ARTICLE

doi:10.1038/nature09871

Copy number variation and selection during reprogramming to pluripotency

Samer M. Hussein[1,2]*, Nizar N. Batada[3]*, Sanna Vuoristo[2], Reagan W. Ching[4], Reija Autio[5,6], Elisa Närvä[5], Siemon Ng[3], Michel Sourour[1], Riikka Hämäläinen[1,2], Cia Olsson[2], Karolina Lundin[2], Milla Mikkola[2], Ras Trokovic[2], Michael Peitz[7], Oliver Brüstle[7], David P. Bazett-Jones[4], Kari Alitalo[8], Riitta Lahesmaa[5], Andras Nagy[1,9] & Timo Otonkoski[2,10]

The mechanisms underlying the low efficiency of reprogramming somatic cells into induced pluripotent stem (iPS) cells are poorly understood. There is a clear need to study whether the reprogramming process itself compromises genomic integrity and, through this, the efficiency of iPS cell establishment. Using a high-resolution single nucleotide polymorphism array, we compared copy number variations (CNVs) of different passages of human iPS cells with their fibroblast cell origins and with human embryonic stem (ES) cells. Here we show that significantly more CNVs are present in early-passage human iPS cells than intermediate passage human iPS cells, fibroblasts or human ES cells. Most CNVs are formed *de novo* and generate genetic mosaicism in early-passage human iPS cells. Most of these novel CNVs rendered the affected cells at a selective disadvantage. Remarkably, expansion of human iPS cells in culture selects rapidly against mutated cells, driving the lines towards a genetic state resembling human ES cells.

Reprogramming somatic cells to pluripotency can be achieved by forced expression of a defined set of factors[1,2]. Several methods have been developed for generating human iPS cells, such as retroviral transduction[1], DNA-transposition-based systems[3,4], transient plasmid delivery[5] and integration/plasmid-free systems[6,7]. To improve efficiency and in an effort to understand the process of reprogramming, several groups have demonstrated that modulating key components of the cell cycle, such as repression of the *Ink4a/Arf* locus or downregulation of the p53–p21 pathway, have marked positive effects on reprogramming efficiency[8–12]. However, p53 suppression can lead to increased levels of DNA damage and genomic instability. These findings suggest that the reprogramming process places a heavy burden on cellular integrity and highlight the importance of further exploring the nature of the DNA damage that is associated with the reprogramming process.

High CNV levels in early-passage human iPS cells

To determine whether reprogramming is associated with *de novo*-generated CNVs, we used the Affymetrix SNP array 6.0 to characterize 22 human iPS cell lines along with 17 human ES cell lines[13], as well as three parental and one unrelated fibroblast lines as controls (Supplementary Table 1). The human iPS cell lines were established either by retroviral[2] or *piggyBac*[3,4] gene delivery methods and confirmed as human iPS cells using established criteria[14] (Supplementary Figs 1–3 and Supplementary Table 2). Nine of the 22 human iPS cell lines were characterized at more than one passage to track CNVs during propagation.

The median number of CNVs in human iPS cell lines (109) was about twofold higher than in human ES cell lines (55) and fibroblasts (53) (Supplementary Fig. 4a and Supplementary Tables 3 and 4). We found that the majority of CNVs (52.4%) in human iPS cells were not present in either human ES cells or fibroblasts (Supplementary Fig. 4b). Interestingly, the number of CNVs negatively correlated with the passage number. This was surprising because fibroblasts and human ES cells showed no significant changes during intermediate length passaging (Supplementary Fig. 4c, d). Both the number and the total size of CNVs in human iPS cell lines decreased during propagation (Fig. 1a and Supplementary Fig. 4e). Neither the reprogramming factor delivery method, fibroblast source or viral integration sites nor the presence or absence of Myc during reprogramming (Fig. 1b, c and Supplementary Fig. 5) influenced these results. This trend was verified in an independent data set on human iPS cell lines derived from four adult skin fibroblasts (Supplementary Table 5), as well as within individual human iPS cell lines analysed at early and later passages (Fig. 1b–d). Our findings indicate that CNVs are generated during the reprogramming process.

Genetic mosaicism in human iPS cells

The decrease in CNVs during passaging could be explained either by DNA repair mechanisms or by mosaicism followed by selection. We propose that DNA repair may not be efficient enough to explain the rapid decrease in CNVs but, instead, that *de novo*-generated CNVs create mosaicism, which is followed by selection favouring less damaged cells during propagation. To obtain direct proof for mosaicism, we established new human iPS cell lines and tested these at very early passages (passage 2 and 3) for CNVs by using fluorescence *in situ* hybridization (FISH). We chose a probe that maps to a locus on chromosome 1 that, according to our single nucleotide polymorphism (SNP) array data, is frequently affected in human iPS cell lines (Fig. 2a). A control probe was selected from a chromosome 1 location that showed normal copy number (2) across all human iPS cell lines that were tested. During early passages, the test probe demonstrated a

[1] Samuel Lunenfeld Research Institute, Toronto, Ontario M5T 3H7, Canada. [2] Research Program Unit, Molecular Neurology, Biomedicum Stem Cell Center, University of Helsinki, Helsinki FI-00014, Finland. [3] Ontario Institute for Cancer Research, Toronto, Ontario M5G 0A3, Canada. [4] Department of Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario M5G 1L7, Canada. [5] Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, Turku 20520, Finland. [6] Department of Signal Processing, Tampere University of Technology, Tampere 33101, Finland. [7] Institute of Reconstructive Neurobiology, Life and Brain Center, University of Bonn and Hertie Foundation, D-53127 Bonn, Germany. [8] Molecular Cancer Biology Laboratory, University of Helsinki and Helsinki University Central Hospital, Helsinki FI-00014, Finland. [9] Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [10] Children's Hospital, University of Helsinki, Helsinki FI-00029, Finland.

*These authors contributed equally to this work.

