

# Derivation and differentiation of haploid human embryonic stem cells

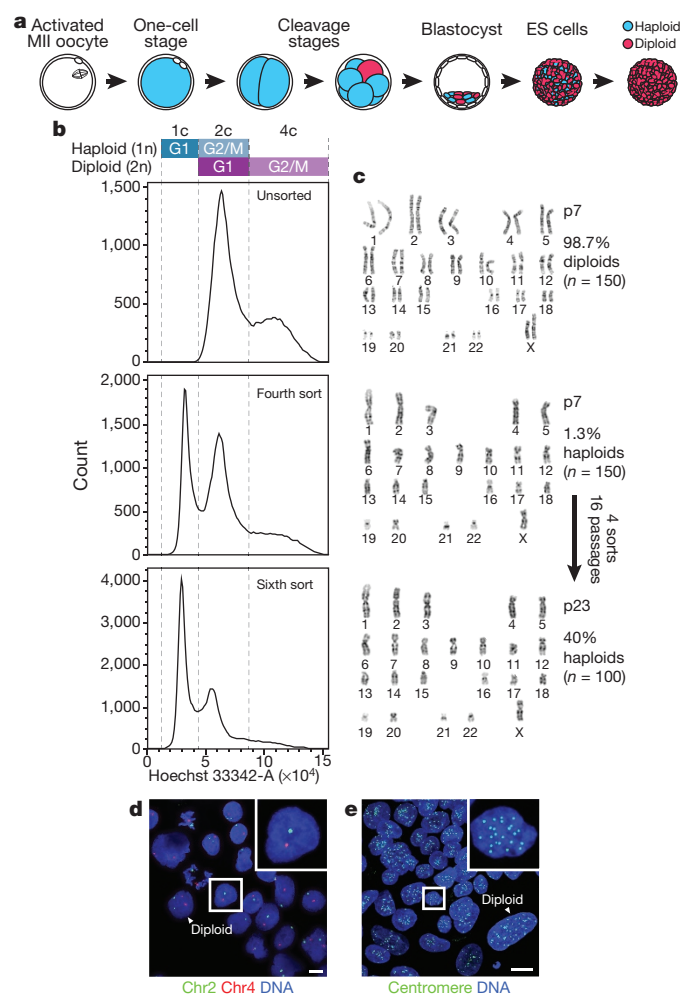
Ido Sagi<sup>1</sup>, Gloryn Chia<sup>2</sup>, Tamar Golan-Lev<sup>1</sup>, Mordecai Peretz<sup>1</sup>, Uri Weissbein<sup>1</sup>, Lina Sui<sup>2</sup>, Mark V. Sauer<sup>3</sup>, Ofra Yanuka<sup>1</sup>, Dieter Egli<sup>2,4</sup> & Nissim Benvenisty<sup>1</sup>

Diploidy is a fundamental genetic feature in mammals, in which haploid cells normally arise only as post-meiotic germ cells that serve to ensure a diploid genome upon fertilization. Gamete manipulation has yielded haploid embryonic stem (ES) cells from several mammalian species<sup>1–6</sup>, but haploid human ES cells have yet to be reported. Here we generated and analysed a collection of human parthenogenetic ES cell lines originating from haploid oocytes, leading to the successful isolation and maintenance of human ES cell lines with a normal haploid karyotype. Haploid human ES cells exhibited typical pluripotent stem cell characteristics, such as self-renewal capacity and a pluripotency-specific molecular signature. Moreover, we demonstrated the utility of these cells as a platform for loss-of-function genetic screening. Although haploid human ES cells resembled their diploid counterparts, they also displayed distinct properties including differential regulation of X chromosome inactivation and of genes involved in oxidative phosphorylation, alongside reduction in absolute gene expression levels and cell size. Surprisingly, we found that a haploid human genome is compatible not only with the undifferentiated pluripotent state, but also with differentiated somatic fates representing all three embryonic germ layers both *in vitro* and *in vivo*, despite a persistent dosage imbalance between the autosomes and X chromosome. We expect that haploid human ES cells will provide novel means for studying human functional genomics and development.

Haploid genetics is useful for delineating genome function. Although haploid genetics has been mostly restricted to unicellular organisms, recent reports of haploid ES cells have extended it into animal species, including mammals<sup>1–7</sup>. Deriving haploid human ES cells by similar approaches has probably been hindered by the limited availability of human oocytes<sup>8</sup>. Artificial activation of unfertilized metaphase II (MII) human oocytes results in efficient development to the blastocyst stage and subsequent derivation of parthenogenetic ES (pES) cells<sup>9–11</sup>. In mouse parthenogenetic embryos haploidy usually persists at the blastocyst stage<sup>12,13</sup>, but diploid cells progressively dominate over increasing cell cycles due to spontaneous and irreversible diploidization<sup>13–15</sup> (Fig. 1a). By estimation, even if diploidization occurs in 1 out of 10 cell cycles, 1% of ES cells may remain haploid at early passages (Extended Data Fig. 1a).

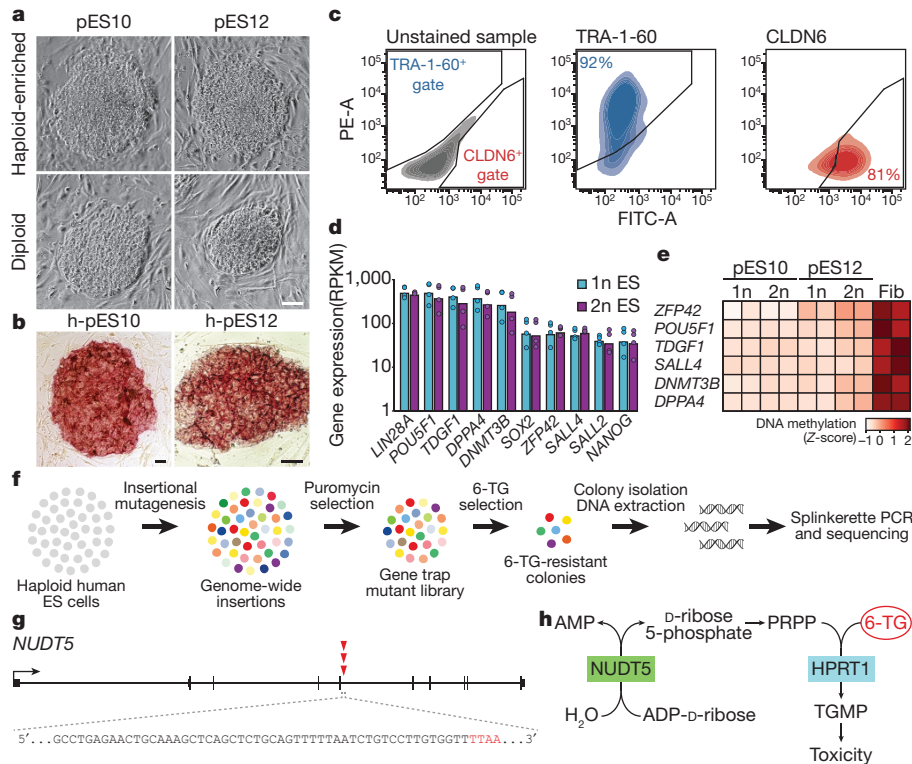
To explore the feasibility of deriving haploid human ES cells, we generated and analysed a collection of 14 early-passage human pES cell lines for the persistence of haploid cells. We initially used chromosome counting by metaphase spreading as a method for unambiguous and quantitative discovery of rare haploid nuclei. Among ten individual pES cell lines, a low proportion of haploid metaphases was found in one cell line, pES10 (1.3%, Extended Data Table 1). In four additional lines, we also used viable fluorescence-activated cell sorting (FACS) with Hoechst 33342 staining, aiming to isolate cells with a DNA content

corresponding to less than two chromosomal copies (2c), leading to the successful enrichment of haploid cells from a second cell line, pES12 (Extended Data Table 2).



**Figure 1 | Derivation of haploid human ES cells.** **a**, Schematic of putative haploidy in pES cells. **b**, DNA content profiles of haploid pES10, established by repeated enrichment of 1c cells. Top to bottom, unsorted diploid cells, partially purified (fourth sort) and mostly purified (sixth sort) haploid cells. **c**, pES10 karyotypes before and after 1c-cell enrichment. **p**, passage. **d**, **e**, DNA FISH (**d**) and centromere staining (**e**) in haploid-enriched pES10 cells. Magnifications show haploid nuclei with a single hybridization signal (**d**) and 23 centromere foci (**e**), respectively. Scale bars, 10  $\mu$ m.

<sup>1</sup>The Azrieli Center for Stem Cells and Genetic Research, Department of Genetics, Silberman Institute of Life Sciences, The Hebrew University, Jerusalem 91904, Israel. <sup>2</sup>Department of Pediatrics, Columbia University, New York, New York 10032, USA. <sup>3</sup>Center for Women's Reproductive Care, College of Physicians and Surgeons, Columbia University, New York, New York 10019, USA. <sup>4</sup>The New York Stem Cell Foundation Research Institute, New York, New York 10032, USA.



**Figure 2 | Haploid human ES cells display pluripotent stem cell characteristics and enable loss-of-function screening.** **a, b,** Colony morphology and alkaline phosphatase staining. Scale bars, 50  $\mu$ m. **c,** Flow cytometry analysis of gated h-pES10 1c cells by co-staining DNA and cell surface markers TRA-1-60 and CLDN6. **d,** Expression levels of pluripotency genes in G1-sorted haploid and diploid ES cells ( $n = 4$  each,

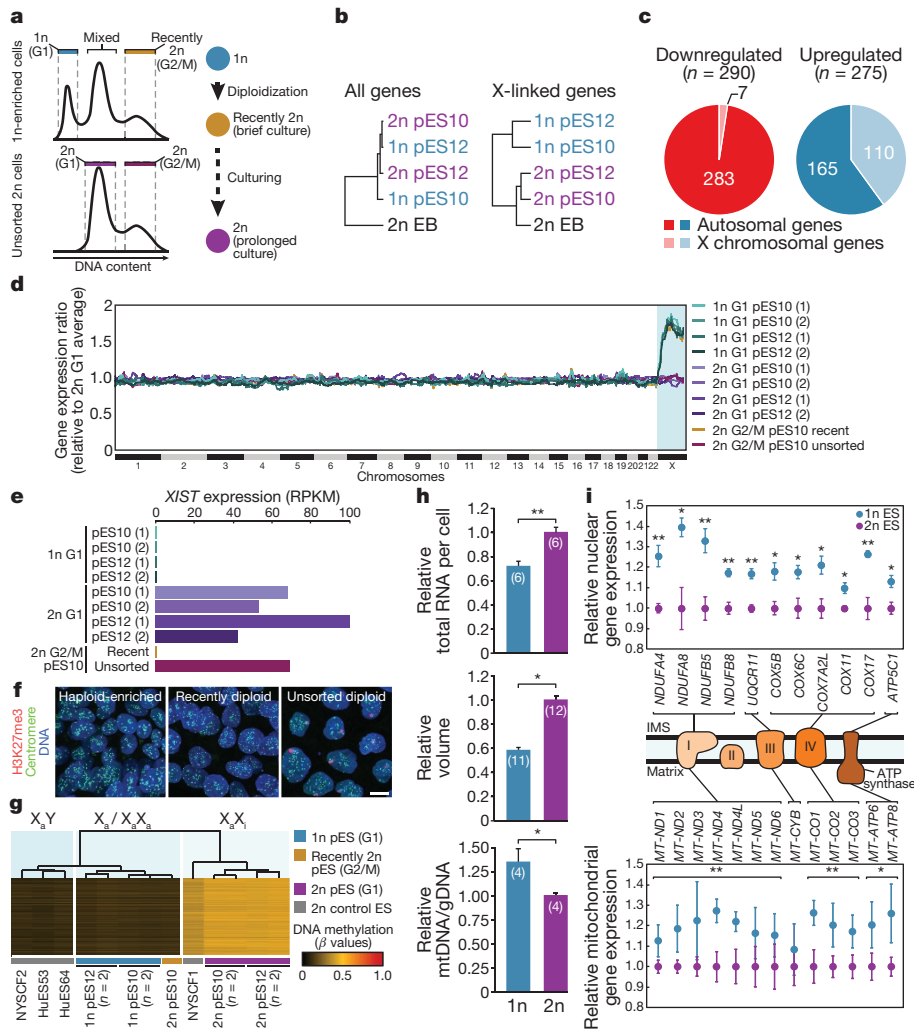
two biological replicates per line, logarithmic scale). Mean indicated by bars. **e,** DNA methylation levels at pluripotency genes in biological duplicates of G1-sorted ES cells and control fibroblasts (Fib). **f,** Schematic of gene trapping and screening for 6-TG-resistance genes. **g,** *NUDT5* insertions (red arrows) detected in three 6-TG-resistant colonies (indicated by TTA). **h,** Metabolic pathway leading to 6-TG toxicity.

