG
GATCAGAAGAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGAT
AGAGATAAAAGACACCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGAC
CACCGCACAGCAAGCGGCCGGCCGCTGATCTTCAGACCTGGAGGAGGAGATATGAGGGA

2. Sequence of 4-gRNA-pYJA5 ($N_{20}$ indicates sgRNA sequence)

CAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCC
ACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTT
CCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGACGGTA
CAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAG
GCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCT
GGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAAAC
TCATTTGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTT
GGAATCACACGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACT
CCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGATAA
ATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTATATAAAATTATTCATA
ATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTTCTATAGTGAATAGAG
TTAGGCAGGGATATTCACCATTATCGTTTCAGACCCACCTCCCAACCCCGAGGGGACCCGAC
AGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCGATTAGT
GAACGGATCGGCACTGCGTGCGCCAATTCTGCAGACAAATGGCAGTATTCATCCACAATTTTA
AAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACA
GACATACAAACTAAAGAATTACAAAAACAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAG
GGACAGCAGAGATCCAGTTTGGTTAGTACCGGGCCCTACGCGTTACTTAACCCTAGAAAGAT
AATCATATTGTGACGTACGTTAAAGATAATCATGCGTAAAATTGACGCATGTGTTTTATCGGTC
TGTATATCGAGGTTTATTTATTAATTTGAATAGATATTAAGTTTTATTATATTTACACTTACATAC