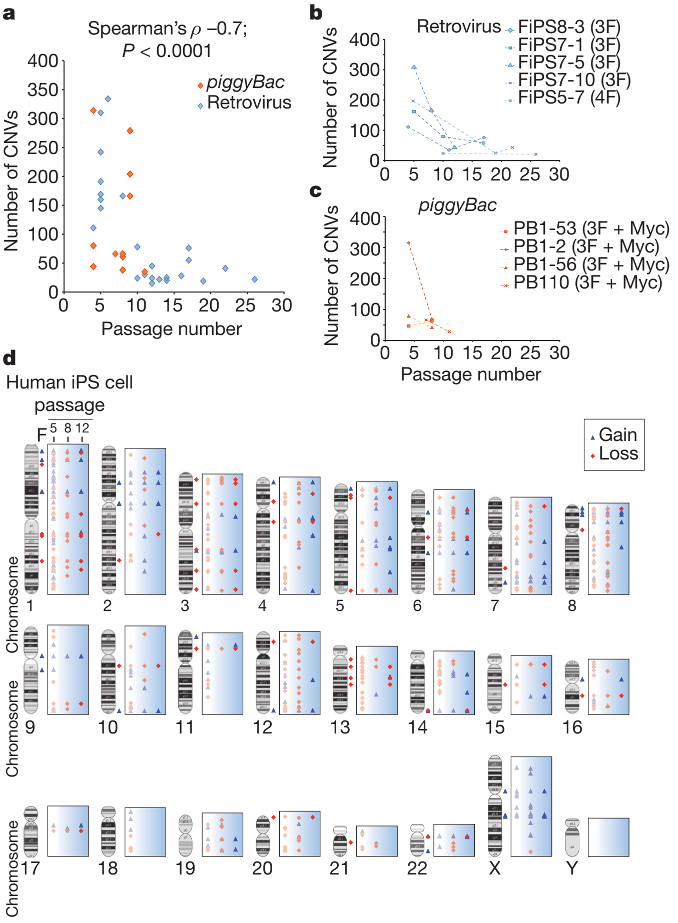

Figure 1 | High mutation level in human iPS cells is reduced through moderate culture. **a**, Number of CNVs in human iPS cell lines with respect to passage number. Each data point represents a sample: blue, retrovirus-derived human iPS cell lines; orange, *piggyBac* lines. Spearman's rank correlation coefficient (ρ) and Student's *t*-distribution were used for statistical analysis and *P*-value calculations. **b**, **c**, With passaging, both retrovirus-derived (**b**) and *piggyBac*-transposon-derived (**c**) human iPS cell lines (listed) show a constant and sharp decrease in the number of CNVs. Numbers in parentheses indicate the number of factors used for generating the corresponding human iPS cell lines: 3F, OCT4, SOX2 and KLF4; 4F, OCT4, SOX2, NANOG and LIN28. **d**, Genomic representation demonstrating the sharp decline in the number of CNVs from early passage (passage 5) to intermediate passage (8) and intermediate–late passage (12) of the human iPS cell line FiPS7-5 relative to the CNVs in the parental fibroblast (F). Blue triangles represent amplifications, and red diamonds represent deletions, with colour intensity varying with passage number.

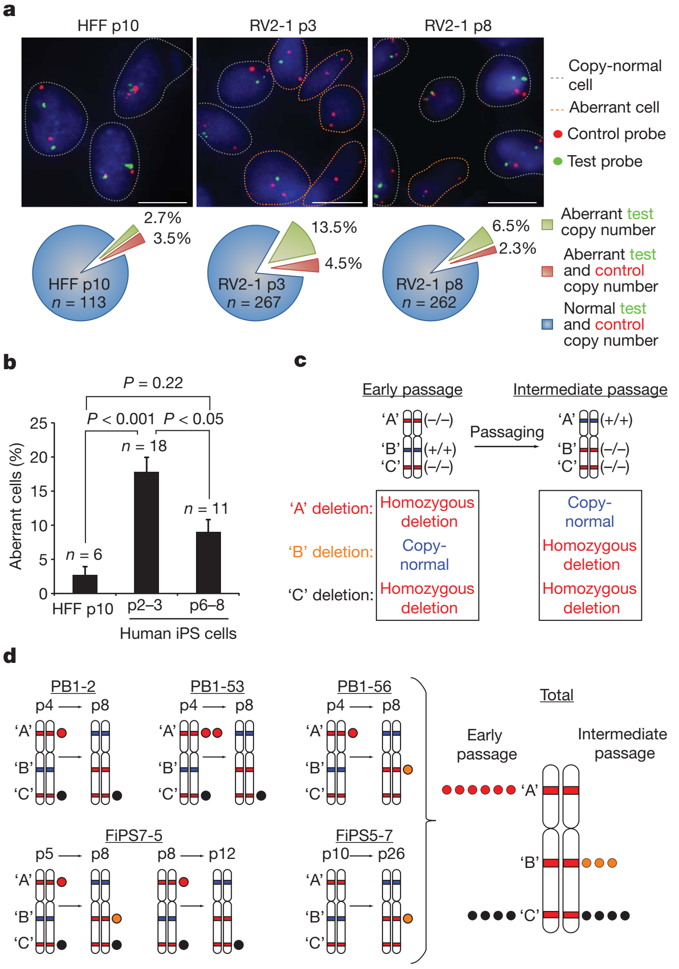

Figure 2 | Increased mosaicism in early-passage human iPS cells. **a**, Merged FISH field images of human foreskin fibroblast (HFF) cells at passage 10 (p10) and the human iPS cell line RV2-1 at p3 and p8. Test (green) and control (red) probes are located on chromosome 1. The pie charts below the images show the percentage of cells that have only aberrant test probe (but not control probe) foci counts or copy number (green), cells that have aberrant test and control probe foci counts (red), and normal cells (blue). Five different field images were counted per sample; n = the total number of cells counted. Dashed grey lines indicate normal cells, and orange lines indicate aberrant cells. Scale bar, 10 µm. **b**, Histogram demonstrates the mean fraction of aberrant cells in fibroblasts (HFF p10) and early-passage (p2–3) and intermediate-passage (p6–8) human iPS cells. Each field contained approximately 20–100 cells; n = total number of fields counted. Error bars, s.e.m. One-way analysis of variance and the Tukey–Kramer post-hoc test were used for statistical analysis and *P*-value calculations. **c**, Three categories of homozygous deletions: type 'A', detected only during early passages; type 'B', appearing in later passages; and type 'C', seen in both early and intermediate passages. −/−, homozygous deletion (red band); +/+, normal copy number (blue band). **d**, Left, non-parental homozygous deletions present in five cell lines passaged from an early passage to an intermediate passage. Each circle represents a homozygous deletion: 'A', red circles; 'B', orange circles; and 'C', black circles. Right, combined total count of homozygous deletions.

significantly higher fraction of cells with aberrant copy number state than the control probe (Fig. 2a). The fraction of aberrant cells was also significantly higher in early-passage human iPS cells (18%) than in fibroblasts (3%) or in later-passage human iPS cells (9%) (Fig. 2b, Supplementary Fig. 6 and Supplementary Table 6).

