RESOURCE
https://doi.org/10.1038/s41556-018-0088-1
nature cell biology

Defining essential genes for human pluripotent stem cells by CRISPR–Cas9 screening in haploid cells

Atilgan Yilmaz[1,2], Mordecai Peretz[1,2], Aviram Aharony[1], Ido Sagi[1] and Nissim Benvenisty[1,*]

The maintenance of pluripotency requires coordinated expression of a set of essential genes. Using our recently established haploid human pluripotent stem cells (hPSCs), we generated a genome-wide loss-of-function library targeting 18,166 protein-coding genes to define the essential genes in hPSCs. With this we could allude to an intrinsic bias of essentiality across cellular compartments, uncover two opposing roles for tumour suppressor genes and link autosomal-recessive disorders with growth-retardation phenotypes to early embryogenesis. hPSC-enriched essential genes mainly encode transcription factors and proteins related to cell-cycle and DNA-repair, revealing that a quarter of the nuclear factors are essential for normal growth. Our screen also led to the identification of growth-restricting genes whose loss of function provides a growth advantage to hPSCs, highlighting the role of the P53–mTOR pathway in this context. Overall, we have constructed an atlas of essential and growth-restricting genes in hPSCs, revealing key aspects of cellular essentiality and providing a reference for future studies on human pluripotency.

Haploid cells allow genetic screening through the generation of a highly enriched hemizygous mutant library, owing to the single set of chromosomes in these cells[1–3]. Much previous work on haploid genetics has been carried out in unicellular organisms, but recent developments have made it possible to extend this field into mammalian cells[1–10].

We recently isolated haploid human embryonic stem cells (hESCs)[11]. These cells exhibit human pluripotent stem cell (PSC) features in their colony morphology, alkaline phosphatase activity, gene expression signatures and epigenetic profiles. Interestingly, haploid hESCs can differentiate into haploid somatic cells in vitro and in vivo, generating cell types representative of the three embryonic germ layers[11]. Haploid hESCs can be grown in standard culture conditions for over 30 passages while retaining a normal haploid karyotype. Therefore, haploid hESCs provide an efficient screening platform to address questions regarding pluripotency on a genome-wide level.

Haploid mammalian cells have recently been used for loss-of-function genetic screens[1,2]. Initial loss-of-function screens in humans utilized a near-haploid leukaemic cell line. This transformed cancer cell line has been used previously to identify the host factors used by human pathogens[12] and, more recently, it has been utilized for a genome-wide loss-of-function screen to identify essential genes in the human genome and for studying synthetic lethality between different genes[13,14].

Here, we performed a genome-wide CRISPR–Cas9-based loss-of-function screen on karyotypically normal haploid hPSCs to define the genes essential for normal growth and survival of human PSCs and the genes that restrict their growth. Our analysis suggests an intrinsic bias of essentiality across cellular compartments, and enables examination of the growth-retardation phenotype of all autosomal-recessive (AR) human disorders. Furthermore, our screen revealed the essentialome of hPSC-specific genes, and highlighted the main pathways that regulate the growth of these cells.

Results

Identification of cell-essential genes in hPSCs. To define the essentialome of hPSCs, we took advantage of our recent discovery of haploid hESCs[1] to build a CRISPR–Cas9-based genome-wide loss-of-function mutant library[13,15] (Fig. 1a). We utilized a human activity-optimized single guide RNA (sgRNA) library that targets more than 18,000 coding genes and contains 10 sgRNAs for about 99% of the target genes[13]. Using this library of about 180,000 guide RNAs (gRNAs), we aimed to identify mutations in essential genes that affect the survival or normal growth of hESCs based on their depletion in the hESC population, as well as mutations in growth-restricting genes that provide a growth advantage to hESCs based on their enrichment over time in culture. We analysed the abundance of sgRNAs within the haploid hESC population at multiple time points after the co-delivery of sgRNAs and *Cas9*, and found gradual depletion and enrichment of numerous sgRNAs (Fig. 1b and Supplementary Fig. 1a). This observation allowed us to analyse two opposing subsets of genes, namely the essential and the growth-restricting genes. To assess the validity of our screen in the context of pluripotency, we followed the temporal changes in sgRNA representation for two well-characterized hESC-enriched and pluripotency-associated genes, *POU5F1* (also known as *OCT4*) and *PRDM14*, as well as their neighbouring hESC-expressed genes (Fig. 1c). sgRNAs targeting both *POU5F1* and *PRDM14* became significantly depleted within three weeks after the delivery of sgRNAs. In contrast, sgRNAs targeting the neighbouring genes, which are not expressed exclusively in hESCs, were not depleted over time. To reveal significant changes in sgRNA representation between the initial and final hESC populations, we calculated a CRISPR score as the ratio of sgRNA abundance between final and initial populations for each gene (Supplementary Fig. 1b)[13,15]. CRISPR scores demonstrated a high degree of reproducibility across replicate experiments (Supplementary Fig. 1c and Supplementary Table 1). Based on this analysis we identified about 9% of the genes in the coding genome as essential for normal growth

[1]The Azrieli Center for Stem Cells and Genetic Research, Department of Genetics, Silberman Institute of Life Sciences, The Hebrew University, Jerusalem, Israel. [2]These authors contributed equally: Atilgan Yilmaz and Mordecai Peretz. *e-mail: nissimb@mail.huji.ac.il

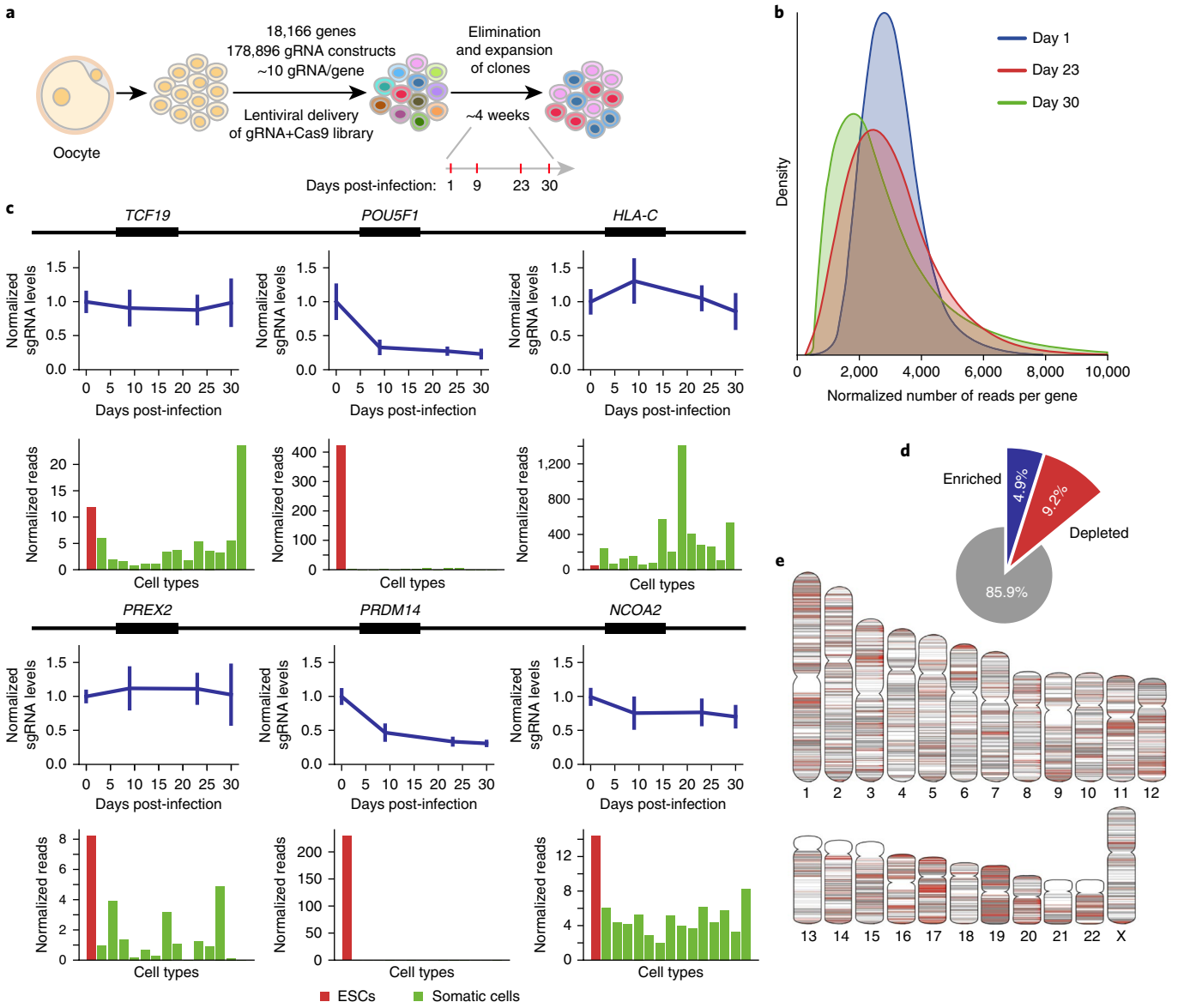

Fig. 1 | Establishment and characterization of a genome-wide CRISPR–Cas9 screen in haploid hPSCs. a, Schematic illustrating generation of the mutant library. **b**, Distribution of the number of gRNA reads per gene at indicated time points after gRNA infection. **c**, Top, Schematic representation of the genomic loci of two pluripotency-associated genes (*POU5F1* and *PRDM14*) and their neighbouring genes expressed in hESCs. Middle, mean ± s.e.m. of sgRNA reads per gene over time in culture ($n = 20$ sgRNAs, two biological replicates of 10 independent sgRNAs per gene). Source data are provided in Supplementary Table 4). Bottom, Expression levels of the genes in ESCs and 14 somatic cell types (from left to right: skin, brain, heart, liver, skeletal muscle, pancreas, lung, stomach, blood, small intestine, kidney, adipose, transformed fibroblasts and transformed lymphocytes). **d**, Percentages of essential genes (red) and growth-restricting genes (blue). Genes with false discovery rate (FDR) less than 0.05 are regarded as significant (Kolmogorov–Smirnov test (KS test), $n = 20$ gRNAs). **e**, Chromosomal distribution of essential genes (red lines) and all other genes targeted in the library (grey lines).

of hESCs, as well as about 5% of the genes as growth-restricting genes (Fig. 1d). Importantly, both essential and growth-restricting genes are distributed across all chromosomes without enrichment in specific chromosomal regions (Fig. 1e and Supplementary Fig. 1d).

We then compared our list of essential genes to those identified in three previous screens performed in human cancer and immortalized lines using a variety of methodologies[13,14,16]. We found a considerable overlap between the different screens, although each study also pointed to a unique set of essential genes (Supplementary Fig. 2a). Clustering these data sets via a principal component analysis (PCA) revealed that they are separated mainly based on mutagenesis methodology, as recently suggested by others[17] (Supplementary Fig. 2b). Thus, in our comparisons, we focused on the cancer lines that were screened for essential genes using the same sgRNA library. Interestingly, even though the essentialome identified in hESCs clustered more closely to that of cancer lines defined using the same sgRNA library, a third of the essential genes identified in hESCs were unique to these cells, indicating that cell identity is also an important factor in shaping the gene essentiality landscape. Although genetic screens using CRISPR–Cas9 technology have been efficiently performed in diploid cells, the use of haploid cells further increases the efficiency of generating complete loss-of-function frameshift mutations (see Methods 'Data analysis' section and Supplementary Fig. 2c,d).