TAATAATAAATTCAACAAACAATTTATTTATGTTTATTTATTTATTAAAAAAAAACAAAAACTCAAA
ATTTCTTCTATAAAGTAACAAAGCaaaaaaaGCACCGACTCGGTGCCACTTTTTCAAGTTGATAA
CGGACTAGCCTTATTTCAACTTGCTACAGCATTTCTGCTGTAGCTCTGAAACNNNNNNNNNNNN
NNNNNNNNNCgaggtacccaagcggcgcacaagctatataaacctgaaggaagtctcaactttacacttaggtcaagttgcttatc
gtactagagcttcagcaggaaatttaactaaaatctaatttaaccagcatagcaaatatcatttattcccaaaatgctaaagtttgagataaa
cggacttgatttccggctgttttgacactatccagaatgccttgcagatgggtggggcatgctaaatactgcagaaaaaaaGCACCCG
ACTCGGGTGCCACTTTTTCAAGTTGTAAACGGACTAGCCTTATTTCAACTTGCTATGCACTCTT
GTGCTTAGCTCTGAAACNNNNNNNNNNNNNNNNNNNNNCgggaaagagtggtctcatacagaacttataagatt
cccaaatccaaagacatttcacgtttatggtgatttcccagaacacatagcgacatgcaaatattgcagggcgccactcccctgtccctcac
agccatcttcctgccagggcgcacgcgcgctgggtgttcccgcctagtgacactgggcccgcgattccttggagcgggttgatgacgtcag
cgttcaaaaaaaGCAGCCGACTCGGCTGCCACTTTTTCAAGTTGTGTACGGACTAGCCTTATTTGA
ACTTGCTATGCAGCTTTCTGCTTAGCTCTCAAACNNNNNNNNNNNNNNNNNNNNCaaacaaggctt
ttctccaagggatatttatagtctcaaaacacacaattactttacagttagggtgagtttccttttgtgctgtttttaaaataataatttagtatttgtat
ctcttatagaaatccaagcctatcatgtaaaatgtagctagtattaaaaagaacagattatctgtcttttatcgcacattaagcctctatagttact
aggaaatattatatgcaaattaaccggggcaggggagtagccgagcttctcccacaagtctgtgcgagggggccggcgcgggcctaga
gatggcggcgtcggatcaaaaaaattaggccacacgttcaagtgcagccacaggataaatttgcactgagcctgggtgggattcggact
cgaccgcatagccttcaggagtgagttttgtgcaataccaaccgacgacttgaccctgccaagcggcaccagatttcttgcgtacgcgatc
ccctaagccaaaggtggcactcaggggaagcgcaaactgccctgcaacgggagcgttggcttcatcgctactttgacccatggtttagttc
ctcaccttgtcgtattatactatgccgatatactatgccgatgattaattgtcaacaaaaaaaGCACCGACTCGGTGCCACTTT
TTCAAGTTGATAACGGACTAGCCTTATTTAAACTTGCTATGCTGTTTCCAGCTTAGCTCTTAAA
CNNNNNNNNNNNNNNNNNNNNNCggtgtttcgtcctttccacaagatatataaagccaagaaatcgaaatactttcaagttac
ggtaagcatatgatagtccattttaaaacataattttaaaactgcaaactacccaagaaattattactttctacgtcacgtattttgtactaatatct
ttgtgtttacagtcaaattaattctaattatctctctaacagccttgtatcgtatatgcaaatatgaaggaatcatgggaaataggccctcTTCC
TGCCCGACCTTGGgGATCCAATTCTACCGGGTAGGGGAGGCGCTTTTCCCAAGGCAGTCTGG
AGCATGCGCTTTAGCAGCCCCGCTGGGCACTTGGCGCTACACAAGTGGCCTCTGGCCTCGC
ACACATTCCACATCCACCGGTAGGCGCCAACCGGCTCCGTTCTTTGGTGGCCCCTTCGCGCC
ACCTTCTACTCCTCCCCTAGTCAGGAAGTTCCCCCCCGCCCCGCAGCTCGCGTCGTGCAGGA
CGTGACAAATGGAAGTAGCAGTCTCACTAGTCTCGTGCAGATGGACAGCACCGCTGAGCAAT
GGAAGCGGGTAGGCCTTTGGGGCAGCGGCCAATAGCAGCTTTGCTCCTTCGCTTTCTGGGCT
CAGAGGCTGGGAAGGGGTGGGTCCGGGGGCGGGCTCAGGGGCGGGCTCAGGGGCGGGGC
GGGCGCCCGAAGGTCCTCCGGAGGCCCGGCATTCTGCACGCTTCAAAAGCGCACGTCTGCC
GCGCTGTTCTCCTCTTCCTCATCTCCGGGCCTTTCGACCTGCATCCATCTAGATCTCGAGCAG
CTGAAGCTTACCATGACCGAGTACAAGCCCACGGTGCGCCTCGCCACCCGCGACGACGTCC
CCAGGGCCGTACGCACCCTCGCCGCCGCGTTCGCCGACTACCCCGCCACGCGCCACACCGT
CGATCCGGACCGCCACATCGAGCGGGTCACCGAGCTGCAAGAACTCTTCCTCACGCGCGTC
GGGCTCGACATCGGCAAGGTGTGGGTCGCGGACGACGGCGCCGCGGTGGCGGTCTGGACC
ACGCCGGAGAGCGTCGAAGCGGGGGCGGTGTTCGCCGAGATCGGCCCGCGCATGGCCGAG
TTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCCTCCTGGCGCCGCACCGG
CCCAAGGAGCCCGCGTGGTTCCTGGCCACCGTCGGgGTCTCGCCCGACCACCAGGGCAAGG
GTCTGGGCAGCGCCGTCGTGCTCCCCGGAGTGGAGGCGGCCGAGCGCGCCGGGGTGCCC
GCCTTCCTGGAGACCTCCGCGCCCCGCAACCTCCCCTTCTACGAGCGGCTCGGCTTCACCG
TCACCGCCGACGTCGAGGTGCCCGAAGGACCGCGCACCTGGTGCATGACCCGCAAGCCCG
GTGCCGGCGGCGGGTCCGGAGGAGAGGGCAGAGGAAGTCTCCTAACATGCGGTGACGTGG
AGGAGAATCCTGGCCCAATGAGCGAGCTGATTAAGGAGAACATGCACATGAAGCTGTACATG
GAGGGCACCGTGGACAACCATCACTTCAAGTGCACATCCGAGGGCGAAGGCAAGCCCTACG
AGGGCACCCAGACCATGAGAATCAAGGTGGTCGAGGGCGGCCCTCTCCCCTTCGCCTTCGA
CATCCTGGCTACTAGCTTCCTCTACGGCAGCAAGACCTTCATCAACCACACCCAGGGCATCC
CCGACTTCTTCAAGCAGTCCTTCCCTGAGGGCTTCACATGGGAGAGAGTCACCACATACGAG
GACGGGGGCGTGCTGACCGCTACCCAGGACACCAGCCTCCAGGACGGCTGCCTCATCTACA
ACGTCAAGATCAGAGGGGTGAACTTCACATCCAACGGCCCTGTGATGCAGAAGAAAACACTC
GGCTGGGAGGCCTTCACCGAGACtCTGTACCCCGCTGACGGCGGCCTGGAAGGCAGAAACG
ACATGGCCCTGAAGCTCGTGGGCGGGAGCCATCTGATCGCAAACATCAAGACCACATATAGA
TCCAAGAAACCCGCTAAGAACCTCAAGATGCCTGGCGTCTACTATGTGGACTACAGACTGGA
AAGAATCAAGGAGGCCAACAACGAGACCTACGTCGAGCAGCACGAGGTGGCAGTGGCCAGA

