nature medicine

LETTERS

https://doi.org/10.1038/s41591-018-0050-6

p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells

Robert J. Ihry[1], Kathleen A. Worringer[1], Max R. Salick[1], Elizabeth Frias[2], Daniel Ho[1], Kraig Theriault[1], Sravya Kommineni[1], Julie Chen[3], Marie Sondey[4], Chaoyang Ye[5], Ranjit Randhawa[1], Tripti Kulkarni[1], Zinger Yang[2], Gregory McAllister[2], Carsten Russ[2], John Reece-Hoyes[2], William Forrester[2], Gregory R. Hoffman[2], Ricardo Dolmetsch[1] and Ajamete Kaykas[1]*

CRISPR/Cas9 has revolutionized our ability to engineer genomes and conduct genome-wide screens in human cells[1–3]. Whereas some cell types are amenable to genome engineering, genomes of human pluripotent stem cells (hPSCs) have been difficult to engineer, with reduced efficiencies relative to tumour cell lines or mouse embryonic stem cells[3–13]. Here, using hPSC lines with stable integration of Cas9 or transient delivery of Cas9-ribonucleoproteins (RNPs), we achieved an average insertion or deletion (indel) efficiency greater than 80%. This high efficiency of indel generation revealed that double-strand breaks (DSBs) induced by Cas9 are toxic and kill most hPSCs. In previous studies, the toxicity of Cas9 in hPSCs was less apparent because of low transfection efficiency and subsequently low DSB induction[3]. The toxic response to DSBs was *P53/TP53*-dependent, such that the efficiency of precise genome engineering in hPSCs with a wild-type *P53* gene was severely reduced. Our results indicate that Cas9 toxicity creates an obstacle to the high-throughput use of CRISPR/Cas9 for genome engineering and screening in hPSCs. Moreover, as hPSCs can acquire *P53* mutations[14], cell replacement therapies using CRISPR/Cas9-engineered hPSCs should proceed with caution, and such engineered hPSCs should be monitored for P53 function.

Human pluripotent stem cells (hPSCs) derived from preimplantation embryos or cellular reprogramming hold great promise for screening and therapeutic applications. hPSCs are genetically intact, expandable and can be differentiated into a wide variety of cell types that are difficult to obtain from human subjects[15]. Developing a practical system for high-throughput genetic engineering of hPSCs has been challenging because hPSCs are recalcitrant to genome modification[8–13]. Several studies have shown that gene targeting in hPSCs is five- to twentyfold lower in efficiency relative to other cell types[3–7]. The cause of this reduced efficiency is unclear, but it has limited both genetic screening and therapeutic editing of hPSCs. A potential solution is the bacterial-derived CRISPR–Cas9 RNA-guided nuclease, which has been repurposed to induce user-defined double-strand breaks (DSBs) in DNA[1–3]. In transformed cells, Cas9 is extremely efficient, with minimal side effects; however, the acute effects of Cas9 have not been extensively studied in non-transformed cells such as hPSCs.

Enhancing the genetic toolkit in hPSCs is necessary to utilize their full potential in genetic screening, disease modelling and cell therapy. To increase indel efficiencies, we improved upon a two-component Cas9 system by consolidating it into a single all-in-one adeno-associated virus integration site 1 (AAVS1) targeting vector with the third-generation doxycycline (dox) inducible system and an insulator to further prevent leaky expression (henceforth iCas9; Fig. 1a, Supplementary Fig. 1a and Supplementary Data 1)[16]. The stable iCas9 lines are properly targeted, have a normal karyotype, and induce Cas9 only with dox (Supplementary Fig. 1b–e)[17].

To determine the average gene disruption (indels), iCas9 cells were infected with lentiviruses to deliver constitutively expressed synthetic-guide RNAs (sgRNAs). H1-iCas9 hPSCs were infected with 47 sgRNAs targeting 16 genes and treated with dox for 8 days in a 96-well plate. Next generation sequencing (NGS) was used to quantify control and mutant allele indel frequencies. NGS analysis of infected cells revealed high percentages of indels (Fig. 1b). The average gene disruption for the 47 sgRNAs was over 90% (Fig. 1c). Despite efficient indel generation, only a small fraction of the hPSCs survived. CRISPR/Cas9 activity caused a sharp decrease in cell number and increased cellular debris. This toxicity created variability across the wells and presents a challenge for density-dependent differentiation protocols (Supplementary Table 1).

To study toxicity in detail, we used the H1-iCas9 line and a lentiviral sgRNA targeting *MAPT*, a neuronal gene not expressed or required for survival in hPSCs. Ten days of dox treatment completely edited the *MAPT* locus and reduced colony size relative to non-targeting controls without a DSB (Fig. 1d,e). To quantify this, confluency was measured over time in cells treated with dox and expressing either a non-targeting or a *MAPT* sgRNA (Fig. 1f). Cells expressing a non-targeting control increased confluency at a steady rate, whereas those expressing a *MAPT* sgRNA decreased confluency. Despite the toxic response, *MAPT* edited cells retained expression of pluripotency proteins TRA-1-60/PODXL, OCT4/POU5F1 and SOX2 (Supplementary Fig. 1f). Initially, to determine if toxicity was related to off-target DSBs, we assayed the top six off-target sites by NGS identified by the CRISPR design tool[4] and detected no mutations at predicted sites (Supplementary Fig. 2a and Supplementary Table 2). We further reduced off-targets by decreasing the exposure to Cas9 and increasing the specificity of Cas9. Transient exposure to Cas9 and *CALM2* targeting sgRNAs by electroporating ribonucleoprotein (RNP) complexes also triggered a toxic response (>80% indels, Fig. 1g–i). The transient nature of RNP delivery minimizes off-target cutting[18] and further supports the hypothesis that DSBs

[1]Department of Neuroscience, Novartis Institutes for Biomedical Research, Cambridge, MA, USA. [2]Department of Chemical Biology and Therapeutics, Novartis Institutes for Biomedical Research, Cambridge, MA, USA. [3]Department of Oncology, Novartis Institutes for Biomedical Research, Cambridge, MA, USA. [4]Abbvie, Cambridge, MA, USA. [5]Blueprint Medicine, Cambridge, MA, USA. *e-mail: ajamete.kaykas@novartis.com

at a single locus are sufficient to cause toxicity. We next generated H1-hESC and 8402-iPSC (induced pluripotent stem cell) lines with a dox-inducible enhanced Cas9 (ieCas9) variant that reduces non-specific DSBs (Supplementary Data 2 and Supplementary Fig. 1)[17,19]. ieCas9 and sgRNAs targeting the neuronal genes *CALM2* and *EMX2* in both hESC and iPSC backgrounds caused a toxic phenotypic response (Supplementary Fig. 2b,c). This suggests that toxicity is not due to effects on other genes or many DSBs, and implies that editing at a single locus is toxic.

To further confirm the toxicity, we generated a second inducible Cas9 based on the Shield1-destabilizing domain (DD) system[20]. H1s were generated with Cas9 fused to a DD tag (ddCas9), which is stabilized in the presence of Shield1 and degraded in its absence (Supplementary Fig. 1 and Supplementary Data 3). To globally test if DNA-targeting sgRNAs are toxic we conducted a large-scale pooled CRISPR screen using both dox- and Shield1-inducible Cas9s. To control for toxicity, we screened a focused 13,000 sgRNA library at high coverage (1,000 cells per sgRNA) across four independent conditions (Fig. 2a). All four conditions were infected with the sgRNA library with two replicates. Two conditions were grown in the absence of Cas9: the parental H1 cells and H1-iCas9 cells (−dox). The remaining two conditions were grown with Cas9 induced by dox or Shield1. A total of ~200 million H1-hESCs were infected at 0.5 multiplicity of infection (MOI; 26 million per replicate). A total of 73 sgRNAs were non-targeting and the remaining targeted ~2,600 genes (5 sgRNAs per gene).

Cells were dissociated to seed new flasks and to be pelleted for DNA isolation every 4 days at 1,000 cells per sgRNA. Cell counts at day 4 demonstrated that iCas9 or ddCas9 hPSCs cultured with dox or Shield1 had little growth compared to H1 and iCas9 hPSCs infected with the same library and density but in the absence of Cas9 induction (Fig. 2b). Exposing the uninduced H1-iCas9 pool of infected cells to dox after passaging reproducibly and severely reduced cell counts (Supplementary Fig. 3). NGS was used to recover spacer sequences, which act as molecular barcodes to count sgRNA-infected cells. All but one of 24 samples recovered 98% of expected spacer sequences, demonstrating that adequate representation at 1,000× coverage per sgRNA was maintained for most sgRNAs. Fold change was calculated for each spacer sequence by comparing each condition to the sequenced lentiviral pool (before infection) using the DESeq2 method[21].