Two individual haploid-enriched ES cell lines were established from pES10 and pES12 (hereafter referred to as h-pES10 and h-pES12) within five to six rounds of 1c-cell enrichment and expansion (Fig. 1b and Extended Data Fig. 1b). These cell lines were cultured in standard conditions for over 30 passages while including cells with a normal haploid karyotype (Fig. 1c and Extended Data Fig. 1c). However, since diploidization occurred at a rate of 3–9% cells per day (Extended Data Fig. 1a; see Methods), sorting at every three to four passages was required to maintain haploid cells. Visualization of ploidy was further enabled by DNA fluorescence *in situ* hybridization (FISH) (Fig. 1d and Extended Data Fig. 1d) and quantification of centromere protein foci (Fig. 1e and Extended Data Fig. 1e; see Supplementary Notes and Extended Data Fig. 2). Besides having an intact karyotype, haploid ES cells did not harbour significant copy number variations (CNVs) relative to their unsorted diploid counterparts (Extended Data Fig. 1f), nor common duplications that would result in pseudo-diploidy, indicating that genome integrity was preserved throughout haploid-cell isolation and maintenance.

Both h-pES10 and h-pES12 exhibited classical human pluripotent stem cell features, including typical colony morphology and alkaline phosphatase activity (Fig. 2a, b). Single haploid ES cells expressed hallmark pluripotency markers as confirmed by centromere foci quantification in essentially pure (>95%) haploid cultures (Extended Data Fig. 3). Selective flow cytometry enabled the validation of the expression of two human ES-cell-specific cell surface markers (TRA-1-60 and CLDN6 (ref. 16)) in single haploid cells (Fig. 2c). Moreover, sorted haploid and diploid ES cells showed highly similar transcriptional and epigenetic signatures of pluripotency genes (Fig. 2d, e). In accordance with a parthenogenetic origin, these cells also featured distinct molecular profiles of maternal imprinting (Extended Data Fig. 4a–d).

Haploid cells are valuable for genetic screening because phenotypically selectable mutants can be identified upon disruption of single-copy alleles. To demonstrate the applicability of this principle in haploid human ES cells, we generated a genome-wide mutant library using a gene trap transposon system (Fig. 2f and Extended Data Fig. 4e; see Methods), and screened for resistance to the purine analogue 6-thioguanine (6-TG). Out of six isolated and analysed 6-TG-resistant colonies, three harboured an identical gene trap insertion at the *NUDT5* autosomal gene (Fig. 2g). The disruption of this gene was recently confirmed to confer 6-TG resistance in human cells<sup>17</sup> (Fig. 2h). Detection of a loss-of-function phenotype due to an autosomal mutation thereby validates that genetic screening is feasible in haploid human ES cells.

The ability of human ES cells to exist as both haploids and diploids led us to investigate whether these two ploidy states differ in certain aspects of gene regulation and cell biology. To analyse haploid and diploid ES cells in the same cell cycle phase, we used FACS to isolate G1-phase haploid cells (1c) and compared them with isogenic G1-phase diploid cells (2c) from unsorted diploid cultures (Fig. 3a and Extended Data Fig. 5a; see Methods). We first aimed to uncover putative ploidy-associated differences by comparing the transcriptomes of haploid and diploid ES cells using RNA sequencing (RNA-seq), considering that observed changes in expression levels would be relative to the total gene expression of each ploidy state, rather than representing absolute differences. On the genome-scale, undifferentiated haploid and diploid ES cells clustered closely and separately from differentiated embryoid bodies (EBs) (Fig. 3b). Nonetheless, we identified 275 relatively upregulated and 290 relatively downregulated genes in haploids compared with diploids (greater than twofold change, false discovery rate (FDR) < 0.05; Extended Data Fig. 5b).



**Figure 3 | Molecular and cellular comparisons of haploid and diploid ES cells.** **a**, Experimental scheme. **b**, RNA-seq-based clustering analyses of isogenic G1-sorted haploid and diploid cells (two biological replicates per line) and diploid pES12-derived EBs. **c**, Pie chart of relatively downregulated and upregulated genes in haploid versus diploid ES cells. **d–g**, Differential X chromosome inactivation in haploid and diploid ES cells. **d**, Genome-wide expression moving median plot. **e**, *XIST* expression levels. **f**, H3K27me3 staining. Scale bar, 10  $\mu$ m. **g**, X chromosome DNA

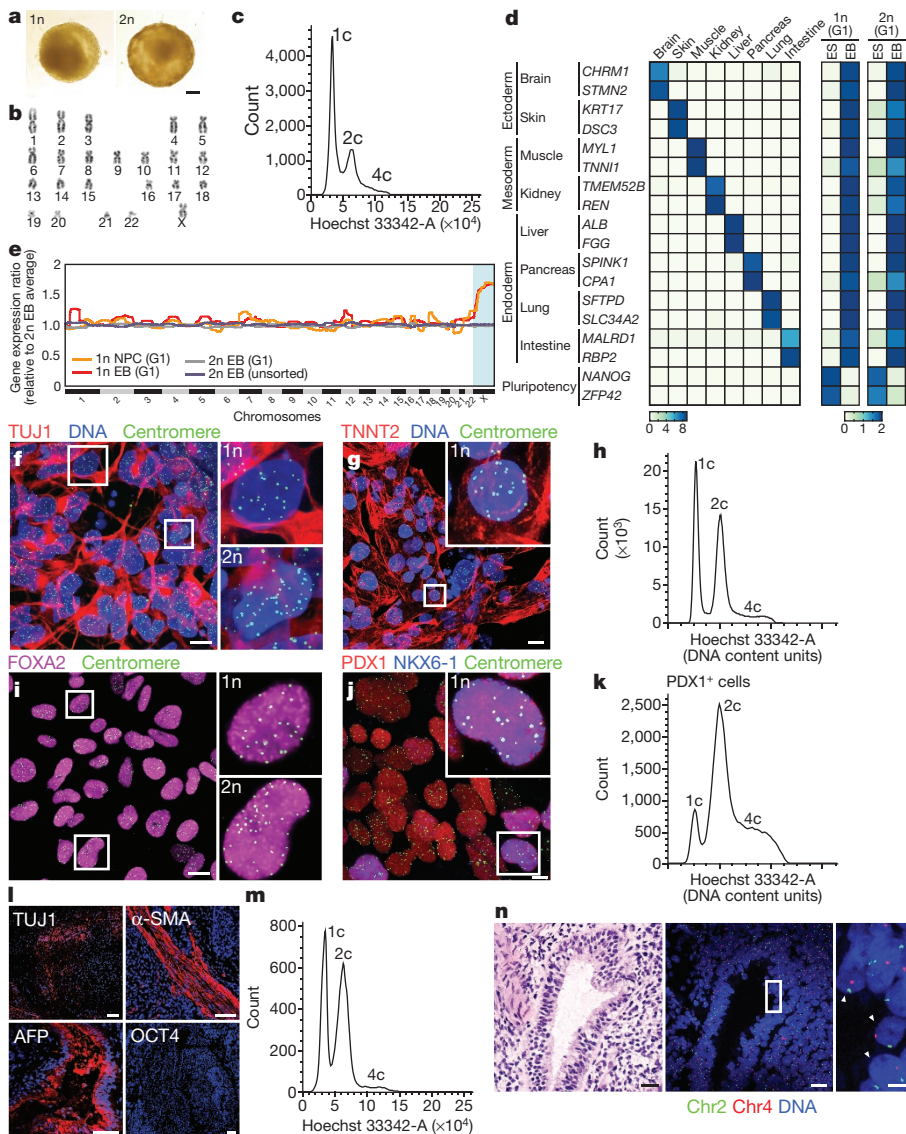
Notably, X chromosomal genes were significantly overrepresented among the relatively upregulated gene set (40%,  $P < 0.001$ ,  $\chi^2$  goodness of fit test) (Fig. 3c), and the expression levels of X chromosomal genes alone clearly distinguished between haploid and diploid ES cells (Fig. 3b). These data correlate with an expected differential status of X chromosome inactivation in haploid and diploid human ES cells: whereas the single-copy X chromosome in haploids is transcriptionally active ( $X_a$ ), one of the two X chromosomes in diploids is often inactivated ( $X_aX_i$ )<sup>18</sup>. Indeed, haploid human ES cells exhibited a relative increase in X chromosomal gene expression and lacked expression of the *XIST* transcript which drives X chromosome inactivation (Fig. 3d, e and Extended Data Fig. 5b–d), as in diploid  $X_aX_a$  human ES cells<sup>19</sup>. X chromosome inactivation is regulated by repressive histone modifications and DNA methylation. H3K27me3 foci were consistently observed in unsorted diploid ES cells, but not in the haploid-enriched counterparts (Fig. 3f). Moreover, methylome analysis showed that the X chromosome DNA methylation signature of haploid ES cells resembles that of diploid male ES cells ( $X_aY$ ), whose single-copy X chromosome is largely hypomethylated, as opposed to the composite pattern of a hypomethylated  $X_a$  and a hypermethylated  $X_i$  in diploid female cells

(Fig. 3g). Interestingly, recently diploidized ES cells (see Methods) remained  $X_aX_a$  soon after diploidization by all assays mentioned earlier (Fig. 3a, d–g). Normalization to total gene expression<sup>20</sup> resulted in seemingly similar expression levels of autosomal genes, but higher levels of X-linked genes in haploid cells (Fig. 3d and Extended Data Fig. 5c). However, assuming that the absolute expression of X-linked genes in haploid  $X_a$  and diploid  $X_aX_i$  cells are equivalent, these data suggest a genome-wide autosomal gene expression level reduction in haploids (Extended Data Fig. 5e, f). In support of this, total RNA amounts isolated from haploid ES cells were significantly lower than those from equal numbers of diploid cells (Fig. 3h). A decrease in total gene expression implied that physical dimensions may also be altered. Indeed, the average haploid:diploid diameter ratio of G1-sorted ES cells was  $\sim 0.8$ , corresponding to  $\sim 0.7$  surface area ratio and  $\sim 0.6$  volume ratio (Fig. 3h and Extended Data Fig. 5g). We subsequently focused on consistent differential regulation within autosomes (see Methods). By transcriptome and methylome analyses, genes relatively downregulated in haploid ES cells were significantly enriched for genes encoding signal-peptide-bearing

methylation levels. **h**, Relative total RNA, cell volume and ratio of mitochondrial to genomic DNA (mtDNA/gDNA) between G1-sorted haploid and diploid ES cells. Number of replicates indicated in parenthesis. Error bars represent s.d. **i**, Mean expression levels  $\pm$  s.e.m. of nuclear and mitochondrial oxidative phosphorylation genes in haploid and diploid ES cells (replicates as in Fig. 2d), and their schematic organization in this pathway. \* $P < 0.05$ ; \*\* $P < 0.01$  (two-tailed unpaired Student's *t* test). Source data for **h** and **i** are available online.

(Fig. 3g). Interestingly, recently diploidized ES cells (see Methods) remained  $X_aX_a$  soon after diploidization by all assays mentioned earlier (Fig. 3a, d–g). Normalization to total gene expression<sup>20</sup> resulted in seemingly similar expression levels of autosomal genes, but higher levels of X-linked genes in haploid cells (Fig. 3d and Extended Data Fig. 5c). However, assuming that the absolute expression of X-linked genes in haploid  $X_a$  and diploid  $X_aX_i$  cells are equivalent, these data suggest a genome-wide autosomal gene expression level reduction in haploids (Extended Data Fig. 5e, f). In support of this, total RNA amounts isolated from haploid ES cells were significantly lower than those from equal numbers of diploid cells (Fig. 3h). A decrease in total gene expression implied that physical dimensions may also be altered. Indeed, the average haploid:diploid diameter ratio of G1-sorted ES cells was  $\sim 0.8$ , corresponding to  $\sim 0.7$  surface area ratio and  $\sim 0.6$  volume ratio (Fig. 3h and Extended Data Fig. 5g).