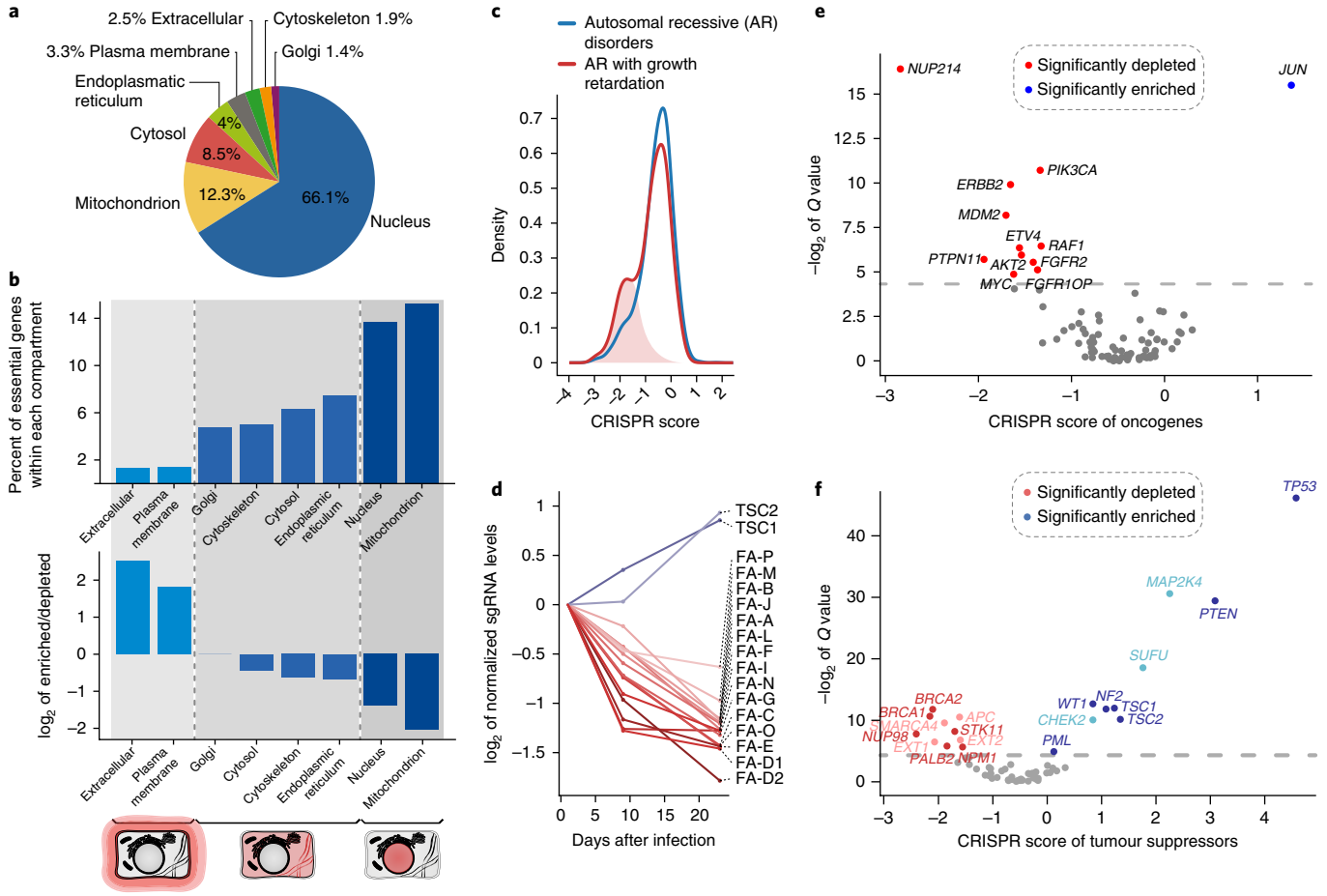

Fig. 2 | Analysis of cell-essential genes. a, Distribution of cell-essential genes across cellular compartments. Essential-gene percentages of the nucleus and mitochondrion compartments were significantly increased over their representation in the library (hypergeometric test (HG test), $n=872$ genes, $P=52\times10^{-65}$ and $P=32\times10^{-12}$, respectively). **b**, Fraction of essential genes within the total number of genes in each cellular compartment (top). Ratio of growth-restricting genes over essential genes in each cellular compartment (bottom). Schematics under the bottom panel illustrate a cell and its compartments. Compartment groups related to graph above are highlighted in red. **c**, CRISPR score represents the average $\log_2$ fold change in the abundance of gRNAs of each gene between final and initial populations. Shown is the distribution of the CRISPR scores of genes associated with AR human disorders (blue curve) and the subset of these genes also associated with a growth retardation phenotype (red curve). **d**, Levels of sgRNA reads per gene over time in a culture for Fanconi anaemia-causing genes (shades of red) and tuberous sclerosis-causing genes (shades of blue). **e**,**f**, Volcano plots representing *Q* value and CRISPR score of canonical oncogenes (**e**) and tumour suppressor genes (**f**). Dashed line: $Q=0.05$ (KS test, $n=20$ gRNAs). Blue: essential genes; red: growth-restricting genes; dark blue: apoptosis-related genes; dark red: genes related to genomic instability and DNA repair.

Cellular localization and disease association of cell-essential genes. We next investigated different aspects relating to cell-essential genes in the context of hESCs, including their cellular localization and their association with AR human genetic disorders and tumour-causing mutations. We found that 66% of the cell-essential genes encode proteins that localize to the nucleus, 12% encode mitochondrial proteins and 8.5% encode cytosolic proteins, while the rest encode proteins that are distributed between the endoplasmic reticulum, plasma membrane, extracellular space, cytoskeleton and the Golgi (Fig. 2a). Analysing the proportion of essential genes among all genes associated with each of these eight cellular compartments revealed three categories: (1) compartments related to the extracellular space showed low proportions of essential genes (about 1%); (2) compartments related to the cytoplasm showed medium proportions of essential genes (5–7%); and (3) the nuclear and mitochondrial compartments showed high proportions of essential genes (14–15%) (Fig. 2b, upper panel). This bias in the cellular localization landscape of the essentialome may suggest different roles for essential genes in the regulation of cell growth and/or different levels of functional redundancy in the various compartments. Interestingly, when we examined the ratio between the number of growth-restricting genes and essential genes within each compartment, we observed that the extracellular space and the plasma membrane had higher representation of growth-restricting genes compared with essential genes, as opposed to the nucleus and mitochondrion (Fig. 2b lower panel and Supplementary Fig. 3a), which may hint at the important involvement of the cellular environment in inhibiting cell growth. The differences in the proportion of essential genes among compartments was equally apparent when only expressed genes (fragments per kilobase of exon per million reads mapped (FPKM) > 1) were analysed (Supplementary Fig. 3b). An analysis of the fraction of essential genes in different cellular compartments in leukaemic KBM7 cells demonstrated a very similar pattern to that observed in hESCs, suggesting that this distribution pattern is shared across different cell types (Fig. 2b upper panel and Supplementary Fig. 3c).

Many of the genes analysed in our screen also underlie human genetic disorders and are mutated in patients. We speculated that some of the genes carrying mutations associated with AR human disorders could be important for the normal growth of hESCs, and

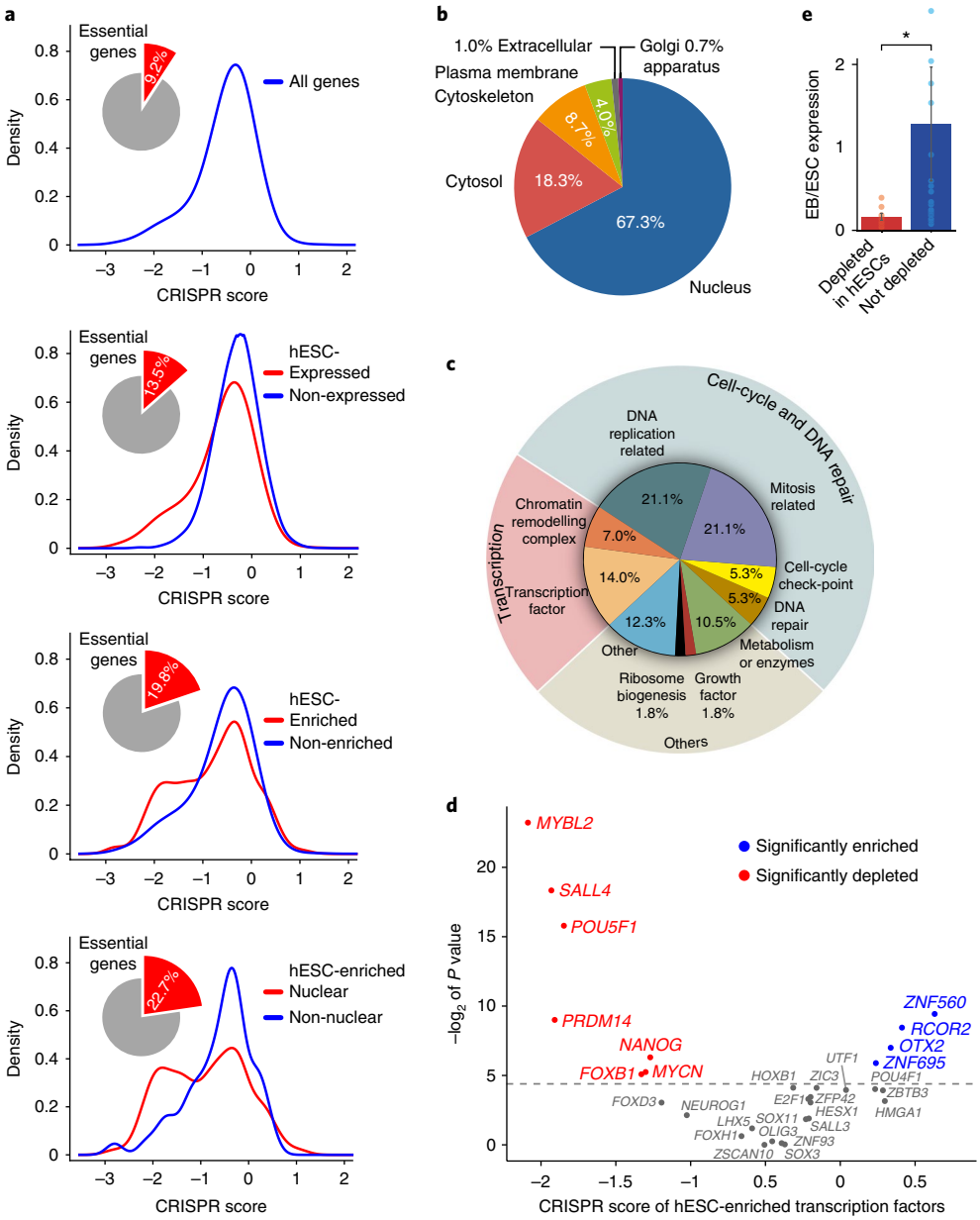

Fig. 3 | Identification and characterization of the hESC essentialome. a, Distribution of CRISPR scores of genes in hESCs. Top to bottom: All genes; hESC-expressed genes; hESC-enriched genes; nuclear hESC-enriched genes. Pie charts show the percentage of essential genes in the gene subsets represented by red curves. **b**, Distribution of the hESC essentialome across cellular compartments. **c**, Functional categorization of the hESC essentialome. **d**, Volcano plot representing significance and CRISPR score of hESC-enriched transcription factors. Dashed line: $P = 0.05$ (KS test, $n = 20$ gRNAs). Blue: essential genes; red: growth-restricting genes. **e**, Expression ratio between embryoid bodies (EBs) and hESCs for the hESC-enriched transcription factors (TFs). TFs whose gRNAs are significantly depleted (red bar) or not depleted (blue bar) are shown (mean ± s.e.m. values, two-tailed *t*-test, $n = 8$ depleted hESC-enriched TFs, $n = 22$ non-depleted hESC-enriched transcription factors, $^*P = 0.027$). Source data are provided in Supplementary Table 4.

hence potentially affect growth in the early human embryo. Of 2,099 human AR-related genes reported in the Online Mendelian Inheritance in Man (OMIM) database[18] that were also represented in our library, 226 (10.8%) were found to be essential for hESC growth. Interestingly, genes responsible for AR disorders that exhibit a growth-retardation phenotype were significantly enriched in essential genes (154 of 766, 20.1%, $P < 0.001$) (Fig. 2c and Supplementary Table 2). A similar analysis in the near-haploid leukaemic KBM7 cell line[6] failed to demonstrate a significant enrichment of essential genes among the genes causing AR disorders with growth-retardation phenotypes in these cells (Supplementary Fig. 3d), suggesting that hESCs provide a more suitable model to study the phenotypes of developmental human disorders. Among AR disorders with a growth-retardation phenotype, we focused on Fanconi anaemia (FA), which was reported to be difficult to model in hPSCs as the growth of the mutant cells was compromised[19,20]. Of 15 genes associated with mutations causing FA, 14 were identified as essential in hESCs (Fig. 2d). In contrast, *TSC1* and *TSC2*, two genes with autosomal dominant mutations associated with tuberous sclerosis and overgrowth in multiple tissues, were identified as growth-restricting genes in these cells[21] (Fig. 2d). Our analysis suggests that the phenotype of growth retardation associated with AR disorders may initiate, in one-fifth of the disorders, at very early stages of embryogenesis. These findings open up an exciting future direction towards

