Published in final edited form as:
Nat Biotechnol. ; 29(12): 1132–1144. doi:10.1038/nbt.2051.

Screening a large, ethnically diverse population of human embryonic stem cells identifies a chromosome 20 minimal amplicon that confers a growth advantage

The International Stem Cell Initiative

Abstract

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, *ID1*, *BCL2L1 and HM13,* expressed in human ES cells, 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,

[28]Singapore Stem Cell Bank, A-STAR, Singapore.

[29]Centre for Neural Engineering, The University of Melbourne, Parkville, Australia.

[30]Department of Surgery, St Vincent's Hospital, The University of Melbourne, Fitzroy, Australia.

[33]Division of Hematology, Brigham and Women's Hospital, Boston, Massachusetts, USA.

[34]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, USA.

[126]Centre for Neural Engineering, The University of Melbourne, Parkville, Australia.

[145]Institute for Integrated Cell-Material Sciences, Kyoto University, Kyoto, Japan.

[146]Yamanaka iPS Cell Special Project, Japan Science and Technology Agency, Kawaguchi, Japan.

Correspondence should be addressed to P.W.A. (p.w.andrews@sheffield.ac.uk)

Supplementary information is available on the Nature Biotechnology website.

AUTHOR CONTRIBUTIONS

Project Co-ordination: PWA. Cytogenetic analyses: DB, KMD and TG-L. Molecular karyotyping by SNP BeadChip: PR. DNA methylation arrays: RMB and PWL. Administration and data curation: AF and PJG. Data analysis and manuscript drafting: PWA, DB, NB, RMB, PJG, BBK, YM, SKWO, MFP and PR. The scientific management of the ISCI project was provided by a steering committee comprising: PWA, NB, BBK, SKWO, MFP, JR and GNS.

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.

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 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.

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

Independence, 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 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, 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.

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 40 lines submitted from the remaining 12 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, 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 Information 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 Information 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 Information 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 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 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 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 in the regions of 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 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 five 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 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 Information 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.

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 Information 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 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, 69%) 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 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 gains 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/

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The International Stem Cell Initiative is funded by The International Stem Cell Forum.

The authors would to acknowledge the following, Medical Research Council, UK. (PWA); Mohammad Pakzad & Adeleh Taei. Royan Institute (HB; -GHS); California Instutute of Regenerative Medicine. (EC); Institute of Medical Biology, A*STAR, Singapore. (JMC); Ministry of Education, Youth and Sports of the Czech Republic (PD and AH); Stem Cell Research Center of the 21st Century Frontier Research Program, Ministry of Education, Science & Technology, Republic of Korea (SC-1140). (DRL); Ministry of Science and Technology of China (863 program 2006AA02A102). (LG); Swedish Research Council, Cellartis. (OH); Department of Biotechnology, Government of India, UK-India Education and Research Initiative and the Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India. (MI); 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. (TI, SY, KT); Swiss National Science Foundation (grant # 4046-114410). (MJ); 06DJ14001 (Shanghai Science and Technology Developmental Foundation), 2007CB948004 (Chinese Ministry of Science of Science and Technology) (YJ); Funding from the North West Science Fund, UK.(SK); California Institute for Regenerative Medicine (CIRM),(PWL); One North East Regional Developmental Agency, Medical Research Council, UK., Newcastle University. (ML); Research funding from the Australian Stem Cell Centre. (AL); Medical Research Council, UK. (HM); The Netherlands Proteomics Consortium grant T4-3. (CM); National BioResource Project, MEXT, Japan. (NN); Singapore Stem Cell Consortium (SSCC) & the Agency for Science Technology and Research (A*STAR). (SKWO); Singapore Stem Cell Consortium (SSCC) & the Agency for Science Technology and Research (A*STAR), Genome Institute of Singapore Core Genotyping Lab.(PR); Academy of Finland, Sigrid Juselius Foundation. (TO); 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). (LVP); Supported by the kind donation of Judy and Sidney Swartz. (BR); Financial support from the Faculty of Medicine, University of New South Wales (UNSW) and the National Health and Medical Research Council (NHMRC) Program Grant # 568969 (Prof Perminder Sachdev), South Eastern Sydney and Illawarra Area Health Service (SEIAHS) for making hESC line, Endeavour-2 available for this study & Henry Chung, Jaemin kim for their help in preparing the samples. (KS); Regional Ministry of Health (GVA), Institute of Health Carlos III (ISCIII), Spanish Ministry of Science and Innovation. (CSV); Academy of Finland (grant: 218050), the Competitive Research Funding of the Tampere University Hospital (grant: 9F217). (HS); Stem Cell Research Center of the 21st Century Frontier Research Program, Ministry of Education, Science & Technology, Republic of Korea (SC-1150). (SKO).

References

1. Baker DE, et al. Adaptation to culture of human embryonic stem cells and oncogenesis in vivo. Nat Biotechnol. 2007; 25:207–215. [PubMed: 17287758]
2. Draper JS, et al. Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. Nat Biotechnol. 2004; 22:53–54. [PubMed: 14661028]
3. Mitalipova MM, et al. Preserving the genetic integrity of human embryonic stem cells. Nat Biotechnol. 2005; 23:19–20. [PubMed: 15637610]
4. Hoffman LM, Carpenter MK. Characterization and culture of human embryonic stem cells. Nat Biotechnol. 2005; 23:699–708. [PubMed: 15940242]
5. Maitra A, et al. Genomic alterations in cultured human embryonic stem cells. Nat Genet. 2005; 37:1099–1103. [PubMed: 16142235]
6. Buzzard JJ, Gough NM, Crook JM, Colman A. Karyotype of human ES cells during extended culture. Nat Biotechnol. 2004; 22:381–382. author reply 382. [PubMed: 15060545]
7. Caisander G, et al. Chromosomal integrity maintained in five human embryonic stem cell lines after prolonged in vitro culture. Chromosome Res. 2006; 14:131–137. [PubMed: 16544187]
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. 2004; 10:461–466. [PubMed: 15044603]

