A Panel of Induced Pluripotent Stem Cells From Chimpanzees: a Resource for Comparative Functional Genomics

Irene Gallego Romero[1,*,†], Bryan J. Pavlovic[1,*], Irene Hernando-Herraez[2], Xiang Zhou[3], Michelle C. Ward[1], Nicholas E. Banovich[1], Courtney L. Kagan[1], Jonathan E. Burnett[1], Constance H. Huang[1], Amy Mitrano[1], Claudia I. Chavarria[1], Inbar Friedrich Ben-Nun[4,5], Yingchun Li[6,7], Karen Sabatini[4,7,8], Trevor R. Leonardo[4,7,8], Mana Parast[6,7], Tomas Marques-Bonet[2,9], Louise C. Laurent[7,8], Jeanne F. Loring[4], and Yoav Gilad[1,†]

1. Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA
2. ICREA, Institute of Evolutionary Biology (UPF-CSIC), PRBB, Barcelona 08003, Spain
3. Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA
4. Center for Regenerative Medicine, Department of Chemical Physiology, The Scripps Research Institute, La Jolla, CA 92037, USA
5. Present address: Lonza Walkersville, Inc. 8830 Biggs Ford Road, Walkersville, MD 21793, USA
6. Department of Pathology, University of California San Diego, San Diego, CA 92103, USA
7. Sanford Consortium for Regenerative Medicine, 2880 Torrey Pines Scenic Drive, La Jolla, CA 92037, USA
8. Department of Reproductive Medicine, University of California San Diego, San Diego, CA 92103, USA
9. Centro Nacional de Análisis Genómico (CNAG), Barcelona, Spain, 08023, Spain

* These authors contributed equally to this work.

† To whom correspondence should be addressed:

email: ireneg@uchicago.edu

email: gilad@uchicago.edu

tel: +1 773 702 8507

Keywords:

iPSC, chimpanzee, induced pluripotent stem cells, evolution, *Pan troglodytes*, cell panel

Abstract

Comparative genomics studies in primates are restricted due to our limited access to samples. In order to gain better insight into the genetic processes that underlie variation in complex phenotypes in primates, we must have access to faithful model systems for a wide range of cell types. To facilitate this, we generated a panel of 7 fully characterized chimpanzee induced pluripotent stem cell (iPSC) lines derived from healthy donors. To demonstrate the utility of comparative iPSC panels, we collected RNA-sequencing and DNA methylation data from the chimpanzee iPSCs and the corresponding fibroblast lines, as well as from 7 human iPSCs and their source lines, which encompass multiple populations and cell types. We observe much less within-species variation in iPSCs than in somatic cells, indicating the reprogramming process erases many inter-individual differences. The low within-species regulatory variation in iPSCs allowed us to identify many novel inter-species regulatory differences of small magnitude.

47 **Introduction**

48 Comparative functional genomic studies of humans and other primates have been
49 consistently hindered by a lack of samples (Gallego Romero et al. 2012). In spite of their clear
50 potential to inform our understanding of both human evolution and disease, practical and ethical
51 concerns surrounding working with non-human primates have constrained the field to using a
52 limited set of cell types collected in a non-invasive or minimally invasive manner, primarily
53 lymphoblastoid cell lines (LCLs) and fibroblasts. Comparative studies of any other primate tissue
54 have been limited to using post-mortem (typically frozen) materials, thereby precluding most
55 experimental manipulation and yielding primarily observational insights (see, for example,
56 Blekhman et al. 2010; Blekhman et al. 2008; Brawand et al. 2011).

57 An alternative has been to use model organisms in an attempt to recapitulate inter-primate
58 regulatory differences. The typical approach involves the introduction of sequences of evolutionary
59 interest into a model system, and then searching for spatial or temporal differences in gene
60 expression that can be ascribed to the introduced sequence (Cotney et al. 2013; Enard et al. 2009).
61 This is a difficult and challenging approach and, perhaps as a result, there are still only a handful of
62 well-described examples of human-specific regulatory adaptations in primates (McLean et al. 2011;
63 Prabhakar et al. 2008) and even fewer cases where the underlying regulatory mechanisms have been
64 resolved (Pollard et al. 2006; Rockman et al. 2005). While these studies are useful and often
65 informative, they also entail assumptions of functional conservation between the model system and
66 the species of interest that may not necessarily be true (Gallego Romero et al. 2012).

67 Induced pluripotent stem cells (iPSCs) can provide a viable means of circumventing these
68 concerns and limitations, at least with respect to the subset of phenotypes that can be studied in *in*
69 *vitro* systems. Reprogramming somatic cell lines to a stable and self-sustaining pluripotent state
70 (Takahashi et al. 2007; Takahashi and Yamanaka 2006) has become routine practice for human and
71 murine cell lines, but extension to other animals, especially non-human primates, is not yet
72 widespread despite some exceptions (e.g. Ben-Nun et al. 2011; Ezashi et al. 2009; Marchetto et al.
73 2013b; Nagy et al. 2011). Instead, the broadest application of iPSCs to date has been the generation of
74 lines derived from patients suffering from a variety of genetic disorders (Cohen and Melton 2011;
75 Israel et al. 2012; Liu et al. 2012; Merkle and Eggan 2013; Wang et al. 2014), with the dual aims of

providing a deeper understanding of disease phenotypes and developing new therapeutic avenues. These cell lines have been shown to display *in vitro* properties corresponding to relevant patient phenotypes observed *in vivo,* both as iPSCs and when differentiated into other pertinent cell types, supporting their utility in clinical applications; more generally, these properties also highlight the tantalizing flexibility of iPSCs as a means of exploring developmental and cell lineage determination pathways.

Thus, the development of an iPSC-based system for comparative genomic studies in primates will allow us to compare regulatory pathways and complex phenotypes in humans and our close evolutionary relatives using appropriate models for different tissues and cell types. This will be a powerful resource with which to examine the contribution of changes in gene regulation to human evolution and diversity. To demonstrate the validity of this approach, we have generated a panel of 7 chimpanzee iPSC lines that are fully characterized and comparable to human iPSC lines in their growth and differentiation capabilities.

Results

We generated a panel of induced pluripotent stem cell lines from seven chimpanzees through electroporation of episomal plasmids expressing *OCT3/4* (also known as *POU5F1*), *SOX2*, *KLF4*, *L-MYC*, *LIN28*, and an shRNA targeting *TP53* (Okita et al. 2011), as well as an *in vitro*-transcribed *EBNA1* mRNA transcript (Chen et al. 2011; Howden et al. 2006) that promotes increased exogenous vector retention in the days following electroporation. Our chimpanzee panel is comprised of seven healthy individuals (4 female, 3 male, further details on these individuals are given in supplementary file 1) ranging from 9 to 17 years old. Fibroblasts from 5 of the 7 individuals were purchased from the Coriell Institute for Medical Research, while the remaining two (C6, C7) were derived from 3 mm skin punch biopsies directly collected from animals at the Yerkes Primate Research Center of Emory University (see methods). All chimpanzee iPSC lines described in this publication are available fully and without restrictions to other investigators upon request to the corresponding authors.

Characterizing the chimpanzee iPSCs

The chimpanzee iPSC lines closely resemble human iPSC lines in morphology (figure 1a; all images shown in main text are from chimpanzee line C4. Similar images of the other lines are available as figure 1-figure supplements 1 to 5). All lines could be maintained in culture for at least 60 passages without loss of pluripotency or self-renewal capability using standard iPSC culture conditions, both on mouse embryonic fibroblast (MEF) feeder cells and in feeder-free conditions. The genomes of all our lines appeared to be cytogenetically stable; all exhibited normal karyotypes after reprogramming and more than 15 passages in culture, ruling out the presence of gross chromosomal abnormalities (figure 1b, figure 1-figure supplement 1).

We confirmed nuclear expression of *OCT3/4*, *SOX2* and *NANOG* in all lines by immunocytochemistry (figure 1c; figure 1-figure supplement 2). The pluripotent cells also express the surface antigens Tra-1-81 and SSEA4, while cells collected from the center of differentiating colonies expressed SSEA1 at levels comparable to differentiating colonies of human iPSC lines (figure 1-figure supplement 3). To confirm that the observed expression of pluripotency-associated genes is of endogenous origin, we performed qPCR with primers designed to specifically amplify the endogenous *OCT3/4*, *SOX2*, *NANOG* and *L-MYC* transcripts (figure 1d; all PCR primers used in this work are listed in supplementary file 2). Indeed, we found no evidence of exogenous gene expression after 10 passages (figure 1-figure supplement 4), and no traces of genomic integration or residual episomal plasmid retention after 15 passages (figure 1e). These observations indicate that self-renewal in our chimpanzee iPSC lines is maintained solely through endogenous gene expression.

To confirm pluripotency and test the differentiation capabilities of our lines, we performed a number of assays. First, we generated embryoid bodies from all 7 chimpanzee iPSC lines and assayed their ability to spontaneously differentiate into the three germ layers by immunocytochemistry. All lines spontaneously gave rise to tissues from the three germ layers (figure 2a; figure 2-figure supplement 1). Second, we carried out directed differentiations to hepatocytes and cardiomyocytes in a subset of the lines using previously published protocols (see methods, figure 2-figure supplement 2 and video 1). Third, we performed teratoma formation assays in four of the lines using Fox Chase SCID-beige and CB17.Cg-*Prkdc*^{scid}*Lyst*$^{bg\text{-}J}$/Crl immunodeficient male mice. All four iPSC lines were capable of generating tumours in mice, and all tumours

examined contained tissues of endodermal, ectodermal and mesodermal origins (figure 2b, figure 2-figure supplement 3). To confirm the chimpanzee origin of these tissues, we extracted and performed Sanger sequencing on mitochondrial DNA from the tumours (figure 2-figure supplement 4).

Finaly, we characterized pluripotency in our lines through PluriTest, a bioinformatic classifier that compares the gene expression profiles of new lines to those obtained from a reference set of over 400 well-characterized human pluripotent and terminally differentiated lines (Müller et al. 2011), modified to accommodate data from both species. All chimpanzee lines have PluriTest pluripotency scores greater than the pluripotency threshold value of 20 (figure 3a, supplementary file 1). We also calculated PluriTest novelty scores for all samples. In human PSCs, novelty values above 1.67 are suggestive of chromosomal duplications or expression of differentiation-associated genes. Human PSCs with high novelty scores are typically either difficult to maintain and expand in culture (because they differentiate spontaneously at a high rate), or cannot be consistently differentiated to all three germ layers. All of our chimpanzee lines had novelty scores above the 1.67 threshold (figure 3b). However, in contrast to human PSCs with high novelty scores, our chimpanzee lines can be both easily maintained in culture and differentiated into all three germ layer lineages, as demonstrated by the embryoid body and teratoma assays detailed above. We thus hypothesize that the observed high novelty scores are likely driven by inter-species gene regulatory differences that the PluriTest assay, which was trained exclusively on human samples, interpreted as abnormal gene expression.

