CRISPR/Cas9 has revolutionized our ability to engineer genomes and conduct genome-wide screens in human cells. 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. 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. The toxic response to DSBs was TP53-dependent, such that the efficiency of precise genome engineering in hPSCs with a wild-type TP53 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 TP53 mutations, cell replacement therapies using CRISPR/Cas9-engineered hPSCs should proceed with caution, and such engineered hPSCs should be monitored for TP53 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. Developing a practical system for high-throughput genetic engineering of hPSCs has been challenging because hPSCs are recalcitrant to genome modification. Several studies have shown that gene targeting in hPSCs is five- to twentyfold lower in efficiency relative to other cell types. 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. 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 adenovirus-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). The stable iCas9 lines are properly targeted, have a normal karyotype, and induce Cas9 only with dox (Supplementary Fig. 1b–e).

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 at predicted sites (Supplementary Fig. 2a and Supplementary Data 1). 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 and further supports the hypothesis that DSBs

---

p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells

Robert J. Ihry1, Kathleen A. Worringer1, Max R. Salick1, Elizabeth Frias2, Daniel Ho1, Kraig Theriault1, Sravya Kommineni1, Julie Chen1, Marie Sondey4, Chaoyang Ye5, Ranjit Randhawa1, Tripti Kulkarni1, Zinger Yang2, Gregory McAllister2, Carsten Russ2, John Reece-Hoyes2, William Forrester2, Gregory R. Hoffman2, Ricardo Dolmetsch1 and Ajamete Kaykas1*
at a single locus are sufficient to cause toxicity. We next generated H1-hESCs 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) system10. H1s were generated with Cas9 fused to a DD tag (dCas9), 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-ieCas9 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 dCas9 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 plating reproducibly and severely reduced cell counts (Supplementary Fig. 3). NSG 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 method21.

Over the 12-day experiment, most sgRNAs remained distributed within ±1 log2(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 (PuroR) 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, 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 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 \(<1.2 \times 10^{-17}\)). 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, \(6.6 \times 10^{-298} \, P_{\text{adj}}\)), 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...
RNP delivery minimize off-target cutting\textsuperscript{14,15}, further supporting that DSBs at a single locus are sufficient to cause a P53-dependent molecular response.

To provide experimental evidence that P53 is functionally involved, we knocked out P53 in H1-iCas9 cells by transiently co-transfecting three chemically synthesized sgRNAs targeting the P53 locus (Supplementary Fig. 5a). The resulting mutant pool was a mixture of control (≤50%) and frameshift alleles (≥50%) at three independent sites within the P53 open-reading frame (ORF) (Supplementary Fig. 5b). The control and P53 mutant pool were then infected with a MAPT sgRNA and grown with or without dox for up to 6 days (Supplementary Fig. 5c). To confirm that the transcriptional response is P53-dependent, mRNA was isolated and quantified using qPCR. At day 2, control cells exhibited a strong induction of P21 and PAS mRNA that was significantly reduced in the P53 mixed mutant pool (Fig. 4a). P53 and P21 proteins were detected using immunofluorescence and high-content imaging. Both P53 and P21 proteins increased in

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 (1,000x coverage). On days 4, 8, and 12, cells were dissociated and counted to maintain 1,000 cells per 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 (1,000x coverage). On days 4, 8, and 12, cells were dissociated and counted to maintain 1,000x 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, Log2(fold change) for the entire 13,000 sgRNA library at the indicated days after dox treatment; n = 12,910 sgRNAs per box blot. d, Log2(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 cell 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 log2(fold change) were calculated for non-targeting sgRNAs that were compared to the initial representation for each of the sgRNAs in the initial pool. 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 = 1,466, 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 log2(fold change) were calculated for each sgRNA normalized to the initial representation for each of the sgRNAs in the initial pool. The log2 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.
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 
cells 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 
hPSCs5–13. We hypothesized that DNA-damage-induced toxicity is in 
direct opposition of engineering efforts. To determine if P53 inhib-
its HDR, we developed an assay to measure precise targeting of a 
transgene into the OCT4/POUSFi locus. We used a pair of sgRNAs

![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 Padj × 10−10. The x axis is log2(fold change). We performed a simple two-group comparison Wald test implemented in the DESeq2 package by default DESeq2 performs a two-sided test. Circles indicate differentially expressed genes (blue circles, Padj < 1.2 × 10−10; grey circles, Padj > 1.2 × 10−10).](image)

**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 Padj × 10−10. The x axis is log2(fold change). We performed a simple two-group comparison Wald test implemented in the DESeq2 package by default DESeq2 performs a two-sided test. Circles indicate differentially expressed genes (blue circles, Padj < 1.2 × 10−10; grey circles, Padj > 1.2 × 10−10). 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 ± 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.
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 DSβ-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 confluence 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/POUSF1 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) 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<sup>ΔD10A</sup> 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<sup>25–31</sup>. The Cas9<sup>ΔD10A</sup>/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. 4c 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\(^{12-15}\). 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\(^{6,17}\). 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\(^{7,26}\). 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\(^8\), transient inhibition is well tolerated in hPSCs\(^{12,26,29}\). 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 dominant negative P53 mutations. Although the mutation rate with clonal expansion could potentially select for dominant negative P53 mutations, Nature 545, 229–233 (2017).