9. Rosler ES, et al. Long-term culture of human embryonic stem cells in feeder-free conditions. Dev Dyn. 2004; 229:259–274. [PubMed: 14745951]
10. Lefort N, et al. Human embryonic stem cells reveal recurrent genomic instability at 20q11.21. Nat Biotechnol. 2008; 26:1364–1366. [PubMed: 19029913]
11. Spits C, et al. Recurrent chromosomal abnormalities in human embryonic stem cells. Nat Biotechnol. 2008; 26:1361–1363. [PubMed: 19029912]
12. Mayshar Y, et al. Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. Cell Stem Cell. 2010; 7:521–531. [PubMed: 20887957]
13. Wang N, Trend B, Bronson DL, Fraley EE. Nonrandom abnormalities in chromosome 1 in human testicular cancers. Cancer Res. 1980; 40:796–802. [PubMed: 7471097]
14. Atkin NB, Baker MC. Specific chromosome change, i(12p), in testicular tumours? Lancet. 1982; 320:1349. [PubMed: 6128640]
15. Rodriguez E, et al. Molecular cytogenetic analysis of i(12p)-negative human male germ cell tumors. Genes Chromosom Cancer. 1993; 8:230–236. [PubMed: 7512366]
16. Skotheim RI, et al. New insights into testicular germ cell tumorigenesis from gene expression profiling. Cancer Res. 2002; 62:2359–2364. [PubMed: 11956097]
17. Mostert M, et al. Comparative genomic and in situ hybridization of germ cell tumors of the infantile testis. Lab Invest. 2000; 80:1055–1064. [PubMed: 10908150]
18. Schneider DT, et al. Genetic analysis of childhood germ cell tumors with comparative genomic hybridization. Klin Padiatr. 2001; 213:204–211. [PubMed: 11528555]
19. Looijenga LH, et al. Comparative genomic hybridization of microdissected samples from different stages in the development of a seminoma and a non-seminoma. J Pathol. 2000; 191:187–192. [PubMed: 10861580]
20. Longo L, Bygrave A, Grosveld FG, Pandolfi PP. The chromosome makeup of mouse embryonic stem cells is predictive of somatic and germ cell chimaerism. Transgenic Res. 1997; 6:321–328. [PubMed: 9322369]
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. 1997; 209:85–91. [PubMed: 9142498]
22. Zody MC, et al. DNA sequence of human chromosome 17 and analysis of rearrangement in the human lineage. Nature. 2006; 440:1045–1049. [PubMed: 16625196]
23. Werbowetski-Ogilvie TE, et al. Characterization of human embryonic stem cells with features of neoplastic progression. Nat Biotechnol. 2009; 27:91–97. [PubMed: 19122652]
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. 2010; 28:371–377. [PubMed: 20351689]
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. 2007; 16:1253–1268. [PubMed: 17409196]
26. Calvanese V, et al. Cancer genes hypermethylated in human embryonic stem cells. PLoS ONE. 2008; 3:e3294. [PubMed: 18820729]
27. Enver T, et al. Cellular differentiation hierarchies in normal and culture-adapted human embryonic stem cells. Hum Mol Genet. 2005; 14:3129–3140. [PubMed: 16159889]
28. Rugg-Gunn PJ, Ferguson-Smith AC, Pedersen RA. Epigenetic status of human embryonic stem cells. Nat Genet. 2005; 37:585–587. [PubMed: 15864307]
29. Adewumi O, et al. Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. Nat Biotechnol. 2007; 25:803–816. [PubMed: 17572666]
30. Rugg-Gunn PJ, Ferguson-Smith AC, Pedersen RA. Status of genomic imprinting in human embryonic stem cells as revealed by a large cohort of independently derived and maintained lines. Hum Mol Genet. 2007; 16(Spec No 2):R243–R251. [PubMed: 17911167]
31. Kim KP, et al. Gene-specific vulnerability to imprinting variability in human embryonic stem cell lines. Genome Res. 2007; 17:1731–1742. [PubMed: 17989250]
32. Andrews PW, et al. The International Stem Cell Initiative: toward benchmarks for human embryonic stem cell research. Nat Biotechnol. 2005; 23:795–797. [PubMed: 16003358]

33. Olariu V, et al. Modeling the evolution of culture-adapted human embryonic stem cells. Stem Cell Res. 2010; 4:50–56. [PubMed: 19837641]
34. Martin GR, Evans MJ. The morphology and growth of a pluripotent teratocarcinoma cell line and its derivatives in tissue culture. Cell. 1974; 2:163–172. [PubMed: 4416368]
35. Andrews PW, Bronson DL, Benham F, Strickland S, Knowles BB. A comparative study of eight cell lines derived from human testicular teratocarcinoma. Int J Cancer. 1980; 26:269–280. [PubMed: 6169654]
36. Chambers I, et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell. 2003; 113:643–655. [PubMed: 12787505]
37. Mitsui K, et al. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell. 2003; 113:631–642. [PubMed: 12787504]
38. Darr H, Mayshar Y, Benvenisty N. Overexpression of NANOG in human ES cells enables feeder-free growth while inducing primitive ectoderm features. Development. 2006; 133:1193–1201. [PubMed: 16501172]
39. Korkola JE, 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. 2006; 66:820–827. [PubMed: 16424014]
40. Widschwendter M, et al. Epigenetic stem cell signature in cancer. Nat Genet. 2007; 39:157–158. [PubMed: 17200673]
41. Frazer KA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007; 449:851–861. [PubMed: 17943122]
42. Li JZ, et al. Worldwide human relationships inferred from genome-wide patterns of variation. Science. 2008; 319:1100–1104. [PubMed: 18292342]
43. Abdulla MA, et al. Mapping human genetic diversity in Asia. Science. 2009; 326:1541–1545. [PubMed: 20007900]
44. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155:945–959. [PubMed: 10835412]
45. Novembre J, et al. Genes mirror geography within Europe. Nature. 2008; 456:98–101. [PubMed: 18758442]
46. Laurent LC, 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. 2011; 8:106–118. [PubMed: 21211785]
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. 2007; 17:1665–1674. [PubMed: 17921354]
48. Assou S, et al. A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. Stem Cells. 2007; 25:961–973. [PubMed: 17204602]
49. Peng JC, et al. Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell. 2009; 139:1290–1302. [PubMed: 20064375]
50. Nottke A, Colaiacovo MP, Shi Y. Developmental roles of the histone lysine demethylases. Development. 2009; 136:879–889. [PubMed: 19234061]
51. Morin RD, et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res. 2008; 18:610–621. [PubMed: 18285502]
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). 2011 (in the press).
53. Lee TI, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. Cell. 2006; 125:301–313. [PubMed: 16630818]
54. Hussein SM, et al. Copy number variation and selection during reprogramming to pluripotency. Nature. 2011; 471:58–62. [PubMed: 21368824]
55. Mosher JT, et al. Lack of population diversity in commonly used human embryonic stem-cell lines. N Engl J Med. 2010; 362:183–185. [PubMed: 20018958]
56. Laurent LC, et al. Restricted ethnic diversity in human embryonic stem cell lines. Nat Methods. 2010; 7:6–7. [PubMed: 20038950]

57. Wu H, et al. Copy number variant analysis of human embryonic stem cells. Stem Cells. 2008; 26:1484–1489. [PubMed: 18369100]

58. Chia NY, et al. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature. 2010; 468:316–320. [PubMed: 20953172]

59. Beroukhim R, et al. The landscape of somatic copy-number alteration across human cancers. Nature. 2010; 463:899–905. [PubMed: 20164920]

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. 2011; 108:3282–3287. [PubMed: 21300885]

61. Martins-Taylor K, et al. Recurrent copy number variations in human induced pluripotent stem cells. Nat Biotechnol. 2011; 29:488–491. [PubMed: 21654665]

62. Deloukas P, et al. The DNA sequence and comparative analysis of human chromosome 20. Nature. 2001; 414:865–871. [PubMed: 11780052]

63. Shaw CJ, Lupski JR. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Hum Mol Genet. 2004; 13(Spec No 1):R57–R64. [PubMed: 14764619]

64. Misceo D, et al. Evolutionary history of chromosome 20. Mol Biol Evol. 2005; 22:360–366. [PubMed: 15496555]

AUTHORS

Katherine Amps[1], Peter W Andrews[2], George Anyfantis[3], Lyle Armstrong[4], Stuart Avery[5], Hossein Baharvand[6], Julie Baker[7], Duncan Baker[8], Maria Barbadillo Munoz[9], Stephen Beil[10], Nissim Benvenisty[11], Dalit Ben-Yosef[12], Juan-Carlos Biancotti[13], Alexis Bosman[14], Romulo Martin Brena[15], Daniel Brison[16], Gunilla Caisander[17], María Vicenta Camarasa[18], Jieming Chen[19], Eric Chiao[20,21], Young Min Choi[22], Andre B H Choo[23], Daniel Collins[24], Alan Colman[25,26], Jeremy Micah Crook[27–31], George Q Daley[32–35],