We subsequently focused on consistent differential regulation within autosomes (see Methods). By transcriptome and methylome analyses, genes relatively downregulated in haploid ES cells were significantly enriched for genes encoding signal-peptide-bearing



**Figure 4 | Differentiation of haploid human cells.** **a**, EBs from haploid-enriched and diploid pES12 cells. Scale bar, 100  $\mu$ m. **b**, Haploid EB cell karyotype. **c**, DNA content profile of h-pES10 EB cells. **d**, Expression of tissue- and pluripotency-specific genes in G1-sorted haploid and diploid ES and EB pES10 cells. **e**, Differential X chromosome inactivation in haploid and diploid EBs and NPCs by genome-wide expression moving median plot. **f, g, i, j**, Centromere and differentiation marker co-staining in h-pES12-derived neurons (**f**), cardiomyocytes (**g**), definitive endoderm cells (**i**) and pancreatic cells (**j**). Scale bars, 10  $\mu$ m. **h, k**, DNA content profiles of h-pES12 cells differentiated into cardiomyocytes (**h**) and PDX1-positive pancreatic cells (**k**). **l**, TUJ1 (ectoderm),  $\alpha$ -SMA (mesoderm), AFP (endoderm) and OCT4 (pluripotency) staining in an h-pES12-derived teratoma. Scale bars, 50  $\mu$ m. **m**, DNA content profile of an h-pES10-derived teratoma. **n**, Serial h-pES12-derived teratoma sections analysed histologically (left; scale bar, 20  $\mu$ m) and by DNA FISH (middle; scale bar, 20  $\mu$ m). Haploid nuclei are shown in magnification (right; scale bar, 5  $\mu$ m).

proteins (Extended Data Fig. 5h). Remarkably, we also detected subtle yet significant relative upregulation of 11 genes involved in oxidative phosphorylation in haploid cells, including representatives encoding subunits of four out of the five complexes comprising this pathway (Fig. 3i and Extended Data Fig. 5i). Furthermore, all 13 mitochondrially encoded oxidative phosphorylation genes were consistently upregulated as well (Fig. 3i), indicating coordinated regulation between nuclear and mitochondrial genes. This coincided with a 32% increase in the mitochondrial DNA (mtDNA) to nuclear DNA ratio between haploids and diploids (Fig. 3h), suggesting that mitochondrial abundance relative to nuclear DNA content is higher in haploid cells.

We next sought to assess the differentiation potential of haploid human ES cells. The 21-day-old EBs generated by spontaneous differentiation of haploid-enriched and diploid ES cells could not be distinguished by their appearance (Fig. 4a), and the morphology of dissociated haploid-cell-derived EB cells was consistent with differentiation (Extended Data Fig. 6a). Notably, metaphase spreading revealed a haploid karyotype (Fig. 4b; 4/4 metaphases), and a largely haploid ( $\sim$ 70%) DNA profile was confirmed by flow cytometry in both h-pES10-derived and h-pES12-derived EB cells (Fig. 4c and Extended Data Fig. 6b). We then compared the gene expression profiles of G1-sorted haploid ES cells and EB cells, focusing on 18 lineage-specific genes across nine cell types (Fig. 4d). Whereas expression levels were

negligible in undifferentiated ES cells, all tissue-specific genes were expressed in haploid and diploid EB cells (Fig. 4d and Extended Data Fig. 6c). Haploid and diploid EB cells showed insignificant expression of pluripotency-specific genes, consistent with efficient acquisition of differentiated cell fates.

To extend this analysis to more specific cell types, we subjected haploid ES cells to directed differentiation assays. Haploid ES cells differentiated towards a neural fate for ten days remained haploid, while giving rise to NCAM1-positive neural progenitor cells (NPCs, with  $\sim$ 90% efficiency) (Extended Data Fig. 7a, b). Sorted haploid NPCs expressed multiple neural-lineage-specific genes but not pluripotency-specific genes (Extended Data Fig. 7c, d). X chromosome inactivation in differentiated diploid female cells results in dosage compensation of 1:2 between the X chromosome and autosomes. As haploid cells do not inactivate their single-copy X chromosome, an X:autosomes dosage imbalance of 1:1 should persist into the differentiated state. Indeed, both haploid NPCs and EB cells showed an  $X_a$  signature contrary to the  $X_aX_i$  signature of diploid EB cells as indicated by whole-genome expression analysis and *XIST* levels (Fig. 4e and Extended Data Fig. 7e).

Neuronal differentiation was not restricted to the progenitor stage as the cells also differentiated with high efficiency ( $>$ 90%) into mature TUJ1 (also known as  $\beta$ -tubulin III)-positive neurons by 20 days, with persistence of haploid cells as shown by co-staining

with centromeres (Fig. 4f; 47% haploids,  $n = 104$ ) and FISH analysis (Extended Data Fig. 7f, g; 46% haploids,  $n = 200$ ). Similarly, haploid cells differentiated into TNNT2-expressing cardiomyocytes (Fig. 4g; 32% haploids,  $n = 97$ ) during an eleven-day protocol resulting in spontaneously beating clusters (Supplementary Video 1), and 39% ( $n = 31$ ) of haploid cells sorted from the whole culture (25% 1c-cells) were confirmed as TNNT2-positive (Fig. 4h and Extended Data Fig. 7h). Next, we differentiated haploid-enriched cultures (~70% haploids) to the pancreatic lineage. Analysing two stages of differentiation by centromere foci analysis, we observed robust differentiation (>90%) of both haploids and diploids into FOXA2-positive definitive endoderm cells (Fig. 4i; 56% haploids,  $n = 112$ ), and differentiation into PDX1-positive pancreatic cells (Fig. 4j; 13% haploid,  $n = 103$ ), some of which were also positive for NKX6-1. In addition, the persistence of haploid PDX1-positive cells was confirmed by flow cytometry (Fig. 4k; 10% PDX1-positive 1c cells; and Extended Data Fig. 7i, j).

Finally, both haploid-enriched human ES cell lines gave rise to teratomas comprising cell types of ectodermal, mesodermal and endodermal origins (Fig. 4l and Extended Data Fig. 8a, b), meeting the most stringent criterion for human pluripotency *in vivo*. Importantly, no residual undifferentiated OCT4-positive cells were detected (Fig. 4l and Extended Data Fig. 8b). Upon dissection, DNA content analysis revealed that a considerable population of h-pES10-derived teratoma cells remained haploid (Fig. 4m). Combined analysis of serial sections from an independent, h-pES12-derived teratoma, by histology and FISH confirmed the existence of *in vivo* differentiated haploid human cells capable of contributing to an organized tissue structure while responding to developmental signals (Fig. 4n). Haploid cells were identified in all analysed teratomas ( $n = 4$ ), albeit with variable proportions, which may be influenced by the initial amount of haploid cells and/or the duration of teratoma formation.

Haploid mammalian cells have proven invaluable for loss-of-function screens<sup>7</sup>. Using a genome-wide library of gene-trapped haploid human ES cells, we demonstrated their potential for biomedically relevant functional genomics by forward genetic screening (Supplementary Discussion). Whereas previous studies on non-human haploid ES cells mostly emphasized the similarity between haploids and diploids, here we also pointed to several transcriptional, epigenetic and physical properties that set them apart (Supplementary Discussion). Interestingly, we did not observe global transcriptional compensation in haploid cells, indicating that it is not required for cellular viability as long as an autosomal balance is preserved. In contrast, autosomal imbalance appears intolerable based on the strict absence of human autosomal monosomies in *in vitro* fertilization ES cells<sup>21</sup>. Remarkably, we found that a haploid human karyotype is not a barrier for ES cell differentiation. As observed in the mouse<sup>2</sup>, haploid human ES cells gave rise to NPCs while remaining haploid. However, while mouse studies showed that haploid cells are lost upon further differentiation<sup>2,14</sup>, we observed specification of human haploid cells into somatic cell fates of all three embryonic germ layers, despite persistent dosage imbalance between the X chromosome and autosomes (Supplementary Discussion).

Throughout evolution, mammalian genomes have been solidified by diploidy-dependent adaptations such as parental imprinting, which restrict the development of haploid uniparental embryos. Nonetheless, haploid cells are capable of directing development in certain animal species<sup>22</sup>. The surprising differentiation potential of haploid human genomes suggests that diploidy-dependent adaptations, rather than haploidy, pose the predominant barriers for uniparental development in humans. The discovery of haploid human ES cells should thus provide novel means to delineate basic aspects of human genetics and development.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 July 2015; accepted 8 February 2016.

Published online 16 March 2016.

1. Leeb, M. & Wutz, A. Derivation of haploid embryonic stem cells from mouse embryos. *Nature* **479**, 131–134 (2011).
2. Elling, U. *et al.* Forward and reverse genetics through derivation of haploid mouse embryonic stem cells. *Cell Stem Cell* **9**, 563–574 (2011).
3. Yang, H. *et al.* Generation of genetically modified mice by oocyte injection of androgenetic haploid embryonic stem cells. *Cell* **149**, 605–617 (2012).
4. Li, W. *et al.* Androgenetic haploid embryonic stem cells produce live transgenic mice. *Nature* **490**, 407–411 (2012).
5. Li, W. *et al.* Genetic modification and screening in rat using haploid embryonic stem cells. *Cell Stem Cell* **14**, 404–414 (2014).
6. Yang, H. *et al.* Generation of haploid embryonic stem cells from *Macaca fascicularis* monkey parthenotes. *Cell Res.* **23**, 1187–1200 (2013).
7. Wutz, A. Haploid mouse embryonic stem cells: rapid genetic screening and germline transmission. *Annu. Rev. Cell Dev. Biol.* **30**, 705–722 (2014).
8. Egli, D. *et al.* Impracticality of egg donor recruitment in the absence of compensation. *Cell Stem Cell* **9**, 293–294 (2011).
9. Revazova, E. S. *et al.* Patient-specific stem cell lines derived from human parthenogenetic blastocysts. *Cloning Stem Cells* **9**, 432–449 (2007).
10. Kim, K. *et al.* Recombination signatures distinguish embryonic stem cells derived by parthenogenesis and somatic cell nuclear transfer. *Cell Stem Cell* **1**, 346–352 (2007).
11. Paull, D. *et al.* Nuclear genome transfer in human oocytes eliminates mitochondrial DNA variants. *Nature* **493**, 632–637 (2013).
12. Tarkowski, A. K., Witkowska, A. & Nowicka, J. Experimental parthenogenesis in the mouse. *Nature* **226**, 162–165 (1970).
13. Kaufman, M. H., Robertson, E. J., Handyside, A. H. & Evans, M. J. Establishment of pluripotential cell lines from haploid mouse embryos. *J. Embryol. Exp. Morphol.* **73**, 249–261 (1983).
14. Leeb, M. *et al.* Germline potential of parthenogenetic haploid mouse embryonic stem cells. *Development* **139**, 3301–3305 (2012).
15. Takahashi, S. *et al.* Induction of the G2/M transition stabilizes haploid embryonic stem cells. *Development* **141**, 3842–3847 (2014).
16. Ben-David, U., Nudel, N. & Benvenisty, N. Immunologic and chemical targeting of the tight-junction protein Claudin-6 eliminates tumorigenic human pluripotent stem cells. *Nat. Commun.* **4**, 1992 (2013).
17. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnol.* **34**, 184–191 (2016).
18. Silva, S. S., Rowntree, R. K., Mekhoubad, S. & Lee, J. T. X-chromosome inactivation and epigenetic fluidity in human embryonic stem cells. *Proc. Natl Acad. Sci. USA* **105**, 4820–4825 (2008).
19. Bruck, T., Yanuka, O. & Benvenisty, N. Human pluripotent stem cells with distinct X inactivation status show molecular and cellular differences controlled by the X-linked ELK-1 gene. *Cell Rep.* **4**, 262–270 (2013).
20. Lovén, J. *et al.* Revisiting global gene expression analysis. *Cell* **151**, 476–482 (2012).
21. Biancotti, J. C. *et al.* The *in vitro* survival of human monosomies and trisomies as embryonic stem cells. *Stem Cell Res.* **9**, 218–224 (2012).
22. Otto, S. P. & Jarne, P. Evolution. Haploids—hapless or happening? *Science* **292**, 2441–2443 (2001).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank all members of the Benvenisty and Egli laboratories for input and support. We thank Y. Avior and W. Breuer for their assistance with experimental procedures. I.S. is supported by the Adams Fellowships Program for Doctoral Students, G.C. is supported by the A\*STAR International Fellowship, U.W. is a Clore Fellow, D.E. is a NYSCF-Robertson Investigator, and N.B. is the Herbert Cohn Chair in Cancer Research. This work was partially supported by The Rosetrees Trust and by The Azrieli Foundation (to N.B.), by the Russell Berrie Foundation Program in Cellular Therapies of Diabetes, by the New York State Stem Cell Science (NYSTEM) IIRP Award number C026184, and by the New York Stem Cell Foundation (to D.E.).