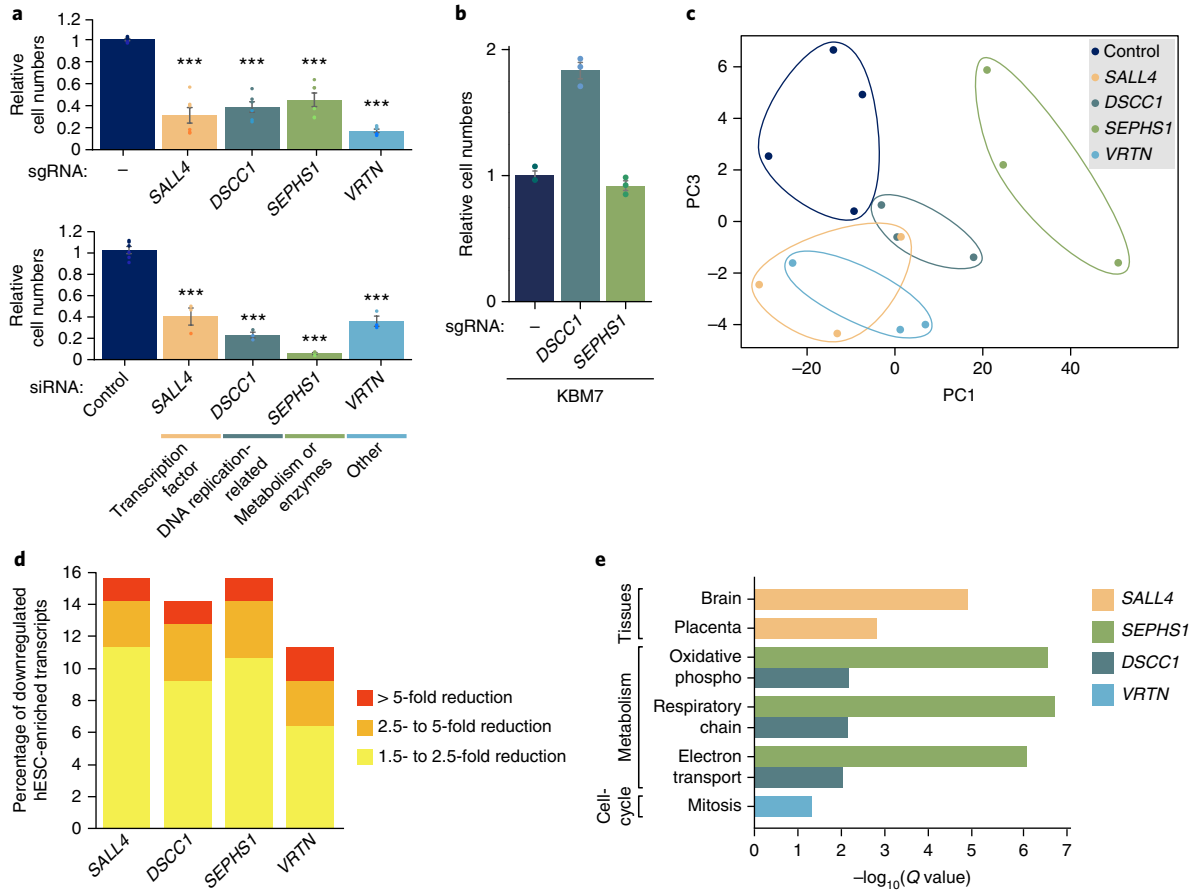

Fig. 4 | Analysis of hESC-essential genes for the survival and pluripotency of hESCs. a, Validation of essential genes representing different functional categories in the hESC essentialome by sgRNA-mediated knockout (upper panel) and siRNA knockdown (lower panel) in diploid hESCs. Either a Cas9-containing sgRNA-free vector or a non-targeting siRNA was used as a control. Shown are mean ± s.e.m. values of the effects of each sgRNA (two-tailed *t*-test, $n = 6$ biological replicates, $P_{SALL4} = 1.6 \times 10^{-4}$, $P_{DSCC1} = 3.1 \times 10^{-5}$, $P_{SEPHS1} = 2.6 \times 10^{-4}$, $P_{VRTN} = 1.9 \times 10^{-11}$) and siRNA on cell growth (two-tailed *t*-test, $n = 3$ biological replicates, $P_{SALL4} = 5.5 \times 10^{-5}$, $P_{DSCC1} = 1.29 \times 10^{-6}$, $P_{SEPHS1} = 2.32 \times 10^{-7}$, $P_{VRTN} = 8.9 \times 10^{-6}$). (***$P < 0.001$ unpaired *t*-test). **b**, Cell viability assay in KBM7 sgRNA-knockout lines for *DSCC1* and *SEPHS1* 4 days after the delivery of sgRNAs and *Cas9*. Control lines received only *Cas9* in the absence of a sgRNA (two-tailed *t*-test, $n = 3$ biological replicates, $P_{DSCC1} = 0.001$, $P_{SEPHS1} = 0.2$). **c**, PCA plot demonstrating the biological replicates of the transcriptome of hESCs with siRNA knockdown for *SALL4*, *DSCC1*, *SEPHS1* and *VRTN* ($n = 60{,}675$ genes). **d**, Percentage of downregulated hESC-enriched transcripts on siRNA knockdown of target genes, divided into different groups of fold reduction. The reduction in expression of pluripotent genes in the cells with knockdown of each of the genes was significant, as calculated by a comparison of the percentage of significantly downregulated genes in the hESC-essential genes to that in control cells (siRNA for *Renilla* luciferase) (two-tailed proportion test, $P_{SALL4} = 0.029$, $P_{DSCC1} = 0.008$, $P_{SEPHS1} = 0.0006$ and $P_{VRTN} = 0.017$). **e**, GO analysis of upregulated genes on siRNA knockdown of target genes ($n_{SALL4} = 922$ genes, $n_{SEPHS1} = 174$ genes, $n_{DSCC1} = 217$ genes, $n_{VRTN} = 539$ genes). Where applicable, data are presented as mean ± s.e.m., and unpaired two-tailed *t*-test was applied (***$P < 0.001$). Source data are provided in Supplementary Table 4.

modelling growth-retardation phenotypes already in hPSCs for a wide group of AR disorders.

Next, we analysed canonical oncogenes and tumour suppressor genes in terms of their essentiality and growth restriction in the context of hESCs[22]. Nearly all oncogenes whose mutations significantly affected the growth of hESCs were classified as essential for normal growth, with the exception of *JUN*, which was found to be growth-restricting (Fig. 2e and Supplementary Fig. 3e). Indeed, c-Jun was shown to interfere with the induction of pluripotency in mouse cells[23]. In contrast, tumour suppressors were divided into essential and growth-restricting gene classes (Fig. 2f and Supplementary Fig. 3f). Gene ontology (GO) analysis revealed that growth-restricting tumour suppressors were enriched in apoptosis-related genes (Fig. 2f, dark blue points), whereas essential tumour suppressor genes were enriched in processes such as genomic instability and DNA repair (Fig. 2f, dark red points). This analysis thus points to distinct roles for tumour suppressor genes in hPSCs.

A comparison of growth restriction by tumour suppressors and the essentiality of oncogenes between hESCs and four cancer cell lines[6] demonstrated that the genetically aberrant lines show marked variation in these genes (Supplementary Fig. 3g). This comparison yielded three groups of genes: (1) genes that were growth-restricting in hESCs but lost this feature in aberrant cells (Supplementary Fig. 3g, left heatmap); (2) genes that are essential in hESCs but lost their essentiality in aberrant lines (Supplementary Fig. 3g, middle heatmap), and (3) genes that were not essential in hESCs but became essential for growth in cancer cells (Supplementary Fig. 3g, right heatmap).

Identification and characterization of the hESC essentialome. The pluripotent state is governed by a set of genes whose expression is enriched in hESCs[24]. Therefore, we hypothesized that hESC-essential genes would be more prevalent within hESC-enriched genes. To test this hypothesis, we performed a gene expression-based analysis in which we divided the genes represented in the library into sub-

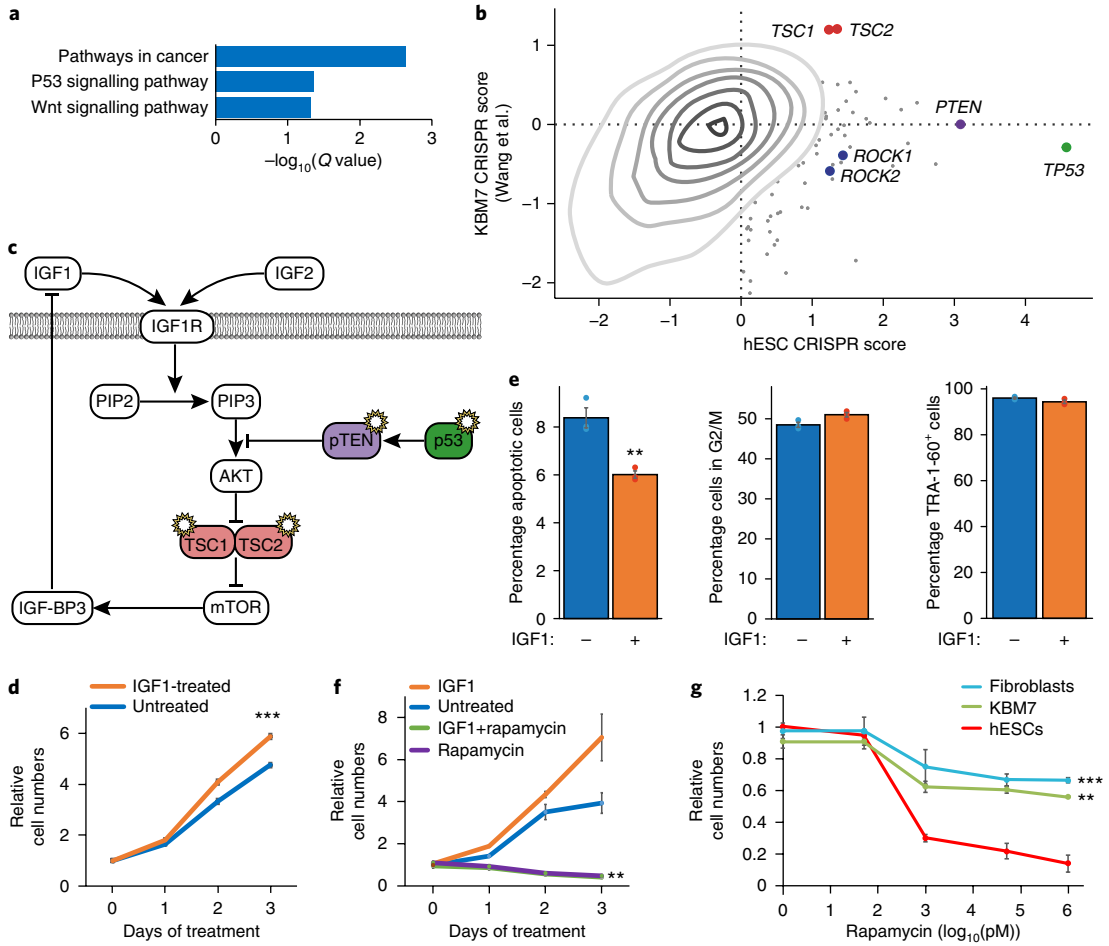

Fig. 5 | Analysis of growth-restricting genes in hESCs. a, GO analysis for the top 50 growth-restricting genes in hESCs ($n=50$ genes). **b**, Comparison of CRISPR scores for all genes from the current haploid hESC screen and a previous screen in the near-haploid leukaemic cell line KBM7[13]. Genes related to P53 and ROCK pathways, which are among the top growth-restricting genes in hESCs, are highlighted. Data from [13] **c**, Schematic representation of the P53–mTOR pathway, highlighting the growth-restricting genes in hESCs. **d**, Growth curves of IGF1-treated (orange) and untreated control (blue) diploid hESCs grown in conditioned medium with 1.2% KSR ($n=4$ biological replicates, $P=1.07\times10^{-5}$). **e**, Percentages of apoptotic cells (left), cells in G2/M phase (middle) and TRA-1-60$^+$ pluripotent cells, in control and IGF1-treated diploid hESCs ($n=3$ biological replicates, $P=0.006$). **f**, Growth curves of untreated control (blue), IGF1-treated (orange), rapamycin-treated (purple) and rapamycin + IGF1-treated (green) diploid hESCs ($n=3$ biological replicates, $P=0.001$). **g**, Relative cell numbers of human foreskin fibroblasts (blue), leukaemic KBM7 cells (green) and diploid hESCs (red) after two days of rapamycin treatment with the indicated concentrations on the x axis ($n=3$ biological replicates, $P_{\text{Fibroblasts}}=8.1\times10^{-5}$, $P_{\text{KBM7}}=4.6\times10^{-3}$). Where applicable, data are presented as mean ± s.e.m., and unpaired two-tailed t-test was applied (**$P<0.01$; ***$P<0.001$). Source data are provided in Supplementary Table 4.

