

HAPLOID HUMAN EMBRYONIC STEM CELL LINES AND SOMATIC CELL LINES AND METHODS OF MAKING THE SAME

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims benefit of priority under 35 U.S.C. §119(e) of U.S. Serial No. 62/198,614, filed July 29, 2015, U.S. Serial No. 62/279,490, filed January 15, 2016, and U.S. Serial No. 62/292,755, filed February 8, 2016, the entire contents of which are incorporated herein by reference in their entireties.

INCORPORATION OF SEQUENCE LISTING

[0002] The material in the accompanying sequence listing is hereby incorporated by reference into this application. The accompanying sequence listing text file, name NYSC1270_3WO_Sequence_Listing, was created on July 27, 2016, and is 2 kb. The file can be assessed using Microsoft Word on a computer that uses Windows OS.

BACKGROUND

[0003] Diploidy is a fundamental genetic feature in mammals, in which haploid cells normally arise only as post-meiotic germ cells that serve to insure a diploid genome upon fertilization. Gamete manipulation has yielded haploid embryonic stem (ES) cells from several mammalian species,¹⁻⁶ but prior to the present invention, not from humans.

[0004] Haploid genetics is a useful tool for delineating genome function, and haploid mammalian cells have proven invaluable through loss-of-function genetic screens, since single-allele mutations are sufficient to induce a phenotype.⁷ The derivation of haploid human ES cell lines has likely been hindered by the limited availability of human oocytes.¹⁰ Therefore, in humans, loss-of-function screens have thus far been facilitated through a near-haploid chronic myeloid leukemia cell line³⁰ and its derivative cells.³¹ Although useful, these are chromosomally aberrant cancer cells representing a single cell type. In that sense, the benefit in utilizing haploid human ES cells for genetic screens relies on the premise of their genomic stability, their capacity to model human early development and their potential to give rise to virtually any cell type of interest.

[0005] Although the increasing efficiency of clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9-mediated mutagenesis can facilitate loss-of-function screens in diploid cells, each allele may be affected differently, making functional gene disruption more efficient in a haploid genome.^{32,33} Moreover, while the use of CRISPR/Cas9 requires predesigned single-guide RNAs (sgRNAs), less biased mutagenesis approaches,

such as the gene trap method, can be readily applied to haploid but not diploid cells for loss-of-function screens.⁷

[0006] 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.³⁴ The surprising differentiation potential of haploid human genomes suggests that diploidy-dependent adaptations, rather than haploidy, may pose the predominant barriers for development in humans. Our discovery of haploid human ES cells thus provides novel means to delineate basic aspects of human genetics and development.

SUMMARY OF THE INVENTION

[0007] Artificial activation of unfertilized metaphase II (MII) human oocytes results in efficient development to the blastocyst stage and subsequent derivation of parthenogenetic ES (pES) cell lines.¹³⁻¹⁵ Although second polar body extrusion at MII results in a haploid egg, these ES cells have been repeatedly reported as diploid. We generated and analyzed 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 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, using a library of gene-trapped haploid human ES cells.

[0008] However, haploid human ES cells also displayed distinct properties from their diploid counterparts, including differential regulation of X chromosome inactivation and 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. Haploid human ES cells provide novel means for studying human functional genomics and development.

[0009] In one embodiment, the invention provides an enriched population of haploid human ES cells in culture, and a composition comprising an enriched population of haploid human ES cells, preferably in a culture medium.

[0010] The invention also provides a method for producing an enriched population of haploid human ES cells, the method comprising identifying haploid metaphases in a sample from a population of ES cells, wherein the ES cells are derived from an artificially activated

human oocyte; and sorting the population of ES cells based on cell ploidy to produce an enriched population of haploid human ES cells. In some embodiments, the method further comprises maintaining the enriched population of ES cells in culture for at least three passages. Preferably, the haploid cells in the sample are identified by metaphase spread analysis or sorting of cells with less than 2 chromosomal copies. In some embodiments, the sorting step based on cell ploidy comprises at least one cycle of flow cytometry, preferably, fluorescence-activated cell sorting (FACS). Haploid cells can also be identified by flow cytometry, centromere protein immunofluorescence staining, or DNA fluorescence in situ hybridization (FISH).

[0011] Preferably, the enriched population comprises at least 5% haploid human ES cells.

[0012] In another aspect, the invention provides a substantially pure population of haploid human ES cells in culture, and a composition comprising a substantially pure population of haploid human ES cells, preferably in a culture medium.

[0013] A further aspect of the invention provides a method for producing a substantially pure population of haploid human ES cells, the method comprising identifying haploid metaphases in a sample from a population of ES cells, wherein the ES cells are derived from an artificially activated human oocyte; and sorting the population of ES cells based on cell ploidy using 2-5 cycles of FACS; thereby producing a substantially pure population of haploid human ES cells. Preferably, the haploid cells in the sample are identified by metaphase spread analysis or sorting of cells with less than 2 chromosomal copies. Haploid cells can also be identified by flow cytometry, centromere protein immunofluorescence staining, or DNA FISH.

[0014] Preferably, the substantially pure population comprises at least 95% haploid human ES cells.

[0015] The invention additionally provides a method for producing a haploid human ES cell line, the method comprising producing an enriched population of haploid human ES cells by a method of the invention; maintaining the enriched population of haploid human ES cells in culture; and sorting the ES cells in culture every three to four passages, wherein the sorting is based on cell ploidy; thereby producing a haploid human ES cell line.

[0016] The invention provides haploid human ES cell lines produced by the methods of the invention.

[0017] Further embodiments of the invention include a population of haploid multipotent human cells in culture and a composition comprising a population of haploid multipotent human cells.

[0018] Another embodiment of the invention is a method for producing a population of haploid multipotent human cells, the method comprising culturing haploid human embryonic stem cells under conditions for directed differentiation, thereby producing a population of haploid multipotent human cells.

[0019] In another aspect, the invention provides a method for producing a population of haploid multipotent human cells, the method comprising culturing haploid human ES cells under conditions to induce embryoid body formation; and dissociating the embryoid body into cells; thereby producing a population of haploid multipotent human cells. In some embodiments, the method further comprises sorting the dissociated cells based on cell surface markers. In some embodiments, the sorting comprises FACS.

[0020] Preferably, the population of haploid multipotent human cells is a substantially pure population. In one embodiment, the population of haploid multipotent human cells comprises endodermal progenitor cells. In one embodiment, the population of haploid multipotent human cells comprises mesodermal progenitor cells. In one embodiment, the population of haploid multipotent human cells comprises ectodermal progenitor cells. In one embodiment, the population of haploid multipotent human cells comprises neural progenitor cells.

[0021] The invention provides a population of haploid differentiated human somatic cells in culture, and a composition comprising a population of haploid differentiated human somatic cells. The invention also provides a method for producing a population of haploid differentiated human somatic cells, the method comprising culturing haploid human ES cells under conditions for directed differentiation, thereby producing a population of haploid differentiated human somatic cells. The invention further provides a method for producing a population of haploid differentiated human somatic cells, the method comprising injecting haploid human ES cells into a non-human mammal under conditions to induce teratoma formation; and dissociating the teratoma into cells; thereby producing a population of haploid multipotent human cells.

[0022] Preferably, the haploid differentiated human somatic cells are selected from the group consisting of neurons, cardiomyocytes, pancreatic cells, skin cells, muscle cells, kidney cells, liver cells, lung cells, and intestinal cells.

[0023] A further aspect of the invention is a method of genetic screening, the method comprising exposing an enriched population of human haploid ES cells to a mutagen to induce at least one mutation in the cells; selecting human haploid ES cells in the enriched population that contain the mutation; and identifying a genotypic and/or phenotypic effect of the mutation in the human haploid ES cells. In one embodiment, the genetic screening is a forward genetic screen. In one embodiment, the mutagen is selected from the group consisting of a physical mutagen, a chemical mutagen, and a biological agent. Preferably, the mutagen is a biological agent, more preferably, a vector.

[0024] In a further aspect the present invention provides a population of genetically-modified human haploid ES cells. Similarly, in another aspect the present invention provides a population of genetically-modified human haploid multipotent cells. In yet another aspect the present invention provides a population of genetically-modified or human haploid differentiated somatic cells. In each of such embodiments the genetically-modified human haploid cells (i.e. ES cells, multipotent cells, or differentiated somatic cells) contain at least one artificially-introduced mutation. In another aspect the present invention provides a library of such genetically-modified human haploid cells (i.e. a library of genetically-modified ES cells, multipotent cells, or differentiated somatic cells), such as, for example, a gene trap library, wherein the library comprises multiple different artificially-introduced mutations. In each of the embodiments involving genetically-modified human haploid cells, the mutations may be introduced by treatment of human haploid cells with a mutagen selected from the group consisting of a physical mutagen, a chemical mutagen, and a biological agent. Similarly, in each of these embodiments, the genetically-modified haploid cells may optionally also comprise one or more marker or reporter genes, for example associated with the artificially-introduced mutation.

BRIEF DESCRIPTION OF THE DRAWINGS/FIGURES

[0025] This patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

[0026] Fig. 1a-1g show derivation of haploid human ES cell lines. Fig. 1a shows a schematic outline of parthenogenetic oocyte activation and potential haploidy in resulting ES cell lines. Second polar body (PB) extrusion at MII without fertilization results in a haploid 1-cell stage embryo and haploid cells are gradually eliminated due to diploidization. Fig. 1b shows a diploidization rate model for a haploid egg with a theoretical diploidization probability of 10%, overlaid with an exponential decay fit (red curve). Approximated cell

cycle numbers for different ES cell line derivation stages is indicated. Fig. 1c shows establishment of the haploid-enriched human ES cell line h-pES10 after repeated sorting and enrichment of 1c-cells. c: chromosomal copies. From top to bottom: DNA content profiles of unsorted diploid cells, partially purified haploid cells at the fourth sort, and mostly purified haploid cells at the sixth sort. Fig. 1d shows diploid and haploid karyotypes of pES10 before and after 4 rounds of haploid cell enrichment and expansion. Fig. 1e shows diploidization dynamics of h-pES10 over seven passages by flow cytometry, overlaid with an exponential fit to the data (red curve). Error bars show standard deviation (s.d.). Fig. 1f shows DNA FISH and Fig. 1g shows centromere staining in haploid-enriched and unsorted diploid pES10 cells. Magnified insets show representative haploid and diploid nuclei with either single or double hybridization signals (Fig. 1f) and 23 or 46 centromeres (Fig. 1g), respectively. White arrows point to diploid nuclei. Scale bars = 10 μ m.

