Published in final edited form as:

Nat Biomed Eng. 2017 October ; 1(10): 826–837. doi:10.1038/s41551-017-0141-6.

Transcriptomic and epigenomic differences in human induced pluripotent stem cells generated from six reprogramming methods

Jared M. Churko^{1,2,3}, Jaecheol Lee^{1,2,3}, Mohamed Ameen^{1,2,3}, Mingxia Gu^{1,2,3}, Meenakshi Venkatasubramanian⁴, Sebastian Diecke^{1,2,3}, Karim Sallam^{1,2,3}, Hogune Im⁴, Gavin Wang⁵, Joseph D. Gold¹, Nathan Salomonis⁴, Michael P. Snyder⁵, and Joseph C. Wu^{1,2,3,*}

¹Institute of Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA

²Stanford Cardiovascular Institute, Stanford University School of Medicine, Stanford, CA 94305, USA

³Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA

⁴Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA

⁵Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA

Abstract

Many reprogramming methods can generate human induced pluripotent stem cells (hiPSCs) that closely resemble human embryonic stem cells (hESCs). This has led to assessments of how similar hiPSCs are to hESCs, by evaluating differences in gene expression, epigenetic marks and differentiation potential. However, all previous studies were performed using hiPSCs acquired from different laboratories, passage numbers, culturing conditions, genetic backgrounds and reprogramming methods, all of which may contribute to the reported differences. Here, by using high-throughput sequencing under standardized cell culturing conditions and passage number, we compare the epigenetic signatures (H3K4me3, H3K27me3 and HDAC2 ChIP-seq profiles) and transcriptome differences (by RNA-seq) of hiPSCs generated from the same primary fibroblast population by using six different reprogramming methods. We found that the reprogramming method impacts the resulting transcriptome and that all hiPSC lines could terminally differentiate, regardless of the reprogramming method. Moreover, by comparing the differences between the

Reprints and permissions information is available at www.nature.com/reprints.

*Correspondence and requests for materials should be addressed to J.C.W. joewu@stanford.edu.

Author contributions

J.D.G., N.S., M.P.S. and J.C.W supervised and planned the project. J.M.C. wrote the manuscript, performed data analysis, generated and cultured hiPSC lines, and performed RNA-seq. N.S. and M.V. performed integration analysis. H.I. helped analyse RNA-seq. J.L. performed ChIP-seq experiments. M.A. and M.G. performed FACS analysis on differentiated cardiomyocytes. G.W. and K. S. helped to culture hiPSC and hESC lines. S.D. generated minicircle hiPSC lines.

Competing interests

The authors declare no competing financial interests.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41551-017-0141-6>.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

hiPSC and hESC lines, we observed a significant proportion of differentially expressed genes that could be attributed to polycomb repressive complex targets.

The introduction of four genes (*OCT4*, *SOX2*, *KLF4* and *MYC*; often referred to as the Yamanaka factors)¹ into somatic cells has been shown to reprogram cells into hiPSCs. The expression of these factors results in dramatic changes in the epigenetic and transcriptomic signatures of the cells being reprogrammed. Human embryonic stem cells (hESCs) are viewed as the gold standard to model pluripotency. For human induced pluripotent stem cells (hiPSCs) to be studied further and used clinically, the transcriptomic and epigenetic status of these cells should be as close as possible to those of hESCs.

Previous studies have compared hESCs with hiPSCs, and have noted that while hiPSCs are very similar to hESCs, hiPSCs still express a unique set of genes not observed in hESCs^{2–6}. Further studies comparing gene expression values in multiple stem cell lines observed that hESCs clustered together and hiPSCs also clustered together, suggesting that these two populations are distinct. In addition, the clustering of hiPSCs depended on the lab in which the hiPSCs were generated^{7,8}. This is intriguing, because the culture conditions and reprogramming method used to generate hiPSCs may be the driving factor that leads to gene expression differences in hiPSCs⁹. Gene expression and clustering may also be influenced by the genetic background of each hiPSC, because single-nucleotide polymorphisms (SNPs) and copy-number variations (CNVs) can also influence gene expression¹⁰.

To determine if the reprogramming method influences the gene expression profile, we performed RNA sequencing (RNA-seq) on twelve cell lines generated by six frequently used reprogramming methods: episomal vectors, mRNA, microRNA/mRNA transfections (microRNA), minicircle vectors, lentivirus and Sendai virus. Because the sex, age, and genetic background of the parental cells could influence the resulting gene expression profile of the hiPSCs generated, we used the same female fibroblast population for each reprogramming method. In addition, to mitigate any transcriptomic differences resulting from culture conditions, we have standardized our hiPSC growing conditions: all lines were grown on Matrigel using chemically defined E8 conditioned media. Moreover, since the passage number can influence gene expression (and the promotion of genetic abnormalities), each line was grown simultaneously, equally passaged, and RNA-seq and chromatin immunoprecipitation sequencing (ChIP-seq) were performed to compare the epigenetic and transcriptomic differences between hiPSCs generated by each reprogramming method. We demonstrate that the reprogramming method may influence the gene expression profile, as the clones from the same method cluster together. Most of the gene differences could be ascribed to an imbalance in polycomb repressive complex (PRC) factors, and may be related to different states of the reprogramming process.

Results

The reprogramming method influences the hiPSC transcriptome profile

Although each reprogramming method uses the same core transcription factors (*OCT4*, *SOX2*, *KLF4* and *c-MYC*), each method differs slightly, and these differences may impact

the resulting hiP-SCs' transcriptome (Table 1). To address this, we reprogrammed the same female fibroblast population using six reprogramming methods and picked two hiPSC clones from each method for further analysis. To verify the generation of bona fide hiPSC lines, each hiPSC line was analysed via immunofluorescence for the expression of known hiPSC markers (Supplementary Fig. 1) and injected into NOD/SCID mice to assess for teratoma formation (Supplementary Fig. 2). As expected, each line expressed hiPSC markers and generated tumours expressing endodermal, mesodermal and ectodermal derivatives. In addition, SNP karyotyping determined that no major karyotypic abnormalities were observed in each cell line (Supplementary Fig. 3).

Next, to determine if the reprogramming method leads to differences in the resulting hiPSC transcriptome, we performed an initial screen with the NHLBI Progenitor Cell Biology Consortium Cell (PCBC) Cincinnati Characterization Core. From this large bank of sequenced hiPSCs, we performed a principal component analysis (PCA) on the gene expression levels from 44 hiPSC lines sequenced from four different reprogramming methods, as well as H9 hESCs (Fig. 1a). In general, the H9 hESCs clustered separately from the hiPSC lines, and the hiPSC lines clustered closer to the lines generated from a similar reprogramming method. However, since these hiPSC lines were made in multiple labs from different somatic cell sources and were analysed at different passage numbers and under distinct growth conditions, we sought to determine if control over each of these aspects (growth media and extracellular matrix attachment, reprogramming environment, and passage) could influence the resulting hiPSC transcriptome and thus the clustering of hiPSC lines. RNA-seq was performed on passage 12 (p12) hiPSCs, because at this time point the nonintegrating Sendai viral vectors would have been cleared from within the cell¹¹. Hierarchical clustering and PCA of gene expression in H7 (p35), H9 (p55), ES02 (p71), HUES9 (p28), LSJ1 (p38) and LSJ2 (p21) hESCs, in the parental fibroblast line (passage 4), as well as in two lines generated from each reprogramming method (p12), show similar clustering of hiPSCs generated from the same reprogramming method (Fig. 1b,c). In addition, cell lines from similar reprogramming techniques (primarily electroporation and transfection-based methods) also clustered closer together, as evident by the clustering pattern of hiPSC clones generated by transfection-based techniques (mRNA and microRNA), viral-infection-based techniques (Sendai and lentivirus) or electroporation-based techniques (minicircle and episomal). However, viral-generated clones were least similar to each other, which may be due to differences in viral load or transgene expression. We therefore mapped our RNA-seq reads to the vectors used to create lentiviral lines, and tested for the presence of Sendai virus within the Sendai-viral-derived lines using real-time PCR (rtPCR). We observed persistent lentivirus-specific transgene expression within both lentiviral clones (Supplementary Fig. 4A). We also assessed for transgene expression within our episomal lines but did not find significant evidence to suggest that episomal integration occurred within the episomal-derived lines (Supplementary Fig. 4B). In addition, using Sendai-viral-specific primers, we observed that the Sendai virus was still present in Sendai virus clone 2 (Sendai_CL2) (Supplementary Fig. 4C). Spearman correlation between each line shows that the gene expression values between hiPSC lines were highly correlated to each other and that hiPSC lines were highly correlated with hESCs (0.86–0.94), with a lower correlation to the parental fibroblasts (0.73–0.80) (Fig. 2a). When assessing which