[1]Centre for Stem Cell Biology, Department of Biomedical Science, The University of Sheffield, Sheffield, UK.
[2]Centre for Stem Cell Biology, Department of Biomedical Science, The University of Sheffield, Sheffield, UK.
[3]Newcastle University, Institute of Genetic Medicine, International Centre for Life, Central Parkway, Newcastle, UK.
[4]Newcastle University, Institute of Genetic Medicine, International Centre for Life, Central Parkway, Newcastle, UK.
[5]Institute of Medical Biology, Singapore.
[6]Royan, Department of Stem Cells and Developmental Biology, Cell Science Research Center, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Islamic Republic of Iran.
[7]Stanford University, Stanford, California, USA.
[8]Sheffield Diagnostic Genetic Services, Sheffield Children's NHS Trust, Sheffield, UK.
[9]Wolfson Centre for Stem Cells, Tissue Engineering & Modelling (STEM), Centre for Biomolecular Sciences, University of Nottingham, University Park, UK.
[10]USC Stem Cell Core Facility, Eli and Edyth Broad Center for Regenerative Medicine and Stem Cell Research, University of Southern California, Keck School of Medicine, Los Angeles, California, USA.
[11]Stem Cell Unit, Department of Genetics, The Institute of Life Sciences, Edmond J. Safra Campus, Givat Ram, The Hebrew University, Jerusalem, Israel.
[12]Racine IVF Unit, Lis Maternity Hospital, Tel Aviv Sourasky Medical Center, Department of Cell & Developmental Biology, Sackler Faculty of Medicine, Tel Aviv University, Israel.
[13]Regenerative Medicine Institute, Cedars-Sinai Medical Institute, Los Angeles, California, USA.
[14]Department of Pathology and Immunology, Faculty of Medicine, Geneva University, Geneva, Switzerland.
[15]USC Epigenome Center, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.
[16]Department of Reproductive Medicine, St Marys's Hospital, Central Manchester NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, UK.
[17]Cellartis AB, Goteborg, Sweden.
[18]Faculty of Life Sciences, University of Manchester, Manchester, UK.
[19]Genome Institute of Singapore, Singapore.
[20]Stanford University, Stanford, USA.
[21]Hoffmann-LaRoche, Nutley, New Jersey, USA.
[22]Department of Obstetrics & Gynaecology, Seoul National University College of Medicine, Seoul, Republic of Korea.
[23]Bioprocessing Technology Institute, Singapore.
[24]Roslin Cells Ltd, Roslin Biocentre, Roslin, Midlothian, UK.
[25]Institute of Medical Biology, A-STAR, Immunos, Singapore.
[26]Singapore Stem Cell Bank, A-STAR, Singapore.
[27]Institute of Medical Biology, A-STAR, Singapore.

Anne Dalton[36], Paul A De Sousa[37,38], Chris Denning[39], Janet Downie[40], Petr Dvorak[41], Karen Dyer Montgomery[42], Anis Feki[43], Angela Ford[44], Victoria Fox[45], Ana M Fraga[46], Tzvia Frumkin[47], Amparo Galán[48], Lin Ge[49], Paul J Gokhale[50], Tamar Golan-Lev[51], Hamid Gourabi[52], Michal Gropp[53], Lu Guangxiu[54], Ales Hampl[55,56], Katie Harron[57], Lyn Healy[58], Wishva Herath[59], Frida Holm[60], Outi Hovatta[61], Johan Hyllner[62], Maneesha S Inamdar[63], Astrid Kresentia Irwanto[64], Tetsuya Ishii[65,150], Marisa Jaconi[66], Ying Jin[67], Susan Kimber[68], Sergey Kiselev[69,70], Barbara B Knowles[71], Oded Kopper[72], Valeri Kukharenko[73], Anver Kuliev[74], Maria A Lagarkova[75], Peter W Laird[76], Majlinda Lako[77],

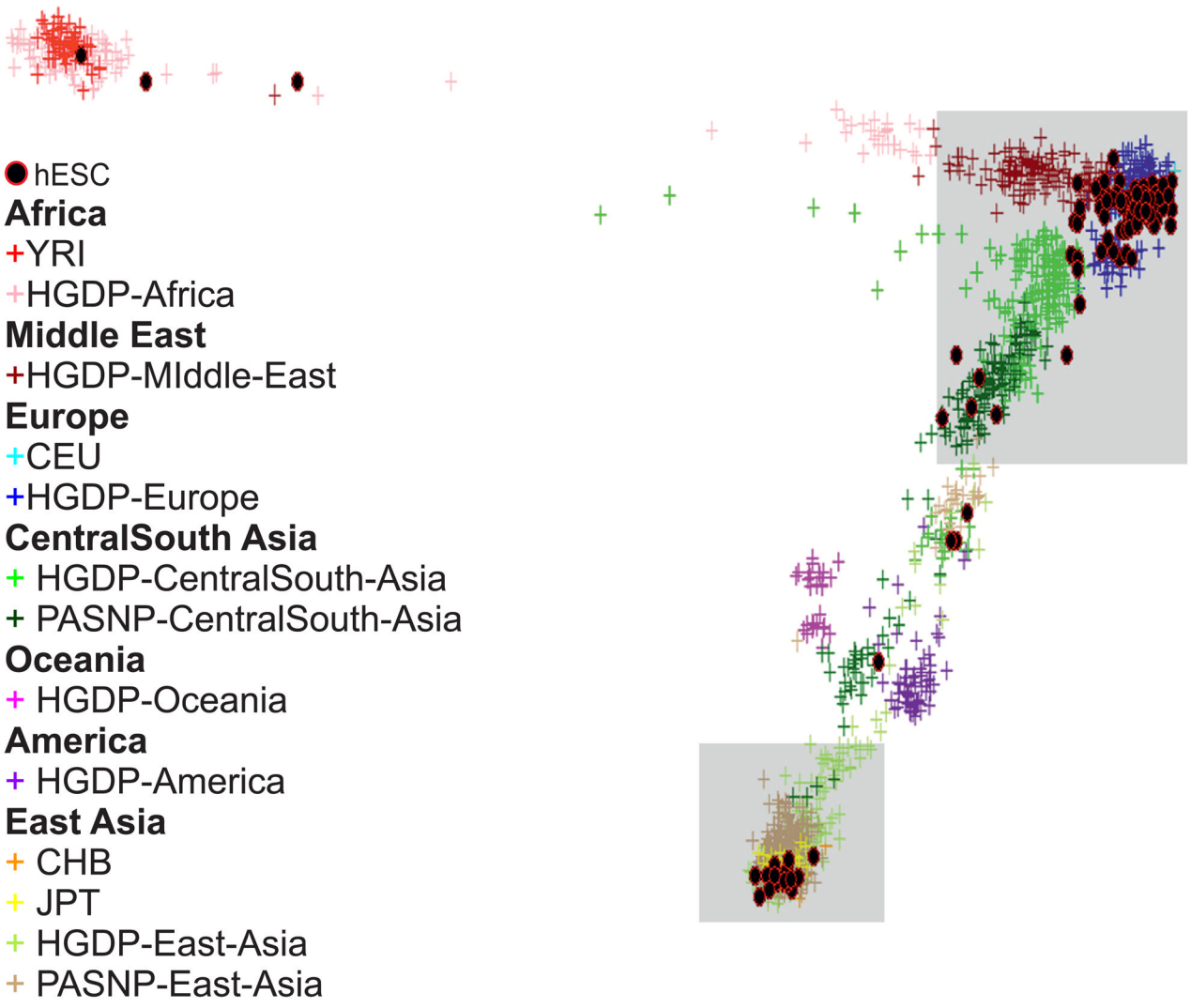
[31]Optics and Nanoelectronics Research Group, NICTA Victorian Research Laboratory, The University of Melbourne, Parkville, Australia.
[32]Stem Cell Transplantation Program, Division of Pediatric Hematology/Oncology, Manton Center for Orphan Disease Research, Howard Hughes Medical Institute, Children's Hospital Boston and Dana Farber Cancer Institute, Boston, Massachusetts, USA.
[35]Harvard Stem Cell Institute, Boston, Massachusetts, USA.
[36]Sheffield Diagnostic Genetic Services, Sheffield Children's NHS Trust, Sheffield, UK.
[37]Medical Research Council Centre for Regenerative Medicine, University of Edinburgh, Edinburgh, UK.
[38]Roslin Cells Ltd., Roslin Biocentre, Roslin, Midlothian, UK.
[39]Wolfson Centre for Stem Cells, Tissue Engineering & Modelling (STEM), Centre for Biomolecular Sciences, University of Nottingham, University Park, UK.
[40]Roslin Cells Ltd, Roslin Biocentre, Roslin, Midlothian, UK.
[41]Department of Biology, Faculty of Medicine, Masaryk University, Brno, Czech Republic.
[42]WiCell Research Institute, Madison, Wisconsin, USA.
[43]Department of Obstetrics and Gynecology, Hôpital Cantonal Fribourgois, Freibourg, Switzerland.
[44]Centre for Stem Cell Biology, Department of Biomedical Science, The University of Sheffield, Sheffield, UK.
[45]USC Stem Cell Core Facility, Eli and Edyth Broad Center for Regenerative Medicine and Stem Cell Research, University of Southern California, Keck School of Medicine, Los Angeles, California, USA.
[46]Laboratório Nacional de Células-Tronco Embrionárias (LaNCE), Depto. Genética e Biologia Evolutiva, Universidade de São Paulo, São Paulo, Brazil.
[47]Racine IVF Unit, Lis Maternity Hospital, Tel Aviv Sourasky Medical Center, Department of Cell & Developmental Biology, Sackler Faculty of Medicine, Tel Aviv University, Israel.
[48]Valencia Node of the Spanish Stem Cell Bank, Prince Felipe Research Center, Valencia, Spain.
[49]Institute of Reproductive & Stem Cell Engineering Central South University, Reproductive & Genetic Hospital CITIC-XIANGYA, Changsha, Hunan, People's Republic of China.
[50]Centre for Stem Cell Biology, Department of Biomedical Science, The University of Sheffield, Sheffield, UK.
[51]Stem Cell Unit, Department of Genetics, Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel.
[52]Royan Institute for Reproductive Biomedicine, Department of Genetics, Tehran, Islamic Republic of Iran.
[53]The Hadassah Human Embryonic Stem Cell Research Center, The Goldyne Savad Institute of Gene Therapy, Hadassah University Medical Center, Jerusalem, Israel.
[54]Institute of Reproductive & Stem Cell Engineering Central South University, Reproductive & Genetic Hospital CITIC-XIANGYA, Changsha, Hunan, People's Republic of China.
[55]Department of Histology and Embryology, Faculty of Medicine, Masaryk University, Brno, Czech Republic.
[56]Institute of Experimental Medicine ASCR, Prague, Czech Republic.
[57]MRC Centre of Epidemiology for Child Health, Institute of Child Health, University College London, London, UK.
[58]UK Stem Cell Bank, Division of Cell Biology and Imaging, National Institute for Biological Standards and Control, South Mimms, Herts., UK.
[59]Genome Institute of Singapore, Singapore.
[60]Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Karolinska University Hospital Huddinge, Stockholm, Sweden.
[61]Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Karolinska University Hospital Huddinge, Stockholm, Sweden.
[62]Cellartis AB, Goteborg, Sweden.
[63]Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India.
[64]Genome Institute of Singapore, Singapore.
[65]Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto, Japan.
[150]Present address: Department of Obstetrics and Gynecology, New York, New York, USA.
[66]Department of Pathology and Immunology, Faculty of Medicine, Geneva University, Geneva, Switzerland.
[67]Key Laboratory of Stem Cell Biology, Institute of Health Sciences, Shanghai Institutes of Biological Sciences, CAS/Shanghai JiaoTong University School of Medicine, Shanghai, China.
[68]Faculty of Life Sciences, University of Manchester, Manchester, UK.
[69]Vavilov Institute of General Genetics, Moscow, Russia.
[70]Stem Cell Department, NRC Kurchatov Institute, Moscow, Russia.
[71]Institute of Medical Biology, Singapore.
[72]Stem Cell Unit, Department of Genetics, Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem, Israel.
[73]Reproductive Genetics Institute, Chicago Illinois, USA.