[0027] Fig. 2a-2i show that haploid human ES cells display classical characteristics of pluripotent stem cells and enable loss-of-function genetic screening. Fig. 2a shows colony morphology of haploid-enriched and matching diploid cell lines. Scale bar is 50 μ m. Fig. 2b shows alkaline phosphatase staining of h-pES10 and h-pES12. Scale bars = 50 μ m. Fig. 2c shows co-staining of pluripotency markers (red), centromeres (green) and DNA (blue) in h-pES10 at colony resolution (upper panel; scale bars = 50 μ m) and single-cell resolution (lower panel; scale bars = 10 μ m). Magnified insets show representative haploid cells with 23 centromeres. Fig. 2d shows flow cytometry analysis of h-pES10 by co-staining DNA and cell surface markers TRA-1-60 and CLDN6, after gating for haploid cells in G1. Fig. 2e shows mean expression levels $\pm$ s.d. of pluripotency genes in haploid (1n) and diploid (2n) pES10 and pES12 cells in G1 ($n = 4$ for each group, with two biological replicates for each cell line, logarithmic scale). RPKM: reads per kilobase per million fragments mapped. Fig. 2f shows DNA methylation levels at pluripotency genes in duplicates of haploid (1n) and diploid (2n) pES10 and pES12 cells in G1, as well as control fibroblasts (Fib). Fig. 2g shows a schematic overview of genome-wide gene trapping in haploid human ES cells and screening for 6-TG-resistance genes. Fig. 2h shows *NUDT5* insertions (red arrows) detected in 3 6-TG-resistant colonies. Upper panel shows gene structure. Lower panel shows genomic sequence of the intronic insertion site (indicated by TTAA) and upstream exonic sequence (in box). Fig. 2i shows a schematic of the metabolic pathway leading to 6-TG toxicity through *NUDT5*-mediated PRPP production. ADP: adenosine diphosphate; AMP: adenosine monophosphate.

[0028] Fig. 3a-3j show molecular and cellular comparisons of haploid and diploid ES cells. Fig. 3a shows an experimental scheme of haploid and diploid ES cell isolation for comparative analyses. Fig. 3b shows RNA-Seq-based hierarchical clustering analysis of isogenic haploid (1n) and diploid (2n) cells in G1 (two biological replicates per cell line), compared with a 2n pES12-derived embryoid body (EB) sample. Fig. 3c shows a pie chart representation of relatively downregulated and upregulated genes in haploid vs. diploid ES cells on autosomes and X chromosome. Fig. 3d shows hierarchical clustering analysis by X chromosomal genes. Fig. 3e-3h show differential X chromosome inactivation (XCI) status in haploid and diploid ES cells. Fig. 3e shows genome-wide gene expression moving median plot (relative to the average of diploids in G1 by RNA-Seq, window size = 100 genes). Fig. 3f shows XIST expression levels. (1) and (2) denote biological replicates. Fig. 3g shows H3K27me3 staining. Scale bar = 10 μ m. Fig. 3h shows DNA methylation levels on the X chromosome. Fig. 3i shows relative total RNA, cell volume and ratio of mitochondrial DNA (mtDNA) to genomic DNA (gDNA) between G1-sorted haploid and diploid ES cells. Numbers of replicates are indicated in parenthesis. Error bars represent s.d. Fig. 3j shows mean expression levels $\pm$ standard error of the mean (s.e.m.) of nuclear (top panel) and mitochondrial oxidative phosphorylation genes (lower panel), upregulated in haploid ES cells relative to diploid ES cells ($n = 4$ for each group, as in Fig. 2e), and schematic representation of their organization in this pathway. IMS: intermembrane space. * $P < 0.05$; ** $P < 0.01$ (two-tailed unpaired Student's t test).

[0029] Fig. 4a-4p show the differentiation potential of haploid human cells. Fig. 4a shows representative images of 21-day EBs from haploid-enriched and diploid pES12 cells. Scale bar = 100 μ m. Fig. 4b shows haploid karyotype of cells dissociated from haploid-enriched EBs in Fig. 4a (plated cells shown in Fig. 10a). Fig. 4c shows the DNA content profile of h-pES10 EB cells. Fig. 4d shows expression of tissue- and pluripotency-specific genes in G1-sorted haploid (1n) and diploid (2n) ES and EB pES10 cells. Fig. 4e shows the DNA content profile of NCAM1-positive h-pES10-derived neural progenitor cells (NPCs). Fig. 4f shows expression of neural- and pluripotency-specific genes (right and left panels, respectively) in G1-sorted haploid pES10 ES cells and NPCs. Color-coded scale shows expression relative to the mean across the NPC sample and an ES cell duplicate. Fig. 4g shows differential XCI status in haploid and diploid pES10-derived EBs and NPCs, as shown by genome-wide gene expression moving median plot (window size = 200 genes). Centromere and differentiation marker co-staining is shown in h-pES12-derived TUJ1-positive neurons (Fig. 4h), TNNT2-positive cardiomyocytes (Fig. 4i), FOXA2-positive

definitive endoderm cells (Fig. 4k) and PDX1-positive and NKX6.1-positive PPCs (Fig. 4l). Magnified insets show representative haploid and diploid nuclei. Scale bars = 10 μ m. DNA content profiles are shown for h-pES12 cells differentiated into cardiomyocytes (Fig. 4j) and PDX1-positive PPCs (Fig. 4m). Fig. 4n shows TUJ1 (ectoderm), α -SMA (mesoderm), AFP (endoderm) and OCT4 (pluripotency) staining in an h-pES12-derived teratoma. Scale bars = 50 μ m. Fig. 4o shows DNA content profile of an h-pES10-derived teratoma. Fig. 4p shows serial h-pES12-derived teratoma sections analyzed by histology with hematoxylin and eosin staining (left panel; scale bar = 20 μ m) and DNA FISH (right panel; scale bar = 20 μ m). Magnified inset shows representative haploid nuclei (scale bar = 5 μ m).

[0030] Fig. 5a-5e show derivation of haploid human ES cell line h-pES12. Fig. 5a shows establishment of a haploid-enriched human ES cell line from pES12 cells after repeated sorting and enrichment of 1c-cells using Hoechst 33342 staining. Shown from top to bottom are the 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: chromosomal copies. Fig. 5b shows karyotypes and haploid metaphase percentage over the course of enrichment and passaging. Fig. 5c shows DNA FISH and Fig. 5d shows centromere protein immunofluorescence staining in h-pES12. Magnified insets show representative haploid nuclei with single hybridization signals (Fig. 5c) and 23 centromeres (Fig. 5d), respectively. Scale bars = 10 μ m. Fig. 5e shows single nucleotide polymorphism (SNP) array-based copy number variation (CNV) analysis comparing haploid (1n) pES10 and pES12 cells to their unsorted diploid (2n) counterparts (logarithmic scale).

[0031] Fig. 6a-6f show determination of ploidy at single-cell level by quantification of centromere foci. Fig. 6a shows that the counted number of centromeres correlates with ploidy. 1n: haploid-enriched pES10 cells grown for 4 passages after the forth sort (n = 33; 76% haploids by this assay); 2n: unsorted diploid pES10 cells (n = 34); 3n: soPS2 cells³⁵ (n = 27); 4n: Hybrid1 cells³⁶ (n = 27). Black horizontal lines indicate mean $\pm$ s.e.m. and dashed lines mark expected chromosome numbers. Fig. 6b shows quantification of haploid (1n) and diploid (2n) cells by DNA FISH in the haploid-enriched (n = 152; 73% haploids by this assay) and diploid (n = 135) cells in Fig. 6a. Fig. 6c shows the DNA content profile of the haploid-enriched cells in Fig. 6a (73% haploids by this assay). c: chromosomal copies. Co-staining of centromeres and either phospho-histone 3 (pH3, Ser10) (Fig. 6d) or 5-ethynyl-2'-deoxyuridine (EdU) (Fig. 6e) distinguishes between different stages of interphase in haploid pES12 cells. DNA staining is shown in blue. Scale bar = 5 μ m. Fig. 6f shows

quantification of centromere counts in the different cell cycle stages shown in Fig. 6d and Fig. 6e. indicated in parenthesis. Black horizontal lines indicate mean $\pm$ s.e.m.

[0032] Fig. 7 shows pluripotent stem cell markers in haploid pES12 cells. Co-staining of pluripotency markers NANOG, OCT4, SOX2, SSEA4 and TRA-1-6 (red), centromeres (green) and DNA (blue) in h-pES12 at colony resolution (upper panel; scale bars = 50 μ m) and single-cell resolution (lower panel; scale bars = 10 μ m) is shown. Magnified insets show representative haploid cells with 23 centromeres.

[0033] Fig. 8a-8e show analysis of parental imprinting and gene trap mutagenesis in haploid human parthenogenetic ES cells. Fig. 8a and Fig. 8b show hierarchical clustering analysis of diploid (2n) in vitro fertilization (IVF) ES cells and G1-sorted haploid (1n) and diploid parthenogenetic ES (pES) cells by expression levels of imprinted genes ($n = 75$, see Table 6) (Fig. 8a) and DNA methylation levels at imprinted differentially methylated regions (iDMRs, $n = 35$)³⁷ (Fig. 8b). (1) and (2) denote biological replicates. Fig. 8c shows relative mean expression levels $\pm$ s.e.m. of representative paternally-expressed imprinted genes across seven chromosomes in the samples shown in Fig. 8a (RPKM ratios). Fig. 8d shows 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 Fig. 8b. β values range from complete hypomethylation (0) to complete hypermethylation (1). Fig. 8e shows a schematic outline of the piggyBac gene trap system. The gene trap vector⁵² 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 Δ tk) and a 3' poly(A) signal (pA). In the presence of the PiggyBac transposase (encoded on a separate plasmid⁵³), 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. ITR: inverted terminal repeat; FRT: flox sites.

[0034] Fig. 9a-9i show comparative analyses of isogenic haploid and diploid human ES cells. Fig. 9a shows sorting purity of haploid (1n) and diploid (2n) ES cells in G1. Fig. 9b shows 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 (>2 -fold 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. Fig. 9c shows 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). Fig. 9d shows a genome-wide moving median plot of the gene expression ratio between haploid and diploid pE10 cells in G1 by expression microarray analysis (window size = 100 genes). Fig. 9e and Fig. 9f show a model for genome-wide autosomal gene level reduction in haploid human ES cell as inferred by differential XCI status. Fig. 9e shows that DNA content, RNA expression levels relative to total RNA, and presumed equality of absolute X chromosomal gene dosage in haploid (Xa) and diploid (XaXi) human ES cells enable the estimation of total RNA levels per haploid cell. Xa and Xi denote active (blue) and inactive (red) X chromosomes, respectively. A: autosomes; X: X chromosome; R: total RNA. Fig. 9f shows schematic genome-wide representation of relative and absolute RNA levels in the cells shown in Fig. 9e. Fig. 9g shows average diameter and calculated surface area and volume of G1-sorted haploid and diploid ES cells. Error bars represent s.d. (n = 4–8). *P < 0.01 (two-tailed unpaired Student's t test). Fig. 9h and Fig. 9i show functional annotation enrichment analysis for relatively downregulated genes and differentially methylated regions (DMRs) (Fig. 9h) as well as relatively upregulated genes (Fig. 9i) in haploid ES cells compared with diploid ES cells.

[0035] Fig. 10a-10c show EB differentiation of haploid human ES cells. Fig. 10a shows a representative image of plated cells dissociated from h-pES12-derived 21-day EBs, the karyotype of which is presented in Fig. 4b. Scale bar = 100 μ m. Fig. 10b shows DNA content profiles of dissociated EBs derived from haploid-enriched and diploid pES12 cells. c: chromosomal copies. Fig. 10c shows expression levels (RPKM) of tissue- and pluripotency-specific genes in undifferentiated (ES) and differentiated (EB) G1-sorted haploid (1n) pES10 cells.