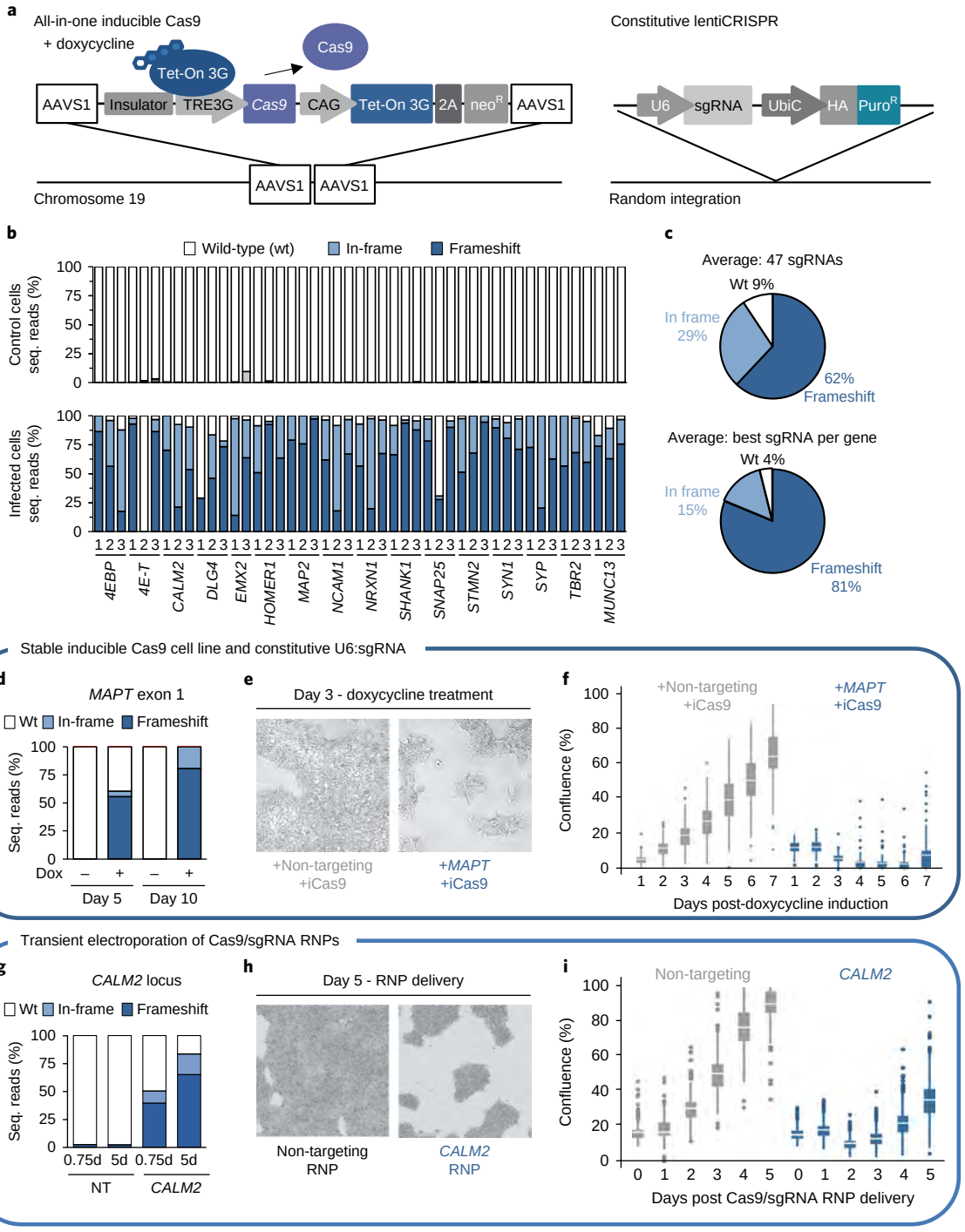
Over the 12-day experiment, most sgRNAs remained distributed within ±1 $\log_2$(fold change) in uninduced conditions (Fig. 2c and Supplementary Data 4). In contrast, the Cas9-induced conditions displayed a time-dependent change in sgRNA representation, which increased the spread of the distribution. Plotting only the non-targeting controls identified a 1.3- to 1.4-fold enrichment specific to the Cas9-induced conditions (Fig. 2d). This indicates that sgRNAs targeting the genome are globally depleted compared to non-targeting controls, and demonstrates that Cas9 toxicity is widespread over a larger number of sgRNAs. To determine if this response is specific to hPSCs, we evaluated the non-targeting controls across pooled CRISPR screens in other cell lines. Fold change was calculated for non-targeting sgRNAs from 14 additional transformed lines using genome-scale sgRNA libraries. Comparing the non-targeting controls from the Cas9-induced conditions with the transformed lines demonstrated a heightened sensitivity to DSBs in hPSCs (Fig. 2e). hPSCs have a greater than 1.3-fold change, while transformed cell lines show little enrichment (0.05- to 0.51-fold change). Finally, we exploited design flaws affecting a subset of the sgRNA library to identify additional evidence for DSB toxicity. Single nucleotide polymorphisms (SNPs) present in the H1-hESC genome disrupted target sites for 250 of the sgRNAs, reducing their ability to create DSBs and causing them to significantly enrich when compared to uninduced or Cas9-free parental lines (Supplementary Fig. 4a). Multiple perfect cut sites were identified for 142 of the sgRNAs, which enhanced their depletion (Supplementary Fig. 4b). Cumulatively, these results demonstrate that hPSCs are extremely sensitive to DSBs and the effect is widespread over many sgRNAs. This toxic effect presents a significant challenge for both engineering and screening efforts.

To further investigate the mechanism by which Cas9 causes toxicity in hPSCs, RNA-seq and differential expression analysis was performed on iCas9 cells expressing either a non-targeting or *MAPT* sgRNA grown in dox for 2 days (Fig. 3a and Supplementary Data 5). Despite the toxic response to DSBs, the expression of the pluripotency mRNAs *OCT4*, *NANOG* and *SOX2* were unchanged relative to non-targeting controls. However, a significant number of genes were induced in cells with a DSB. Gene ontology analysis of the top 100 hits identified 25 genes with roles in programmed cell death (STRING-db, FDR 1.92E-08), including genes encoding components of the intrinsic and extrinsic death pathways such as *BAX*, *BBC3*, *FAS* and *TNFRSF10B*. Consistent with this, high content imaging and analysis of DSB-induced iCas9 cells identified increases in DNA damage and apoptotic proteins, including phospho-histone H2A.X/H2AFX (pH2A.X), cleaved PARP1 (cPARP1) and cleaved caspase-3 (CASP3, referred to as CC3) (Fig. 3b).

Fig. 1 | Efficient Cas9 gene disruption is toxic to hPSCs. a, Schematic diagram of the two-component Cas9 system using an all-in-one inducible Cas9 (iCAS9) construct and lentiviral delivery of constitutive sgRNA. The AAVS1 donor vector contains a CAS9 expression cassette under control of the dox-inducible TRE3G promoter and a separate expression cassette with the Tet-On 3G activator and the neomycin resistance gene. The donor was targeted to the AAVS1 site in hPSC lines. The sgRNA lentiviral vector has a U6 promoter driving expression of the gRNA and a separate cassette for an HA-tagged version of puromycin resistance gene ($Puro^R$) for detection and selection. **b**, Editing efficiency in iCas9 cells. iCas9 cells were left untreated (control cells, top). iCas9 cells infected with lentivirus containing sgRNAs and treated with dox for 8 days (bottom). Two to three separate sgRNAs each, labelled 1, 2, 3 for 16 genes (47 sgRNAs total) were tested. The results of >10,000 reads per sample of sequencing of PCR amplicons across the sgRNA cut sites are plotted ($n=1$). Wild-type (wt) reads without mutations are represented by white bars, in-frame mutations by light blue bars and frameshift mutations by dark blue bars. **c**, Summary of efficiency and indel types generated by the 47 sgRNAs. Averages are shown for all 47 sgRNAs and the best sgRNA per gene. **d**, Indel quantification in exon 1 of the *MAPT* locus at days 5 and 10 of dox treatment; >100,000 reads per sample ($n=1$). **e**, Bright-field images of live iCas9 cells cultured in the presence of a non-targeting or *MAPT* sgRNA after 3 days of dox treatment. **f**, Quantification of the toxic response to Cas9-induced DSBs in live cells, as assessed by the percent cell confluence of cells expressing a non-targeting or *MAPT* sgRNA grown in dox for the indicated number of days. Images were taken from individual wells with either a *MAPT* sgRNA ($n=88$) or non-targeting sgRNA ($n=96$); one image per well. One representative experiment of a total of three is shown. **g**, Indel quantification at the *CALM2* locus at 0.75 and 5 days after electroporation of non-targeting (NT) and *CALM2*-targeting Cas9/sgRNA RNP complexes. Averaged data from three independent electroporations are shown; $n=3$, >3,000 reads per sample. **h**, Bright-field images of non-targeting and *CALM2* treated cells in **g** 5 days after electroporation. **i**, Quantification of the toxic response to Cas9-induced DSBs in live cells, as assessed by the percent cell confluence expressing a non-targeting- or *CALM2*-sgRNA-CAS9 RNP complex. Data are the average of three independent electroporations per sgRNA; 121 images per electroporation × three replicates, $n=363$. One representative experiment of a total of two is shown. For each box plot the mean is depicted by a white line flanked by a rectangle spanning Q1–Q3 (interquartile range, IQR). The thin vertical line marks the greatest or lowest values falling within upper/lower adjacent values (UAV/LAV). UAV is the greatest value ≤ Q3 + 1.5 × IQR. LAV is the lowest value ≤ Q1 − 1.5 × IQR. Outliers are defined as being Q3 + or Q1 − (3 × IQR) and are displayed as dots.

To identify the key pathways involved, an in silico interactome analysis was performed on the top 100 differentially expressed genes (adjusted P value cutoff of $<1.2\times10^{-17}$). Causal reasoning algorithms consistently identified *P53* as one of the top-ranking hypotheses, along with *MYC*, *SP1* and *EP300*[22,23]. These hypotheses are tightly interconnected and further investigation was focused on *P53* because of its well-established role in the DNA damage response (DDR)[24]. The one-step *P53* hypothesis accurately explained 33 of the 100 input genes (Fig. 3c) and was consistent with *P21/CDKN1A*, a canonical P53 transcriptional target, being the most differentially expressed gene[25]. *P53* mRNA was unchanged, consistent with studies demonstrating post-transcriptional regulation of *P53*[26,27].

The most differentially expressed gene was *P21* (6.12-fold, 6.6×10^{-298} P_{adj}), a cell cycle regulator and *P53* target with known roles in the DDR[28]. To confirm these results, iCas9 cells were infected with seven independent sgRNAs and treated with dox for 2 days. *P21* mRNA was then measured by qPCR (Fig. 3d). The expression of *P21* was increased between 3- and 10-fold in the targeting sgRNAs compared to a non-targeting EGFP control sgRNA. Transient exposure from electroporating Cas9 and sgRNA containing RNPs triggered a toxic response and increased *P21* expression (Fig. 3e). Additionally, the use of ieCas9 did not abrogate the induction of *P21* mRNA during DSB induction in hESCs or iPSCs, which is consistent with the toxic phenotype (Fig. 3f,g and Supplementary Fig 2b,c). Both enhanced Cas9 and transient Cas9



