Supplemental Material−Extended Methods

Cell Culture

mESCs were a provided by A. Smith (Ying et al. 2003). Cell were cultured at 37 °C and with 5% CO2 on gelatin coated plates in mESC growth medium composed of knockout Dulbecco's modified Eagle's medium (DMEM) supplemented with 15% ES cell-qualified Fetal bovine serum, 0.1 mM non-essential amino acids, 100 μM β-mercaptoethanol, 50 μg ml−1 penicillin, 50 μg ml−1 streptomycin, 100 units ml−1 Leukemia inhibitory factor, 2 mM L-glutamine, 3uM CHIR99021 and 1 uM PD0325901.

Virus production

Virus were packed in 293T cells. Transfection was performed using Polyethylenimine, and the plasmids pCMV-VSVG (a gift from Bob Weinberg; Addgene plasmid # 8454; http://n2t.net/addgene:8454; RRID:Addgene_8454) (Stewart et al. 2003) and psPAX2 (a gift from Didier Trono; Addgene plasmid # 12260; http://n2t.net/addgene:12260; RRID:Addgene_12260). The viruses were harvested at 48 and 72h post transfection. For the CAS9 virus, the lentiCas9-Blast was used (a gift from Feng Zhang; Addgene plasmid # 52962; http://n2t.net/addgene:52962; RRID:Addgene_52962) (Sanjana et al. 2014), and for the knockdown library the Brie pooled library was used (a gift from David Root and John Doench; Addgene #73633) (Doench et al. 2016).

Generation of CAS9 expressing cells

mESCs were transfected with a lentivirus containing the CAS9 expression vector at increasing multiplicity of infection (MOI) from 0.5 to 4 in four different wells. The cells

were then treated with Blasticidin for 7 days, and the well with the highest number of surviving cells was maintained (MOI 4).

Screen for essential genes using pooled CRISPR library

The Brie knockout library was used in the screen. The library targets 19,674 genes, with 4 gRNAs targeting each gene. In addition, the library contains 1000 control gRNAs which allows for an estimation of random drift effects on gRNA abundance. CAS9 expressing mESC were transfected with lentiviruses containing the Brie pooled library at MOI ~ 0.4. The cells were treated with puromycin for 7 days. Following the antibiotic selection, the cells were allowed to proliferate for additional 10 days. Cells were passaged at days 8, 11, 15 and 18 post transfection. For each passage a minimum number of 31 million cells was retained for sequencing and additional 31 million cells were re-plated, this allowed for a maintenance of adequate library representation (an average of ~400 cells per gRNA).

DNA extraction, amplification and sequencing

Genomic DNA was extracted from the cells using Zeymo quick-DNA miniprep kit. The extracted DNA was amplified using Q5® High-Fidelity 2X Master Mix, during the amplification adaptors and barcodes were attached. In addition to DNA samples from the transfected cells, a sample of the plasmid library used for the lentivirus production was also sequenced. To minimize variance originating from amplification biases, each sample was amplified in 3-4 different PCR reactions and the products were pooled. The PCR products were purified using AMPure beads. Samples were sequenced on a NextSeq machine (Illumina). Reads were counted by first locating the CACCG sequence

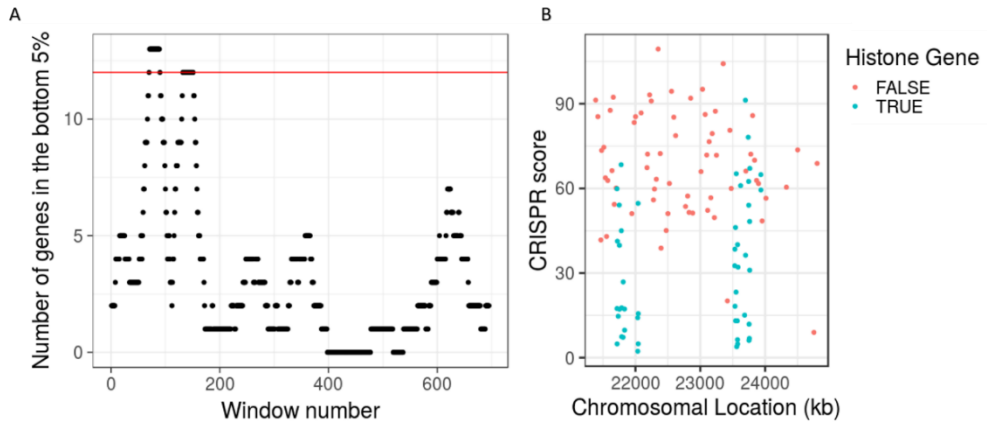
that appears in the vector 5' in all gRNA inserts. The next 20 bases are the gRNA insert, which were then mapped to a reference file of all possible gRNAs present in the library using bowtie2 (Sanjana et al. 2014).