[0036] Fig. 11a-11i show directed differentiation of haploid human ES cells. Fig. 11a and Fig. 11b show flow cytometry analysis with co-staining of DNA and NCAM1 in h-pES10 cells following neural differentiation. Fig. 11a shows gating for NCAM1-positive cells (right panel) based on a negative secondary-antibody-stained control sample (left panel). Fig. 11b shows DNA content profile of the entire cell population (related to Fig. 4e). c: chromosomal copies. Fig. 11c shows expression levels (RPKM) of neural-specific genes in G1-sorted haploid (1n) pES10 ES cells and NPCs. Fig. 11d shows XIST expression levels in haploid and diploid (2n) pES10-derived EBs and NPCs. Fig. 11e shows TUJ1 staining in h-pES12-derived neurons. Scale bar = 100 μ m. Fig. 11f shows DNA FISH on the neurons

shown in Fig. 11e. Magnified insets show representative haploid and diploid nuclei with single and double hybridization signals, respectively. Scale bar = 10 μ m. Fig. 11g shows TNNT2 staining in G1-sorted haploid pES12-derived cardiomyocytes. Scale bar = 10 μ m. Fig. 11h and Fig. 11i show flow cytometry analysis with co-staining of DNA and PDX1 in h-pES10 cells following pancreatic differentiation. Fig. 11h shows gating for PDX1-positive cells (right panel) based on a negative secondary-antibody-stained control sample (left panel). Fig. 11i shows DNA content profile of the entire cell population (related to Fig. 4m). c: chromosomal copies.

[0037] Fig. 12a and Fig. 12b show in vivo differentiation of haploid human ES cells. Fig. 12a shows hematoxylin and eosin histological sections of teratomas derived from h-pES10 and h-pES12. Scale bar = 50 μ m. Fig. 12b shows TUJ1 (ectoderm), α -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 μ m.

DETAILED DESCRIPTION OF THE INVENTION

[0038] Many aspects of the present invention are described in the above Summary of the Invention section of this patent application, as well as in the Drawings/Figures, the Brief Description of the Drawings/Figures, and the Claims sections of this patent application. This Detailed Description section provides certain additional description relating to the present invention and is intended to be read in conjunction with all other sections of this patent application.

[0039] The invention provides haploid human parthenogenetic ES cell lines following activation of haploid human oocytes. Earlier studies on mouse oocyte activation have demonstrated that haploidy likely persists at least partially in the inner cell mass (ICM) of the resulting embryo.^{8,9} Nonetheless, diploids cells progressively dominate over increasing cell cycles, due to spontaneous and irreversible diploidization events (Fig. 1a).^{9,16,17} Based on a diploidization kinetics model, we estimated that even if diploidization should occur in 1 out of 10 cell cycles, 1% of ES cells may remain haploid at early passages (Fig. 1b). To model diploidization dynamics of ES cells in culture or of early embryonic cells, we considered the following definitions and assumptions:

[0040] (1) If H_n is the total number of haploid cells at the n^{th} cell division cycle, and p is the probability of undergoing diploidization at a single cell division cycle, the total number of diploidization events (d) at the n^{th} cell division cycle is:

$$d_n = p \cdot H_{n-1}$$

[0041] (2) The total number of diploid cells (D) at the at the n th cell division cycle is:

$$D_n = 2D_{n-1} + d_n$$

[0042] (3) The total number of cells (T) at the at the n th cell division cycle is:

$$T_n = 2T_{n-1} - d_n$$

[0043] (4) And the haploid fraction (h) at the n th cell division cycle is thus:

$$h_n = \frac{H_n}{T_n} = \frac{T_n - D_n}{T_n} = 1 - \frac{D_n}{T_n} = 1 - \frac{2D_{n-1} + d_n}{2T_{n-1} - d_n} = 1 - \frac{2D_{n-1} + p \cdot H_{n-1}}{2T_{n-1} - p \cdot H_{n-1}}$$

[0044] For simplification, we further assumed that:

(1) Diploidization occurs due to failed cell division cycle at a constant probability, resulting in one diploid daughter cell rather than two haploid daughter cells.

(2) Normal cell division cycles produce two daughter cells, and are synchronous on average.

(3) Any selective advantage in favor of either haploid or diploid cells is negligible, and cases of cell death and aneuploidy are not considered.

[0045] Simulation of this model throughout 100 cell division cycles generated the data points plotted in Fig. 1b, fitting an exponential decay function.

[0046] The frequency of haploid cells, which decreases over time due to a gradual and irreversible acquisition of a diploid genome, required the analysis of over 2,000 metaphases and a total of 14 parthenogenetic ES cell lines at passages 4–10, to allow the establishment of two individual haploid ES cell lines with a 14% success rate. Overall, we have utilized four independent methodologies, namely metaphase spread analysis, flow cytometry, FISH and centromere quantification, to determine ploidy in undifferentiated and differentiated cells.

[0047] Haploid and diploid human ES cells shared many similarities, including classical pluripotent stem cells attributes as well as multilineage differentiation potential in vitro and in vivo, and could not be distinguished based on their relative global gene expression profiles. While other studies on non-human haploid ES cells mostly emphasized their resemblance, we aimed to identify putative transcriptional, epigenetic and physical properties that set these two ploidy states apart. XCI, which is readily observed in diploid human ES cells,¹⁹ does not occur in haploid ES cells. Notably, the absence of XCI also extends into recently diploidized ES cells, which may later become XaXi in culture. These findings allowed us to infer a reduction in absolute gene expression levels in haploid human

ES cells, a conclusion that could not have been drawn in undifferentiated mouse ES cells, where XCI does not occur.²⁶ This suggests that global transcriptional compensation is not a prerequisite 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.²⁷ Still, the discrepancy between the haploid:diploid ratios of physical parameters such as diameter (~0.8), surface area (~0.7) and volume (~0.6) implies that regulatory robustness involves specific compensatory mechanisms. Interestingly, we observed subtle yet consistent relative upregulation of genes involved in oxidative phosphorylation involving a coordinated crosstalk between nuclear and mitochondrial genomes, which is likely the reflection of a relative increase in mitochondria abundance in haploid cells. This robustness could reflect a more permissive regulatory phase in the early embryo, particularly in light of X chromosome dosage and elevated oxidative phosphorylation activity, which are both consistent with an early preimplantation epiblast identity.²⁸

[0048] We show that a haploid human karyotype is not a barrier for ES cell differentiation. In particular, we show that haploid human ES cells give rise to neural progenitor cells while remaining haploid, as has also been observed in the mouse². However, while mouse studies showed that haploid cells are lost upon further differentiation,^{2,16} we observed specification of human haploid cells into somatic cell fates of all three embryonic germ layers, despite a dosage imbalance between the X chromosome and autosomes that persisted from the pluripotent state into the differentiated state.

[0049] Haploid human ES cells of the invention are identified by metaphase spread analysis or sub-2c (where “c” stands for chromosomal copy number) cell sorting of a population of ES cells derived from artificially activated human oocytes. Haploid human ES cells can also be identified by flow cytometry, preferably FACS, centromere protein immunofluorescence staining, or DNA FISH. Methods for carrying out metaphase spread analysis and sub-2c sorting, centromere protein immunofluorescence, and DNA FISH are known in the art and described herein.

[0050] Methods of artificial activation are known in the art and include, but are not limited to, parthenogenetic activation and androgenetic activation. Parthenogenetic techniques involve the activation of the oocyte using an electrical pulse, a calcium ionophore, a kinase inhibitor, a translation inhibitor or a combination of these.⁴¹ Androgenetic techniques involve the fertilization of an enucleated oocyte with a sperm, typically by intracytoplasmic sperm injection. The genome of the oocyte is removed before or after

fertilization of the oocyte to generate a cell that contains only the sperm genome. The oocyte may be exposed to an activation stimulus as for parthenogenesis.

[0051] Flow cytometry, including FACS, can be used in the methods of the invention to identify and/or sort cells based on ploidy, cell surface markers, or other phenotypic characteristic(s). Other cell sorting techniques known in the art, for example, magnetic-activated cells sorting (MACS), can also be used in the methods of the invention.

[0052] The haploid human ES cells and cells lines of the invention can be kept “in culture,” which refers, by a non-limiting example, to standard human ES cell growth conditions. Namely, culture on a feeder layer of arrested mouse embryonic fibroblasts in gelatin-coated plates in medium containing Knockout Dulbecco’s Modified Eagle’s Medium (Gibco, Life Technologies) supplemented with 15% Knockout Serum Replacement (KSR; Gibco, Life Technologies), 2 mM L-glutamine, 0.1 mM nonessential amino acids, penicillin and streptomycin (50 units mL⁻¹ and 50 µg mL⁻¹, respectively), 0.1 mM β-mercaptoethanol and 8 ng mL⁻¹ basic fibroblast growth factor. Cells can be maintained in a humidified incubator at 37 °C and 5% CO₂ and passaged every 3–5 days using Trypsin Solution A without EDTA. Preferably, haploid human ES cells are maintained in culture for at least three passages, at least four passages, at least five passages, at least seven passages, at least ten passages, at least twenty passages, or at least thirty passages. Preferably, haploid human ES cells are maintained in culture for at least about ten days, at least about twenty days, at least about thirty days, at least about forty-five days, at least about sixty days, at least about three months, at least about four months, or at least about six months.

[0053] “Multipotent” haploid human cells of the invention are progenitor or stem cells that have the potential to develop into multiple, but not all, cell types. Neural stem cells, hematopoietic stem cells, and mesenchymal stem cells are non-limiting examples of multipotent cells. The inventors have demonstrated that multipotent haploid human cells can be produced from embryoid bodies differentiated from haploid human ES cells, or by directed differentiation of haploid human ES cells toward a particular lineage.

[0054] The term “cell lines” refers to cells that can grow in culture for many passages, and can be enriched for haploid cells by cell sorting. In accordance with one example of the invention, the cell lines are cultured under standard human ES cell growth conditions and occasional enrichment of the haploid fraction by sorting every 3 to 5 passages.

[0055] As used herein, the term “enriched population” refers to a percentage of haploid cells in a total cell population that is greater than 1%, preferably greater than 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50% , 55%, 60% , 65%, 70% , 75% , 80%,

85% , or 90%. Typically, an enriched population can be obtained after a single cycle of sorting, such as FACS.

[0056] The term “substantially pure” refers to a percentage of haploid cells in a total cell population that is above 90%, preferably above 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%. Most preferably, a substantially pure population is a confluent population of haploid cells.

[0057] The haploid human cells of the invention can be used, for instance, in developmental, genetic, and cellular biology research to study basic mechanisms such as haploidy vs. diploidy, X chromosome inactivation, parental imprinting and hybridism. Because the haploid human cells can be engineered with desired homozygosity and immunogenicity properties, they are also useful for therapeutic purposes, such as in regenerative medicine. The genomes of haploid parthenogenetic ES cells or of haploid differentiated human cells could potentially be used in human reproduction, to replace the genome of an oocyte.