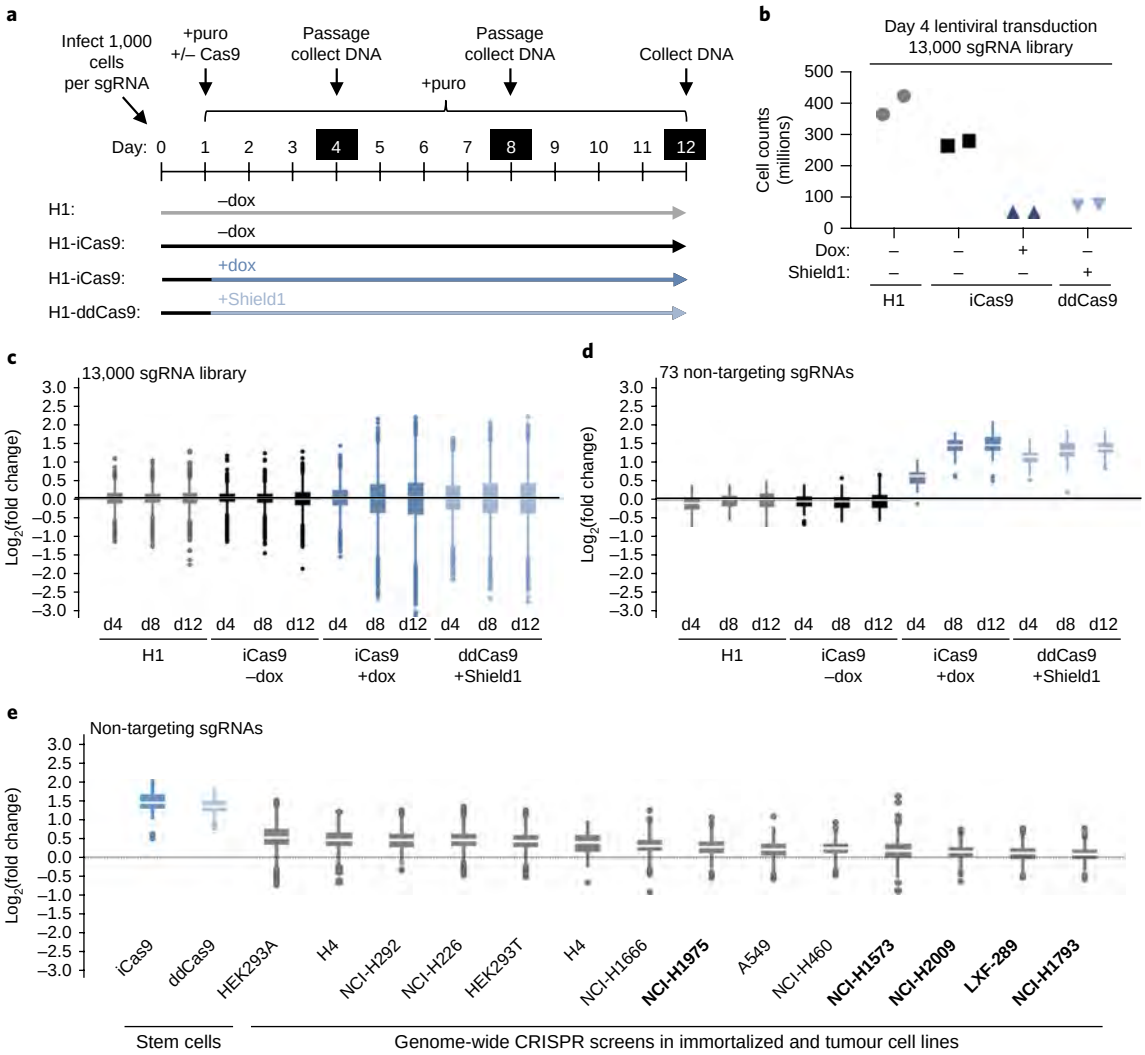


Fig. 2 | CRISPR screen identifies a hPSC-specific toxic response to Cas9-induced DSBs. a, Experimental paradigm for the pooled screen in hPSCs, in which 13,000 sgRNAs in four independent cell lines were tested. H1 parental cells, light grey; H1-iCas9 minus dox, black; H1-iCas9 plus dox, blue; H1-ddCas9 plus Shield1, light blue. For each condition 2.6×10^7 cells (two replicates) were infected at 0.5 MOI to ensure that each cell expresses a single sgRNA. At 24 h after lentiviral infection, non-transduced cells were killed by puromycin (puro). Each individual sgRNA was screened in at least 1,000 cells (1,000× coverage). On days 4, 8, and 12, cells were dissociated and counted to maintain 1,000× coverage of each sgRNA. **b**, Cell counts at day 4 in the indicated cell lines are shown; $n=2$ biologically independent replicates of sgRNA library expressing hPSCs. Replicate results are shown in Supplementary Fig. 3. **c**, Log_2(fold change) for the entire 13,000 sgRNA library at the indicated days after dox treatment; $n=12{,}910$ sgRNAs per box blot. **d**, Log_2(fold change) for 73 non-targeting control sgRNAs at the indicated days after dox treatment; $n=73$ non-targeting sgRNAs per box blot. **e**, Comparison of DSB sensitivity of human stem cell lines to transformed cells lines. Two hPSC lines (same data as in **d**) at day 12 and 14 additional transformed lines, are shown. Lines with *P53* mutations are indicated in bold. The *y* axis box plots depicting log_2(fold change) were calculated for non-targeting sgRNAs that were compared to the initial representation for each of the sgRNAs in the initial input sgRNA library. Number of non-targeting sgRNAs per screen from left to right: $n=73$, $n=73$, $n=1{,}524$, $n=943$, $n=508$, $n=508$, $n=1{,}524$, $n=146$, $n=508$, $n=508$, $n=508$, $n=508$, $n=508$, $n=508$, $n=508$, $n=508$. For **c**–**e**, the representation of each sgRNA is as determined by barcode counting of genome integrated sgRNAs, as assessed by NGS. The *y* axis box plots depicting log_2(fold change) were calculated for each sgRNA normalized to the initial representation for each of the sgRNA in the initial input sgRNA library. $n=2$ biologically independent replicates of sgRNA library expressing hPSCs. For each box plot, the mean is depicted by a white line flanked by a rectangle spanning Q1–Q3 (IQR). The thin vertical line marks the greatest or lowest values falling within upper/lower adjacent values (UAV/LAV). UAV is the greatest value ≤ Q3 + 1.5 × IQR. LAV is the lowest value ≤ Q1 − 1.5 × IQR. Outliers are defined as being Q3 + or Q1 − (3 × IQR) and are displayed as dots.

RNP delivery minimize o ff-target c utting[18,19], f urther s upporting that DSBs at a single locus are sufficient to cause a *P53*-dependent molecular response.

To provide experimental evidence that *P53* is functionally involved, w e k nocked o ut *P53* in H1-iC as9 ce lls b y t ransiently co-transfecting t hree c hemically sy nthesized sgRN As t argeting the *P53* locus (Supplementary Fig. 5a). The resulting mutant pool was a mixt ure o f co ntrol (≤50%) a nd f rameshift a lleles (≥50%) at three independent si tes within t he *P53* open-reading f rame (ORF) (S upplementary Fig . 5b). Th e co ntrol a nd *P53* mu tant pool were then infected with a *MAPT* sgRNA and grown with or without do x for u p t o 6 d ays (S upplementary Fig . 5c). T o con-firm t hat t he t ranscriptional r esponse i s *P53*-dependent, mRNA was isolated and quantified u sing qPCR. At day 2, control cells exhibited a strong induction of *P21* and *FAS* mRNA that was sig-nificantly r educed in t he *P53* mix ed m utant p ool (Fig. 4a). P53 and P21 p roteins w ere det ected u sing imm unofluorescence a nd high-content im aging. B oth P53 a nd P21 p roteins in creased in