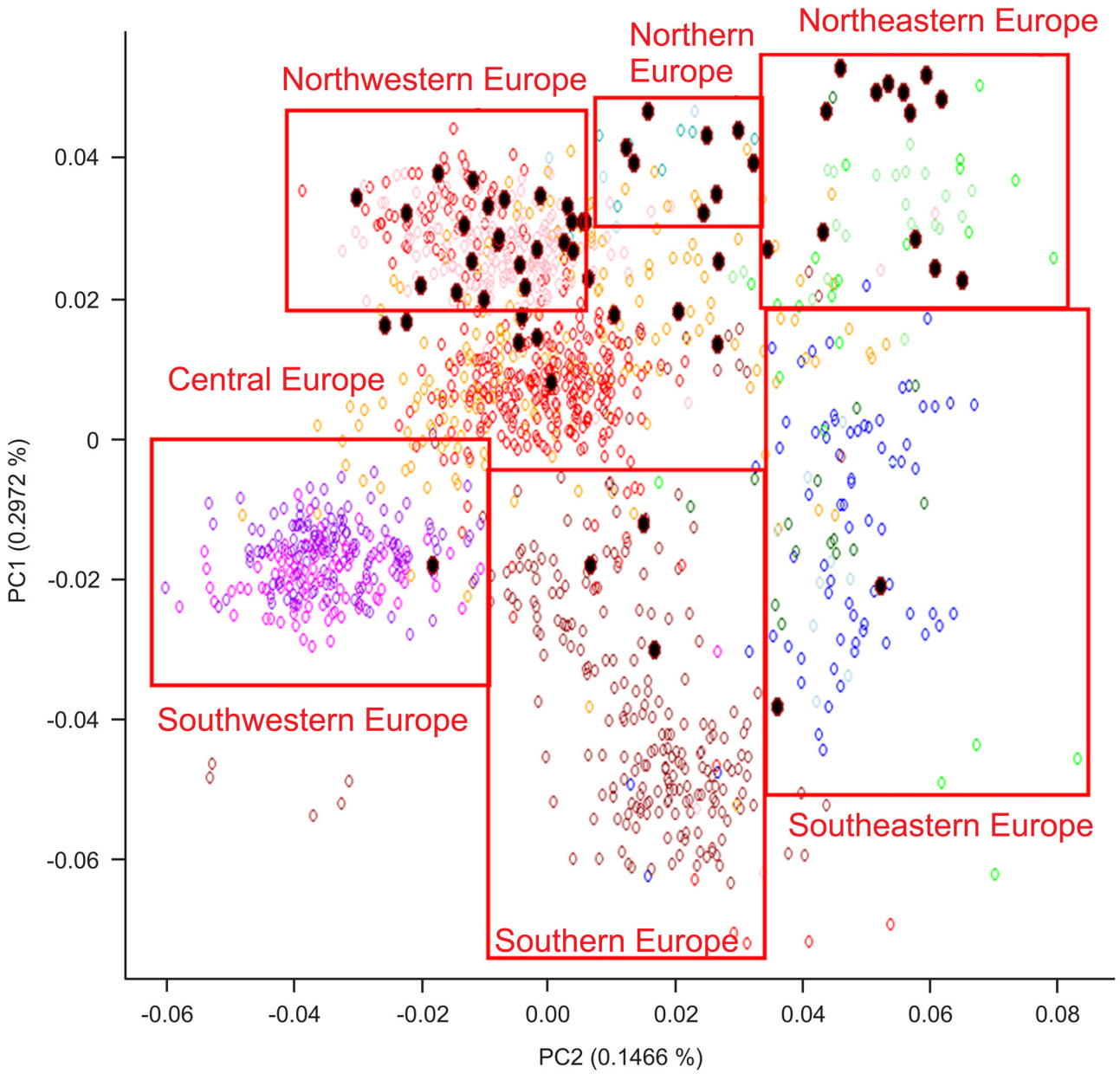
Andrew Laslett[78,79], Neta Lavon[80], Dong Ryul Lee[81], Jeoung Eun Lee[82], Chunliang Li[83], Linda S Lim[84], Tenneille E Ludwig[85], Yu Ma[86], Edna Maltby[87], Ileana Mateizel[88], Yoav Mayshar[89], Maria Mileikovsky[90], Stephen L Minger[91,92], Takamichi Miyazaki[93], Shin Yong Moon[94], Harry Moore[95], Christine Mummery[96], Andras Nagy[97], Norio Nakatsuji[98], Kavita Narwani[99], Steve K W Oh[100], Sun Kyung Oh[101], Cia Olson[102], Timo Otonkoski[103,104], Fei Pan[105], In-Hyun Park[106], Steve Pells[107], Martin Pera[108], Lygia V Pereira[109], Ouyang Qi[110], Grace Selva Raj[111], Benjamin Reubinoff[112], Alan Robins[113], Paul Robson[114], Janet Rossant[115], Ghasem H Salekdeh[116], Eva Sánchez[117], Thomas C Schulz[118], Karen Sermon[119], Jameelah Sheik Mohamed[120], Hui Shen[121], Eric Sherrer[122],