categories related to their expression and enrichment in hESCs[25]. The CRISPR score distribution of hESC-expressed genes shifted towards more depleted values as compared with the distribution of genes that are not expressed (or expressed at low levels) in hESCs (Fig. 3a). The percentage of essential genes increased from 9.2% among all genes to 13.5% among hESC-expressed genes, to 19.8% among genes enriched in hESCs, and up to 22.7% among nuclear hESC-enriched genes (Fig. 3a). This stepwise analysis led to the identification of a subset of hESC-enriched genes, constituting the hESC essentialome (Supplementary Table 3). Of these genes, 67% are nuclear, 18% localize to cytosol, and the remainder are distributed across the cytoskeleton, plasma membrane, extracellular space and Golgi (Fig. 3b). Importantly, the hESC-specific essentialome is significantly depleted of genes localized to mitochondria (HG test, $n=50$ genes, $P=12\times10^{-4}$), even though this compartment was significantly enriched among the cell-essential genes (Fig. 2a). Functional categorization of the hESC-specific essentialome revealed that the majority of genes are related to two main functional groups: cell-cycle and DNA-repair (~53%) and transcription (21%) (Fig. 3c).

Transcription factor (TF) networks have been classically studied within the context of pluripotency[24]. Therefore, from ~2,000 annotated human TFs[25,26], we focused on those that showed enriched expression in hESCs and analysed their CRISPR scores. We found that the hESC-essential TFs include well-characterized pluripotency factors such as *SALL4*, *POU5F1*, *PRDM14* and *NANOG*, as well as *MYBL2*, *FOXB1* and *MYCN*, but not pluripotency-associated factors such as *UTF1* and *ZFP42* (Fig. 3d). Interestingly, we identified a subset of growth-restricting TFs such as *ZNF560*, *RCOR2*, *OTX2* and *ZNF695*. A comparison between hESCs and the leukaemic KBM7 cells for these hESC-essential TFs revealed that these TFs were indeed essential exclusively in hESCs, with the exception of *MYBL2* (Supplementary Fig. 4a). We reasoned that the essentiality difference between the essential and dispensable hESC-enriched TFs might be due to their expression levels in hESCs as compared with immediate progenitor cells. To test this, we compared the

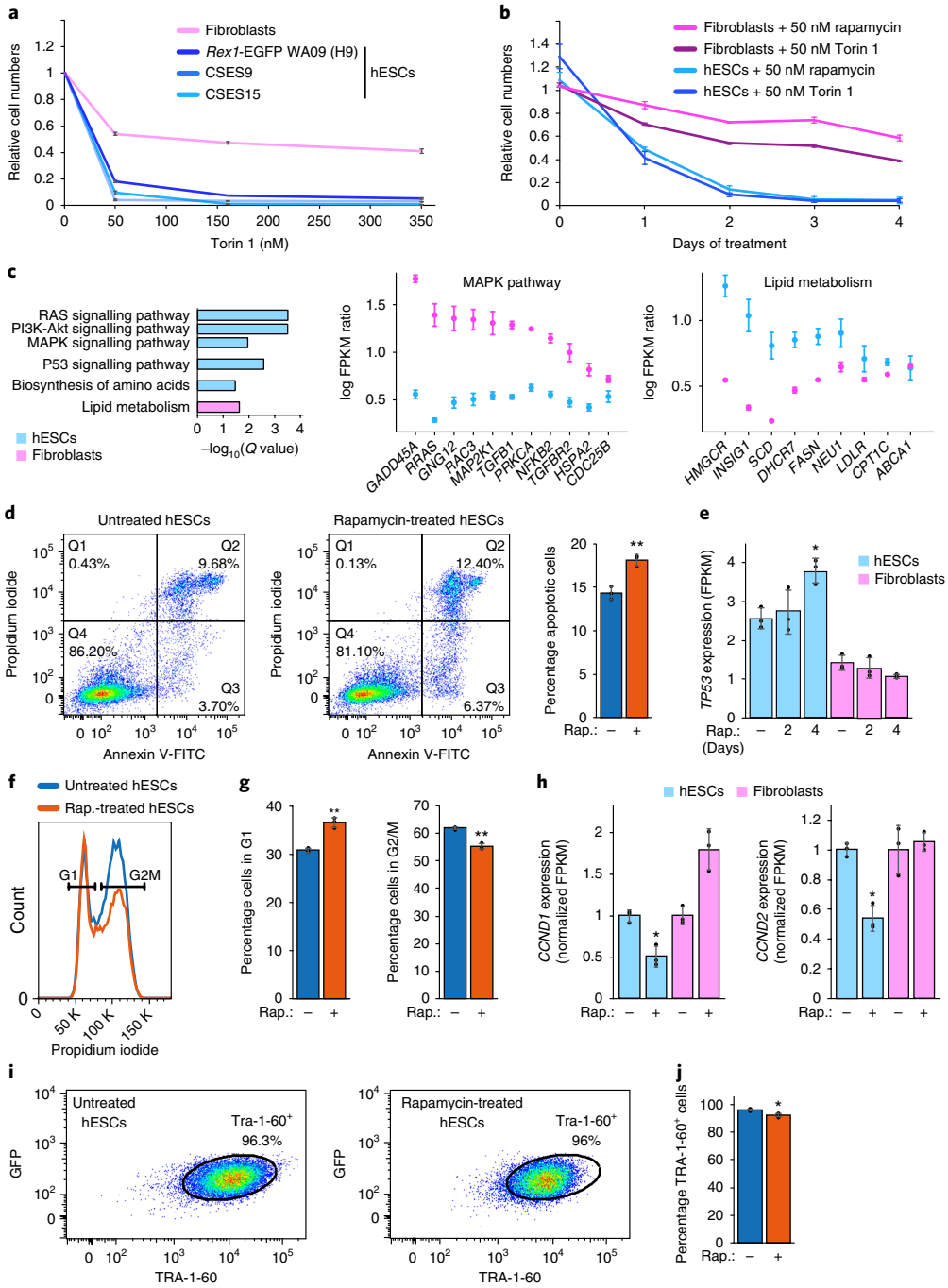

Fig. 6 | Characterization of the selective sensitivity of hESCs to mTOR inhibition. a, Dose response of Torin 1-treated human fibroblasts (pink) and three different hESC lines (shades of blue) at the indicated concentrations on the *x*-axis after two days of treatment ($n = 3$ biological replicates). **b**, Growth curves of rapamycin- and Torin 1-treated fibroblasts (pink) and hESCs (blue). Shown are the values normalized to the untreated controls at the corresponding time points ($n = 3$ biological replicates for each time point). **c**, GO analysis for significantly downregulated genes after two days of rapamycin treatment of hESCs (blue, $n = 593$ genes) and fibroblasts (pink, $n = 243$ genes) (left). Also shown is the relative expression of MAPK pathway members (middle) or lipid metabolism-related genes (right) in rapamycin-treated hESCs (blue) and fibroblasts (pink) for genes expressed in both cell types (FPKM > 1). **d**, Flow cytometry analysis of apoptotic hESCs (diploid CSES9 cell line) following two days of rapamycin treatment. Shown are representative analyses out of three biological replicates of untreated (left) and rapamycin-treated hESCs (middle). Also shown is the percentage of apoptotic cells in untreated and rapamycin-treated (Rap.) hESCs ($n = 3$ biological replicates, $P = 0.009$) (right). **e**, Expression levels of *TP53* in hESCs (blue) and fibroblasts (pink) on rapamycin treatment for the indicated durations. FPKM values were normalized to the expression levels of the corresponding untreated controls in each cell type ($n = 3$ biological replicates, $P = 0.02$). **f**, Cell-cycle analysis of hESCs by flow cytometry following two days of rapamycin treatment. Shown are representative analyses from three biological replicates. **g**, Percentages of hESCs in G1 (left) and G2/M (right) phases in untreated and rapamycin-treated hESCs ($n = 3$ biological replicates, $P_{G1} = 0.004$, $P_{G2/M} = 0.004$). **h**, Expression levels of *CCDN1* and *CCDN2* in hESCs (blue) and fibroblasts (pink) following two days of rapamycin treatment ($n = 3$ biological replicates, $P_{CCDN1} = 0.01$, $P_{CCDN2} = 0.01$). **i**, Analysis of TRA-1-60$^+$ hESCs after four days of rapamycin treatment. Shown are representative analyses from three biological replicates. **j**, Percentage of TRA-1-60$^+$ hESCs in untreated and rapamycin-treated hESCs ($n = 3$ biological replicates, $P = 0.02$). Where applicable, data are presented as mean ± s.e.m., and unpaired two-tailed *t*-test was applied. Source data are provided in Supplementary Table 4.

expression of TFs in hESCs with that in embryoid bodies (EBs) that had undergone differentiation for 20 days (Fig. 3e). Indeed, the EB/hESC expression ratio was found to be smaller for the essential hESC-enriched TFs, suggesting that cell-type-specific gene expression is an important factor in determining essentiality.

Next, we aimed to validate the hESC essentialome we defined in haploid hESCs. To this end, we used both RNA interference (RNAi) and CRISPR–Cas9 mutagenesis in diploid hESCs, focusing on genes that represent different functional categories. siRNA-mediated knockdown, as well as sgRNA-mediated knockout, of the pluripotency-associated TF *SALL4*[27], the DNA replication factor *DSCC1*[28], the selenium metabolism enzyme *SEPHS1*[29] and the putative DNA-binding, nuclear protein *VRTN*[30] inhibited the growth of normal diploid hESCs (Fig. 4a and Supplementary Fig. 4b). Reduction in transcript levels in the siRNA and sgRNA experiments was confirmed for these four genes (Supplementary Fig. 4c,d).

To demonstrate the concept of hESC-specific essential genes, we compared the growth rates of hESCs and the near-haploid leukaemic KBM7 cells mutated in genes that were found as essential in hESCs. We thus chose to analyse four genes that are expressed in both cell types: *DSCC1* and *SEPHS1* (FPKM values were 9.1 and 5.6 for KBM7 and 40.7 and 203.1 for hESCs, respectively), discussed above, and two additional genes, the oncogene *PIK3CA* and the endoplasmic reticulum gene *PDIA4* (FPKM values were 5 and 59.7 for KBM7 and 4.4 and 214.8 for hESCs, respectively). Mutations in either *DSCC1*, *SEPHS1*, *PIK3CA* or *PDIA4* did not perturb the growth of KBM7 cells, but significantly inhibited the growth of hESCs (Fig. 4b and Supplementary Fig. 4e-g).

We then aimed to unravel some of the pathways affected by hESC-essential genes. We thus analysed the transcriptome of cells with knockdown in *SALL4*, *DSCC1*, *SEPHS1* or *VRTN*. PCA based on RNA sequencing demonstrated that inhibition of expression of these genes caused separation from the control siRNA-treated conditions at different degrees, suggesting distinct functions for these genes (Fig. 4c). Importantly, 12–16% of the hESC-enriched genes were downregulated upon knockdown of each of these four genes (Fig. 4d). GO analysis[22] revealed that hESC-essential genes affect different aspects of hESC biology: *SALL4* knockdown upregulated genes related to differentiation to ectodermal brain and trophectodermal placenta cells, suggesting that its inhibition induced differentiation of the pluripotent cells. Interestingly, knockdown of either *SEPHS1* or *DSCC1* affected energy metabolism through oxidative phosphorylation, whereas *VRTN* knockdown showed a modest effect on mitosis (Fig. 4e).