To provide evidence for selection, we focused on regions containing homozygous deletions, which DNA repair mechanisms cannot correct. Although we could detect only a small number of such deletions, our detection rate for homozygous deletions was very reproducible, detecting 98% of the deletions in three to four replicates (Supplementary Table 7). Our false discovery rate was 9.7% for detecting other types of CNV (Supplementary Table 7 and Supplementary Fig. 7), suggesting low error in calling CNVs and robust detection of homozygous deletions. We focused on homozygous deletions found only in human iPS cell lines and not their parental fibroblasts, and we categorized these into three groups: type 'A' homozygous deletions, which are present only in early passages; type 'B' homozygous deletions, which are detected only in later passages; and type 'C' homozygous deletions, which remain during passaging (Fig. 2c). Five of the cell lines presented with non-parental homozygous deletions at an early or intermediate passage (Fig. 2d). In four of the lines, we identified homozygous deletions that were selected against during passaging (type A). We also found type B and type C deletions, suggesting that selection pressure is bidirectional, selecting both for and against CNVs (Fig. 2d).

Novel CNVs in early-passage human iPS cells

We obtained a list of 6,596 non-overlapping common CNVs identified in 270 healthy individuals from two combined studies in the HapMap Project[15,16]. These common CNVs could be considered to be functionally the most neutral[17,18]. We designated the set of CNVs that were identified in our study and that do not belong to common CNVs as novel CNVs. They accounted for 15% of the total CNVs in fibroblasts and 25% in human ES cells but 37% in human iPS cells (Supplementary Fig. 8a). The novel CNV fraction was significantly higher in early-passage human iPS cell lines than in later passages, in which it decreased to levels similar to those found in human ES cells and in fibroblasts (Supplementary Fig. 8b). Only a minority of non-parental human iPS cell CNVs that overlapped human ES cell CNVs were novel (Supplementary Fig. 8c).

Selection against highly damaged human iPS cells

In a mosaic population, CNVs can be identified with SNP arrays only if they are present above the detection threshold level, which depends on the type and the size of CNVs. For example, with the Affymetrix SNP array 6.0 and Genotyping Console 3.0 software algorithm, a trisomy might not be fully detectable if the mutant cell contribution is less than 40% (refs 13, 19). Owing to the multiple probe-based and threshold-based nature of CNV identification, a single large CNV found within a subpopulation of the cells (for example, type L cells in Fig. 3a) could be misrepresented as multiple small, consecutive CNVs, providing a false representation of the data. If the CNV is selected for during maintenance, these type L cells will become more prevalent in the population. Consequently, the detection of this large CNV would be more accurate in intermediate passages (that is, a relatively larger size and number of overlapping CNVs within the intermediate-passage cells would be shared with early-passage cells), and the number of 'false' consecutive CNVs would decrease. In the case of selection against a mutation, the CNV number would still show a decrease with passaging, but this mechanism would not affect the size of overlapping CNVs between early and intermediate passages. The overlap in CNVs between early and intermediate passages would be minimal. We therefore investigated whether this putative error component could account for the observed high number of novel CNVs in early-passage lines (Fig. 3a).

Focusing on novel CNVs present only in human iPS cells, we found that the overlapping CNV size was equivalent among different passages and clones (Fig. 3b), but the number of overlapping (shared) novel CNVs in human iPS cells was relatively small (Fig. 3c), indicating that most CNVs in early-passage lines are not the product of type L cell subpopulations and are indeed 'true' CNVs. Moreover, the rate of selection against novel CNVs in a passage interval (change in novel CNV number divided by number of passages) was significantly higher between the relatively early passages and the intermediate passages than it was between the intermediate and late passages. The latter rate was comparable to the selection rate of human ES cells (Fig. 3d), suggesting that early-passage human iPS cells endure strong selection pressure and lose the majority of their *de novo* mutations. These results demonstrate that most of the novel CNVs in human iPS cells are generated during the reprogramming process.

Novel CNVs recur within fragile regions

To investigate possible sources of negative selection, we asked whether the high level of *de novo* mutations led to functional consequences, such as an increase in senescence or apoptosis or a decrease in self-renewal. We assessed mutations within genes that may affect differentiation, proliferation or maintenance of pluripotency. In early passages but not in later passages, several deletions were found in genes and regions essential for maintaining an undifferentiated state (Supplementary Table 8). Such mutations included deletions in the genes