[74]Reproductive Genetics Institute, Chicago Illinois, USA.
[75]Vavilov Institute of General Genetics, Moscow, Russia.
[76]USC Epigenome Center, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.
[77]Newcastle University, Institute of Genetic Medicine, International Centre for Life, Newcastle, UK.
[78]CSIRO Materials Science and Engineering, Clayton, Australia.
[79]Department of Anatomy and Developmental Biology, Monash University, Clayton, Australia.
[80]Regenerative Medicine Institute, Cedars-Sinai Medical Institute, Los Angeles, California, USA.
[81]Department of Biomedical Science, CHA Stem Cell Institute, CHA University, Gangnam-gu, Seoul, Korea.
[82]CHA Stem Cell Institute, CHA University, Gangnam-gu, Seoul, Korea.
[83]Shanghai Stem Cell Institute, Shanghai JiaoTong University School of Medicine, Shanghai, China.
[84]Genome Institute of Singapore, Singapore.
[85]WiCell Research Institute, Madison, Wisconsin, USA.
[86]Shanghai Stem Cell Institute, Shanghai JiaoTong University School of Medicine, Shanghai, China.
[87]Sheffield Diagnostic Genetic Services, Sheffield Children's NHS Trust, Sheffield, UK.
[88]Department of Embryology and Genetics, Vrije Universiteit Brussel, Brussels, Belgium.
[89]Stem Cell Unit, Department of Genetics, Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel.
[90]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada.
[91]Wolfson Centre for Age-Related Diseases, King's College London, London, UK.
[92]GE Healthcare, Cardiff, UK.
[93]Laboratory of Embryonic Stem Cell Research, Stem Cell Research Center, Institute for Frontier Medical Sciences, Kyoto University, Japan.
[94]Department of Obstetrics & Gynaecology, Seoul National University College of Medicine, Seoul, Republic of Korea.
[95]Centre for Stem Cell Biology, Department of Biomedical Science, The University of Sheffield, Sheffield, UK.
[96]Department of Anatomy & Embryology, Leiden University Medical Center, Leiden, The Netherlands.
[97]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada.
[98]Institute for Integrated Cell-Material Sciences, Kyoto University, Ushinomiya-cho, Yoshida, Sakyo-ku, Kyoto, Japan.
[99]Regenerative Medicine Institute, Cedars-Sinai Medical Institute, Los Angeles, California, USA.
[100]Bioprocessing Technology Institute, Singapore.
[101]Institute of Reproductive Medicine & Population, Medical Research Center, Seoul National University, Seoul, Republic of Korea.
[102]Research Programs Unit, Molecular Neurology, Biomedicum Stem Cell Centre, University of Helsinki.
[103]Research Programs Unit, Molecular Neurology, Biomedicum Stem Cell Centre, University of Helsinki.
[104]Children's Hospital, University of Helsinki and Helsinki University Central Hospital, Finland.
[105]USC Epigenome Center, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.
[106]Department of Genetics, Yale Stem Cell Center, Yale School of Medicine, New Haven, Connecticut, USA.
[107]MRC Centre for Regenerative Medicine, University of Edinburgh, Edinburgh, UK.
[108]Eli and Edyth Broad Center for Regenerative Medicine and Stem Cell Research, University of Southern California, Keck School of Medicine, Los Angeles, California, USA.
[109]Laboratório Nacional de Células-Tronco Embrionárias (LaNCE), Depto. Genética e Biologia Evolutiva, Universidade de São Paulo, São Paulo, Brazil.
[110]Institute of Reproductive & Stem Cell Engineering Central South University, Reproductive & Genetic Hospital CITIC-XIANGYA, Changsha, Hunan, People's Republic of China.
[111]Singapore Stem Cell Bank, A-STAR, Singapore.
[112]The Hadassah Human Embryonic Stem Cell Research Center, The Goldyne Savad Institute of Gene Therapy and the Department of Obstetrics and Gynecology, Hadassah University Medical Center, Jerusalem, Israel.
[113]Viacyte, Athens Georgia, USA.
[114]Genome Institute of Singapore, Singapore.
[115]Program for Developmental Biology. The Hospital for Sick Children, Toronto, Ontario, Canada.
[116]Department of Molecular Systems Biology, Cell Science Research Center, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Islamic Republic of Iran.
[117]Valencia Node of the Spanish Stem Cell Bank. Prince Felipe Research Center, Valencia, Spain.
[118]Viacyte, Athens Georgia, USA.
[119]Department of Embryology and Genetics, Vrije Universiteit Brussel, Brussels, Belgium.
[120]Genome Institute of Singapore, Singapore.
[121]USC Epigenome Center, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.
[122]Viacyte, Athens Georgia, USA.

Kuldip Sidhu[123], Carlos Simon Vallés[124], Shirani Sivarajah[125–127], Heli Skottman[128], Claudia Spits[129], Glyn Stacey[130], Raimund Strehl[131], Nick Strelchenko[132], Hirofumi Suemori[133], Bowen Sun[134], Riitta Suuronen[135], Kazutoshi Takahashi[136], Timo Tuuri[137], Parvathy Venu[138], Yuri Verlinsky[139,151], Dorien Ward-van Oostwaard[140], Daniel J Weisenberger[141], Yue Wu[142,143], Shinya Yamanaka[144–147], Lorraine Young[148] & Qi Zhou[149]

[123]Stem Cell Laboratory, Faculty of Medicine, University of New South Wales, Australia.
[124]Valencia Node of the Spanish Stem Cell Bank. Prince Felipe Research Center, Valencia, Spain.
[125]Singapore Stem Cell Bank, A-STAR, Singapore.
[127]Optics and Nanoelectronics Research Group, NICTA Victorian Research Laboratory, The University of Melbourne, Parkville, Australia.
[128]Institute for Regenerative Medicine, University of Tampere, Tampere, Finland.
[129]Department of Embryology and Genetics, Vrije Universiteit Brussel, Brussels, Belgium.
[130]UK Stem Cell Bank, Division of Cell Biology and Imaging, National Institute for Biological Standards and Control, South Mimms, Herts., UK.
[131]Cellartis AB, Goteborg, Sweden.
[132]Reproductive Genetics Institute, Chicago Illinois, USA. [(2)]Present address: Department of Obstetrics and Gynecology, New York, New York, USA.
[133]Laboratory of Embryonic Stem Cell Research, Stem Cell Research Center, Institute for Frontier Medical Sciences, Kyoto University.
[134]Shanghai Stem Cell Institute, Shanghai JiaoTong University School of Medicine, Shanghai, China.
[135]Institute for Regenerative Medicine, University of Tampere, Tampere, Finland.
[136]Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto, Japan.
[137]Research Programs Unit, Molecular Neurology, Biomedicum Stem Cell Centre, University of Helsinki.
[138]Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India.
[139]Reproductive Genetics Institute, Chicago Illinois, USA.
[151]Deceased.
[140]Department of Anatomy & Embryology, Leiden University Medical Center, Leiden, The Netherlands.
[141]USC Epigenome Center, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.
[142]Wolfson Centre for Age-Related Diseases, King's College London, London, UK.
[143]GE Healthcare, Cardiff, UK.
[144]Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto, Japan.
[147]Gladstone Institute of Cardiovascular Disease, San Francisco, California, USA.
[148]Wolfson Centre for Stem Cells, Tissue Engineering & Modelling (STEM), Centre for Biomolecular Sciences, University of Nottingham, University Park, UK.
[149]CSIRO Materials Science and Engineering, Clayton, Australia.