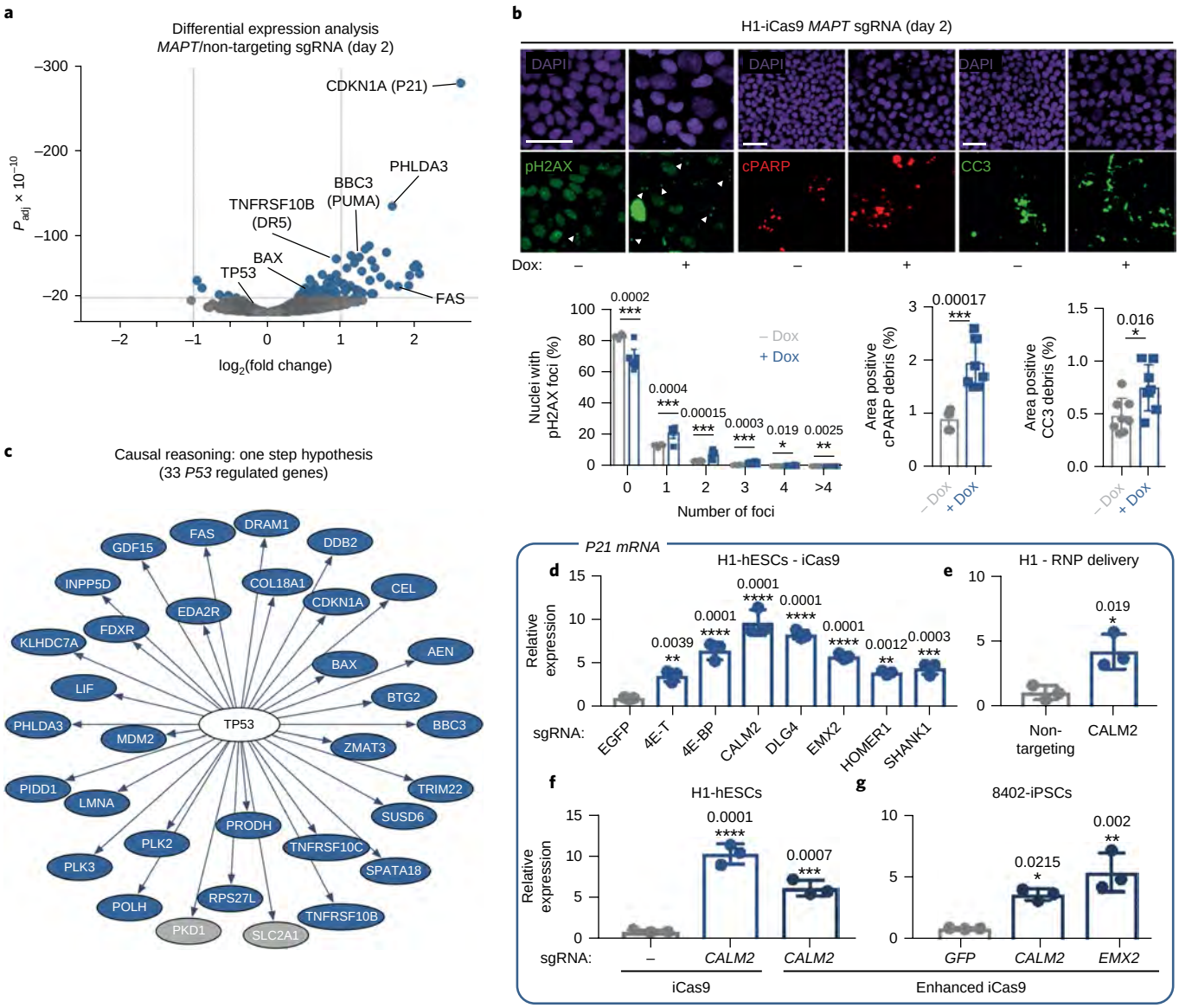


Fig. 3 | Characterization of the Cas9 DSB-induced transcriptional response. a, Volcano plot depicting differential expression in H1-iCas9 after infection with *MAPT* or non-targeting control sgRNAs, as assessed by RNA-seq at 2 days after dox treatment; $n=3$ biologically independent samples per condition. The *y* axis is P_{adj} ($\times 10^{-10}$). The *x* axis is $\log_2$(fold change). We performed a simple two-group comparison Wald test implemented in the DEseq2 package[40]; by default DESeq2 performs a two-sided test. Circles indicate differentially expressed genes (blue circles, $P_{adj} < 1.2 \times 10^{-17}$; grey circles, $P_{adj} > 1.2 \times 10^{-17}$). **b**, High-content image analysis of *MAPT* sgRNA infected H1-iCas9 cells cultured with (+) or without (−) dox. Top, Representative images of cell stained with pH2AX (columns 1 and 2, green), cPARP (columns 3 and 4, red) and CC3 (columns 5 and 6, green). DAPI-stained nuclei are in purple. Bottom, Quantitation of images for the number of nuclei with pH2AX foci, the percent area of cPARP debris, and the percent area of CC3 debris. White triangles indicate pH2AX-positive nuclei with foci. pH2AX: via multiple Welch's unpaired two-tailed *t*-test. cPARP: via Welch's unpaired two-tailed *t*-test. CC3: via Welch's unpaired two-tailed *t*-test. Bars represent average percent and error bars at ±1 standard deviation; $n=8$ wells per condition with 16 images per well. **c**, Interactome analysis identifies P53-dependent changes in expression caused by Cas9-induced DSBs. Upregulated genes are depicted in blue and downregulated genes in grey. **d**–**g**, qPCR quantification of *P21* mRNA in cells treated with DSB-inducing Cas9. The *y* axis is relative expression and each bar represents mean relative expression for sgRNAs targeting the indicated genes; $n=3$ independent mRNA samples per sgRNA; error bars indicate ±1 s.d. One-way ANOVA, equal variances for panels with >2 groups and unpaired two-tailed *t*-test with equal variance for panels with only two groups compared. Specific *P* values are given above each condition. $^*P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$, $^{****}P<0.0001$.

DSB-induced controls, and this is significantly reduced in the *P53* mutant pool (Fig. 4b and Supplementary Fig. 5d). Finally, the toxic response was quantified by measuring confluency during editing in the control and *P53* mutant pool. Dox-treated controls had a strong toxic response, whereas the *P53* mutant pool continued to grow despite DSB induction (Fig. 4c and Supplementary Fig. 5). In addition, 5 of 14 transformed lines had mutations in *P53* and displayed reduced Cas9-induced toxicity (Fig. 2e, bold entries). These results demonstrate that P53 is required for the toxic response to DSBs induced by Cas9.

Inserting a transgene into a specific locus by using Cas9 to stimulate homology directed repair (HDR) is a challenging task in hPSCs[5–13]. We hypothesized that DNA-damage-induced toxicity is in direct opposition of engineering efforts. To determine if P53 inhibits HDR, we developed an assay to measure precise targeting of a transgene into the *OCT4/POU5F1* locus. We used a pair of sgRNAs

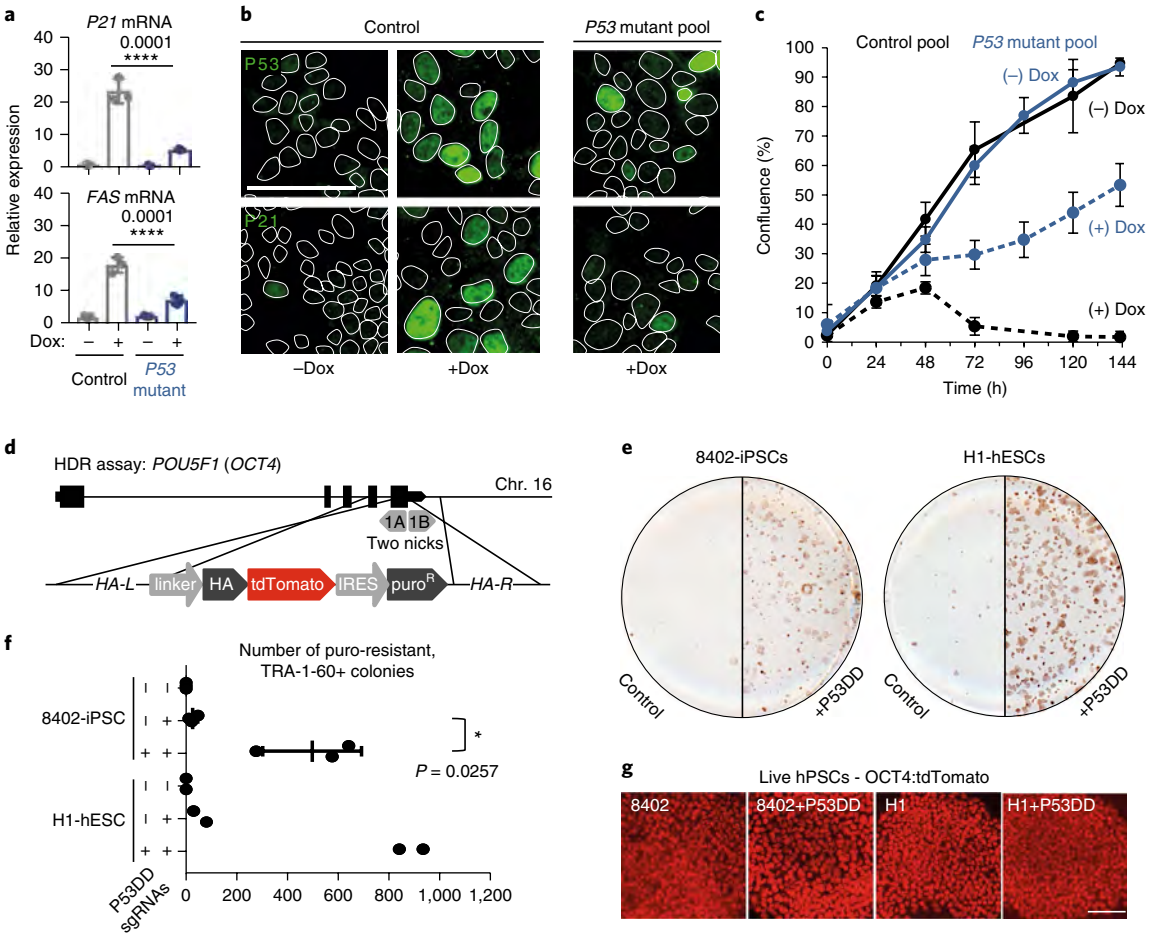


Fig. 4 | P53-dependent toxicity inhibits Cas9 genome engineering in hPSCs. a, qPCR detects an induction of *P21* and *FAS* mRNA in dox-treated controls expressing the *MAPT* sgRNA. Relative expression (*y* axis) is calculated by comparing to untreated control cells. Each bar is mean relative expression. Genotype and dox treatment are labelled on the *x* axis. $n = 3$ independent mRNA samples per condition; error bars indicate ±1 s.d. One-way ANOVA, equal variances. ****$P = 0.0001$. **b**, In control *MAPT* sgRNA infected cells, immunofluorescence staining detects DSB-dependent (+dox) increases in P53 and P21 protein. P53 and P21 are shown in green. DAPI co-stained nuclei are outlined in white. Scale bar, 100 µm Representative images from $n = 4$ independent wells per condition; 16 images per well. Quantification of images is presented in Supplementary Fig. 5d. **c**, Cas9-induced toxic response is *P53*-dependent. Live imaging of confluence in *MAPT* sgRNA-expressing iCas9 cells ± dox in control or *P53* mutant pool. Black lines, control; blue lines, *P53* mutant pool. Solid lines, without dox; dashed lines, with dox. Mean confluency is depicted by coloured circles. Error bars indicate ±1 s.d. $n = 24$ independent wells per condition; 1 image per well. **d**, Schematic of the HDR assay targeting the *OCT4/POU5F1* locus. Nickase and two sgRNAs targeting the stop codon were used to introduce a gene trap fusing an HA tagged tdTomato to the *OCT4* ORF and an internal ribosome entry site (IRES) to drive the expression of the puromycin resistance (puro^R) gene from the *OCT4* promoter. **e**, P53-induced toxicity inhibits the efficiency and yield of HDR in hPSCs. Stem-cell-specific TRA-1-60 antibodies conjugated to HRP were used to visualize colonies surviving puromycin selection following the electroporation of OCT4 donor, dual nickases and ±P53DD plasmid. One out of three representative experiments are shown. **f**, Quantification of number of colonies for independent biological replicates conducted on different weeks in both 8402-iPSCs and H1-hESCs. Unpaired, one-sided Welch's *t*-test with unequal variance: 8402-iPSCs, $n = 3$; H1-hESCs, $n = 2$. Colonies were too large for accurate quantification in a third experiment (see Supplementary Fig. 6 for all data). Mean is represented by a vertical centre line; error bars indicate ±1 s.d. **g**, Live imaging of nuclear OCT4:tdTomato in both control and P53DD treated hPSCs. Representative images from $n = 3$ experiments. Scale bar, 100 µm.

spanning the stop codon and a Cas9^{D10A} nickase to trigger a DSB. The donor plasmid transgene has a gene trap design and does not contain a promoter or nuclear localization signal of its own (Fig. 4d and Supplementary Data 6). As a result, only correctly targeted cells will express a nuclear OCT4-tdTomato fusion protein and gain resistance to puromycin. P53 signalling was transiently blocked by overexpressing a dominant negative P53DD transgene that inhibits the *P53* DSB response and has been routinely used to increase reprogramming efficiency of iPSCs without causing genome instability[29–31]. The Cas9^{D10A}/sgRNA(s) and *OCT4* gene trapping plasmids were co-electroporated with or without the P53DD plasmid and scored for the number of puromycin-resistant colonies expressing nuclear tdTomato (Fig. 4e–g). P53 inhibition greatly increased the number and size of TRA-1-60-positive colonies surviving the engineering and selection process in both 8402-iPSCs and H1-hESCs (Fig. 4e and Supplementary Fig. 6). Multiple independent experiments showed that control 8402-iPSCs and H1-hESCs had an average of 26.3 and 54.5 colonies and that p53DD significantly boosted this average to 500 and 892, respectively (Fig. 4f). P53 inhibition resulted in a 19-fold increase in successful insertions for 8402-iPSCs and a 16-fold increase for H1-hESCs, dramatically improving the efficiency of genome engineering in hPSCs.