Analysis of growth-restricting genes highlights the role of the P53-mTOR pathway in hESC growth. We next analysed the group of growth-restricting genes in hESCs. GO analysis[31] of the highest-scoring 50 growth-restricting genes showed enrichment in pathways related to cancer, the P53 signalling pathway and Wnt signalling pathway (Fig. 5a). Importantly, the tumour suppressor genes *TP53* and *PTEN* were identified as the highest scoring growth-restricting genes. We then compared the CRISPR scores of all genes in our hESC screen to those in a previous screen performed in the near-haploid leukaemic cell line KBM7[13] (Fig. 5b). This analysis demonstrated that the members of P53 and ROCK pathways, which were identified among the highest scoring growth-restricting genes in hESCs, were absent among the growth-restricting genes in KBM7 cells, suggesting that the P53 pathway may already be mutated in this cancer cell line. Interestingly, among 13 distinct P53 target pathways, we found an enrichment of highest scoring growth-restricting genes of hESCs in the branch inhibiting the IGF1/mTOR pathway (Fig. 5c and Supplementary Fig. 5). Therefore, we aimed to validate the role of this pathway in the regulation of hESC growth. We found that, especially under conditions with low levels of knockout serum replacement (KSR), addition of insulin-like growth factor 1 (IGF1) significantly increased the growth rate of diploid hESCs (Fig. 5d and Supplementary Fig. 6a-d). We reasoned that the IGF1-mediated increase in growth rate could be attributed to the regulation of cell death through apoptosis, to regulation of proliferation rate through changes in the cell-cycle or to differences in spontaneous differentiation. Therefore, we measured the percentage of apoptotic cells, cells in G2/M phase and TRA-1-60^{+}-pluripotent cells after IGF1 treatment, and found that IGF1 significantly decreases the percentage of apoptotic hESCs while not affecting their cell-cycle or differentiation dynamics (Fig. 5e and Supplementary Fig. 6e–g). To demonstrate the direct involvement of mTOR in the growth-regulation of hESCs, we treated hESCs with rapamycin, a selective inhibitor of mTOR, in the presence and absence of IGF1. Inhibition of mTOR caused drastic growth-inhibition of hESCs, and IGF1 failed to rescue this inhibition, suggesting that IGF1 acts upstream of mTOR (Fig. 5f).

hESCs are more sensitive to growth regulation by the mTOR pathway than somatic cells. Interestingly, hESCs were more sensitive to growth inhibition under various doses of rapamycin than human foreskin fibroblasts, as well as KBM7 cells, suggesting that the IGF1/mTOR pathway regulates hESC growth in a cell-type-selective manner (Fig. 5g). mTOR kinase is associated with two major complexes, mTOR complexes 1 and 2 (mTORC1 and mTORC2)[32]. Rapamycin was shown to inhibit mTORC1 but not mTORC2[32] (Supplementary Fig. 6h,i). To assess the contribution of these complexes to the selective sensitivity of hESCs for the mTOR pathway, we utilized a catalytic inhibitor of mTOR, Torin 1, which can inhibit both mTORC1 and mTORC2 (Supplementary Fig. 6h). We found that Torin 1 treatment completely abolished the growth and survival of hESCs after two days of treatment at low nanomolar concentrations, whereas the growth of fibroblasts was inhibited up to 60% under the same conditions (Fig. 6a). Interestingly, inhibition of mTORC1 alone induced dramatic growth inhibition in hESCs, whereas inhibition of both mTORC1 and 2 had only a partial effect on cell growth in actively proliferating fibroblasts (Fig. 6b and Supplementary Fig. 6j,k).

To unravel the molecular mechanism of the sensitivity of hESCs to mTORC1-inhibition, we analysed the transcriptomes of rapamycin-treated hESCs and fibroblasts. We found that in hESCs, two days of rapamycin treatment downregulated genes associated with several growth- or apoptosis-related pathways such as the RAS/PI3K-AKT/MAPK signalling pathways, the P53 signalling pathway, and amino-acid biosynthesis pathways (Fig. 6c). In contrast, in fibroblasts, rapamycin treatment did not significantly downregulate genes enriched in these pathways, but downregulated genes related to lipid metabolism (Fig. 6c). A similar difference in the downregulated pathways between hESCs and fibroblasts was also observed after four days of rapamycin treatment (Supplementary Fig. 6l).

We next examined the effect of mTORC1 inhibition in hESCs at the cellular level, namely in the context of apoptosis, cell-cycle and differentiation. Rapamycin treatment increased the percentage of apoptotic hESCs (Fig. 6d and Supplementary Fig. 6e), accompanied by an increase in the levels of *TP53* (Fig. 6e). Interestingly, *TP53* did not increase in rapamycin-treated fibroblasts. Rapamycin treatment also had a significant effect on the cell-cycle dynamics of hESCs by increasing the percentage of hESCs in G1 phase and decreasing the percentage of hESCs in G2/M phases (Fig. 6f,g and Supplementary Fig. 6f). This observation was supported by the decrease in expression of the G1-S transition factors *CCND1* and *CCND2* (Fig. 6h). In contrast to hESCs, fibroblasts did not downregulate *CCND1* and *CCND2* upon rapamycin treatment. Although the major effects of mTORC1 inhibition were an increase in apoptosis and a decrease in proliferation, probably mediated through a G1-arrest, we also found that rapamycin treatment caused a modest decrease in the fraction of hESCs that were positive for the expression of the pluripotent cell marker TRA-1-60 (Fig. 6i,j and Supplementary Fig. 6g).

Discussion

The recent isolation of haploid hESCs enables a unique way of genome-wide loss-of-function screening in hPSCs. Our results on the cell essentialome unravel interesting aspects of cell biology. We show that essentiality decreases substantially among genes associated with the extracellular matrix and cell membrane, whereas these compartments have higher ratios of growth-restricting genes (Fig. 2b). The low levels of essentiality may suggest that there is a high degree of functional redundancy among these genes, as in the case of growth factor families and their receptors. Interestingly, we identified a single growth factor gene, *TDGF1*, which has an enriched expression in hESCs and is essential for the growth of hESCs, as suggested in previous studies[33].

hPSCs are commonly used to model human genetic disorders[34]. In many cases the cells need to be differentiated into somatic cells in order to analyse the disease phenotypes. Our analysis of all genetic disorders with a growth-retardation phenotype shows that the effect of growth can be documented in 20% of the disorders even in undifferentiated cells, without requiring laborious differentiation protocols (Fig. 2c). The ability to preferentially analyse growth-retardation phenotypes of developmental disorders was specific to hESCs, as it could not be demonstrated in cancer cells (Supplementary Fig. 3d).

Genetic screening in hESCs was also valuable for documenting the essential genes among oncogenes, and to show that tumour suppressor genes comprise both growth-restricting and essential genes (Fig. 2e,f). The comparison of essential genes identified in hESCs and in previously screened cancer cells demonstrated a significant overlap across cell types, with the exception of the phenotypes of tumour suppressor genes and oncogenes (Supplementary Figs. 2a and 3g). We have shown that a group of tumour suppressor genes lost their growth-restricting or essential phenotype in cancer cells, probably because of mutations that directly or indirectly affected their phenotype, as in the case of *TP53* mutations, which appear in the four examined transformed cell lines[35,36]. In addition, certain tumour-related genes became essential for growth in the cancer cells, probably as a result of gain-of-function mutations in these tumours. For example, the *ABL1* and *BRC* genes, which are not essential in normal cells, became essential in two chronic myelogenous leukaemia cells (KBM7 and K562 lines[13]) as a result of a fusion between these two genes, known as the Philadelphia chromosome[37].

The classical description of pluripotency highlights a complex TF network that governs the gene expression profile of this highly versatile cell state[38]. Previous gene expression studies suggested several TFs as markers of pluripotency[24,39]. When combined with such gene expression analyses, our essential gene screen in hESCs reveals that the majority of these pluripotency-associated TFs are dispensable for the growth and survival of hESCs (Fig. 3d). Interestingly, some of these TFs, for example, *ZPF42*, have been previously suggested to also be dispensable for pluripotency in mouse ESCs[40]. We identified seven essential TFs with enriched expression in hESCs. Two of these factors were the oncogenic factors *MYBL2* and *MYCN*. Another oncogenic factor, c-*Myc*, has been used to increase the efficiency of induced pluripotency[41]. Future studies may focus on whether *MYBL2* and *MYCN* can replace c-*Myc* in the reprogramming factor cocktails to yield more authentic induced PSCs.

The finding that the growth of hESCs is regulated in a cell-type-selective manner by the P53-mTOR pathway also highlights key avenues of research in light of a recent report demonstrating that the most prevalent mutations among hPSC lines occur in *TP53*[42] (Fig. 5). Our results suggest that the selective advantage of *TP53* mutations might be overridden by providing chemical mTOR-activators in culture and hence preventing the overgrowth of *TP53* mutants. Conversely, cell-type-selective sensitivity to mTORC1 inhibition may be used to eliminate undesired pluripotent cells from terminally differentiated cultures. The cell-type-specific unusual sensitivity of hESCs to inhibition of mTORC1 seems to be mediated by several pathways and mainly through inhibition of the MAPK pathway.

mTORC1 inhibition in hESCs has also been suggested to cause endoderm and mesoderm differentiation[43]. Although we identified a modest downregulation of the surface expression of TRA-1-60 by mTORC1 inhibition, the most profound effects of this inhibition were on the levels of apoptosis and cell-cycle regulation (Fig. 6d-j).

In conclusion, our approach identified cell-essential and hESC-essential genes in karyotypically stable hPSCs. Our characterization of the hESC essentialome extends the definition of pluripotency beyond the TF-centric view and suggests that genes regulating cell-cycle and DNA repair, which are enriched in hESCs, are also essential for their normal growth and play a vital role in pluripotent cell identity. This present work lays the ground for future studies investigating a broad range of genes essential to human pluripotency, growth regulation in hPSCs and disease modelling using hPSCs.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41556-018-0088-1.

Received: 24 December 2017; Accepted: 20 March 2018;
Published online: 16 April 2018

References

1. Wutz, A. Haploid mouse embryonic stem cells: rapid genetic screening and germline transmission. *Annu. Rev. Cell Dev. Biol.* **30**, 705–722 (2014).
2. Yilmaz, A., Peretz, M., Sagi, I. & Benvenisty, N. Haploid human embryonic stem cells: half the genome, double the value. *Cell Stem Cell* **19**, 569–572 (2016).
3. Sagi, I. & Benvenisty, N. Haploidy in humans: an evolutionary and developmental perspective. *Dev. Cell* **41**, 581–589 (2017).
4. Tarkowski, A. K., Witkowska, A. & Nowicka, J. Experimental parthenogenesis in the mouse. *Nature* **226**, 162–165 (1970).
5. Leeb, M. & Wutz, A. Derivation of haploid embryonic stem cells from mouse embryos. *Nature* **479**, 131–134 (2011).
6. Elling, U. et al. Forward and reverse genetics through derivation of haploid mouse embryonic stem cells. *Cell Stem Cell* **9**, 563–574 (2011).
7. Yang, H. et al. Generation of genetically modified mice by oocyte injection of androgenetic haploid embryonic stem cells. *Cell* **149**, 605–617 (2012).
8. Li, W. et al. Androgenetic haploid embryonic stem cells produce live transgenic mice. *Nature* **490**, 407–411 (2012).
9. Li, W. et al. Genetic modification and screening in rat using haploid embryonic stem cells. *Cell Stem Cell* **14**, 404–414 (2014).
10. Yang, H. et al. Generation of haploid embryonic stem cells from *Macaca fascicularis* monkey parthenotes. *Cell Res.* **23**, 1187–1200 (2013).
11. Sagi, I. et al. Derivation and differentiation of haploid human embryonic stem cells. *Nature* **532**, 107–111 (2016).
12. Carette, J. E. et al. Haploid genetic screens in human cells identify host factors used by pathogens. *Science* **326**, 1231–1235 (2009).
13. Wang, T. et al. Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
14. Blomen, V. A. et al. Gene essentiality and synthetic lethality in haploid human cells. *Science* **350**, 1092–1096 (2015).
15. Shalem, O. et al. Genome-scale CRISPR–Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
16. Hart, T. et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526 (2015).
17. Hart, T. et al. Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. *Genes Genomes Genet.* **7**, 2719–2727 (2017).
18. *OMIM—Online Mendelian inheritance in man* (John Hopkins University School of Medicine, accessed 26 April 2017); https://www.omim.org/
19. Raya, Á. et al. Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. *Nature* **460**, 53–59 (2009).
20. Tulpule, A. et al. Knockdown of Fanconi anemia genes in human embryonic stem cells reveals early developmental defects in the hematopoietic lineage. *Blood* **115**, 3453–3462 (2010).
21. Henske, E. P., Jóźwiak, S., Kingswood, J. C., Sampson, J. R. & Thiele, E. A. Tuberous sclerosis complex. *Nat. Rev. Dis. Prim.* **2**, 16035 (2016).
22. Walker, E. J. et al. Monoallelic expression determines oncogenic progression and outcome in benign and malignant brain tumors. *Cancer Res.* **72**, 636–644 (2012).

