Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage

Collaborators: Bosman, Alexis; Feki, Anis; Jaconi, Marisa

How to cite

International Stem Cell Initiative. Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. In: Nature biotechnology, 2011, vol. 29, n° 12, p. 1132–1144. doi: 10.1038/nbt.2051

This publication URL: https://archive-ouverte.unige.ch//unige:25638
Publication DOI: 10.1038/nbt.2051

nature biotechnology

Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage

The International Stem Cell Initiative[1]

The International Stem Cell Initiative analyzed 125 human embryonic stem (ES) cell lines and 11 induced pluripotent stem (iPS) cell lines, from 38 laboratories worldwide, for genetic changes occurring during culture. Most lines were analyzed at an early and late passage. Single-nucleotide polymorphism (SNP) analysis revealed that they included representatives of most major ethnic groups. Most lines remained karyotypically normal, but there was a progressive tendency to acquire changes on prolonged culture, commonly affecting chromosomes 1, 12, 17 and 20. DNA methylation patterns changed haphazardly with no link to time in culture. Structural variants, determined from the SNP arrays, also appeared sporadically. No common variants related to culture were observed on chromosomes 1, 12 and 17, but a minimal amplicon in chromosome 20q11.21, including three genes expressed in human ES cells, *ID1*, *BCL2L1 and HM13,* occurred in >20% of the lines. Of these genes, *BCL2L1* is a strong candidate for driving culture adaptation of ES cells.

In human ES cell cultures, somatic mutations that generate a selective advantage, such as a greater propensity for self-renewal, can become fixed over time[1]. This selection may be the reason for the nonrandom genetic changes found in human ES cells maintained for long periods in culture. These changes, mostly detected by karyotypic analyses, commonly involve nonrandom gains of chromosomes 12, 17, 20 and X, or fragments of these chromosomes[2–12]. The embryonal carcinoma (EC) stem cells of human teratocarcinomas, the malignant counterparts of ES cells, though typically highly aneuploid, always contain amplified regions of the short arm of chromosome 12 and, commonly, gains of chromosomes 1, 17 and X[13–16]. Gain of chromosome 20q has also been noted in yolk sac carcinoma and nonseminomatous germ cell tumors, which contain EC cells[17–19]. Such observations suggest that these specific genetic changes in ES cells may be related to the nature of pluripotent stem cells themselves rather than the culture conditions. Mouse ES cells also undergo karyotypic changes upon prolonged passage[20], often with gain of mouse chromosomes 8 and 11 (ref. 21); mouse chromosome 11 is highly syntenic with human chromosome 17 (ref. 22).

Structural variants in otherwise karyotypically normal human ES cells have also been described[10,11,23,24]. These structural variants include gains on chromosome 4, 5, 15, 18 and 20 and losses on chromosome 10, although only gains on chromosome 20 were commonly observed in multiple cell lines.

Marked epigenetic changes have also been noted on prolonged passage; studies of global DNA methylation in human ES cells found considerable instability with time in culture[25,26]. Functional gain of the X chromosome, resulting from loss of X-chromosome inactivation in culture-adapted ES cells with two karyotypically normal X chromosomes has been reported[27]. On the other hand, some imprinted genes retain their monoallelic expression over long-term culture of human ES cells, although this stability is not invariant for all loci[28–31].

Because stem cells can adopt alternative fates (that is, self-renewal, differentiation or death), it might be expected that those maintained in the pluripotent state for many passages would be subject to strong selection favoring variants that enhance the probability of self-renewal[32]. Viewed in this light, the increased frequency of genetic variants in ES cell cultures over time might be considered inevitable[33]. Indeed, ES cell lines do often show progressive 'adaptation' to culture, with the result that late-passage cells may be maintained more easily, showing enhanced plating efficiencies[27]. Similarly, some mouse and human EC cell lines derived from germ cell tumors are nullipotent, as if selected for the capacity for self-renewal exclusively[34,35]. Taken together, these observations suggest that acquisition of extra copies of portions of chromosomes 12, 17, 20 and X by human ES and EC cells is driven by increased dosage of a gene or genes that favor self-renewal, independent of culture conditions.

However, there are also reports of human ES cell lines that have been maintained for many passages *in vitro* without overt karyotypic changes. It has been argued that some culture techniques, such as manual 'cutting and pasting' of ES cell colonies, favor maintenance of cells with a diploid karyotype[3,6]. As the appearance of a genetic variant in an ES cell culture must involve both mutation and selection, the low population size in cultures maintained by these methods may simply beat the mutation frequency[33]. Nevertheless, culture conditions themselves might influence the mutation rate independently of selection, and a population bottleneck, such as cloning, could allow a viable genetic variant to dominate in the absence of a selective advantage.

[1]A full list of authors and affiliations is provided at the end of the paper.

Received 6 September; accepted 26 October; published online 27 November 2011; doi:10.1038/nbt.2051

Candidate genes from the commonly amplified regions can be posited to provide the driving force for selection of variant ES cells, but direct evidence for the involvement of any specific gene is lacking. For example, *NANOG,* on human chromosome 12p, promotes the self-renewal of ES cells when overexpressed[36–38], but one of the two minimal amplicons of chromosome 12p in EC cells has been reported to exclude the *NANOG* locus[39]. It is also unclear to what extent changes affecting different loci are selected independently of one another or whether alterations at some loci act synergistically. Further, overexpression of disparate genes affecting a common pathway(s) could lead to an increased proliferative potential. Although the frequent gain of chromosomes 12, 17, 20 and X in both ES and EC cells argues for a selective advantage independent of culture conditions, changes affecting other regions might be more likely to depend upon culture conditions.

To provide better insight into the frequency and types of genetic changes affecting human ES cells on prolonged passage, the International Stem Cell Initiative (ISCI) surveyed by karyology and high-resolution SNP array 125 independent human ES cell lines, provided by 38 laboratories in 19 countries around the world, particularly to identify the common genetic changes that occur during prolonged culture (**Supplementary Table 1**). An opportunity was also taken to screen the samples against a specialized custom DNA methylation array focused on polycomb-target genes. These likely play a role in controlling ES cell differentiation and could be primary targets for the types of epigenetic change observed in cancer cells[40]. Thus, they may provide a source of selective advantage for variant stem cells. In most cases, each line was analyzed at both an early- and a late-passage level, using all three types of assay. The scale and design of this screen helped ensure that the ES cell lines sampled were representative of the world population. A group of 11 human iPS cell lines from three laboratories was also included to provide a pilot comparison of these pluripotent cells derived by reprogramming. Our results indicate that the common gains of chromosomes 12 and 17 in human ES cells are unlikely to be driven by the gain of single genes, but that the gain of chromosome 20 may be driven by the gain of a single gene, *BCL2L1*.

RESULTS

Diversity and population structure of the cell lines surveyed

To define the range of ethnicity represented by the human ES cell lines included in this study, we first analyzed the SNP calls identified in the SNP array data by referencing them to ethnically defined human genotyping data sets. Of the samples submitted for SNP analysis, three cell lines were included twice, and four pairs of ES cell lines and a set of three lines were identified as having a full sibling relationship (**Supplementary Table 1**). After accounting for these, 112 genetically unrelated ES cell lines passed SNP quality-control criteria. Subsequent analysis allowed us to determine whether specific structural variants found in particular cell lines are limited to the population from which they were derived or common to all human ES cell lines studied.

For population structure analysis, the international breadth of this study required the use of a diverse set of reference samples to compare to these 112 genetically unrelated cell lines. The reference samples were pooled from the HapMap[41], the human genome diversity panel (HGDP)[42] and the Pan-Asian SNP Initiative[43] to generate an ethnically diverse set of 1,868 reference samples. We performed cluster analysis[44] of the human ES cell samples against these reference samples, using the CEU (European), Chinese, Japanese and African HapMap populations as references, to arrive at the population structure of the human ES cell lines analyzed (**Fig. 1a**).

Of the 112 genetically unrelated ES cell lines, 61 (54%) were of European ancestry (excluding Middle East–East European and Central-South Asia–South European), 31 (28%) of Asian ancestry, 3 (3%) of African ancestry, 12 (10%) of Middle East and East European ancestry, and 4 (4%) of Central-South Asian and South European ancestry (**Table 1**). The European ES cell lines were further stratified using a recently described comprehensive European reference set[45] and were found to match subpopulations from many different regions of Europe (**Fig. 1b**). The cell lines of Asian descent were stratified into those of East Asian origin, including those of Han Chinese, Korean, Japanese and Indian origin, and those of Central or Central-South Asian origin (**Fig. 1c**,**d**). Five of the cell lines classified as Middle East and East European clustered with one another but not particularly close to any of the reference samples used in this study, namely clusters belonging to HGDP-Central/South-Asia, HGDP-Middle East and the HGDP-European samples (**Fig. 1d**). Four of these five lines were derived in Iran, and are most likely of Persian ancestry, a population not represented in the reference samples. It is notable that the nine ES cell lines most commonly cited in the scientific literature are representative of the genetic backgrounds of populations from northern, northwestern and central European, Han Chinese, Indian and Middle Eastern populations (**Table 1**).

Karyotype analysis

Stability of the cell lines. Analyses were carried out on all 120 human ES cell lines (including duplicate and sibling cell lines) provided for karyotyping at both early- and late-passage levels ('paired' lines), as well as on five additional lines that were provided only in early passage (**Supplementary Table 1**). Among this total of 125 lines, 42 (34%) had abnormal karyotypes (defined as at least two metaphases with identical, abnormal karyotypes of at least 30 metaphases screened) in at least one passage level. The data from this study confirm that human ES cells are commonly diploid soon after derivation, and that many do retain a normal karyotype after many passages (**Fig. 2a**).

Late-passage cultures of the paired lines were approximately twice as likely to have a chromosome abnormality (39/120, 33%) as those from the early-passage cultures (17/120, 14%). Among the five lines submitted only at an early-passage level, one (20%) had an abnormal karyotype with an extra copy of chromosome 17q. Of the 39 paired lines with abnormal karyotypes at late passage, 24 were normal at the early passage, whereas the remaining 15 also had abnormalities at both passage levels. In this case, the abnormalities seen at the late passage were mostly similar to those seen at the early passage. About half of all the abnormalities involved combinations of chromosomes 1, 12, 17 and 20 (**Fig. 2a**).

A number of cultures were mosaic with, mostly, two populations of cells, one with a normal karyotype and one with a particular abnormal karyotype; 10 of 24 with abnormalities only at late passage, and 8 out of 15 with abnormalities at both passage levels were mosaic (**Supplementary Table 1**). Five lines that were mosaic at early passage showed an increase in the abnormal cell population at late passage. In all of these cases, the abnormality involved extra copies of chromosomes 1, 12, 17, 20 or X. One pair showed additional chromosome changes in the late passage and one pair had unrelated abnormal karyotypes at each passage level. Two lines were scored as abnormal in early passage but normal at late passage. However, both were mosaic, with 3/30 metaphases in one case with a translocated chromosome t(2:19), and 5/30 metaphases in the other with a duplication on chromosome 13. Both chromosomal rearrangements were unique to these lines and most likely represent random changes that were outcompeted by the normal cells over time.