Genome engineering of hPSCs using Cas9 has great potential. However, to fully exploit this we need to increase the editing efficiency

and reduce toxicity. We developed a highly efficient Cas9 system in hPSCs that is useful for screening and making engineered cells. We found that DSBs induced by Cas9 triggered a *P53*-dependent toxic response that reduces the efficiency of engineering by at least an order of magnitude.

Several groups have demonstrated that multiple cuts induced by Cas9 causes death in transformed cells[32–35]. In contrast, targeting a single locus is sufficient to kill the majority of hPSCs. hPSCs are most similar to cells of the early embryo, and the extreme sensitivity to DSBs may serve as a mechanism to prevent the development of aberrant cells[36,37]. The heightened *P53*-dependent toxic response provides an explanation for the long-standing observation that hPSCs have inefficient rates of genome engineering. Several studies comparing indel and HDR efficiencies across cell lines identified a 5- to 20-fold reduction in hPSCs relative to transformed lines[3–7]. These results agree with our observation that P53 inhibits HDR efficiency by an average of 17-fold in hPSCs. While long-term P53 inhibition can lead to increased mutational burden[38], transient inhibition is well tolerated in hPSCs [12,29,39]. P53 inhibition may facilitate the generation of large collections of engineered hPSCs by increasing efficiency and reducing variable yields.

The toxic response to Cas9 activity has important implications for hPSC-based therapies. P53 inhibition could alleviate toxicity but has the potential to increase off-target mutations and poses a risk for cancer. For ex vivo engineering, Cas9 toxicity combined with clonal expansion could potentially select for *P53* mutant cells more tolerant of DNA damage. Although the mutation rate of *P53* remains to be determined for other clinically relevant cell types, this is a serious concern for hPSCs. The basal *P53* mutation rate in hESCs is significant, and in ref. [14] it was found that 3.5% of independent hESC lines and up to 29% of hESCs commonly used in RNA-seq databases have *P53* mutations. Before engineering patient cells, the risks and benefits must be fully evaluated. It will be imperative to determine the spontaneous mutation rate of *P53* in engineered cells as well as the mutational burden associated with transient *P53* inhibition. As hPSC-based cell therapies using genome-edited cells move into the clinic, it will be critical to ensure that patient cells have a functional *P53* before and after engineering.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41591-018-0050-6.

Received: 11 September 2017; Accepted: 28 March 2018;
Published online: 11 June 2018

References

1. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–822 (2012).
2. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
3. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
4. Hsu, P. D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
5. He, X. et al. Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Res.* **44**, e85 (2016).
6. Lombardo, A. et al. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat. Biotechnol.* **25**, 1298–1306 (2007).
7. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e04766 (2014).
8. Zwaka, T. P. & Thomson, J. A. Homologous recombination in human embryonic stem cells. *Nat. Biotechnol.* **21**, 319–321 (2003).
9. Hockemeyer, D. et al. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol.* **27**, 851–857 (2009).
10. Liu, Y. & Rao, M. Gene targeting in human pluripotent stem cells. *Methods Mol. Biol.* **767**, 355–367 (2011).
11. Hockemeyer, D. & Jaenisch, R. Induced pluripotent stem cells meet genome editing. *Cell Stem Cell* **18**, 573–586 (2016).
12. Song, H., Chung, S. K. & Xu, Y. Modeling disease in human ESCs using an efficient BAC-based homologous recombination system. *Cell Stem Cell* **6**, 80–89 (2010).
13. Merkle, F. T. et al. Efficient CRISPR-Cas9-mediated generation of knockin human pluripotent stem cells lacking undesired mutations at the targeted locus. *Cell Rep.* **11**, 875–883 (2015).
14. Merkle, F. T. et al. Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations. *Nature* **545**, 229–233 (2017).
15. Avior, Y., Sagi, I. & Benvenisty, N. Pluripotent stem cells in disease modelling and drug discovery. *Nat. Rev. Mol. Cell Biol.* **17**, 170–182 (2016).
16. González, F. et al. An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* **15**, 215–226 (2014).
17. Wells, M. F. et al. Genetic ablation of AXL does not protect human neural progenitor cells and cerebral organoids from Zika virus infection. *Stem Cell* **19**, 703–708 (2016).
18. Liang, X. et al. Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J. Biotechnol.* **208**, 44–53 (2015).
19. Slaymaker, I. M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2015).
20. Banaszynski, L. A., Chen, L., Maynard-Smith, L. A., Ooi, A. G. L. & Wandless, T. J. A. Rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**, 995–1004 (2006).
21. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
22. Chindelevitch, L. et al. Causal reasoning on biological networks: interpreting transcriptional changes. *Bioinformatics* **28**, 1114–1121 (2012).
23. Jaeger, S. et al. Causal network models for predicting compound targets and driving pathways in cancer. *J. Biomol. Screen.* **19**, 791–802 (2014).
24. Lane, D. P. p53, Guardian of the genome. *Nature* **358**, 15–16 (1992).
25. El-Deiry, W. S. et al. *WAF1*, a potential mediator of p53 tumor suppression. *Cell* **75**, 817–825 (1993).
26. Canman, C. E. et al. Activation of the ATM kinase by ionizing radiation and phosphorylation of p53. *Science* **281**, 1677–1679 (1998).
27. Vassilev, L. T. et al. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**, 844–848 (2004).
28. Cazzalini, O., Scovassi, A. I., Savio, M., Stivala, L. A. & Prosperi, E. Multiple roles of the cell cycle inhibitor p21(CDKN1A) in the DNA damage response. *Mutat. Res.* **704**, 12–20 (2010).
29. Schlaeger, T. M. et al. A comparison of non-integrating reprogramming methods. *Nat. Biotechnol.* **33**, 58–63 (2015).
30. Hong, H. et al. Suppression of induced pluripotent stem cell generation by the p53–p21 pathway. *Nature* **460**, 1132–1135 (2009).
31. Hagiyama, H. et al. Signaling through the antigen receptor of B lymphocytes activates a p53-independent pathway of c-Myc-induced apoptosis. *Oncogene* **18**, 4091–4098 (1999).
32. Wang, T. et al. Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
33. Munoz, D. M. et al. CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. *Cancer Discov.* **6**, 900–913 (2016).
34. Aguirre, A. J. et al. Genomic copy number dictates a gene-independent cell response to CRISPR-Cas9 targeting. *Cancer Discov.* **2641**, 617–632 (2016).
35. Hart, T. et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526 (2015).
36. Dumitru, R. et al. Human embryonic stem cells have constitutively active Bax at the golgi and are primed to undergo rapid apoptosis. *Mol. Cell* **46**, 573–583 (2012).
37. Liu, J. C. et al. High mitochondrial priming sensitizes hESCs to DNA-damage-induced apoptosis. *Cell Stem Cell* **13**, 483–491 (2013).
38. Hanel, W. & Moll, U. M. Links between mutant p53 and genomic instability. *J. Cell. Biochem.* **113**, 433–439 (2012).
39. Qin, H. et al. Regulation of apoptosis and differentiation by p53 in human embryonic stem cells. *J. Biol. Chem.* **282**, 5842–5852 (2007).
40. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

Acknowledgements

The authors thank F. Sigoillot for access to the list of sgRNAs with multiple perfect binding sites, M. Morris and A. Hill for help with interactome analysis, and M. Hild for constructive feedback on the project.

Author contributions

R.J.I. and A.K. designed all the experiments and wrote the manuscript. R.D. revised the manuscript. R.J.I. designed iCas9 constructs. R.J.I. and S.K. made transgenic cell lines and characterized them. D.H. and C.Y. developed and performed indel analysis of mutated DNA samples. M.S. packaged the 47 individual sgRNAs, and K.A.W. tested them. K.T. helped with live cell imaging of confluence. E.F., G.R.H. and G.M. helped with the design of the pooled screen, execution and analysis. J.R.-H. generated sgRNA libraries. C.R. sequenced pooled screen samples. G.R.H., G.M., Z.Y. and W.F. provided access and analysed non-targeting control data across transformed cell lines. T.K. identified sgRNAs with SNPs in the H1-hESC genome. J.C. prepped RNA samples for RNA-seq experiments. R.R. performed RNA-seq and interactome analysis. T.K. analysed SNP and multicutter data. M.R.S. conducted high-content image analysis. K.A.W. helped design and performed the *OCT4* HDR assay

Competing interests

All authors were employees of Novartis Institutes for Biomedical Research when the research was conducted.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41591-018-0050-6.

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

Correspondence and requests for materials should be addressed to A.K.

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

Methods

hPSC cell culture. Cells were grown in TeSR-E8 medium (Stemcell Tech., 05940) on tissue-culture plates coated with vitronectin (Gibco, A14700). Passaging for maintenance was performed using ReLeSR (Stemcell Tech., 05873) to dissociate cell clumps to be replated in E8 plus thiazovivin (Selleckchem, S1459) at 0.2 µM. For lentiviral transduction, electroporation, pooled screening and live imaging of confluence, accutase (Gibco, A1110501) was used to create a single cell suspension, which was counted to accurately replate speciàc numbers of cells in E8 plus thiazovivin at 0.8 µM. Karyotyping was performed by Cell Line Genetics. H1-hESCs (WA01-NIHhESC-10-0043) were obtained from WiCell. 8402-iPSCs originated from GW08402 åbroblasts from the Coriell Institute and reprogrammed as described in ref. [41]. hPSC lines were free of myoplasma and were tested using the Mycoalert detection kit (Lonza). SNP ångerprinting conàrmed the identify of hPSC lines used.

Inducible Cas9 cell line generation. Inducible Cas9 plasmids used in this study were synthesized by GenScript, and sequences are available in Supplementary Data 1 and 2. The AAVS1 TALE-Nuclease KIT(GE60xA-1) was obtained from System Biosciences (SBI). The iCas9 plasmid with a neomycin resistance gene was knocked in to the AAVS1 locus of H1-hESCs by co-electroporation with AAVS1 targeting TALENs using the NEON system (ThermoFisher). One million cells with 1.5 µg for each TALEN plasmid and 4 µg for the donor plasmid were used for each electroporation at 1,050 V 30 ms (two pulses). After G418 selection (200 µg ml^{-1}), clones were picked, expanded and screened by treating with dox for 24 h and staining for FLAG-tagged Cas9. Clones with strong expression of Cas9 were subsequently banked, karyotyped and tested for proper targeting using junction PCR. KOD Xtreme Hot Start DNA Polymerase (Millipore-71975) was used with the step-down cycling protocol as recommend by the manufacturer for junction PCR. Primers are listed in Supplementary Table 3. This strategy is generally adaptable to hPSCs, and stable iCas9 lines using additional selectable markers (blasticidin, zeomycin) have been generated in multiple hPSCs backgrounds H1/H9: human embryonic stem cells (hESCs) from WiCell and 8402[42] and HDFn (control)[43] induced pluripotent stem cells (iPSCs). ddCas9 (Supplementary Data 3) in a piggyBAC vector was electroporated using 4 µg of the piggyBAC donor and 1 µg of the piggyBAC dual helper plasmid. G418 resistant clones were selected by anti-FLAG stain and karyotyped. iCas9 and ddCas9 cell lines were maintained in media containing 200 µg ml^{-1} G418 (Millipore, 345812) to ensure proper transgene expression.