23. Liu, J. et al. The oncogene c-Jun impedes somatic cell reprogramming. *Nat. Cell Biol.* **17**, 856–867 (2015).
24. De Los Angeles, A. et al. Hallmarks of pluripotency. *Nature* **525**, 469–478 (2015).
25. GTEx Consortium. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
26. Ravasi, T. et al. An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* **140**, 744–752 (2010).
27. Zhang, J. et al. Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nat. Cell Biol.* **8**, 1114–1123 (2006).
28. Merkle, C. J., Karnitz, L. M., Henry-Sanchez, J. T. & Chen, J. Cloning and characterization of hCTF18, hCTF8, and hDCC1: human homologs of a *Saccharomyces cerevisiae* complex involved in sister chromatid cohesion establishment. *J. Biol. Chem.* **278**, 30051–30056 (2003).
29. Low, S. C., Harney, J. W. & Berry, M. J. Cloning and functional characterization of human selenophosphate synthetase, an essential component of selenoprotein synthesis. *J. Biol. Chem.* **270**, 21659–21664 (1995).
30. Lenz, M. et al. Epigenetic biomarker to support classification into pluripotent and non-pluripotent cells. *Sci. Rep.* **5**, 8973 (2015).
31. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
32. Guertin, D. A. & Sabatini, D. M. The pharmacology of mTOR inhibition. *Sci. Signal.* **2**, pe24 (2009).
33. Fiorenzano, A. et al. Cripto is essential to capture mouse epiblast stem cell and human embryonic stem cell pluripotency. *Nat. Commun.* **7**, 12589 (2016).
34. Avior, Y., Sagi, I. & Benvenisty, N. Pluripotent stem cells in disease modelling and drug discovery. *Nat. Rev. Mol. Cell Biol.* **17**, 170–182 (2016).
35. Sen, S., Takahashi, R., Rani, S., Freireich, E. J. & Stass, S. A. Expression of differentially phosphorylated Rb and mutant p53 proteins in myeloid leukemia cell lines. *Leuk. Res.* **17**, 639–647 (1993).
36. ATCC. ATCC cell lines by gene mutation (*American Type Culture Collection,* Manassas, 2018); http://bit.ly/2GApeiM
37. Faderl, S. et al. The biology of chronic myeloid leukemia. *N. Engl. J. Med.* **341**, 164–172 (1999).
38. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**, 1049–1061 (2008).
39. Assou, S. et al. A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. *Stem Cells* **25**, 961–973 (2007).
40. Masui, S. et al. Rex1/Zfp42 is dispensable for pluripotency in mouse ES cells. *BMC Dev. Biol.* **8**, 45 (2008).
41. Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
42. Merkle, F. T. et al. Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations. *Nature* **545**, 229–233 (2017).
43. Zhou, J. et al. mTOR supports long-term self-renewal and suppresses mesoderm and endoderm activities of human embryonic stem cells. *Proc. Natl Acad. Sci. USA* **106**, 7840–7845 (2009).

Acknowledgements

The authors thank E. Meshorer and all members of The Azrieli Center for Stem Cells and Genetic Research for their input and critical reading of the manuscript. The authors also thank O. Yanuka, T. Golan-Lev and A. Petcho for assistance with tissue culture. This work was partially supported by the US–Israel Binational Science Foundation (grant no. 2015089), by the Israel Science Foundation (grant no. 494/17) and by the Azrieli Foundation. A.Y. is supported by the Lady Davis Postdoctoral Fellowship. I.S. is supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities, and N.B. is the Herbert Cohn Chair in Cancer Research.

Author contributions

A.Y., M.P. and N.B. designed the experiments, interpreted the data and wrote the manuscript, with input from all authors. A.Y. performed the experiments and analysed the data. M.P. performed the bioinformatics analyses. A.A. assisted with the bioinformatics analyses. I.S. assisted with the characterization of haploid hESCs. N.B. supervised the study.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41556-018-0088-1.

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

Correspondence and requests for materials should be addressed to N.B.

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

Methods

Cell lines, vectors and reagents. The following cell lines were used in this study—haploid hESCs: h-pES10 cell line, recently isolated by us[11]; *REX1*-EGFP cells: hESCs carrying the eGFP gene under the *REX1* promoter[44]; 293T cells: obtained from R. Weinberg (Whitehead Institute); BJ human fibroblasts: purchased from Clontech; KBM7 cells: purchased from Horizon Discovery. The activity-optimized Human CRISPR Pooled Library (a gift from D. Sabatini and E. Lander, cat. no. 1000000067), pCMV-VSV-G (a gift from B. Weinberg, cat. no. 8454), psPAX2 (a gift from D. Trono, cat. no. 12260) and LentiCRISPR v2 (a gift from Feng Zhang, cat. no. 52961) were purchased from Addgene. IGF1 was obtained from PeproTech. Rapamycin was obtained from Cayman Chemical Company. Torin 1 was obtained from Cell Signaling.

Cell culture. Haploid hESCs were cultured at 37 °C and 5% CO_2 on feeder layer growth-arrested mouse embryonic fibroblasts (MEFs) in standard hESC growth medium, composed of knockout Dulbecco's modified Eagle's medium (DMEM) supplemented with 15% knockout serum replacement (KSR, Thermo Fisher Scientific), 2 mM L-glutamine, 0.1 mM nonessential amino acids, 50 units ml^{-1} penicillin, 50 µg ml^{-1} streptomycin, 0.1 mM β-mercaptoethanol and 8 ng ml^{-1} basic fibroblast growth factor (bFGF). Cells were passaged by a quick trypsinization using trypsin-EDTA (Biological Industries) and plated in the presence of 10 µM ROCK inhibitor Y-27632 (Stemgent) for 1 day after splitting. BJ fibroblasts, feeder layer MEFs and 293T cells were cultured in DMEM supplemented with 10% fetal bovine serum (Invitrogen), 2 mM L-glutamine, 50 units ml^{-1} penicillin and 50 µg ml^{-1} streptomycin. *REX1*-EGFP WA09, CSES9 and CSES15 hESC lines were cultured in feeder-free conditions on matrigel-coated plates (Corning) in mTeSR1 (STEMCELL Technologies). KBM7 cells were cultured in Iscove's modified Dulbecco's medium (IMDM) supplemented with 10% fetal bovine serum (Invitrogen), 2 mM L-glutamine, 50 units ml^{-1} penicillin and 50 µg ml^{-1} streptomycin. Cell lines were free of mycoplasma.

Enrichment of haploid hESCs. Haploid hESCs were enriched as described previously[45]. Briefly, cells were washed with phosphate buffered saline (PBS), trypsinized using TrypLE Select (Thermo Fisher Scientific) and stained with 10 µg ml^{-1} Hoechst 33342 (Sigma Aldrich) in hESC growth medium at 37 °C for 30 min. Cells were then centrifuged and resuspended in PBS containing 10% KSR and 10 µM ROCK inhibitor Y-27632, filtered through a 70 µm cell strainer (Corning) and sorted by a 405 nm laser in BD FACSAria III (BD Biosciences). On plating the sorted cells, 10 µM ROCK inhibitor Y-27632 was added to the medium for one day.

Library plasmid amplification, virus production and transduction of haploid hESCs. sgRNA library cloned into *Cas9*-containing lentiCRISPR v1 plasmids[13] was transformed into Endura electrocompetent cells (Lucigen). Transformed cells were plated on ampicillin-containing agar plates (Sigma) and used for plasmid isolation. To maintain the diversity of the sgRNA library, more than 100-fold coverage of the size of the sgRNA library was achieved in the number of transformed colonies (>18 million colonies).

To produce virus library for 181,131 sgRNAs, 293T cells in forty 15 cm culture plates with around 70–80% confluency were transfected with sgRNA-containing lentiCRISPR v1, pCMV-VSV-G and psPAX2 plasmids at a ratio of 13.3:6.6:10 (30 µg total per plate), respectively, in the presence of polyethylenimine 'Max' (Polysciences). Transfection medium was exchanged with 0.5% BSA-containing 293T growth medium after 16 h, and lentiviral particle-containing culture supernatant was harvested 60–65 h after transfection. Culture supernatant was spun down at 3,000 r.p.m. for 10 min at 4 °C and then filtered through 0.45 µm cellulose acetate filters (Millipore). Filtered supernatant was centrifuged in a swing bucket rotor (Beckman Coulter) at 24,000 r.p.m. for 2 h at 4 °C. The pellet was very briefly dried, then reconstituted in cold hESC growth medium (<1 ml) and frozen in aliquots at −70 °C. Virus titres were measured as described previously[15]. A total of 378 million haploid-enriched hESCs were transduced with the virus library at a multiplicity of infection (MOI) of 0.3, resulting in infection of 30% of the cells and hence leading to a 700-fold coverage of the sgRNA library size. An MOI of 0.3 ensures a high enrichment in the proportion of cells that are infected with only one viral particle and therefore carry a single mutation. Haploid hESCs can be sorted to purity from a mixed population of haploid and diploid cells[1,27]. We transduced hESCs a week after haploid cell enrichment, when about 90% of the cells are still haploid. After the introduction of sgRNAs, diploidization would lead to the generation of homozygous mutations through endoduplication of mutant cells, and hence it is not expected to affect the effectiveness of loss-of-function mutations.

For transduction, haploid hESCs were trypsinized with trypsin-EDTA, centrifuged and resuspended in hESC growth medium supplemented with 10 µM ROCK inhibitor Y-27632 and 8 µg ml^{-1} polybrene (Sigma). The viruses were then added to the cell suspension. Transduced haploid hESCs were densely plated on feeder layer MEFs overnight (3 million cells in 1.5 ml hESC medium for one well of a six-well plate). At 24 h after transduction, cells were passaged on a feeder layer of DR3 MEFs at a ratio of 1:3 in the presence of 5 µM ROCK inhibitor Y-27632. During this passaging, 35 million cells were harvested for DNA extraction and sgRNA analysis for the 'Day 1' time point after infection.

At 12 h after this passaging, the medium of the cells was replaced with puromycin-containing medium (0.3 mg ml^{-1}, Sigma). Cells were kept under puromycin selection for 7 days and then passaged again. At 9 days after initial transduction, 55 million cells (300-fold the size of the sgRNA library) were collected and mixed for DNA extraction and sgRNA analysis for the 'Day 9' time point after infection. Transduced haploid hESCs were passaged every 4–5 days while maintaining at least 400-fold representation of the sgRNA library. Fifty-five million cells were also collected for each time point 'Day 23' and 'Day 30' after infection.

DNA extraction, PCR amplification of sgRNAs and high-throughput DNA sequencing. Genomic DNA was extracted with a Blood & Cell Culture DNA Midi Kit (QIAGEN) according to the manufacturer's instructions. The region containing the sgRNA integration was amplified with the following primers, which also contain overhang sequences compatible for Nextera DNA library preparations (Illumina):

5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGCTTTATA-TATCTTGTGGAAAGGACG-3′ (forward) and 5′-GTCTCGTGGGCTCGGAGAT GTGTATAAGAGACAGACGGACTAGCCTTATTTTAACTTGC-3′ (reverse).

The total genomic DNA for each time point was divided into 50 µl PCR reactions with 4 µg DNA input. The PCR settings have been described previously[46]. After purification of the 160-base-pair (bp) product, a second PCR reaction was performed using Nextera adapter primers to generate a Nextera DNA library according to the manufacturer's instructions (Illumina). DNA libraries containing sgRNA constructs from two replicate experiments were sequenced using NextSeq 500 (Illumina).