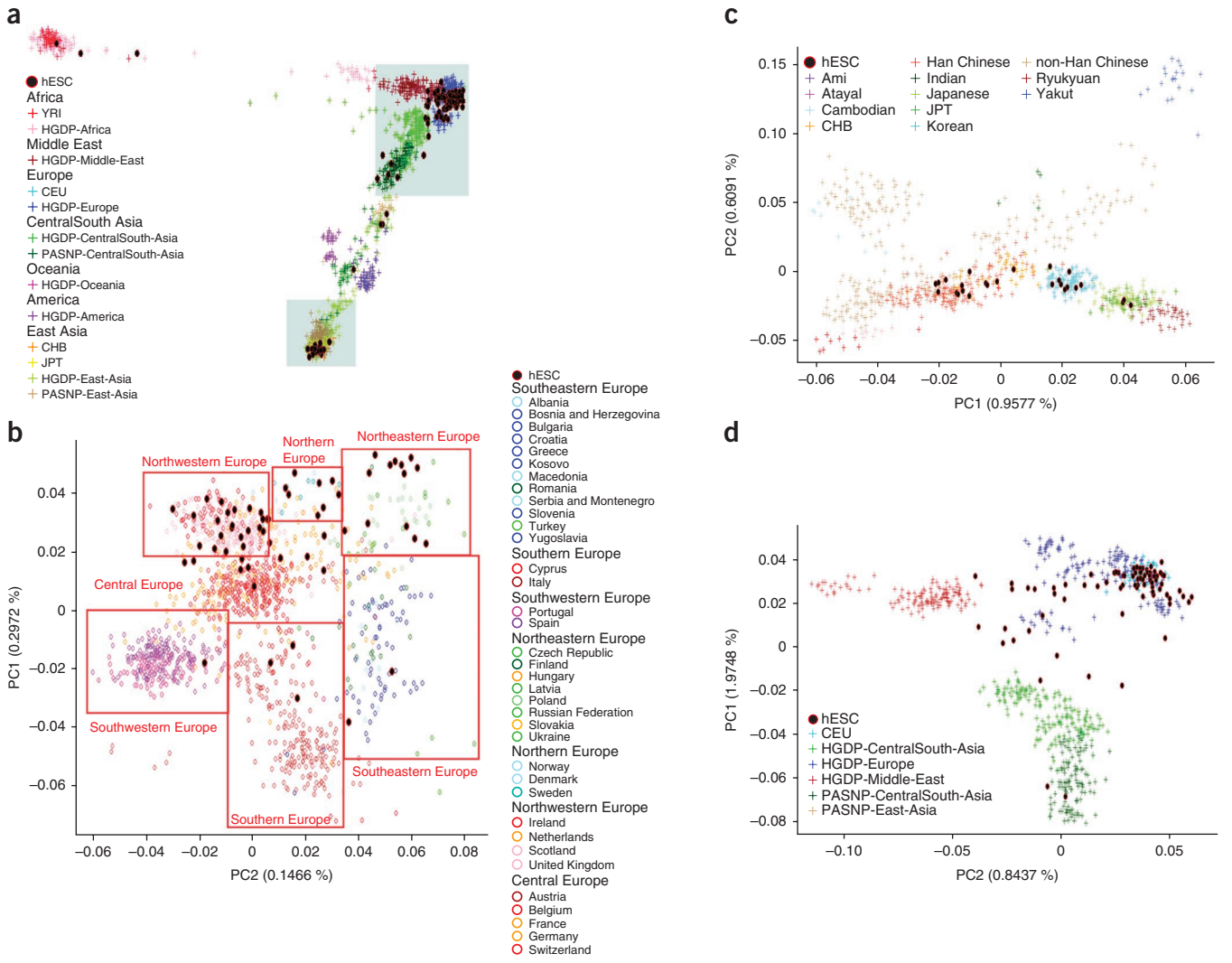

Figure 1 Population structure of the human ES cell lines analyzed. Principal component (PC) analyses were conducted on the entire final merged data set. PC1 and PC2 are plotted on the *y* and *x* axes, respectively. (**a**) The overall distribution of the human ES cell lines studied compared to the major ethnic groups identified in the HapMap study[41], the human genome diversity panel (HGDP)[42] and the Pan-Asian SNP Initiative[43]. (**b**–**d**) The cell lines were further subdivided to show their relationships to European (**b**), East Asian and Indian (**c**) and Middle East-European–Central South Asian populations (**d**).

Among the 11 iPS cell lines examined, three exhibited chromosome abnormalities, a frequency (27%) comparable to that found in ES cell lines. Of these, one line (RR01) exhibited trisomy 12 at both early and late passage. The other two lines were provided only at one passage level; one had a trisomy 12 (RR05) and one an inversion on chromosome 5 (RR03). None of these abnormalities were present in the somatic cells from which they were derived. These results are consistent with recent analysis of human iPS cell chromosomal instability both in the general frequency of aberrations and over-representation of chromosome 12 alterations[12,46].

A common source of cells with abnormal karyotypes. The proportion of cell lines with abnormal karyotypes did increase with delta, the difference in estimated number of population doublings ($P = 0.048$) (**Fig. 2b**). There was also a marked variation in the proportion of abnormal ES cell lines submitted by the different participating laboratories. The 42 abnormal lines were among cell lines submitted by 21 laboratories, whereas no abnormal lines were found among the other 38 lines submitted from the remaining 11 laboratories. This was not directly linked to the delta of the submitted lines and might simply reflect the stochasticity of mutation, or could imply a laboratory effect. The cell lines in each category were from diverse ethnic origins, and were cultured under very similar conditions, although a role for subtle variations in culture technique cannot be excluded. Nevertheless, consistent with suggestions that enzymatic mass-passaging techniques may favor the generation of abnormalities, a twofold higher proportion of the paired lines that had an initially normal karyotype but became abnormal at late passage were passaged by enzymatic methods (18/58, 31%), relative to those passaged by the manual cut-and-paste technique (6/43, 14%) (χ^2, $P = 0.009$). This effect is significant even after adjusting for delta ($P = 0.017$).

Candidate regions/genes. Aberrations of all chromosomes with the exception of chromosome 4 were observed (**Fig. 3**). However, most chromosomes were affected in very few instances, and four cell lines with particularly abnormal karyotypes accounted for many of these sporadic changes (**Supplementary Table 1**). In addition, there were three instances of balanced rearrangements seen as sole aberrations,

Table 1 Ethnic origin of human ES cell lines analyzed indicating ancestry of those most often cited

Ancestry	Number of cell lines[a]	Most commonly used cell lines	No. citations (2008 to 2009)[b]
European	**63 (61[c])**		
Italian	4		
Southwestern European	2		
Southeastern European	2		
Northeastern European	14[d]		
Northern European	8	BG01	13
Northwestern European	24[d]	HUES7	18
Central European	11	H1	95
Asian	**33 (32[c])**		
Central Asian	3		
Central-South Asian	1		
Han Chinese	14	HES2	16
		HES3	14
Japanese	3		
Korean	9		
Indian	3[d]	HES-1	6
African	**4 (3[c])**		
East African	1		
West African	3[d]		
Middle East–East European	**14[e] (12[c])**		
		H9	122
		H7	25
		HSF-6	12
Central-South Asia South European	**4**		
Total cell lines	**118 (112[c])**		

[a]The numbers of cell lines shown includes only those that passed quality control for SNP analysis. [b]UMass Stem Cell Registry (http://www.umassmed.edu/iscr/hESCusage.aspx). [c]Total number of genetically unrelated cell lines. [d]Includes two cell lines from siblings. [e]Includes three cell lines from siblings.

a translocation between 2 and 19 in an early-passage human ES cell culture, an inversion of 11 in a late-passage culture, for which the early passage was normal, and a Robertsonian translocation between chromosome 21 and 22 in both passages of one line. There were also abnormalities affecting chromosome 7 in seven ES cells, but five came from one laboratory, suggesting an unknown cause particularly associated with that group, perhaps related to their derivation of ES cells from prenatal genetic screening material. By contrast, in most abnormal lines (25/42), the changes involved one or more of chromosomes 1, 12, 17 and 20. Of the 17 lines that were abnormal in early passage, eight had abnormalities involving these chromosomes, whereas, of the 24 lines that acquired abnormalities between early and late passage, 16 lines had changes involving acquisition of one or more of these chromosomes (**Fig. 2a**). Among the gains, there were minimal amplicons affecting the telomeric region of chromosome 17 (17q25) in two lines, and another affecting 20q11.2 was apparent in another line (**Fig. 3**). Gains of only the short arm of chromosome 12 were found in three cell lines.

The large differential in frequency between gain and loss of chromosomes is remarkable. In contrast to the 39 ES cell lines that showed gains of chromosomal material in late passage, 20 lines showed losses of chromosomal material. However, only two lines exhibited chromosomal deletions that were not caused by unbalanced translocations (one, UU03, had two unrelated deletions of chromosomes 6 and 18), although even in these there were also unrelated chromosome gains. Excepting the deletions on chromosome 7, which only occurred in the lines from one laboratory, three regions showed recurrent loss, 10p13-pter (five cases), 18q21-qter (five cases) and 22q13-qter (three cases); in several cases these were the sole changes (**Fig. 3**).

Structural changes determined by molecular karyotyping

Identification of ES cell–associated structural variants. As genomic structural changes do occur below the ~5 MB detectable limit of karyotyping, we used SNP data to identify structural variants and detect structural changes down to a minimum of 1 kb in length. We identified structural variants for all samples that passed quality control, but restricted our detailed analyses to those cells judged to have a normal karyotype, because of the difficulty of ascribing functional significance to a small structural genomic change in a background of a much larger karyotypic abnormality. Nevertheless, we did examine the breakpoints in six cases of balanced rearrangements (PP-107, NN-12, J-02, CC-05, AA-03, RR-03) but found no evidence of structural variants associated with these (**Supplementary Table 1**). In addition, although loss of heterozygosity can be detected with the SNP platform, we focused our attention primarily on structural variant analysis as this is the more likely structural change to lead to a selective advantage. Nonetheless, we provide a spreadsheet of overlapping loss of heterozygosity across the 225 human ES cell samples and an associated .bed file with all loss-of-heterozygosity calls (**Supplementary Data Sets 1** and **2**). Structural variants were identified in the 200 DNA samples from karyotypically normal ES cells that passed quality control by comparison with the reference genome (hg18). Further quality controls removed one sample due to an extremely high number of structural variants called and two more for extremely high total length of structural variants (**Supplementary Fig. 1**). A total of 27,409 deletions with an average size of 40.2 kb, and 7,413 duplications with an average size of 95.4 kb, were detected. The sizes of these structural variants and the total number of differences between deletions and duplications are consistent with previous structural variant studies of human populations[47]. As structural variants are a common feature of variation between individuals, the majority of structural variants detected in the human ES cells most likely represent the condition of the genomes of the respective embryos from which they were derived, and are unrelated to human ES cell culture.

To aid in distinguishing culture-associated structural variants, we compared the human ES cell structural variants to those identified using the same platform to analyze a set of 267 HapMap samples (raw data directly supplied by Illumina). Though relatively restricted in population diversity compared to our human ES cell data set, the HapMap samples provide a set of common reference structural variants. Our subsequent analyses focused only on variant regions enriched in human ES cell lines over the HapMap samples. We identified 504 regions of gain and 860 regions of deletion in the karyotypically normal ES cell lines as 'ES cell associated' (**Supplementary Data Set 3** and **Supplementary Table 2**).