Interspecies analysis of gene expression and methylation data from iPSCs

To better examine gene expression and regulatory differences between human and chimpanzee iPSCs, we generated genome-wide RNA-sequencing and DNA methylation data (see methods) from all chimpanzee iPSC lines, as well as from 7 human iPSC lines also generated and validated in our laboratory. While all of the chimpanzee iPSCs were derived from fibroblast cell lines (supplementary file 1), the human iPSCs were derived from both fibroblasts and immortalised lymphoblastoid cell lines (LCLs) from Caucasian and Yoruba individuals (see supplementary file 3 for additional details). We designed the comparative study this way in order to demonstrate that

regulatory differences between human and chimpanzee iPSCs cannot be explained by technical differences due to culturing conditions or the cell type of the somatic precursor cells used for reprograming.

To prevent biases due to genetic divergence between the two species, we chose to restrict our gene expression analyses to a curated set of genes with one-to-one orthology between humans and chimpanzees (Blekhman 2012; Blekhman et al. 2010). Following assessment of quality control metrics (see methods), we obtained normalised RPKM estimates for 12,171 genes that were expressed in at least 4 iPSC lines from either one of the species (see methods). We similarly restricted our DNA methylation analyses to a set of 335,307 high quality probes with a high degree of sequence conservation between humans and chimpanzees (as in Hernando-Herraez et al. 2013; see methods).

To examine broad patterns in the data, we used principal component analysis (PCA). We observed clear and robust separation of human and chimpanzee iPSC lines along the first principal component (PC) in both the gene expression and DNA methylation data (figures 4a, 4b; regression of PC1 by species; $P < 10^{-13}$ for the expression data; $P < 10^{-12}$ for the DNA methylation data). Within the human samples, PC2 appears to be driven by ethnicity, as we observe all Caucasian samples consistently clustering together despite their different cell types of origin ($P = 0.005$ for the association between PC2 and human ethnicity in the expression data, $P = 0.044$ in the DNA methylation data).

We then analysed regulatory differences between the species by first focusing on the gene expression data. At an FDR of 1%, we identified 4,609 genes (37.9%) as differentially expressed (DE) between the iPSCs of the two species (supplementary file 4; see methods for details). The majority of DE genes do not exhibit large inter-species fold-change differences in expression levels (figure 4-figure supplement 1, 4-2). An analysis of functional annotation of the DE genes reveals that no Gene Ontology Biological Process terms (GO BP; Ashburner et al. 2000) are significantly overrepresented among these genes at an FDR of 5% (supplementary file 5), although we identified 123 overrepresented terms if we limit our analysis to the 546 genes with absolute $\log_2$ fold-change difference > 2 (supplementary file 5). Additionally, we tested for concordance between our list of DE genes and a list of 2,730 genes that were previously classified as DE between human and non-

192 human primate iPSC lines (Marchetto et al. 2013b). Given our stringent approach to consider
193 orthologous genes, only 2,081 (76%) genes could be analysed across the two studies. Of these, 1,495
194 genes are detectably expressed in our lines, and 1,079 (72.2%) are classified as DE between the
195 species in both data sets (a highly significant enrichment; χ^2 $P < 10^{-16}$). Expression trends within
196 these DE genes are in the same direction in both data sets in 1,060 of cases (98.24%).
197 Next, we used a similar approach to identify differentially methylated probes and regions
198 between the iPSCs of both species (see methods). We identified 63,791 probes that are differentially
199 methylated (DM) between the two species at an FDR of 1%, 26,554 of which have a mean intergroup
200 β difference ≥ 0.1, our arbitrary effect size threshold for retaining probes for differentially
201 methylated region (DMR) identification and downstream analyses. Of these, 10,460 probes could be
202 further grouped into 3,529 regions of 2 or more DM probes within 1 kb, which we designated DMRs;
203 (supplementary file 6); the numbers of probes and regions identified as DM at a range of mean
204 interspecies β thresholds are given in supplementary file 7.
205 In order to consider the DNA methylation and gene expression data jointly, we focused on a
206 subset of 2,348 DMRs that could be associated with a single Ensembl gene. Overall, these DMRs
207 were associated with 2,141 genes, of which 1,350 were also detectably expressed in the iPSCs, and
208 558 (41.3%) were classified as differentially expressed between the species, a slightly higher
209 proportion than expected by chance alone (P = 0.1). We further classified the DMRs as either
210 'promoter', 'genic' or 'mixed' depending on their position relative to annotated gene transcripts (see
211 methods). The overall set of DMRs, as well as genic DMRs, are significantly associated with 4 and 79
212 GO BP terms respectively (FDR < 5%), including terms related to neurogenesis and skeletal system
213 development. Enrichment of several terms related to neurogenesis and skeletal system development
214 is likewise marginally significant amongst promoter and mixed DMRs (supplementary file 8).
215 However, the subset of inter-species DE genes that are also associated with DMRs are not
216 significantly enriched with annotation for any GO BP or MF terms.
217
218 ***Comparative histone modification data***
219 We used ChIP-seq to characterize the genome-wide distribution of two types of histone
220 modifications (H3K27me3 and H3K27ac) in three of our chimpanzee iPSCs (see methods). We

compared the chimpanzee data to histone modification data from three human iPSC lines from the Roadmap Epigenomics project (figure 5). To do so, we downloaded raw sequence files from GEO and processed data from both species using the same pipeline (see methods). We identified ChIP-seq peaks using MACS or RSEG, as appropriate, and accounted for differences in genome sequence between the species as well as for incomplete power to identify peaks across species (see methods). To relate the ChIP-seq data to genes (and integrate over data from all peaks that are in proximity to a given gene), we then generated enrichment ChIP scores for a set of previously defined 26,115 orthologous transcription start sites (TSSs, from Zhou et al. 2014). The enrichment score (see methods for details, also supplementary file 9), reflects the ratio of mapped ChIP-seq read counts across all peaks within a 4kb window centred on an orthologous TSS, relative to the genome-wide read count average after adjusting expectations based on the input control sample. We chose to classify as 'enriched' any region where the mean enrichment score across all three individuals in the species was larger than 1. This cut-off is arbitrary, but we confirmed that our qualitative results are robust by additionally testing enrichment cut-offs of 2, 5, and 10.

Using this approach, we first examined genome-wide patterns of H3K27me3 enrichment in chimpanzee and human iPSCs. Overlap across the two species is considerably higher than expected by chance (figure 5a, χ^2 $P < 10^{-16}$), but it is somewhat unclear how to interpret this observation with respect to the expectation that human and chimpanzee iPSCs would have similar pluripotency potential. We thus focused on a set of 3,913 genes (Li et al. 2013) previously annotated as bivalently modified in human PSCs – i.e., genes known to be associated with both high H3K4me3 and H3K27me3, indicative of a 'poised' or 'primed' state (Bernstein et al. 2006). We expect the vast majority of these genes to also be associated with similar modifications in chimpanzee iPSCs. Only 2,910 of the known bivalent genes were associated with clear orthologous TSSs and could be tested using our comparative H3K27me3 ChIP-seq data. Of these, 306 were not associated with the modification in either species, whereas of the 2,604 genes that were associated with H3K27me3 in at least one species, 2,368 (90.1%) were enriched for H3K27me3 in both species (figure 5b, χ^2 $P < 10^{-16}$).

We then examined H3K27ac enrichment patterns in both species. This mark is indicative of active promoters and gene transcription. Overall, we find good agreement between human and chimpanzee genes enriched for H3K27ac, with 95.8% human genes associated with the mark also

enriched in chimpanzees (figure 5c). However, there is a clear excess of genome-wide H3K27ac signal in chimpanzee iPSCs relative to humans, possibly due to an overall more sensitive ChIP enrichment in the chimpanzee samples (figure 5-figure supplements 1 and 2).

We proceeded by focusing on a list of 22 core pluripotency transcription factors (taken from Ng and Surani 2011; Orkin and Hochedlinger 2011), where we expect to find H3K27ac signal shared across the two species at a higher rate than in the genome-wide data, given the role of these factors in maintaining pluripotency. Due to our stringent requirements for establishing orthology, we were initially able to examine data from 14 of those genes; 11 of which were associated with H3K27ac in both species (figure 5d) - one of the discrepancies is *REX1 (*also known as *ZFP42*), which we discuss further below. We extended our analysis to include the full set of 22 pluripotency transcription factors regardless of orthology, by testing solely for absence or presence of signal peaks identified by MACS (that is, without considering enrichment scores; see methods). We again found a high overlap in H3K27ac enrichment across species, with 15 of the 22 genes associated with H3K27ac enrichment in both species (including the three master regulators of pluripotency, *OCT4*, *SOX2*, and *NANOG;* figure 5-figure supplement 3). Of the remaining 7 genes, one (*DAX1*) was not found to be associated with H3K27ac in either species, four genes (*ESSRB*, *KLF2*, *KLF4*, and *KLF5*) were associated with H3K27ac only in chimpanzee (although this observation may reflect incomplete power to detect peaks in the human data), and only two genes (*ZFX* and *REX1)* were associated with H3K27ac in human but not in chimpanzee iPSCs.

REX1 may be dispensable for chimpanzee pluripotency

In order to further consider inter-species differences in the core pluripotency regulatory network, we examined expression levels in our chimpanzee and human iPSCs in the same list of 22 core pluripotency TFs described above. Expression values in all iPSC lines are shown in figure 6a (see also figure supplement 6-1). Given the stringency of our interspecies analysis approach with respect to unique read mapping, we are unable to calculate RNA-seq-based expression estimates for six of these TFs, including *OCT4* or *NANOG*, both of which have multiple pseudogenes that can confound mapping algorithms (however, as shown in figure 1d, our qPCR results demonstrate that expression of those 2 genes is similar amongst all chimpanzee iPSC lines, and marginally higher

than in our human iPSC control line). Of the 16 TFs with expression data for iPSCs from both species, 4 (*E2F1, ESRRB, SALL4* and *REX1)* are differentially expressed between human and chimpanzee iPSCs at an FDR of 1%. Of these, *ESRRB* and *REX1* are associated with absolute inter-species expression $\log_2$ fold-changes > 1. However, because *ESRRB* is expressed at very low levels across all samples (mean RPKM across all 14 samples = 0.47), we focused our subsequent analyses on *REX1*, which is expressed at low or undetectable levels in 6 of our 7 chimpanzee iPSCs (mean RPKM = 0.667), but at high levels in all human iPSC lines (mean RPKM = 180.58) and a single chimpanzee iPSC, C6 (figure 6a). Our DNA methylation data is consistent with this gene expression pattern: all 10 probes located in the 5' UTR or up to 1,500 bp upstream from the *REX1* TSS are highly methylated in the six chimpanzee lines (mean β across all promoter probes = 0.87), but exhibit intermediate or low levels of DNA methylation in all of the human iPSC lines and the *REX1*-expressing C6 line (figure 6b); the entire region is a DMR (supplementary file 6). Consistent with these findings, *REX1* is also differentially enriched for H3K27ac signal in the two species – we identified no H3K27ac peaks at the *REX1* TSS in the three chimpanzee lines, which did not include C6 (figure 5d, figure 5-figure supplement 3).