reprogramming techniques produced hiPSCs with a more similar gene expression profile to hESCs, Spearman correlation values were the highest for viral-derived hiPSC clones (lentivirus, $r = 0.96$ and Sendai virus, $r = 0.95$) (Fig. 2b). RPKMs (reads per kilobase of transcript per million mapped reads) of each Ensembl ID were calculated using AltAnalyze (Supplementary Table 1). Significant gene expression differences (Bayes moderated t -test P value (unpaired) assuming unequal variance, $P < 0.05$ with 2-fold difference) (Supplementary Table 2), and significant splicing events (687 events) between hiPSCs and hESCs (Supplementary Fig. 5 and Supplementary Table 3) were also calculated. We graphed the number of genes called as significantly different, and Spearman correlation values between each reprogramming method, reprogramming technique (virus-, transfection- and electroporation-based techniques), as well as hiPSCs compared to hESCs (Fig. 2a–c). We also identified genes highly enriched (markers) in each type of reprogramming method (Fig. 2d). Given the possibility that sodium butyrate, which is included in some reprogramming methods, could lead to the reprogramming-specific differences observed, we sought to assess whether treatment of hiPSCs with sodium butyrate does indeed lead to reprogramming-method-specific gene changes. However, genes activated in hiPSCs cultured for five passages with sodium butyrate were not specific to the lines reprogrammed using sodium butyrate (Supplementary Table 2). We further combined our RNA-seq reprogramming-specific gene list with RNA-seq data generated from the PCBC to identify genes that were consistently identified as reprogramming-specific (Supplementary Table 2). Furthermore, we assessed the expression of selected genes (*ZNF732*, *PRPS1*, *AMFR* and *CXXC5*) that were identified as significant between lentiviral and Sendai-viral-derived lines in 17 independent hiPSC lines generated within our laboratory and another laboratory by rtPCR (Supplementary Fig. 6). We validated *ZNF732* to be significantly enriched within lentiviral-derived lines and *AMFR* within Sendai-derived lines (Supplementary Fig. 6B). By comparing the gene expression levels in hiPSC lines to hESCs, 562 protein-coding genes were found to be significantly different between hiPSCs and hESCs (Fig. 2c). Similarly, we compared the genes called as significantly different between hiPSCs and hESCs in the PCBC dataset, in our dataset, as well as in a study where genetically matched hiPSCs were reprogrammed from differentiated hESCs, RNA-sequenced, and compared⁵. We found that three overlapping genes (*DDIT4*, *BHLHE40* and *SLC2A1*) within each of these datasets were identified as significantly different between hESCs and hiPSCs.

Pathway and GO analysis identifies that PRC/HDAC2 activity and other core transcriptional networks are different between hESCs and each reprogramming method

To determine which pathways are significantly enriched between hiPSC and hESCs lines, we performed gene ontology (GO) analysis on the 562 genes that are significantly different between hiPSCs and hESCs (Fig. 3a), as well as between the type of reprogramming method and hESCs (Supplementary Fig. 7A–D). Significant GO molecular functions were cytoskeletal-protein binding (hypergeometric test, $P = 2.59 \times 10^{-5}$) and troponin-T binding (hypergeometric test, $P = 8.49 \times 10^{-5}$). As for GO biological function, the most significant pathways were: establishment of protein localization to the plasma membrane, cellular macromolecular-complex assembly, protein-complex subunit organization, and plasma-membrane organization (hypergeometric test, $P = 9.18 \times 10^{-7}$, 1.09×10^{-6} , 2.86×10^{-6} and 3.35×10^{-6} , respectively).

To determine if the differences in gene expression between each reprogramming method are due to differential transcription factor binding activity, we next performed a transcription-factor/DNA binding-target enrichment analysis on the gene expression differences, and found targets within the *SOX2/NANOG* network as well as within the polycomb repressive complex PRC (Supplementary Fig. 7E,F) network to be significantly different between hiPSCs and hESCs. For the enriched transcription/DNA binding-factors analysis, we report the quantification of the number of genes associated with each transcription/DNA binding factor as a percentage of the total number of differentially expressed genes. We found significant enrichment of *YY1*, *SUZ12*, *ELK1*, *NRF1*, *MYC:MAX*, *SOX2*, *NANOG*, *EED*, *EZH2*, *LEF1*, *HIF1a*, *TCF3* and *HDAC2* activity (Fig. 3b). Interestingly, interactions within the PRC2/HDAC2 network could account for over 35% of the differentially expressed genes between hiPSCs and hESCs.

Interplay between HDAC2 ChIP-seq binding and MYC targets in hESCs

PRC2/HDAC2 activity may be a critical component in gene expression differences between hiPSCs and hESCs. Histone deacetylases (HDACs) initiate silencing of genomic regions by removing acetyl groups and the PRC methylates Lys27 at histone H3. The PRC continually methylates until it is trimethylated at H3 at Lys27 (H3K27me3). Thus, *HDAC2* activity and H3K27me3 marks suggest transcriptionally silent chromosomal regions. Given that targets of the PRC/HDAC2 pathway are significantly different between hiPSCs and hESCs, we sought to understand how *HDAC2* activity might be playing a role in differences between hiPSCs and hESCs. ChIP-seq using anti-HDAC2 was performed on five hESC and five hiPSC lines, and MACS (model-based analysis of ChIP-seq) was used to identify differential peaks between hiPSCs and hESCs (Fig. 3c). MACS found 809 differential peaks unique to the hESCs and 370 peaks unique to hiPSCs corresponding to *HDAC2* localization (Supplementary Table 4). GREAT (genomic regions enrichment of annotations tool) analysis performed on hESC enriched peaks strongly suggest that these peaks were also associated with *MYC* activity, since these regions corresponded to *MAX:MYC* associated factor X (ID: V\$MYCMAX_03) from MSigDB predicted promoter motifs (hypergeometric test, $P = 9.7912 \times 10^{-5}$) as well as 'Cluster 6: genes down-regulated in B493-6 cells (B lymphocytes) by *MYC* (GeneID = 4609)' from the MSigDB Perturbation database (hypergeometric test, raw P value 3.7981×10^{-3}) (Fig. 3d). In addition, *MYC* expression (by RNA-seq) (Fig. 3e) as well as genes regulated by *MYC* (Fig. 3f) was observed to be significantly higher in hESCs versus hiPSCs. This suggests that *MYC* and *HDAC2* activity facilitates and contributes to the differential epigenetic differences between hESCs and hiPSCs.