Genome-of-origin variants. The apparent ES cell–associated structural variants most likely include some rare and/or localized variants absent in the HapMap set, yet unrelated to human ES cell culture selection. There are a number of examples in which a particular variant occurs in a single line in both the early and late passage. Although we cannot exclude that such variants arose in culture before the early-passage samples being obtained, it is more likely they represent rare and/or localized variants present in the genomes of the donated embryos. We did see such a case among the iPS cell lines for which we have SNP data from the parental somatic cell line. For instance, in three iPS cell lines derived from the same parental fibroblast, the same rare gain (chr12:106,928,902-107,008,902) was detected in both the early and late passages and the parental line (**Supplementary Data Set 3**). Also, among the sibling human ES cells lines, we found recurring rare variants specific to each family. For instance, a gain at chr3:45,220,749-45,263,539 was found in the early and late passages of human ES cell lines G02 and G05, although this allele was absent in G04, the third of these sibling lines. At another

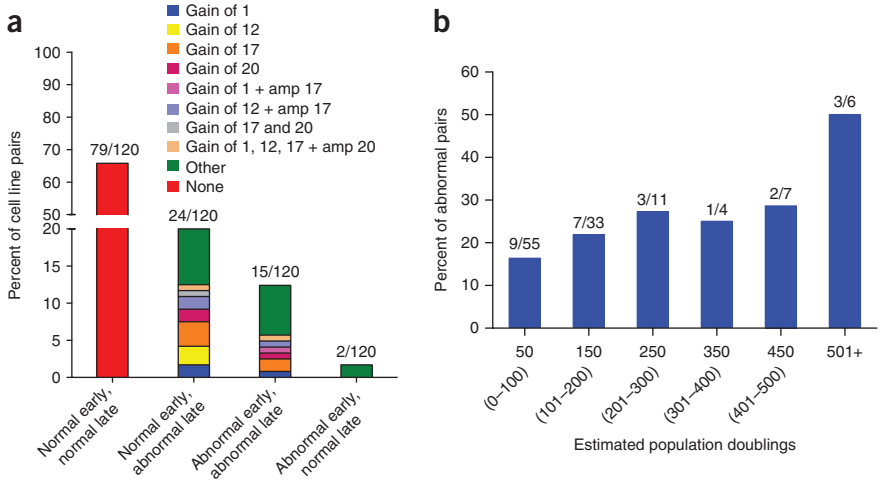

Figure 2 Cytogenetic changes occurring during prolonged passage of human ES cells. (**a**) Percentage of human ES cell line pairs that exhibited a karyotypic abnormality in either early or late passages, or both. Cell lines were excluded if they were known to be derived from karyotypically abnormal embryos. The ES cell pairs are grouped according to whether the chromosome change was observed at late passage only (normal early, abnormal late), both at early and late passages (abnormal early, abnormal late) or early passage only (abnormal early, normal late) and no chromosomal change (normal early, normal late). The percentage of cell lines that have individual gains of chromosomes 1, 12, 17 and 20, gain of chromosomes 1 and 17, or gain of chromosomes 1, 12, 17 and 20 are highlighted. Chromosome changes not involving 1, 12, 17 and 20 are indicated as 'Other'. The numbers above each bar indicate the total number of lines that fall into the four categories out of the total number of pairs of lines analyzed. Two cell lines (CO2 and CC05) in the 'abnormal early, abnormal late' category were known to be derived from karyotypically abnormal embryos (a trisomy 13 and ring chromosome 18). One abnormal cell line (AA06) has been excluded from this figure as only one passage was available for analysis. (**b**) Proportion of pairs of lines that acquired karyotypic abnormalities over different periods in culture. The pairs of lines are grouped according to 'Delta', the difference in estimated population doublings between the early and late passages. Only those lines that had a normal karyotype at the early-passage level were included in the analysis, and of those only 115 pairs could reliably be assigned an estimated population doubling time estimate.

location, chr3:167,536,633-167,837,107, a gain occurs in the early and late passage of all three of these sibling lines. For the purposes of this study, we have assumed that none of these rare variants arose during ES cell culture, and we define them as 'genome-of-origin' variants (**Supplementary Table 2**).

Dynamically changing variants. Some structural variants that were detected are represented in the HapMap population and change dynamically in ES cell culture, suggesting the labile nature of at least some genomic elements. For example, 18 human ES cell lines had a gain at chr17:75,289,455-75,296,305 (**Supplementary Table 2**, labile structural variant), but this was also present in four HapMap samples. Among the human ES cell samples, this structural variant was present in the late but not early passage of four lines, but in five other definitive cases it was present in the early but not late passage. Thus, this represents a dynamically changing variant with no evidence for positive selection in human ES cell culture but provides an example of the labile nature of the human genome.

Structural variants enriched in late-passage cultures. In the subset of structural variants enriched in the ES cells, there was no overall trend toward a gain of total structural variant numbers between early-passage and late-passage samples: that is, there was an increase in the total number in the late passage of some lines, but a decrease in others (**Supplementary Table 2**). Among the particular structural variants that did show increases in several lines in a late passage, a number encompassed regions known to encode genes that may be relevant to human ES cell behavior, but they were isolated instances. For example, a deletion variant spanning the *SOX21* locus, a gene encoding a transcription factor associated with differentiation of human ES cells, was found in one line (UU03-E), and a minimal deleted region at chr4:983425-997875, which spans the promoter and first exon of *FGFRL1*, was present in the late but not early passage of two lines (L03-l, TT20-l). FGFRL1 is expressed in human ES cells and may act as an inhibitory sink for FGF2, which is important for human ES cell maintenance[48]. Late-passage samples of both the MM01 and TT20 lines share a minimal overlapping deletion variant of ~540 bp at chr3:196,472,618-196,473,157. This spans a highly conserved open reading frame (C3orf21) that is expressed in human ES cells but has no known function[48].

Structural variants in karyotypically normal ES cells

We next analyzed structural variants in regions subject to common karyotypic abnormalities. In one region of chromosome 1q, two cell lines (V09 and FF01) in late, but not early, passage, have an overlapping 3.1 MB gain (chr1:199,985,282-203,092,388), which spans *JARID1B*, a polycomb-related gene encoding a histone H3 lysine-4-demethylase[49,50]. On chromosome 12, two cell lines (B02 and F04) have an overlapping gain of 1.1 MB in chr12:5,592,150-6,749,326 in the late-passage samples. This structural variant is within a minimal amplicon identified by karyology (12p13.31) and includes *NANOG*, *CD9* and *GAPDH*, all of which are expressed in human ES cells. There was little evidence for repeated occurrence of gains below the resolution of standard banding techniques in regions of chromosome 17 (**Supplementary Fig. 2**).

In contrast, there was a striking occurrence of a structural variant gain within chromosome 20 in 22 karyotypically normal cell lines. Notably, these gains, many validated by quantitative PCR (**Supplementary Fig. 3**), are within the minimal amplicon 20q11.2 found by karyology (**Fig. 4**). Among these 22 cell lines, there were five instances where the gain was detected in both early and late passage but 17 instances where it was detected only in the late passage. There were no instances of this gain in early passage but absence in late passage of the same cell line. This gain was also present in an ES cell line (J01) that had an abnormal karyotype at late passage and in an iPS cell line (RR01) that contained an extra copy of chromosome 12 (**Supplementary Table 1**). This strongly suggests that once genomic rearrangements occur in this region, they provide a positive selective advantage during subsequent culture. The least difference in passage number between the early and late passage from the same cell line, which had the gain in the late passage alone, was 22. The apparent strong positive selection for gain of this region suggests that a gene providing a cell-autonomous functional advantage under normal human ES cell culture conditions is encoded within the DNA of the

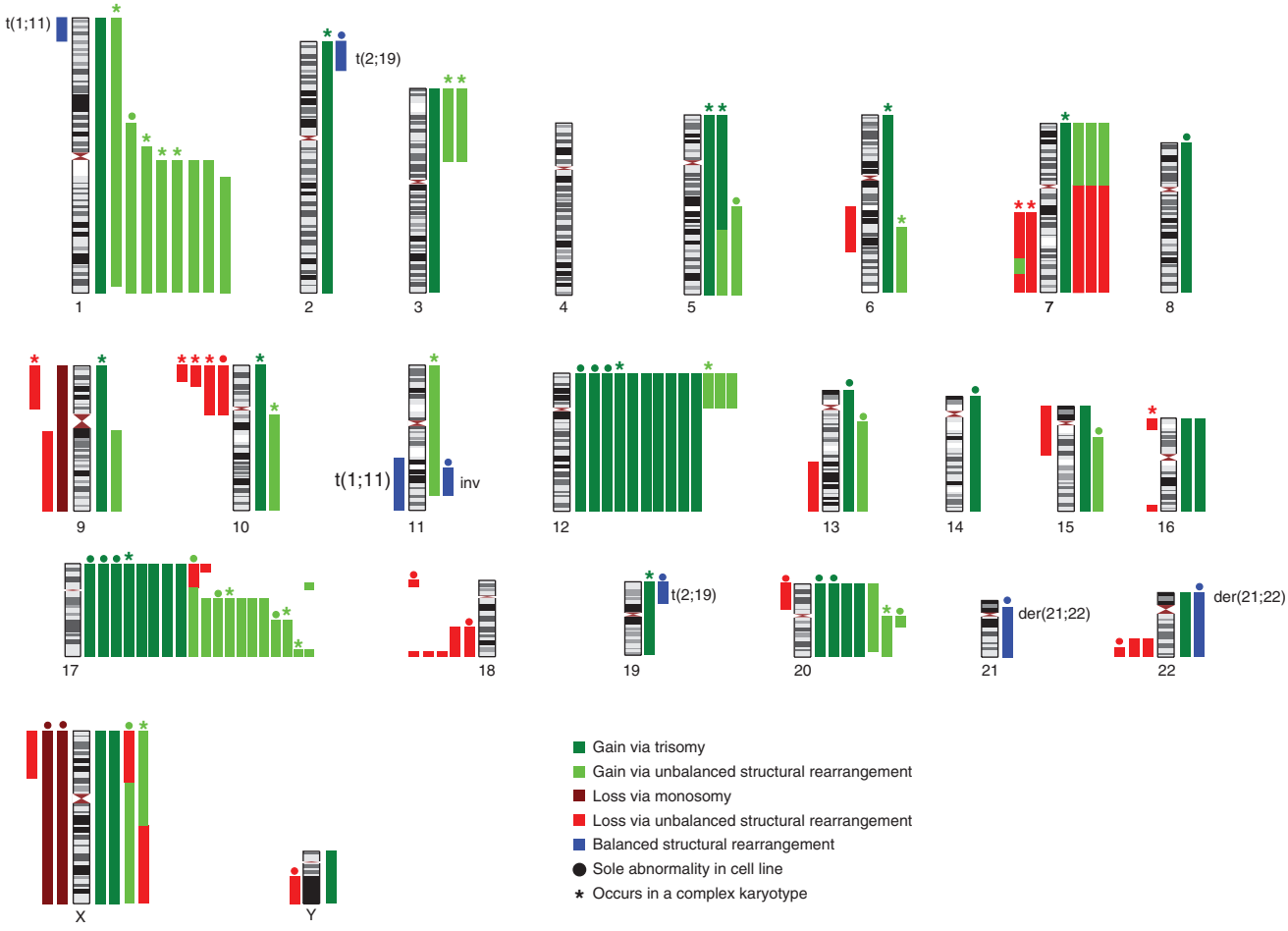

Figure 3 Ideogram demonstrating the chromosome changes found in this study. Each colored bar represents one chromosome change occurrence in one cell line. Chromosome losses and gains are shown to the left and right of the ideogram, respectively, except that those instances where a single chromosome rearrangement results in a gain and a loss the colored bars are shown together for clarity. The cytogenetic changes are color coded: Maroon, loss of a whole chromosome (monopsony); red, loss via a structural chromosome rearrangement (unbalanced translocation or interstitial deletion); dark green, gain of a whole chromosome (trisomy); light green is gain via a structural chromosome rearrangement (unbalanced translocation or interstitial duplication); blue represents the occurrence of an apparently balanced rearrangement the nature of which is labeled. Instances in which a change affected only a single chromosome are denoted by ●, whereas changes associated with complex karyotypes (>5 unrelated chromosome aberrations) are denoted by ★. Two cell lines (C02 and CC05) were known to be derived from karyotypically abnormal embryos and contain a trisomy 13 and ring chromosome 18 respectively. iPS cell lines are excluded from this figure. Based upon these studies the minimal critical chromosomal regions subject to gain in culture adapted human ES cell lines were 1q21-qter, 12p11-pter, 17q21.3-qter and 20q11.2. The minimal regions subject to loss were 10p13-pter, 18q21-qter and 22q13-qter.

shared overlapping region. Moreover, three cell lines (F-01, Q-02 and K-05) that had a normal karyotype and a 20q11.21 structural variant gain in early passage acquired an abnormal banded karyotype in samples from later passage. The late-passage abnormal karyotypes of F-01, Q-02 and K-05 were 46,XX,der(15)t(15;17)(p11;q21); (47,XX,+der(1)(t(1;1)(p?21.2;q11); and 47,XX,t(1;11)(p?36;q13),trp(17)(p11.2),+20, respectively. This preliminary evidence suggests that early gains in 20q11.21 might promote further subsequent genetic change.