The *REX1* genes codes for a transcription factor present in all placental mammal species, which has long been established as a marker of pluripotency in human and mouse PSCs (Brivanlou et al. 2003). Multiple publications have suggested that this gene plays an important role in maintaining pluripotency and inhibiting differentiation into the three primary tissue germ layers (Masui et al. 2008; Scotland et al. 2009; Son et al. 2013), with multiple mechanisms of action having been proposed. However, *REX1*-knockout mouse ESC lines can give rise to chimeric animals, and homozygous F2 *REX1* null mice are viable (Masui et al. 2008), suggesting that *REX1* may not be indispensable for murine pluripotency. In humans, loss of *REX1* expression in ESCs following shRNA knockdown has been associated with a rapid loss of pluripotency, as well as a decrease in glycolytic activity and a lack of observable mature mesodermal structures in teratoma formation assays (Son et al. 2013).

To determine the consequences of a lack of *REX1* expression in chimpanzee iPSCs, we considered gene expression data from all human iPSC lines and the 6 chimpanzee iPSC lines that do not express *REX1*. We asked whether there is an excess of DE genes among those thought to be

directly regulated by, or downstream of, *REX1* (figure 6c, d; see methods), but failed to find enrichment in all categories except for genes associated with GO term BP:0006096, glycolysis, where 19 of 34 testable genes were DE at an FDR of 1% between the two species ($p < 0.01$ from 100,000 permutations). The direction of this effect ran contrary to previous reports, however, with genes highlighted by (Son et al. 2013) as downregulated following *REX1* knockdown, such as *PGAM1* or *LDHA*, having significantly higher expression in chimpanzee iPSCs than in human iPSCS (figure 6d). Furthermore, the *REX1*-expressing line C6 is not an outlier amongst the other chimpanzee iPSC lines (figure 6d), suggesting that the observed inter-species regulatory differences cannot be attributed to differences in *REX1* expression between the species.

We note that both the teratomas and EBs generated from chimpanzee iPSC lines that do not express *REX1* gave rise to mature structures from all three germ layers similar to those observed in *REX1*-expressing line C6 (figure 2-figure supplements 1 and 3). Furthermore, and consistent with our observations, *REX1* is either absent or expressed at low levels in one replicate of either of the two retrovirally reprogrammed bonobo (*Pan paniscus*, sister species to chimpanzees) iPSC lines generated by (Marchetto et al. 2013b), although it is expressed in both replicates of both chimpanzee iPCS lines from the same group (figure supplement 6-2). Together, these findings suggest that that the variable loss of *REX1* expression in chimpanzee and bonobo iPSCs does not impair pluripotency, and that its regulatory functions of in humans may be being fulfilled in chimpanzee iPSCs by other regulatory mechanisms.

Comparison of iPSCs and other tissues

We collected RNA-sequencing data from all cell lines used to generate both the chimpanzee and human iPSCs (supplementary file 10). Following quality control and normalisation steps, we obtained RPKM values for 13,147 genes across all 28 iPSC and precursor samples (see methods). We also obtained DNA methylation profiles from all samples at the same 335,307 probes described above. PCA of both data sets show that the first PC was significantly associated with tissue type in both data sets ($P < 10^{-27}$ for the expression data; $P < 10^{-17}$ for the DNA methylation data; see figure 7 and supplementary files 11-12), while human and chimpanzee samples are separated by species along PC2 ($P = 0.001$ for the expression data; $P < 10^{-4}$ for the methylation data). However, given the

absence of chimpanzee LCLs in our dataset, it is not possible to determine whether the separation is driven by tissue type, species, or both.

Overall, chimpanzee iPSCs have significantly higher levels of DNA methylation compared to the somatic lines they were generated from ($P < 10^{-15}$; figure supplement 7-1), an observation that extends to all genomic features we tested (figure supplement 7-2); similar observations have been previously made in human PSCs (Bock et al. 2011; Nazor et al. 2012). Remarkably, both DNA methylation and gene expression levels in iPSCs are relatively homogeneous within species, far more so than in their corresponding precursor cells (figures 6b, 6d; $P < 10^{-14}$ when comparing overall pairwise distances within all chimpanzee iPSCs and within all chimpanzee fibroblasts in the methylation data; $P < 10^{-9}$ for the same comparison in the gene expression data). DNA methylation levels in iPSCs also have significantly reduced coefficients of variation relative to their precursor lines (range of CVs for chimpanzee iPSCs = 0.78 - 0.80, for chimpanzee fibroblasts = 0.87 - 0.90; $P < 10^{-06}$). We observed the same pattern in the human data, although in this case the multiple somatic origins of the cell lines of origin contribute to the higher level of variation.

We then performed analyses of gene expression and DNA methylation differences in the combined iPSC and somatic precursor dataset. First, we carried out a comparison of the iPSCs and the precursor cells within each species (see methods) and classified 9,235 genes as DE between chimpanzee fibroblasts and the corresponding iPSCs. In humans the number of DE genes is 7,765 if we consider all iPSC lines and their somatic precursors, 8,087 if we only consider those derived from LCLs (n = 5), and 5,489 if we only consider those derived from fibroblasts (n = 2; supplementary file 13). Similarly, we identified 18,029 DMRs between chimpanzee fibroblasts and iPSCs, and 12,078 DMRs between all human somatic precursors and all human iPSCs (supplementary files 14-15). No GO categories are significantly overrepresented in any of these data sets.

Next, we focused on a comparison of inter-species differences in gene expression and DNA methylation levels across cell types. Following joint normalisation and modelling of data from all samples (see methods), we classified 5,663 genes as DE between the chimpanzee precursor fibroblasts and the collection of human precursor LCLs and fibroblasts, as well as 84,747 DM probes and 9,107 DMRs (always at an FDR of 1%). Most of these regulatory differences, however, reflect variation across cell types rather than across species (6,324 genes and 70,312 probes are DE or DM

between the human fibroblasts and LCLs, respectively). We thus considered only data from the fibroblast precursors in the two species. Only 2 of the human iPSCs were reprogrammed from fibroblasts, leading to a loss in power; we were nonetheless able to identify 1,236 DE genes and 25,456 DM probes between human and chimpanzee fibroblasts, and 1,118 DE genes and 16,392 DM probes between the corresponding iPSCs of the two species. None of these gene sets were significantly enriched for functional annotations using GO BP terms. Although the overlap of inter-species DE genes and DM probes between the iPSCs and the precursors is considerable (13.6% of DE genes and 11.8% of DM probes), a large number of regulatory differences are only observed between the iPSC lines of the two species (figure supplement 7-3). This observation is robust with respect to different approaches to normalising and modelling the data (figure supplement 7-4), strongly suggesting that many of the differences we observe between our chimpanzee and human iPSC lines may be intrinsic features of the pluripotent state in these two species.

Discussion

Induced pluripotent stem cells have the potential to transform our understanding of the biology of non-model organisms and facilitate functional comparative studies. To this end, we have generated a panel of 7 fully characterized chimpanzee iPSCs. All lines are capable of spontaneously giving rise to the three tissue germ layers *in vitro* and *in vivo* and meet all currently established criteria for pluripotency. The chimpanzee iPSC lines provide a tantalising avenue for investigating how changes in gene expression and regulation underlie the architecture of complex phenotypic traits in humans and our closest living relatives (Gallego Romero et al. 2012; Marchetto et al. 2013a). In particular, we believe that through the use of directed differentiation protocols, functional studies could be performed in cell types where strong *a priori* hypotheses support a role for selective pressure underlying inter-species divergence (e.g. liver, heart, kidney (Blekhman et al. 2010; Blekhman et al. 2008)). In that sense, we hope that this panel of cell lines will be a useful tool to researchers interested in overcoming current limitations of comparative studies in primates. To that purpose, all chimpanzee iPSC described in this publication the panel are available fully and without restrictions to other investigators upon request to the corresponding authors.

Other groups have previously generated pluripotent stem cells from primates (Ben-Nun et al. 2011; Chan et al. 2010; Deleidi et al. 2011; Hong et al. 2014; Liu et al. 2008; Okamoto and Takahashi 2011; Tomioka et al. 2010; Wu et al. 2012; Wu et al. 2010; Wunderlich et al. 2014). Indeed, a recent publication (Marchetto et al. 2013b) reported the generation of two chimpanzee and two bonobo (*Pan paniscus*) iPSC lines through the use of retroviral vectors. However, in the course of our work we have found that retroviral vector silencing in chimpanzee iPSCs was not as stable as in human iPSC lines generated at the same time using the same method (see methods and figure 1-figure supplement 5). Our use of episomal vectors circumvents this problem, and more broadly the problems of both random exogenous gene reactivation and disruption of the host genome through retroviral integration (Sommer et al. 2012).

More generally, while the sum total of primate PSC generation efforts so far has resulted in a sizable number of lines being established from various donors and species, these have been generated through various reprogramming protocols and source cell types. We have generated iPSCs from a panel of seven individuals using a consistent protocol and cell type of origin. Given the panel size, it is powerful enough to robustly detect inter-species differences in gene expression, splicing and regulation. The fact that our panel contains both female and male lines also allows for future studies of sex-specific differences in gene expression in various cell types. Indeed, we have previously shown that this can be accomplished using as few as six individuals from each species (Blekhman et al. 2010).

Beyond its future applications, however, our panel has already yielded insights into the pluripotent state in chimpanzees and humans. On the one hand, both at the transcriptional and epigenetic level, our iPSCs are remarkably homogeneous both within and between species, significantly more so than their precursors cells. This finding aligns with our current understanding of the reprogrammed pluripotent state as a complex, highly regulated state (Jaenisch and Young 2008), deviations from which are likely to result in loss of pluripotency and lineage commitment. Additional support for this notion was provided by the strong overlap in H3K27me3 signal between the two species, especially in known bivalent genes. It is remarkable that we have been able to observe this considerable conserved chromatin signature despite the obvious confounding technical batch effect in these comparative data.

On the other hand, we were also able to identify over 4,500 genes that are differentially expressed between human and chimpanzee iPSCs, as well as over 3,500 differentially methylated regions between the two species. These numbers are greater than what has been previously observed in comparisons of other tissues across humans and chimpanzees with similar sample sizes (Blekhman et al. 2010; Blekhman et al. 2008). We believe that the reasons for this difference are likely to primarily stem from increased power to detect DE genes and DMRs in our sample relative to previous work. Given the small amount of intra-species variability we observed in RNA-seq and methylation relative to other tissues, we expect to have greater power to detect small, statistically significantly inter-species differences that would have been missed by studies that consider more variable tissue samples. This notion is supported by the fact that the fraction of genes with log FC < 2 we detect as DE between human and chimpanzee iPSCs is greater than in other comparison we have performed with any other tissue (figure 4-figure supplement 2). Though small in magnitude, we expect that a subset of these regulatory differences may be biologically relevant (e.g., we find that inter-species regulatory differences in methylation levels are enriched in regions associated with developmental processes; supplementary files 5 and 6).