[0058] Importantly, the haploid human cells of the invention are useful in genetic screening, preferably forward genetics. One example is in homozygote loss-of-function screens to identify targets of various diseases, and in drug screens to identify candidate compounds for treating these diseases. A genetic screen can include the use of a mutagen to introduce one or more mutations into haploid human cells. Mutagens suitable for use in the present invention include physical mutagens, such as ionizing radiation (X-rays, gamma rays, ultraviolet rays, etc.); chemical mutagens, such as alkylating agents; and biological agents, such as plasmid, phage, or viral vectors. Examples of biological agents include insertional vectors, for example, gene trap vectors, and technologies for site-directed mutagenesis, such as zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), or the CRISPR/Cas9 system.

[0059] As used in this specification and the appended claims, the singular forms “a,” “an,” and “the” include plural referents, unless the context clearly dictates otherwise. The terms “a” (or “an”) as well as the terms “one or more” and “at least one” can be used interchangeably.

[0060] Furthermore, “and/or” is to be taken as specific disclosure of each of the two specified features or components with or without the other. Thus, the term “and/or” as used in a phrase such as “A and/or B” is intended to include A and B, A or B, A (alone), and B (alone). Likewise, the term “and/or” as used in a phrase such as “A, B, and/or C” is intended

to include A, B, and C; A, B, or C; A or B; A or C; B or C; A and B; A and C; B and C; A (alone); B (alone); and C (alone).

[0061] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention is related. For example, The Dictionary of Cell and Molecular Biology (5th ed. J.M. Lackie ed., 2013), the Oxford Dictionary of Biochemistry and Molecular Biology (2d ed. R. Cammack et al. eds., 2008), and The Concise Dictionary of Biomedicine and Molecular Biology (2d ed. P-S. Juo, 2002) can provide one of skill with general definitions of some terms used herein.

[0062] Units, prefixes and symbols are denoted in their Système International de Unites (SI) accepted form. Numeric ranges are inclusive of the numbers defining the range. The headings provided herein are not limitations of the various aspects or embodiments of the invention, which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification in its entirety.

[0063] Wherever embodiments are described with the language “comprising,” otherwise analogous embodiments described in terms of “consisting of” and/or “consisting essentially of” are included.

[0064] All of the references cited in this disclosure are hereby incorporated by reference in their entireties. In addition, any manufacturers’ instructions or catalogues for any products cited or mentioned herein are incorporated by reference. Documents incorporated by reference into this text, or any teachings therein, can be used in the practice of the present invention. Documents incorporated by reference into this text are not admitted to be prior art.

[0065] The invention is further described in the following non-limiting Examples.

EXAMPLES

Example 1. Methods

Human oocyte manipulation and parthenogenetic ES cell line derivation.

[0066] Human oocyte donation and pES and swaPS cell line derivation procedures were described previously.^{15,35} Briefly, mature MII oocytes were activated using a calcium ionophore and/or an electrical pulse, followed by 4 hour 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.¹⁵ ES cell lines were derived by laser-ablation of the trophectoderm³⁸ and addition of ROCK inhibitor Y-27632 at 10 μ M to the derivation medium.³⁵ Two to three days after plating, remaining trophectoderm cells were laser-ablated, and ICM cells were allowed to grow for 10–14 days until manual picking of the outgrowth was feasible.

[0067] Cell culture.

[0068] Unless otherwise stated, human ES cells were cultured on a feeder layer of growth-arrested mouse embryonic fibroblasts 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⁻¹ penicillin, 50 μ g mL⁻¹ streptomycin, 0.1 mM β -mercaptoethanol and 8 ng mL⁻¹ basic fibroblast growth factor (bFGF). Cells were free of mycoplasma and maintained in a humidified incubator at 37 °C and 5% CO₂. 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 μ M ROCK inhibitor Y-27632 (Stemgent) for 1 day or up to 2 days 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. Rapid expansion of the outgrowth allows isolation of haploid ES cells as early as passage 3.

[0069] Isolation and maintenance of haploid human ES cell lines.

[0070] Following identification of haploid cells in human parthenogenetic ES cell lines at passages 6–8 by either metaphase spread analysis or sub-2c cell sorting (Table 1 and Table 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.

[0071] TABLE 1: Identification of haploid cells in early-passage human parthenogenetic ES cell lines by metaphase spread analysis

pES cell line	Oocyte donor	Passage no. at analysis	No. of haploid metaphases	No. of diploid metaphases	Total no. 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	0	13	13	0
		10	0	140	140	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
Total:			2	2148	2150	

[0072] The derivation of cell lines pES1–6 was reported previously.^{15,35}

[0073] 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 1st sort	Passage number at 2nd sort	Presence of haploid cells at 2nd sort
6	1105	6	10	Undetected
12	1160	6	10	Yes
swaPS cell line*				
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.³⁵

[0074] Metaphase spread analysis.

[0075] For induction of mitotic arrest, growing cells were incubated for 40 min in the presence of 100 ng mL⁻¹ colcemid (Biological Industries), added directly to the culture medium, in a humidified incubator at 37 °C with 5% CO₂. The cells were then trypsinized, centrifuged at 1000 RPM at room temperature and gently resuspended in 37 °C-warmed hypotonic solution (2.8 mg mL⁻¹ KCl and 2.5 mg mL⁻¹ sodium citrate) followed by 20-min 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 analyzed metaphases (minimum of 20 metaphases per analysis).

[0076] Live ES cell sorting by DNA content.

[0077] Cells were washed with phosphate buffered saline (PBS), dissociated using either TrypLE Select or TrypLE Express (Thermo Fisher Scientific) and stained with 10 $\mu\text{g mL}^{-1}$ Hoechst 33342 (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 μM ROCK inhibitor Y-27632, filtered through a 70- μm 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 μM ROCK inhibitor Y-27632 for 24 hours. 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 G2/M-phase peak in haploid-enriched cultures and compared with G2/M-phase diploid cells from unsorted diploid cultures. Note that haploid-enriched cultures also consist of a mixed population of G2/M-phase haploids and G1-phase diploids. Sorting purity was confirmed by rerunning a fraction of sorted samples through the instrument.

[0078] Flow cytometry.

[0079] 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 analyzing 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 $\mu\text{g mL}^{-1}$ 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 **Table 3**. In all flow cytometry procedures, samples were

filtered through a 70- μ m cell strainer (Corning Life Sciences) and analyzed in either BD FACSaria III or BD Influx (BD Biosciences).

TABLE 3: Primary antibodies used in the Examples

Antibody	Host species	Type	Isotype	Dilution	Vendor	Catalog no.	Assay
Anti-centromere	Human	Polyclonal	NA	1:50	Antibodies Incorporated	15-235-0001	IF
Anti-pH3	Rabbit	Polyclonal	NA	1:1,000	EMD Millipore	06-570	IF
Anti-human NANOG	Goat	Polyclonal	IgG	1:100	R&D Systems	AF1997	IF
Anti-human OCT4	Goat	Polyclonal	IgG	1:100	Santa Cruz Biotechnology	sc-8628	IF
Anti-human SOX2	Rabbit	Polyclonal	IgG	1:100	Stemgent	09-0024	IF
Anti-human/mouse SSEA4	Mouse	Monoclonal	IgG	1:500	R&D Systems	MAB1435	IF
Alexa Fluor® 488-conjugated anti-human TRA-1-60	Mouse	NA	IgM	1:100	BD Biosciences	560173	IF
PE-conjugated anti-human TRA-1-60	Mouse	Monoclonal	IgM	1:40	BD Biosciences	560884	FC
Anti-human CLDN6	Mouse	Monoclonal	IgG	1:40	R&D Systems	MAB3656	FC
Anti-H3K27me3	Rabbit	Polyclonal	IgG	1:400	EMD Millipore	07-449	IF
Anti-human NCAM-1/CD56	Goat	Polyclonal	IgG	1:150	R&D Systems	AF2408	FC
Anti-human FOXA2	Rabbit	Polyclonal	NA	1:200	Cell Signaling Technology	3143	IF
Anti-human TUJ1	Rabbit	NA	NA	1:300	Sigma-Aldrich	T2200	IF
Anti-human TNNT2	Rabbit	Polyclonal	IgG	1:100	Abcam	ab45932	IF
Anti-human NKX6.1	Mouse	Monoclonal	IgG	1:300	DSHB	F55A10	IF
Anti-human PDX1	Goat	Polyclonal	IgG	1:200	R&D Systems	AF2419	IF, FC
Anti-human NKX6.1	Mouse	Monoclonal	IgG	1:300	DSHB	F55A10	IF
Anti-human α -SMA	Mouse	Monoclonal	IgG	1:500	Dako	M0851	IF
Anti-human AFP	Rabbit	Polyclonal	NA	1:500	Dako	A0008	IF

NA: not available. IF: immunofluorescence staining. FC: flow cytometry.

DNA fluorescence *in situ* hybridization.

[0080] DNA FISH was performed as described elsewhere³⁹ 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 MATEK glass plates for several passages prior to analysis.

[0081] Alkaline phosphatase and immunofluorescence staining.

[0082] 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 (Table 3) 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 a 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. 1g and Fig. 5d were grown on Matrigel-coated MATEK glass plates for several passages prior to analysis.

[0083] 6-TG Resistance Screen

[0084] To generate a gene trap mutant library, 9 replicates of $4\text{--}5 \times 10^6$ haploid pES10 cells (within one passage after 1c-cell enrichment) were co-transfected with 20 μg 5'-PTK-3' gene trap vector⁵² and 20 μg pCyL43 piggyBac transposase plasmid⁵³ using Bio-Rad Gene Pulser (suspended in 800 μL Opti-MEM, 4-mm cuvettes, 320 V, 250 μF), 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 $\mu\text{g mL}^{-1}$ puromycin starting 48 hours 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 μM 6-TG (Sigma-Aldrich) on DR4 MEFs for 18 days, during which 6 resistant colonies were independently isolated and characterized. Genomic DNA was extracted (NucleoSpin Tissue Kit, MACHEREY-NAGEL) and insertion sites were detected using splinkerette PCR as described previously,⁵⁴ 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.

[0085] Total DNA was isolated using the NucleoSpin Tissue Kit (MACHEREY-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.

[0086] Genome integrity analysis.

[0087] Copy number variation (CNV) analysis was carried out on DNA samples of G1-sorted haploid and diploid pES10 and pES12 cells (Table 4) 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.