**Author Contributions** I.S., D.E. and N.B. designed the study and wrote the manuscript with input from all authors. I.S. isolated and characterized haploid human ES cell lines, performed differentiation experiments and analysed the data. G.C. developed and performed the centromere quantification analysis and carried out neuronal differentiation. T.G.-L. assisted in tissue culture and performed karyotype analyses and tissue sectioning. I.S., M.P., U.W. and O.Y. were involved in the genetic screening. M.P. and U.W. assisted with teratoma assays. L.S. assisted with pancreatic differentiation. M.V.S. was involved in all aspects of oocyte donation and research. D.E. derived human pES cell lines from haploid oocytes. D.E. and N.B. supervised the study.

**Author Information** All high-throughput data have been deposited at the Gene Expression Omnibus (GEO) under accession number GSE71458. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.E. ([de2220@cumc.columbia.edu](mailto:de2220@cumc.columbia.edu)) or N.B. ([nissimb@cc.huji.ac.il](mailto:nissimb@cc.huji.ac.il)).

## METHODS

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

**Human oocyte manipulation and parthenogenetic ES cell line derivation.** Human oocyte donation and pES and swaPS cell line derivation procedures were described previously<sup>11,23</sup>. Oocyte donors gave informed consent. Experiments were approved by the embryonic stem cell research oversight committee and the institutional review board at Columbia University Medical Center. Briefly, mature MII oocytes were activated using a calcium ionophore and/or an electrical pulse, followed by 4 h of culture with puromycin. Polar body extrusion and the presence of a single pronucleus indicating haploidy were confirmed, and oocytes were allowed to develop to the blastocyst stage. swaPS cells were derived following activation of an oocyte whose nuclear genome had been swapped with that of another oocyte<sup>11</sup>. ES cell lines were derived by laser ablation of the trophectoderm<sup>24</sup> and addition of ROCK inhibitor Y-27632 at 10  $\mu$ M to the derivation medium<sup>23</sup>. Then 2–3 days after plating, remaining trophectoderm cells were laser ablated, and inner cell mass cells were allowed to grow for 10–14 days until manual picking of the outgrowth was feasible.

**Cell culture.** Unless otherwise stated, human ES cells were cultured on a feeder layer of growth-arrested mouse embryonic fibroblasts (MEFs) in standard human ES cell medium composed of Knockout Dulbecco's Modified Eagle's Medium supplemented with 15% Knockout Serum Replacement (KSR, Thermo Fisher Scientific), 2 mM L-glutamine, 0.1 mM nonessential amino acids, 50 units ml<sup>-1</sup> penicillin, 50  $\mu$ g ml<sup>-1</sup> streptomycin, 0.1 mM  $\beta$ -mercaptoethanol and 8 ng ml<sup>-1</sup> basic fibroblast growth factor (bFGF). Cells were free of mycoplasma and maintained in a humidified incubator at 37°C and 5% CO<sub>2</sub>. Passaging was carried out either mechanically with gentle trypsinization using trypsin solution A without EDTA (Biological Industries), or enzymatically using TrypLE Express (Thermo Fisher Scientific) with addition of 10  $\mu$ M ROCK inhibitor Y-27632 (Stemgent) for 1 day after splitting. Haploid ES cells could also be grown in feeder-free conditions on Matrigel-coated plates (Corning) in mTeSR1 (STEMCELL Technologies) or StemFitN.AK03 (Ajinomoto) media.

**Isolation and maintenance of haploid human ES cell lines.** Following identification of haploid cells in human parthenogenetic ES cell lines at passages 6–7 by either metaphase spread analysis or sub-2c-cell sorting (see below and Extended Data Tables 1 and 2), haploid ES cell lines were established by sorting the 1c-cell population, with diploid cells serving as a reference. Haploid ES cell cultures were further maintained by enrichment rounds of 1c-cell sorting every 3–4 passages.

**Metaphase spread analysis.** For induction of mitotic arrest, growing cells were incubated for 40 min in the presence of 100 ng ml<sup>-1</sup> colcemid (Biological Industries), added directly to the culture medium in a humidified incubator at 37°C with 5% CO<sub>2</sub>. The cells were then trypsinized, centrifuged at 1,000 r.p.m. at room temperature and gently resuspended in 37°C warmed hypotonic solution (2.8 mg ml<sup>-1</sup> KCl and 2.5 mg ml<sup>-1</sup> sodium citrate) followed by 20 min of incubation at 37°C. Cells were fixed by addition of fixative solution (3:1 methanol:acetic acid) and incubation for 5 min at room temperature. Fixation was repeated at least three times following centrifugation and resuspension in fixative solution. Metaphase spreads were prepared on slides and stained using the standard G-banding technique. Karyotype integrity was determined according to the International System for Human Cytogenetic Nomenclature (ISCN) based on the observation of a normal karyotype in at least 80% of analysed metaphases (minimum of 20 metaphases per analysis).

**Live ES cell sorting by DNA content.** Cells were washed with phosphate buffered saline (PBS), dissociated using either TrypLE Select or TrypLE Express (Thermo Fisher Scientific) and stained with 10  $\mu$ g ml<sup>-1</sup> Hoechst 33342 (ref. 2) (Sigma-Aldrich) in human ES cell medium at 37°C for 30 min. Following centrifugation, cells were resuspended in PBS containing 15% KSR and 10  $\mu$ M ROCK inhibitor Y-27632, filtered through a 70- $\mu$ m cell strainer (Corning) and sorted using the 405 nm laser in either BD FACSAria III or BD Influx (BD Biosciences). For continued growth, sorted cells were plated with fresh medium containing 10  $\mu$ M ROCK inhibitor Y-27632 for 24 h. For comparative analyses, G1-phase cells were sorted from isogenic haploid-enriched and unsorted diploid cultures. Cells that had undergone diploidization relatively recently in culture (within 3 passages after haploid cell enrichment) were isolated by sorting the 4c peak in haploid-enriched cultures and compared with 4c diploid cells from unsorted diploid cultures. Note that haploid-enriched cultures also consist of a mixed 2c-cell population of G2/M-phase haploids and G1-phase diploids. Sorting purity was confirmed by rerunning a fraction of sorted samples through the instrument.

**Flow cytometry.** All DNA content profiles were generated based on flow cytometry with Hoechst 33342 staining. Haploid cell proportion was estimated based

on the percentage of 1c cells and the relative contribution of G1 cells with regards to other phases of the cell cycle. Estimation of diploidization rate was based on the proportion of haploid cells between consecutive enrichment rounds as well as experimental analysis of h-pES10 diploidization kinetics throughout 7 passages (30 days) by analysing the DNA content of 2–3 replicates at each passage using flow cytometry with propidium iodide in methanol-fixed and RNase-treated cells. Diploidization rate was estimated by fitting the data to an exponential decay curve. For simultaneous flow cytometry analysis of DNA content and cell surface molecules, cells were washed, dissociated and incubated on ice for 30 min in the presence of 10  $\mu$ g ml<sup>-1</sup> Hoechst 33342 (Sigma-Aldrich) and either a conjugated antibody or a secondary antibody diluted 1:200 following a 60 min incubation with a primary antibody. For simultaneous flow cytometry analysis of DNA content and intracellular PDX1, dissociated cells were treated as described for immunofluorescence procedures, with Hoechst 33342 for DNA staining. Primary antibodies are detailed in Supplementary Table 1. In all flow cytometry procedures, samples were filtered through a 70- $\mu$ m cell strainer (Corning Life Sciences) and analysed with either BD FACSAria III or BD Influx (BD Biosciences).

**DNA fluorescence *in situ* hybridization.** DNA FISH was performed as described elsewhere<sup>25</sup> using probes for human chromosomes 2 and 4 and DNA staining with 4',6-diamidino-2-phenylindole (DAPI). Haploidy and diploidy were respectively determined per nucleus based on single or double hybridization signals. ES cells subject to FISH were grown on Matrigel-coated plates in StemFitN.AK03 medium for several passages before analysis.

**Alkaline phosphatase and immunofluorescence staining.** Alkaline phosphatase staining was performed using the Leukocyte Alkaline Phosphatase Kit (Sigma-Aldrich). For immunofluorescence staining, samples were washed with PBS, fixed with 4% paraformaldehyde for 10 min, and permeabilized and blocked in blocking solution (0.1% Triton X-100 and 5% donkey serum in PBS). Cells were incubated with primary antibodies (detailed in Supplementary Table 1) and secondary antibodies diluted 1:500 in blocking solution, and DAPI was used for DNA staining. Cells were washed twice with PBS subsequently to fixation and each incubation step. Images were taken using Zeiss LSM 510 Meta Confocal Microscope. Centromere quantification was carried out by manually counting centromere foci across individual planes along the z axis. EdU staining was performed using the Click-iT EdU Alexa Fluor 488 Imaging Kit (Thermo Fisher Scientific). ES cells subject to centromere staining in Fig. 1e and Extended Data Fig. 1e were grown on Matrigel-coated plates in StemFitN.AK03 for several passages before analysis.

**6-TG resistance screen.** To generate a gene trap mutant library, 9 replicates of approximately 4  $\times$  10<sup>6</sup> haploid pES10 cells (within one passage after 1c-cell enrichment) were co-transfected with 20  $\mu$ g 5'-PTK-3' gene trap vector<sup>26</sup> and 20  $\mu$ g pCyl43 piggyBac transposase plasmid<sup>27</sup> using Bio-Rad Gene Pulser (suspended in 800  $\mu$ L Opti-MEM, 4-mm cuvettes, 320 V, 250  $\mu$ F), and replated on a 100  $\times$  20 mm dish with DR3 MEFs and ROCK inhibitor Y-27632. Selection for insertions into expressed loci was carried out using 0.3  $\mu$ g ml<sup>-1</sup> puromycin starting 48 h post transfection, followed by pooling into a single library, represented by approximately 16,000 resistant colonies. Transfection with 5'-PTK-3' only was used as a negative control. To screen for 6-TG-resistant mutants, the mutant library was grown in the presence of 6  $\mu$ M 6-TG (Sigma-Aldrich) on DR4 MEFs for 18 days, during which 6 resistant colonies were independently isolated and characterized. Analysis of a resistant clone showed persistence of haploid cells. Genomic DNA was extracted (NucleoSpin Tissue Kit, MACHERY-NAGEL) and insertion sites were detected using splinkerette PCR as described previously<sup>28</sup>, followed by PCR product purification and Sanger sequencing (ABI PRISM 3730xl DNA Analyzer (Applied Biosystems)). Sequences were mapped to the human genome (GRCh38/hg38) using UCSC BLAT search tool.

**Isolation of total DNA and RNA.** Total DNA was isolated using the NucleoSpin Tissue Kit (MACHERY-NAGEL). Total RNA was isolated using Qiagen RNeasy Kits according to the manufacturer's protocols. To determine total RNA levels per cell, haploid and diploid cells were isolated from the same cultures by sorting the 1c (haploid G1) and 4c (diploid G2/M) populations, respectively. Following growth for 2 passages, cells were harvested and counted, and RNA was isolated from triplicates of 400,000 cells from each cell line and ploidy state (pES10 and pES12, haploid and diploid; 12 samples in total). RNA amounts were quantified using NanoDrop. **Genome integrity analysis.** Copy number variation (CNV) analysis was carried out on DNA samples of G1-sorted haploid and diploid pES10 and pES12 cells (see Supplementary Table 2) using Infinium Omni2.5Exome-8 BeadChip single nucleotide polymorphism (SNP) arrays (Illumina) following the manufacturer's protocols. Raw data were processed using Genome Studio Genotyping Module (Illumina) to obtain log R ratios values for analysis using R statistical programming language. As expected, diploid pES10 and pES12 cells were homozygous across all chromosomes.

**RNA sequencing.** For a detailed list of samples analysed by RNA-seq, see Supplementary Table 3. Total RNA samples (200 ng–1 µg, RNA integrity number (RIN) >9) were enriched for mRNAs by pulldown of poly(A)<sup>+</sup> RNA. RNA-seq libraries were prepared using the TruSeq RNA Library Prep Kit v2 (Illumina) according to the manufacturer's protocol and sequenced using Illumina NextSeq 500 to generate 85 bp single-end reads.