The duplicated regions of chromosome 20 in the various cell lines and the minimal amplicon are diagrammed in **Figure 4b**. The proximal ends of each of the structural variant gains within chromosome 20 are presumed to lie in a nonbridged sequencing gap sized at 1 MB near the centromeric region of the long arm. The most proximal SNP identified in all these gains is the first occurring after this gap, at position 29,267,954. The distal end of the gain varies across the lines. The longest gain extends to 31,793,485 with a measured length of 2.5 MB, similar to the shortest karyotypically identified gain in this region, dup(20)q11.21 in cell line UU01 (**Fig. 3**). The shortest gain is 0.55 MB extending to 29,821,940 and contains 13 genes (**Fig. 4c**). Three of these genes, *ID1*, *BCL2L1* and *HM13*, are known to be expressed in human ES cells based on mRNA-Seq data (**Fig. 4c**) and published microarray data[27]. Although a single RefSeq-annotated microRNA lies in this region, there is no evidence for its expression in human ES cells[51]. Further, combined with the mRNA-Seq data, ChIP-Seq data from H1 human ES cells of histone modifications, considered universal predictors of enhancer and promoter activity (H3K4me3, H3K27ac), do not suggest additional functional regions other than those associated with the three RefSeq genes identified as expressed in human ES cells (**Fig. 4c**).

When four pairs of cell lines with and without the chromosome 20 gain were analyzed, there was no clear correlation between increased expression and the presence of the 20q11.21 gain for these three expressed genes (**Fig. 4d**). Nevertheless, preliminary results indicated a strong selective advantage in culture for cells with the gain

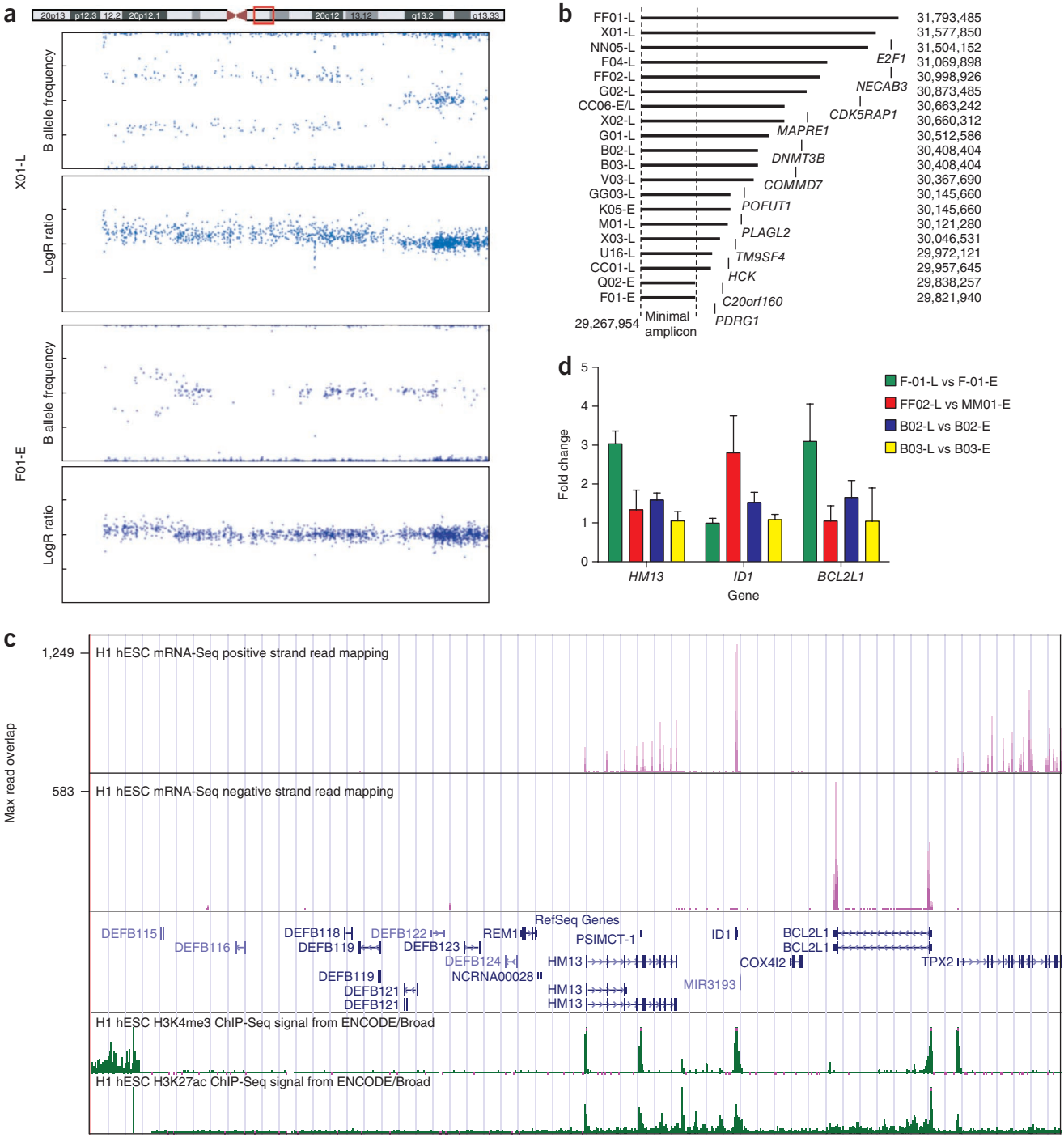

Figure 4 Copy number variation occurrence in human ES cell lines during prolonged passage. (**a**) 20q11.21 gain. The region on chromosome 20 frequently found to experience gain over extended human ES cell culture is indicated by the red boxed region in the chromosome ideogram. Also shown are the B allele frequency and logR ratio plots representing instances of one of the longest and one of the shortest 20q11.21 structural variants. (**b**) Length representation of all individual occurrences of gains in the 20q11.21 region. Samples from which the structural variant was derived are indicated on the left-hand column. The invariant 5′ region and the variable 3′ positions are indicated. Position of genes outside of the minimal amplicon that show greater than 20 RPKM level of expression in human ES cells are shown (RPKM = number of reads that map per kilobase of exon model per million mapped reads for each gene). (**c**) Expression, RefSeq gene, and regulation tracks in the minimal amplicon. Positive and negative strand mRNA-Seq data from H1 human ES cells indicating polyA RNA transcripts expressed within the minimal amplicon region (chr20:29,267,954-29,853,264) are shown together with H1 human ES cell ChIP-Seq data of histone modifications considered universal predictors of enhancer and promoter activity. (**d**) Comparison of expression levels of three genes (*HM13*, *ID1*, *BCL2L1*) contained within the identified minimal 20q11.2 amplicon between early- (normal) and late-passage (20q11.2 CNV carrying) samples. MM01 and FF02 are genetically identical sub-lines from two separate laboratories, MM01 has no amplification at 20q11.2, whereas FF02 possesses a copy number change at 20q11.2 that includes the identified minimal amplicon (**b**).

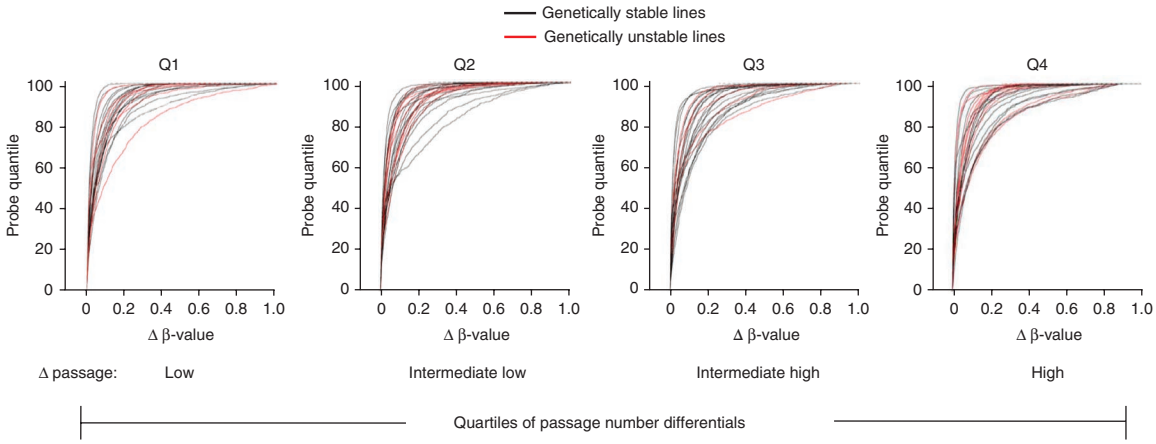

Figure 5 Cumulative distribution function of methylation changes in human ES cells in this study. The change in DNA methylation is represented by empirical CDF curves of the absolute difference in DNA methylation between early- and late-passage cell-line pairs for all 1,536 analyzed probes. The black curves denote genetically stable lines; the red curves denote genetically unstable lines. All analyzed lines were divided into quartiles based on the passage-number difference between the early and late member of each pair. The first quartile contains the lines with the lowest difference in passage number between the early and late sample (range 4 to 47), whereas the fourth quartile contains the lines with the highest difference in estimated population doublings (range 210 to 1,482).

over those without (**Supplementary Fig. 4**). It has also been recently reported that Bcl-X_L, the long, anti-apoptotic isoform encoded by the *BCL2L1* locus, can suppress apoptosis in human ES cells and increase their cloning efficiency[52]. Further, when we transfected MM01 ES cells with a constitutive vector encoding Bcl-X_L, the predominant isoform expressed in human ES cells, these cells showed a distinct growth advantage with respect to the parental cells (**Supplementary Fig. 4**).

DNA methylation analysis

To examine whether cell lines that are genetically unstable at the karyotype level tend to show higher levels of epigenetic instability, we analyzed DNA methylation patterns, focusing on developmentally relevant genes known to be targets of abnormal promoter DNA methylation in cancer[40], and thus most likely to be subjected to selection for altered expression during culture adaptation. For this we used a custom GoldenGate DNA methylation array developed to interrogate DNA methylation changes in known polycomb group protein (PcG) targets in human ES cells[53]. In general, the DNA methylation patterns of the human ES cells tended to be unstable, with both increases or decreases depending upon the locus (**Fig. 5** and **Supplementary Data Set 4**). **Table 2** summarizes those genes that were most frequently subject to gain or loss of methylation during passage, or that showed the least change. Overall, we did not observe any hot spots for DNA methylation at the ~1,500 loci interrogated in the array used in this study, and chromosomes 12, 17 and 20 were not any more methylated, on average, than the rest of the genome.

Table 2 The top 20 genes that were most frequently gained, lost or showed no change in DNA methylation levels in the 120 ES cell lines analyzed at early and late passage

Gained DNA methylation	Lost DNA methylation	No change in DNA methylation
GPC3	*CBLN4*	*NR4A3*
RAB9B	*HIST1H3C*	*EPHA4*
TCEAL4	*LY6H*	*COL12A1*
IL1RAPL2	*HIST1H4L*	*TIGD3*
ESX1	*ANKRD20B*	*SNX7*
TCEAL3	*HIST1H4F*	*PIP5K1B*
AMMECR1	*DMRT2*	*KCNJ2*
MGC39900	*TTLL7*	*T*
LRCH2	*FOXD4L1*	*ZBTB7A*
ZCCHC12	*FOXD4L2*	*IL20RA*
REPS2	*ONECUT1*	*GNAO1*
SOX3	*MAL*	*EPB41L4A*
RP13-360B22.2	*SYT6*	*VDR*
TSC22D3	*BHLHB4*	*HS6ST3*
NHS	*HIST1H3I*	*VGLL2*
TCEAL7	*XTP7*	*SIX1*
MGC4825	*NEUROG1*	*SFT2D2*
GPR50	*TFAP2D*	*BCAN*
BCL2L10	*DRD5*	*ELMOD1*
CDX4	*ASCL2*	*PTGER4*

GPC3 gained more than 5% DNA methylation (range: 98–5%) in over 70% of the samples analyzed, whereas *CBLN4* lost more than 5% DNA methylation (range: 70–5%) in over 60% of them. The genes listed in the "No change" column showed fluctuations in DNA methylation <1% in all samples profiled.