TABLE 4: Samples analyzed by SNP arrays and DNA methylation arrays

Sample no.	Sample name	Cell line	Passage	Description	Assay*
1	pES10 h-G1 rep1	h-pES10	15	Haploid ES cells in G1 (1c), 3 rd sort, biological replicate 1	M, S
2	pES10 h-G1 rep2	h-pES10	19	Haploid ES cells in G1 (1c), 4 th sort, biological replicate 2	M
3	pES10 d-G1 rep1	d-pES10	17	Diploid ES cells in G1 (2c), technical replicate 1	M, S
4	pES10 d-G1 rep2	d-pES10	17	Diploid ES cells in G1 (2c), technical replicate 2	M
5	h-pES10 d-G2/M	h-pES10	15	Diploid ES cells in G2/M from a mixed haploid-diploid culture (4c), 3 rd sort	M
6	pES12 h-G1 rep1	h-pES12	18	Haploid ES cells in G1 (1c), 4 th sort, biological replicate 1	M, S
7	pES12 h-G1 rep2	h-pES12	22	Haploid ES cells in G1 (1c), 5 th sort, biological replicate 2	M
8	pES12 d-G1 rep1	d-pES12	18	Diploid ES cells in G1 (2c), biological replicate 1	M, S
9	pES12 d-G1 rep2	d-pES12	18	Diploid ES cells in G1 (2c), biological replicate 2	M
10	NYSCF2	NYSCF2	20	Diploid IVF ES cell line	M

Sample no.	Sample name	Cell line	Passage	Description	Assay*
11	HuES53	HuES53	< 20	Diploid IVF ES cell line	M
12	HuES64	HuES64	< 20	Diploid IVF ES cell line	M

* M: DNA methylation analysis. S: SNP array analysis.

RNA sequencing.

[0088] Total RNA samples (200 ng–1 µg, RNA integrity number (RIN) > 9) were enriched for mRNAs by pull-down of poly(A)⁺ 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.

[0089] **Table 5** provides a detailed list of samples analyzed by RNA-Seq.

TABLE 5: Samples analyzed by RNA-Seq

Sample no.	Sample name	Cell line	Passage	Description
1	pES10 h-G1 rep1	h-pES10	15	Haploid ES cells in G1 (1c), 3 rd sort, biological replicate 1
2	pES10 h-G1 rep2	h-pES10	19	Haploid ES cells in G1 (1c), 4 th sort, biological replicate 2
3	pES10 d-G1 rep1	d-pES10	13	Diploid ES cells in G1 (2c), biological replicate 1
4	pES10 d-G1 rep2	d-pES10	17	Diploid ES cells in G1 (2c), biological replicate 2
5	h-pES10 d-G2/M	h-pES10	15	Diploid ES cells in G2/M from a mixed haploid-diploid culture (4c), 3 rd sort
6	d-pES10 d-G2/M	d-pES10	13	Diploid ES cells in G2/M from a diploid culture (4c)
7	pES12 h-G1 rep1	h-pES12	18	Haploid ES cells in G1 (1c), 4 th sort, biological replicate 1
8	pES12 h-G1 rep2	h-pES12	22	Haploid ES cells in G1 (1c), 5 th sort, biological replicate 2
9	pES12 d-G1 rep1	d-pES12	18	Diploid ES cells in G1 (2c), biological replicate 1
10	pES12 d-G1 rep2	d-pES12	18	Diploid ES cells in G1 (2c), biological replicate 2
11	pES10 NPC h-G1	h-pES10	21	Haploid NCAM1 ⁺ NPCs in G1 (1c) derived from ES cells after 4 th sort + 10 days culture + 10 days differentiation
12	pES10 EB h-G1	h-pES10	25	Haploid EB cells in G1 (1c) derived from ES cells after the 5 th sort + 13 days culture + 21 days differentiation
13	pES10 EB d-G1	d-pES10	12	Diploid EB cells in G1 (2c) derived from ES cells after 21 days differentiation
14	pES10 EB d- unsorted	d-pES10	12	Unsorted diploid EBs derived from ES cells after 21 days differentiation
15	pES12 EB d- unsorted	d-pES12	19	Unsorted diploid EBs derived from ES cells after 21 days differentiation

Sample no.	Sample name	Cell line	Passage	Description
16	HuES53	HuES53	< 20	Diploid IVF ES cell line
17	HuES64	HuES64	< 20	Diploid IVF ES cell line

Transcriptome analysis.

[0090] 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 >2-fold 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 (*see* the Geneimprint website), listed in **Table 6**. RNA-Seq data from control ES cell line NYSCF1 were published elsewhere³⁷ (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.⁴⁰ Color-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 previously⁴¹ by using Affymetrix Human Gene 1.0 ST arrays.

TABLE 6: Imprinted genes used for hierarchical clustering analysis

Gene	ID	Locus	Gene	ID	Locus
<i>TP73</i>	ENSG00000078900	chr1:3547330-3663900	<i>SLC22A18</i>	ENSG000000110628	chr11:2909009-2946476
<i>RNU5D-1</i>	ENSG000000200169	chr1:45196726-45196842	<i>PHLDA2</i>	ENSG000000181649	chr11:2949502-2950685
<i>DIRAS3</i>	ENSG000000162595	chr1:68167148-68698803	<i>OSBPL5</i>	ENSG000000021762	chr11:3108345-3187969
<i>LRRTM1</i>	ENSG000000162951	chr2:79384131-80875905	<i>WT1</i>	ENSG000000184937	chr11:32409320-32480315
<i>GPR1</i>	ENSG000000183671	chr2:207040039-207082771	<i>ANO1</i>	ENSG000000131620	chr11:69924407-70035634
<i>ZDBF2</i>	ENSG000000204186	chr2:207139386-207179148	<i>ZC3H12C</i>	ENSG000000149289	chr11:109964086-110042566
<i>NAP1L5</i>	ENSG000000177432	chr4:89442135-89629693	<i>NTM</i>	ENSG000000182667	chr11:131240372-132206716
<i>FAM50B</i>	ENSG000000145945	chr6:3832166-3855971	<i>RBP5</i>	ENSG000000139194	chr12:7276279-7281538
<i>LIN28B</i>	ENSG000000187772	chr6:105404922-105531207	<i>RB1</i>	ENSG000000139687	chr13:48877886-49056122
<i>AIM1</i>	ENSG000000112297	chr6:106959729-107018326	<i>DLK1</i>	ENSG000000185559	chr14:101192041-101201539
<i>PLAGL1</i>	ENSG000000118495	chr6:144261436-144385735	<i>MEG3</i>	ENSG000000214548	chr14:101245746-101327368
<i>SLC22A2</i>	ENSG000000112499	chr6:160592092-160698670	<i>RTL1</i>	ENSG000000254656	chr14:101346991-101351184
<i>SLC22A3</i>	ENSG000000146477	chr6:160769299-160932156	<i>MKRN3</i>	ENSG000000179455	chr15:23810453-23873064
<i>DDC</i>	ENSG000000132437	chr7:50526133-50633154	<i>MAGEL2</i>	ENSG000000254585	chr15:23888690-23891175
<i>GRB10</i>	ENSG000000106070	chr7:50657759-50861159	<i>NDN</i>	ENSG000000182636	chr15:23930564-23932450
<i>MAGI2</i>	ENSG000000187391	chr7:77646392-79100524	<i>NPAP1</i>	ENSG000000185823	chr15:24920540-24928593
<i>TFPI2</i>	ENSG000000105825	chr7:93220884-93540577	<i>SNRPN</i>	ENSG000000128739	chr15:25068793-25492435
<i>SGCE</i>	ENSG000000127990	chr7:94214541-94285521	<i>SNURF</i>	ENSG000000273173	chr15:25068793-25492435
<i>PEG10</i>	ENSG000000242265	chr7:94285636-94299007	<i>UBE3A</i>	ENSG000000114062	chr15:25497371-25684128
<i>PPP1R9A</i>	ENSG000000158528	chr7:94536513-94925727	<i>ATP10A</i>	ENSG000000206190	chr15:25922419-26110317
<i>DLX5</i>	ENSG000000105880	chr7:96649703-96654409	<i>NAA60</i>	ENSG000000262621	chr16:3415098-3627401
<i>CPA4</i>	ENSG000000128510	chr7:129932973-129964020	<i>ZNF597</i>	ENSG000000167981	chr16:3415098-3627401
<i>MEST</i>	ENSG000000106484	chr7:130125882-130148500	<i>TCEB3C</i>	ENSG000000183791	chr18:44388352-44627658
<i>KLF14</i>	ENSG000000174595	chr7:130417400-130418888	<i>DNMT1</i>	ENSG000000130816	chr19:10244020-10341962
<i>DLGAP2</i>	ENSG000000198010	chr8:1449531-1656642	<i>MIR371A</i>	ENSG000000199031	chr19:54290850-54291423
<i>ZFAT</i>	ENSG000000066827	chr8:135490030-135725292	<i>NLRP2</i>	ENSG000000022556	chr19:55434876-55512510

Gene	ID	Locus	Gene	ID	Locus
<i>ZFAT-AS1</i>	ENSG00000248492	chr8:135490030-135725292	<i>PEG3</i>	ENSG00000198300	chr19:57202053-57352097
<i>KCNK9</i>	ENSG00000169427	chr8:140613080-140715299	<i>ZIM2</i>	ENSG00000269699	chr19:57202053-57352097
<i>GLIS3</i>	ENSG00000107249	chr9:3824126-4348392	<i>MIMT1</i>	ENSG00000268654	chr19:57352269-57359924
<i>INPP5F</i>	ENSG00000198825	chr10:121485608-121588652	<i>BLCAP</i>	ENSG00000166619	chr20:36120873-36156333
<i>H19</i>	ENSG00000130600	chr11:2016405-2022700	<i>NNAT</i>	ENSG00000053438	chr20:36120873-36156333
<i>IGF2</i>	ENSG00000167244	chr11:2150341-2182571	<i>MIR296</i>	ENSG00000268649	chr20:57392186-57392780
<i>IGF2-AS</i>	ENSG00000099869	chr11:2150341-2182571	<i>MIR298</i>	ENSG00000216031	chr20:57393280-57393368
<i>INS</i>	ENSG00000254647	chr11:2150341-2182571	<i>GNAS</i>	ENSG00000087460	chr20:57393973-57486247
<i>KCNQ1</i>	ENSG00000053918	chr11:2465913-2882798	<i>GNAS-AS1</i>	ENSG00000235590	chr20:57393973-57486247
<i>KCNQ1OT1</i>	ENSG00000269821	chr11:2465913-2882798	<i>DGCR6</i>	ENSG00000183628	chr22:18893540-18924066
<i>KCNQ1DN</i>	ENSG00000237941	chr11:2891262-2893335	<i>DGCR6L</i>	ENSG00000128185	chr22:20301798-20307603
<i>CDKN1C</i>	ENSG00000129757	chr11:2904442-2907111			

DNA methylation analysis.

[0091] DNA methylation analysis was performed on genomic DNA from the samples detailed in **Table 4** using Infinium HumanMethylation450 BeachChips (Illumina) following the Infinium HD Methylation Protocol as described previously.³⁷ DNA methylation data from control ES cell line NYSCF1 were published before (GEO accession number GSE61657).³⁷ 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 analyzed as described before.³⁷ For analysis of DNA methylation levels on the X chromosome, probes with average β 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 analyzed for functional annotation enrichment using DAVID (using the Benjamini method to determine statistical significance).

Cell size analysis.

[0092] 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.