Data analysis. The numbers of reads obtained from sequencing were 80.5 million and 68.6 million reads for day 1, 71.5 million reads for day 9, 84.1 million and 70.4 million reads for day 23, and 43.2 million reads for day 30 after introduction of the gRNAs. Mapping was performed by aligning the sgRNA sequences to the reads (treating the reads as a reference genome) using the bowtie2 program[47], and analysing only complete 20-base matches. The minimal number of reads mapped to any of the sgRNAs under any of the conditions was 4. The count table was then normalized relative to the total number of reads in each of the conditions, and replicates were averaged. CRISPR scores are the average of the $\log_2$ ratios of the abundance of all sgRNAs for each gene between final (day 23) and initial (day 1) populations (Supplementary Fig. 1b). Statistical significance was determined by the Kolmogorov–Smirnov test for two samples, using ks_2samp from python's scipy.stats module. In doing so, each population of sgRNAs belonging to a gene was compared to the general distribution of sgRNAs from the same condition. The Benjamini–Hochberg FDR correction was accomplished with the multipletests feature from python's statsmodels.sandbox.stats.multicomp module.

Comparison of genetic screens in haploid versus diploid cells. Induction of loss of function, using the CRISPR–Cas9 methodology, may be more efficient in haploid than diploid cells for several reasons. The Cas9 endonuclease guided by the sgRNA creates a double-strand break that can lead to nucleotide insertions or deletions (indels) due to the non-homologous-end-joining mechanism. In the majority of cases, these indels will be one or two nucleotides, creating a frameshift and a loss of function in the allele. However, in some cases, allelic loss of function does not occur, mainly due to indels in multiples of three nucleotides preserving the reading frame, but also due to implementation of the homologous-recombination repair mechanism or the lack of a double-strand-break reaction. Let L be the allelic loss-of-function rate, then, for a given sgRNA, the probability of successfully targeting both alleles in a diploid cell is L^2, while in a haploid cell this rate is L. Hence, for a specific sgRNA, we would expect the loss-of-function rate to be $\frac{L}{L^2}=\frac{1}{L}$ times higher in haploids than in diploids. Thus, if indels in the multiples of three nucleotides occur in a third of the mutations, frameshift mutations will occur 50% more frequently in a haploid allele than in diploid alleles.

The assumption that complete loss-of-function alleles are more prevalent in haploid than in diploid chromosomes is also supported by data from the near-haploid KBM7 cells, where chromosome 8 is the only full diploid chromosome[6]. In the analysis in ref. [6], the first percentile CRISPR scores in diploid chromosome 8 and the other haploid autosomes in KBM7 cells were determined. As shown in Supplementary Fig. 3c, chromosome 8 shows significantly different values than the other chromosomes, suggesting that it is more efficient to achieve loss-of-function mutation in haploid chromosomes. Analysing values for chromosome 8 in three all-diploid cell lines[6] (K562, Jiyoye and Raji) or in the all-haploid cell line (hESCs), showed that chromosome 8 is not different from the other autosomes (Supplementary Fig. 3c,d). The data support the notion that although genetic screens using CRISPR–Cas9 technology are fairly efficient in diploid cells, the use of haploid cells provides a further advantage in generating complete loss-of-function frameshift mutations.

Analysis of cellular compartments. Localization of proteins into cellular compartments was retrieved from the Subcellular Localization Database[48] website (http://compartments.jensenlab.org/About), where each gene is given a number of compartments with matching confidence level scores. For each of the genes we defined the maximal confidence score, and assigned it with the compartments.

We then analysed the genes that were associated with a single compartment. Among 18,099 genes in our study, 17,242 (95%) were assigned to compartments and 10,932 (63%) were associated with a single compartment. Statistical significance for the enrichment of nuclear and mitochondrial compartments among essential genes was assessed by the hypergeometric test.

Analysis of AR disorders and cancer-related genes. To analyse the involvement of genes responsible for genetic disorders in the growth of hESCs we utilized the database of the Online Mendelian Inheritance in Man (OMIM)[18] (https://www.omim.org/), which lists diseases associated with genes, their pattern of inheritance, and their clinical synopsis. Of the annotated genes in OMIM, gRNAs for 3,592 genes appear in the library, and 2,099 of them have an AR inheritance. Of these AR inheritance genes, 766 also had a growth retardation-related phenotype. The FA genes are one example of such genes: *FANCA* (FA-A), *FANCB* (FA-B), *FANCC* (FA-C), *BRCA2* (FA-D1), *FANCD2* (FA-D2), *FANCE* (FA-E), *FANCF* (FA-F), *FANCG* (FA-G), *FANCI* (FA-I), *BRIP1* (FA-J), *FANCL* (FA-L), *FANCM* (FA-M), *PALB2* (FA-N), *RAD51C* (FA-O) and *SLX4* (FA-P).

The list of canonical oncogenes and tumour suppressor genes were retrieved from a previous study[13].

Analysis of hESC-enriched genes. Defining genes with enriched expression in hESCs was performed by comparing expression data from 26 tissues to that of 10 hESC lines from four different studies (SRR2038465, SRR2038466, SRR2038467, SRR2038469, SRR2038474, SRR2038475, SRR2038476, SRR2038477, SRR2453342, SRR2453346, SRR2453356, SRR2453360, SRR2453365, SRR2453368, SRR2453370, SRR3382655, SRR3575052). The tissue expression data was retrieved from the GTEx Portal version V6p (GTEx_Analysis_v6p_RNA-seq_RNA-SeQCv1.1.8_gene_reads.gct)[25], normalized together with the data from hESC lines, and averaged over similar tissues. Overall, the average expression data represented the following number of samples of each category: 17 hESCs, 285 transformed fibroblasts, 119 transformed lymphocytes, 815 brain, 689 oesophagus, 609 skin, 579 adipose, 431 skeletal muscle, 414 heart, 394 whole blood, 347 colon, 324 thyroid, 321 lung, 605 nerve, 215 breast, 194 stomach, 172 pancreas, 146 adrenal gland, 120 liver, 107 prostate, 105 spleen, 104 pituitary, 89 small intestine, 84 uterus, 58 salivary gland, 33 kidney, 12 bladder. For a gene to be considered as expressed in hESC, its averaged FPKM level has to surpass 1. Genes that were considered enriched in hESC are at least 10 times more expressed in hESC than any other tissue and at least 3 times more expressed relative to transformed cell lines.

Comparison between hESC and EB expression was performed using expression data from hESC and EB samples from the same haploid cell line (SRR2131924, SRR2131925, SRR2131926, SRR2131927, SRR2131929, SRR2131937).

siRNA knockdown, cell viability and growth curve assays. All genes were targeted with commercial pooled siRNAs to increase the specificity of the knockdown. esiRNAs for *SALL4*, *DSCC1* and *SEPHS1* and the control esiRNA for Renilla luciferase were obtained from Sigma (cat. nos. EHU037061, EHU021301, EHU107861 and EHURLUC, respectively). siRNA for *VRTN* was obtained from GE Healthcare Dharmacon (cat. no. L-021159-02-0005). Briefly, 30–50 nM siRNA was mixed with 0.14 µl of DharmaFECT 1 transfection reagent (GE Healthcare Dharmacon) in Opti-MEM I reduced serum medium (Thermo Fisher Scientific) for 30 min. The transfection mix was added on a matrigel-coated well of a 96-well plate and 3,000 hESCs were plated on the transfection mix in the presence of mTeSR 1 medium. The cells were subsequently grown for 3–4 days and the mTeSR 1 medium was replaced every 24 h. Cell viability was assessed by a CellTiter-Glo luminescent cell viability assay according to the manufacturer's instructions (Promega). Luminescence reads for the target genes were normalized to control siRNA conditions, and the replicate experiments were averaged. For growth curves of IGF1-treated and control hESCs, cells were plated in matrigel-coated wells with equal numbers and their density was measured for four consecutive days with CellTiter-Glo luminescent cell viability assay (Promega). The density measured one day after plating was considered Day 0, after which cells were switched to MEF-conditioned medium containing 100 ng ml^{-1} IGF1. Every day was normalized to Day 0 and replicate experiments were averaged. For rapamycin and Torin 1 treatment experiments, cells were cultured in mTeSR 1 medium and the drug-containing medium was replaced every 24 h.

Generation of stable sgRNA cell lines. sgRNA sequences used for cloning into lentiCRISPR v2 lentiviral vector were as follows:
5′-GCGCTCTTCAGATCCACGAG-3′ for *SALL4*,
5′-GCAGAGTGTTCCTGAAGGAA-3′ for *DSCC1*,
5′-CACGTGGTAAACAGATCAGA-3′ for *SEPHS1*,
5′-GCACTGGCGGTGTCAAGCCC-3′ for *VRTN*,
5′-ACAGCCACACACTACATCAG-3′ for *PIK3CA* and
5′-GCTGGCCAGCACAGACGCTG-3′ for *PDIA4A*. Viruses containing these constructs were packaged as described above. For the control lines, lentiCRISPR v2 vector without any sgRNA was used. REX1-EGFP hESC and KBM7 cells were transduced with the viral supernatant and the transduced cells were selected with puromycin (0.3 mg ml^{-1} for hESCs, 2-4 mg ml^{-1} for KBM7) one day after infection. Three days after infection the cells were plated on 96-well plates for the cell viability assay, and collected for analysis of the wild-type transcript levels of targeted genes. Cell viability was assessed by CellTiter-Glo luminescent cell viability assay (Promega) four days after infection.

RNA isolation, RNA sequencing and quantitative real-time PCR. For high-throughput RNA sequencing experiments with siRNA knockdown, cells were collected 48 h after transfection of siRNAs for *SALL4* and *VRTN* and 72 h for *DSCC1* and *SEPHS1*. Total RNA was isolated from three independent biological replicates with RNeasy Mini or Micro Kit (QIAGEN) and the mRNA fraction of total RNA was enriched by pulldown of poly(A)-RNA. RNA sequencing libraries were generated using SENSE Total RNA-Seq Library Prep Kit (LEXOGEN) according to the manufacturer's protocol and sequenced using Illumina NextSeq 500 with 85 bp single-end reads. For rapamycin experiments, RNA sequencing libraries were generated using the Illumina TruSeq RNA prep kit v2 according to the manufacturer's protocol and sequenced using Illumina NextSeq 500 with 84 bp single-end reads. For transcriptome analysis, reads were mapped to the GRCh38 reference genome using STAR. Statistical significance was then determined by two-tailed unpaired Student's *t*-test, and GO enrichment analysis was done by DAVID. Statistical significance was determined using the Benjamini correction.

For qRT–PCR experiments, total RNA was reverse-transcribed into first-strand complementary DNA (cDNA) (Quantabio). The qRT–PCR reaction consisted of initial incubation at 50 °C for 2 min and denaturation at 95 °C for 10 min. The cycling parameters were as follows: 95 °C for 15 s and 60 °C for 30 s. After 40 cycles, the reactions underwent a final dissociation cycle as follows: 95 °C for 15 s, 60 °C for 1 min, 95 °C for 15 s and 60 °C for 15 s. Expression of each gene was normalized to *GAPDH* expression. The primer sequences used in qRT–PCR reactions to test the siRNA knockdowns were as follows: 5′-TTGAGGGGAGATGGGTACTG-3′ and 5-AATAAGATGGGGACAGGGTTG-3′ for *SALL4*, 5′-TTAGCCTTCCACCCAAACTG-3′ and 5-TCCCAAAGCGCATGTCTAC-3′ for *DSCC1*, 5′-AGGCATTACCCGTAGTCGTG-3′ and 5-TCCAGAAAACCATTCAAGACG-3′ for *SEPHS1*, 5′-TGAGGCACTGGAGATCACTG-3′ and 5-GGGCCATAATCTGCAAACAG-3′ for *VRTN*, 5-AGCCACATCGCTCAGACACC-3′ and 5-GTACTCAGCGCCAGCATCG-3′ for *GAPDH*. qRT–PCR primers to detect the wild-type transcript levels in stable sgRNA lines were designed to have their 3′ ends around the Cas9 cut-site of the genes of interest and were as follows: 5′-GCGCTCTTCAGATCCACGAG-3′ and 5-CCCGTGTGTCATGTAGTGAAC-3′ for *SALL4*, 5′-GCAGAGTGTTCCTGAAGGAA-3′ and 5-CTCAGGTAAATCATCTACTTTCAGC-3′ for *DSCC1*, 5′-GAGGAACGAGGTGTCGTTTG-3′ and 5-CACGTGGTAAACAGATCAGA-3′ for *SEPHS1*, 5′-GCACTGGCGGTGTCAAGCCC-3′ and 5-ATAAGTGGACCGTGAGATGC-3′ for *VRTN*, 5′-ACAGCCACACACTACATCAG-3′ and 5-TTGTGACGATCTCCAATTCC-3′ for *PIK3CA* and 5′-GCAGTTTGCTCCGGAATATG-3′ and 5-GCTGGCCAGCACAGACGCTG-3′ for *PDIA4*.