Wells, M. F. et al. Genetic ablation of AXL does not protect human progenitor cells and cerebral organoids from Zika virus infection. Stem Cell 19, 703–708 (2016).


Sundar, N. et al. Human embryonic stem cells have constitutively active B at the golgi and are primed to undergo rapid apoptosis. Mol. Cell 46, 573–583 (2012).


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

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 cells from the plates to be resuspended in E8 plus triazavirin (Selleckchem, S1459) at 2 μg/ml. For lentiviral transduction, electroporation, pooled screening and live imaging of confluence, accutase (Gibco, A110501) was used to create a single-cell suspension, which was counted to accurately replate specific numbers of cells in media containing 200 μg/ml of puromycin spanning 0.3 μg/ml to 2 μg/ml puromycin (Carin, 61-385-RA). At the onset of each megagenealogy experiment, Shield1 (Clontech, 631037, 0.5 μM) or dox (Clontech, 631311, 2 μg/ml) 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_PuroR backbone by GenScript. The 13,000 sgRNA library was designed, cloned into the pNGx_LV_g003 (Supplementary Table 3) and cloned into the pNGx_LV_g003 backbone (Supplementary Table 4) and cloned into the pNGx_LV_g003 (Supplementary Table 4) and cloned into the pNGx_LV_g003 backbone. Next-generation library construction, sequencing and data analysis were performed as described in ref. 48. Non-hPSCs pooled screening data are available but restricted to non-targeting sgRNA sequences.

Crispr in vivo screen for Cas9 optimization. During REPLAP, lentivirusCRISPRs were added to a single-cell suspension of 200,000 cells in 2 ml E8 for lentiviral delivery of sgRNA for Cas9 mutagenesis. To ensure proper transgene expression, on the day of each mutagenesis experiment, Shield1 (Clontech, 631037, 0.5 μM) was added to the cultures. Cells were dissociated with accutase and resuspended in 100 μg/ml puromycin for the remainder of the screen. Dox and Shield1 were added to the Cas9-positive conditions from day 1 throughout 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 log2-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. 48.

Lentiviral delivery of sgRNAs for Cas9 mutagenesis. Replification lentivirusCRISPRs were added to a single-cell suspension of 200,000 cells in 2 ml E8 (Stemcell Tech., 05940) with 0.8 μg/ml triazavirin (Selleckchem, S1459). After 24 h, cells were resuspended in 2 μg/ml puromycin (Carin, 61-385-RA). For qPCR, mRNA concentration was measured using a Nanodrop 2000 (ThermoFisher). Relative expression was calculated using the ΔΔCT method.

Lentiviral delivery of synthetic crRNA/tracrRNAs for Cas9 mutagenesis. icas9 cells were treated with dox for 24 h before transfection of synthetic CRISPR RNP (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 triazavirin. The amount of tracrRNA/tracrRNA mixture was then diluted in 100 μl Opti-MEM (ThermoFisher, 13190088) and incubated for 5 min at RT. Parallel 6 μl of RNAmix (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 RNAmix/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 in vivo screen for Cas9 optimization. In this application, the DNA template for PCR during CRISPR RNP (crRNA)trans-activating crRNA (tracrRNA) pairs. Cells were dissociated with accutase and resuspended in 100 μg/ml puromycin for the remainder of the screen. Dox and Shield1 were added to the Cas9-positive conditions from day 1 throughout 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 log2-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. 48. 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 replicates in the same microcentrifuge tube. Total mRNA was isolated using the RNeasy Miniprep kit plus (Qiagen, 74134). An Agilent 2100 bioanalyzer and Nano 6000 kit (Agilent, 5067-1512) 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 76bp 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) with the STAR algorithm v2.5.1b (Gene Expression and Transcript Quantification (TPM) was performed using HTSeq-count (v0.6.0) and RSEM (v1.2.28) respectively. The gene counts were then used for differential expression analysis using DESeq2. Ninety-three percent of variance was explained with principal component analysis, and confirmed samples had 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, 18800051). cDNA was diluted 1:5 in H2O before analysis using tagman gene expression arrays and the 2X Fast Start Universal Probe master mix (ROX) (Roche, 4913957001). qPCR plates (384-well) were run on a ViA 7 Real-Time PCR System (ThermoFisher). Relative expression was calculated using the ΔCT method as described in ref. 44, and BACTIN was used as the reference gene. TaqMan gene expression arrays PAM-MGB (ThermoFisher, 4331182): P21/Cdkn1a (Hs00355782_ml), BACTIN (Hs01060665_g1) and FAS (Hs01063653_ml). 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 for another analysis. The knowledge base used was a combination of a MetaBase (a manually curated commercial database of mammalian biology from Clarivate Analytics) and STRING (v10).