Identification of essential genes

For all samples, library sizes were normalized using calcNormFactors function in edgeR, which uses the trimmed mean of the M values method (Sanjana et al. 2014). This normalization corrects for under estimation of gRNAs abundances due to the presence of a few highly represented gRNAs. Following the normalization, the log fold change of each gRNA in each proliferation day was calculated relative to the initial counts in the plasmid library. In order to determine which genes are under significant positive or negative selection we used a simulation-based approach that is based on comparing the fold change of gRNAs targeting each gene to the fold change of randomly selected control gRNAs. This approach allows the detection of negative and positive selection that effect only gene targeting gRNAs in the presence of random drift. For each proliferation day we ranked all gRNAs by their representation fold change relative to the library. For each gene we then calculated the sum of ranks of its gRNA across all days. This score was compared to an empirical null distribution generated by randomly selecting 4 control gRNA and calculating their sum of ranks (10,000 simulation). The *P*-value for each gene was calculated by dividing the number of time its score was lower or equal to the scores obtained by the random simulation of control gRNAs. *P*-values were corrected for multiple testing using Benjamini-Hochberg false discovery rate (FDR) procedure. CRISPR score were defined as the sum of ranks for each gene divided by 10^4.

Testing for a possible bias arising from amplified genomic regions

Recent studies reported that CAS9-mediated cleavage of amplified genomic regions could cause a DNA damage response that results in cell death (Ruan et al. 2008; Ihry et al. 2018). Thus, low CRISPR scores of genes located in those regions will not necessarily reflect essentiality. In order to test this in our data, we used a sliding window approach to identify contiguous stretches of low CRISPR scores. To this end, for each chromosome we sorted the genes according to their location and used a sliding window of 40 genes. Windows in which more than 30% of the genes had a CRISPR sore in the bottom 5% were further examined. We detected two such gene clusters on chromosome 13 (see figure below). When examining the clusters closely, we found that they contain mostly histone genes and that non-histone genes in those regions did not have low CRISPR scores. Therefore, we concluded that the clusters probably reflect true essentials and not a DNA damage response.



Amplified genomic regions do not generate false-positive hits. (A) The number of genes in the bottom 5% of the CRISPR scores for each sliding window for genes in chromosome 13, red line indicates a threshold of 30% of the genes in the window. **(B)** The CRISPR scores of the two clusters in chromosome 13 containing genes with low crisper scores are shown for histone genes (blue) and non-histone genes (red).

Overlap of essential genes with previous published dataset

The overlap between genes found to be under significant negative selection (FDR corrected $P < 0.05$) in the screens was tested using Fisher's exact test. Welch's t-test was used to test the significance for the difference in the quantitative fold change (day 15 post transfection) between genes found to be under significant or non-significant negative selection in previous screen (Tzelepis et al. 2016). A Combined *P*-value for the two screens were obtained using the sum z (Stouffer's) method (Stouffer et al. 1949; Whitlock 2005). In this method p-values are converted to Z values, combined, and then converted back into *P*-values.