Lentiviral delivery of sgRNAs for Cas9 mutagenesis. During replating, lentiCRISPRs were added to a single-cell suspension of 200,000 cells in 2 ml E8 (Stemcell Tech., 05940) with 0.8 µM thiazovivin (Selleckchem, S1459). After 24 h, cells were maintained in 2 µg ml^{-1} puromycin (Corning, 61-385-RA). At the onset of each mutagenesis experiment, Shield1 (Clontech, 631037, 0.5 µM) or dox (Clontech, 631311, 2 µg ml^{-1}) was added to induce Cas9. The 47 sgRNAs in Fig. 1 were designed using the sgRNA Designer (Broad Institute) (Supplementary Table 4) and cloned into the pNGx_LV_g003_HA_Puro^{R} backbone by GenScript. The 13,000 lentiviral sgRNA library was designed, cloned into the pNGx_LV_g003_TagRFP_T2A_Puro^{R} backbone and packaged as described in ref. [44]. For pooled screening, viral titre was determined by exposing cells to a 12-point dose response of each lentiviral stock. A total of 200,000 cells were plated into a single well of a six-well plate (21,000 cells per cm^2). Four days after infection, percent red fluorescent protein (RFP) was assayed by FACS (SONY SH800Z), and the data were used to calculate the amount of virus needed for 0.5 MOI. Puromycin concentration was optimized by infecting at 0.5 MOI and testing a dose–response of puromycin spanning 0.3 µg ml^{-1} to 2 µg ml^{-1}. At 2 µg ml^{-1} puromycin, 100% of surviving cells are RFP-positive.

Cas9/sgRNA RNP complex delivery. A 1 µl volume of 61 µM Cas9 protein (IDT, 1074182) was complexed with 1 µl of 100 µM full-length synthetic modified sgRNAs (Synthego) for 5 min at room temperature. Following incubation, RNP complexes were diluted with 100 µl R buffer. Diluted RNPs were mixed with one million pelleted hPSCs and electroporated at 1,100 V for 2 20 ms pulses using the Neon electroporation system (Thermo)[45].

Lipid delivery of synthetic crRNA/tracrRNAs for Cas9 mutagenesis. iCas9 cells were treated with dox for 24 h before transfection of synthetic CRISPR RNA (crRNA)/trans-activating crRNA (tracrRNA) pairs. Cells were dissociated with accutase and replated at a density of 200,000 cells per well in a six-well plate in 2 ml E8 plus thiazovivin. The amount of tracrRNA(90 nM)/crRNA(30 nM) used was calculated by referring to a final concentration diluted in 2 ml of stem cell medium (one well of a six-well plate). tracrRNAs/crRNAs were synthesized by IDT and resuspended in H_2O at 100 µM. tracrRNA (3×) and three crRNAs (1× each) targeting *P53* were incubated together in H_2O for 5 min at room temperature (RT). The crRNA/tracrRNA mixture was then diluted in 100 µl Opti-MEM (ThermoFisher, 31985088) and incubated for 5 min at RT. In parallel 6 µl of RNAimax (ThermoFisher, 13778150) was diluted in 100 µl Opti-MEM for 5 min at RT. Each tube was mixed and incubated for 15 min at RT. A 200 µl volume of the RNAimax/crRNA/tracrRNA/Opti-MEM was added dropwise to a well of a six-well plate with freshly seeded iCas9 cells pretreated with dox. Cells were maintained in E8 medium with dox for 3 days following transfection.

CRISPR indel analysis from genomic DNA. The DNA template for PCR during library construction can be either cell lysate or genomic DNA purified using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, 69506) following the manufacturer's protocol. For direct lysis, cells grown in a 96-well plate (Fig. 1b) were washed once with 100 µl 1× PBS. Following removal of PBS, 40 µl lysis buffer was added (10 mM Tris-HCl, 0.05% SDS, 25 µg ml^{-1} proteinase K) to the cells, then incubated at 37 °C for an hour, followed by 85 °C proteinase inactivation for 15 min. The lysate was directly used as a template for subsequent PCR. Each target was amplified using locus-specific primers (Supplementary Table 3) followed by a second amplification with Illumina index containing primers. Libraries were quantified and sequenced on the Illumina MiSeq. For sequence analysis, raw reads were aligned to a reference sequence and binned into one of three categories: control, in-frame and frameshift.

Live imaging of confluency. An IncuCyte zoom (Essen Biosciences) was used to quantify confluency in live cells each day post medium change. The confluence processing analysis tool (IncuCyte Zoom Software v2016A) calculated confluency for each sample. Infected cells were allowed to recover after puromycin selection and were maintained in the absence of dox so as to not induce mutations prior to the growth assay. For each population of sgRNAs, cells were counted and plated in medium with or without dox at a density of 21,000 cells per cm^2 at the start of the experiment (day 0).

Pooled screening. The 13,000 sgRNA sub-genome library, included sgRNAs targeting 2,600 genes (5 sgRNAs per gene) and non-targeting controls, was designed, synthesized, cloned and packaged as described in ref. [44]. To infect at 1,000 cells per sgRNA, five T225 flasks were seeded at 21,000 cells per cm^2 and infected at 0.5 MOI for each condition. Two independent replicates were maintained for each condition. At 24 h after infection, cells were treated with 0.5 µg ml^{-1} puromycin for the remainder of the screen. Dox and Shield1 were added to the Cas9-positive conditions from day 1 through day 12. At each passage, cells were counted to maintain 1,000 cells per sgRNA for both the newly seeded flask and the pellet for DNA isolation. To generate $\log_2$(fold change) values, DNA was isolated from pelleted cells and PCR-amplified with primers targeting the lentiviral sgRNA backbone. Next-generation library construction, sequencing and data analysis were performed as described in ref. [44]. Non-hPSCs pooled screening data are available but restricted to non-targeting sgRNA sequences.

RNA-seq and qPCR. To detect the signal from dying cells, samples were collected by pelleting both the cellular debris in the media as well as the dissociated, formerly adherent, cells from an entire well per replicate in the same microcentrifuge tube. Total mRNA was isolated using the RNeasy Mini kit plus (Qiagen, 74134).

An Agilent 2100 bioanalyser and Nano 6000 kit (Agilent, 5067-1511) were used to quantify and check the quality of each mRNA sample. High-quality RNA (RIN 10, 240 ng) was used for PolyA + RNA-seq. Libraries were made using a Hamilton automated protocol with the TruSeq Stranded mRNA LT sample prep kit (Illumina, RS-122-2101) and sequenced on the Illumina HiSeq 2500. An average of more than 50 million 76 bp paired-end reads were obtained per sample. Processing was conducted using open-source software. Raw fastq files were aligned to a human reference genome (GRCh37.74) using the STAR aligner (v2.5.1b)[46]. Gene counts and transcript quantification values (TPM) was performed using HTSeq-count (v0.6.0)[47] and RSEM (v1.2.28)[48] respectively. The gene counts were then used for differential expression analysis using DESeq2[40]. Ninety-three percent of variance was explained with principal component analysis, and confirmed samples have similar variance.

For qPCR, mRNA concentration was measured using a Nanodrop 2000 (Thermo Scientific). RNA (200 µg) was used as a template for cDNA synthesis using the SuperScript III first-strand synthesis system (ThermoFisher, 18080051). cDNA was diluted 1:5 in H_2O before analysis using taqman gene expression arrays and the 2× Fast Start Universal Probe master mix (ROX) (Roche, 04913957001). qPCR plates (384-well) were run on a ViiA 7 Real-Time PCR System (ThermoFisher). Relative expression was calculated using the ΔΔCT method as described in ref. [49], and bACTIN was used as the reference gene. TaqMan gene expression arrays FAM-MGB (ThermoFisher, 4331182): *P21/CDKN1A* (Hs00355782_m1), *bACTIN* (Hs01060665_g1) and *FAS* (Hs00163653_m1). A custom TaqMan gene expression assay was ordered to detect Cas9 mRNA.

Interactome analysis. The Clarivate Analytics (previously Thomson Reuters) Computational Biology Methods for Drug Discovery (CBDD) toolkit implements several published algorithms (in R) for network and pathway analysis of -omics data. An internal R wrapper functioned to facilitate the use of the CBDD toolkit to run causal reasoning algorithms[50,51]. The knowledge base used was a combination of a MetaBase (a manually curated commercial database of mammalian biology from Clarivate Analytics) and STRING[52].

Immunofluorescence and microscopy. Cells were fixed in 4% PFA in PBS for 10 min at room temperature and washed with 0.1% triton X-100 in PBS after

fixation. Cells were blocked in 2% goat serum, 0.01% BSA and 0.1% triton X-100 in PBS for 1 h at room temperature. Primary antibodies were diluted in blocking solution and incubated with cells overnight at 4 °C. Cells were washed three times before incubation with secondary antibodies or fluorescently conjugated primary antibodies at RT for 1.5 h. Cell were washed three times and incubated with DAPI 1:1,000 for 5 min at RT before imaging. Primary antibodies: 1:250 P21 (12D1) (CST-2947), 1:250 P53 (7F5) (CST-2527), 1:300 FLAG (M2) (Sigma-F1804), 1:200 CC3 (Asp175) (CST-9661), 1:100 phospho-histone H2A.X (Ser139/Y142) (CST-5438), 1:50 cleaved PARP-647 (Asp214) (D64E10) (CST-6987). Secondary antibodies: 1:500 goat anti-mouse IgG (H+L) AF488 (ThermoFisher, A-11029), 1:500 goat anti-rabbit IgG (H+L) AF488 conjugate (ThermoFisher, A-11008). For OCT4 targeting assay, live cells were imaged for tdTomato fluorescence and then fixed, permeabilized, washed, incubated with peroxidase suppressor (Thermo) for 30 min, washed twice, and then blocked for 30 min (5% goat serum/0.1% Tween-20/PBS). Cells were incubated at 37 °C for 2 h with anti-TRA-1-60 (MAB4360, Millipore, 1:300 dilution), washed three times, and then for 1 h with anti-IgM conjugated to Horse Radish Peroxidase (HRP) (31440, Thermo, 1:250). A metal enhanced DAB substrate kit was used for detection (34065, Thermo). Live and fixed immunofluorescent images were taken using the ×10 and ×20 objectives on an Axio Observer.D1 (Ziess). Images for high content analysis were taken on an Incell 6000 (GE Healthcare Life Sciences). P53, P21, H2AX, cPARP and CC3 immunoflurescence quantification was conducted using Cell Profiler (version 2.1.1, revision 6c2d896) software. For P53 and P21 proteins, average immunofluorescent intensity was determined for each nucleus, and a positive-expression threshold was set based on the no-secondary control. To quantify H2AX foci, the number of individual foci were detected within each nucleus via CellProfiler's object detection module. To quantify cPARP and CC3, positive regions were detected via thresholding, and the area of this region was normalized to total plate area covered by colonies.