We specifically highlighted an inter-species difference in *REX1* expression levels. This gene is considered an indispensible pluripotency marker in human and mouse PSCs, but our observations suggest that it may not be the case in chimpanzees. Although only one chimpanzee iPSC line expresses *REX1*, we were unable to identify any systematic differences between our human and chimpanzee iPSCs that would indicate a reduction in pluripotency. We also examined *REX1* expression levels in 73 human iPSC lines generated in-house from Caucasian individuals using the Illumina HT12v4 array (figure 6-figure supplement 3). All lines had PluriTest pluripotency scores > 20, yet 3 of 73 lines (4.1%) showed levels of *REX1* expression that were indistinguishable from background signal, suggesting that *REX1* may not be expressed in these individuals despite their high pluripotency scores. We also examined methylation status at the *REX1* locus in previously published human ESCs and iPSCs from (Ziller et al. 2011), and found that although all ESC lines examined exhibited consistent levels of low methylation at the *REX1* promoter, human iPSC lines analysed in exhibited either hemi- or hyper-methylated *REX1* promoter regions (figure 6-figure supplement 4). In the absence of publicly available *REX1* expression data from either of the hiPSC

lines with hypermethylated promoters we cannot be certain that the gene is not expressed in these lines, but the combination of these findings with our observations above and previous literature suggest that *REX1* may be important in regulation and maintenance of pluripotency in ESCs, but not necessarily so in iPSCS.

Additionally, in chimpanzees, the *REX1* gene has undergone multiple deletions and insertions relative to the human sequence, most significantly a 647 bp insertion in its first intron, and two insertions in the 3' UTR region of approximately 300 bp each that may disrupt the local regulatory landscape; the gene has also been duplicated, with a second copy retrotransposed into chromosome 14; none of these changes are shared with gorillas or orang-utans. Although it is currently unclear whether some or all of these changes are also present in the bonobo, these findings might explain why we observed low or no *REX1* expression more frequently in chimpanzees than in humans, and suggests that the gene may not be necessary for maintaining pluripotency in the *Pan* lineage.

PSCs have been used to study developmental pathways *in vitro* (for example, Paige et al. 2012; Rada-Iglesias et al. 2012; Wamstad et al. 2012; Xie et al. 2013). Although optimization of existing differentiation protocols will likely be necessary for application in the chimpanzee system, our panel of iPSCs makes it possible to carry out comparative developmental studies between humans and chimpanzees, and firmly test the hypothesis that changes in gene regulation and expression, especially during development, underlie phenotypic differences between closely related species, especially primates (Britten and Davidson 1971; Carroll 2005; Carroll 2008; Jacob 1977; King and Wilson 1975). In addition, we should be able to recreate and test the effect of inter-species regulatory changes in the correct cell type and species environment, enabling studies that cannot otherwise be performed in humans and non-human primates. The use of panels of iPSCs including lines from both humans and non-human primates will thus allow us to gain unique insights into the genetic and regulatory basis for human-specific adaptations.

Materials and Methods:

Isolation and culture of fibroblasts

All biopsies and animal care were conducted by the Yerkes Primate Research Center of Emory University under protocol 006-12, in full accordance with IACUC protocols. Skin punch biopsies (3 mm) were rinsed in DPBS containing Primocin (Invivogen) and penicillin/streptomycin (Pen/Strep, Corning) and manually dissected into 10-15 smaller pieces. The tissue was digested in 0.5 % collagenase B (Roche) for 1-2 hours until cells were released from the extracellular matrix. Dissociated cells were pelleted by centrifugation at 250 x g, and the supernatant was spun a second time at 700 x g to pellet any cells that had not been completely released from the extracellular matrix. Cell pellets were resuspended in a 1:1 mixture of α-MEM and F12 (both from Life Technologies) supplemented with 10% FBS (JR Scientific), NEAA, GlutaMAX (both from Life Technologies), 1% Pen/Strep, 64 mg/L L-ascorbic acid 2-phosphate sesquimagnesium salt hydrate (Santa Cruz Biotech) and Primocin. Cells were plated in a single well of a 6-well plate coated with 4 μg/cm^2 human fibronectin (BD Biosciences) and 2 μg/cm^2 mouse laminin (Stemgent). Cultures were grown at 5% CO_2/5% O_2 until confluent and then split using 0.05% trypsin. For routine passaging cells were cultured at 5% CO_2 and atmospheric oxygen in primate fibroblast media, which is the same as plating media but does not contain F12 base media.

Generation of retrovirally-reprogrammed iPSC lines (a failed attempt)

We initially attempted to generate lines by retroviral transduction through transfection with pMXs- vectors encoding the human *OCT3/4*, *SOX2*, *KLF4*, *L-MYC* and *NANOG* sequences (Addgene plasmids 17217, 17218, 17219, 26022 and 18115) as well as vectors encoding the MSCV-VSV.G envelope protein (Addgene plasmid 14888) and MSCV gag-pol (Addgene plasmid 14887). 15 μg of each vector was transfected into 293FT cells (Life Technologies) using Lipofectamine 2000 (Life Technologies) as directed by the manufacturer. We collected virus-containing supernatant from the 293FT cells 48 and 72 h after transfection and immediately used this viral media to transduce chimpanzee fibroblasts, alongside 10 ug/mL of polybrene (Sigma Aldrich H9268). To aid viral penetration, we centrifuged the cells at 1800 RPM for 45 minutes following each transduction. 24 h after the second transduction, we replaced the viral media with A-MEM + 10% FBS, NEAA and Glutamax. Transduced fibroblasts were allowed to recover for a further 2 days and then seeded on γ-irradiated, CF-1-derived mouse embryonic fibroblasts (MEF) at a density of 10,000 cells/cm^2, and

maintained in hESC media (DMEM/F12 supplemented with 20% KOSR, 0.1 mM NEAA, 2 mM GlutaMAX, 1% Pen/Strep, 0.1 mM BME and 25 ng/mL human bFGF) supplemented with 0.5 mM valproic acid (Stemgent) until day 14. We obtained iPSCs from 5 chimpanzees by using this protocol. Yet, when we performed quality control and pluripotency checks on these lines we found that the exogenous transfected genes were still expressed (figure 1-figure supplement 5). Pluripotency in these lines could not be maintained exclusively through endogenous expression. We discarded all 5 lines and proceeded with a different reprograming strategy as detailed below.

Generation of episomally-reprogrammed iPSC lines

Fibroblasts were grown at 5% CO_2/atmospheric O_2 in primate fibroblast media until 70-80% confluence and released by trypsinisation for transfection. 1.5×10^6 cells were transfected with 1.5 μg per episomal vector containing the following genes: *OCT3/4*, *SHp53*, *SOX2*, *KLF4*, *LIN28*, and L-MYC (Addgene plasmids 27077, 27078, 27080 and 27082; (Okita et al. 2011)). To boost the initial retention of vectors following transfection, 3 μg of *in vitro* transcribed ARCA capped/polyadenylated EBNA1 mRNA was cotransfected with the vectors (see below). Transfected cells were seeded at 15,000/cm^2 on tissue culture plates precoated with 1 μg/cm^2 vitronectin (Stemcell Technologies). Cells were grown in Essential 8 media (made in house as previously described in (Chen et al. 2011)) without *TGFβ1,* supplemented with 0.5 mM sodium butyrate (NaB, Stemgent)) and 100 nM hydrocortisone (Sigma Aldritch). Hydrocortisone was used between days 1-12, or until cell density exceeded >70% confluence. At day 12, cells were detached using TrypLE (Life Technologies) and replated at a density of 5,000 cells/cm^2 on cell culture dishes precoated with 0.01 mg/cm^2 (1:100) of hESC-grade Matrigel (BD Sciences) and grow in Essential 8 media without *TGFβ1* or NaB. Colonies began to form at days 18-22 and were picked between days 24-30 onto dishes coated with γ-irradiated CF-1 derived MEF and subsequently grown in hESC media (as described above) supplemented with 100 ng/mL human bFGF (Milteny Biotech). Clones were routinely split using Rho-associated kinase (ROCK) inhibitor Y27632 (Tocris) at a concentration of 10 μM. Cells were migrated to 1:100 hESC Matrigel (BD Sciences) and maintained on Essential 8 media after a minimum of 15 passages on MEF. Feeder free cells were passaged using EDTA-based cell release solution as in (Chen et al. 2011).

538

539 ***Generation of* EBNA1 *mRNA***

540 To generate a template for in vitro transcription, an EBNA1 template was designed using the
541 wild type HHV4 *EBNA1* as a reference sequence (NCBI accession YP_401677.1). The reference
542 sequence was modified by replacing the GA repeat region and domain B (amino acids 90-375) with a
543 second, tandem, chromatin-binding domain (domain A, amino acids 27-89), similar to what was
544 done by (Howden et al. 2006). The nuclear localization signal (amino acids 379-386) was removed
545 and replaced with the sequence GRSS. Using the amino acid sequence as the starting template, the
546 corresponding DNA sequence was generated by reverse translation and optimized for expression in
547 human cell lines using Genscript's OptimumGene codon algorithm. This sequence was synthesized
548 by Genscript and provided in the pUC57 cloning vector; the EBNA1 coding sequence was subcloned
549 into pcDNA3.1+ (Life Technologies) using the restriction enzymes BamHI and HindIII. Capped and
550 poly(A) mRNA transcripts were generated using the mMESSAGE mMACHINE T7 ULTRA kit (Life
551 Technologies) with 1 μg of BamHI linearized pcDNA3.1+EBNA1 as the template. The plasmids
552 encoding the wild type and modified EBNA1 sequences have been deposited to Addgene as
553 plasmid ID#s 59199 and 59198 for the wild type and modified sequences respectively.

554

555 ***iPSC characterization by immunocytochemistry***

556 iPSC colonies were cultured on MEF for 4-6 days and fixed using PBS containing 4% PFA
557 (Santa Cruz BioTech) for 15 minutes at room temperature. After rinsing with PBS, fixed cells were
558 blocked and permeabilised for one hour in PBS containing 0.3% triton and 5% BSA. Primary
559 antibodies: OCT3/4 (SC-5279), SOX2 (SC-17320), NANOG (SC-33759), SSEA-4 (SC-21704), and Tra-1-
560 81 (SC-21706), all from Santa Cruz BioTech, were diluted 1:100 in blocking solution. Fixed cells were
561 incubated with the primary antibody solution overnight on a rocker at 4 °C. After washing out the
562 primary antibody solution, fixed cells were incubated with secondary antibodies (labeled with either
563 Alexa-488 or Alexa-594, 1:400, Life Technologies) diluted in blocking for 1 hour on a rocker at room
564 temperature. Nuclei were counterstained using 1 μg/mL Hoechst 33342 (ThermoFisher). All
565 fluorescence imaging was conducted using an AMG EVOS FL (Life Technologies).