As shown by cumulative distribution function (CDF) curves, most cell lines underwent extensive DNA methylation changes during their time in culture (Online Methods). However, there was a marked difference between the cell lines. For example, in some cell lines there were few changes observed even if there was a large difference in passage level between the early- and late-passage samples (**Fig. 5** Q4 and **Supplementary Table 3**), whereas with other pairs there were large differences observed even when the passage-level difference between the samples was small (**Fig. 5** Q1 and **Supplementary Table 3**). However, the causes of the variation in methylation stability between the lines were not evident. There was no obvious laboratory effect, and the karyotypically abnormal cell lines were not any more unstable than their karyotypically normal counterparts. This suggests that genetic instability played little to no role in the epigenetic instability of the cell lines analyzed. In addition, the DNA methylation patterns of the sibling ES cell lines were as different between themselves as they were between unrelated lines (**Supplementary Data Set 4**), suggesting that the genetic background of human ES cells plays a minor role in the degree of their epigenetic instability.

DISCUSSION

The occurrence of genetic and epigenetic change in human ES cells on prolonged passage is clearly important with respect to their use in regenerative medicine. Understanding the key genes involved and the mechanisms that drive change is important, not only for minimizing the impact of such variants in applications of ES and iPS cells, but also

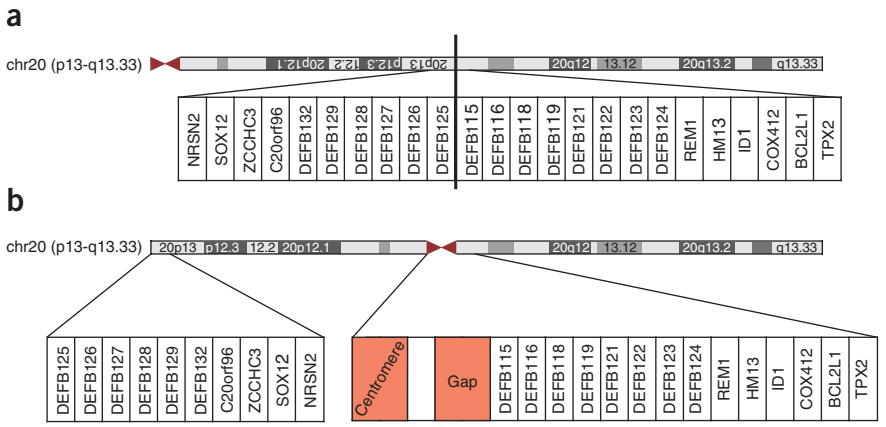

Figure 6 Recent pericentric inversion associated with 20q11.21 susceptibility to gain. (**a**) The ancestral condition of chromosome 20 before a pericentric inversion in the last common ancestor of the gorilla, chimp and human. (**b**) Structure of human chromosome 20 with the location of the gap indicated in which the proximal end of all 20q11.21 amplicons lie.

for exploring the mechanisms that control the fate decisions of pluripotent cells between self-renewal, death and differentiation. Nevertheless, given the scale of the present study, it is striking that most of the ES cell lines studied (79/120 pairs, 66%) remained karyotypically normal, even after many passages, whereas it was only with respect to chromosome 20 that evidence for structural variants in a specific region offering a strong selective advantage could be deduced. Among the small number of iPS cell samples studied, 3 out of 11 had abnormal karyotypes, with 1 of the 3 having the 20q11 gain in the late-passage sample.

Since the first reports of nonrandom chromosomal gain in human ES cells, many studies by standard karyology and by various molecular techniques, including CGH and SNP arrays, have found that, indeed, certain regions of the genome of both ES and, more recently, iPS cells are particularly subject to such genetic change upon prolonged passage in culture. Recently, it was also shown that iPS cells acquire mutations during their derivation, although many such mutations are lost on subsequent passaging[54]. It is commonly assumed that those genetic changes that repeatedly appear in pluripotent stem cells provide variant cells with a growth advantage, but the nature of the selective advantage is unclear. At the molecular karyotype level, it is difficult to disentangle changes that simply reflect variants existing in the human population from those acquired during culture. To address this, we explicitly sought to compare the genomes of a large set of human ES cell lines at two different passage levels and from as diverse a set as possible of the principal laboratories isolating these cells around the world. Although the number of human ES cell lines that have been derived worldwide is uncertain, the 125 ES cell lines analyzed in this study represent a substantial proportion of those commonly available. Notably, our data show that these lines include representatives of most major ethnic groups, reflecting far greater ethnic diversity than previously reported[55,56].

One feature of the human genome emphasized by the current study is that some regions are especially dynamic, particularly but not exclusively those including repetitive elements. In the current panel of ES cells, many regions showed gains or losses between the passage levels, but with no consistency, suggesting that there is no common selection pressure driving the copy number changes. That such dynamically variable regions were readily detected suggests that human ES cell cultures may go through population size restrictions more often than appreciated. Indeed, the cell cycle time of human ES cells is about 18–20 h, but common culture practice involves splitting cultures at low split ratios every 4–5 d or longer. This implies a very large proportion of undifferentiated cells, maybe as many as 90%, are lost between passages of stock cultures[33].

Likewise, the DNA methylation status of the ES cell lines also appeared to change dynamically. Although there was a marked increase in differential DNA methylation with time, indicated by the greater number of DNA methylation changes in the cell lines with the highest differences in passage number, there was also a substantial variation between lines that had undergone similar differences in passage numbers. Thus, human ES cells change not only genetically, but also epigenetically in culture. This conclusion is consistent with several other smaller scale studies that have interrogated human ES cells with respect to either general DNA methylation[25], or imprinting[29,31]. These studies all found DNA methylation and imprinting changes that appeared to be variable between lines and were locus dependent. However, we could not identify specific recurring regions subject to methylation in the genome and there was no observed correlation between DNA methylation changes and chromosomal abnormalities. This suggests that, in general, changes in DNA methylation may be a dynamic process and not necessarily associated with adaptation as such. This point is reinforced by the observation that DNA methylation is markedly different between sibling lines.

In addition to these apparently stochastic and dynamic changes in the genome and epigenome, we did detect marked nonrandom changes in certain parts of the genome. The karyotypic changes seen in the current study match well with other published reports (**Supplementary Fig. 5**)[1]. Gains of chromosomes 1, 12, 17 and 20 and losses of chromosomes 10p and 18q are common in both data sets, and it is only gains of chromosomes 12, 17 and 20 that are often seen as a sole karyotypic change. However, recurrent deletion of chromosome 22q is a novel finding. On the other hand, the gain of chromosome X is a relatively common finding in published studies, whereas only two instances of gain and three instances of loss were observed in the present study. In the light of their relatively frequent occurrence, the minimal amplicons 1q21-qter, 12p11-pter, 17q25-qter and 20q11.2, and perhaps minimal deletions 10p13-pter, 18q21-qter and 22q13-qter deserve special attention as being likely to harbor genes of particular importance for the culture adaptation of human ES cells.

The frequent nonrandom gain of chromosomes 1, 12, 17 and 20 suggests that these chromosomes include a gene(s) that, when overexpressed, confers a growth advantage. Yet, it is striking that in our current extensive study, as in previous studies, structural variant analysis did not point to any frequent repetitive minimal amplicon occurring on chromosomes 1, 12 and 17. Obvious candidate genes are located on these chromosomes—for example, *NANOG* on chromosome 12—but none seems to be more subject to structural variants than other genes on these chromosomes in the absence of karyotypic change. We did see gains spanning the neighboring *SLC2A3*/*NANOGP1* region described in a recent study[46] but this is just as prevalent, if not more so, within our reference samples and spread across most major ethnic groups, suggesting it is a common structural variant in the human population rather than specific to human ES cells. Together, these observations suggest that the selective advantage attributable to the

gain of chromosomes 1, 12 and 17 may depend upon overexpression of genes or genetic elements at multiple, spatially separated loci, or upon the combination of a structural gene with a long range *cis*-acting regulatory element such that both units must be amplified together to yield an increased function. Alternatively, the appearance of gains within smaller regions may be restricted by chromosomal structure less susceptible to this form of mutation.

By contrast, and in agreement with other studies[5,10,11,23,46,57], our karyotypic and structural variant data point to a region (20q11.21) that, when amplified, apparently drives selection. In this study, because of the much larger number of cell lines and our ability to compare early and late passage, we were able to map the gain to a specific region. Other studies have also reported that gains in this region are associated with enhanced growth characteristics[23], and at least some of the lines in the present study were reported by their contributors to have increased population growth rates (data not shown). The frequency of this gain (25% of the karyotypically normal cell lines), combined with the enrichment in late-passage samples, clearly indicates its selective advantage in human ES cell culture. The mechanism for the selective advantage presumably lies in the minimal region shared by all 22 affected lines, a region containing 13 genes, only three of which are known to be expressed in human ES cells: *HM13*, *ID1* and *BCL2L1*.

A recent genome-wide RNA interference (RNAi) screen highlights the functional importance of *BCL2L1*, an anti-apoptotic factor, in human ES cell biology[58]. This RNAi screen ranked *BCL2L1* twenty-second of 21,121 genes in reducing proliferation after knockdown, whereas *HM13* and *ID1* were ranked 6,679th and 4,224th, respectively[58]. Additionally, a recent structural variant screen of >3,000 specimens from two dozen cancer cell types similarly identified a reoccurring gain on 20q11.21 in which *BCL2L1* was also contained within the minimal amplicon, and knockdown experiments indicated a role for *BCL2L1* in cancer cell proliferation[59]. Recently, it has also been reported that overexpression of the related anti-apoptotic gene, *BCL2*, enhances the survival of human ES cells[60], although *BCL2* is encoded with the region of chromosome 18 subject to recurrent loss in the current data set. Taken together, these observations suggest that similar mutations shared between ES and cancer cells lead to a selective advantage during clonal evolution. The temporal component of our study, where we see[17] instances of early/normal to late/mutated transitions, provides additional support for the notion that the 20q11.21 mutation is the driver mutation in the clonal evolution of these adapted stem cells. Although a role for *ID1* (ref. 61) and *HM13* cannot be excluded, enhanced cell survival due to elevated expression levels of *BCL2L1* offers the most likely mechanism.

The repeated appearance of a structural variant across multiple lines requires both a selective advantage for the variant (e.g., increased expression of *BCL2L1*), and a predisposition for the respective mutation to occur. It is noteworthy that the proximal end of all human ES cell 20q11.21 gains lies within a gap region of the current human assembly[62]. The presumption is that the highly repetitive sequence within this gap predisposes the region to structural rearrangement. With the link between genome rearrangements, primate evolution and disease association[63], it is notable that this gap coincides with a recent chromosomal rearrangement, a pericentric inversion[64], occurring in the last common ancestor of gorilla, chimp and human (**Fig. 6**). The gap region, possibly a centromeric remnant of a tandem duplication[62], introduces the repetitive sequence creating 20q11.21 rearrangement (or amplification) susceptibility. The frequency of appearance that is created by this combination of mutability and the decreased apoptosis warrants routine surveillance similar to that now done in karyotypic analysis.