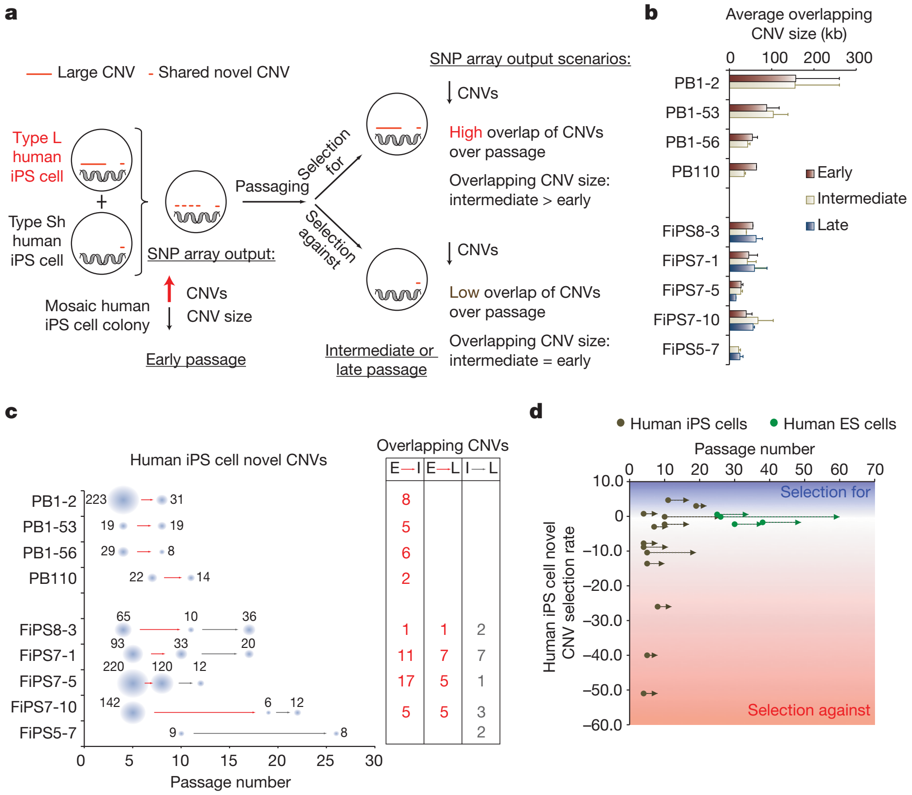

Figure 3 | High selection pressure against *de novo* mutations suggests reprogramming as the source of novel CNVs. **a**, Possible error in SNP array output from a mosaic human iPS cell colony containing type L cells (cells with a large CNV) and type Sh cells (cells with shared novel CNVs). **b**, Average overlapping novel CNV size (kilobases, kb) in human iPS cell lines at early, intermediate and late passage. Error bars, s.e.m. **c**, Left, plot showing the change in the number of novel CNVs in human iPS cells with passaging. Blue circle size corresponds to the number of CNVs observed at the passage tested, with the number listed next to the circle. Right, the number of overlapping CNVs. Values for early (E) compared with intermediate (I) or late (L) passage are shown in red, and for I compared with L are shown in grey. **d**, Selection rate of human iPS cells. The rate is calculated as change in novel CNV number divided by change in passage number. Four human ES cell lines were used as controls. Arrows indicate the start and the end of the passage range tested for each sample.

encoding the epidermal growth factor receptor, fibroblast growth factor receptor 2, β-catenin (also known as CTNNB1) and polycomb-bound regions, all of which have been implicated in human ES cell maintenance[20–23]. We also found that six early-passage human iPS cell lines had deletions in the regions encoding the microRNAs let-7c and miR-125b, which affect the expression of genes known to be involved in human ES cell differentiation and maintenance, such as those encoding Myc, Ras, p53 and ERBB3 (refs 24–27).

To explore possible mechanisms behind reprogramming-induced CNVs, we investigated mutations reminiscent of those induced by replication stress, such as DNA replication fork stalling and collapse, and we assessed CNVs found in regions of genomic fragility, such as common fragile sites (CFSs) and subtelomeric regions[28–30]. CFSs contain late-replicating sequences and are a major target for genomic rearrangements in oncogene-expressing and pre-neoplastic cells[31–34]. We compiled a list of CFSs from published reports (Supplementary Table 9) and measured the fraction of recurring deletions within CFS regions compared with the whole genome. Deletions recurred more frequently in CFSs than in the generic part of the genome, and more specifically they recurred more frequently in human iPS cells than in human ES cells and fibroblasts (Fig. 4a and Supplementary Table 9). Furthermore, this recurring CNV fraction consisted mainly of novel CNVs in human iPS cells (Fig. 4a), suggesting that a higher level of novel CNVs may result in part from replication stress[28,29]. This observation was consistent with previous reports demonstrating increases in the level of reactive oxygen species during reprogramming. Reactive oxygen species are prevalent in cells undergoing replication stress and may contribute to the incidence of mutations in other parts of the genome as well[35].

To examine mutations that correlate with senescent or apoptotic cells, we focused on deletions incurred in subtelomeric regions because these areas have been shown to be highly sensitive to DNA double strand breaks[36,37] and because deletions within these regions are a major cause of chromosomal instability[37]. We compared the average deletion size with those seen in CFSs and the whole genome. The average deletion size within subtelomeric regions was significantly larger in early-passage lines than in later passages, while remaining unchanged in the generic part of the genome and in CFSs (Supplementary Fig. 9a). We also found that several early-passage human iPS cell lines had deletions in the subtelomeric region nearest to the telomeres (25 kilobases away) (Supplementary Fig. 9b). The increased selection against large subtelomeric deletions is consistent with the idea that CNVs in these areas probably lead to a higher level of phenotypic change because these areas are gene rich and prone to genomic instability[38,39].

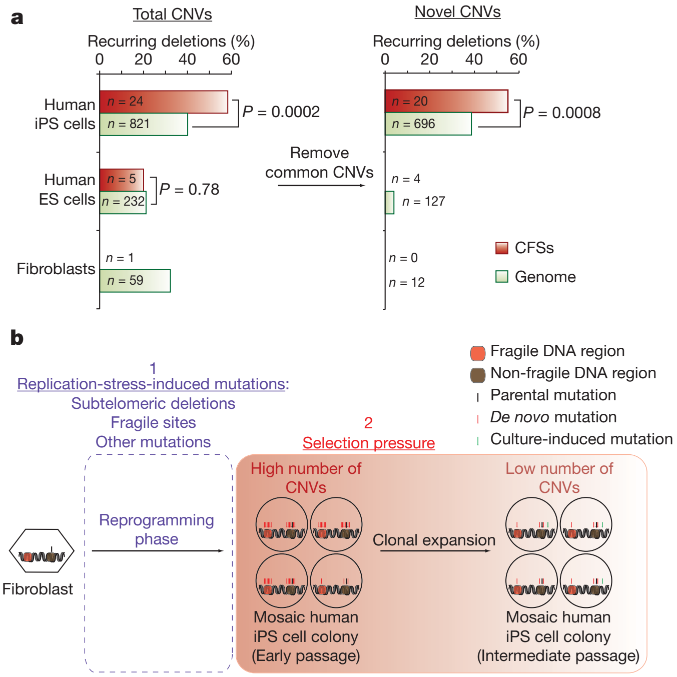

Figure 4 | Frequent mutations in fragile genomic regions influence selection during the expansion of human iPS cells. a, Recurring deletions as a proportion of total CNV deletions and novel CNV deletions, within either CFSs or the whole genome. Deletions are considered to be recurring if they are found in more than one sample (*n* = the total number of deletions observed). For human iPS cells, only non-parental deletions are considered. The chi-squared test was used for statistical analysis and *P*-value calculations. **b**, Summary model illustrating the increase in the number of CNVs that results from replication stress during reprogramming, followed by a selection phase that occurs after reprogramming and eliminates unstable human iPS cells containing high numbers of CNVs.

Discussion

Two recent studies[40,41] report on the observation of specific genomic aberrations associated with the pluripotent state in human ES cells and human iPS cells. One group carried out a meta-analysis of large numbers of gene expression profiles that had been determined for pluripotent stem cells by different laboratories[40]. They showed that human iPS cells are subject to the type of culture adaptations that have been shown to affect the karyotypic integrity of human ES cells[42]. Their data also suggest that a distinct category of genomic aberrations may be associated with the early phase of human iPS cell establishment. Their conclusion is in line with the second recent report, in which SNP arrays were used to compare CNVs in a large number of normal somatic cell lines, human ES cell lines and human iPS cell lines[41]. Interestingly, human ES cells were found to contain more gains, and human iPS cells more deletions, than somatic cell samples. This finding further substantiates the differences between these two types of pluripotent cell. It also underscores the differences in the selection forces that affect human ES cells and human iPS cells (at least during the establishment period) and that could affect the quality of the final products. Data from the second study[41] suggest that the reprogramming process is associated with selection for deletions that affect tumour-suppressor genes, whereas maintenance of the cell lines selects for duplications in oncogenic genes.