[0093] Relative mtDNA abundance was analyzed by quantitative PCR (qPCR) by using primers for the mitochondrial gene ND2 (forward primer: 5'–TGTTGGTTATACCCTTCCCGTACTA–3' (SEQ ID NO: 1); reverse primer: 5'–CCTGCAAAGATGGTAGAGTAGATGA–3' (SEQ ID NO: 2)) and normalization to nuclear DNA by using primers for the nuclear gene BECN1 (forward primer: 5'–CCCTCATCACAGGGCTCTCTCCA–3' (SEQ ID NO: 3); reverse primer: 5'–GGGACTGTAGGCTGGGAAGTATGC–3' (SEQ ID NO: 4)), as described elsewhere.⁴² 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 Table 4 (n = 4 for each group, with two biological replicates for each cell line).

[0094] All high-throughput data have been deposited at the Gene Expression Omnibus (GEO) under accession number GSE71458.

Embryoid body differentiation.

[0095] 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^{–1} 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.

[0096] Differentiation into neural progenitor cells.

[0097] NCAM1-positive ES cell-derived NPCs were obtained using a 10-days protocol for efficient neural differentiation⁴³ with slight modification.⁴⁴ 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 and a Cy3-conjugated secondary antibody (Jackson ImmunoResearch Laboratories) diluted 1:200.

[0098] Neuronal differentiation.

[0099] Differentiation into neurons was carried out by following a published protocol⁴⁵ based on synergistic inhibition of SMAD signaling⁴⁶ 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 μ M SB431542 (Selleckchem) and 2.5 μ M LDN-193189 (Stemgent) for 4 days. Subsequently, cells were kept in N2 medium⁴⁵ supplemented with 10 μ M SB431542 and 2.5 μ M LDN-193189 for additional 4 days, followed by 2 days in N2 medium supplemented with B-27 (Thermo Fisher Scientific) and 10 μ M DAPT (Stemgent). The cells were then dissociated and replated on 0.01% poly-l-ornithine– (Sigma-Aldrich) and laminin–coated (4 μ g/ml, Thermo Fisher Scientific) plates in the presence of 10 μ M 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–1 BDNF (R&D) until analysis by immunostaining and FISH on day 20.

[00100] Cardiomyocyte differentiation.

[00101] 80–90% confluent ES cells grown on Matrigel-coated plates (Corning) in mTeSR1 (STEMCELL Technologies) were subject to an 11-days regimen⁴⁷ 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.

[00102] Differentiation toward the pancreatic lineage.

[00103] The protocol utilized here was developed based on several recent publications.^{48–50} 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 progenitor cells (PPCs). Differentiation was initiated within 2 passages after 1c-cell enrichment.

[00104] Teratoma formation assay.

[00105] All experimental procedures in animals were approved by the ethics committee of the Hebrew University. ES cells were trypsinized and approximately 2×10^6 cells were resuspended in 100 μ L human ES cell medium and 100 μ L Matrigel (BD Biosciences), followed by subcutaneous injection into NOD-SCID Il2rg–/– immunodeficient

mice (Jackson Laboratory). Eight to twelve weeks after injection, tumors were dissected and subjected to further analysis. Histological slides were prepared from tumor slices cryopreserved in O.C.T. compound (Sakura Finetek) using Leica CM1850 cryostat (Leica Biosystems, 10- μ m sections), followed by immunostaining, hematoxylin and eosin staining or FISH analysis. Flow cytometry with Hoechst 33342 staining was performed on dissociated cells from freshly dissected tumors.

[00106] Example 2. Determination of ploidy at single-cell level by quantification of centromere foci.

[00107] We devised a methodology for determining ploidy at single-cell resolution based on centromere protein immunofluorescence staining. As each chromosome normally has one centromere, we reasoned that being able to detect and enumerate centromeres would provide a means to visualize ploidy in individual cells, while also allowing to define cellular identity by co-staining for specific markers.

[00108] We first tested this method on cell lines of known ploidies, including haploid-enriched and diploid pES10 cells, triploid soPS2 cells³⁵ and tetraploid Hybrid1 cells,³⁶ demonstrating a correlation between ploidy and the counted number of centromeres (Fig. 6a). Centromere counts were within with the expected range of chromosome number. 76% of the haploid-enriched cells showed 15–25 centromere foci, whereas the remaining cells showed 30–48 foci, similar to the range documented in diploid cells (34–51). Importantly, this percentage of haploid cells was consistent with that estimated by DNA FISH (73%, Fig. 6b) and DNA content flow cytometry (73%, Fig. 6c), indicating that centromere foci quantification is a reliable method for identifying haploid ES cells.

[00109] The accuracy of counting centromeres decreased with increasing ploidy, due to centromere clustering, which would lead to an underestimation of the actual number of individual centromeres, as well as difficulties in counting large numbers of centromeres in single cells. Observing higher numbers of foci than expected could be explained by visual artifacts or aneuploidy in rare cells. To address whether cell cycle progression altered centromere foci numbers or affected their quantification, we co-stained haploid ES cells for centromere protein and either phospho-histone 3 (pH3, Ser10) or 5-ethynyl-2'-deoxyuridine (EdU) (marking cells entering mitosis and undergoing DNA replication, respectively), to quantify the number of centromeres at different stages of the cell cycle (Fig. 6d–6f). Evidently, centromere foci numbers did not increase during DNA replication, confirming that haploid cells can be accurately detected throughout interphase by centromere staining.

[00110] Example 3. Derivation of Haploid Human ES Cells

[00111] We generated and analyzed a collection of 14 early-passage (passage ≤ 9) human pES cell lines for the persistence of haploid cells. All cell lines originated from activated oocytes displaying second polar body extrusion and a single pronucleus. We initially utilized chromosome counting by metaphase spreading and G-banding 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 exclusively in a single cell line, pES10 (1.3%, Table 1). We also used viable FACS with Hoechst 33342 staining, aiming to isolate cells with a DNA content corresponding to less than two chromosomal copies (2c) from four additional lines, leading to the successful enrichment of haploid cells from a second cell line, pES12 (Table 2).

[00112] Two individual haploid-enriched ES cell lines were established from both pES10 and pES12 (hereafter referred to as h-pES10 and h-pES12) within five to six rounds of 1c-cell FACS enrichment and expansion (Fig. 1c (pES10), Fig. 5a (pES12)). These cell lines were grown in standard culture conditions for over 30 passages while including cells with a normal haploid karyotype (Fig. 1d, Fig. 5b). However, since diploidization occurred at a rate of 3–9% of the cells per day (Fig. 1e), cell sorting at every three to four passages was required for maintenance and analysis of haploid cells. Further, visualization of ploidy in adherent conditions was enabled by DNA fluorescence in situ hybridization (FISH) (Fig. 1f, Fig. 5c) and quantification of centromere protein foci (Fig. 1g, Fig. 5d; Fig. 6). In addition to their intact karyotype, haploid ES cells did not harbor significant copy number variations (CNVs) relative to their unsorted diploid counterparts (Fig. 5e). Importantly, we did not observe common duplications of specific regions in the two cell lines that would result in pseudo-diploidy. Therefore, genome integrity was preserved throughout haploid-cell isolation and maintenance. As expected, single nucleotide polymorphism (SNP) array analysis demonstrated complete homozygosity of diploid pES10 and pES12 cells across all chromosomes.

[00113] Both h-pES10 and h-pES12 exhibited classical human pluripotent stem cell features, including typical colony morphology and alkaline phosphatase activity (Fig. 2a, Fig. 2b). Single haploid ES cells expressed various hallmark pluripotency markers (NANOG, OCT4, SOX2, SSEA4 and TRA1-60), as confirmed in essentially pure haploid cultures by centromere foci quantification (>95% haploids) (Fig. 2c, Fig. 7). Notably, selective flow cytometry enabled to validate the expression of two human ES-cell-specific cell surface markers (TRA-1-60 and CLDN618) in single haploid cells (Fig. 2d). Moreover, sorted haploid and diploid ES cells showed highly similar transcriptional and epigenetic signatures

of pluripotency genes (Fig. 2e, Fig. 2f). Since the haploid ES cells were derived as parthenotes, they featured distinct transcriptional and epigenetic profiles of maternal imprinting, owing to the absence of paternally-inherited alleles and paternal epigenetic profiles (Fig. 8).

[00114] Haploid cells are valuable for loss-of-function genetic screening because phenotypically-selectable mutants can be identified upon disruption of a single allele. To demonstrate the applicability of this principle in haploid human ES cells, we generated a genome-wide mutant library using a piggyBac transposon gene trap system that targets transcriptionally active loci (Fig. 2g, Fig. 8e), and screened for resistance to the purine analog 6-thioguanine (6-TG). Out of six isolated and analyzed 6-TG-resistant colonies, three harbored a gene trap insertion localizing to the nucleoside diphosphate linked moiety X-type motif 5 (NUDT5) autosomal gene (Fig. 2h). NUDT5 disruption was recently confirmed to confer 6-TG resistance in human cells,⁵¹ by acting upstream to the production of 5-phospho-d-ribose-1-pyrophosphate (PRPP), which serves as a phosphoribosyl donor in the hypoxanthine phosphoribosyltransferase 1 (HPRT1)-mediated conversion of 6-TG to thioguanosine monophosphate (TGMP) (Fig. 2i). Detection of a loss-of-function phenotype due to an autosomal mutation validates that genetic screening is feasible in haploid human ES cells.

[00115] Example 4. Molecular and Cellular Comparisons of Haploid and Diploid ES Cells

[00116] The ability of human ES cells to exist both as haploids and diploids led us to investigate whether these two ploidy states may differ in certain aspects of gene regulation and cell biology. To analyze haploid and diploid ES cells in the same phase of the cell cycle, 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, Fig. 9a). 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 with one another and separately from differentiated embryoid bodies (EBs), indicating resemblance that extends beyond the effects of genetic background (Fig. 3b). Nonetheless, a total of 565 differentially expressed genes were identified (>2-fold change, false discovery rate (FDR) < 0.05), corresponding to 275

relatively upregulated genes and 290 relatively downregulated genes in haploids compared with diploids (Fig. 9b).

[00117] Notably, X chromosomal genes were significantly overrepresented among the relatively upregulated gene set (40%, $P < 0.001$, χ^2 goodness of fit test) (Fig. 3c), and the expression levels of X chromosomal genes alone clearly distinguished between haploid and diploid ES cells; the latter clustering even more closely with their differentiated derivatives than their undifferentiated haploid counterparts (Fig. 3d). These data are in line with an expected differential status of X chromosome inactivation (XCI) in haploid and diploid human ES cells: while the single X chromosome in haploids is transcriptionally active (Xa), one of the two X chromosomes in diploids often undergoes XCI (XaXi)¹⁹ as in female somatic cells. Indeed, haploid human ES cells exhibited a relative increase in X chromosomal gene expression compared with diploids by both RNA-Seq and expression microarray analysis, and lacked expression of the XCI-driving transcript XIST (Fig. 3e, Fig. 3f, Fig. 9b–9d), as observed in diploid XaXa human ES cells²⁰. XCI is an epigenetic phenomenon, regulated by repressive histone modifications and DNA methylation. H3K27me3 foci were consistently observed in unsorted diploid ES cells, but not in their haploid-enriched counterparts (Fig. 3g). Moreover, methylome analysis showed that the X chromosome DNA methylation signature of haploid ES cells resembles that of diploid male ES cells (XaY), whose single-copy X chromosome is largely hypomethylated, as opposed to the composite pattern of a hypomethylated Xa and a hypermethylated Xi in diploid female cells (Fig. 3h). Interestingly, recently diploidized ES cells remained XaXa soon after diploidization (within 3 passages after haploid cell enrichment) by all the above-mentioned assays (Fig. 3a, Fig. 3e–3h).