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 imaging.
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 (12D11) (CST-2947), 1:250 P53 (7F5) (CST-2527), 1:300 FLAG (M2) (Sigma-F1804), 1:200 CC3 (Asp175) (CST-9661), 1:500 phospho-histone H2AX (Ser139/Y142) (CST-5438), 1:500 cleaved PARP-647 (Asp214) (D64E10) (CST-6987). Secondary antibodies: 1:500 goat anti-mouse IgG (H+L) AFF488 conjugate (ThermoFisher, A-11092), 1:500 goat anti-rabbit IgG (H+L) AFF488 conjugate (ThermoFisher, A-11090). For OCT4 targeting assay, live cells were imaged for tdTomato fluorescence and then fixed, permeablized, washed, and 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 was used for detection (34065, Thermo). Live and fixed immunofluorescent images were taken using the x10 and x20 objectives on an Axio Observer.D1 (Zeiss). Images for high content analysis were taken on an Incell 6000 (GE Healthcare Life Sciences). P53, P21, H2AX, cPARP and CC3 immunofluorescence quantification was conducted using Cell Profiler (version 2.1.1, revision 6cd2896) software. For P53 and P21 proteins, average immunofluorescent intensity was determined for each nucleus, and a positive expression threshold was set based on the quantified 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-100 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 antibody (C30A3) antibody (5263, CST) and 1:50 647 FITC-conjugated anti-TRA-1-60 antibody (FCMAB115F, Millipore), 1:50 647 conjugated anti-Sox2 antibody (C2063, CST) and 1:50 647 conjugated anti-Sox2 antibody (C2063, 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 SH8002. iCas9 cells were infected with lentiCRISPRv2 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 Cas9D10A nickases36bp apart spanning the stop codon to trigger a DSB with 5’ overhangs. The donor plasmid contained a transgene flanked by two ~700bp homology arms (Fig. 4d and Supplementary Data 6). Before genome engineering, hPSCs were pretreated with 1μM thiazovivin for at least 2 h and collected using accutase. A mixture of 4μg of OCT4-tdTomato-puro3 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 2μg of either an episomal vector for P33DD (pCE-mP353D3, Episomal iPS Reprogramming Kit cat. no. A15960, addgene no. 41856) or EBNA alone (pCXB-EBNA1, Epis Prepare kit: cat. no. A15960, addgene no. 41857) were electroporated into iX10® 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) and thiazovivin. After 48 h, cells were selected with 0.3μM 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 variances 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 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.

For Supplementary Fig. 6: 8402-iPSCs unpaired two-tailed Welch’s t-test, unequal variance (t=3.14, d.f. = 2.008, P = 0.0445). H1-iHESCs, unpaired two-tailed Welch’s t-test, unequal variance (t=3.809, d.f. = 2.019, P = 0.0299).

H2A.X and H2AX repair with an OCT4 gene trapping vector was performed using PRISM software (version 7.0c). For Supplementary Fig. 6: 8402-iPSCs unpaired two-tailed Welch’s t-test, unequal variance (t=3.14, d.f. = 2.008, P = 0.0445). H1-iHESCs, unpaired two-tailed Welch’s t-test, unequal variance (t=3.809, 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 precision40. 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). Results for FAS mRNA were replicated in Fig. 4a.

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). Results for FAS mRNA were replicated in Fig. 4a.

References
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).
Experimental design

1. Sample size

Describe how sample size was determined.

Specific sample 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 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.


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).

<table>
<thead>
<tr>
<th>n/a</th>
<th>Confirmed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

n/a | 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.

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, SSD
- P53 (7F5) (CST-2527) lot 8
  1:250 dilution – Induced by Cas9 DDR Fig 3B, Reduced in P53 mutant* with Fig. 4B, SSD
- 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
- POUSF1/OCT4 gRNA1 CAACAATGAAAATCTTCAGG
- SOX2 gRNA2 CGTTCATCGACAGGCTAAG

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

11. Description of research animals
   Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines
   No animals used in this study

12. Description of human research participants
   Policy information about studies involving human research participants
   This study did not involve human research participants
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. ☒