FACS. Cells were dissociated using accutase for 10 min at 37 °C to create a single-cell suspension that was subsequently fixed in 4% PFA in PBS for 10 min at room temperature on a rocker. Cells were spun down at 300 relative centrifugal force (RCF) for 3 min between each subsequent solution change. Cells were washed with 0.1% Triton-X in PBS after fixation and blocked in 2% goat serum, 0.01% BSA and 0.1% Triton-X in PBS for 1 h at RT. Conjugated primary antibodies were diluted in blocking solution and incubated with cells on a rocker overnight at 4 °C: 1:50 FITC-conjugated anti-TRA-1-60 antibody (FCMAB115F, Millipore), 1:50 647 conjugated anti-OCT4 (C30A3) antibody (5263, CST) and 1:50 647 conjugated anti-Sox2 (D6D9) antibody (5067, CST). Cells were washed and resuspended in PBS and transferred to a 5 ml flow cytometry tube with strainer cap before FACS analysis on a Sony SH800Z. iCas9 cells were infected with lentiCRISPRs targeting *MAPT*, *OCT4* and *SOX2*, and were cultured for 8 days in the presence of dox before FACS analysis.

OCT4 targeting assay. We designed a pair of $Cas9^{D10A}$ nickases[53] 36 bp apart spanning the stop codon to trigger a DSB with 5′ overhangs. The donor plasmid contained a transgene flanked by two ~700 bp homology arms (Fig. 4d and Supplementary Data 6). Before genome engineering, hPSCs were pretreated with 1 μM thiazovivin for at least 2 h and collected using accutase. A mixture of 4 μg of OCT4-tdTomato-puro^R targeting vector (Supplementary Data 6), 1 μg of each sgRNA cloned into a vector that co-expresses Cas9-D10A (or a vector lacking gRNAs as a control) and 2 μg of either an episomal vector for P53DD (pCE-mP53DD, Epi5 Episomal iPSC Reprogramming Kit cat. no. A15960, addgene no. 41856) or EBNA1 alone (pCXB-EBNA1, Epi5 kit: cat. no. A15960, addgene no. 41857) were electroporated into 1×10^6 cells using a Neon electroporation system (Thermo). Cells were deposited into one well of a six-well dish coated with matrigel containing 50% fresh mTESR:50% conditioned mTESR supplemented with bFGF (10 ng ml^{-1}) and thiazovivin. After 48 h, cells were selected with 0.3 μg ml^{-1} puromycin in the presence of thiazovivin.

Statistics. *qPCR.* Statistical analysis of relative expression values generated by qPCR for *P21* and *FAS* mRNAs was conducted using PRISM software (version 7.0c). Relative expression values for control and treated samples were tested for statistical differences by conducting an ordinary one-way analysis of variance (ANOVA) test with equal variances for analysis with more than two groups for multiple comparisons to a control group. An unpaired two-tailed *t*-test with equal variance was performed for analysis with two groups. $n = 3$ for control, and treated mRNA samples isolated from three independent wells are plotted as individual dots. Bars indicate mean and error bars depict standard deviation. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$. Figures 3d–g and 4a replicate the induction of *P21* mRNA observed in RNA-seq data in Fig. 3a.

For Fig. 3d, *P21* mRNA, ANOVA ($F = 40.27$, degrees of freedom (d.f.) = 7, $P < 0.0001$).

Dunnett's multiple comparisons test:
GFP versus 4E-T, $P = 0.0039$
GFP versus 4E-BP, $P = 0.0001$
GFP versus CALM2, $P = 0.0001$
GFP versus DLG4, $P = 0.0001$
GFP versus EMX2, $P = 0.0001$
GFP versus HOMER1, $P = 0.0012$
GFP versus SHANK1, $P = 0.0003$

For Fig. 3e, *P21* mRNA, unpaired two-tailed *t*-test, equal variance ($t = 3.783$, d.f. = 4, $P = 0.0194$).

For Fig. 3f, *P21* mRNA, ANOVA ($F = 83.22$, d.f. = 2, $P < 0.0001$).
Dunnett's multiple comparisons test:
iCas9 control versus iCas9 CALM2, $P = 0.0001$
iCas9 control versus ieCas9 CALM2, $P = 0.0007$

For Fig. 3g, *P21* mRNA, ANOVA ($F = 17.46$, d.f. = 2, $P = 0.0032$).
Dunnett's multiple comparisons test:
GFP versus CALM2, $P = 0.0215$
GFP versus EMX2, $P = 0.0020$

For Fig. 4a, *P21* mRNA, ANOVA ($F = 100.8$, d.f. = 3, $P < 0.0001$).
Dunnett's multiple comparisons test:
Control (+) versus control (−), $P = 0.0001$
Control (+) versus mutant (−), $P = 0.0001$
Control (+) versus mutant (+), $P = 0.0001$

For Fig. 4a, *FAS* mRNA, ANOVA ($F = 83.55$, d.f. = 3, $P < 0.0001$).
Dunnett's multiple comparisons test:
Control (+) versus control (−), $P = 0.0001$
Control (+) versus mutant (−), $P = 0.0001$
Control (+) versus mutant (+), $P = 0.0001$

High-content imaging. Statistical analysis of high-content imaging data quantified using Cell Profiler (version 2.1.1, revision 6c2d896) was conducted using PRISM software (version 7.0c). For cPARP, CC3 and pH2AX proteins an unpaired two-tailed Welch's *t*-test with unequal variances was performed. For P21 and P53 proteins an ordinary one-way ANOVA test with equal variance was performed. Images were taken from independently stained wells on the same plate ($n = 4$ wells for P21/P53 and $n = 8$ wells for cPARP, CC3 and pH2AX). Each well was plotted as an individual dot. Sixteen images were taken per well. Bars indicate mean and error bars depict standard deviation. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

For Fig. 3b, pH2AX, unpaired two-tailed Welch's *t*-test, unequal variance:
0 foci ($t = 6.55$, d.f. = 7.53, $P = 0.00023$)
1 focus ($t = 6.044$, d.f. = 7.55, $P = 0.00038$)
2 foci ($t = 6.99$, d.f. = 7.57, $P = 00015$)
3 foci ($t = 6.00$, d.f. = 8.06, $P = 0.0003$)
4 foci ($t = 2.89$, d.f. = 8.29, $P = 0.0194$)
5 foci ($t = 4.04$, d.f. = 9.74, $P = 0.0025$)

For Fig. 3b, cPARP, unpaired two-tailed Welch's *t*-test, unequal variance ($t = 6.17$, d.f. = 8.93, $P = 0.00017$).

For Fig. 3b, CC3, unpaired two-tailed Welch's *t*-test, unequal variance ($t = 2.76$, d.f. = 12.95, $P = 0.016$).

For Supplementary Fig. 5d, P21, ANOVA ($F = 24.27$, d.f. = 3, $P < 0.000022$).
Tukey's multiple comparisons test:
wt − dox versus wt + dox, $P = 0.00003$
wt − dox versus p53−/− pool − dox, $P = 0.5584$
wt − dox versus p53−/− pool + dox, $P = 0.7352$
wt + dox versus p53−/− pool − dox, $P = 0.00019$
wt + dox versus p53−/− pool + dox, $P = 0.00012$
p53−/− pool − dox versus p53−/− pool + dox, $P = 0.9895$

For Supplementary Fig. 5d, P53, ANOVA ($F = 42.64$, d.f. = 3, $P < 0.000001$.
Tukey's multiple comparisons test:
wt − dox versus wt + dox, $P = 0.000007$
wt − dox versus p53−/− pool − dox, $P = 0.4181$
wt − dox versus p53−/− pool + dox, $P = 0.4013$
wt + dox versus p53−/− pool − dox, $P = 0.000001$
wt + dox versus p53−/− pool + dox, $P = 0.000053$
p53−/− pool − dox versus p53−/− pool + dox, $P = 0.0324$

HDR assay. Statistical analysis of hPSC colonies undergoing homology-directed repair with an OCT4 gene trapping vector was performed using PRISM software (version 7.0c). An unpaired, one-sided Welch's *t*-test with unequal variance was performed. TRA-1-60-positive colonies from each independent electroporation were counted in six-well plates. Data were replicated in two different cell lines by conducting independent experiments on three consecutive weeks. Each electroporation is represented by individual dots. 8402-iPSCs $n = 3$; H1-hESCs is $n = 2$ (colonies were too large for accurate colony quantification in a third experiment; see Supplementary Fig. 6). To calculate the percent confluency, images were filtered via Gaussian smoothing, then normalized and threshold filtered via MATLAB R2017a ($n = 3$ for 8402-iPSCs and H1-hESCs)

Mean is represented by a vertical centre line. Error bars represent standard deviation. $^*P < 0.05$.

For Fig. 4f: 8402-iPSCs, unpaired two-tailed Welch's *t*-test, unequal variance ($t = 4.17$, d.f. = 2.037, $P = 0.0257$).

For Supplementary Fig. 6: 8402-iPSCs unpaired two-tailed Welch's *t*-test, unequal variance ($t = 3.114$, d.f. = 2.008, $P = 0.0445$); H1-hESCs, unpaired two-tailed Welch's *t*-test, unequal variance ($t = 3.869$, d.f. = 2.019, $P = 0.0299$).

RNA expression analysis. Differential gene expression analyses were performed in R with DESeq2, which is suitable for studies with fewer replicates ($n = 3$), and consistently shows high sensitivity and precision[40]. Gene counts were normalized for library size differences using the geometric mean, and modelled with a negative binomial distribution. Log_2(fold change) and *P* values for each gene are provided in Supplementary Data 5. RNA was isolated from three independent wells of a six-well plate for both control and DSB-treated samples ($n = 3$). Results for *P21* mRNA were replicated in independently isolated samples with nine different sgRNAs in Figs. 3d–g and 4a. Results for *FAS* mRNA were replicated in Fig. 4a.