[00118] Normalization to total gene expression, which is inherent to conventional relative gene expression analyses,²¹ resulted in seemingly similar expression levels of autosomal genes but higher levels of X-linked genes in haploid compared with diploid ES cells (Fig. 3e, Fig. 9c). However, assuming that the absolute expression of X-linked genes in haploid Xa and diploid XaXi cells are equivalent, these data suggest a genome-wide autosomal gene level reduction in haploid cells (Fig. 9e, Fig. 9f). In support of this notion, we found that total RNA amounts isolated from haploid ES cells were significantly lower than those obtained from the same numbers of diploid cells (Fig. 3i). An overall decrease in total gene expression implied that the physical dimensions of these cells may also be altered. Indeed, the average diameter ratio between sorted haploid and diploid ES cells in G1 was

around 0.8 (9.6 and 11.5 μm , respectively), corresponding to haploid:diploid ratios of around 0.7 in surface area and around 0.6 in volume (Fig. 3i, Fig. 9g).

[00119] We subsequently focused on consistent differential regulation within autosomes. Based on transcriptional and DNA methylation analyses, we found significant enrichment of genes encoding proteins with signal peptides to be relatively downregulated in haploid ES cells (Fig. 9h). 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. 3j, Fig. 9i). Furthermore, all 13 mitochondrial genes involved in oxidative phosphorylation were consistently upregulated in haploid cells as well (Fig. 3j), indicating coordinated regulation between these nuclear and mitochondrial genes. This coincides with a 32% increase in the mitochondrial DNA (mtDNA) to nuclear DNA ratio between haploids and diploids (Fig. 3i), suggesting that mitochondrial abundance relative to the nuclear DNA content is relatively higher in haploid cells.

[00120] Example 5. Differentiation of Human Haploid ES Cells

[00121] We next sought to assess the differentiation potential of haploid human ES cells of parthenogenetic origin. Although mammalian parthenogenetic development is restricted due to the non-equivalence of parental genomes,^{22,23} diploid human parthenogenetic pluripotent stem cells are functionally pluripotent as evident by their ability to give rise to all embryonic lineages.^{13,24,25} To address whether human parthenogenetic ES cells are capable of multilineage differentiation as haploids, we performed several differentiation assays, followed by ploidy and differentiation status characterizations of the resulting cells. 21-day-old EBs generated by spontaneous differentiation of haploid-enriched and diploid ES cells could not be distinguished by appearance (Fig. 4a), and the morphology of dissociated haploid-cell-derived EB cells was consistent with differentiation (Fig. 10a). Notably, metaphase spread analysis revealed a haploid karyotype (Fig. 4b; 4/4 metaphases), and a largely haploid DNA profile (~70% haploids) was confirmed by flow cytometry in both h-pES10- and h-pES12-derived EB cells (Fig. 4c, Fig. 10b). We then compared the gene expression profiles of G1-sorted haploid ES and EB cells, focusing on 18 genes that showed clear specificity across eight tissues and pluripotent stem cells. For example, our gene set included CHRM1 (cholinergic receptor), KRT17 (keratine), MYL1 (myosin), REN (renin), ALB (albumin), CPA1 (carboxypeptidase), SFTPD (surfactant) and MALRD1 (MAM and LDL receptor), which are expressed in the brain, skin, muscle, kidney, liver, pancreas, lung and intestine, respectively (Fig. 4d). Whereas the expression of these lineage-specific genes

was negligible in undifferentiated ES cells, all were expressed in haploid and diploid EB cells (Fig. 4d, Fig. 6c). In addition, haploid and diploid EB cells showed insignificant expression of pluripotency-specific genes, consistent with efficient differentiation and acquisition of somatic cell fates of all three embryonic germ layers.

[00122] To extend this analysis to more specific and potentially more mature cell types, we subjected haploid ES cells to directed differentiation assays. Haploid ES cells undergoing directed differentiation towards a neural fate for ten days remained haploid while efficiently giving rise to neural cell adhesion molecule 1 (NCAM1)-positive neural progenitor cells (NPCs, ~90% efficiency) (Fig. 4e, Fig. 11a, Fig. 11b). Sorted haploid NPCs expressed multiple neural-lineage-specific genes but not pluripotency-specific genes (Fig. 4f, Fig. 11c), indicating a robust exit from the pluripotent state while taking on a neural fate. XCI is imperative in diploid differentiated female cells, resulting in dosage compensation and a ratio of 1:2 between the X chromosome and autosomes. Since haploid ES cells are incapable of inactivating their single-copy X chromosome, an X:autosomes dosage imbalance of 1:1 should persist into the differentiated state. Indeed, both haploid NPCs and haploid EB cells showed an Xa signature contrary to the XaXi signature of diploid EB cells, as indicated by whole-genome expression analysis and XIST levels (Fig. 4g, Fig. 11d).

[00123] Neuronal differentiation was not restricted to the progenitor stage as the cells also differentiated with high efficiency (>90%) into mature TUJ1 (also known as β -tubulin III)-positive neurons by 20 days with notable persistence of haploid cells, as shown by both co-staining with centromeres (Fig. 4h; 47% haploids, n = 104) and FISH analysis (Fig. 11e, Fig. 11f; 46% haploids, n = 200). Similarly, haploid cells differentiated into cardiac troponin T type 2 (TNNT2)-expressing cardiomyocytes (Fig. 4i; 32% haploids, n = 97) during an 11-day protocol resulting in spontaneously beating clusters and 39% (n = 31) of haploid cells sorted from the whole culture (25% 1c cells) were confirmed as TNNT2-positive (Fig. 4j, Fig. 11g). Next, we differentiated haploid-enriched cultures (~70% haploids) to the pancreatic lineage, analyzing two stages of differentiation by centromere foci analysis, namely, specification to definitive endoderm and further into pancreatic cells. We observed robust differentiation (>90%) of both haploids and diploids into forkhead box A2 (FOXA2)-positive definitive endoderm cells (Fig. 4k; 56% haploids, n = 112), and into pancreatic and duodenal homeobox 1 (PDX1)-positive pancreatic cells (Fig. 4l; 13% haploid, n = 103), some of which were also positive for NK6 homeobox 1 (NKX6.1). In addition to centromere analysis, the persistence of haploid PDX1-positive cells was also confirmed by flow cytometry (Fig. 4m; 10% PDX1-positive 1c cells; Fig. 11h, Fig. 11i).

[00124] Finally, both haploid-enriched human ES cell lines gave rise to teratomas comprising cell types of ectodermal, mesodermal and endodermal origins as shown by histological and immunostaining analyses with TUJ1, α -smooth muscle actin (α -SMA) and α -fetoprotein (AFP) (Fig. 4n, Fig. 12a, Fig. 12b), meeting the most stringent criterion for human pluripotency in vivo. Importantly, no residual undifferentiated OCT4-positive cells could be detected (Fig. 4n, Fig. 12b). Upon dissection, DNA content analysis revealed that a considerable population of h-pES10-derived teratoma cells remained haploid (Fig. 4o). 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 able to contribute to an organized tissue structure while responding to developmental signals (Fig. 4p). It is worth noting that haploid cells were identified in all analyzed teratomas (n = 4), although with variable proportions, which may be influenced by the initial amount of haploid cells and/or the time length of differentiation.

REFERENCES

1. Leeb, M. & Wutz, A. Derivation of haploid embryonic stem cells from mouse embryos. *Nature* **479**, 131–4 (2011).
2. Elling, U. *et al.* Forward and reverse genetics through derivation of haploid mouse embryonic stem cells. *Cell Stem Cell* **9**, 563–74 (2011).
3. Yang, H. *et al.* Generation of genetically modified mice by oocyte injection of androgenetic haploid embryonic stem cells. *Cell* **149**, 605–17 (2012).
4. Li, W. *et al.* Androgenetic haploid embryonic stem cells produce live transgenic mice. *Nature* **490**, 407–11 (2012).
5. Li, W. *et al.* Genetic modification and screening in rat using haploid embryonic stem cells. *Cell Stem Cell* **14**, 404–14 (2014).
6. Yang, H. *et al.* Generation of haploid embryonic stem cells from *Macaca fascicularis* monkey parthenotes. *Cell Res.* **23**, 1187–200 (2013).
7. Wutz, A. Haploid mouse embryonic stem cells: rapid genetic screening and germline transmission. *Annu. Rev. Cell Dev. Biol.* **30**, 705–22 (2014).
8. Tarkowski, A. K., Witkowska, A. & Nowicka, J. Experimental parthenogenesis in the mouse. *Nature* **226**, 162–5 (1970).
9. 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**, 249–61 (1983).
10. Egli, D. *et al.* Impracticality of egg donor recruitment in the absence of compensation. *Cell Stem Cell* **9**, 293–4 (2011).
11. Leeb, M. & Wutz, A. Haploid genomes illustrate epigenetic constraints and gene dosage effects in mammals. *Epigenetics Chromatin* **6**, 41 (2013).
12. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196–9 (2007).
13. Revazova, E. S. *et al.* Patient-specific stem cell lines derived from human parthenogenetic blastocysts. *Cloning Stem Cells* **9**, 432–49 (2007).
14. Kim, K. *et al.* Recombination signatures distinguish embryonic stem cells derived by parthenogenesis and somatic cell nuclear transfer. *Cell Stem Cell* **1**, 346–52 (2007).
15. Paull, D. *et al.* Nuclear genome transfer in human oocytes eliminates mitochondrial DNA variants. *Nature* **493**, 632–7 (2013).
16. Leeb, M. *et al.* Germline potential of parthenogenetic haploid mouse embryonic stem cells. *Development* **139**, 3301–5 (2012).

17. Takahashi, S. *et al.* Induction of the G2/M transition stabilizes haploid embryonic stem cells. *Development* **141**, 3842–7 (2014).
18. 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).
19. 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. U. S. A.* **105**, 4820–5 (2008).
20. 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**, 262–70 (2013).
21. Lovén, J. *et al.* Revisiting global gene expression analysis. *Cell* **151**, 476–82 (2012).
22. McGrath, J. & Solter, D. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **37**, 179–83 (1984).
23. Barton, S. C., Surani, M. A. & Norris, M. L. Role of paternal and maternal genomes in mouse development. *Nature* **311**, 374–6 (1984).
24. Mai, Q. *et al.* Derivation of human embryonic stem cell lines from parthenogenetic blastocysts. *Cell Res.* **17**, 1008–19 (2007).
25. Stelzer, Y., Yanuka, O. & Benvenisty, N. Global analysis of parental imprinting in human parthenogenetic induced pluripotent stem cells. *Nat. Struct. Mol. Biol.* **18**, 735–41 (2011).
26. Minkovsky, A., Patel, S. & Plath, K. Concise review: Pluripotency and the transcriptional inactivation of the female Mammalian X chromosome. *Stem Cells* **30**, 48–54 (2012).
27. Biancotti, J. C. *et al.* The in vitro survival of human monosomies and trisomies as embryonic stem cells. *Stem Cell Res.* **9**, 218–24 (2012).
28. Zhou, W. *et al.* HIF1 α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J.* **31**, 2103–16 (2012).
29. Shuai, L. *et al.* Durable pluripotency and haploidy in epiblast stem cells derived from haploid embryonic stem cells in vitro. *J. Mol. Cell Biol.* 1–29 (2015). doi:10.1093/jmcb/mjv044
30. Carette, J. E. *et al.* Haploid genetic screens in human cells identify host factors used by pathogens. *Science* **326**, 1231–5 (2009).