From our study, we conclude that the reprogramming process is associated with high mutation rates, causing increased levels of CNVs and genetic mosaicism in the resultant early-passage human iPS cell lines. Our data also suggest that *de novo* CNVs are the consequence of replication stress (Fig. 4b). Using our approach, we failed to find evidence suggesting that other mechanisms operate. Of the 116 DNA-repair-related and/or checkpoint-related genes that we investigated, we found only four cell lines in which a CNV might have affected a single gene (Supplementary Table 10). However, because our study was limited to CNVs, we could not exclude the possibility of other types of mutation that lead to perturbations of checkpoints or repair of DNA double strand breaks. Such mutations could lead to non-allelic homologous recombination (NAHR)-based rearrangements and/or non-homologous end-joining (NHEJ)-based rearrangements. Both NAHR and NHEJ have been reported to be involved in CNV formation[43].

In summary, because most *de novo* mutations confer a growth or survival disadvantage to the cells, they are selected against, eventually leading to a CNV load similar to that found in human ES cells. This negative selection, however, does not exclude the possibility that certain hazardous aberrations give the cell a selective advantage over cells with an intact genome. Our results highlight the importance of understanding the molecular mechanisms underlying the reprogramming of somatic cells to a pluripotent state, with particular emphasis on forces that negatively affect the integrity of the genome. With a better understanding of the reprogramming process, we will increase the likelihood of finding ways to counteract the pitfalls and create human iPS cells that can safely be used for cell-based therapies in the future.

METHODS SUMMARY

Human fibroblast lines were reprogrammed by retroviral transduction[1] and *piggyBac* transposition as previously described[3]. Human iPS cell lines were expanded and characterized as previously described[14]. *In vitro* differentiation of human iPS cells was carried out using embryoid body, neuronal and endodermal differentiation protocols as described in the Methods. Teratomas were generated as described elsewhere[44]. Supplementary Table 2 lists the details of the characterization of each human iPS cell clone and the factors used for reprogramming. Bisulphite sequencing of *NANOG* and *OCT4* promoters was performed as previously described[45]. Splinkerette PCR was used to identify viral integration sites in three human iPS cell lines as previously described[46]. FISH protocols are provided in the Supplementary Information. Samples were run on Affymetrix SNP array 6.0, and Genotyping Console 3.0.2 was used to analyse and determine CNV levels, genotype calls and loss of heterozygosity detection as detailed in the Methods.

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

Received 30 March 2010; accepted 27 January 2011.

1. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
2. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
3. Woltjen, K. *et al. piggyBac* transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* **458,** 766–770 (2009).
4. Kaji, K. *et al.* Virus-free induction of pluripotency and subsequent excision of reprogramming factors. *Nature* **458,** 771–775 (2009).
5. Yu, J. *et al.* Human induced pluripotent stem cells free of vector and transgene sequences. *Science* **324,** 797–801 (2009).
6. Warren, L. *et al.* Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7,** 618–630 (2010).
7. Kim, D. *et al.* Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. *Cell Stem Cell* **4,** 472–476 (2009).
8. Utikal, J. *et al.* Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. *Nature* **460,** 1145–1148 (2009).
9. Marión, R. M. *et al.* A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature* **460,** 1149–1153 (2009).
10. Li, H. *et al.* The *Ink4/Arf* locus is a barrier for iPS cell reprogramming. *Nature* **460,** 1136–1139 (2009).
11. Kawamura, T. *et al.* Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature* **460,** 1140–1144 (2009).
12. Hong, H. *et al.* Suppression of induced pluripotent stem cell generation by the p53–p21 pathway. *Nature* **460,** 1132–1135 (2009).
13. Närvä, E. *et al.* High-resolution DNA analysis of human embryonic stem cell lines reveals culture-induced copy number changes and loss of heterozygosity. *Nature Biotechnol.* **28,** 371–377 (2010).
14. Chan, E. M. *et al.* Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. *Nature Biotechnol.* **27,** 1033–1037 (2009).
15. McCarroll, S. A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature Genet.* **40,** 1166–1174 (2008).
16. Conrad, D. F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **464,** 704–712 (2010).
17. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305,** 525–528 (2004).
18. Iafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nature Genet.* **36,** 949–951 (2004).
19. Hagenkord, J. M. *et al.* Array-based karyotyping for prognostic assessment in chronic lymphocytic leukemia: performance comparison of Affymetrix 10K2.0, 250K Nsp, and SNP6.0 arrays. *J. Mol. Diagn.* **12,** 184–196 (2010).
20. Eiselleova, L. *et al.* A complex role for FGF-2 in self-renewal, survival, and adhesion of human embryonic stem cells. *Stem Cells* **27,** 1847–1857 (2009).
21. Brill, L. M. *et al.* Phosphoproteomic analysis of human embryonic stem cells. *Cell Stem Cell* **5,** 204–213 (2009).
22. Wang, L. *et al.* Self-renewal of human embryonic stem cells requires insulin-like growth factor-1 receptor and ERBB2 receptor signaling. *Blood* **110,** 4111–4119 (2007).
23. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441,** 349–353 (2006).
24. Melton, C., Judson, R. L. & Blelloch, R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature* **463,** 621–626 (2010).
25. Scott, G. K. *et al.* Coordinate suppression of ERBB2 and ERBB3 by enforced expression of micro-RNA *miR-125a* or *miR-125b*. *J. Biol. Chem.* **282,** 1479–1486 (2007).
26. Maimets, T., Neganova, I., Armstrong, L. & Lako, M. Activation of p53 by nutlin leads to rapid differentiation of human embryonic stem cells. *Oncogene* **27,** 5277–5287 (2008).
27. Johnson, S. M. *et al.* RAS is regulated by the *let-7* microRNA family. *Cell* **120,** 635–647 (2005).
28. Arlt, M. F. *et al.* Replication stress induces genome-wide copy number changes in human cells that resemble polymorphic and pathogenic variants. *Am. J. Hum. Genet.* **84,** 339–350 (2009).
29. Durkin, S. G. *et al.* Replication stress induces tumor-like microdeletions in *FHIT/FRA3B*. *Proc. Natl Acad. Sci. USA* **105,** 246–251 (2008).
30. Sfeir, A. *et al.* Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell* **138,** 90–103 (2009).
31. Gorgoulis, V. G. *et al.* Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434,** 907–913 (2005).
32. Bartkova, J. *et al.* DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **434,** 864–870 (2005).
33. Casper, A. M., Nghiem, P., Arlt, M. F. & Glover, T. W. ATR regulates fragile site stability. *Cell* **111,** 779–789 (2002).
34. Arlt, M. F., Durkin, S. G., Ragland, R. L. & Glover, T. W. Common fragile sites as targets for chromosome rearrangements. *DNA Repair (Amst.)* **5,** 1126–1135 (2006).
35. Esteban, M. A. *et al.* Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. *Cell Stem Cell* **6,** 71–79 (2010).
36. Kulkarni, A., Zschenker, O., Reynolds, G., Miller, D. & Murnane, J. P. Effect of telomere proximity on telomere position effect, chromosome healing, and sensitivity to DNA double-strand breaks in a human tumor cell line. *Mol. Cell. Biol.* **30,** 578–589 (2010).
37. Zschenker, O. *et al.* Increased sensitivity of subtelomeric regions to DNA double-strand breaks in a human cancer cell line. *DNA Repair (Amst.)* **8,** 886–900 (2009).
38. Linardopoulou, E. V. *et al.* Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature* **437,** 94–100 (2005).
39. Riethman, H., Ambrosini, A. & Paul, S. Human subtelomere structure and variation. *Chromosome Res.* **13,** 505–515 (2005).
40. Mayshar, Y. *et al.* Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. *Cell Stem Cell* **7,** 521–531 (2010).
41. 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).
42. Baker, D. E. *et al.* Adaptation to culture of human embryonic stem cells and oncogenesis *in vivo*. *Nature Biotechnol.* **25,** 207–215 (2007).
43. Hastings, P. J., Lupski, J. R., Rosenberg, S. M. & Ira, G. Mechanisms of change in gene copy number. *Nature Rev. Genet.* **10,** 551–564 (2009).
44. Mikkola, M. *et al.* Distinct differentiation characteristics of individual human embryonic stem cell lines. *BMC Dev. Biol.* **6,** 40 (2006).
45. Deb-Rinker, P., Ly, D., Jezierski, A., Sikorska, M. & Walker, P. R. Sequential DNA methylation of the Nanog and Oct-4 upstream regions in human NT2 cells during neuronal differentiation. *J. Biol. Chem.* **280,** 6257–6260 (2005).
46. Horn, C. *et al.* Splinkerette PCR for more efficient characterization of gene trap events. *Nature Genet.* **39,** 933–934 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by funding for the ESTOOLS consortium under the Sixth RFP of the EU (S.M.H., T.O., R.L. and O.B.), the Academy of Finland (T.O. and R.A.), the Finnish Cancer Organizations (R.L.), the Research Funds of the Helsinki University Hospital (T.O.), the Sigrid Jusélius Foundation (T.O.), the Stem Cell Network Canada, GL2 and CIHR (A.N.). S.M.H. is a recipient of a McEwen Post-doctoral Fellowship from the McEwen Centre for Regenerative Medicine. N.N.B. is the recipient of a New Investigator Award from the Ontario Institute for Cancer Research, through the Ontario Ministry of Research and Innovation. A.N. is Tier 1 Canada Research Chair in Stem Cells and Regeneration. We are grateful to I. Rogers, M. Faiz and K. Nagy for critically reading and editing the manuscript, and J. Ustinov, H. Sariola and J. Palgi for skilled assistance in the characterization of human iPS cells. We thank N. Rahkonen for sample processing, as well as the Finnish Microarray and Sequencing Centre (http://www.btk.fi/), and the Institute of Human Genetics, University of Bonn, for array processing.

Author Contributions S.M.H. coordinated and performed most of the experiments in this project, analysed and interpreted the data, and prepared the manuscript. N.N.B. analysed and interpreted the data and prepared an early version of the manuscript. S.V. provided essential experimental assistance and performed some experiments in this study. R.W.C. performed, and advised on the planning of, FISH experiments. R.A. and E.N. analysed and interpreted the data and coordinated the sample processing of the arrays. O.B. and M.P. provided analysed data and samples from Illumina arrays for validation. M.S., S.N., R.H., R.T., M.M., C.O. and K.L. contributed experimentally. D.P.B-J. and K.A. provided essential experimental material and support. R.L. directed the SNP array data generation and analysis. T.O. directed, coordinated and supervised the first stage of the project and contributed to the preparation of the manuscript. A.N. prepared the manuscript, supervised the second stage of the project and helped to interpret the data. K.A. and R.L. contributed equally as co-authors. A.N. and T.O. contributed equally as senior authors.

Author Information Affymetrix SNP array 6.0 data from each individual sample have been deposited with the NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE26173. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.O. (timo.otonkoski@helsinki.fi) or A.N. (nagy@lunenfeld.ca).

METHODS

SNP array 6.0 analysis. Sample handling and hybridization were performed as previously described[13]. All human ES cell line analysis files, with the exception of CA1 and CA2, were obtained from a previous study[13]. For detecting CNVs and genotype calls, the Affymetrix Genotyping Console 3.0.2 and the Birdseed (v2) algorithm were used, respectively. CNV locations are based on the human genome assembly of March 2006 (NCBI36/hg18). Samples were normalized to 40 International HapMap samples hybridized on the same platform to decrease technical variation (refer to Supplementary Table 11 for HapMap sample codes and SNP profiles)[13]. For CNV calls, regional GC correction, 10-kilobase (kb) size cut-off value, and a minimum of ten markers were used as analysis configurations. All of the array samples passed quality control requirements, having contrast QC (quality control) and MAPD (median absolute pairwise difference) values within the boundaries (Supplementary Table 12). All identified CNVs were included, except for CNVs spanning centromeric regions (the average marker distribution within these regions is too large (>40kb)) and the Y chromosome in female samples, which was considered as false positive and excluded from the analysis. R (v2.9.2) software and the program Microsoft Excel 2008 (v12.2.3) were used for *in silico* data analysis and CNV data parsing. R and StatPlus for Microsoft Excel (v5.8.3.8) were used for statistical analysis and *P*-value calculations.