GO terms and KEGG pathway enrichment analysis

GO terms enrichment of essential genes was performed using the Gorilla tool (Eden et al. 2009) on the background of all genes tested in the screens. GO terms significant at FDR corrected value $P < 0.05$ were summarized using reviGO (Supek et al. 2011). Terms that were not specific and contained more than 20% of the tested genes were removed. In addition, terms for which all the essential genes overlapped other more enriched terms were removed. Comparison of GO term enrichment between essential genes in the fast and gradual declining group and between genes essential specifically in mESCs or specifically in hESCs relative to genes essential in both mESCs and hESCs was performed using Fisher's exact test. Comparison of GO term enrichment between all mESCs essential genes and ESCs specifically essential genes was performed on all the terms significantly enriched for ESCs essential genes and the top 10 terms enriched for all essential genes. P values were calculated using a permutation test by sampling 187 genes (the number of ESCs essential genes) from the mESCs essential gene list and

testing their enrichment for each GO term. P values were corrected for multiple testing using the FDR correction. KEGG pathways enrichment analysis of essential genes was done using the clusterProfiler R package (Yu et al. 2012). Comparison of KEGG pathways enrichment between essential genes in the fast and gradual declining groups and between genes essential specifically in mESCs or specifically in hESCs relative to genes essential in both mESCs and hESCs was performed using Fisher's exact test

Analysis of paralog genes

Paralog genes were identified using ensemble biomart (Smedley et al. 2015) and TreeFam (Ruan et al. 2008) databases. The significance of the difference between the CRISPR score distribution of genes with and without a paralog in the top KEGG pathways was tested using a Mann-Whitney test. The enrichment for genes without a paralog for genes in the fast and gradual declining clusters and for non-essential genes was tested using a fisher's exact test.

Cluster identification based on gRNA kinetics

Clustering of essential genes in mESC was performed based on the correlation between the depletion rates for all essential genes. The correlation matrix was then used for hierarchical clustering using R hclust function with the default settings. The dendrogram branches were cut to obtain two main clusters.

Gene expression in ESCs

Gene expression for mESCs was obtained from the study of Tesar et al. (Tesar et al. 2007) and for hESCs from the study of de Leemput et al. (van de Leemput et al. 2014). For mESCs microarray data was normalized using quantile normalization and the mean expression across 3 samples for each gene. For hESCs normalized counts from RNA-seq data were used, and for each gene the mean expression across 4 samples was calculated. To test for significance differences in gene expression between groups we used the Welch's t-test in case of two groups, and Tukey test for 3 groups.

Difference in mean half-life between genes in the gradual and fast declining group

Data on protein half-life was obtained from Mathieson et al. (Mathieson et al. 2018) and Schwanhausser et al. (Schwanhäusser et al. 2011). The difference in the mean log10 half-life for genes in the fast and gradual declining groups was determined using Welch's t-test.

Comparison of essential genes between mESCs, hESCs and cancer cell lines

Data on essential genes in haploid hESCs grown on feeder cells was obtained from Yilmaz et al. (Yilmaz et al. 2016); data on essential genes in diploid hESCs grown on feeder cells was obtained from Mair et al. (Mair et al. 2019). Data on essential genes in human cancer cell lines was obtained from project Achilles (Meyers et al. 2017). Genes essential in cancer cell lines were defined as genes in the common essential gene list (genes in the top ranked essential genes in >90% of cell lines). Haploid hESCs were infected with a CRISPR knockout library composed of ~180,000 gRNAs (10 gRNAs per gene), cells were collected on days 7,23 and 30 (Yilmaz et al. 2018). Diploid hESCs

were infected with a CRISPR knockout library composod of 71,090 gRNAs (4 gRNAs per gene), cells were collected on days 4,8 and 12(Mair et al. 2019).

Human mouse paralogs were identified using biomart (Ensembl genes 97)(Smedley et al. 2015). Genes without direct 1:1 orthologs were filtered out and not used in the human-mouse comparison. Genes essential only in ESCs were defined as genes essential in mESCs and in at least one of the hESCs (FDR corrected $P < 0.05$) but not present in the human cancer cell lines common essential gene list . Genes essential specifically in mESCs were defined as genes not essential in both hESCs . Genes essential specifically in hESCs were defined as genes essential in at least one of the hESCs lines (FDR corrected $P < 0.05$) and not essential in mESCs.