31. Carette, J. E. *et al.* Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* **477**, 340–3 (2011).
32. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–4 (2014).
33. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–7 (2014).
34. Otto, S. P. & Jarne, P. Evolution. Haploids--hapless or happening? *Science* **292**, 2441–3 (2001).
35. Noggle, S. *et al.* Human oocytes reprogram somatic cells to a pluripotent state. *Nature* **478**, 70–5 (2011).
36. Cowan, C. A., Atienza, J., Melton, D. A. & Eggan, K. Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science* **309**, 1369–73 (2005).
37. 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).
38. 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**, 103–6 (2009).
39. Rao, P. H., Nandula, S. V & Murty, V. V. Molecular cytogenetic applications in analysis of the cancer genome. *Methods Mol. Biol.* **383**, 165–85 (2007).
40. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–5 (2013).
41. Yamada, M. *et al.* Human oocytes reprogram adult somatic nuclei of a type 1 diabetic to diploid pluripotent stem cells. *Nature* **510**, 533–6 (2014).
42. Wanet, A. *et al.* Mitochondrial remodeling in hepatic differentiation and dedifferentiation. *Int. J. Biochem. Cell Biol.* **54**, 174–85 (2014).
43. Kim, D.-S. *et al.* Robust enhancement of neural differentiation from human ES and iPS cells regardless of their innate difference in differentiation propensity. *Stem Cell Rev.* **6**, 270–81 (2010).
44. Stelzer, Y., Sagi, I. & Benvenisty, N. Involvement of parental imprinting in the antisense regulation of onco-miR-372-373. *Nat. Commun.* **4**, 2724 (2013).
45. Wang, L. *et al.* Differentiation of hypothalamic-like neurons from human pluripotent stem cells. *J. Clin. Invest.* **125**, 796–808 (2015).

46. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* **27**, 275–80 (2009).
47. Lian, X. *et al.* Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/ β -catenin signaling under fully defined conditions. *Nat. Protoc.* **8**, 162–75 (2013).
48. Hua, H. *et al.* iPSC-derived β cells model diabetes due to glucokinase deficiency. *J. Clin. Invest.* **123**, 3146–53 (2013).
49. Pagliuca, F. W. *et al.* Generation of functional human pancreatic β cells in vitro. *Cell* **159**, 428–39 (2014).
50. Rezania, A. *et al.* Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat. Biotechnol.* **32**, 1121–33 (2014).
51. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* (2016). doi:10.1038/nbt.3437.
52. Cadiñanos, J. & Bradley, A. Generation of an inducible and optimized piggyBac transposon system. *Nucleic Acids Res.* **35**, e87 (2007).
53. Wang, W. *et al.* Chromosomal transposition of PiggyBac in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 9290–5 (2008).
54. Chen, L. *et al.* Transposon activation mutagenesis as a screening tool for identifying resistance to cancer therapeutics. *BMC Cancer* **13**, 93 (2013).

[00125] The foregoing description of the specific embodiments will fully reveal the general nature of the invention such that others can, without undue experimentation, apply knowledge that is within the ordinary skill of those in the art to readily modify and/or adapt such specific embodiments for various applications without departing from the general concept of the present invention. Therefore, such adaptations and modifications are intended to be within the meaning and range of equivalents of the disclosed embodiments, based on the teaching and guidance presented herein. It is to be understood that the phraseology or terminology herein is for the purpose of description and not of limitation, such that the terminology or phraseology of the present specification is to be interpreted by the skilled artisan in light of the teachings and guidance. The present invention is further described by the following claims.

SEQUENCE LISTING

<110> THE NEW YORK STEM CELL FOUNDATION
YISSUM RESEARCH DEVELOPMENT COMPANY OF THE HEBREW
UNIVERSITY OF JERUSALEM
EGLI , Dietrich M.
BENVENISTY , Nissim
SAGI , Ido

<120> HAPLOID HUMAN EMBRYONIC STEM CELL LINES AND SOMATIC CELL LINES
AND METHODS OF MAKING THE SAME

<130> NYSC1270-3W0

<150> US 62/198,614

<151> 2015-07-29

<150> US 62/279,490

<151> 2016-01-15

<150> US 62/292,755

<151> 2016-02-08

<160> 4

<170> PatentIn version 3.5

<210> 1

<211> 25

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Mitochondrial gene ND2

<220>

<221> misc_feature

<223> Forward primer

<400> 1

tgttggttat acccttcccg tacta

25

<210> 2

<211> 25

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Mitochondrial gene ND2

<220>
<221> misc_feature
<223> Reverse Primer

<400> 2
cctgcaaaga tggtagagta gatga

25

<210> 3
<211> 23
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Nuclear gene BECN1

<220>
<221> misc_feature
<223> Forward Primer

<400> 3
ccctcatcac agggctctct cca

23

<210> 4
<211> 24
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Nuclear gene BECN1

<220>
<221> misc_feature
<223> Reverse Primer

<400> 4
gggactgtag gctgggaact atgc

24

CLAIMS

1. A method for producing an enriched population of haploid human ES cells comprising:

- a) identifying haploid metaphase cells in a sample from a population of ES cells, wherein the ES cells are derived from an artificially activated human oocyte; and
- b) sorting the population of ES cells based on cell ploidy to produce an enriched population of haploid human ES cells.

2. The method of claim 1, further comprising maintaining the enriched population of ES cells in culture for at least three passages.

3. The method of claim 1, wherein the haploid metaphase cells in the sample are identified by metaphase spread analysis or sub-2c cell sorting.

4. The method of claim 1, wherein the haploid metaphase cells in the sample are identified by flow cytometry, centromere protein immunofluorescence staining, or DNA fluorescence in situ hybridization.

5. The method of claim 1, wherein the sorting step comprises at least one cycle of fluorescence-activated cell sorting (FACS).

6. A method for producing a substantially pure population of haploid human ES cells comprising:

- a) identifying haploid metaphase cells in a sample from a population of ES cells, wherein the ES cells are derived from an artificially activated human oocyte; and
- b) sorting the population of ES cells based on cell ploidy using 2-5 cycles of FACS;

thereby producing a substantially pure population of haploid human ES cells.

7. The method of claim 6, wherein the haploid metaphase cells in the sample are identified by metaphase spread analysis or sub-2c cell sorting.

8. An enriched population of haploid human embryonic stem (ES) cells *in vitro*.

9. A composition comprising an enriched population of haploid human ES cells.

10. The enriched population of claim 8, the composition of claim 9, or the method of claim 1, wherein the enriched population comprises at least 5% haploid human ES cells.

11. A substantially pure population of haploid human ES cells in culture.

12. A composition comprising a substantially pure population of haploid human ES cells.

13. The substantially pure population of claim 11, the composition of claim 12, or the method of claim 6, wherein the substantially pure population comprises at least 95% haploid human ES cells.

- 14.** An isolated haploid human ES cell line.
- 15.** A method for producing a haploid human ES cell line comprising:
- a) producing an enriched population of haploid human ES cells by the method of claim 1;
 - b) maintaining the enriched population of haploid human ES cells in culture; and
 - c) sorting the ES cells in culture every three to four passages, wherein the sorting is based on cell ploidy;
- thereby producing a haploid human ES cell line.
- 16.** A population of haploid multipotent human cells *in vitro*.
- 17.** A composition comprising a population of haploid multipotent human cells.
- 18.** A method for producing a population of haploid multipotent human cells comprising culturing haploid human embryonic stem cells under conditions for directed differentiation, thereby producing a population of haploid multipotent human cells.
- 19.** A method for producing a population of haploid multipotent human cells, comprising:
- a) culturing haploid human ES cells under conditions to induce embryoid body formation; and
 - b) dissociating the embryoid body into cells;
- thereby producing a population of haploid multipotent human cells.
- 20.** A method for producing a population of haploid multipotent human cells comprising:
- a) injecting haploid human ES cells into a non-human mammal under conditions to induce teratoma formation; and
 - b) dissociating the teratoma into cells;
- thereby producing a population of haploid multipotent human cells.
- 21.** The method of claim 19 or claim 20, further comprising:
- c) sorting the dissociated cells based on cell surface markers.
- 22.** The method of claim 21, wherein the sorting comprises FACS.
- 23.** The population of claim 16, the composition of claim 17, or the method of any one of claims 18, 19, or 20, wherein the population of haploid multipotent human cells is a substantially pure population.
- 24.** The population of claim 16, the composition of claim 17, or the method of any one of claims 18, 19, or 20, wherein the population of haploid multipotent human cells comprises endodermal progenitor cells.

- 25.** The population of claim 16, the composition of claim 17, or the method of any one of claims 18, 19, or 20, wherein the population of haploid multipotent human cells comprises mesodermal progenitor cells.
- 26.** The population of claim 16, the composition of claim 17, or the method of any one of claims 18, 19, or 20, wherein the population of haploid multipotent human cells comprises ectodermal progenitor cells.
- 27.** The population of claim 16, the composition of claim 17, or the method of claim 18, wherein the population of haploid multipotent human cells comprises neural progenitor cells.
- 28.** A population of haploid differentiated human somatic cells in culture.
- 29.** A composition comprising a population of haploid differentiated human somatic cells.
- 30.** A method for producing a population of haploid differentiated human somatic cells comprising culturing haploid human ES cells or haploid multipotent human cells under conditions for directed differentiation, thereby producing a population of haploid differentiated human somatic cells.
- 31.** The population of claim 28, the composition of claim 29, or the method of claim 30, wherein the haploid differentiated human somatic cells are selected from the group consisting of neurons, cardiomyocytes, pancreatic cells, skin cells, muscle cells, kidney cells, liver cells, lung cells, and intestinal cells.
- 32.** A method of genetic screening comprising:
- a) exposing an enriched population of human haploid ES cells to a mutagen to induce at least one mutation in the cells;
 - b) selecting human haploid ES cells in the enriched population that contain the mutation; and
 - c) identifying a genotypic and/or phenotypic effect of the mutation in the human haploid ES cells.
- 33.** The method of claim 32, wherein the genetic screening is a forward genetic screen.
- 34.** The method of claim 32, wherein the mutagen is selected from the group consisting of a physical mutagen, a chemical mutagen, and a biological agent.