**Transcriptome analysis.** RNA-seq reads were aligned to the human reference genome (GRCh37/hg19) using TopHat (version 2.0.8b) allowing 5 mismatches. Reads per kilobase per million fragments mapped (RPKM) values were quantified using Cuffquant and normalized using Cuffnorm in Cufflinks (version 2.1.1) to generate relative gene expression levels. Hierarchical clustering analyses were performed on RPKM values using Pearson correlation and average linkage. Analysis of differential gene expression relative to total RNA in haploid and diploid human ES cells ( $n = 4$  in each group) was carried out by two complementary strategies, as follows: first, we used Cuffdiff with default parameters, considering differences of greater than twofold with FDR < 0.05 as significant; second, to identify possibly subtle yet consistent transcriptional differences, we tested for genes whose minimal expression levels across all replicates of a certain group were higher than their maximal expression level across all replicates of the other group. Statistical significance was then determined by two-tailed unpaired Student's *t*-test. Functional annotation enrichment analysis was done by DAVID (using the Benjamini method to determine statistical significance). Imprinting analyses included 75 human imprinted genes (<http://www.geneimprint.com/>), listed in Supplementary Table 4. RNA-seq data from control ES cell line NYSCF1 were published elsewhere<sup>29</sup> (GEO accession number GSE61657). Genome-wide gene expression moving median plots were generated using the R package zoo (version 1.7–12) after removal of genes that were not expressed in the averaged reference diploid sample by flooring to 1 and setting an expression threshold of above 1. RNA-seq data from different tissues were retrieved from the Genotype-Tissue Expression (GTEx) portal (<http://www.gtexportal.org/>)<sup>30</sup>. Colour-coded scales in Fig. 4d correspond to gene expression levels relative to the mean across tissues (left scale) and across each set of ES cell duplicate and EB sample (right scale). Expression microarray analysis was performed as previously<sup>31</sup> by using Affymetrix Human Gene 1.0 ST arrays.

**DNA methylation analysis.** DNA methylation analysis was performed on genomic DNA from the samples detailed in Supplementary Table 2 using Infinium HumanMethylation450 BeachChips (Illumina) following the Infinium HD Methylation Protocol as described previously<sup>29</sup>. DNA methylation data from control ES cell line NYSCF1 were published before<sup>29</sup> (GEO accession number GSE61657). Data were processed and normalized by using subset-quantile within array normalization (SWAN) and adjusted for batch effects using the R package ChAMP (version 1.4.0). DNA methylation levels at CpG sites associated with pluripotency-specific genes and iDMRs were analysed as described before<sup>29</sup>. For analysis of DNA methylation levels on the X chromosome, probes with average  $\beta$  values of less than 0.4 were filtered out. DMR analysis was facilitated by the lasso function in ChAMP using default settings. DMRs were then assigned to genes by proximity and analysed for functional annotation enrichment using DAVID (using the Benjamini method to determine statistical significance).

**Cell size analysis.** Following sorting of haploid and diploid cell populations in G1, the diameter ( $2r$ ) of viable single cells was measured by Countess Automated Cell Counter (Invitrogen) and their surface area and volume were calculated as  $4\pi r^2$  and  $4/3\pi r^3$ , respectively. Analysis included 7, 4, 8 and 4 technical replicates for 1n pES10, 1n pES12, 2n pES10 and 2n pES12, respectively.

**Mitochondrial DNA abundance analysis.** Relative mtDNA abundance was analysed by quantitative PCR (qPCR) by using primers for the mitochondrial gene *MT-ND2* (forward primer: 5'-TGTTGGTTATACCCCTCCCGTACTA-3'; reverse primer: 5'-CCTGCAAAGATGGTAGAGTAGATGA-3') and normalization to nuclear DNA by using primers for the nuclear gene *BECN1* (forward primer: 5'-CCCTCATCACAGGGCTCTCTCCA-3'; reverse primer: 5'-GGGACTGTAGGCTGGGAAGTATGC-3'), as described elsewhere<sup>32</sup>. Analysis was performed using Applied Biosystems 7300 Real-Time PCR System with PerfeCTa SYBR Green FastMix (Quanta Biosciences). Analysis included all G1-sorted samples detailed in Supplementary Table 2 ( $n = 4$  for each group, with two biological replicates for each cell line).

**Embryoid body differentiation.** EB differentiation was carried out by detaching ES cell colonies with Trypsin solution A without EDTA (Biological Industries), followed by resuspension and further culture of cell aggregates in human ES cell medium without bFGF on low attachment plates. Differentiation of haploid ES cells was initiated within 2 passages after 1c-cell enrichment. After 21 days, EB RNA was extracted from unsorted and/or sorted EB cells in G1 following dissociation and staining with 10 µg ml<sup>-1</sup> Hoechst 33342 (Sigma-Aldrich) at 37 °C for 30 min.

Metaphase spread analysis was performed on dissociated EB cells plated on 0.2% gelatin and expanded in human ES cell medium without bFGF.

**Differentiation into neural progenitor cells.** NCAM1-positive ES cell-derived neural progenitor cells were obtained using a 10-days protocol for efficient neural differentiation<sup>33</sup> with slight modification<sup>34</sup>. Differentiation was initiated within 2 passages after 1c-cell enrichment. RNA was extracted from sorted haploid NCAM1-positive cells in G1 by co-staining with Hoechst 33342 and an anti-human NCAM-1/CD56 primary antibody (see Supplementary Table 1) and a Cy3-conjugated secondary antibody (Jackson ImmunoResearch Laboratories) diluted 1:200.

**Neuronal differentiation.** Differentiation into neurons was carried out by following a published protocol<sup>35</sup> based on synergistic inhibition of SMAD signalling<sup>36</sup> with modification, as follows: differentiation was initiated within 2 passages after 1c-cell enrichment with fully confluent ES cells cultured on Matrigel-coated plates in mTeSR1 by replacing the medium with human ES cell medium without bFGF, containing 10 µM SB431542 (Selleckchem) and 2.5 µM LDN-193189 (Stemgent) for 4 days. Subsequently, cells were kept in N2 medium<sup>35</sup> supplemented with 10 µM SB431542 and 2.5 µM LDN-193189 for an additional 4 days, followed by 2 days in N2 medium supplemented with B-27 (Thermo Fisher Scientific) and 10 µM DAPT (Stemgent). The cells were then dissociated and replated on 0.01% poly-L-ornithine coated (Sigma-Aldrich) and laminin coated (4 µg ml<sup>-1</sup>, Thermo Fisher Scientific) plates in the presence of 10 µM ROCK inhibitor Y-27632 (Selleckchem), and further cultured in the same medium without Y-27632 for the next 4 days. Neuronal cultures were maintained in N2 medium supplemented with B-27 and 20 ng ml<sup>-1</sup> BDNF (R&D) until analysis by immunostaining and FISH on day 20.

**Cardiomyocyte differentiation.** 80–90% confluent ES cells grown on Matrigel-coated plates in mTeSR1 were subject to an 11-day regimen<sup>37</sup> based on consecutive GSK3 and WNT inhibition with CHIR99021 and IWP-2 (Selleckchem), respectively. Differentiation was initiated within 2 passages after 1c-cell enrichment. On day 11 of differentiation, 1c cells were sorted and plated for immunostaining.

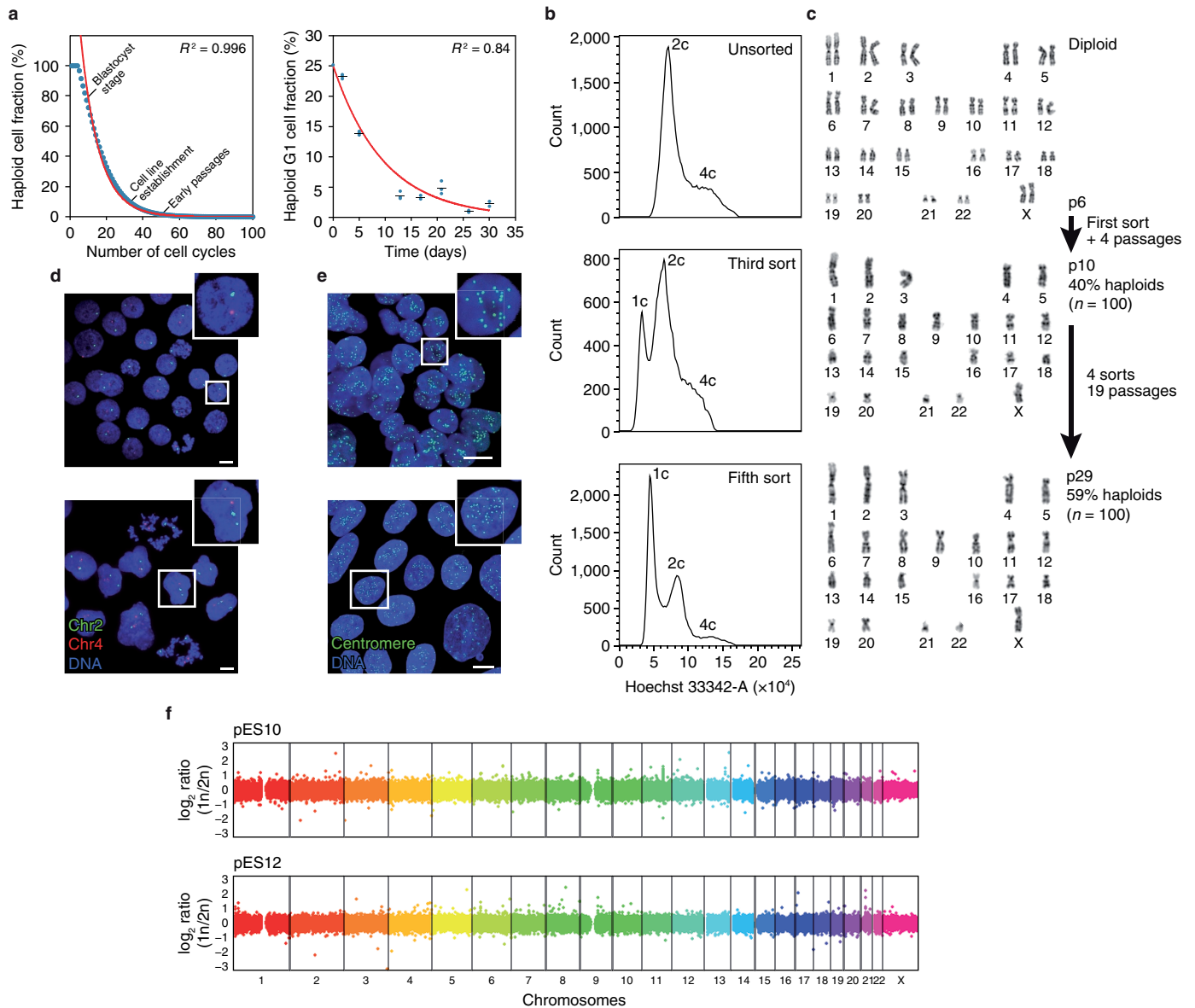
**Differentiation towards the pancreatic lineage.** The protocol used here was developed based on several recent publications<sup>38–40</sup>. ES cells grown in feeder-free conditions were differentiated into definitive endoderm by using STEMdiff Definitive Endoderm Kit (StemCell Technologies) for 3–4 days. Subsequent specification was achieved by a step-wise protocol involving treatment with recombinant human KGF/FGF7 (R&D Systems), LDN-193189 (Stemgent), KAAD-cyclopamine (Stemgent) and retinoic acid (Stemgent). On days 8–11, EGF (R&D System) was used to induce pancreatic cells. Differentiation was initiated within as few as 2 passages after 1c-cell enrichment.

**Teratoma formation assay.** All experimental procedures in animals were approved by the ethics committee of the Hebrew University. ES cells were trypsinized and approximately  $2 \times 10^6$  cells were resuspended in 100 µl human ES cell medium and 100 µl Matrigel (BD Biosciences), followed by subcutaneous injection into NOD-SCID *Il2rg*<sup>-/-</sup> immunodeficient mice (Jackson Laboratory). 8–12 weeks after injection tumours were dissected and subjected to further analysis. Histological slides were prepared from tumour slices cryopreserved in O.C.T. compound (Sakura Finetek) using Leica CM1850 cryostat (Leica Biosystems, 10-µm sections), followed by immunostaining, haematoxylin and eosin staining or FISH analysis. Flow cytometry with Hoechst 33342 staining was performed on dissociated cells from freshly dissected tumours.