The identification of genes that drive both cancer progression of EC cells in germ cell tumors and the progressive culture adaptation of ES cells has been a goal since the first clear recognition that gain of sections of the short arm of chromosome 12 is an invariant feature of EC cells[14]. The commonality of the changes in the tumors and in the ES cell in culture suggests common underlying mechanisms. However, the identification of a specific driver gene on chromosomes 1, 12 and 17 has been elusive, suggesting that more than one gene may be involved in the growth advantage of the aneuploid cells. Our present results now point to a specific gene subject to gain, most likely the anti-apoptotic gene, *BCL2L1*, on chromosome 20, that may promote the survival of ES cells *in vitro* and EC cells *in vivo*, thereby providing a strong growth advantage, whether in cancers or *in vitro*.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturebiotechnology/.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

The International Stem Cell Initiative is funded by The International Stem Cell Forum. The authors would like to acknowledge the following: Medical Research Council, UK (P.W.A., H.M.); Mohammad Pakzad & Adeleh Taei, Royan Institute (H.B., G.H.S.); California Institute for Regenerative Medicine (CIRM) (E.C., P.W.L.); Institute of Medical Biology, A*STAR, Singapore (J.M.C.); Ministry of Education, Youth and Sports of the Czech Republic (P.D., A.H.); Stem Cell Research Center of the 21st Century Frontier Research Program, Ministry of Education, Science & Technology, Republic of Korea (SC-1140) (D.R.L., S.K.O.); Ministry of Science and Technology of China (863 program 2006AA02A102) (L.G.); Swedish Research Council, Cellartis (O.H.); Department of Biotechnology, Government of India, UK-India Education and Research Initiative and the Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India (M.I.); Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation, Leading Project of the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program) of the Japan Society for the Promotion of Science (JSPS), Grants in-Aid for Scientific Research of JSPS and MEXT (T.I., S.Y., K.T.); Swiss National Science Foundation (grant no. 4046-114410) (M.J.); Shanghai Science and Technology Developmental Foundation (06DJ14001), Chinese Ministry of Science and Technology (2007CB948004) (Y.J.); funding from the North West Science Fund, UK (S.K.); One North East Regional Developmental Agency, Medical Research Council, UK, Newcastle University (M.L.); research funding from the Australian Stem Cell Centre (A.L.L.); The Netherlands Proteomics Consortium grant T4-3 (C.M.); Stem Cell Network, Canada (A.N.); National BioResource Project, MEXT, Japan (N.N.); Singapore Stem Cell Consortium (SSCC) & the Agency for Science Technology and Research (A*STAR) (S.K.W.O., P.R.) and the Genome Institute of Singapore Core Genotyping Lab (P.R.); Academy of Finland, Sigrid Juselius Foundation (T.O.); Conselho Nacional de Desenvolvimento Científico e Tecnológico/Departamento de Ciência e Tecnologia do Ministério da Saúde (CNPq/MS/DECIT), and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (L.V.P.); supported by the kind donation of Judy and Sidney Swartz (B.R.); financial support from the Faculty of Medicine, University of New South Wales (UNSW) and the National Health and Medical Research Council (NHMRC) Program Grant no. 568969 (Perminder Sachdev), South Eastern Sydney and Illawarra Area Health Service (SEIAHS) for making hES cell line Endeavour-2 available for this study, and H. Chung and J. Kim for their help in preparing the samples (K.S.); Academy of Finland (grant 218050), the Competitive Research Funding of the Tampere University Hospital (grant 9F217) (H. Skottman).

AUTHOR CONTRIBUTIONS

Project coordination: P.W.A. **Cytogenetic analyses**: D.B., A.D., E.M., K.D.M. and T.G.-L. **Molecular karyotyping by SNP BeadChip**: P.R. **DNA methylation arrays**: R.M.B. and P.W.L. **Administration and data curation**: A. Ford and P.J.G. **Data analysis and manuscript drafting**: P.W.A., S.A., D.B., N.B., R.M.B., P.J.G., K.H., L.H., B.B.K., Y. Mayshar, S.K.W.O., M.F.P. and P.R. **The scientific management of the ISCI project was provided by a steering committee comprising:** P.W.A., N.B., B.B.K., S.K.W.O., M.F.P., J.R. and G.N.S. **Sample contribution**: A. Colman, A. Robins, A. Hampl, A. Bosman, A.M. Fraga, A. Nagy, A.B.H. Choo, A.L. Laslett, A. Feki, A. Kuliev, A. Kresentia Irwanto, B. Reubinoff, B. Sun, C. Denning,

C. Mummery, C. Li, C. Olson, C. Spits, D. Ben-Yosef, D. Collins, D.J. Weisenberger, D. Ryul Lee, D. Ward-van Oostwaard, E. Chiao, E. Sherrer, Fei Pan, F. Holm, G. Anyfantis, G.Q. Daley, G.H. Salekdeh, G. Selva Raj, G. Caisander, H. Gourabi, H. Moore, H. Skottman, H. Suemori, H. Baharvand, H. Shen, I. Mateizel, In-Hyun Park, J. Sheik Mohamed, J. Downie, J. Eun Lee, J.M. Crook, J. Chen, J. Hyllner, J.-C. Biancotti, J. Baker, K. Sermon, K. Amps, K. Narwani, K. Takahashi, K. Sidhu, L. Ge, L.S. Lim, L. Young , Q. Zhou, L. Guangxiu, L.V. Pereira, L. Armstrong, M. Lako, M.S. Inamdar, M.A. Lagarkova, M.B. Munoz, M. Mileikovsky, M.V. Camarasa, M. Jaconi, M. Gropp, N. Lavon, N. Strelchenko, N. Nakatsuji, O. Kopper, O. Hovatta, O. Qi, P. Venu, P.A. De Sousa, P. Dvorak, R. Strehl, R. Suuronen, S. Kiselev, S. Yong Moon, S. Yamanaka, S. Sivarajah, S. Beil, S.L. Minger, S.K.W. Oh, S. Pells, S. Kyung Oh, S. Kimber, T. Miyazaki, T.E. Ludwig, T. Ishii, T.C. Schulz, T. Otonkoski, T. Tuuri, T. Frumkin, V. Kukharenko, V. Fox, W. Herath, Y. Jin, Y. Min Choi, Y. Ma, Y. Wu and Y. Verlinsky.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/nbt/index.html.

Published online at http://www.nature.com/nbt/index.html.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Baker, D.E. *et al.* Adaptation to culture of human embryonic stem cells and oncogenesis *in vivo*. *Nat. Biotechnol.* **25**, 207–215 (2007).
2. Draper, J.S. *et al.* Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nat. Biotechnol.* **22**, 53–54 (2004).
3. Mitalipova, M.M. *et al.* Preserving the genetic integrity of human embryonic stem cells. *Nat. Biotechnol.* **23**, 19–20 (2005).
4. Hoffman, L.M. & Carpenter, M.K. Characterization and culture of human embryonic stem cells. *Nat. Biotechnol.* **23**, 699–708 (2005).
5. Maitra, A. *et al.* Genomic alterations in cultured human embryonic stem cells. *Nat. Genet.* **37**, 1099–1103 (2005).
6. Buzzard, J.J., Gough, N.M., Crook, J.M. & Colman, A. Karyotype of human ES cells during extended culture. *Nat. Biotechnol.* **22**, 381–382, author reply 382 (2004).
7. Caisander, G. *et al.* Chromosomal integrity maintained in five human embryonic stem cell lines after prolonged in vitro culture. *Chromosome Res.* **14**, 131–137 (2006).
8. Inzunza, J. *et al.* Comparative genomic hybridization and karyotyping of human embryonic stem cells reveals the occurrence of an isodicentric X chromosome after long-term cultivation. *Mol. Hum. Reprod.* **10**, 461–466 (2004).
9. Rosler, E.S. *et al.* Long-term culture of human embryonic stem cells in feeder-free conditions. *Dev. Dyn.* **229**, 259–274 (2004).
10. Lefort, N. *et al.* Human embryonic stem cells reveal recurrent genomic instability at 20q11.21. *Nat. Biotechnol.* **26**, 1364–1366 (2008).
11. Spits, C. *et al.* Recurrent chromosomal abnormalities in human embryonic stem cells. *Nat. Biotechnol.* **26**, 1361–1363 (2008).
12. Mayshar, Y. *et al.* Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. *Cell Stem Cell* **7**, 521–531 (2010).
13. Wang, N., Trend, B., Bronson, D.L. & Fraley, E.E. Nonrandom abnormalities in chromosome 1 in human testicular cancers. *Cancer Res.* **40**, 796–802 (1980).
14. Atkin, N.B. & Baker, M.C. Specific chromosome change, i(12p), in testicular tumours? *Lancet* **320**, 1349 (1982).
15. Rodriguez, E. *et al.* Molecular cytogenetic analysis of i(12p)-negative human male germ cell tumors. *Genes Chromosom. Cancer* **8**, 230–236 (1993).
16. Skotheim, R.I. *et al.* New insights into testicular germ cell tumorigenesis from gene expression profiling. *Cancer Res.* **62**, 2359–2364 (2002).
17. Mostert, M. *et al.* Comparative genomic and in situ hybridization of germ cell tumors of the infantile testis. *Lab. Invest.* **80**, 1055–1064 (2000).
18. Schneider, D.T. *et al.* Genetic analysis of childhood germ cell tumors with comparative genomic hybridization. *Klin. Padiatr.* **213**, 204–211 (2001).
19. Looijenga, L.H. *et al.* Comparative genomic hybridization of microdissected samples from different stages in the development of a seminoma and a non-seminoma. *J. Pathol.* **191**, 187–192 (2000).
20. Longo, L., Bygrave, A., Grosveld, F.G. & Pandolfi, P.P. The chromosome make-up of mouse embryonic stem cells is predictive of somatic and germ cell chimaerism. *Transgenic Res.* **6**, 321–328 (1997).
21. Liu, X. *et al.* Trisomy eight in ES cells is a common potential problem in gene targeting and interferes with germ line transmission. *Dev. Dyn.* **209**, 85–91 (1997).
22. Zody, M.C. *et al.* DNA sequence of human chromosome 17 and analysis of rearrangement in the human lineage. *Nature* **440**, 1045–1049 (2006).
23. Werbowetski-Ogilvie, T.E. *et al.* Characterization of human embryonic stem cells with features of neoplastic progression. *Nat. Biotechnol.* **27**, 91–97 (2009).
24. Narva, E. *et al.* High-resolution DNA analysis of human embryonic stem cell lines reveals culture-induced copy number changes and loss of heterozygosity. *Nat. Biotechnol.* **28**, 371–377 (2010).
25. Allegrucci, C. *et al.* Restriction landmark genome scanning identifies culture-induced DNA methylation instability in the human embryonic stem cell epigenome. *Hum. Mol. Genet.* **16**, 1253–1268 (2007).
26. Calvanese, V. *et al.* Cancer genes hypermethylated in human embryonic stem cells. *PLoS ONE* **3**, e3294 (2008).
27. Enver, T. *et al.* Cellular differentiation hierarchies in normal and culture-adapted human embryonic stem cells. *Hum. Mol. Genet.* **14**, 3129–3140 (2005).
28. Rugg-Gunn, P.J., Ferguson-Smith, A.C. & Pedersen, R.A. Epigenetic status of human embryonic stem cells. *Nat. Genet.* **37**, 585–587 (2005).
29. Adewumi, O. *et al.* Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. *Nat. Biotechnol.* **25**, 803–816 (2007).
30. Rugg-Gunn, P.J., Ferguson-Smith, A.C. & Pedersen, R.A. Status of genomic imprinting in human embryonic stem cells as revealed by a large cohort of independently derived and maintained lines. *Hum. Mol. Genet.* **16**Spec No. 2, R243–R251 (2007).
31. Kim, K.P. *et al.* Gene-specific vulnerability to imprinting variability in human embryonic stem cell lines. *Genome Res.* **17**, 1731–1742 (2007).
32. Andrews, P.W. *et al.* The International Stem Cell Initiative: toward benchmarks for human embryonic stem cell research. *Nat. Biotechnol.* **23**, 795–797 (2005).
33. Olariu, V. *et al.* Modeling the evolution of culture-adapted human embryonic stem cells. *Stem Cell Res.* **4**, 50–56 (2010).
34. Martin, G.R. & Evans, M.J. The morphology and growth of a pluripotent teratocarcinoma cell line and its derivatives in tissue culture. *Cell* **2**, 163–172 (1974).
35. Andrews, P.W., Bronson, D.L., Benham, F., Strickland, S. & Knowles, B.B. A comparative study of eight cell lines derived from human testicular teratocarcinoma. *Int. J. Cancer* **26**, 269–280 (1980).
36. Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643–655 (2003).
37. Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642 (2003).
38. Darr, H., Mayshar, Y. & Benvenisty, N. Overexpression of NANOG in human ES cells enables feeder-free growth while inducing primitive ectoderm features. *Development* **133**, 1193–1201 (2006).
39. Korkola, J.E. *et al.* Down-regulation of stem cell genes, including those in a 200-kb gene cluster at 12p13.31, is associated with in vivo differentiation of human male germ cell tumors. *Cancer Res.* **66**, 820–827 (2006).
40. Widschwendter, M. *et al.* Epigenetic stem cell signature in cancer. *Nat. Genet.* **39**, 157–158 (2007).
41. Frazer, K.A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
42. Li, J.Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008).
43. Abdulla, M.A. *et al.* Mapping human genetic diversity in Asia. *Science* **326**, 1541–1545 (2009).
44. Pritchard, J.K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
45. Novembre, J. *et al.* Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008).
46. Laurent, L.C. *et al.* Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. *Cell Stem Cell* **8**, 106–118 (2011).
47. Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17**, 1665–1674 (2007).
48. Assou, S. *et al.* A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. *Stem Cells* **25**, 961–973 (2007).
49. Peng, J.C. *et al.* Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* **139**, 1290–1302 (2009).
50. Nottke, A., Colaiacovo, M.P. & Shi, Y. Developmental roles of the histone lysine demethylases. *Development* **136**, 879–889 (2009).
51. Morin, R.D. *et al.* Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.* **18**, 610–621 (2008).
52. Bai, H. *et al.* Bcl-xL enhances single-cell survival and expansion of human embryonic stem cells without affecting self-renewal. *Stem Cell Res. (Amst.)* (in the press).
53. Lee, T.I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006).
54. Hussein, S.M. *et al.* Copy number variation and selection during reprogramming to pluripotency. *Nature* **471**, 58–62 (2011).
55. Mosher, J.T. *et al.* Lack of population diversity in commonly used human embryonic stem-cell lines. *N. Engl. J. Med.* **362**, 183–185 (2010).
56. Laurent, L.C. *et al.* Restricted ethnic diversity in human embryonic stem cell lines. *Nat. Methods* **7**, 6–7 (2010).
57. Wu, H. *et al.* Copy number variant analysis of human embryonic stem cells. *Stem Cells* **26**, 1484–1489 (2008).
58. Chia, N.Y. *et al.* A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* **468**, 316–320 (2010).
59. Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).
60. Ardehali, R. *et al.* Overexpression of BCL2 enhances survival of human embryonic stem cells during stress and obviates the requirement for serum factors. *Proc. Natl. Acad. Sci. USA* **108**, 3282–3287 (2011).
61. Martins-Taylor, K. *et al.* Recurrent copy number variations in human induced pluripotent stem cells. *Nat. Biotechnol.* **29**, 488–491 (2011).
62. Deloukas, P. *et al.* The DNA sequence and comparative analysis of human chromosome 20. *Nature* **414**, 865–871 (2001).
63. Shaw, C.J. & Lupski, J.R. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum. Mol. Genet.* **13** Spec No 1, R57–R64 (2004).
64. Misceo, D. *et al.* Evolutionary history of chromosome 20. *Mol. Biol. Evol.* **22**, 360–366 (2005).