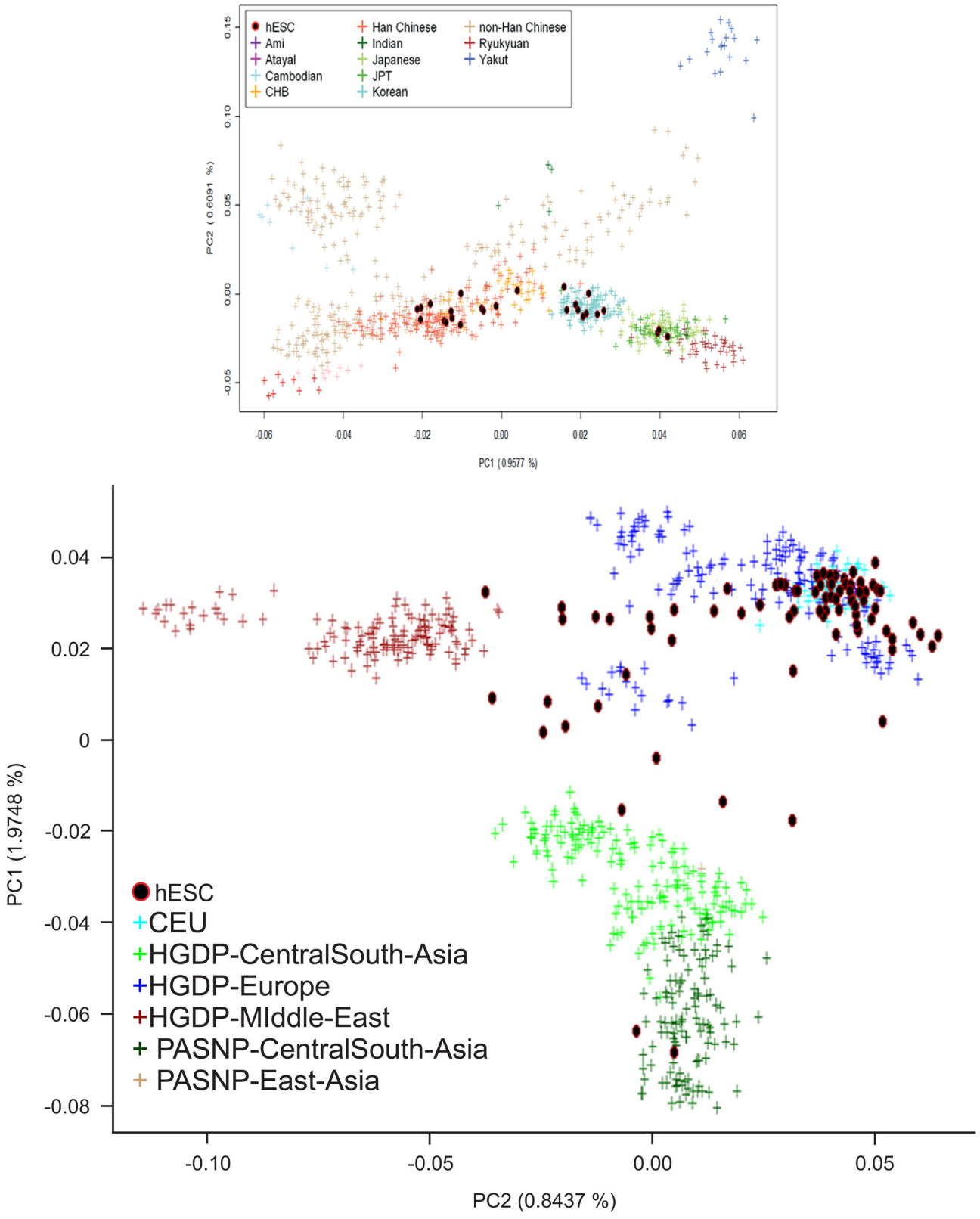
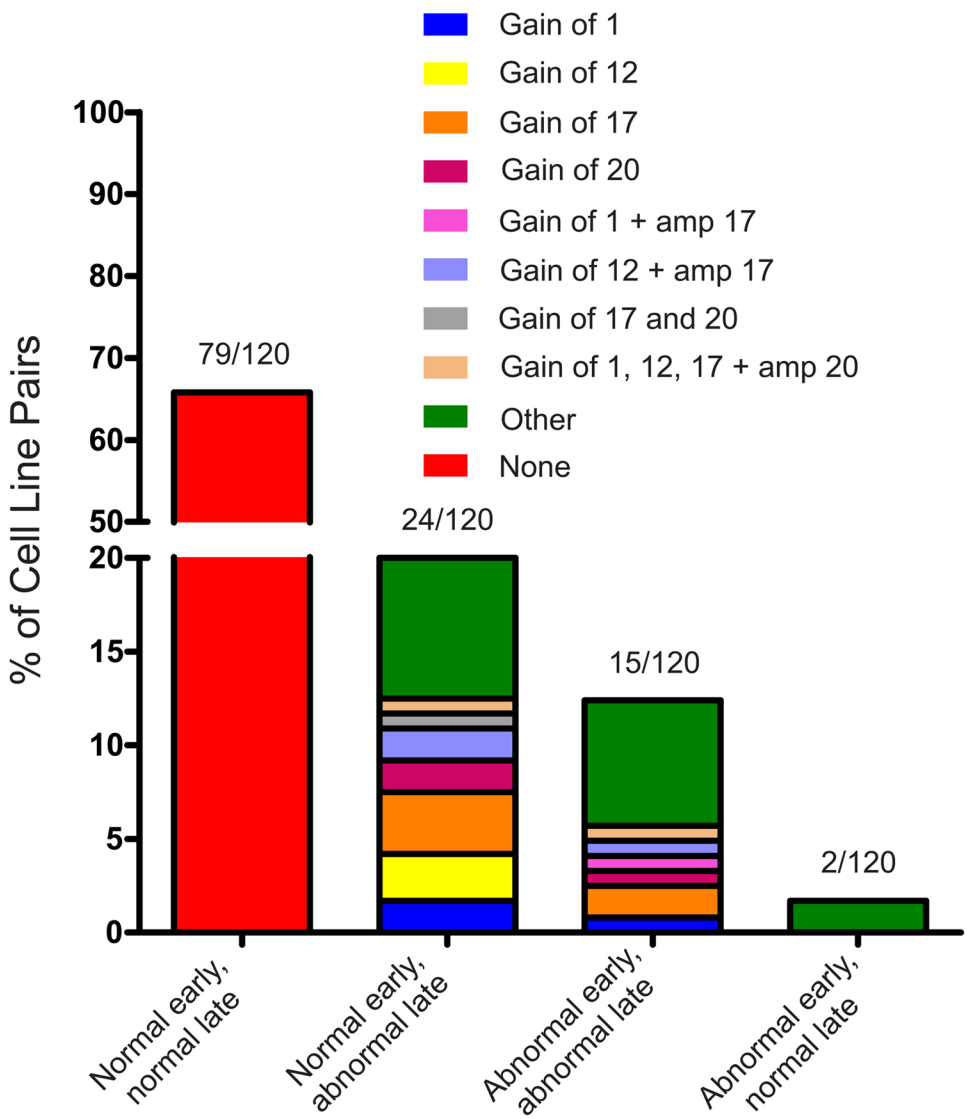


Figure 1.

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 (**c**).