Differential expression analysis between mESCs and EpiSCs

Microarray gene expression data for 3 mESCs samples and 3 EpiSCs samples was obtained from Zhou et al. (Zhou et al. 2012). The data was normalized by quintile normalization and differential expression was performed using limma (Ritchie et al. 2015). Association with genes significantly upregulated in mESCs relative to EpiSCs was determined using Fisher's exact test.

Overlap with mouse embryonic lethal genes and human LoF intolerant genes

A list of genes which have a knockout mouse with a phenotype of prenatal lethality, abnormal survival (excluding extended life span), decreased prenatal and postnatal growth (, excluding increased body size, weight gain and diet related phenotypes) was obtained from the mouse Genome Informatics (MGI) database(Table S5) (Bult et al. 2018). The overlap between genes identified as essential in mESC and the list of genes

associated with growth and lethality in mice was tested only for genes that have a knockout mouse information in the MGI database. The significance of the overlap was tested using Fisher's exact test. A list of genes defined as human LoF intolerant was obtained from Lek et al. (Lek et al. 2016). The significance of the overlap between human LoF intolerant genes and essential genes in mESCs and hESCs was calculated using Fisher's exact test.

Association of essential genes with human phenotypes

Analysis of human phenotypes for genes essential in mESCs or hESCs was based on phenotypes in the Human Phenotype Ontology (Köhler et al. 2019). The significance of the association was calculated by Fisher's exact test and *P*-values were corrected for multiple testing by FDR procedure. Significant phenotypes with more than 90% overlap of genes with a more significant phenotype were filtered out. Association with developmental and neurodevelopmental phenotypes for genes essential in mESCs or hESCs, was based on the DDG2P dataset from the deciphering developmental disorders project (Wright et al. 2015). Neurodevelopmental phenotypes were defined as any developmental phenotype involving the brain. The significance of the association was calculated using Fisher's exact test.

Gene expression during in vitro human and mouse corticogenesis

Neurodevelopmental disorders risk genes, as defined by the deciphering developmental disorders project (Wright et al. 2015) or genes present in the developmental brain disorders database (tires 1 & 2) (Gonzalez-Mantilla et al. 2016), were clustered according to their expression patterns during *in vitro* corticogenesis of

mouse (Hubbard et al. 2013) or human (van de Leemput et al. 2014) cells. Clustering was performed using R hclust function with the default settings (the complete method).

References

Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE, Anagnostopoulos A, Asabor R, Baldarelli RM, Beal JS, Bello SM, et al. 2018. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res*.

Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, et al. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**: 184–91.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**: 48.

Gonzalez-Mantilla AJ, Moreno-De-Luca A, Ledbetter DH, Martin CL, A M-D-L, M Y-A, J M, SE M, D M-D-L, F B, et al. 2016. A Cross-Disorder Method to Identify Novel Candidate Genes for Developmental Brain Disorders. *JAMA Psychiatry* **73**: 275.

Hubbard KS, Gut IM, Lyman ME, McNutt PM. 2013. Longitudinal RNA sequencing of the deep transcriptome during neurogenesis of cortical glutamatergic neurons from murine ESCs. *F1000Research* **2**: 35.

Ihry RJ, Worringer KA, Salick MR, Frias E, Ho D, Theriault K, Kommineni S, Chen J, Sondey M, Ye C, et al. 2018. p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nat Med* **24**: 939–946.

Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gourdine J-P, Gargano M, Harris NL, Matentzoglu N, McMurry JA, et al. 2019. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res* **47**: D1018–D1027. http://www.ncbi.nlm.nih.gov/pubmed/30476213 (Accessed February 28, 2019).

Lek M, Karczewski KJ, Minikel E V., Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**: 285–291.