566

567 ***Quantitative PCR for endogenous and exogenous gene expression***

568 RNA was extracted using Qiagen RNA miniprep columns from cell pellets collected from
569 fibroblasts, day 7 post transfection and feeder free (Matrigel and Essential 8) iPSC lines at passage 10
570 or higher for both the retroviral and episomal reprogrammings; 1 μg of total RNA was reverse
571 transcribed using the Maxima first strand cDNA synthesis kit (Thermo Scientific). Quantitative PCR
572 was performed using a 1:96 dilution of cDNA and SYBR Select master mix (Life Technologies) with
573 both forward and reverse primers at a concentration of 0.2 μM.). Data was collected and analysed
574 using the Viia7 (Life Technologies). Primer sequences are shown in supplementary file 2, exogenous
575 gene expression melt curves are shown in figure 1-figure supplement 5.
576
577 ***Generation of embryoid bodies and immunofluorescence***

578 Colonies growing on MEF were detached using Dispase/Collagenase IV (1 mg/ml each; both
579 from Life Technologies) in DMEM/F12 and grown as a suspension culture on low adherent plates
580 using hESC media without bFGF. After one week of suspension growth, cells were transferred to 12
581 or 24-well plates coated with 0.1% gelatin and grown in DMEM supplemented with 20% FBS, 0.1
582 mM nonessential amino acids, 2 mM GlutaMAX, 1% Pen/Strep and 64 μg/mL L-Ascorbic acid 2-
583 phosphate sesquimagnesium salt hydrate. Embryoid bodies were grown for 1-2 weeks prior to
584 fixation and immunofluorescence staining. Cultures were fixed and stained as described above
585 using the following antibodies: AFP (1:200, SC-130302, Santa Cruz Biotech), FOXA2 (1:200, SC-6554,
586 Santa Cruz Biotech), α-smooth muscle actin (1:1500, CBL171, Millipore) and MAP2 (1:200, sc-20172
587 and sc-74420, Santa Cruz Biotech).
588
589 ***Integration analysis***

590 To test for genomic integration and residual retention of episomal plasmids, each iPSC line
591 was migrated to feeder free conditions and grown beyond passage 15 on hESC-qualified Matrigel
592 (1:100 dilution, BD Biosciences) coated plates in Essential 8 media (Life Technologies). DNA was
593 extracted from feeder free cultures using DNeasy Blood and Tissue Kits (Qiagen). PCR was
594 performed using 100 ng of genomic DNA, an annealing temperature of 72°C and 25 cycles using
595 primers designed to amplify a region common to all episomal vectors used (supplementary file 2).

Genomic DNA (100 ng) isolated from day 7 cultures, and 1 pg of each episomal vector were used as positive controls. PCR products were run on a 1% agarose gel and visualised using ethidium bromide.

Karyotyping

After 15 passages on MEF and hESC media, cells were migrated to 1:100 hESC Matrigel (BD Sciences) and maintained on Essential 8 media for upwards of 6 passages. Feeder-free adapted cells were sent to Cell Line Genetics Inc (Madison, WI) for karyotyping as described in (Meisner and Johnson 2008).

Teratoma formation assays

In vivo developmental potential of the reprogrammed cell lines was examined. Monolayer iPSCs from three chimpanzee lines were grown on Matrigel (1:100) in E8 medium (Life Technologies) and collected by EDTA treatment (Life Technologies). Cells were counted and resuspended at a ratio of 1:1 cell volume to Matrigel and kept on ice until the injection. Six-week-old CB17.Cg-*Prkdc*^{scid}*Lyst*$^{bg\text{-}J}$/Crl immunodeficient male mice were obtained (Charles River Laboratories) and approximately one million iPSCs for each clone were injected into the testis-capsule. After five to eight weeks teratomas were isolated, weighed, measured, dissected, and fixed in 10% formalin. The specimens were embedded in paraffin, stained with hematoxylin and eosin, and analyzed by a histopathologist. All animal work was conducted under the approval of the Institutional Care and Use Committee of UCSD (Protocol# S09090).

In addition, live feeder free iPSC cultures maintained in Essential 8 media on Matrigel iPSCs from C4955 (passage 15+7) were provided to Applied Stem Cell Inc. (Menlo Park, CA) for teratoma analysis as previously described (Chen et al. 2012).

Species-of-origin identity of teratoma samples

DNA was extracted from frozen teratoma tissue using DNeasy Blood and Tissue Kits (Qiagen). For teratomas derived from individual C4955, core sections were isolated from FFPE embedded teratomas tissue using a 3 mm dermal punch tool; DNA was extracted from core samples

using a QIAamp DNA FFPE Tissue Kit (Qiagen). PCR was performed using universal mitochondrial primers ((Kocher et al. 1989) supplementary file 2) amplifying cytochrome b (*Cytb*, chimpanzee reference sequence NC_001643:bp 14233-14598) or the 12S ribosomal gene (*12S*, NC_001643:bp 484-915) with 250-500 ng of genomic DNA as the starting template. Two-step PCR was conducted with an annealing temperature of 50°C for 1 minute and an extension step at 72°C for 4 minutes for a total of 30 cycles. DNA was purified using a Wizard SV gel and PCR Clean-up kit (Promega); dye terminator cycle sequencing was conducted by the University of Chicago Comprehensive Cancer Center using 60 ng of purified PCR template and 4 μM of either the forward or reverse primer. Alignment to the chimpanzee, human (NC_012920) and mouse (NC_005089) reference sequences was accomplished using CLC Main Workbench 6.9 (Qiagen) and MUSCLE (Edgar 2004).

Directed differentiation of chimpanzee iPSCs to hepatocytes and cardiomyocytes

In order to demonstrate that chimpanzee iPSCs can be directly differentiated into other cell types, we differentiated C2 iPSC into hepatocytes and C7 into cardiomyocytes using the published protocols of (Cheng et al. 2012) and (Lian et al. 2013) respectively, with the following modifications: In both cases we plated iPSCs at 0.35 x 10^6 cells/cm^2 in 0.44 mL/cm^2 and cultured them in Essential 8 media 24 hours prior to initiating all differentiations. To increase hepatocyte differentiation efficiency, 1 μM of sodium butyrate was added during the first 24 hours of differentiation. After 24 days of differentiation, cells were immunostained as described above with a primary antibody for albumin (1:200, A6684, Sigma Aldrich; figure 2-figure supplement 2).

After 10 days of differentiation, differentiated C7 cultures were enriched for cardiomyoctes by culture in RPMI based media without glucose supplemented with 5 mM sodium DL-lactate for 10 days as described previously (Burridge et al. 2014; Tohyama et al. 2013). After day 20 purified cardiomyocytes were cultured in media lacking glucose supplemented with 10 mM galactose (Rana et al. 2012). After 25 days of cardiac differentiation, we characterized calcium flux in and out of iPSC-derived cardiomyocytes by treating cultures with 5 μM Fluo-4 AM (F-14217, Life Technologies) for 15 minutes, washing cultures once and imaging them with an AMG EVOS FL microscope (video 1).

654 ***Microarray genotyping and PluriTest***

655 RNA from passage ≥15 iPSCs was extracted using the QIAGEN RNeasy kit according to the
656 manufacturer's instructions. Quality of the extracted RNA was assessed using an Agilent
657 Bioanalyzer 2100 (RIN scores for all samples ranged from 9.9 to 10), and RNA was processed into
658 biotinylated cRNA and hybridized to the HT12v4 array using standard Illumina reagents as directed
659 by the manufacturer. Arrays were scanned using an Illumina HiScan, and data processed using
660 Illumina's GenomeStudio software. Using these data, we carried out PluriTest as previously
661 described (Müller et al. 2011). Additionally, we mapped all detected HT12v4 probe sequences (n =
662 46,297) to the chimpanzee (panTro3) genome using BWA 0.6.3 (Li and Durbin 2009). Probes that
663 mapped to a single genomic location with no mismatches were retained (n = 21, 320, 46.2% of all
664 probes) for the analysis that was restricted only to the chimpanzee lines.

665 When we considered data from human and chimpanzee iPSCs together, without excluding
666 probes based on sequence matches to the chimpanzee genome, all chimpanzee lines in the panel had
667 pluripotency scores slightly below the pluripotency threshold (figure supplement 3-1, lighter points).
668 However, low pluripotency scores could stem from differences in our ability to estimate gene
669 expression levels in the chimpanzee compared to the human due to attenuated hybridization caused
670 by sequence divergence (Gilad et al. 2005). Indeed, when we subset the array to retain only those
671 detected probes that map to the chimpanzee genome with no ambiguity or mismatches, all
672 chimpanzee lines have pluripotency scores greater than the pluripotency threshold value of 20
673 (figure 3-figure supplement 1, darker points).

674

675 ***RNA sequencing and differential expression testing between iPSCs***

676 50bp single-end RNA sequencing libraries were generated from RNA extracted from 7
677 chimpanzee and 7 human iPSC lines using the Illumina TruSeq kit as directed by the manufacturer
678 (San Diego, CA), as well as from their precursor fibroblast or LCL cell lines. All iPSC samples were
679 multiplexed and sequenced on four lanes of an Illumina HiSeq 2500; while the precursor cell lines
680 were multiplexed and sequenced on six lanes of the same sequencer. We generated a minimum of
681 28,010,126 raw reads per sample (supplementary file 10), and confirmed the raw data were of high
682 quality using FastQC (available online at http://www.bioinformatics.babraham.ac.uk/projects/

683 fastqc/). We mapped raw reads to the chimpanzee (panTro3) or human (hg19) genome as
684 appropriate using TopHat 2.0.8 (Trapnell et al. 2009), allowing for a maximum of 2 mismatches in
685 each read. Due to the relatively poor annotation of the chimpanzee genome and to prevent biases in
686 expression level estimates due to differences in mRNA transcript size and genetic divergence
687 between the two species, we limited the analysis to reads that mapped to a list of orthologous
688 metaexons across 30,030 Ensembl genes drawn from hg19 and panTro3, as in (Blekhman et al. 2010).
689 Following mapping, gene level read counts were generated using *featureCounts* 1.4.4 as implemented
690 in Subread (Liao et al. 2013). Due to mapping biases between human and chimpanzee ribosomal
691 proteins and pseudogenes, we removed all genes associated with the Gene Ontology Cellular
692 Compartment category 'ribosome' (GO:0005840, n = 141) and all annotated pseudogenes in Ensembl
693 release 65 (n = 3170, December 2011, the oldest available archival version of Ensembl) from the data
694 at this point.

695 We considered two normalization approaches in our analysis. In one instance, we examined
696 only RNA-sequencing data from chimpanzee and human iPSCs, and retained 12,171 genes with at
697 least 4 observations in one of the two species of $\log_2$ CPM > 1. CPM were then loess normalized by
698 species within individuals with *voom* (Law et al. 2014). As the orthologous genes are not constrained
699 to be the same length in both species, we computed RPKM for each gene before carrying out any
700 inter-species comparisons. We then used the R/Bioconductor package *limma* 3.20.3 (Smyth 2004) to
701 test for differential expression in our RNA-seq data, with a model that included only a species effect.
702 Finally, we tested for an enrichment of GO categories amongst DE genes using the R package *topGO*
703 2.16.0 (Alexa et al. 2006). These normalised values were used only to identify genes differentially
704 expressed between iPSCs of the two species.