Cell culture. Human fibroblast lines were cultured in 10% FBS (PromoCell) and GlutaMAX in DMEM (Gibco). Human iPS cells were cultured on mitotically inactivated mouse embryonic fibroblasts (MEFs) in KnockOut DMEM supplemented with 20% KnockOut Serum Replacement (Gibco), 0.1 mM 2-mercaptoethanol (Gibco), 1× GlutaMAX (Gibco), 1× non-essential amino acids (Gibco), 1× ITS liquid media supplement (Sigma) and 6 ng ml^{-1} FGF2 (Sigma). Human iPS cells were passaged using 20 U ml^{-1} type IV collagenase (Gibco), approximately every 5 days. Human ES cells were cultured and maintained as previously described[13,44].

Bisulphite sequencing. Bisulphite conversion was carried out on each DNA sample (1 μg) using the EpiTect Bisulfite Kit (QIAGEN). *OCT4* and *NANOG* promoters were amplified using previously published[45] bisulphite-specific primers (Supplementary Table 13) and a PCR protocol consisting of an initial 1-min denaturation step followed by 35 cycles of 95 °C for 15 s, 54 °C for 30 s and 72 °C for 45 s. The resultant PCR product was sequenced using either the appropriate forward primer or the reverse primer at the Centre for Applied Genomics (Toronto). At CG dinucleotides, cytidine–guanine was scored as methylated CG, whereas thymidine–guanine was considered to be an unmethylated CG. Ambiguous CGs were scored using control fibroblasts as a methylated reference.

Splinkerette PCR and quantitative PCR. Genomic DNA was extracted using the GenElute Mammalian Genomic DNA Miniprep kit (Sigma). Splinkerette PCR was performed as described previously[46]. Splinkerette primers are listed in Supplementary Table 13, and the start position and location of viral integration sites are listed in Supplementary Table 14. Total RNA was extracted using a NucleoSpin RNA II kit (Macherey-Nagel), with on-column DNase treatment. The amount of RNA was quantified using a Nanodrop (NanoDrop Technologies), and RNA was separated on 1% agarose gels to check its quality. Highly pure RNAs were reverse transcribed using the QuantiTect Reverse Transcription Kit (QIAGEN) as per the manufacturer's protocol. Supplementary Table 13 lists all PCR-amplified genes and CNVs and their corresponding primers. Annealing temperatures of 55–58 °C were used for most primers. For quantitative PCR (Q-PCR), we used LuminoCt SYBR Green qPCR ReadyMix (Sigma), a JANUS automated liquid handling robot (PerkinElmer) and the CFX384 real-time PCR detection system (Bio-Rad).

False discovery estimation and CNV validation. The false positive estimate for the samples was studied by hybridizing three HapMap samples in four replicates (Supplementary Table 7). By using analysis settings identical to those for the main data, we found that, on average, 76.2% of total CNV size was detected in all four replicates, 15.4% in three, 4.8% in two and 3.6% only in one of the replicates. By contrast, for homozygous deletions, no CNVs were detected in only one replicate, indicating very low or negligible false positive detection for homozygous deletions. These values are analogous to those from an earlier study[47]. For further validation of CNVs, CNVs from three ES cell lines were also confirmed by running the same samples on an Illumina Human 610-Quad Chip platform. The CNVs from the Illumina data matched 75% (on average) of the CNVs observed in the Affymetrix data (Supplementary Table 5). The Illumina Data were analysed for log Bayes factors greater than 10 using QuantiSNP software (http://www.well.ox.ac.uk/QuantiSNP). Q-PCR was also used to validate some of the discovered CNVs and to estimate the false discovery rate (see Supplementary Fig. 7 for details).

Human iPS cell generation. Human foreskin fibroblasts (HFFs; CRL-2429, ATCC) and human lung embryonic fibroblasts (IMR90; CCL-186, ATCC) were reprogrammed to human iPS cells as previously described[1]. Briefly, retroviral constructs—pMXs–OCT4, pMXs–SOX2, pMXs–KLF4, pMXs–NANOG and pMXs–LIN28—were obtained by cloning the human cDNA encoding each of the factors into the pMXs retroviral vector. pMXs constructs were transfected separately into the 293-GPG packaging cell line[48] (10^6 cells per 100-mm-diameter culture dish) to produce retroviral supernatant. Fibroblast lines, seeded overnight, were infected twice with different, but equally mixed, combinations of viral supernatants (0.5 ml each supernatant, 4×10^5 cells per 60-mm-diameter dish), over the course of 2 days (see Supplementary Table 2 for the different combinations). The following day, the medium was changed to fibroblast medium. On day 4, infected cells were collected and reseeded on mitotically inactivated MEFs. The next day, the medium was changed to human ES cell medium containing FGF2 as described elsewhere[44]. Medium was replenished every 2 days. At 20–30 days post transduction, depending on colony size, colonies with human ES-cell-like morphology were picked and expanded for further analysis. For the new *piggyBac*-transposon-generated human iPS cell lines, HFF cells were seeded in 60-mm-diameter plates at a density of 4×10^5 cells per plate. After 24 h culturing, cells were trypsinized, and they were then electroporated using a 100-μl tip and program number 20 in the Neon Transfection System (Invitrogen) with 250 ng each transposon construct[3], 500 ng PB-rtTA construct[4] and 500 ng pCyL43 PB transposase plasmid[3]. After 24 h, the medium was supplemented with doxycycline (day 0) and was then changed to human ES cell medium at 48 h after transfection. Cells were fed every 2 days with doxycycline-containing medium (1.5 μg ml^{-1}) for 20–30 days. Doxycycline was removed one passage after picking human iPS cell clones. Human iPS cell colonies were picked and cultured as described above for retrovirus-derived human iPS cells. For sample collection and genomic DNA extraction, cells were scraped in collagenase or dispase (1 mg ml^{-1}) and centrifuged twice at low speed to pellet the cells as small colonies and remove the majority of MEFs, which remain as single cells in suspension and are aspirated with the medium.

Pluripotent stem cell differentiation. For embryoid body formation, the cells were detached by collagenase IV treatment and plated onto ultra-low attachment dishes (Corning) in human ES cell medium without FGF2. The culture medium was changed every 3 days. After 10 days, the embryoid bodies were collected for further analysis. Teratomas were generated as described elsewhere[44].