Mair B, Tomic J, Masud SN, Tonge P, Weiss A, Usaj M, Tong AHY, Kwan JJ, Brown KR, Titus E, et al. 2019. Essential Gene Profiles for Human Pluripotent Stem Cells Identify Uncharacterized Genes and Substrate Dependencies. *Cell Rep* **27**: 599-615.e12.

Mathieson T, Franken H, Kosinski J, Kurzawa N, Zinn N, Sweetman G, Poeckel D,

Ratnu VS, Schramm M, Becher I, et al. 2018. Systematic analysis of protein turnover in primary cells. *Nat Commun* **9**: 689.

Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, Dharia N V, Montgomery PG, Cowley GS, Pantel S, et al. 2017. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat Genet* **49**: 1779–1784.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**: e47–e47.

Ruan J, Li H, Chen Z, Coghlan A, Coin LJM, Guo Y, Hériché J-K, Hu Y, Kristiansen K, Li R, et al. 2008. TreeFam: 2008 Update. *Nucleic Acids Res* **36**: D735-40.

Sanjana NE, Shalem O, Zhang F. 2014. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**: 783–784. http://www.ncbi.nlm.nih.gov/pubmed/25075903 (Accessed February 7, 2017).

Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. 2011. Global quantification of mammalian gene expression control. *Nature* **473**: 337–342.

Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, Arnaiz O, Awedh MH, Baldock R, Barbiera G, et al. 2015. The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res* **43**: W589–W598.

Stewart SA, Dykxhoorn DM, Palliser D, Mizuno H, Yu EY, An DS, Sabatini DM, Chen ISY, Hahn WC, Sharp PA, et al. 2003. Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA* **9**: 493–501.

Stouffer S, Suchman E, DeVinney L, Star S. 1949. The American soldier: Adjustment during army life.(Studies in social psychology in World War II), Vol. 1. http://psycnet.apa.org/record/1950-00790-000 (Accessed February 4, 2019).

Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms ed. C. Gibas. *PLoS One* **6**: e21800.

Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, Gardner RL, McKay RDG. 2007. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**: 196–199.

Tzelepis K, Koike-Yusa H, De Braekeleer E, Li Y, Metzakopian E, Dovey OM, Mupo A, Grinkevich V, Li M, Mazan M, et al. 2016. A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Rep* **17**: 1193–1205.

van de Leemput J, Boles NC, Kiehl TR, Corneo B, Lederman P, Menon V, Lee C, Martinez RA, Levi BP, Thompson CL, et al. 2014. CORTECON: A Temporal Transcriptome Analysis of In Vitro Human Cerebral Cortex Development from Human Embryonic Stem Cells. *Neuron* **83**: 51–68.

Whitlock MC. 2005. Combining probability from independent tests: The weighted Z-method is superior to Fisher's approach. *J Evol Biol* **18**: 1368–1373.

Wright CF, Fitzgerald TW, Jones WD, Clayton S, McRae JF, van Kogelenberg M, King DA, Ambridge K, Barrett DM, Bayzetinova T, et al. 2015. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet* **385**: 1305–1314.

Yilmaz A, Peretz M, Aharony A, Sagi I, Benvenisty N. 2018. Defining essential genes for human pluripotent stem cells by CRISPR–Cas9 screening in haploid cells. *Nat Cell Biol* **20**: 610–619.

Yilmaz A, Peretz M, Sagi I, Benvenisty N. 2016. Haploid Human Embryonic Stem Cells: Half the Genome, Double the Value. *Cell Stem Cell* **19**: 569–572. http://linkinghub.elsevier.com/retrieve/pii/S1934590916303496 (Accessed November 22, 2016).

Ying Q-L, Stavridis M, Griffiths D, Li M, Smith A. 2003. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol* **21**: 183–186. http://www.nature.com/articles/nbt780 (Accessed May 5, 2018).

Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omi A J Integr Biol* **16**: 284–287.

Zhou W, Choi M, Margineantu D, Margaretha L, Hesson J, Cavanaugh C, Blau CA, Horwitz MS, Hockenbery D, Ware C, et al. 2012. HIF1α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J* **31**: 2103–2116.