705 For the dataset containing RNA-sequencing data from iPSCs and their precursors, we again
706 only retained 13,147 genes with at least 4 observations in one of the four groups (chimpanzee iPSCs,
707 chimpanzee precursors, human iPSCs or human precursors) of $\log_2$ CPM > 1. Gene counts were then
708 loess normalised within individuals by tissue, after correcting for the lack of independence within
709 different tissues from the same individual, through the function corfit. As above, we then computed
710 species-specific RPKM values, and used *limma* and *topGO* to test for differential expression and GO
711 category enrichment, respectively. In this instance, we used a model design with 6 parameters for

712 the main effect (chimpanzee iPSC, human LCL-derived iPSC, human fibroblast-derived iPSC,
713 chimpanzee fibroblast, human LCL and human fibroblast) and no additional covariates.
714 To confirm that our conclusions are robust with respect to the choice of normalization
715 procedure, in both cases, we also tried a variety of other normalization schemes, including
716 correcting for %GC content as in (Risso et al. 2011), none of which had a substantial effect on the
717 final results (supplementary file 16). Finally, we built neighbor joining trees using Manhattan
718 distances calculated from RPKM values at all 13,147genes using the nj function in the R library *ape*
719 (Paradis et al. 2004). All analyses were performed at a false discovery rate (Benjamini and Hochberg
720 1995) threshold of 1% unless otherwise noted, using R 3.1.0 (R Core Team 2013) and Bioconductor
721 2.14 (Gentleman et al. 2004).
722
723 ***DNA Methylation arrays***
724 To analyze DNA methylation, we extracted DNA from all chimpanzee and human iPSC lines
725 described above, as well as from the source fibroblast or lymphoblastoid cell lines. In all cases, 1000
726 ng of genomic DNA were bisulphite-converted and hybridized to the Infinium
727 HumanMethylation450 BeadChip at the University of Chicago Functional Genomics facility as
728 directed by the manufacturer. Since the probes on the array were designed using the human
729 reference genome, we followed the approach described in (Hernando-Herraez et al. 2013) to
730 compare humans and chimpanzees. We retained those probes that had either a perfect match to the
731 chimpanzee reference genome, or had 1 or 2 mismatches in the first 45 bp but no mismatches in the
732 3′ 5 bp closest to the CpG site being assayed. We also removed all probes that contained human
733 SNPs (MAF ≥0.05) or chimpanzee SNPs (MAF ≥0.15) within the last 5 bp of their binding site closest
734 to the CpG being assayed. Within each individual, probes with a detection $P > 0.01$ were excluded.
735 This resulted in the retention of 335,307 autosomal probes, and an additional 8,210 X chromosome
736 probes, which we normalized and analyzed separately by sex. In all cases we performed a two-color
737 channel signal adjustment, quantile normalization and β-value recalculation as implemented in the
738 lumi package (Du et al. 2008). Because the HumanMethylation450 BeadChip contains two assay
739 types which utilize different probe designs, we performed a BMIQ (beta mixture quantile method)
740 normalization (Teschendorff et al. 2013) on the quantile-normalized autosomal data set. We did not

perform this step on the X chromosome data, due to its methylation patterns. We built neighbor joining trees using Manhattan distances at all 335,307 probes using the nj function as above.

In order to identify differentially methylated probes we used an identical approach to that described above for the identification of DE genes. First, we identified probes that were differentially methylated between the iPSCs of both species using *limma* by using a reduced data set and model containing only data from the iPSCs themselves. Then, we fit a linear model to the data using limma with 6 parameters corresponding to the 6 tissue/species combinations in the data, classifying probes as differentially methylated at an FDR of 1%. As with the expression data, the reduced model has more power to identify DM probes between the two iPSC groups than the full model; however, there is great concordance between the two sets of results (figure supplement 7-5). We excluded all probes with mean β inter-group differences < 0.1 in order to group DM probes into DMRs, which we define as 2 or more DM probes separated by < 1kb, with the additional requirement that the effect be in the same direction in all DM probes within the region. Finally, to examine the content of these DMRs, we used annotation files for the HumanMethylation450 Bead Chip provided by the manufacturer and discarded all DMRs associated with either multiple or no genes. We tested for enrichment of GO BP categories amongst the genes contained in the DMRs by using the R package topGO 2.16.0 (Alexa et al. 2006), using as a background set all genes in which it is theoretically possible to detect DMRs.

H3K27ac and H3K27me3 ChIP-seq data

ChIP-seq assays were performed as previously described (Schmidt et al. 2009), with slight modifications. Specifically, approximately 60 million iPSCs from three chimpanzee individuals (C2, C5 and C7) were cross-linked with 1% formaldehyde for 10 minutes. Cells were lysed and chromatin sheared with a Covaris S2 (settings: 4 minutes, duty cycle 10%, 5 intensity, 200 cycles per burst in 4 6x16mm tubes per individual). H3K27ac- and H3K27me3-enriched regions were isolated using 5 ug of either H3K27ac antibody (ab4729, Abcam, Cambridge, MA, USA) or H3K27me3 antibody (07-449, Millipore, Billerica, MA, USA). ChIP and input DNA from each individual were end-repaired, A-tailed and ligated to Illumina Truseq sequencing adapters before 18 cycles of PCR amplification. 200-300 bp DNA fragments were selected for sequencing. Input libraries were multiplexed and

770 sequenced on one lane of an Illumina HiSeq2500 using the rapid run mode, ChIP libraries were
771 multiplexed and sequenced on three lanes of an Illumina HiSeq2500 using the rapid run mode.
772 For comparison purposes, we downloaded ChIP input, H3K27ac and H3K27me3 data from 3
773 human iPSC lines (iPS 6.9, iPS-18a, and iPS.20b, all of them release 5) generated by the Roadmap
774 Epigenomics Consortium (Roadmap Epigenomics Consortium et al. 2015) from the NIH GEO
775 database (supplementary file 10). Human and chimpanzee samples were mapped to either hg19 or
776 panTro3 using BWA 0.7.9 (Li and Durbin 2009); reads that mapped outside chromosomes 1-22 + X
777 were discarded, as were reads that did not map uniquely to a single genomic region with less than 2
778 mismatches, or reads that were marked by Picard (http://picard.sourceforge.net) as originating from
779 PCR duplicates.
780 After mapping and filtering, we used MACS 1.4.4 (Zhang et al. 2008) and RSEG 0.4.4 (Song
781 and Smith 2011) to identify peaks in the H3K27ac and H3K27me3 data respectively. Our analyses in
782 this section follow those of (Zhou et al. 2014). Briefly, for MACS, we specified an initial P-value
783 threshold of H3K27ac, 0.001, and used each line's ChIP input file for comparison. For RSEG, we used
784 the 'rseg-diff' function to compare H3K27me3 enrichment against each individual's ChIP input file,
785 with the recommended 20 maximum iterations for hidden Markov model training. We then filtered
786 enriched regions or peaks identified by either program by retaining only those that overlapped a
787 previously defined set of 200 bp orthologous windows (Zhou et al. 2014), where at least 80% of
788 bases are mappable across species using liftOver. We define mappability as the ability of each 20bp
789 kmer beginning in that window to be uniquely mapped to the genome.
790 To ensure that sequence divergence did not confound our analyses, we mapped each
791 identified region or peak in humans to the chimpanzee genome, and vice versa, using liftOver, and
792 excluded regions and peaks where 80% or greater of bases in the enriched peaks or regions failed to
793 align to the other genome. To further minimise the number of false positive results in our
794 interspecies comparison (due to incomplete power), we applied a two-step cutoff (Cain et al. 2011)
795 to the list of enriched regions and peaks. For H3K27ac, we retained all peaks that were identified
796 with a first, stringent cutoff of FDR < 5% in one species and a, second, relaxed cutoff of FDR < 15% in
797 the other, as in (Zhou et al. 2014). Because RSEG does not report FDR values for enriched regions,
798 we used each region's domain score, which is the sum of the posterior scores of all bins within the

799 domain, and set a first, stringent cutoff of 20 in one species, and a second, relaxed threshold
800 demanding only that the region be classified as 'enriched' by RSEG, without a specific score
801 requirement.

802 Having done this, we integrated data from multiple peaks (when present) to generate a gene-
803 level metric of ChIP signal in each individual. Specifically, we computed an enrichment score for
804 each histone mark in each individual in a set of previously defined 26,115 orthologous transcription
805 start sites (Zhou et al. 2014) by dividing RPKM values at each TSS at gene *i* for either mark minus
806 RPKM values in TSS at gene *i* for ChIP input, all of it over the genome-wide average RPKM for
807 either mark minus the genome-wide average RPKM for ChIP input. Given the way in which we
808 have defined this enrichment score, a score > 0 indicates those genes where we detected more
809 histone mark reads than input reads, while a score > 1 indicates a gene with an excess of histone
810 mark reads greater than what we would expect given the genome-wide distribution.

811 Because 8 of the 22 genes in the list of pluripotency master genes used to generate figure 5d
812 do not have clearly defined orthologous TSSs, we also examined whether MACS identified peaks in
813 the 2kb +/- TSS for all 22 genes and their orthologous position in the chimpanzee genome, identified
814 solely through liftOver – that is, without taking into account whether there is evidence for a TSS at
815 that position in the chimpanzee genome. To generate figure 5-figure supplement 3, we simply asked
816 how many of the 22 genes had at least 1 peak at an FDR < 5% in at least 1 individual in either species,
817 regardless of orthology and sequence conservation.

818 We note that since different labs produced the human and chimpanzee data, we expect a
819 considerable technical batch effect to be completely confounded with species annotation. Given this
820 study design, we expect the technical batch effect to result in the appearance of inter-species
821 differences; yet, our goal is to demonstrate similarity across species. Thus, our conclusions (of high
822 overlap across species), are conservative with respect to the technical batch effect.

823

824 ***REX1 expression and function***

825 To examine the possible consequences of reduced expression of *REX1* in chimpanzee iPSCs,
826 we retrieved genes that responded to a *REX1* knockdown in mESCs from supplementary files 2 and
827 3 of (Scotland et al. 2009) and converted Affymetrix MG 430 2.0 probe IDs to ENSM and ultimately

828 to orthologous ENSG identifiers using Biomart release 66 (to control for deprecated identifiers).
829 Because (Masui et al. 2008; Scotland et al. 2009; Son et al. 2013) have highlighted *REX1*'s function in
830 controlling cell cycle progression, glycolysis and cellular differentiation, we additionally retrieved
831 genes associated with these terms to generate figure 6c as follows: The core set of pluripotency TFs
832 are those described by (Orkin and Hochedlinger 2011) and (Ng and Surani 2011). Cell cycle and
833 glycolysis categories contain all genes associated with GO BP:0007049 and BP:0006096 respectively,
834 whereas cell fate contains genes associated with any GO term that contains the words "ectoderm",
835 "mesoderm" or "endoderm". We also examined individual examples of cell fate differentiation: CNS
836 development genes are associated with BP:0007417 or any of its offspring; cardiovascular system
837 development genes are associated with BP:0072358 or any of its offspring; hepatobiliary system
838 development genes are associated with BP:0055123 or any of its offspring. Confidence intervals
839 around the null hypothesis were generated independently for each category from 100,000
840 permutations in R.