Various reprogramming methods generate highly similar yet unique hiPSC epigenetic H3K27me3 and H3K4me3 peak profiles when compared to hESCs

To compare the epigenetic similarities or differences between our hiPSC and hESC lines, we next performed ChIP-seq on each hiPSC and hESC line using an antibody against a histone mark strongly correlated with gene activity (H3K4me3), as well as a histone mark negatively correlated with gene activity (H3K27me3)¹². Since H3K27me3 and H3K4me3 marks are more likely to affect gene expression when they are localized within 2 kb of the transcriptional start site (TSS), we quantified the total number of reads within ± 2 kb of the TSS for all annotated genes, normalized the read count per million mapped reads (peak

density), and normalized the peak density to that of the control H3 ChIP-seq profiles from the hiPSC and hESC lines. Spearman correlation on the $-\log$ peak density values of each sample shows a high correlation between the hiPSC lines in the H3K4me3 and H3K27me3 datasets (Fig. 4a). We identified 251 significant TSS peak density differences within the H3K4me3 ChIP-seq set, and 281 significant TSS peak density differences in the H3K27me3 ChIP-seq profile between hESCs and hiPSCs (Fig. 4b and Supplementary Tables 5 and 6). Intriguingly, whereas the majority of differences observed within the H3K4me3 profile were distributed throughout each chromosome, the vast majority of H3K27me3 peaks were identified on the X chromosome within the hiPSC lines (Fig. 5a). To investigate if X inactivation could be different among our hiPSC lines, we then identified the levels of the X-chromosome-encoded long non-coding RNA *XIST* (Fig. 5b). We note that hESCs (combined coverage) possess low levels of active H3K4me3 mark in the promoter region of *XIST*; however, high levels could be observed in hiPSC lines. In addition, low levels of *XIST* expression were found in the hESCs, but were abundant in hiPSC lines. Interestingly, when we assessed the levels of *XIST* per clone, one line (microRNA_CL2) did not express *XIST* (Fig. 5c). In addition, we assessed the levels of *XIST* at passages 12 and 40 by rtPCR. Four hiPSC lines at passage 40 did not express *XIST*, and in the remaining lines expressing *XIST*, we observed a 10-fold decrease in the levels of *XIST* (Fig. 5d). Finally, we performed a gene-set enrichment analysis (GSEA) for transcription factors that are known to regulate the genes found to be significantly different from our epigenetic TSS peak density analysis (Fig. 5e) between hiPSCs and hESCs. Here, *Tcf3* and *LEF1* binding sites were enriched within the H3K4me3 TSS peak density profile, and *YY1* and *LEF1* within the H3K27me3 ChIP-seq profile, when comparing hESCs to hiPSCs peak differences.

Cardiac differentiation potential of hiPSCs from multiple reprogramming methods

Differential H3K27me3 ChIP-seq and *XIST* profiles within our hiPSCs suggests that hiPSCs possess unique epigenetic profiles. In addition, our transcription factor enrichment analysis suggests that activity of the PRC is differentially regulated. It has previously been shown that disruption of a member of the PRC (Ezh2) in mice leads to cardiac abnormalities¹³. To assess whether these detected differences can functionally impact our hiPSCs, we next performed cardiomyocyte differentiation on all of our lines with equal seeding density, reagent stock concentrations, passage, and culture conditions in parallel (Supplementary Fig. 8), and performed fluorescence-activated cell sorting (FACS) of the cardiomyocyte-specific marker cardiac troponin-T after differentiation. All hiPSC lines were capable of differentiating into cardiomyocytes, with a very similar efficiency regardless of the method of reprogramming (Fig. 6a,b). However, the microRNA_CL2 produced more cardiac troponin-T-positive cells than the Sendai_CL1 lines. Next, to determine if the gene expression of hiPSCs could correlate with the observed differences between the hiPSC lines differentiation potential, we performed supervised hierarchical clustering of all lines with the genes known to play a role in cardiac differentiation from the Cardiovascular Gene Ontology Annotation Initiative (Fig. 6c). Although the hiPSCs clustered together and the hESCs clustered together, microRNA_CL2 did not cluster closer to the hESC lines. To determine if epigenetic differences can be used to link a hiPSC line with differentiation potential, we clustered the TSS peak density profile of genes from the Cardiovascular Gene Ontology Annotation Initiative from the H3K27me3 ChIP-seq data. While hiPSCs clustered together

and hESCs clustered together, clustering within hiPSCs was observed to be more dependent on the reprogramming method. However, numerous H3K27me3 differences between hiPSCs and hESCs (23 genes, dotted line) were identified in the TSS region of genes critical for cardiac differentiation (that is, *TBX5* and *GATA6*) (Fig. 6d), suggesting that epigenetic differences on cardiac genes may impact the ability of hiPSCs to differentiate into cardiomyocytes. This is supported by observing variable H3K27me3 peak density values in these cardiac genes between the two hiPSC lines that demonstrated the largest differences in our cardiomyocyte differentiation potential assay ($74.94 \pm 4.95\%$ microRNA_CL2 versus $43.57 \pm 7.50\%$ Sendai_CL1 TNNT2 expressing cells) (Fig. 6e). Moreover, in performing a GSEA for transcription factors that may regulate different H3K27me3 cardiac gene TSS profiles between hiPSCs and hESCs, activity of *TCF3* was highly enriched (CAGGTG_V \$E12_Q6 $P = 1.926 \times 10^{-4}$). Complementary to this observation, the gene expression of *TCF3* (by RNA-seq) is significantly higher in the hESCs compared to hiPSCs (Fig. 6f). In addition, this gene set was also enriched for PRC targets (hypergeometric test, EED targets $P = 2.288 \times 10^{-20}$, PRC2 targets $P = 1.283 \times 10^{-16}$ and Suz12 targets $P = 2.194 \times 10^{-15}$).

Discussion

With the discovery of hiPSCs, various reprogramming methods have been developed, which differ in efficiencies, cost, and safety concerns (non-integrative versus integrative, DNA-based versus RNA-based). Of these methods, Sendai-virus-based reprogramming remains the best method to generate hiPSCs, given that this method is non-integrating and has been shown to yield little to no epigenetic and transcriptome differences¹⁴ or genetic changes¹⁵. However, when comparing the transcriptome of multiple hiPSC lines generated by multiple reprogramming methods, hiPSC lines were observed to cluster more similar to the lab in which each line was derived^{7,8}, and the culturing conditions, passage number, genetic background and reprogramming method may have contributed to these transcriptomic and epigenetic differences. Here we have reprogrammed the same fibroblasts with six different reprogramming methods, and standardized the media requirements, plating matrix, passaging enzyme, seeding density, library preparation method, and sequencing conditions. In addition, we have standardized the passage number to passage 12, as hiPSCs may still express the Sendai viral transgenes, early passage hiPSCs have a tendency to spontaneously differentiate at lower passages, and many consortiums are banking hiPSCs at passage 12 to ensure stable pluripotency.

Although all of our lines produced teratomas and expressed the ‘classic’ hiPSC markers, RNA-seq could still identify a high number of protein-coding genes differentially expressed when compared to hESCs (for example, episomal lines). However, when applying a genome-wide transcriptome analysis, hiPSC clones within the same reprogramming method clustered together, and hiPSCs made by similar reprogramming techniques also showed a closer clustering potential. When analysing the transcriptome differences compared to hESCs, we determined that the viral hiPSC lines expressed a more similar transcriptome, and that the transfection-based reprogramming lines were the least similar. Given that the passage number was standardized, the differences observed between these methods could be due to the more potent ability of the lentiviral and Sendai virus methods to reprogram fibroblasts, therefore resulting in gene expression levels reaching an equilibrium state earlier

than required for the other methods. Since lentivirus integrates into the host genome and Sendai virus has been noted to persist within hiPSC lines up to passage 12, this prolonged exposure may be promoting the rapid reprogramming process.

The genetic background of hiPSCs is known to contribute to gene differences reported between hiPSCs and hESCs¹⁶. Our study minimized this difference by using twelve hiPSC lines generated from the same fibroblast population. We identified potential lentiviral, Sendai virus, plasmid and mRNA-based gene differences between our lines, and identified overlapping reprogramming-specific genes found within the PCBC RNA-seq data. Furthermore, we compared our RNA-seq data and the PCBC RNA-seq data for genes differentially expressed between hiPSCs and hESCs with those genes in genetically matched hESCs and hiPSCs⁵. Within all three datasets, three genes were commonly called as being different between hiPSCs and hESCs (*DDIT4*, *BHLHE40* and *SLC2A1*). Interestingly, *DDIT4*, *BHLHE40* and *SLC2A1* all play a role in energy metabolism and the hypoxic response^{17–19}; however, further effort will be required to understand the significance behind these gene differences.