TRA-1-60 immunocytochemistry. hESCs were trypsinized with TrypLE Select. Cells were collected in cold PBS containing 10% KSR, centrifuged at 300*g* for 5 min and resuspended in 200 µl PBS containing 10% KSR. PE-conjugated mouse anti-human TRA-1-60 antibody (1:40, BD Biosciences) was then incubated with the cells for 30 min at 4 °C. Cells were washed with PBS containing 10% KSR twice, centrifuged at 300*g* at 4 °C and resuspended in PBS with 10% KSR. Immunolabelled cells were filtered through a 70 µm cell strainer and analysed in BD FACSAria III for the proportion of TRA-1-60-positive cells.

Apoptosis assay. hESCs were trypsinized gently with TrypLE Select. Apoptotic cells were labelled with Annexin V and propidium iodide (PI) using the MEBCYTO Apoptosis Kit according to the manufacturer's instructions (MBL). Labelled cells were filtered through a 70 µm cell strainer and analysed in BD FACSAria III for the proportion of annexin V-positive cells.

PI staining. PI staining was performed as described previously[45]. Briefly, hESCs were trypsinized with TrypLE Select and fixed with cold methanol (4 °C). Fixed cells were treated with 200 µg ml^{-1} RNase A (Sigma) for 30 min and stained with 50 µg ml^{-1} PI for 5 min. Stained cells were filtered through a 70 µm cell strainer and analysed in BD FACSAria III for their cell-cycle profile.

Western blotting. hESCs were washed with PBS, lysed in sample buffer (100 nM Tris at pH 6.8, 200 mM dithiothreitol (DTT), 4% SDS, 0.2% bromophenol blue, 20% glycerol) and boiled for 5 min. Total protein originating from an equal number of cells was separated by 12% SDS–PAGE and transferred to a nitrocellulose membrane (Pall Corporation). Membranes were blocked with 8% BSA for 1.5 h at room temperature, sliced into strips for each primary antibody at the corresponding molecular weight ranges, and incubated with the primary antibodies (in TBS-T with 5% BSA) overnight at 4 °C. Membranes were washed three times with TBS-T and incubated with the secondary antibody (in TBS-T with 5% BSA) for an hour at room temperature. Following the three washes with TBS-T, membranes were incubated with EZ-ECL (Biological Industries,

cat. no. 20-500-120). Signals were detected using X-ray films (Fujifilm, cat. no. 47410). Working dilutions of the primary and secondary antibodies were as follows: anti-phospho-AKT antibody at 1:1,000 (Cell Signaling Technology, cat. no. 4060), anti-GAPDH antibody at 1:30,000 (Cell Signaling Technology, cat. no. 21185), anti-phospho-4E-BP1 antibody at 1:1,000 (Cell Signaling Technology, cat. no. 23684) and anti-rabbit-HRP at 1:5,000 (Santa Cruz Biotechnology, cat. no. SC-2004).

Data reporting. No statistical methods were used to predetermine the sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

Code availability. All custom scripts used in this study are available from the corresponding author on reasonable request.

Statistics and reproducibility. Statistical analysis was performed using Python, R and Microsoft Office Excel. Data are presented as mean-centred and with the standard error of the mean. An unpaired two-tailed *t*-test was performed for comparisons of two groups unless otherwise stated. FDR was controlled using the Benjamini–Hochberg correction using $P < 0.05$ as statistical significance. Statistics source data for the repeats are shown in Supplementary Table 4. The exact *P* values and number of replicates per condition are stated in the figure legends with the statistical method used. Data presented in Figs. 1b–e, 2a–f, 3a–d and 5a,b, Supplementary Figs. 1a,d, 2a–d, 3a,b,d–g, 4a,b and 5 were derived by averaging two independent genome-wide screens with strongly correlated results ($r = 0.88$, Supplementary Fig. 1c). All other experiments were repeated with at least three independent biological repeats, unless otherwise stated.

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

Data availability. RNA-seq data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession codes GSE103846 and GSE107965. DNA-seq data that support the findings of this study have been deposited in the GEO database under accession code GSE111309. Previously published sequencing data that were re-analysed here are available under accession codes GSE62772, GSE73211, GSE80264 and GSE81791 and from the GTEx Portal version V6p at link https://www.gtexportal.org/home/datasets. Registered users can access the files using the dbGaP accession no. phs000424.v6.p1. CRISPR score tables and significance values are provided in Supplementary Tables 1–3. Source data for Figs. 3, 4, 5 and 6 and Supplementary Figs. 4 and 6 are provided as Supplementary Table 4. All other data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

44. Eiges, R. et al. Establishment of human embryonic stem cell-transfected clones carrying a marker for undifferentiated cells. *Curr. Biol.* **11**, 514–518 (2001).
45. Sagi, I., Egli, D. & Benvenisty, N. Identification and propagation of haploid human pluripotent stem cells. *Nat. Protoc.* **11**, 2274–2286 (2016).
46. Wang, T., Lander, E. S. & Sabatini, D. M. Viral packaging and cell culture for CRISPR-based screens. *Cold Spring Harb. Protoc.* http://doi.org/gcw7js (2016).
47. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
48. Binder, J. X. et al. COMPARTMENTS: unification and visualization of protein subcellular localization evidence. *Database* **2014**, bau012 (2014).

nature research

Corresponding author(s): Nissim Benvenisty

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	DNA and RNAseq data were collected using bcl2fastq software v2.18.0.12. qRT-PCR data were collected using the Applied Biosystems 7300 System Software v1.4.0. Optical density measurements were collected using the Gen5 Software v2.01.14. Flow cytometry data was collected using the BD FACSDiva Software v8.0.1.
Data analysis	Python version 3.5, HTSeq version 0.9.1 and Bowtie2 version 2.3.4.1 have been used for the analysis of RNA and DNA sequencing analysis. FlowJo software Version 7.6.5 has been used for the analysis of flow cytometry experiments. Microsoft Office Excel (Microsoft Office Professional Plus 2016) and R Studio version 3.4.3 were used to calculate mean, standard deviation, P value and to perform statistical analyses. The custom scripts used in this study are available on request from the corresponding author.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNA-seq data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession codes GSE103846 and GSE107965. DNA-seq data that support the findings of this study have been deposited in GEO database under accession code GSE111309. Previously published sequencing data that were re-analyzed here are available under accession codes GSE62772, GSE73211, GSE80264, GSE81791 and from the GTEx Portal version V6p under the link https://www.gtexportal.org/home/datasets. Registered users can access the files using the dbGaP accession number phs000424.v6.p1. CRISPR score tables and the significance values are provided in Supplementary Tables 1-3. Source data for Fig. 3, 4, 5, 6 and Supplementary Fig. 4, 6 have been provided as Supplementary Table 4. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences

Study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical method was used to determine the sample size. Sample size was chosen based on standards in the field. Samples size and number of independent experiments are stated in figure legends or in Methods section. Statistical analysis was only performed for a minimum of three biologically independent samples.
Data exclusions	No data were excluded from the analysis.
Replication	All attempts at replication were successful.
Randomization	No animals and/or human participants were used and no randomization was performed for the experimental groups.
Blinding	Investigators were not blinded to group allocation during data collection and/or analysis.

Materials & experimental systems

Policy information about availability of materials

n/a	Involved in the study
☐	☒ Unique materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Research animals
☒	☐ Human research participants

Unique materials

Obtaining unique materials	All unique materials used are readily available from the authors or from standard commercial sources. Please see the Methods section.

Antibodies

Antibodies used	PE-conjugated mouse anti-Human TRA-1-60 antibody, Supplier: BD Biosciences, Catalog No. 560884, Clone: TRA-1-60, Lot: 5114784 and 6183856, Dilution for flow cytometry analysis: 1:40 Rabbit anti-phospho-4E-BP1 (Thr37/46) monoclonal antibody, Supplier: Cell Signaling Technology, Catalog No. 2855, Clone: 236B4, Lot: 23, Dilution for western blotting: 1:1000 Rabbit anti-phospho-Akt (Ser473) antibody, Supplier: Cell Signaling Technology, Catalog No. 4060, Lot: 16, Dilution for western blotting: 1:1000

	Rat anti-GAPDH antibody, Supplier: Cell Signaling Technology , Catalog No. 2118S, Lot: 0008, Dilution for western blotting: 1:30,000
Validation	The use and validation of antibodies were mainly based on the statement of the manufacturers and were also validated by our laboratory by the use of negative and/or positive controls such as undifferentiated vs. differentiated cells or serum-activated vs. serum-deprived cells.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Haploid hESCs -h-pES10 cell line was previously isolated by us (Please see Sagi et al., Nature, 532, 107–111 (2016)). Rex1-EGFP hESCs were previously generated by us via stable transfection of H9 cell line (J. Itskovitz-Eldor). 293T cell line was originally from R. Weinberg, BJ human fibroblasts were from Clontech. KBM7 cell line was from Horizon Discovery. CSES9 and CSES15 cell lines were previously generated by us.
Authentication	Commercially available cell lines used in this study were not authenticated by ourselves. Haploid human embryonic stem cells were validated by their morphology, gene expression patterns, karyotype analyses and teratoma formation.
Mycoplasma contamination	All cell lines tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in the study.

Method-specific reporting

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ Magnetic resonance imaging

Flow Cytometry

Plots

Confirm that:

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

☒ 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).

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

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

Methodology

Sample preparation	For the enrichment of haploid cells, cells were washed with phosphate buffered saline (PBS), trypsinized using TrypLE Select (Thermo Fisher Scientific) and stained with 10 µg ml-1 Hoechst 33342 (Sigma Aldrich) in hESC growth medium at 37 0C for 30 min. Cells were then centrifuged and resuspended in PBS that contains 10% KSR and 10 µM ROCK inhibitor Y-27632, filtered through a 70-µm cell strainer (Corning) and sorted by a 405 nm laser in BD FACSAria III (BD Biosciences). For TRA-1-60 immunocytochemistry, hESCs were trypsinized with TrypLE Select. Cells were collected in cold PBS containing 10% KSR, centrifuged at 300 g for 5 minutes and resuspended in 200 µl PBS containing 10% KSR. PE-conjugated mouse anti-human TRA-1-60 antibody (1:40, BD Biosciences) was then incubated with the cells for 30 minutes at 4 0C. Cells were washed with PBS containing 10% KSR twice, centrifuged at 300 g at 4 0C and resuspended in PBS with 10% KSR. Immunolabeled cells were filtered through a 70-µm cell strainer and analyzed in BD FACSAria III for the proportion of TRA-1-60-positive cells. For apoptosis assay, hESCs were trypsinized gently with TrypLE Select. Apoptotic cells were labeled with Annexin V and Propidium Iodide (PI) using the MEBCYTO Apoptosis Kit according to manufacturer's instructions (MBL). Labeled cells were filtered through a 70-µm cell strainer and analyzed in BD FACSAria III for the proportion of Annexin V-positive cells. Finally for the PI staining, hESCs were trypsinized with TrypLE Select and fixed with cold methanol (4 0C). Fixed cells were treated with 200 µg ml-1 RNase A (Sigma) for 30 minutes and stained with 50 µg ml-1 PI for 5 minutes. Stained cells were filtered through a 70-µm cell strainer and analyzed in BD FACSAria III for their cell cycle profile.
Instrument	BD Biosciences FACSAria III
Software	FlowJo software Version 7.6.5 has been used for analysis.
Cell population abundance	When the haploid cells were sorted, the post-sort fraction for haploid cells was nearly 100%.
Gating strategy	Cells were first gated for single cells using forward and side scatter gates. Subsequently, boundaries between positive and negative cell populations were made based on the staining of untreated control samples. Supplementary Figures 6e, f and g exemplify the gating strategies used for Annexin-V, propidium iodide and TRA-1-60 stainings.

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