Katherine Amps[1], Peter W Andrews[1], George Anyfantis[2], Lyle Armstrong[2], Stuart Avery[3], Hossein Baharvand[4], Julie Baker[5], Duncan Baker[6], Maria B Munoz[7], Stephen Beil[8], Nissim Benvenisty[9], Dalit Ben-Yosef[10,11], Juan-Carlos Biancotti[12], Alexis Bosman[13], Romulo Martin Brena[14], Daniel Brison[15], Gunilla Caisander[16], María V Camarasa[17], Jieming Chen[18], Eric Chiao[5,19], Young Min Choi[20], Andre B H Choo[21], Daniel Collins[22], Alan Colman[3,23], Jeremy M Crook[3,23–26], George Q Daley[27–30], Anne Dalton[6], Paul A De Sousa[22,31], Chris Denning[7], Janet Downie[22], Petr Dvorak[32], Karen D Montgomery[33], Anis Feki[34], Angela Ford[1], Victoria Fox[8], Ana M Fraga[35], Tzvia Frumkin[10], Lin Ge[36], Paul J Gokhale[1], Tamar Golan-Lev[9], Hamid Gourabi[4], Michal Gropp[37], Lu Guangxiu[36], Ales Hampl[38,39], Katie Harron[40], Lyn Healy[41], Wishva Herath[18], Frida Holm[42], Outi Hovatta[42], Johan Hyllner[16], Maneesha S Inamdar[43], Astrid Kresentia Irwanto[18], Tetsuya Ishii[44,73], Marisa Jaconi[13], Ying Jin[45], Susan Kimber[17], Sergey Kiselev[46,47], Barbara B Knowles[3], Oded Kopper[9], Valeri Kukharenko[48], Anver Kuliev[48], Maria A Lagarkova[47], Peter W Laird[14], Majlinda Lako[2], Andrew L Laslett[49,50], Neta Lavon[12], Dong Ryul Lee[51], Jeoung Eun Lee[52], Chunliang Li[53], Linda S Lim[18], Tenneille E Ludwig[33], Yu Ma[53], Edna Maltby[6], Ileana Mateizel[54], Yoav Mayshar[9], Maria Mileikovsky[55], Stephen L Minger[56,57], Takamichi Miyazaki[58], Shin Yong Moon[20], Harry Moore[1], Christine Mummery[59], Andras Nagy[55], Norio Nakatsuji[60], Kavita Narwani[12], Steve K W Oh[21], Sun Kyung Oh[61], Cia Olson[62], Timo Otonkoski[62,63], Fei Pan[14], In-Hyun Park[64], Steve Pells[31], Martin F Pera[65], Lygia V Pereira[35], Ouyang Qi[36], Grace Selva Raj[23], Benjamin Reubinoff[37], Alan Robins[66], Paul Robson[18], Janet Rossant[67], Ghasem H Salekdeh[68], Thomas C Schulz[66], Karen Sermon[54], Jameelah Sheik Mohamed[18], Hui Shen[14], Eric Sherrer[66], Kuldip Sidhu[69], Shirani Sivarajah[23–25], Heli Skottman[70], Claudia Spits[54], Glyn N Stacey[41], Raimund Strehl[16], Nick Strelchenko[48,73], Hirofumi Suemori[58], Bowen Sun[53], Riitta Suuronen[70], Kazutoshi Takahashi[44], Timo Tuuri[62], Parvathy Venu[43], Yuri Verlinsky[48,74], Dorien Ward-van Oostwaard[59], Daniel J Weisenberger[14], Yue Wu[56,57], Shinya Yamanaka[44,60,71,72], Lorraine Young[7] & Qi Zhou[49]

[1]Centre for Stem Cell Biology, Department of Biomedical Science, The University of Sheffield, Sheffield, UK. [2]North East England Stem Cell Institute at Life, International Centre for Life, Newcastle upon Tyne, UK. [3]Institute of Medical Biology, A-STAR, Immunos, Singapore. [4]Royan Institute for Reproductive Biomedicine, Department of Genetics, Tehran, Islamic Republic of Iran. [5]Stanford University, Stanford, California, USA. [6]Sheffield Diagnostic Genetic Services, Sheffield Children's NHS Trust, Sheffield, UK. [7]Wolfson Centre for Stem Cells, Tissue Engineering & Modelling (STEM), Centre for Biomolecular Sciences, University of Nottingham, UK. [8]USC Stem Cell Core Facility, The Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research at USC, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. [9]Stem Cell Unit, Department of Genetics, Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem, Israel. [10]Racine IVF Unit, Lis Maternity Hospital, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel. [11]Department of Cell Developmental Biology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel. [12]Regenerative Medicine Institute, Cedars-Sinai Medical Institute, Los Angeles, California, USA. [13]Department of Pathology and Immunology, Faculty of Medicine, Geneva University, Geneva, Switzerland. [14]USC Epigenome Center, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. [15]Department of Reproductive Medicine, St. Marys's Hospital, Central Manchester NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, UK. [16]Cellartis AB, Goteborg, Sweden. [17]Faculty of Life Sciences, University of Manchester, Manchester, UK. [18]Genome Institute of Singapore, Singapore. [19]Hoffmann-LaRoche, Nutley, New Jersey, USA. [20]Department of Obstetrics & Gynaecology, Seoul National University College of Medicine, Seoul, Republic of Korea. [21]Bioprocessing Technology Institute, Singapore. [22]Roslin Cells Ltd., Roslin Biocentre, Roslin, Midlothian, UK. [23]Singapore Stem Cell Consortium, A-STAR, Singapore. [24]Centre for Neural Engineering, The University of Melbourne, Parkville, Australia. [25]Optics and Nanoelectronics Research Group, NICTA Victorian Research Laboratory, The University of Melbourne, Parkville, Australia. [26]Department of Surgery, St. Vincent's Hospital, The University of Melbourne, Fitzroy, Australia. [27]Stem Cell Transplantation Program, Division of Pediatric Hematology/Oncology, Manton Center for Orphan Disease Research, Howard Hughes Medical Institute, Children's Hospital Boston and Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [28]Division of Hematology, Brigham and Women's Hospital, Boston, Massachusetts, USA. [29]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, USA. [30]Harvard Stem Cell Institute, Boston, Massachusetts, USA. [31]Medical Research Council Centre for Regenerative Medicine, University of Edinburgh, Edinburgh, UK. [32]Department of Biology, Faculty of Medicine, Masaryk University, Brno, Czech Republic. [33]WiCell Research Institute, Madison, Wisconsin, USA. [34]Department of Obstetrics and Gynecology, Hopital Cantonal Fribourgois, Freibourg, Switzerland. [35]National Laboratory for Embryonic Stem Cell Research (LaNCE), Department of Genetics and Evolutionary Biology, University of São Paulo, São Paulo, Brazil. [36]Institute of Reproductive & Stem Cell Engineering, Central South University, Reproductive & Genetic Hospital CITIC-XIANGYA, Changsha, Hunan, People's Republic of China. [37]The Hadassah Human Embryonic Stem Cell Research Center, The Goldyne Savad Institute of Gene Therapy, Hadassah University Medical Center, Jerusalem, Israel. [38]Department of Histology and Embryology, Faculty of Medicine, Masaryk University, Brno, Czech Republic. [39]Institute of Experimental Medicine ASCR, Prague, Czech Republic. [40]MRC Centre of Epidemiology for Child Health, Institute of Child Health, University College London, London, UK. [41]UK Stem Cell Bank, Division of Cell Biology and Imaging, National Institute for Biological Standards and Control, South Mimms, Herts, UK. [42]Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Karolinska University Hospital Huddinge, Stockholm, Sweden. [43]Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India. [44]Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto, Japan. [45]Key Laboratory of Stem Cell Biology, Institute of Health Sciences, Shanghai Institutes of Biological Sciences, CAS/Shanghai JiaoTong University School of Medicine, Shanghai, People's Republic of China. [46]Stem Cell Department, NRC Kurchatov Institute, Moscow, Russia. [47]Vavilov Institute of General Genetics, Moscow, Russia. [48]Reproductive Genetics Institute, Chicago, Illinois, USA. [49]CSIRO Material Science and Engineering, Clayton, Australia. [50]Department of Anatomy and Developmental Biology, Monash University, Clayton, Australia. [51]Department of Biomedical Science, CHA Stem Cell Institute, CHA University, Gangnam-gu, Seoul, Republic of Korea. [52]CHA Stem Cell Institute, CHA University, Gangnam-gu, Seoul, Republic of Korea. [53]Shanghai Stem Cell Institute, Shanghai JiaoTong University School of Medicine, Shanghai, People's Republic of China. [54]Department of Embryology and Genetics, Vrije Universiteit Brussel, Brussels, Belgium. [55]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. [56]Wolfson Centre for Age-Related Diseases, King's College London, London, UK. [57]GE Healthcare, Cardiff, UK. [58]Laboratory of Embryonic Stem Cell Research, Stem Cell Research Center, Institute for Frontier Medical Sciences, Kyoto University, Kyoto, Japan. [59]Department of Anatomy & Embryology, Leiden University Medical Center, Leiden, The Netherlands. [60]Institute for Integrated Cell-Material Sciences, Kyoto University, Ushinomiya-cho, Yoshida, Sakyo-ku, Kyoto, Japan. [61]Institute of Reproductive Medicine & Population, Medical Research Center, Seoul National University, Seoul, Republic of Korea. [62]Research Programs Unit, Molecular Neurology, Biomedicum Stem Cell Centre, University of Helsinki, Finland. [63]Children's Hospital, University of Helsinki and Helsinki University Central Hospital, Finland. [64]Department of