For endodermal differentiation, cells were differentiated as described elsewhere[49]. In brief, 80–90% confluent cells were cultured on a mitotically inactivated MEF layer for 24 h in RPMI 1640 medium (Gibco) supplemented with GlutaMAX, 100 ng ml^{-1} recombinant human activin A (provided by M. Hyvönen) and 10% (v/v) WNT3A-conditioned medium (DMEM supplemented with 10% (v/v) KnockOut Serum Replacement and GlutaMAX, conditioned for 7 days on L Wnt-3A cells (ATCC)). The cells were cultured for another 2 days in RPMI 1640 with GlutaMAX, 100 ng ml^{-1} activin A and 0.2% (v/v) FBS to the definitive endoderm (DE) stage. DE-stage cells were then cultured for 3 days in RPMI 1640 supplemented with GlutaMAX, 2% (v/v) FBS and 50 ng ml^{-1} KGF (R&D Systems) to the primitive gut tube (PG) stage. The cells were cultured for another 3 days with DMEM supplemented with GlutaMAX, 1% (v/v) B-27 supplement (Gibco), 2 μM all-*trans* retinoic acid (Sigma), 0.25 μM KAAD-cyclopamine (Toronto Research Chemicals) and 50 ng ml^{-1} noggin (R&D Systems) to the posterior foregut (PF) stage. Finally, the cells were cultured for another 3 days in DMEM supplemented with GlutaMAX and 1% (v/v) B-27 supplement to the pancreatic endoderm (PE) stage. The medium was changed every day, and RNA samples were collected at the end of every stage for Q-PCR and immunocytochemistry.

For neuronal differentiation, cultured cells were detached with type IV collagenase and transferred as small colonies in a 1/1 ratio to ultra-low binding six-well plates (Costar) in NSE medium (Euromed medium supplemented with sodium pyruvate (Gibco), GlutaMAX, N-2 supplement (Gibco), B-27 supplement, 25 μg ml^{-1} human insulin (Sigma), non-essential amino acids, 0.1 mM 2-mercaptoethanol and 0.05% (v/v) BSA (Gibco)). After 6 days in suspension culture, the spheres were transferred onto plates coated with 1/100-diluted growth-factor-reduced Matrigel (BD Biosciences) in a 1/1 ratio of NSE and NB medium. NB medium consists of neurobasal medium (Gibco) supplemented with GlutaMAX, non-essential amino acids, 2% (v/v) B-27 supplement, 2 μg ml^{-1} heparin (Sigma), 0.1 mM 2-mercaptoethanol and 0.05% (v/v) BSA. The cells were cultured for another 10 days, and the medium was changed every other day. The cells were then immunostained for βIII-tubulin and nestin.

Immunocytochemistry. Samples were washed with PBS and fixed in 4% paraformaldehyde (Electron Microscopy Sciences) for 15 min at room temperature. After three washes in PBS, cells were permeabilized in 0.2% Triton X-100 in PBS for 12 min and were subsequently washed three times with PBS. Samples were then blocked with Protein Block for 10 min, washed three times with PBS and incubated with primary antibodies overnight at 4 °C. The next day, cells were washed twice with Tween-20–PBS and twice with PBS. Secondary antibodies—Alexa Fluor 594 anti-goat IgG or Alexa Fluor 488 anti-rabbit IgG (both from Invitrogen)—were diluted 1/500 in 0.2% Triton X-100 in PBS, and cells were incubated with antibodies for 30 min at 4 °C. Primary antibodies were anti-NANOG (Santa Cruz Biotechnology), anti-OCT4 (Santa Cruz Biotechnology), anti-brachyury (Santa Cruz Biotechnology), anti-FOXA2 (Santa Cruz Biotechnology), anti-SOX17 (Santa Cruz Biotechnology), anti-TRA-1-60 (Millipore), anti-βIII-tubulin (R&D Systems), anti-PDX1 (Beta Cell Biology Consortium), anti-NKX6.1 (Beta Cell Biology Consortium) and anti-nestin (Chemicon) antibodies.

Three-dimensional FISH. Human iPS cells were cultured on glass slides seeded with MEF feeder cells. Samples were fixed in 2% paraformaldehyde in PBS for 5 min, washed three times with PBS, permeabilized with 0.5% Triton X-100 in PBS for 20 min, and washed three more times with PBS. The slides were then placed in a solution of 20% glycerol in PBS overnight at 4 °C. Slides were frozen in liquid nitrogen, allowed to partly thaw and then placed back into the 20% glycerol solution. This process was repeated five times. After the freeze–thaw procedure, the slides were washed three times in PBS and then placed in a solution of 0.1 M HCl for 5 min. Slides were then washed with 2× SSC and left overnight at 4 °C in a solution of 50% formamide in 2× SSC. Before hybridization, the slides were denatured in a solution of 70% formamide in 2× SSC at 75 °C for 3 min and then immediately placed in a separate container containing the same denaturation solution that had been kept on ice. Control (Bac clone RP11-788E9) and test (Bac clone RP11-58E1) probes were obtained from the Centre for Applied Genomics (Toronto). Test and control probe region coordinates were chr1: 146,828,351–147,150,258 and chr1: 104,629,600–104,808,778, respectively, based on the human genome assembly of March 2006 (NCBI36/hg18). The test probe was selected based on a cluster of CNVs consisting of mainly deletions within a frequently affected region in chromosome 1 (coordinates Chr1: 145,797,568–147,958,358). The probes were directly labelled with either spectrum green or orange fluorophore-conjugated nucleotides. A hybridization mixture consisting of labelled probe and human Cot-1 DNA in a 2/1 ratio in hybridization buffer (50% formamide, 10% dextran sulphate, 50 nM sodium phosphate buffer, pH 7.0, in 2× SSC) was prepared and denatured at 80 °C for 5 min and then allowed to partially reanneal at 37 °C for 20 min. This mixture was then applied to the slides that had been kept on ice during the previous step and left to hybridize overnight at 37 °C. After hybridization, the slides were washed in 50% formamide in 2× SSC three times at 42 °C, then once in a solution of 0.5× SSC at 60 °C, and finally in a solution of 2× SSC at room temperature. Slides were mounted with VECTASHIELD containing DAPI (Vector Laboratories) before fluorescence imaging. Images were collected using an IX81 inverted brightfield microscope (Olympus) equipped with a Cascade 512 camera (Photometrics) using a ×60, 1.32 NA, oil-immersion objective and Immersion Oil Type DF (Cargille Labs) imaging medium. Images were collected using MetaMorph Premier 7.7 (Molecular Devices) and analysed with ImageJ (National Institutes of Health).

47. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444,** 444–454 (2006).
48. Ory, D. S., Neugeboren, B. A. & Mulligan, R. C. A stable human-derived packaging cell line for production of high titer retrovirus/vesicular stomatitis virus G pseudotypes. *Proc. Natl Acad. Sci. USA* **93,** 11400–11406 (1996).
49. Kroon, E. *et al.* Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells *in vivo. Nature Biotechnol.* **26,** 443–452 (2008).