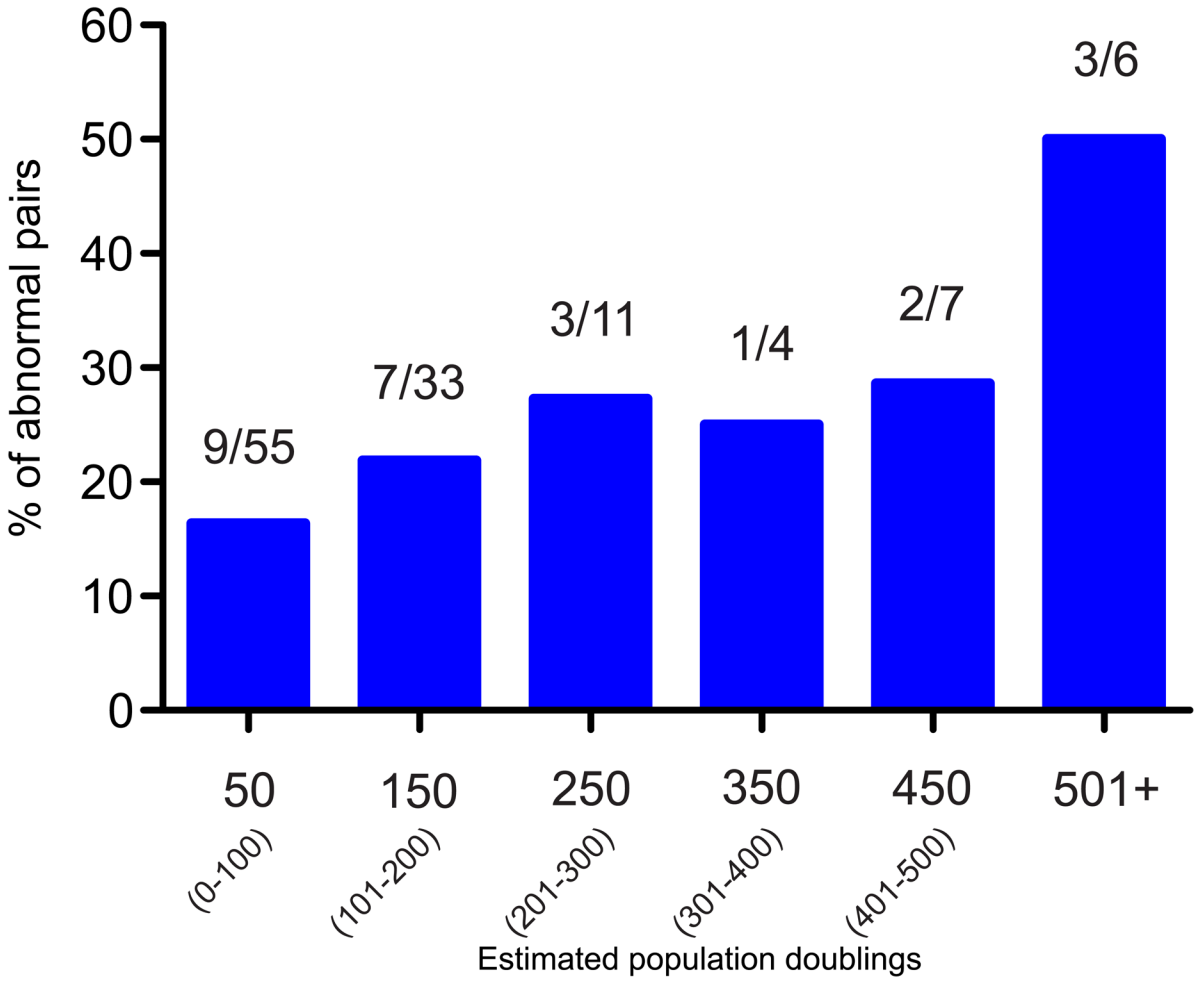


Figure 2.

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, abnormal 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 (C02 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 117 pairs could reliably be assigned an estimated population doubling time estimate.

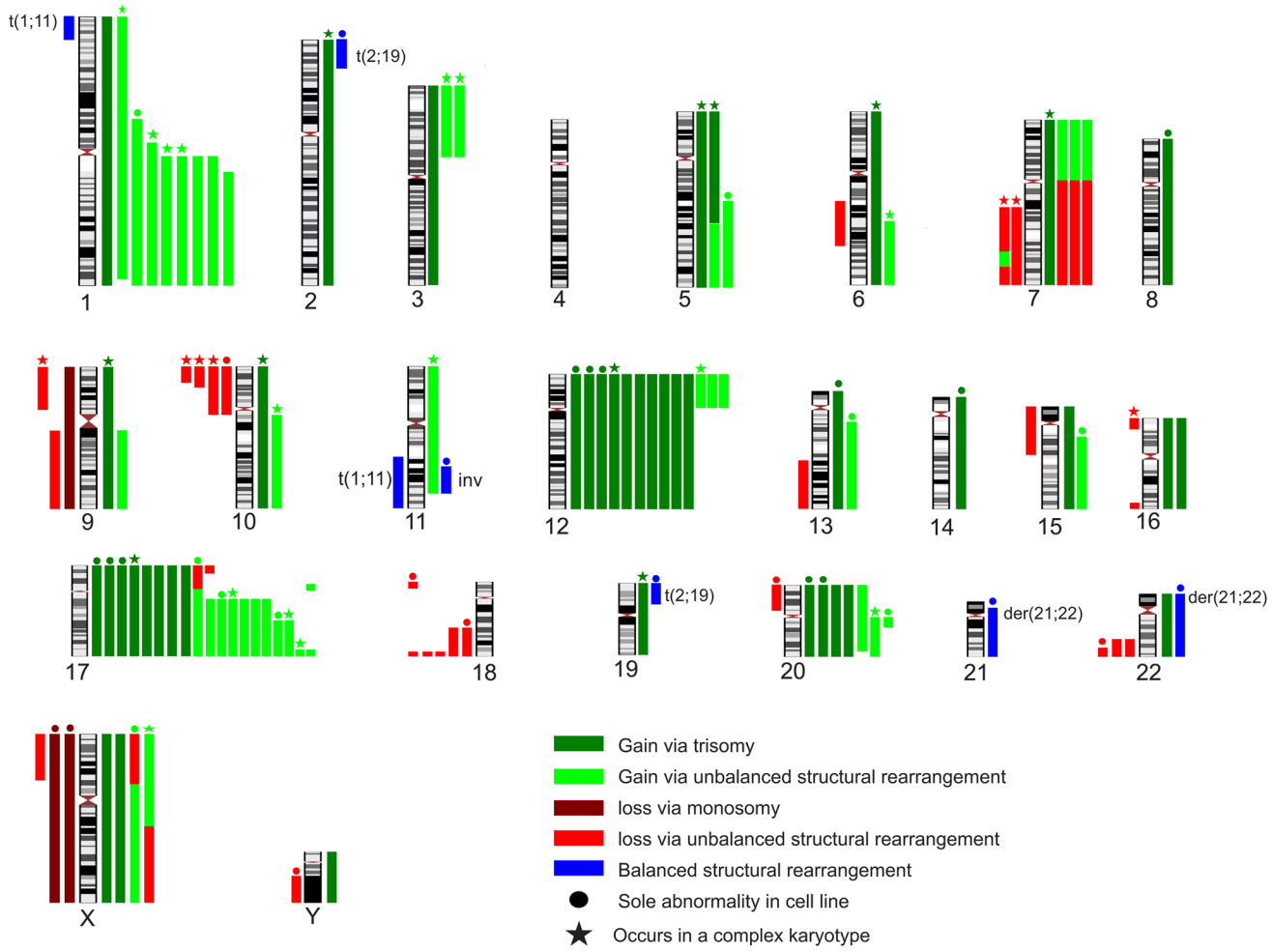


Figure 3.