TACTGCGACCTCCCTAGCAAACTGGGGCACAAGCTTAATTGAGCGGCCGCTAGGTACCTTTA
AGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTG
GAAGGGCTAATTCACTCCCAAAGAAGTCAAGATCTGCTTTTTGCCTGTACTGGGTCTCTCTGG
TTAGACCAGAGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGG
AACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTG
TTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGC
AGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCC
CTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGA
GGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGG
ACAGCAAGGGGGAGGATTGGGAAGTCAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTAT
GGGCGGCCGTTAATGATATCTATAACAAGAAAATATATATATAATAAGTTATCACGTAAGTAGA
ACATGAAATAACAATATAATTATCGTATGAGTTAAATCTTAAAAGTCACGTAAAAGATAATCATG
CGTCATTTTGACTCACGCGGTCGTTATAGTTCAAAATCAGTGACACTTACCGCATTGACAAGC
ACGCCTCACGGGAGCTCCAAGCGGCGACTGAGATGTCCTAAATGCACAGCGACGGATTCGC
GCTATTTAGAAAGAGAGAGCAATATTTCAAGAATGCATGCGTCAATTTTACGCAGACTATCTTT
CTAGGGTTAAATTAAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATG
GTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGG
gAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCG
CTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAAC
GCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCT
GCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTAT
CCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAG
GAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATC
ACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCG
TTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCT
GTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCA
GTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGA
CCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGC
CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGA
GTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGaACAGTATTTGGTATCTGCGCTCT
GCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCG
CTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAG
AAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGA
TTTTGGTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCG
CACATTTCCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAA
ATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCA
AAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAG
AACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGA
ACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAA
AGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGG
GAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGT
AACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCATTCAGG
CTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGA
AAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGT
TGTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGACTAGTTA
TTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAA
CTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAAT
GACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTT
ACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGA
CGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCC
TACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTA
CATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACG
TCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCG
CCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGCGCGTTT

TGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAG
GGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTC
TGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTA
GCAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGACGC
AGGACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACG
CCAAAAATTTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAA
GCGGGGGAGAATTAGATCGCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAAT
ATAAATTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCC
TGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAG
GATCAGAAGAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGAT
AGAGATAAAAGACACCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGAC
CACCGCACAGCAAGCGGCCGGCCGCTGATCTTCAGACCTGGAGGAGGAGATATGAGGGA

**Four sgRNA primer sequence (5'-3', $N_{20}$ in sgRNA1 primer Fwd, sgRNA2 primer Fwd, sgRNA3 primer Fwd is exactly the sgRNA sequence, however, in sgRNA4 primer Rev it should be the reverse complement sequence of the sgRNA sequence):**

sgRNA1 primer Fwd: ttgtggaaaggacgaaacaccG$N_{20}$GTTTAAGAGCTAAGCTG

sgRNA2 primer Fwd: cttggagaaaagccttgtttG$N_{20}$GTTTGAGAGCTAAGCAGA

sgRNA3 primer Fwd: gtatgagaccactctttcccG$N_{20}$GTTTCAGAGCTAAGCACA

sgRNA4 primer Rev: ATTTCTGCTGTAGCTCTGAAAC$N_{20}$Cgaggtacccaagcggc

**Common primers sequences (5'-3'):**

mU6 Rev: CAAACAAGGCTTTTCTCCAAGGG

M Rev: Cgggaaagagtggtctcataca

**Constant template sequences (5'-3')**

C1 sequence

GTTTAAGAGCTAAGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTG
AAAAAGTGGCACCGAGTCGGTGCttttttttgttgacaattaatcatcggcatagtatatcggcatagtataatacgacaaggtg
aggaactaaaccatgggtcaaagtagcgatgaagccaacgctcccgttgcagggcagtttgcgcttcccctgagtgccacctttggcttag
ggatcgcgtacgcaagaaatctggtgccgcttggcagggtcaagtcgtcggttggtattgcacaaaactcactcctgaaggctatgcggt
cgagtccgaatcccacccaggctcagtgcaaatttatcctgtggctgcacttgaacgtgtggcctaattttttttgatccgacgccgccatctcta
ggcccgcgccggcccccctcgcacagacttgtgggagaagctcggctactcccctgccccggttaatttgcatataatatttcctagtaactat
agaggcttaatgtgcgataaaagacagataatctgttctttttaatactagctacattttacatgataggcttggatttctataagagatacaaat
actaaattattattttaaaaaacagcacaaaaggaaactcaccctaactgtaaagtaattgtgtgttttgagactataaatatcccttggagaa
aagccttgtttG

M sequence

GTTTGAGAGCTAAGCAGAAAGCTGCATAGCAAGTTCAAATAAGGCTAGTCCGTACACAACTTG

AAAAAGTGGCAGCCGAGTCGGCTGCttttttgaacgctgacgtcatcaacccgctccaaggaatcgcgggcccagtgtc
actaggcgggaacacccagcgcgcgtgcgccctggcaggaagatggctgtgagggacaggggagtggcgccctgcaatatttgcatg
tcgctatgtgttctgggaaatcaccataaacgtgaaatgtctttggatttgggaatcttataagttctgtatgagaccactctttcccG

C2s sequence

GTTTCAGAGCTAAGCACAAGAGTGCATAGCAAGTTGAAATAAGGCTAGTCCGTTTACAACTTG
AAAAAGTGGCACCCGAGTCGGGTGCttttttctgcagtatttagcatgccccacccatctgcaaggcattctggatagtgtc
aaaacagccggaaatcaagtccgtttatctcaaactttagcattttgggaataaatgatatttgctatgctggttaaattagattttagttaaatttc
ctgctgaagctctagtacgataagcaacttgacctaagtgtaaagttgagacttccttcaggtttatatagcttgtgcgccgcttgggtacctcG