The top 100 differential expressed genes had an adjusted *P* value of $<1.2 \times 10^{-17}$ and were used as input for interactome analysis. Causal reasoning algorithms highly ranked MYC (rank = 1, SigNet score = 29.32), SP1 (rank = 1, SigNet score = 30.81), EP300 (rank = 3, SigNet score = 24.03) and P53 (rank = 9, SigNet score = 23.87) as hypotheses explaining the top 100 differentially expressed genes. MYC, SP1 and EP300 hypotheses are connected and include P53. The P53 results from the interactome analysis were functionally validated in Fig. 4a–c, which tested *P53* mutant cells.

SNPs and multi-cutter sgRNA representation. Log_2(fold change) for 249 sgRNAs affected by SNPs in the H1-hESC genome and 142 sgRNAs with one or more perfect cut sites were compared between Cas9-negative and Cas9-positive conditions. Two-tailed, unpaired Welch's *t*-tests with unequal variance were performed using the R open-source package (version 3.3.0). $n = 2$ independent DNA samples sequenced for each condition. ddCas9 + Shield1 and iCas9 + dox (Cas9+) versus H1 parental and iCas9 minus dox (Cas9−). $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$; NS, not significant.

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

Code availability. RNA-seq data processing and pooled CRISPR NGS were conducted using open-source software (see Supplementary Methods). Interactome analysis was performed using the commercially available Computational Biology Methods for Drug Discovery (CBDD) toolkit provided by Clarivate Analytics (previously Thomson Reuters). CBDD provides both data and algorithms, which were accessed using an internal wrapper in R.

Data availability. Raw data and results generated for hPSCs are available upon request. Supplementary Data 1 includes log_2(fold change) of sgRNAs for the pooled CRISPR screen in hPSCs presented in Fig. 1c,d. CRISPR screening data generated for cancer cell lines are restricted to non-targeting sgRNAs. Supplementary Data 2 includes differential expression analysis of mRNA for the DSB-induced hPSCs depicted in Fig. 3a.

References

41. Sun, Y. et al. A deleterious Nav1.1 mutation selectively impairs telencephalic inhibitory neurons derived from Dravet syndrome patients. *eLife* **5**, e13073 (2016).
42. Wells, M. F. et al. Genetic ablation of AXL does not protect human neural progenitor cells and cerebral organoids from Zika virus infection. *Stem Cell* **19**, 703–708 (2016).
43. Bidinosti, M. et al. CLK2 inhibition ameliorates autistic features associated with SHANK3 deficiency. *Science* **351**, 1199–1203 (2016).
44. Dejesus, R. et al. Functional CRISPR screening identifies the ufmylation pathway as a regulator of SQSTM1/p62. *eLife* **5**, e17290 (2016).
45. Liang, X. et al. Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J. Biotechnol.* **208**, 44–53 (2015).
46. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
47. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data HTSeq. *Bioinformatics* **31**, 166–169 (2015).
48. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
49. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, 16–21 (2001).
50. Chindelevitch, L. et al. Causal reasoning on biological networks: interpreting transcriptional changes. *Bioinformatics* **28**, 1114–1121 (2012).
51. Jaeger, S. et al. Causal network models for predicting compound targets and driving pathways in cancer. *J. Biomol. Screen.* **19**, 791–802 (2014).
52. Szklarczyk, D. et al. STRINGv10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).
53. Ran, F. A. et al. Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).

nature research

Corresponding author(s): Ajamete Kaykas

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

Specific samples sizes for each experiment are indicated in figure legends. n=3 was chosen because it is sufficient to conduct t-tests and ANOVA to generate p-values to determine if results are significant.

For qPCR 3 independent samples isolated from 3 independent wells. Three independent mRNA samples (n=3) were used to calculate the mean crossing point for each condition as outlined by Pfaffl et. al., 2001. DESeq2 was used to analyze RNA-seq and is designed to work with at least 3 replicates per condition (n=3) (Love et al., 2014). High-content imaging of immunofluorescent hPSCs was conducted on n= 4 to 8 stained wells, 16 images per well. Quantification of live imaging and was performed on n=3 to 96 wells with 1 to 121 images per well dependent on plate type.

Pooled CRISPR screening used two independent replicates with high coverage (1000 cells per sgRNA). This 1000x coverage per sgRNA provides high sensitivity to detect small changes in fitness in an internally controlled setting. The scale of pooled screen is large and stem cell media is prohibitively expensive and only two replicates per condition were possible.

CRISPR indels quantified by NGS >10,000 reads per sample (n=1) for for a total 72 independent sgRNAs. This demonstrates a high on-target mutation rate across irrespective of sgRNAs sequence. High editing efficiencies with MAPT lentiCRISPR were replicated in 1D, S2A, and S5E.

Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).

Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29, 16–21 (2001).

2. Data exclusions

Describe any data exclusions.

No data was excluded

3. Replication

Describe whether the experimental findings were reliably reproduced.

Critical experiments were independently replicated at later times with different reagents. Figure legends indicate number of times replicated.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Sample randomization is not applicable for in vitro cell culture experiments where large batches of homogeneous cultures can be easily tested in parallel. In this setting it is difficult to plate down and accurately maintain randomized samples in tissue culture plates. We went to great lengths to minimize the potential effects of a particular reagent such as a sgRNA, Cas9 variant/expression system, or genetic background.

In figure 2, 13 million cells were infected (.5 MOI) with 13,000 sgRNAs (1000

independent cells per sgRNA). This allowed us to globally quantify the change in fitness caused by a double strand breaks, irrespectively of targeting sequence, relative to controls cells with non-targeting sgRNAs.

Cas9 was expressed by several different methods; stable shield or dox inducible cell lines, transient DNA expression plasmids, and transient RNPs. In addition enhanced Cas9 variants with low off target effects were also tested.

Lastly, phenotypic, molecular and genome engineering assays were used in two distinct pluripotent stem cell backgrounds. H1-hESCs (human embryonic stem cells) 8402-iPSCs (induced pluripotent stem cells)

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

All computational analysis was conducted by bioinformaticians not involved the design of the study. In this case they had no preconceptions about the data and were therefore blind to the outcome. This types of analysis applies to pooled CRISPR analysis, interactome analysis, RNA-seq and high-content image analysis.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Open source code was used for pooled CRISPR screening NGS (Bowtie v1.0.1-Langmead et al., 2009, DESeq2 v1.10.1 - Love et al., 2014).

Open source code was used for RNA-seq (STAR aligner v2.5.1b, HTSeq-count v0.6.0, RSEM v1.2.28, DESeq2 V 1.16.1 - Love et al., 2014).

Graphs were generated in Spotfire TIBCO 6.5.4.6, PRISM (version 7.0c) and Microsoft Excel v 15.41.

Interactome (CBDD version 5) analysis is commercially available through Clarivate Analytics and was accessed using an internal wrapper.

IncuCyte Zoom Software v2016A was used to quantify confluency.

Cell Profiler (Version: 2.1.1 revision 6c2d896) was used for high content image analysis.

SONY SH800Z v2.1.3 software was used use for FACS.

To calculate the percent confluency (for p53DD), the images were filtered via Gaussian smoothing, then normalized and threshold filtered via MATLAB R2017a.

PRISM (version 7.0c) and R open source package (version 3.3.0) was used for statistical analysis.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No unique materials.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

Primary antibodies used in human pluripotent stem cells:

P21 (12D1) (CST-2947) lot 9
1:250 dilution – Induced by Cas9 DDR Fig. 3B, Reduced in P53 mutant* with Fig. 4B, S5D
P53 (7F5) (CST-2527) lot 8
1:250 dilution – Induced by Cas9 DDR Fig. 3B, Reduced in P53 mutant* with Fig. 4B, S5D
Cleaved caspase-3 (Asp175) (CST-9661) lot 43
1:200 dilution - Induced by Cas9 DDR Fig. 3B
Phospho-histone H2A.X (Ser139/Y142) (CST-5438) lot 1
1:100 dilution - Induced by Cas9 DDR Fig. 3B
Cleaved PARP-647 (Asp214) (D64E10) (CST-6987) lot 7
1:50 dilution - Induced by Cas9 DDR Fig. 3B
anti-TRA-1-60 (Millipore-MAB4360) 2723570
1:300 dilution – Stains hPSCs Fig 4E
FLAG (M2) (Sigma-F1804)
1:300 dilution – Stains FLAG tagged Cas9 Fig S1C
anti-TRA-1-60 antibody FITC conjugated (Millipore-FCMAB115F) lot 2664435
1:50 dilution – Reduced in OCT4 and SOX2 mutants* Fig S1F
anti-OCT4 (C30A3) antibody 647 conjugated (CST-5263) lot 3
1:50 dilution - Reduced in OCT4 mutants* Fig S1F
anti-SOX2 (D6D9) antibody 647 conjugated (CST-5067) lot 5
1:50 dilution - Reduced in SOX2 mutants* Fig S1F

sgRNA used to validate antibodies*
P53 gRNA1 GAAGGGACAGAAGATGACAG
P53 gRNA2 GAAGGGACAGAAGATGACAG
P53 gRNA4 GAGCGCTGCTCAGATAGCGA
POU5F1/OCT4 gRNA1 CAACAATGAAAATCTTCAGG
SOX2 gRNA2 CGTTCATCGACGAGGCTAAG

Antibodies were used from Sigma, ThermoFisher, Cell Signaling Technologies (CST) and Millipore. The antibodies are commonly used and catalogs numbers can be used to look up additional validation experiments.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

H1-hESCs - WiCell, 8402-iPSCs originated from GW08402 fibroblasts - Coriell

b. Describe the method of cell line authentication used.

H1-hESCs and 8402 iPSCs were authenticated by SNP fingerprinting, Karyotype analysis and staining for pluripotent stem cell markers.

c. Report whether the cell lines were tested for mycoplasma contamination.

hPSC lines were free of Myoplasma and tested using the Mycoalert Detection kit (Lonza).

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

none

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

No animals used in this study

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

This study did not involve human research participants

nature research

Corresponding author(s): Ajamete Kaykas

☐ Initial submission ☒ Revised version ☑ Final submission

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

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

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

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

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

Methodological details

5. Describe the sample preparation.	human embryonic stem cells, H1-hESCs (WiCell)
6. Identify the instrument used for data collection.	Sony SH800Z cell sorter
7. Describe the software used to collect and analyze the flow cytometry data.	Sony SH800Z cell sorter software v2.12
8. Describe the abundance of the relevant cell populations within post-sort fractions.	N/A - samples were analyzed only and not sorted
9. Describe the gating strategy used.	Gates were based on the positive control infected with a sgRNA targeting MAPT, a neuronal non-essential gene not expressed in stem cells. Boundaries were set to capture double positive cells for both TRA-1-60/OCT4 and TRA-1-60/SOX2.

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