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.

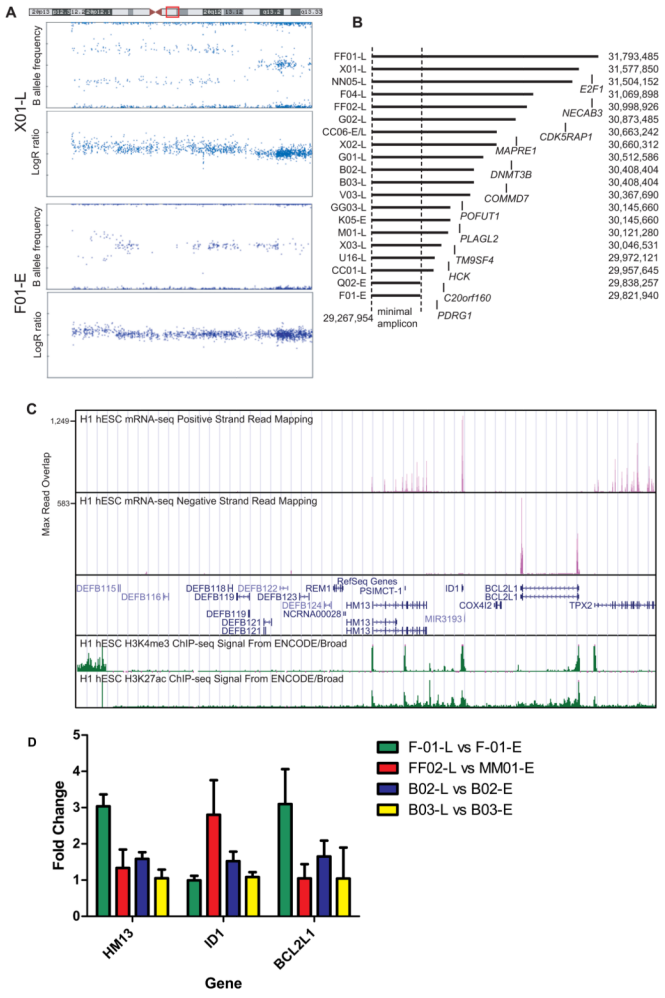


Figure 4.

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**).

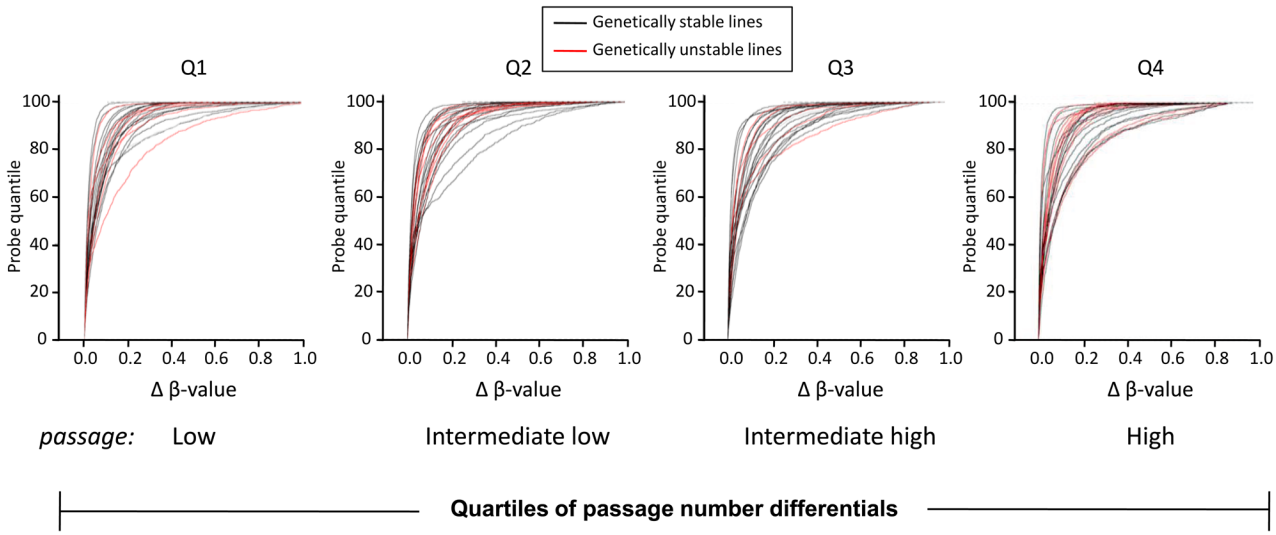


Figure 5.

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).

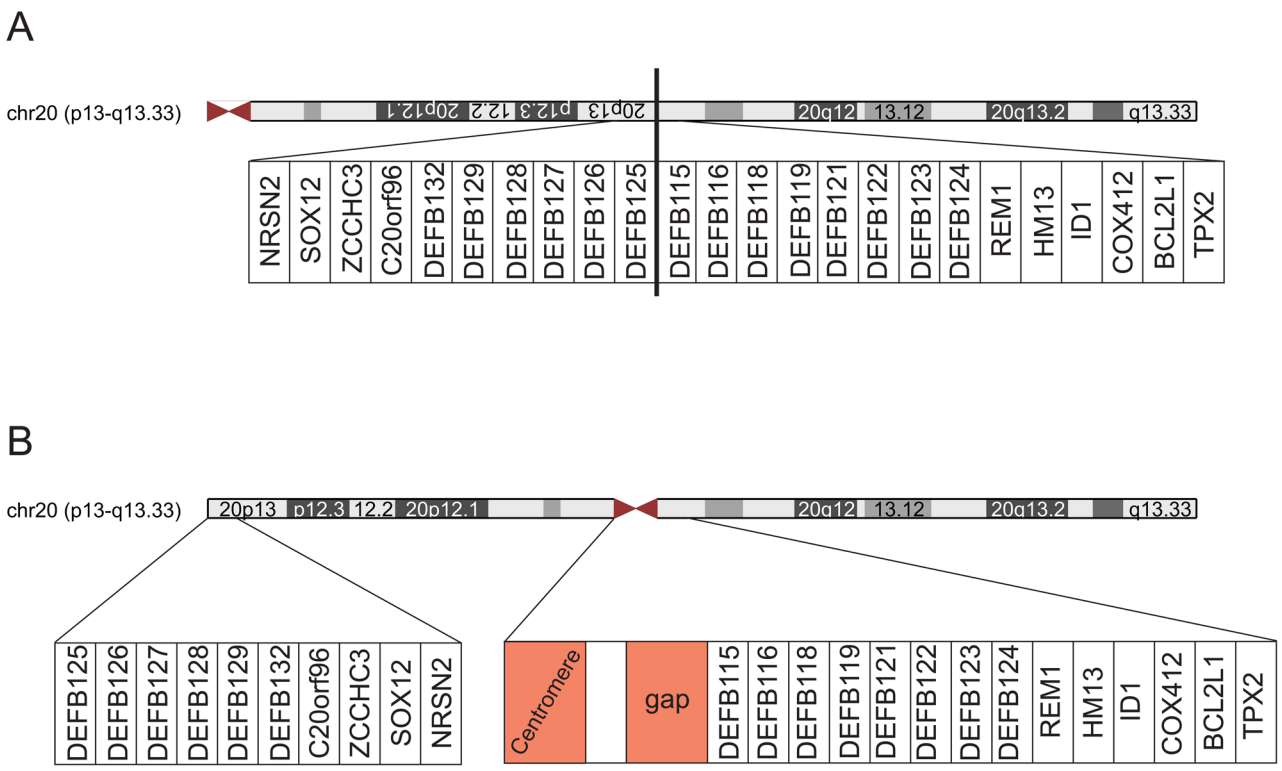


Figure 6.

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.

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	Citations from 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.

Table 2

The top 20 genes that were most frequently gained, lost or showed no change in DNA methylation levels in the 122 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.