Our study suggests that hiPSCs possess unique epigenetic signatures. This is corroborated by our identification of PRC and *HDAC2* targets that are differentially regulated between hiPSCs and hESCs. HDACs would be recruited to deacetylate a target histone, and in combination with a transcription factor (*YY1* and *MYC*) and the PRC2 (*EED*, *SUZ12* and *EZH2*) transcriptional silencing (methylation) of the target site would occur. *MYC* binding with *HDAC2* is not only demonstrated in our study by co-occupancy of *HDAC2* with *MYC* binding regions but has also been supported by other work^{20,21}. In addition, PRC2 and HDAC inhibitors have been demonstrated to promote hiPSC generation^{22,23}, and prolonged *HDAC2* inhibition within the first couple of passages could also promote a more rapid reprogramming of hiPSCs long after first-passage colony isolation.

Because it is likely that the gene differences that we observed in each of our reprogramming methods may be at least partly due to epigenetic memory or activated during reprogramming²⁴, we also sought to perform ChIP-seq on a histone modification highly correlated with gene activity (H3K4me3) and on a histone modification negatively correlated with gene activity (H3K27me3). Consistent with our RNA-seq findings, we observed differences in these H3K4me3 and H3K27me3 marks in the TSS region of genes corresponding to genes regulated by *TCF3* and *LEF1*. In addition, we observed high levels of XIST in lower-passage hiPSCs, whereas XIST expression was decreased in higher passage (p40) hiPSCs. This indicates that epigenetic changes are still occurring in passage 12 hiPSCs, and thus prolonged culture may promote a more similar hESC epigenetic state. Nevertheless, in assessing whether the epigenetic differences observed between each line could affect hiPSC cardiac differentiation²⁴, we did not observe a significant impact in the differentiation potential between each line. This suggests that the reprogramming method used to generate each line does not play a significant role in influencing the differentiation potential of hiPSCs.

Methods

Fibroblast isolation and culture

Skin punch biopsy (2 × 2 mm) was performed on patients following Stanford Institutional Review Board (IRB) approval and Stem Cell Research Oversight (SCRO) approval at Stanford University and under informed consent. The skin biopsy was manually minced with scissors and incubated in collagenase type IV (Life Technologies, Grand Island, NY, USA) for four hours. Large debris was removed and fibroblasts were centrifuged and resuspended in DMEM with Glutamax (Invitrogen) supplemented with 10% FBS. Fibroblasts were expanded in culture at 37 °C, 20% O₂, and 5% CO₂ in a humidified incubator over two weeks and were frozen at passages 3 and 4 until required for reprogramming.

hiPSC generation

Due to technical difficulties in handling twelve hiPSC lines and six hESC lines in parallel and performing RNA-seq and ChIP-seq using four antibodies (H3, HDAC2, H3K4me3 and H3K27me3), we restricted our hiPSC generation from each reprogramming method to two clones per patient. All hiPSC lines were tested to be mycoplasma negative using the Mycoalert Mycoplasma testing kits (LT07-318, Lonza).

Episomal-based hiPSC generation

Dermal fibroblasts were grown in fibroblast media (DMEM supplemented with 10% FBS and 1% pen/strep) on gelatin-coated T225 flasks. Fibroblasts were dissociated using trypsin and 3 × 10⁶ fibroblasts were reseeded the day before electroporation. On the day of electroporation, one million fibroblasts were electroporated with a total of 10 µg of episomal vectors²⁵ (Supplementary Fig. 9A) using the Neon Transfection System (Invitrogen). Fibroblasts were plated onto a Matrigel-coated (356231, BD Biosciences) 10 cm dish in fibroblast media supplemented with hydrocortisone (Sigma H0396 diluted in DMEM/F12) for three days. Fibroblasts were then switched to essential 7 (E7) media (essential 6 (Invitrogen) with FGF2 (50 µg l⁻¹)) supplemented with sodium butyrate (0.2 mM NaB) (B5887 diluted in DMSO, Sigma) for 13 days. Fibroblasts were then switched to essential 8 (E8) media (Invitrogen) until hiPSC colonies were picked (day 30 after electroporation). Picked colonies were passaged at a 1:10 ratio using Accutase (Invitrogen) and reseeded in E8 containing a 10 µM ROCK inhibitor for the first day (Sigma, Y27632). Each day, hiPSCs were fed with fresh E8 media. All further culturing and passaging of hiPSCs was performed by this method.

Minicircle hiPSC lines

Reprogramming using the minicircle technique was performed as previously described²⁶, but with an optimized minicircle backbone and changes in the reprogramming procedure (Supplementary Fig. 9B). To increase the cell survival after the electroporation, the minicircle DNA was purified using Zymoclean Gel DNA Recovery Kit. One day before the reprogramming experiment, 1 × 10⁶ human fibroblasts were plated into a 10 cm dish. On the following day, cells were trypsinized and electroporated with 12 µg of a newly developed codon optimized minicircle construct CoMiP 4in1 (OKSM_IRES_TOMATO) using the

Invitrogen Neon system (1600 volts, 10 ms and 3 pulses; transfection efficiency should be higher than 50%). Thereafter, the cells were equally distributed onto two Matrigel coated plates (5.5×10^5 cells) in fibroblast media. On the following morning (day 1), the media was changed to fibroblast media containing 0.2 mM NaB and $50 \mu\text{g ml}^{-1}$ ascorbic acid. On day 5, the media was changed to fibroblast media and E7 media in a 50:50 ratio supplemented with 0.2 mM NaB and $50 \mu\text{g ml}^{-1}$ ascorbic acid. After day 7, the media was switched to E7 plus 0.2 mM NaB and $50 \mu\text{g ml}^{-1}$ ascorbic acid. After the first hiPSC-like colonies appeared (day 30), six individual hiPSC clones were picked and expanded in E8 medium.

mRNA and microRNA/mRNA hiPSC lines

Fibroblasts were sent to the Pluripotent Stem Cell Shared Resource Facility at the Icahn School of Medicine for the generation of mRNA and microRNA/mRNA lines. Reprogramming was performed using the Stemgent mRNA Reprogramming System (Stemgent, Cambridge, MA) following the manufacturer's protocol²⁷. Briefly, 50,000 live cells were seeded into a well of a Matrigel-coated 6-well plate and 1 μg of a mixture of GFP and human Oct4, Klf4, Sox2, Lin28 and c-Myc mRNA was transfected using the Stemfect RNA Transfection Kit (Stemgent, Cambridge, MA) daily on days 1–5, as well as days 7–17 following the initial seeding in NuFF-conditioned Plurition Medium. For microRNA lines, 3.5 μl of the microRNA cluster mix was transfected on day 1 and day 4 after initial seeding as previously described^{28,29}. Thirteen clones were picked on day 21 after initial seeding and were frozen at passage 0 (p0). p0 clones were shipped to Stanford and thawed into E8 media on Matrigel-coated plates. One additional round of colony picking was performed and hiPSCs were expanded until p4. At p4, all lines were frozen until all other lines were made.

Sendai virus hiPSC lines

Sendai virus reprogramming was performed according to the Cytotune Sendai Viral reprogramming protocol (Invitrogen). Briefly, fibroblasts were grown in DMEM with 10% fetal bovine serum and were trypsinized the day before infection and 1×10^6 fibroblasts were plated into a Matrigel-coated well of a 6-well plate. The following day, Sendai virions containing *POU5F1*, *SOX2*, *KLF4* and *cMYC* genes were added to the well and the tissue culture plate was spun for 15 min at 800g. For four days after viral infection, fibroblasts were grown in fibroblast media. On day 5 after infection, fibroblasts were cultured in E7 media combined with NaB for 12 additional days. Fibroblasts were switched to E8 media on day 14 after infection and colonies were picked on day 20. After expansion, Sendai viral clones were frozen at p4 until all reprogramming lines were created.