- Noggle, S. *et al.* Human oocytes reprogram somatic cells to a pluripotent state. *Nature* **478**, 70–75 (2011).
- Chen, A. E. *et al.* Optimal timing of inner cell mass isolation increases the efficiency of human embryonic stem cell derivation and allows generation of sibling cell lines. *Cell Stem Cell* **4**, 103–106 (2009).
- Rao, P. H., Nandula, S. V. & Murty, V. V. Molecular cytogenetic applications in analysis of the cancer genome. *Methods Mol. Biol.* **383**, 165–185 (2007).
- Cadiñanos, J. & Bradley, A. Generation of an inducible and optimized piggyBac transposon system. *Nucleic Acids Res.* **35**, e87 (2007).
- Wang, W. *et al.* Chromosomal transposition of PiggyBac in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* **105**, 9290–9295 (2008).
- Chen, L. *et al.* Transposon activation mutagenesis as a screening tool for identifying resistance to cancer therapeutics. *BMC Cancer* **13**, 93 (2013).
- Johannesson, B. *et al.* Comparable frequencies of coding mutations and loss of imprinting in human pluripotent cells derived by nuclear transfer and defined factors. *Cell Stem Cell* **15**, 634–642 (2014).
- Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nature Genet.* **45**, 580–585 (2013).
- Yamada, M. *et al.* Human oocytes reprogram adult somatic nuclei of a type 1 diabetic to diploid pluripotent stem cells. *Nature* **510**, 533–536 (2014).
- Wanet, A. *et al.* Mitochondrial remodeling in hepatic differentiation and dedifferentiation. *Int. J. Biochem. Cell Biol.* **54**, 174–185 (2014).
- Kim, D.-S. *et al.* Robust enhancement of neural differentiation from human ES and iPSC cells regardless of their innate difference in differentiation propensity. *Stem Cell Rev.* **6**, 270–281 (2010).

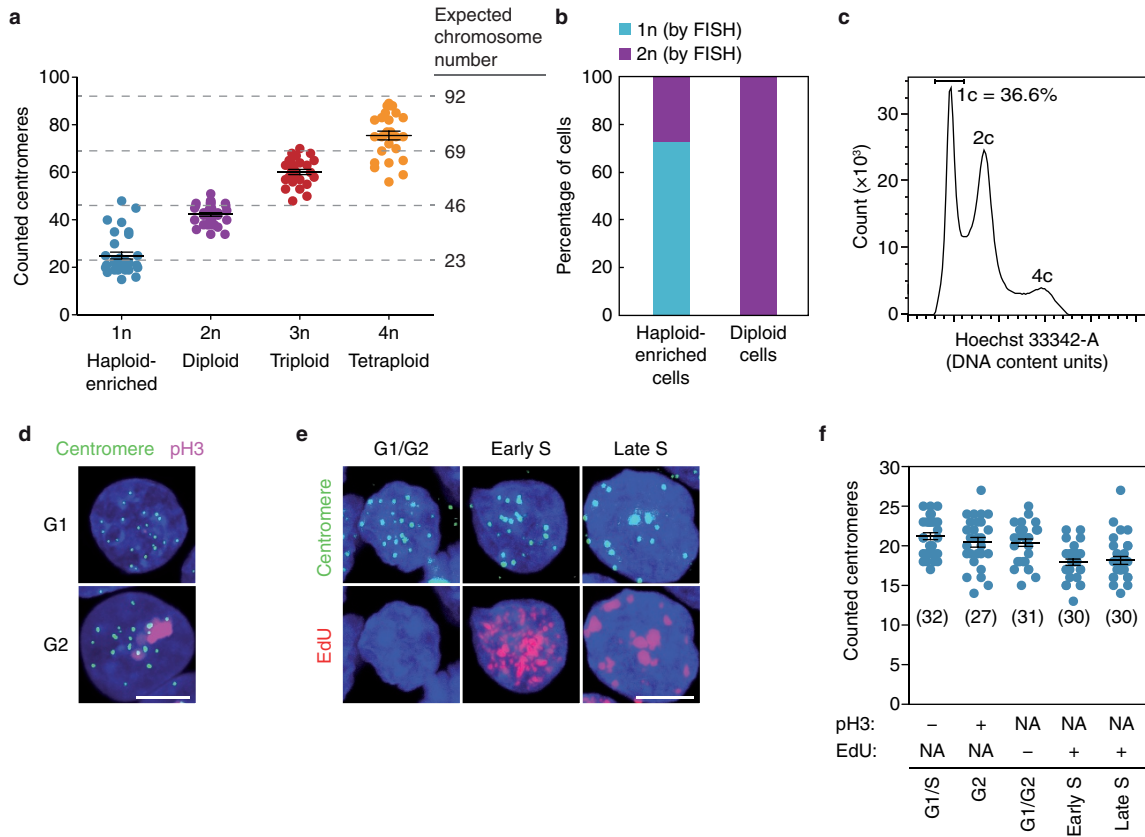
34. Stelzer, Y., Sagi, I. & Benvenisty, N. Involvement of parental imprinting in the antisense regulation of onco-miR-372-373. *Nat. Commun.* **4**, 2724 (2013).
35. Wang, L. *et al.* Differentiation of hypothalamic-like neurons from human pluripotent stem cells. *J. Clin. Invest.* **125**, 796–808 (2015).
36. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature Biotechnol.* **27**, 275–280 (2009).
37. Lian, X. *et al.* Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/ $\beta$ -catenin signaling under fully defined conditions. *Nature Protocols* **8**, 162–175 (2013).
38. Hua, H. *et al.* iPSC-derived  $\beta$  cells model diabetes due to glucokinase deficiency. *J. Clin. Invest.* **123**, 3146–3153 (2013).
39. Pagliuca, F. W. *et al.* Generation of functional human pancreatic  $\beta$  cells *in vitro*. *Cell* **159**, 428–439 (2014).
40. Reznica, A. *et al.* Reversal of diabetes with insulin-producing cells derived *in vitro* from human pluripotent stem cells. *Nature Biotechnol.* **32**, 1121–1133 (2014).
41. Cowan, C. A., Atienza, J., Melton, D. A. & Eggan, K. Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science* **309**, 1369–1373 (2005).





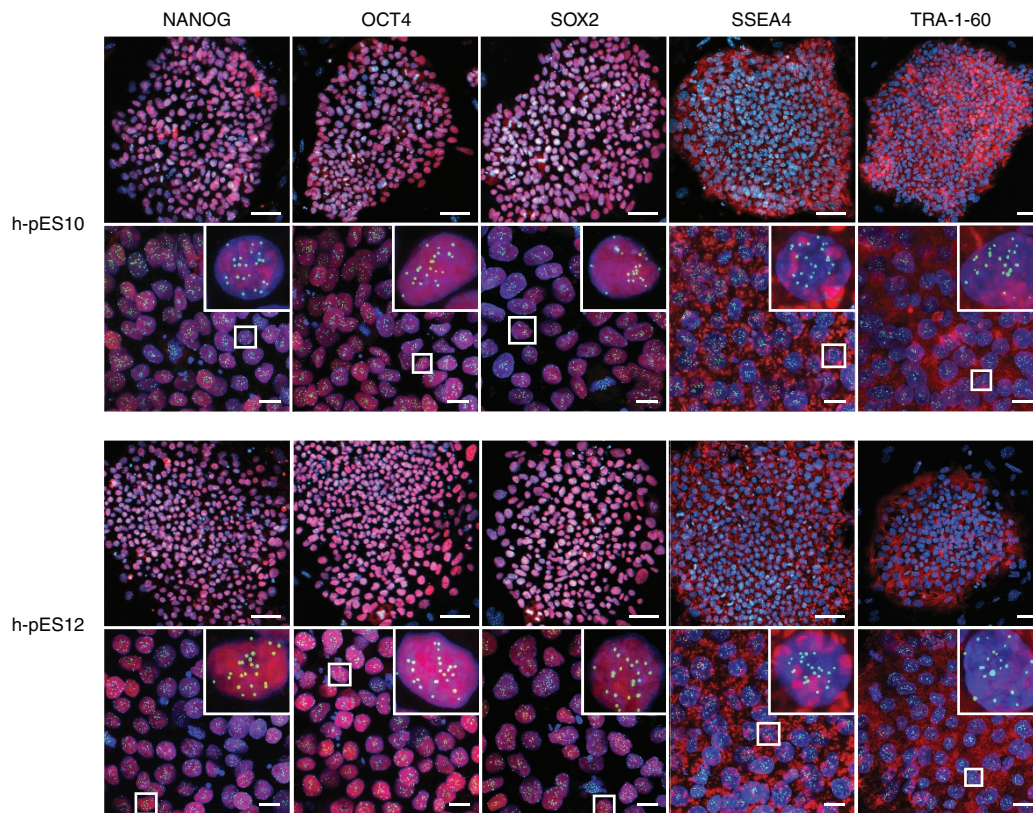
**Extended Data Figure 1 | Derivation of haploid human ES cell lines h-pES10 and h-pES12.** **a**, Left panel, diploidization rate model for a haploid egg with a theoretical diploidization probability of 10%, overlaid with an exponential decay fit (red curve, see Supplementary Notes). Approximated cell cycle numbers for different ES cell line derivation stages are indicated. Right panel, diploidization dynamics of h-pES10 over seven passages by flow cytometry, overlaid with an exponential fit to the data (red curve). Bars indicate the mean across biological replicates. **b**, Establishment of a haploid-enriched human ES cell line from pES12 cells after repeated sorting and enrichment of 1c cells using Hoechst 33342 staining. Top to bottom, DNA content profiles of

unsorted diploid cells, partially purified haploid cells at the third sort, and mostly-purified haploid cells at the fifth sort. **c**, Karyotypes and haploid metaphase percentage over the course of enrichment and passaging. **d, e**, DNA FISH (**d**) and centromere protein immunofluorescence staining (**e**) in haploid-enriched pES12 cells (upper panels) and unsorted diploid pES10 cells (lower panels). Magnified insets show representative haploid and diploid nuclei with single or double hybridization signals (**d**) and 23 or 46 centromeres (**e**), respectively. Scale bars, 10  $\mu$ m. **f**, SNP array-based CNV analysis comparing haploid pES10 and pES12 cells with their unsorted diploid counterparts (logarithmic scale).