841 Finally, to compare our data with a previously published set of chimpanzee and bonobo
842 iPSC lines (Marchetto et al. 2013b), we downloaded fastq files from GEO (Series GSE47626) and
843 mapped only the first mate from all reads using the same approach as above, but allowing 4
844 mismatches in the entire 100bp read. We normalised expression estimates jointly with our own cell
845 lines to generate figure supplement 6-2, and used all data points from a given species, irrespective of
846 origin, to generate boxplots.

847 To generate figure 6-figure supplement 3, we took Illumina HT12v4 array data from 73
848 human iPSC lines generated in house and calculated Pluritest pluripotency and novelty scores as
849 above, using the full probe set. Independently, we normalised and background-corrected the raw
850 array intensities using the lumiExpresso function in *lumi* and extracted expression values for all 73
851 human iPSC lines at the single array probe associated with *REX1*.

852 In order to examine *REX1* methylation levels in other human PSCs, we obtained Infinium
853 HumanMethylation450 BeadChip for the lines reported in (Ziller et al. 2011) from the authors, and
854 normalised it jointly with our own data as described above. We then extracted normalised
855 methylation β levels at the 13 probes that map to *REX1* in both chimpanzees and humans to
856 generate figure 6-figure supplement 4.

857

858 ***Other indicators of genomic stability***

859 Finally, we assessed two broad indicators of stability in our chimpanzee lines. All iPSC lines
860 derived from female chimpanzees, and 3 of 4 lines derived from human females, show strong
861 evidence for elevated expression of XIST relative to male lines (FDR-adjusted $P = 0.0010$; figure 7-
862 figure supplement 6) and maintenance of X-chromosome inactivation during pluripotency. X-
863 chromosome methylation patterns in females corroborate these observations, with the majority of
864 probes mapping to the X-chromosome in our data being either hemimethylated ($0.2 < \beta < 0.8$) or
865 hypermethylated ($\beta \geq 0.8$) in females but not in males (figure 7-figure supplement 7). We also used a
866 list of 168 imprinted probes from (Ma et al. 2014) to check for maintenance of genomic imprinting
867 after reprogramming. We find that the majority of imprinted loci remain hemimethylated following
868 reprogramming in both human and chimpanzee iPSC lines (figure 7-figure supplement 8). However,
869 we identify two sets of probes that are consistently hypermethylated in pluripotent lines but were
870 hemimethylated in their precursor cells. The first cluster contains 5 probes that are hypermethylated
871 across both chimpanzee and human iPSCs; these probes are associated with the genes *KCNK9,*
872 *ANKRD11* and *MKRN3*. The second cluster is comprised of 21 probes that are hypermethylated in
873 all human iPSCs but only 2 chimpanzee iPSCs in our data, and is associated with the gene *PEG3-*
874 *ZIM2*, which has been previously shown to be abnormally methylated in both hESCs and hiPSCs
875 (Lund et al. 2012).

876

877 ***Data Access*:**

878 All novel RNA-sequencing, DNA methylation and ChIP-seq data are available at the GEO
879 under SuperSeries number GSE61343. Additionally, a table with *P*-values for all hypothesis testing
880 performed using the methylation data (by probe) is available on the Gilad lab website
881 (http://giladlab.uchicago.edu/ Data.html).

882 All chimpanzee iPSC lines described in this publication are available fully and without
883 restrictions to other investigators upon request to the corresponding authors.

884

885

Acknowledgements

We thank members of the Gilad and Marques-Bonet labs for helpful discussions; Julien Roux for help with differential gene expression analysis; Alexander Meissner for sharing methylation data from human iPSC and ESC lines, Fred Gage for providing a list of differentially expressed genes in retrovirally-reprogrammed chimpanzee iPSCs, and an anonymous reviewer for suggesting that we add comparative ChIP-seq data from H3K27ac and H3K27me3. This work was supported by NIH grant GM077959 to YG as well as by grants from the California Institute for Regenerative Medicine (CIRM) CL1-00502 and TR01250 to JFL, and ERC Starting Grant 260372 and MICINN (Spain) BFU2011-28549 to TM-B. IGR is supported by a Sir Henry Wellcome Postdoctoral Fellowship; IHH is supported by FI Generalitat Catalunya; MCW is supported by an EMBO Long-Term Fellowship (ALTF 751-2014) and the European Commission Marie Curie Actions; NEB is supported by an NIH training grant (GM007197) and an NIH pre-doctoral award (F31 AG 044948); LCL is supported by a WRHR Career Development award (NIH K12 HD001259) and the UCSD Department of Reproductive Medicine. We also acknowledge the generous support of the Yerkes Primate Center through their grant ORIP/OD P51OD011132.

Disclosure Declaration:

The authors declare no conflicts of interests.

Figure legends:

Figure 1: Characterization of chimpanzee iPSC lines. a. Phase contrast image of representative chimpanzee iPSC line. Scale bar: 1000 μm. b. Representative karyotype from chimpanzee iPSC line after > 15 passages, showing no abnormalities. c. ICC staining of iPSC lines with antibodies for pluripotency markers as indicated. Scale bar: 200 μm. d. Quantitative PCR testing for expression of endogenous pluripotency factors in all 7 chimpanzee iPSC lines. Line H20961 is a male human iPSC line generated in-house used as reference. e. PCR gel showing an absence of exogenous episomal reprogramming factors in all 7 chimpanzee iPSC lines. All PCRs were carried out on templates extracted from passage >15 with the exception of C3651*, which is from passage 2. Fib – is a negative fibroblast control (from individual C8861) prior to transfection, day 12 + is a positive control 12 days after transfection, 27077 + to 27082 + are the plasmids used for reprogramming.

Figure 2: a. ICC staining of differentiated embryoid bodies with antibodies for the three germ layers as indicated. Scale bar: 200 μm. b. Histological staining of teratomas derived from iPSC line C4955, showing generation of tissues from all three germ layers. Scale bar: 100 μm.

Figure 3: a. PluriTest pluripotency scores in the 7 chimpanzee lines and 4 human reference iPSC lines. Purple circles denote chimpanzees; yellow squares, humans. b. PluriTest results after removal of probes not mapping to the chimpanzee genome. All samples in the top left quadrant are human and have satisfactory pluripotency and novelty scores. Samples in the top right quadrant correspond to our chimpanzee iPSC panel, and have consistently high pluripotency yet high novelty scores.

Figure 4: Principal component analysis plots of data from the iPSCs. a. PCA generated from expression data of 12,171 orthologous genes. b. PCA generated from DNA methylation data measured by 335,307 filtered probes.

Figure 5: Overlap of H3K27me3 and H3K27ac signal between chimpanzee and human iPSCs at orthologous TSSs. a. H3K27me3 enrichment near all genes with an orthologous TSS. b. H3K27me3 enrichment near 2,910 genes previously identified as bivalent in human PSCs. c. H3K27ac enrichment near all genes with an orthologous TSS. d. H3K27ac peaks near 14 known pluripotency master regulators with orthologous TSSs.

Figure 6: *REX1* may be dispensable for pluripotency in chimpanzee iPSCs. In both panels REX1-expressing chimpanzee iPSC line is coloured red, significant interspecies differences are indicated along the left-hand side, and purple boxes indicate chimpanzee lines, yellow boxes indicate human lines. a. Expression values of 16 core pluripotency transcription factors in all human and chimpanzee iPSC lines. b. Methylation status of 13 CpG sites associated with *REX1* in all human and chimpanzee iPSCs. Location of the probe relative to the gene sequence is indicated along the right hand side. c. Fraction of DE genes in multiple categories downstream of *REX1* in human and mouse ESCs. 1: Genes associated with any Gene Ontology term that contains the words "ectoderm", "mesoderm" or "endoderm". 2. CNS development genes are associated with GO:0007417 or any of its offspring. 3: cardiovascular system development genes are associated with GO:0072358 or any of its offspring. 4: hepatobiliary system development genes are associated with GO:0055123 or any of its offspring. d. Expression levels of 34 genes associated with GO:0006096, glycolysis, in all human and chimpanzee iPSC lines. All reported p-values were calculated after excluding C6.

Figure 7: Relationships of iPSCs to their precursors. a. PCA of gene expression data from all iPSCs and their precursor cell lines. b. Neighbour-joining tree of Euclidean distances between all samples generated based on the gene expression data. c. PCA of DNA methylation data from all iPSCs and their precursor cell lines. d. Neighbour-joining tree of Euclidean distances between all samples generated based on the DNA methylation data.

Figure supplement legends:

Figure 1-figure supplement 1: Karyotypes for the 6 chimpanzee iPSC lines not shown in main text figures, generated after >15 passages in culture. Passage number for each line represents passages on MEF feeders plus additional passages on Matrigel.

Figure 1-figure supplement 2: ICC staining of the 6 chimpanzee iPSC lines not shown in main text figures with antibodies for pluripotency markers as indicated. Scale bar: 200 μm.

Figure 1-figure supplement 3: ICC staining showing SSEA1 expression in chimpanzee iPSC culture plates, clearly distinct from NANOG expression.

Figure 1-figure supplement 4: Melt curves showing a lack of exogenous reprogramming gene expression in episomally reprogrammed chimpanzee iPSCs after > 10 passages.

Figure 1-figure supplement 5: Exogenous gene expression in retrovirally reprogrammed chimpanzee iPSCs after various passages. All values are relative to expression in a day-7-post-transfection chimpanzee fibroblast.

Figure 2-figure supplement 1: ICC staining of differentiated embryoid bodies derived from the 6 chimpanzee iPSC lines not shown in main text figures, with antibodies for the three germ layers as indicated. Scale bar: 200 μm.

Figure 2-figure supplement 2: ICC staining of directly differentiated hepatocytes from line C2, with antibodies as indicated. Scale bar: 200 μm.

Figure 2-figure supplement 3: Histological staining of teratomas derived from three additional chimpanzee iPSC lines, showing generation of tissues from all three germ layers. Scale bar: 500 μm.

Figure 2-figure supplement 4: Sequencing traces from teratomas generated from chimpanzee iPSC lines for the mitochondrial genes *12S* (C3649, C4955) and *cytb* (C8861, C40210). All traces show clear evidence of the presence of chimpanzee tissue in the teratoma.

Figure 3-figure supplement 1: The effects of probe sub-setting in PluriTest pluripotency score calculations. Lighter shades indicate pluripotency scores before the removal of probes not mapping to the chimpanzee genome, darker shades indicate pluripotency after probe removal. Purple circles denote chimpanzees; yellow squares, humans.

Figure 4-figure supplement 1: Volcano plot showing the distribution of differentially expressed genes between iPSCs of chimpanzee and human origin.

Figure 4-figure supplement 2: Density plots of $\log_2$ FC change values amongst DE genes for the main comparisons presented in the text. The area bounded by the grey lines represents $\log_2$ FC changes with an absolute magnitude < 2.