The presence of Sendai virus was detected using the TaqMan iPSC Sendai Detection Kit (A13640, Life Technologies) (SeV: forward: GGA TCA CTA GGT GAT ATC GAG C; reverse: ACC AGA CAA GAG TTT AAG AGA TAT GTA TC) at passage 12.

Lentivirus hiPSC lines

Generation of hiPSCs was performed as previously described³⁰. In brief, 6×10^6 293TN cells were seeded one day before transfection with the packaging plasmids pPAX2 (Plasmid 12260) and pMD2.G (Plasmid 12259) from Addgene, and a codon optimized 4-in-1 lentiviral vector (*OCT4*, *KLF4*, *SOX2*, *c-MYC*, plus td TOMATO) (Supplementary Fig. 9C-

F)³¹ using Lipofectamine 2000 in the presence of 25 mM chloroquine. Supernatants were harvested 24, 48 and 72 h after transfection, filtered, and concentrated using Lenti-X Concentrator (631231, Clontech Laboratories). To generate hiPSC lines, 3×10^4 human fibroblasts were seeded one day before starting the reprogramming into each well of a 6-well dish. The next day, 30 μ l of concentrated virus supernatant were added to one well of a 6-well dish and spin inoculated (700g) for one hour at 37 °C in the presence of transduction media (fibroblast media + polybrene (10 μ g ml⁻¹, sc-134220, Santa Cruz Biotechnology) + 50 μ g ml⁻¹ ascorbic acid). The next day, the media was replaced with fibroblast media + NaB + 50 μ g ml⁻¹ ascorbic acid. Thereafter, the media was changed every other day until day 5, when the cells were harvested and plated at different densities on Matrigel in the presence of E8 supplemented with NaB + 50 μ g ml⁻¹ ascorbic acid. At day 14, hiPSC colonies were visible and these colonies were picked at day 20 to be further amplified and purified.

High-throughput sequencing

RNA-sequencing—Total RNA was isolated using the miRNeasy Micro Kit (Qiagen). 100 ng of RNA was converted to cDNA using the Ovation RNA-Seq System V2 kit (NuGEN). cDNA was fragmented to an average fragment size of 300 bp using the Covaris S2 and sheared cDNA was purified using Agencourt AMPure XP beads. End repair, dA tailing and adapter ligation was performed using the NEBNext Ultra DNA Library Prep with 500 ng sheared cDNA input. Size selection was performed on the Pippin Prep (Sage Science) set to isolate a 300–360 bp fragment. Six bar-coded RNA-seq libraries were pooled and hybridized to one lane of a HiSeq flow cell at the Stanford Stem Cell Institute Genome Center or to Macrogen (South Korea) to obtain 2×100 paired end reads on the Illumina HiSeq 2000 platform at an average depth of 80 million reads per sample.

ChIP-sequencing—5 μ g of H3K4me3 (39159, Active Motif), H3K27me3 (39155, Active Motif), control H3 (61475, Active Motif) and HDAC2 (2545S, Cell Signaling) antibodies were used. Each antibody was incubated with dynabeads (10003D, Invitrogen) for 12 h at 4 °C. A small portion of the crosslinked, sheared chromatin was saved as the input, and the remainder was incubated with the antibody conjugated dynabeads. After overnight incubation at 4 °C, the incubated beads were rinsed with sonication buffer (50 mM HEPES pH 7.9, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS and 0.5 mM PMSF), high salt buffer (50 mM HEPES, pH 7.9, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS and 0.5 mM PMSF) and LiCl buffer (20 mM Tris, pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate and 0.5 mM PMSF). The washed beads were incubated with elution buffer (50 mM Tris, pH 8.0, 1 mM EDTA, 1% SDS and 50 mM NaHCO₃) for 1 h at 65 °C and then de-crosslinked with 5 M NaCl for overnight at 65 °C. The immunoprecipitated DNA was treated with RNase A and proteinase K, and purified by ChIP DNA clean and concentrator (D5205, Zymo Research). Before sequencing, ChIP samples were evaluated by quantitative (qPCR) for the primers to the promoter regions of *NANOG* and *NKX2-5* (Supplementary Fig. 10). 10 ng was used as input DNA into the NEBNext Ultra DNA Library Prep kit to add on sequencing adapters and a fragment size of 420–580 bp was isolated using the Pippin Prep (Sage Science). Multiplexed libraries were sequenced using the Illumina HiSeq 2000 platform. Bowtie was used to map the raw sequencing reads to the reference Ensembl GRCh38

genome. Correlation between each line was performed by counting the number of reads which fell within 2 kb of the transcription start site from H3K4me3, H3K27me3 and H3 ChIP-seq profiles. ChIP-seq raw fastq files are stored under the Gene Expression Omnibus (GEO) accession number GSE69626. The number of reads within each TSS was quantified using Bedtools 2.19.1³² and was normalized to the total number of reads within each sample sequenced (per million mapped reads). No enriched peaks for H3K27me3 samples ES02 and Lentivirus_CL1 were found and these two samples were removed for further analysis. Enrichment of H3K4me3 and H3K27me3 marks was performed by dividing the total number of reads within each TSS from the H3K4me3 and H3K27me3 dataset with the number of mapped reads per million in each sample's corresponding H3 pull down. Graphical representation of the correlation data was performed using R by calculating the Spearman rank correlation coefficients with cor function and using the corrplot package (<https://github.com/taiyun/corrplot>). Circular localization of H3K4me3 and H3K27me3 marks per chromosome was created using Circos³³.

HDAC2 ChIP-seq analysis—Bowtie was used to map the raw sequencing reads to the reference Ensembl GRCh38 genome and differential peak calling was performed using the callpeaks and bdgdiff modules in MACS v2.0.10³⁴. Samples with few peaks called were removed from the analysis. After differential peaks between hESCs and hiPSCs were identified, HOMER 4.6 (<http://biowhat.ucsd.edu/homer/ngs/index.html>) was used to annotate peaks to the nearest transcriptional start site (TSS). To ascertain biological significance of HDAC2 peak differences, GREAT analysis was performed as previously described³⁵.

Transgene expression quantification—Quantification of the possible expression of lentiviral, episomal and minicircle transgenes was performed using RNA-seq fastq files from each lentiviral, episomal, minicircle and mRNA reprogramming method specific clones. A modified human transcriptome index was created by taking the Ensembl 72 human cDNA fasta and appending lentiviral and episomal sequences. Quantification was performed using Kallisto with the paired-end fastq files and the modified index file as reference. Transcripts per million values (tpm) were calculated and graphed.

Cardiomyocyte differentiation and FACS analysis—All lines were thawed at p10 and grown for two passages before beginning cardiomyocyte differentiation. Each line was seeded on the same day at 1.2×10^5 per well in a Matrigel-coated 6-well plate. For four days, hiPSCs were fed with E8 media until 80% confluence was reached. E8 media was aspirated and replaced with B27 without insulin (A1895601, Life Technologies) in RPMI supplemented with 6 μ M CHIR-99021 (CT99021, Selleckchem) to initiate differentiation. Cultures were moved to normoxic conditions for the duration of cardiomyocyte differentiation. Two days after CHIR treatment, RPMI supplemented with B27 without insulin was added for one day followed by a two-day treatment of 5 μ M IWR-1 (I0161, Sigma) in RPMI supplemented with B27 without insulin. After IWR-1 treatment, cells were fed for two days with RPMI containing B27 without insulin and five days with RPMI with B27 with insulin (17504-044, Life Technologies) (Supplementary Fig. 8). Cardiomyocyte differentiation was repeated three additional times with all lines in parallel at the same

passage number. For cell sorting by flow cytometry, differentiated cardiomyocytes were trypsinized (12605-010, Life Technologies) for 20 min and centrifuged at 300g for 4 min. Cardiomyocytes were resuspended in 1 ml of 1% paraformaldehyde (15713S, Electron Microscope) and incubated for 20 min at room temperature. 4 ml of PBS was added to the cell solution and cells were centrifuged and resuspended in a cold (−20 °C) methanol/acetone solution (80:20) for 5 min. Methanol/acetone permeabilized cells were washed with FACS wash buffer (1% BSA in PBS) and filtered into 35 µm filter FACS tubes (352235, BD Falcon). Cardiac troponin-T antibody (1:200 MS-295-P, Thermo-Scientific) was added for 30 min and cells were washed (centrifuged at 300g for 4 min and resuspended in FACS wash buffer) three times with 5 ml of FACS wash buffer. Fluorescent secondary antibody labelling was performed using anti-mouse IgG (Molecular Probes) at a dilution of 1:1000 for 30 min and washed three times with FACS wash buffer. FACS was then performed using the FACSaria III cell sorter (BD Biosciences) and data was collected by a researcher blinded to the sample labels.