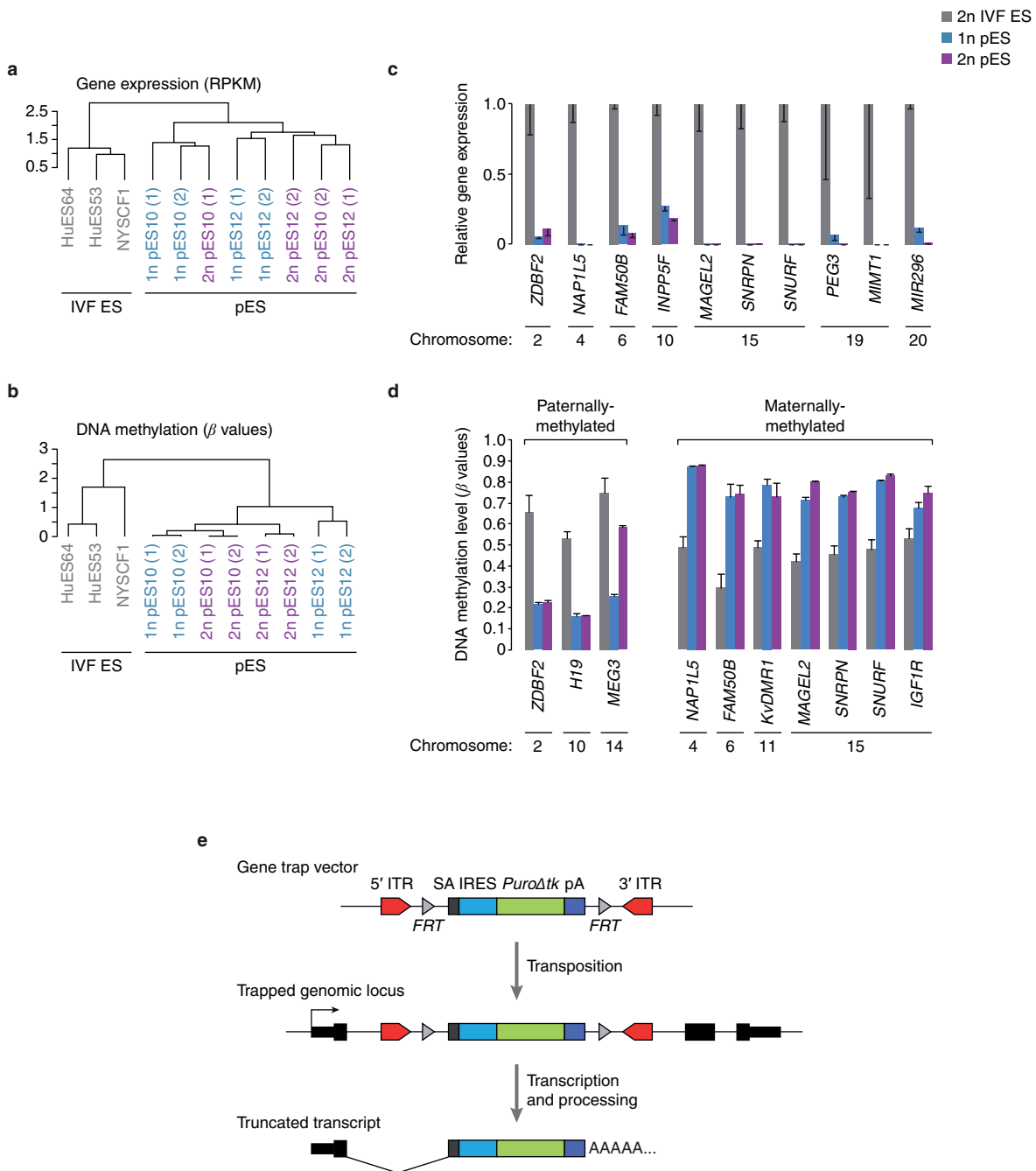
**Extended Data Figure 2 | Determination of ploidy at single-cell level by quantification of centromere foci.** **a**, The counted number of centromeres correlates with ploidy. 1n, haploid-enriched pES10 cells grown for 4 passages after the fourth sort ( $n = 33$ ; 76% haploids by this assay); 2n, unsorted diploid pES10 cells ( $n = 34$ ); 3n, soPS2 cells<sup>23</sup> ( $n = 27$ ); 4n, Hybrid1 cells<sup>41</sup> ( $n = 27$ ). Black horizontal lines indicate mean  $\pm$  s.e.m. and dashed lines mark expected chromosome numbers. **b**, Quantification of haploid and diploid cells by DNA FISH in the haploid-enriched ( $n = 152$ ; 73% haploids by this assay) and diploid ( $n = 135$ ) cells in **a**. **c**, DNA

content profile of the haploid-enriched cells in **a** (73% haploids by this assay). **c**, chromosomal copies. **d**, **e**, Co-staining of centromeres and either phospho-histone 3 (pH3, Ser10) (**d**) or 5-ethynyl-2'-deoxyuridine (EdU) (**e**) for distinguishing between different stages of interphase in haploid pES12 cells. In blue, DNA staining. Scale bar, 5  $\mu$ m. **f**, Quantification of centromere counts in the different cell cycle stages shown in **d** and **e**.  $n$  indicated in parenthesis. Black horizontal lines indicate mean  $\pm$  s.e.m. See Supplementary Notes for details.



**Extended Data Figure 3 | Pluripotent stem cell markers in haploid human ES cells.** Co-staining of pluripotency markers NANOG, OCT4, SOX2, SSEA4 and TRA-1-60 (red), centromeres (green) and DNA (blue)

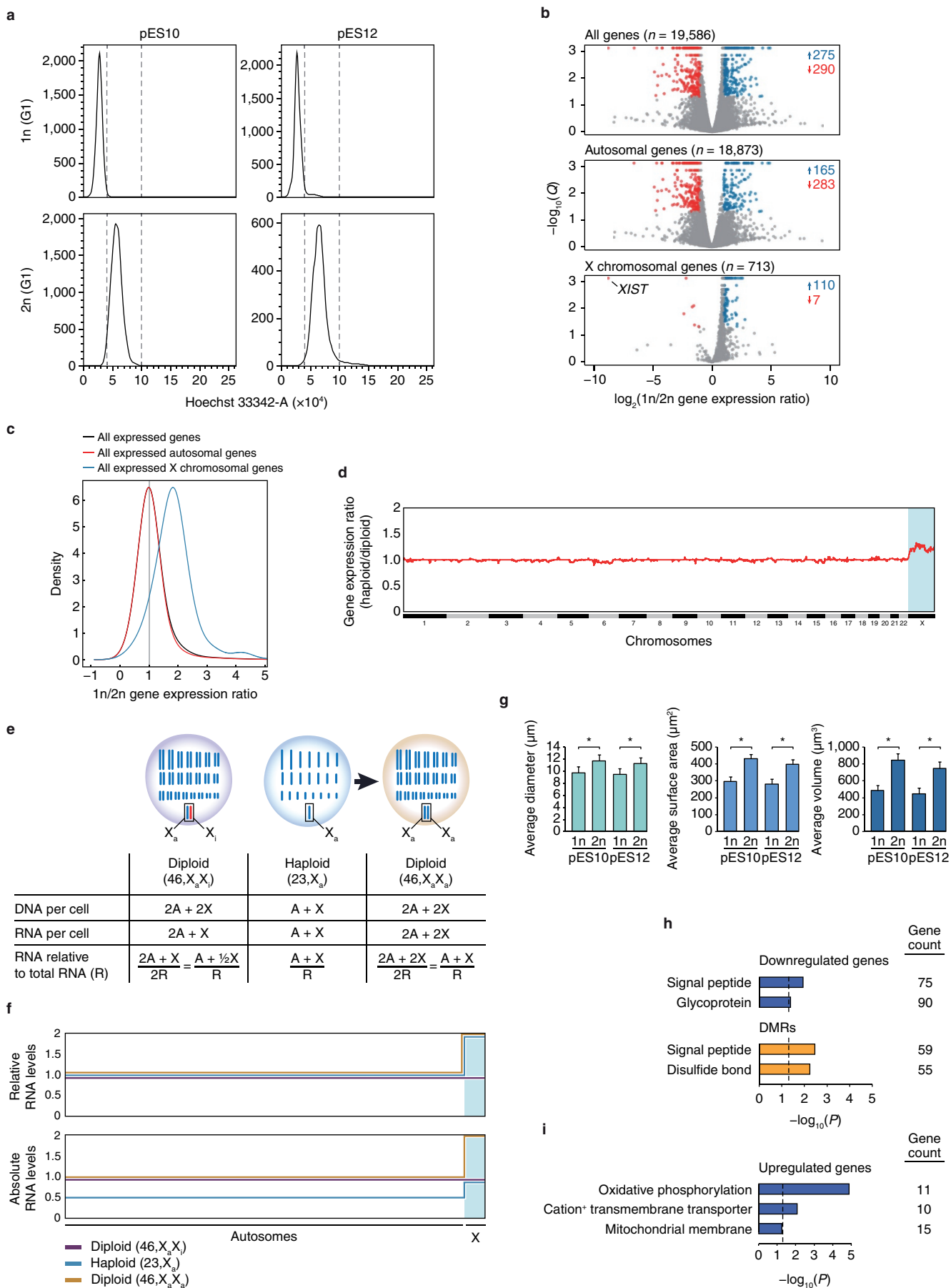
in h-pES10 and h-pES12 at colony resolution (upper panels; scale bars, 50  $\mu\text{m}$ ) and single-cell resolution (lower panels; scale bars, 10  $\mu\text{m}$ ). Magnified insets show representative haploid cells with 23 centromeres.



#### Extended Data Figure 4 | Analysis of parental imprinting and gene trap mutagenesis in haploid human parthenogenetic ES cells.

**a, b**, Clustering analysis of diploid *in vitro* fertilization (IVF) ES cells and G1-sorted haploid and diploid parthenogenetic ES (pES) cells by expression levels of imprinted genes ( $n = 75$ , see Supplementary Table 4) (**a**) and DNA methylation levels at imprinted differentially methylated regions (iDMRs,  $n = 35$ )<sup>29</sup> (**b**). The symbols (1) and (2) indicate biological replicates. **c**, Relative mean expression levels  $\pm$  s.e.m. of representative paternally expressed imprinted genes across seven chromosomes in the samples shown in **a** (RPKM ratios). **d**, Mean DNA methylation levels  $\pm$  s.e.m. at representative paternally methylated and maternally methylated iDMRs (typically intermediately methylated in bi-parental

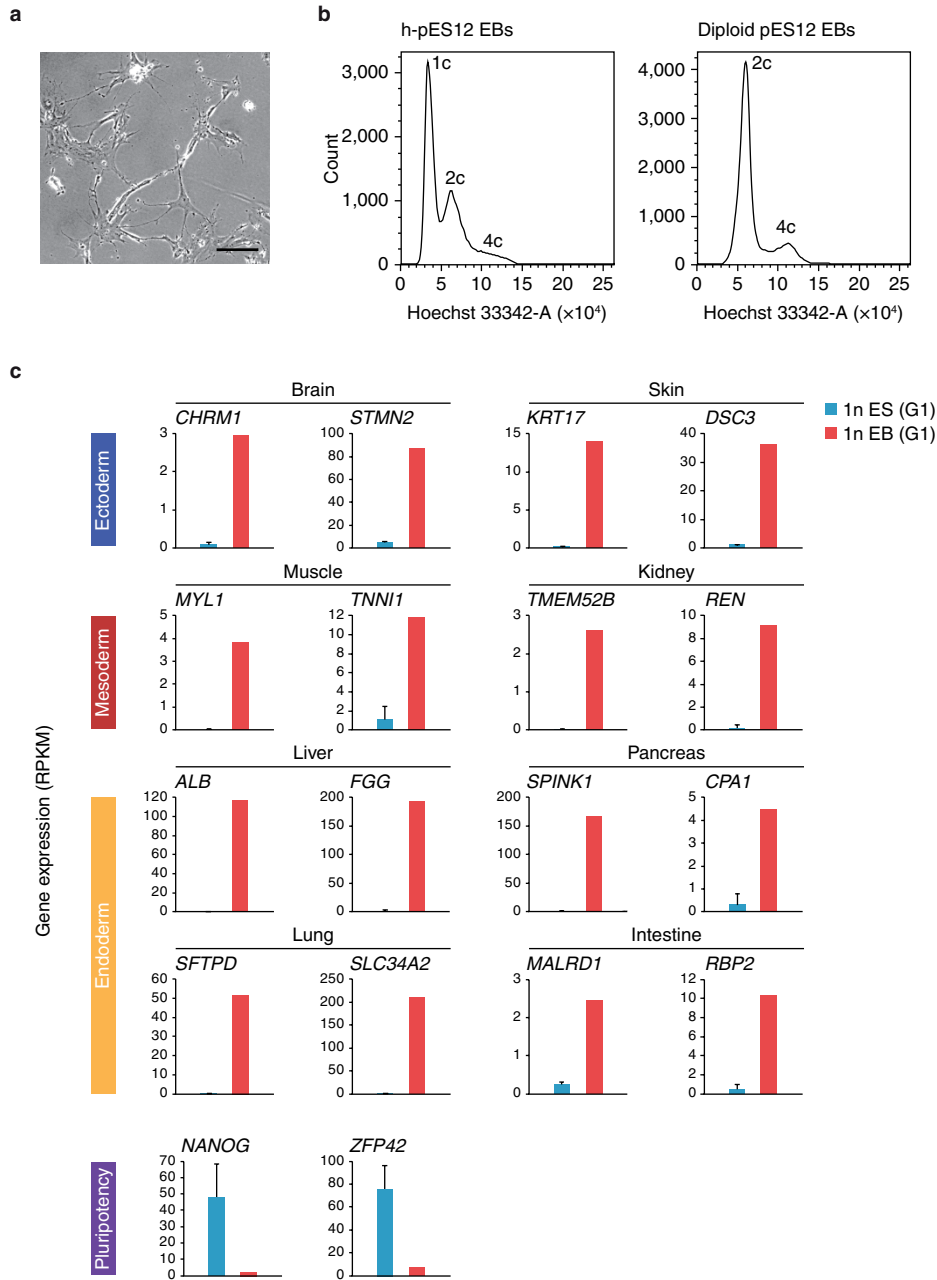
control cells, and respectively hypomethylated and hypermethylated in parthenogenetic cells) in the samples shown in **b**.  $\beta$  values range from complete hypomethylation (0) to complete hypermethylation (1). **e**, Schematic outline of the *piggyBac* gene trap system. The gene trap vector<sup>26</sup> is flanked by *piggyBac* inverted terminal repeats (ITRs) and *FRT* sites, and carries a 5' splice acceptor (SA), an internal ribosome entry site (IRES) element followed by a promoterless puromycin resistance gene (*Puro $\Delta$ tk*) and a 3' poly(A) signal (pA). In the presence of the *PiggyBac* transposase (encoded on a separate plasmid<sup>27</sup>, not shown), the gene trap vector undergoes random transposition into the genome. Insertion into a transcriptionally active gene results in truncation of the endogenous transcript and introduction of resistance to puromycin.