Genetics, Yale Stem Cell Center, Yale School of Medicine, New Haven, Connecticut, USA. [65]The Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research at USC, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. [66]Viacyte, Athens, Georgia, USA. [67]Program for Developmental Biology, The Hospital for Sick Children, Toronto, Ontario, Canada. [68]Department of Molecular Systems Biology, Cell Science Research Center, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Islamic Republic of Iran. [69]Stem Cell Laboratory, Faculty of Medicine, University of New South Wales, Australia. [70]Institute for Regenerative Medicine, University of Tampere, Tampere, Finland. [71]Yamanaka iPS Cell Special Project, Japan Science and Technology Agency, Kawaguchi, Japan. [72]Gladstone Institute of Cardiovascular Disease, San Francisco, California, USA. [73]Present addresses: Japan Science and Technology Agency, Tokyo, Japan (T. Ishii) and Department of Obstetrics and Gynecology, New York University Langone Medical Center, New York, New York, USA (N. Strelchenko). [74]Deceased.

ONLINE METHODS

Design of study. Laboratories that agreed to contribute cell lines to the study were asked to provide four preparations of DNA from a culture of each cell line to be analyzed, two preparations from as early a passage as possible and two from as late a passage as possible (**Supplementary Note 1**, protocol for DNA preparation). The samples were coded and shipped directly, one pair of early- and late-passage DNA for SNP analysis to the Genome Institute, A*STAR, Singapore, and one pair of samples for DNA methylation analysis to the University of Southern California. The laboratories were also asked to prepare fixed hypotonic-treated samples of cells at both passage levels, following culture in colcemid, according to a specified protocol (**Supplementary Note 2**). These samples were shipped to the University of Sheffield for preparation and karyotypic analysis of metaphase spreads. Ideally the laboratories were asked to provide the material for both DNA analyses and the cytogenetic analysis from the same cultures of the cells. In a few cases this could not be done, for example, when a first sample did not meet quality control criteria and additional samples had to be prepared, but all samples were provided from as close a passage as possible. The full details of each cell line included in the study, the passage levels of each sample and culture details, including split ratios, culture time between passages, and medium and subculture methodology, are provided in **Supplementary Table 1**.

Karyology. Chromosome analysis was carried out in accord with the general principles developed by the International Stem Cell Banking Initiative[65]. Briefly, cells were to be cultured in the presence of 0.1 μg/ml colcemid (Invitrogen) for up to 4 h, followed by dissociation with 0.25% trypsin/versene (Invitrogen). The cells were pelleted by centrifugation and resuspended in prewarmed 0.0375 M KCl hypotonic solution and incubated for 10 min. After centrifugation the cells were resuspended in fixative (3:1 methanol/acetic acid) and sent to the central karyotyping facility, where metaphase spreads were prepared on glass microscope slides and G-banded by brief exposure to trypsin and stained with 4:1 Gurr's/Leishmann's stain (Sigma-Aldrich). (See **Supplementary Note 2** for detailed protocols.) A minimum of 10 metaphase spreads were analyzed in detail and a further 20 counted and scored from both the early- and late-passage cultures. A 30-cell examination can exclude mosaicism at the 10% level with 95% confidence[66]. Analysis was performed by a Health Professionals Council registered Clinical Cytogeneticist in a CPA accredited laboratory (Duncan Baker DipRCPath, Sheffield Diagnostic Genetic Services, Sheffield Children's NHS Foundation Trust and the CSCB, University of Sheffield). A representative image of each cell line was captured using Applied Imaging's Cytovision system. All abnormal karyotype results were confirmed by two other experienced analysts (Tamar Golan-Lev, Hebrew University, Jerusalem, and Karen Montgomery, WiCell).

Cell lines were described as abnormal if at least two cells were found with the same chromosome aberration. Fluorescent *in situ* hybridization (FISH) was used to characterize abnormalities when appropriate. Abnormalities confined to a single cell were recorded (**Supplementary Table 2**). For those in which the single cell abnormality involved chromosomes 1, 12, 17 or 20, interphase FISH analysis with appropriate probes (**Supplementary Table 2**) was performed to confirm or exclude low level mosaicism. FISH analysis was also performed, on the early-passage cultures of those cell lines with normal early cultures and abnormal late cultures to confirm or exclude low-level mosaicism in the early passage.

SNP array analysis. To analyze genomic structural variations below the resolution of standard chromosomal banding analysis, we chose the Illumina 1M Quad SNP array technology. In addition to providing the ability to detect structural variants by virtue of appropriately spaced invariant genetic features, the SNP features allowed characterization of the population structure between all human ES cell lines analyzed. The samples received were run on the Illumina 1M Quad platform and the data were subjected to structural variant quality control (QC) assessment in which the minimum number SNP/structural variant = 10, the minimum length of structural variant = 1 kb, the minimum confidence score for structural variant call = 10, the Log R Ratio s.d. < 0.35, the B allele freq s.d. < 0.06 and the waviness factor −0.04 < WF < 0.04.

PLINK[67] was used to detect genetically related samples (early- and late-passage paired samples and other sibling relationships) and the related samples with the lower call-rate were removed. QC was performed on these samples, by excluding structural variant probes, nonautosomal SNPs, and samples and SNPs with call-rates <95%. Thereafter, 982,351 SNPs from 114 unrelated samples were merged with three publicly available data sets: HGDP, Hapmap Phase 2 (release 23) and the Pan-Asian SNP Initiative (PASNPI). Additional QC was then performed on the merged data set by exclusion of samples and SNPs with call-rates <95% and monomorphic SNPs. Furthermore, we excluded the extremely diverse southeast-Asian samples in PASNPI, to focus the analyses on human ES cell origins in Africa, Europe, Central-south Asia and East Asia. Ultimately, the merged data set contained 1,967 samples and 11,279 SNPs. Subsequently, a series of PC analyses (PCAs) were performed. The PCAs were performed using smartpca from the software EIGENSTRAT[68] (found in EIGENSOFT, http://genepath.med.harvard.edu/~reich/Software.htm).

The first PCA was performed on the entire final merged data set. The scatter plot of PC2 against PC1 of this PCA was then dissected into four sectors, based primarily on the global regions of Africa (PC2 < −0.04), Middle East-Europe-Central-South Asia (PC1 > 0.005 and PC2 > 0.008), Oceania-America-Central-South Asia (−0.015 < PC1 < 0.005) and East Asia (PC1 < −0.015). (**Fig. 1a**). PCAs were then reperformed on the samples found in these four sectors. Notably, most of the human ES cell samples were concentrated in the European sector (with the HGDP-Europe and HapMap European cluster).

Hence, to further elucidate the origins of the human ES cell lines in Europe, we removed 16 human ES cell European outliers in the Central-south Asia-Europe-Middle East sector (**Fig. 1d**); these are considered to be of dubious ancestry, suspected to originate from two groups: Central-south Asian or Southern European or mixed from both; Middle Eastern or Eastern European or mixed from both. Lack of reference populations and insufficient SNPs could be possible reasons for this uncertainty in determination of ancestry. The remaining human ES cell samples (found in the European cluster) were then merged with 1,385 samples and 168,352 SNPs from the POPRES European data set[45]. (The sample and SNP lists were obtained through correspondence with the author.) This second merged data set between human ES cell and POPRES European data yielded 1,448 samples and 55,972 SNPs, after excluding samples and SNPs with call rates <95% and monomorphic SNPs. PCA was then performed again on this data set (**Fig. 1b**). Color coding was deliberately implemented in a similar manner to that of reference [45].

The human ES cell lines were classified into broad categories of European, Asian, African and other ancestries and then further subclassified into ethnicities listed in **Table 1**. Human ES cell samples that are full siblings of the 114 samples were included, resulting in a total of 120 human ES cell lines.

Frequency and mapping analysis of SNP data. The PennCNV[47] Hidden-Markov model algorithm was used to identify structural variants in the ISCI samples with human hg18 as the reference genome. High-quality structural variant calls were filtered as follows: first, samples were checked for overall quality using the following criteria from the PennCNV output: 0.01 < BAF_drift < 0.01; −0.05 < WF < 0.05; LRR_SD < 0.35. Second, individual structural variant calls for samples passing QC were assessed as follows: minimum SNPs per structural variant > 10; structural variant Length > 1K; PennCNV confidence threshold > 10. Most samples that failed quality control either exhibited extensive karyotypic abnormalities and/or bad SNP call rates, both of which could contribute to difficulties in structural variant detection. The samples that did pass quality control identified 39,926 deletions of an average size of 23.1 kb and 14,351 duplications of an average size of 117.4 kb in length. These size and total number differences between deletions and duplications is consistent with previous structural variant studies of human populations[47]. The sensitivity of detection was in the 5–10% range based on CNV calls in regions identified as amplified by karyotype analysis.

DNA methylation analysis. 1 μg of genomic DNA was treated with sodium bisulfite using the Zymo EZ-96 DNA Methylation Kit (Zymo Research) following the manufacturer's recommended protocol. DNA methylation measurements were generated using the Illumina GoldenGate DNA methylation platform as previously described[69], at the University of Southern California Epigenome Center production facility. The sequences assayed were previously identified[53] as targets of SUZ12, a subunit of the Polycomb Repressive Complex 2, in the nonrepetitive portion of the genome in human

doi:10.1038/nbt.2051

ES cells (**Supplementary Table 4** array loci, sequences, etc). DNA methylation measurements (β-values) were generated for 1,452 autosome and 84 X-linked loci.

To obtain a global view of the DNA methylation changes undergone by each cell line pair from the early to the late passage, CDF curves were generated. CDF curves were chosen because they are able to concomitantly capture and visually represent the changes detected in all loci analyzed in a single line. To test whether increased passage number correlated with increase in the number of loci gaining DNA methylation, we calculated the difference in passage number between the early and the late member of each cell line pair, ranked these differences and divided the samples into quartiles based on passage-number difference. To investigate whether sibling lines behaved similarly in terms of DNA methylation changes, we calculated the number of loci with an absolute DNA methylation difference ≥10% between the early and late member of each pair of sibling lines. We then compared whether the total number of loci with the specified absolute change in DNA methylation differed significantly from the number of loci undergoing a similar change in pairs of unrelated lines using a Wilcoxon signed-rank test. The data were analyzed using the R statistical programming package version 2.12.1 (http://www.r-project.org/).

Cell line availability. Supplementary Table 5 provides contact information and conditions of availability for the cell lines described in this study.

65. The International Stem Cell Banking Initiative. Consensus guidance for banking and supply of human embryonic stem cell lines for research purposes. *Stem Cell Rev.* **5**, 301–314 (2009).
66. Hook, E.B. Exclusion of chromosomal mosaicism: tables of 90%, 95% and 99% confidence limits and comments on use. *Am. J. Hum. Genet.* **29**, 94–97 (1977).
67. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
68. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
69. Bibikova, M. *et al.* High-throughput DNA methylation profiling using universal bead arrays. *Genome Res.* **16**, 383–393 (2006).