Immunofluorescence labelling of hiPSCs and hiPSC-CMs—hiPSCs were plated onto Matrigel-coated glass coverslips and were fixed using 5% paraformaldehyde for 10 min as previously described³⁶. The coverslips were washed with IF wash buffer (3% PBS containing 0.1% tween-20) for 5 min and blocked using PBS containing 3% BSA for 30 min. Primary antibodies for hiPSC makers (1:100 SOX2, 1:250 Tra-1-81 (mab4381, Millipore), 1:500 Tra-1-60, 1:100 NANOG (SC-33759, Santa Cruz), and 1:100 POU5F1 (OCT4) (SC-9081, Santa Cruz)) were incubated overnight at 4 °C in PBS containing 3% BSA. For immunofluorescent imaging of hiPSC-CMs, anti-Troponin T (1:200 MS-295-P, Thermo-Scientific) and anti- α -sarcomeric actin (1:500, MA1-21597, Thermo-Scientific) were incubated overnight at 4 °C. Coverslips were washed with IF wash buffer 3 time for 5 min and the appropriate secondary was added for 30 min 1:400 (Life technologies, A-11037 or Alexa Fluor 488 goat anti-rabbit IgG H+L, A-11001Alexa Fluor 488 goat anti-mouse IgG H+L). Coverslips were washed for an additional 3 times for 5 min and mounted using Faramount Aqueous Mounting Media (S3025, Dako, Carpinteria, CA). Images were then taken using LSM 510 confocal microscope (Zeiss).

Teratoma formation—For teratoma formation, each hiPSC line was detached from its culture plates using EDTA (2 mM)-supplemented PBS for 3 min. Detached hiPSCs were centrifuged at 400g for 4 min, resuspended in PBS and counted. One million cells from each line were added to 50 µl of cold Matrigel and kept on ice until injection. Each line was injected into the hindlimb of NOD/SCID mice and after approximately 45 days, explanted tumors were embedded into paraffin, sectioned and processed for haematoxylin & eosin (H&E) staining as previously described^{36,37}.

hiPSC karyotype analysis—Each hiPSC line from p12 was detached from the culture play using trypsin for 10 min and centrifuged at 300g for 3 min. hiPSC pellets were resuspended in 200 µl of PBS and DNA was isolated using the DNeasy blood and tissue kit (Qiagen, Valencia, CA). DNA was submitted to the Stanford Functional Genomics Facility for SNP karyotyping using the HuCytoSNP-12 chip (Illumina). CNV and SNP visualization was performed using KaryoStudio v1.4 (Illumina) (Supplementary Fig. 3).

Minicircle production—One clone from a minicircle parental plasmid (Figure S9B) transformed bacterial plate was added to 5 ml LB Broth media with 50 $\mu\text{g ml}^{-1}$ kanamycin in a shaker at 250 rpm at 37 °C until the OD reached between 3.75 and 4.25. Approximately, 400 ml of minicircle induction mix (1 volume of fresh LB Broth (1200 ml), 4% 1N NaOH (48 ml) and 1% L-arabinose (12 g)) was added to each flask and incubated at 32 °C at 250 rpm for 6–8 hours. Bacteria was harvested by centrifugation and minicircles were isolated according to the Qiagen Plasmid Plus Maxi kit instructions. Confirmation of successful minicircle induction was performed by digestion using HindIII and NdeI since NdeI linearizes only the parental backbone but not the minicircle (expected band around 7.7 kb).

Lentivirus titrations—Lentiviral production was performed as previously described³⁰. To ensure optimal viral load, lentivirus titrations were initially performed by making viral dilutions (1:100, 1:200, 1:400, 1:800, 1:1,600 and 1:3,200) from the stock viral solution. 50 μl of each dilution was added to a 24-well containing 1×10^5 HEK 293TN cells. After 72 h, HEK cells were trypsinized and resuspended in 0.5 ml of PBS for FACS analysis to determine the percentage of TOMATO-positive cells. Biological titre was calculated using $\text{TU } \mu\text{l}^{-1} = (p \times n/100 \times v) \times 1/\text{DF}$ where TU = transfection units, p = % TOMATO⁺ cells, v = volume of dilution added to each well = 40 μl , n = number of cells at time of transduction, and DF = dilution factor. The viral dilution which produced over 80% TOMATO⁺ cells (Supplementary Fig. 9F) was used for making the lentiviral hiPSC lines.

Western blots—Cells were harvested in RIPA lysis buffer (Sigma) supplemented with protease (Complete Mini; Roche) and phosphatase inhibitors (phosphatase inhibitor cocktail I and II; Sigma). 20 μg of total protein were separated on 4–12% SDS gels (Invitrogen) by SDS-PAGE and transferred to polyvinylidene fluoride membranes following the protocol of the manufacturer (Invitrogen XCell SureLock Mini Cell Electrophoresis and Wet Blot System). Membranes were incubated overnight with the OCT4 primary antibody (AF1759, R&D Systems), washed, and incubated with anti-goat horseradish peroxidase secondary antibody (Jackson Laboratories) for 45 min. Chemiluminescence reaction was carried using the Amersham ECL Prime Western Blotting Detection Reagent kit (GE Healthcare) on a Kodak image station 4000R.

RNA-seq gene expression and splicing analysis—As input for gene expression and splicing analyses, Tophat junction BED files were imported into AltAnalyze 2.0.8³⁸ to build a BED exon coordinate reference file from EnsMart72³⁹. Exon counts were performed from the reference BED files using Bedtools³² from BAM files generated using Tophat⁴⁰. AltAnalyze was run on each junction and exon BED file using default values to produce RPKM values. Differentially expressed genes among fibroblasts, hESCs, and hiPSCs made by each reprogramming technique were calculated using an empirical Bayes moderated t-test $P < 0.05$ with a 2-fold expression difference. To determine the predicted gene expression markers for each method of reprogramming (transfection-, electroporation- and viral-based) the MarkerFinder algorithm was used as described (www.AltAnalyze.org). R was used with the cor function to acquire Spearman rank correlation coefficients and the package corrplot was used graph each correlation matrix as well as the number of genes called as significant between each group. Alternative splicing and alternative promoter regulation for known and

novel exons were obtained using a joint analysis of reciprocally expressed exon-exon junctions (ASPIRE algorithm) and exons along (splicing index) yielded thousands of alternative splicing events for all pairwise comparisons (fibroblasts versus hiPSCs, fibroblasts versus hESCs, or hiPSCs versus hESCs).

RNA-seq GO analysis and transcription factor enrichment—AltAnalyze 2.0.8³⁸, Toppgene suite⁴¹ and gene set enrichment analysis (GSEA)⁴² were used to acquire transcription factor enrichment terms, and to determine which genes corresponded to activity from each transcription factor. GO terms were acquired by submitting a list of genes into the online Toppgene suite. This suite performs the GO term enrichment analysis and provides a p value for each term. Genome Viewer v2.3 was used to create exon and junction-level coverage maps using the Sashimi Plot View.