Extended Data Figure 5 | See next page for caption.

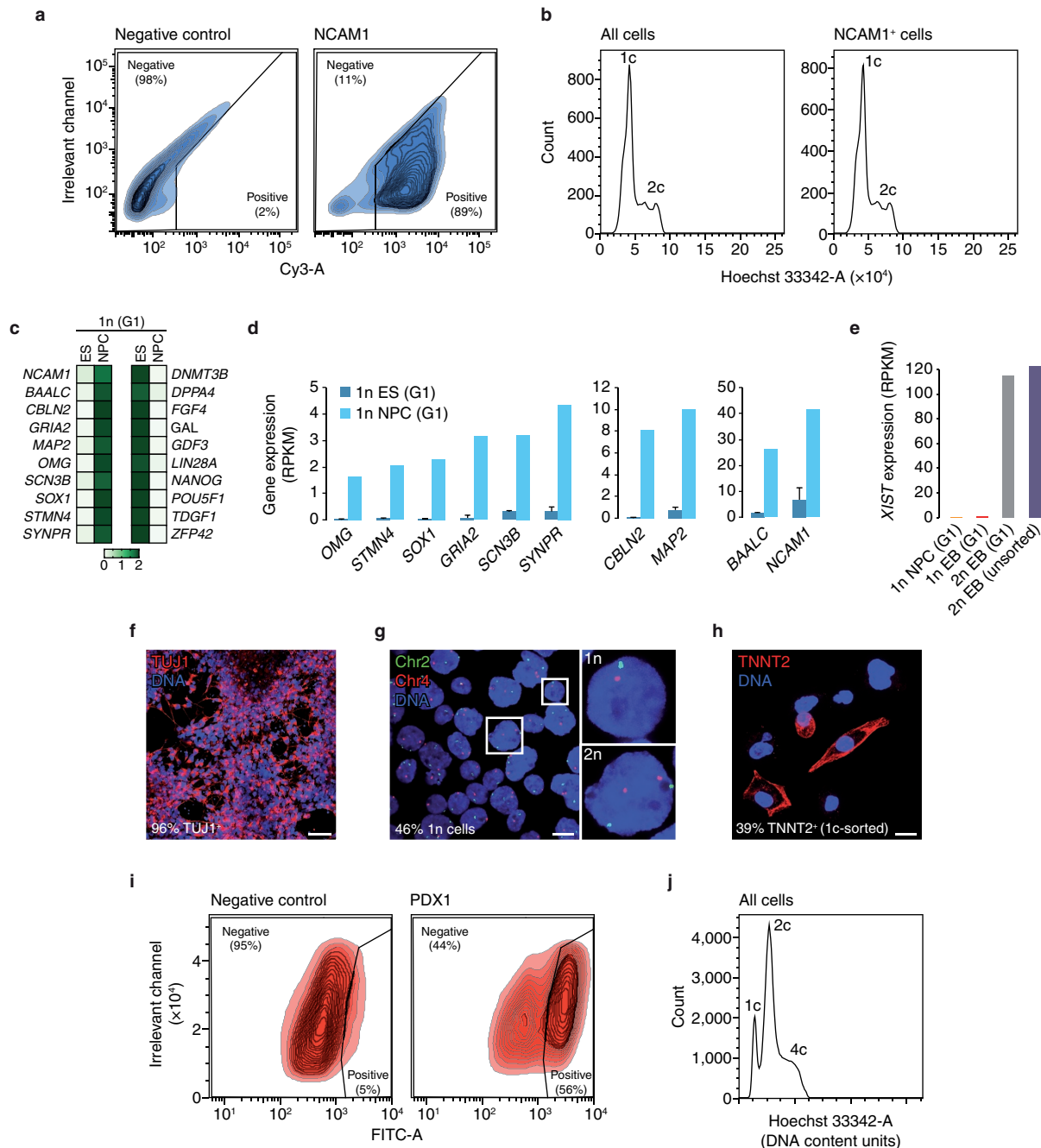
**Extended Data Figure 5 | Comparative analyses of isogenic haploid and diploid human ES cells.** **a**, Sorting purity of haploid and diploid ES cells in G1. **b**, Log-scaled volcano plots of relative differential gene expression between haploid and diploid human ES cells, divided into panels by all genes (top), autosomal genes (middle) and X chromosomal genes (bottom).  $Q$ , false discovery rate (FDR). Significantly downregulated and upregulated genes (greater than twofold change,  $Q < 0.05$ ) in haploid cells are marked in red and blue, respectively, and their totals are indicated to the right. Note that *XIST* is the most downregulated transcript in haploid cells. **c**, Smoothed distributions of the  $1n/2n$  gene expression ratios for all expressed genes, all expressed autosomal genes and all expressed X chromosomal genes (expression threshold, mean RPKM  $> 0.1$ ). **d**, Genome-wide moving median plot of the gene expression ratio between haploid and diploid pES10 cells in G1 by expression microarray analysis (window size = 100 genes). **e**, **f**, Model for genome-wide autosomal

gene expression level reduction in haploid human ES cell as inferred by differential X chromosome inactivation status. **e**, DNA content, RNA expression levels relative to total RNA and presumed equality of absolute X chromosomal gene dosage in haploid ( $X_a$ ) and diploid ( $X_aX_i$ ) human ES cells, enable the estimation of total RNA levels per haploid cell.  $X_a$  and  $X_i$  denote active (blue) and inactive (red) X chromosomes, respectively. A, autosomes; X, X chromosome; R, total RNA. **f**, Schematic genome-wide representation of relative and absolute RNA levels in the cells shown in **e**. **g**, Average diameter and calculated surface area and volume of G1-sorted haploid and diploid ES cells. Error bars represent s.d. \* $P < 0.01$  (two-tailed unpaired Student's  $t$ -test). **h**, **i**, Functional annotation enrichment analysis for relatively downregulated genes and differentially methylated regions (DMRs) (**h**), as well as relatively upregulated genes (**i**) in haploid ES cells compared with diploid ES cells.



**Extended Data Figure 6 | EB differentiation of haploid human ES cells.**  
**a**, Representative image of plated cells dissociated from h-pES12-derived 21-day EBs. Karyotype is shown in Fig. 4b. Scale bar, 100  $\mu$ m. **b**, DNA content profiles of dissociated EBs derived from haploid-enriched and

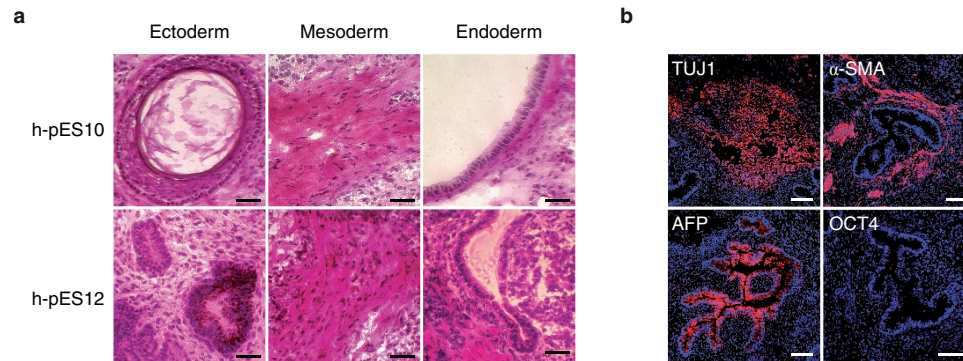
diploid pES12 cells. **c**, Expression levels (RPKM) of tissue- and pluripotency-specific genes in undifferentiated (ES) and differentiated (EB) G1-sorted haploid (1n) pES10 cells.



**Extended Data Figure 7 | Directed differentiation of haploid human ES cells.** **a, b**, Flow cytometry analysis with co-staining of DNA and NCAM1 in h-pES10 cells following neural differentiation. **a**, Gating for NCAM1-positive cells (right) based on a negative secondary antibody stained control sample (left). **b**, DNA content profiles of the entire cell population (left) and NCAM1-positive cells (right). **c**, chromosomal copies. **c, d**, Expression levels of neural- and pluripotency-specific genes in haploid G1-sorted ES cells and NPCs. **e**, *XIST* expression levels in haploid and diploid pES10-derived EBs and NPCs. **f**, TUJ1 staining in h-pES12-derived neurons. Scale bar, 100  $\mu$ m. **g**, DNA FISH in the

neurons shown in **f**. Magnified insets show representative haploid and diploid nuclei with single and double hybridization signals, respectively. Scale bar, 10  $\mu$ m. **h**, TNNT2 staining in G1-sorted haploid pES12-derived cardiomyocytes. Scale bar, 10  $\mu$ m. **i, j**, Flow cytometry analysis with co-staining of DNA and PDX1 in h-pES10 cells following pancreatic differentiation. **i**, Gating for PDX1-positive cells (right) based on a negative secondary antibody stained control sample (left). **j**, DNA content profile of the entire cell population (related to Fig. 4k). **c**, chromosomal copies.





**Extended Data Figure 8 | *In vivo* differentiation of haploid human ES cells.** **a**, Haematoxylin and eosin histological sections of teratomas derived from h-pES10 and h-pES12. Scale bar, 50  $\mu\text{m}$ . **b**, TUJ1 (ectoderm),  $\alpha$ -SMA

(mesoderm), AFP (endoderm) and OCT4 (pluripotency) staining in an h-pES10-derived teratoma. DNA staining is shown in blue. Note the absence of nuclear OCT4 staining. Scale bars, 100  $\mu\text{m}$ .

Extended Data Table 1 | Identification of haploid cells in early-passage human parthenogenetic ES cell lines by metaphase spread analysis

| pES cell line | Oocyte donor | Passage number at analysis | Number of haploid metaphases | Number of diploid metaphases | Total number of metaphases | % Haploid metaphases |
|---------------|--------------|----------------------------|------------------------------|------------------------------|----------------------------|----------------------|
| 1             | 8/31/09      | 4                          | 0                            | 233                          | 233                        | 0                    |
| 2             | 1043         | 6                          | 0                            | 278                          | 278                        | 0                    |
| 3             | 1058         | 6                          | 0                            | 273                          | 273                        | 0                    |
| 4             | 1058         | 6                          | 0                            | 222                          | 222                        | 0                    |
| 5             | 1058         | 9                          | 0                            | 50                           | 50                         | 0                    |
| 6             | 1105         | 5<br>10                    | 0<br>0                       | 13<br>140                    | 13<br>140                  | 0<br>0               |
| 8             | 1151         | 4                          | 0                            | 361                          | 361                        | 0                    |
| 9             | 1157         | 6                          | 0                            | 234                          | 234                        | 0                    |
| 10            | 1160         | 7                          | 2                            | 150                          | 152                        | 1.32                 |
| 11            | 1160         | 8                          | 0                            | 194                          | 194                        | 0                    |
| <b>Total:</b> |              |                            | 2                            | 2148                         | 2150                       |                      |

The derivation of cell lines pES1–6 was reported previously<sup>11,23</sup>.

**Extended Data Table 2 | Isolation of haploid cells from early-passage human parthenogenetic ES cell lines by sub-2c-cell sorting**

| pES cell line           | Oocyte donor | Passage number at first sort | Passage number at second sort     | Presence of haploid cells at second sort |
|-------------------------|--------------|------------------------------|-----------------------------------|--|
| 6                       | 1105         | 6                            | 10                                | Undetected                               |
| 12                      | 1160         | 6                            | 10                                | Yes                                      |
| <b>swaPS cell line*</b> |              |                              |                                   |  |
| 4                       | 1126         | 6                            | 11                                | Undetected                               |
| 5                       | 1155         | 5                            | 9                                 | Undetected                               |
| 11                      | 1175         | 4                            | Sorted population did not survive |  |

\*swaPS cells are parthenogenetic ES cells derived following activation of an oocyte whose nuclear genome had been swapped with that of another oocyte<sup>11</sup>.