Data availability

The authors declare that all data supporting the findings of this study are available within the paper and the Supplementary Information. RNA-seq and ChIP-seq raw fastq files are stored under the GEO accession number GSE69626.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was funded by the Canadian Institute of Health Research 201210MFE-289547 (J.M.C.), National Institutes of Health 1K99HL128906 (J.M.C.), PCBC_JS_2014/4_01 (J.M.C.), National Research Foundation of Korea 2012R1A6A3A03039821 (J.L.), the Burroughs Wellcome Foundation, National Institutes of Health R01 HL123968, HL128170, R01 HL126527 (J.C.W.), and P01 GM099130 (M.P.S.). The authors would like to thank the Stanford Stem Cell Institute Genome Center for their sequencing knowledge, V. Sebastiano for hESC culturing, and B. Huber for his help with the teratoma assay. We would also like to thank J. Brito and B. Wu for their help in editing the manuscript.

References

1. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006; 126:663–676. [PubMed: 16904174]
2. Chin MH, et al. Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell*. 2009; 5:111–123. [PubMed: 19570518]
3. Wernig M, et al. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature*. 2007; 448:318–324. [PubMed: 17554336]
4. Bock C, et al. Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell*. 2011; 144:439–452. [PubMed: 21295703]
5. Choi J, et al. A comparison of genetically matched cell lines reveals the equivalence of human iPSCs and ESCs. *Nat Biotechnol*. 2015; 33:1173–1181. [PubMed: 26501951]
6. Ruiz S, et al. Identification of a specific reprogramming-associated epigenetic signature in human induced pluripotent stem cells. *Proc Natl Acad Sci USA*. 2012; 109:16196–16201. [PubMed: 22991473]
7. Newman AM, Cooper JB. Lab-specific gene expression signatures in pluripotent stem cells. *Cell Stem Cell*. 2010; 7:258–262. [PubMed: 20682451]
8. Guenther MG, et al. Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. *Cell Stem Cell*. 2010; 7:249–257. [PubMed: 20682450]

9. Wang Y, et al. A transcriptional roadmap to the induction of pluripotency in somatic cells. *Stem Cell Rev.* 2010; 6:282–296. [PubMed: 20336394]
10. Kim K, et al. Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. *Nat Biotechnol.* 2011; 29:1117–1119. [PubMed: 22119740]
11. Fusaki N, Ban H, Nishiyama A, Saeki K, Hasegawa M. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proc Jpn Acad Ser B Phys Biol Sci.* 2009; 85:348–362.
12. Gifford CA, et al. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell.* 2013; 153:1149–1163. [PubMed: 23664763]
13. Delgado-Olguin P, et al. Epigenetic repression of cardiac progenitor gene expression by Ezh2 is required for postnatal cardiac homeostasis. *Nat Genet.* 2012; 44:343–347. [PubMed: 22267199]
14. Kyttila A, et al. Genetic variability overrides the impact of parental cell type and determines iPSC differentiation potential. *Stem Cell Rep.* 2016; 6:200–212.
15. Bhutani K, et al. Whole-genome mutational burden analysis of three pluripotency induction methods. *Nat Commun.* 2016; 7:10536. [PubMed: 26892726]
16. Rouhani F, et al. Genetic background drives transcriptional variation in human induced pluripotent stem cells. *PLoS Genet.* 2014; 10:e1004432. [PubMed: 24901476]
17. Heilig C, et al. Implications of glucose transporter protein type 1 (GLUT1)-haplodeficiency in embryonic stem cells for their survival in response to hypoxic stress. *Am J Pathol.* 2003; 163:1873–1885. [PubMed: 14578187]
18. Janaszak-Jasiecka A, et al. miR-429 regulates the transition between hypoxia-inducible factor (HIF)1A and HIF3A expression in human endothelial cells. *Sci Rep.* 2016; 6:22775. [PubMed: 26954587]
19. Wang C, et al. Hypoxia inhibits myogenic differentiation through p53 protein-dependent induction of Bhlhe40 protein. *J Biol Chem.* 2015; 290:29707–29716. [PubMed: 26468276]
20. Bhandari DR, et al. The regulatory role of c-MYC on HDAC2 and PcG expression in human multipotent stem cells. *J Cell Mol Med.* 2011; 15:1603–1614. [PubMed: 20716118]
21. Marshall GM, et al. Transcriptional upregulation of histone deacetylase 2 promotes Myc-induced oncogenic effects. *Oncogene.* 2010; 29:5957–5968. [PubMed: 20697349]
22. Zhang Z, Wu WS. Sodium butyrate promotes generation of human induced pluripotent stem cells through induction of the miR302/367 cluster. *Stem Cells Dev.* 2013; 22:2268–2277. [PubMed: 23534850]
23. Huangfu D, et al. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat Biotechnol.* 2008; 26:795–797. [PubMed: 18568017]
24. Kim K, et al. Epigenetic memory in induced pluripotent stem cells. *Nature.* 2010; 467:285–290. [PubMed: 20644535]
25. Okita K, et al. A more efficient method to generate integration-free human iPS cells. *Nat Methods.* 2011; 8:409–412. [PubMed: 21460823]
26. Narsinh KH, et al. Generation of adult human induced pluripotent stem cells using nonviral minicircle DNA vectors. *Nat Protoc.* 2011; 6:78–88. [PubMed: 21212777]
27. Warren L, et al. Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell.* 2010; 7:618–630. [PubMed: 20888316]
28. Anokye-Danso F. Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell.* 2011; 8:376–388. [PubMed: 21474102]
29. Liao B, et al. MicroRNA cluster 302–367 enhances somatic cell reprogramming by accelerating a mesenchymal-to-epithelial transition. *J Biol Chem.* 2011; 286:17359–17364. [PubMed: 21454525]
30. Sharma A, et al. The role of SIRT6 protein in aging and reprogramming of human induced pluripotent stem cells. *J Biol Chem.* 2013; 288:18439–18447. [PubMed: 23653361]
31. Warlich E, et al. Lentiviral vector design and imaging approaches to visualize the early stages of cellular reprogramming. *Mol Ther.* 2011; 19:782–789. [PubMed: 21285961]
32. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010; 26:841–842. [PubMed: 20110278]

33. Krzywinski MI, et al. Circos: An information aesthetic for comparative genomics. *Genome Res.* 2009; 19:1639–1645. [PubMed: 19541911]
34. Zhang Y, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008; 9:R137. [PubMed: 18798982]
35. McLean CY, et al. GREAT improves functional interpretation of *cis*-regulatory regions. *Nat Biotechnol.* 2010; 28:495–501. [PubMed: 20436461]
36. Sun N, et al. Patient-specific induced pluripotent stem cells as a model for familial dilated cardiomyopathy. *Sci Transl Med.* 2012; 4:130ra147.
37. Huber BC, et al. Costimulation-adhesion blockade is superior to cyclosporine A and prednisone immunosuppressive therapy for preventing rejection of differentiated human embryonic stem cells following transplantation. *Stem Cells.* 2013; 31:2354–2363. [PubMed: 24038578]
38. Emig D, et al. AltAnalyze and DomainGraph: analyzing and visualizing exon expression data. *Nucleic Acids Res.* 2010; 38:W755–W762. [PubMed: 20513647]
39. Kasprzyk A, et al. EnsMart: a generic system for fast and flexible access to biological data. *Genome Res.* 2004; 14:160–169. [PubMed: 14707178]
40. Trapnell C, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc.* 2012; 7:562–578. [PubMed: 22383036]
41. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* 2009; 37:W305–W311. [PubMed: 19465376]
42. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005; 102:15545–15550. [PubMed: 16199517]

Fig. 1. RNA-seq expression differences between hiPSC lines, fibroblasts and hESCs

a, 3D PCA of the log-fold gene expression profile of episomal, Sendai, lentivirus and mRNA lines, as well as H9 hESCs, was used to compare 42 lines sequenced by the PCBC Cell Characterization Core in Cincinnati. Similar reprogramming techniques clustered together. **b**, 3D PCA of the log-fold gene expression differences in minicircle, episomal, Sendai, lentivirus, mRNA and microRNA lines demonstrates that hESCs and fibroblasts clustered separately from the hiPSC lines, while clones from each reprogramming method clustered closer together. In addition, the type of reprogramming method (viral, transfection,

or electroporation) tended to cluster closer c, RPKM values were calculated and filtered to remove genes with a RPKM of less than 3, and RPKM values were transformed into log values. Values were clustered using the Euclidian column metric, weighted column clustering method, city block row clustering metric and weighted row clustering method.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 2. Comparison of RNA-seq expression differences among fibroblasts, hESCs and hiPSCs reprogrammed by different methods

a–c, Significant gene expression differences were calculated (bottom left (**a**) or left (**b,c**) matrices: number of protein coding genes with an empirical Bayes moderated t -test P value of $P < 0.05$ and 2-fold difference) and Spearman rank correlation (top right (**a**) or right (**b,c**) matrices) for each method of reprogramming were graphed for each clone (**a**), type of reprogramming method (**b**) and combined hESCs and hiPSC lines (**c**). **d**, AltAnalyze was run on all hiPSCs, hESCs and fibroblasts, and the marker prediction algorithm was applied to generate a heatmap (Euclidian distance hierarchical clustering of the $\log_2$ fold expression) of the predicted markers. Below each reprogramming group (electroporation, fibroblasts, hESCs, virus or transfections) is the list of the genes enriched within that group.

Fig. 3. Gene ontology (GO) enrichment, transcription factor target enrichment, and HDAC2 activity between hESCs and hiPSCs

a, GO biological and molecular function enrichment was performed on the genes that were significantly different between hiPSCs and hESCs. The GO terms with the smallest P value (unadjusted P using the hypergeometric test) from each reprogramming method compared to hESCs was graphed. **b**, Transcription factor (TF) enrichment between hiPSCs and hESCs was predicted to be associated with *HDAC2*, *EZH2*, *SOX2*, *NANOG*, *EED*, *MYC/MAX*, *NRF1*, *ELK1*, *TCF3*, *HIF1A*, *LEF1*, *SUZ12*, and *YY1* activity. Quantification of the percentage of genes (called as significantly different between hiPSCs and hESCs) that are annotated as targets of each transcription factor was divided as genes either elevated in hESCs (red bars) or elevated in hiPSCs (blue bars). **c**, $\log_2$ peak density values were

calculated in the peak regions identified by MACS in the *HDAC2* ChIP-seq dataset. A heatmap was used to represent $\log_2$ fold peak density differences between hiPSCs and hESCs. **d**, Activity of *LEF1* and *MYC* was predicted to be associated with *HDAC2* binding sites (GREAT analysis). **e,f**, RNA-seq demonstrates elevated *MYC* expression in hESCs versus hiPSCs (**e**; hiPSCs, 4.45 ± 0.72 ; hESCs, 13.78 ± 0.54 ; empirical Bayes moderated *t*-test $P < 0.05$, mean $\pm$ s.e.m.) and higher levels of genes regulated by MYC (**f**; red represents higher expression in hESCs while blue represents higher expression in hiPSCs).

Fig. 4. Comparison of H3K4me3 and H3K27me3 ChIP-seq between each reprogramming method

a. Spearman correlation values were calculated for the peak density within 2 kb of the TSS for the H3K4me3 and H3K27me3 dataset. High correlations were observed within hESCs and within hiPSCs being compared. **b.** Comparison of H3K4me3 and H3K27me3 identified 251 gene regions (annotated TSS) in the H3K4me3 dataset and 281 gene regions in the H3K27me3 dataset (with a >5-fold change in $\log_2$ peak density values between hiPSCs and hESCs).

Fig. 5. X chromosome H3K27me3 and *XIST* differences in hiPSCs

a, Localization of H3K4me3 and H3K27me3 differences between hiPSCs and hESCs along the genome identified that the vast majority of identified H3K27me3 differences were enriched in the hiPSCs along the X chromosome (dotted box). **b**, Combined coverage of reads by RNA-seq and H3K4me3 of *XIST* (chrX: 73032849-73083969) reveals the active H3K4me3 marks predominantly observed in the hiPSCs and RNA-seq reads predominantly observed in the hiPSCs. **c**, Expression of *XIST* from the RNA-seq data revealed low levels of *XIST* expression in the microRNA_CL2 line but not the other hiPSC lines. **d**, *XIST* expression was significantly lower at passage 40 (p40) than at passage 12 (p12) in hiPSCs (p12, 1.69 ± 0.35 ; p40, 0.122 ± 0.05 ; mean $\pm$ s.e.m.). **e**, Transcription factor enrichment analysis of differential ChIP-seq targets in **b** corresponded to genes involved *TCF3* and *LEF1* activity in the H3K4me3 dataset and *YY1* and *LEF1* activity in the H3K27me3 dataset.

Fig. 6. Comparison of cardiac differentiation potential between each reprogramming method

a, To assess the cardiomyocyte differentiation potential, each line was differentiated in parallel ($N = 9$) and FACS analysis was performed for the cardiac-specific marker troponin-T (TNNT2). **b**, Quantification of the percentage of cardiomyocytes present after differentiation did not reveal any significant differences in the cardiac differentiation potential between lines generated by different reprogramming methods (mean $\pm$ s.e.m, one-way ANOVA, $P > 0.05$). **c**, Supervised clustering was performed using the RNA-seq expression from cardiovascular genes annotated by the Cardiovascular Gene Ontology Annotation Initiative. **d**, Heatmap of supervised clustering performed using H3K27me3 TSS log₂ peak density values found to be different (>5 -fold enrichment) with the genes from the

Cardiovascular Gene Ontology Annotation Initiative. Largest fold differences in H3K27me3 peak density (dotted box) and the corresponding gene's TSS were plotted (peak density values) in hESCs, and two lines demonstrated the largest differences between their cardiac differentiation potential (microRNA_CL2 and Sendai_CL1). **e**, These genes are critical for cardiomyocyte differentiation and were also demonstrated to be enriched in *TCF3* targets (unadjusted raw *P* using the hypergeometric test). **f**, RNA-seq RPKM values of *TCF3* demonstrates significant difference between the expression levels of *TCF3* in hESCs versus hiPSCs (hiPSCs, 4.63 ± 0.89 ; hESCs, 7.55 ± 0.47 ; mean $\pm$ s.e.m.).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Summary comparison of the six reprogramming methods used in this study

Reprogramming factors	Integrating	Reprogramming efficiency	Effort	Cost	Promoter	order of factors
Lentivirus OCT4 SOX2 KLF4 c-MYC (codon optimized)	Yes	High	Low	Low	SFFV	One vector (OKSM)
Sendai OCT4 SOX2 KLF4 c-MYC	No	High	Low	High	NA	Separate virions
MiniCircle OCT4 SOX2 KLF4 c-MYC	Maybe	Low	Medium	Low	SFFV	One vector (OKSM)
Episomal OCT4 SOX2 KLF4 L-MYC Lin28 p53-shRNA	Maybe	Low	Medium	Low	CaG	Three vectors
mRNA OCT4 SOX2 KLF4 c-MYC Lin28	No	Low	High	High	NA	Separate mRNA
microRNA OCT4 SOX2 KLF4 c-MYC microRNA 302 and microRNA 367 clusters	No	Low	High	High	NA	Separate mRNA

The reprogramming factors, possibility of transgene integration, reprogramming efficiency, effort, cost, promoter used, and order of reprogramming factors are compared among the lentiviral, Sendai virus, minicircle, episomal, mRNA, and microRNA/mRNA (microRNA) reprogramming methods used in this study.