Figure 5-figure supplement 1: Density plots of H3K27ac enrichment scores at orthologous TSSs in the entire data set and at 3,572 genes enriched only in chimpanzee iPSCs.

Figure 5-figure supplement 2: Density plots of mean RPKM in chimpanzee iPSCs in all 12,171 genes with expression data and in the subset of 1,737 genes with expression data and H3K27ac signal enrichment solely in chimpanzee iPSCs.

Figure 5-figure supplement 3: H3K27ac peaks observed in at least 1 chimpanzee or human iPSC, as identified by MACS at 22 known pluripotency master regulators. In the case of all three genes that differ between this figure and figure 5d – *KLF5*, *NR5AD* and *SMAD1* – processed enrichment signal after accounting for orthology is weak, and falls very close to our normalised enrichment score threshold of 1, explaining the difference between the two.

Figure 6-figure supplement 1: Expression values of 15 core pluripotency transcription factors in all human and chimpanzee iPSC lines. The data used to generate this figure are identical to those used to generate figure 5a except that expression levels of REX1 are not included in the calculation. REX1-expressing chimpanzee iPSC line C6 is coloured red, significant interspecies differences are indicated along the left-hand side, and purple boxes indicate chimpanzee lines, yellow boxes indicate human lines.

Figure 6-figure supplement 2: Expression levels of *REX1* in human, chimpanzee and bonobo iPSC lines generated in this study and in (Marchetto et al. 2013b). Data for this figure were jointly normalised.

Figure 6-figure supplement 3: Plot of PluriTest pluripotency scores versus normalised *REX1* intensity in 73 human iPSC lines derived in-house.

Figure 6-figure supplement 4: Methylation status of 13 CpG sites associated with *REX1* in chimpanzee and human iPSCs from this study and human PSCs from (Ziller et al. 2011). REX1-expressing chimpanzee iPSC line is coloured red; location of the probes relative to the gene sequence is indicated along the right hand side.

Figure 7-figure supplement 1: Boxplots of methylation beta values at 335,307 probes across all samples. Plots are colored by tissue type: light blue: chimpanzee iPSCs; dark blue: human iPSCs; light orange: chimpanzee fibroblasts; dark orange: human fibroblasts; turquoise: human LCLs.

Figure 7-figure supplement 2: Boxplots of methylation beta values across all samples, grouped by potency and genomic features. Boxes are colored by tissue type: light blue: chimpanzee iPSCs; light orange: chimpanzee fibroblasts.

Figure 7-figure supplement 3: Venn diagrams showing overlap in interspecies differences before and after reprogramming. a. Overlap in DE genes between chimpanzee and human fibroblasts, and

chimpanzee and human fibroblast-derived iPSCs. b. Overlap in DM probes between chimpanzee and human fibroblasts, and chimpanzee and human fibroblast-derived iPSCs.

Figure 7-figure supplement 4: Venn diagram showing overlap of genes identified as DE between iPSCs of the two species when we normalize the iPSC data independently and alongside data from the precursors.

Figure 7-figure supplement 5: Venn diagram showing overlap of probes identified as DM between iPSCs of the two species under the full and reduced *limma* models.

Figure 7-figure supplement 6: Normalized *XIST* expression values in 7 chimpanzee and human iPSCs. Circles denote chimpanzee iPSCs, squares indicate human iPSCS.

Figure 7-figure supplement 7: Quantile-normalized methylation beta values at 8,210 X-chromosome probes in 7 chimpanzee iPSCs and 7 human iPSCs. The colour bar beneath the dendrogram indicates sex of the individuals: purple: female; yellow: male. Sample names ending with _FB indicate fibroblast lines used to generate the corresponding iPSC line, samples ending with _LCL indicate LCL lines used to generate the corresponding iPSC line.

Figure 7-figure supplement 8: Normalized methylation beta values at 168 assayable probes known to be subject to parental imprinting effects, from (Ma et al. 2014). Sample names ending with _FB indicate fibroblast lines used to generate the corresponding iPSC line, samples ending with _LCL indicate LCL lines used to generate the corresponding iPSC line.

Video legends:

Video 1: Calcium transient flux in and out (GFP labelled) and contractility of directly differentiated cardiomyocytes from chimpanzee iPSC line C7.

Supplementary file headings:

Supplementary file 1: Descriptive data for all chimpanzee cell lines used in this work.

Supplementary file 2: Origin and purpose of all primers used.

Supplementary file 3: Descriptive data for all human iPSC lines used.

Supplementary file 4: Normalized RPKM values and DE genes between chimpanzee and human iPSCs.

Supplementary file 5: Gene Ontology BP terms associated with genes DE between chimpanzee and human iPSCs.

Supplementary file 6: DMRs identified between chimpanzee and human iPSCs.

Supplementary file 7: Numbers of DM probes and DMRs between chimpanzee and human iPSCs identified under various mean β difference thresholds.

Supplementary file 8: Gene Ontology BP terms associated with genes within DMRs between chimpanzee and human iPSCS.

Supplementary file 9: H3K27ac and H3K27me3 enrichment scores in 3 chimpanzee and 3 human iPSCs around 26,115 orthologous TSSs.

Supplementary file 10: RNA- and ChIP-sequencing reads generated and mapped for all samples in this work.

Supplementary file 11: Correlations between principal components and selected covariates in the expression data.

Supplementary file 12: Correlation between principal components and selected covariates in the methylation data.

Supplementary file 13: Normalized RPKM values and DE genes identified under the full *limma* DE testing framework.

Supplementary file 14: DMRs identified between chimpanzee iPSCs and their precursor fibroblasts.

Supplementary file 15: DMRs identified between human iPSCs and their precursor cells.

Supplementary file 16: Effects of different normalization schemes on the number of genes classified as DE in the full data set.

References

Alexa, A., Rahnenführer, J., and Lengauer, T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics (Oxford, England)* **22:** 1600-1607.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* **25:** 25-29.

Ben-Nun, I.F., Montague, S.C., Houck, M.L., Tran, H.T., Garitaonandia, I., Leonardo, T.R., Wang, Y.-C., Charter, S.J., Laurent, L.C., Ryder, O.A. et al. 2011. Induced pluripotent stem cells from highly endangered species. *Nature Methods* **8:** 829-831.

Benjamini, Y. and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to Multiple Testing. *Journal of the Royal Statistical Society B* **57:** 289-300.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K. et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125:** 315-326.

Blekhman, R. 2012. A database of orthologous exons in primates for comparative analysis of RNA-seq data. *Nature Precedings*.

Blekhman, R., Marioni, J.C., Zumbo, P., Stephens, M., and Gilad, Y. 2010. Sex-specific and lineage-specific alternative splicing in primates. *Genome Research* **20:** 180-189.

Blekhman, R., Oshlack, A., Chabot, A.E., Smyth, G.K., and Gilad, Y. 2008. Gene regulation in primates evolves under tissue-specific selection pressures. *PLoS Genetics* **4:** e1000271.

Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M.J., Croft, G.F., Amoroso, M.W., Oakley, D.H. et al. 2011. Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144:** 439-452.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M. et al. 2011. The evolution of gene expression levels in mammalian organs. *Nature* **478:** 343-348.

Britten, R.J. and Davidson, E.H. 1971. Repetitive and non-repetitive DNA sequences and a specualation on the origins of evolutionary novelty. *The Quarterly Review of Biology* **46:** 111-138.

Brivanlou, A.H., Gage, F.H., Jaenisch, R., Jessell, T., Melton, D., and Rossant, J. 2003. Stem cells. Setting standards for human embryonic stem cells. *Science* **300:** 913-916.

Burridge, P.W., Matsa, E., Shukla, P., Lin, Z.C., Churko, J.M., Ebert, A.D., Lan, F., Diecke, S., Huber, B., Mordwinkin, N.M. et al. 2014. Chemically defined generation of human cardiomyocytes. *Nature Methods* **11**.

Cain, C.E., Blekhman, R., Marioni, J.C., and Gilad, Y. 2011. Gene expression differences among primates are associated with changes in a histone epigenetic modification. *Genetics* **187:** 1225-1234.

Carroll, S.B. 2005. Evolution at two levels: on genes and form. *PLoS Biology* **3:** e245.

Carroll, S.B. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134:** 25-36.

Chan, A.W.S., Cheng, P.-H., Neumann, A., and Yang, J.-J. 2010. Reprogramming Huntington monkey skin cells into pluripotent stem cells. *Cellular reprogramming* **12:** 509-517.

1146 Chen, G., Gulbranson, D.R., Hou, Z., Bolin, J.M., Ruotti, V., Probasco, M.D., Smuga-Otto, K., Howden,
1147 S.E., Diol, N.R., Propson, N.E. et al. 2011. Chemically defined conditions for human iPSC
1148 derivation and culture. *Nature Methods* **8:** 424-429.
1149 Chen, K.G., Mallon, B.S., Hamilton, R.S., Kozhich, O.A., Park, K., Hoeppner, D.J., Robey, P.G., and
1150 McKay, R.D. 2012. Non-colony type monolayer culture of human embryonic stem cells.
1151 *Stem Cell Res* **9:** 237-248.
1152 Cheng, X., Ying, L., Lu, L., Galvão, A.M., Mills, J.a., Lin, H.C., Kotton, D.N., Shen, S.S., Nostro, M.C.,
1153 Choi, J.K. et al. 2012. Self-renewing endodermal progenitor lines generated from human
1154 pluripotent stem cells. *Cell Stem Cell* **10:** 371-384.
1155 Cohen, D.E. and Melton, D. 2011. Turning straw into gold: directing cell fate for regenerative
1156 medicine. *Nature Reviews Genetics* **12:** 243-252.
1157 Cotney, J., Leng, J., Yin, J., Reilly, S.K., DeMare, L.E., Emera, D., Ayoub, A.E., Rakic, P., and Noonan,
1158 J.P. 2013. The evolution of lineage-specific regulatory activities in the human embryonic
1159 limb. *Cell* **154:** 185-196.
1160 Deleidi, M., Hargus, G., Hallett, P., Osborn, T., and Isacson, O. 2011. Development of
1161 histocompatible primate-induced pluripotent stem cells for neural transplantation. *Stem*
1162 *Cells* **29:** 1052-1063.
1163 Du, P., Kibbe, W.A., and Lin, S.M. 2008. lumi: a pipeline for processing Illumina microarray.
1164 *Bioinformatics* **24:** 1547-1548.
1165 Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
1166 throughput. *Nucleic Acids Res* **32:** 1792-1797.
1167 Enard, W., Gehre, S., Hammerschmidt, K., Hölter, S.M., Blass, T., Somel, M., Brückner, M.K.,
1168 Schreiweis, C., Winter, C., Sohr, R. et al. 2009. A humanized version of Foxp2 affects
1169 cortico-basal ganglia circuits in mice. *Cell* **137:** 961-971.
1170 Ezashi, T., Telugu, B.P.V.L., Alexenko, A.P., Sachdev, S., Sinha, S., and Roberts, R.M. 2009.
1171 Derivation of induced pluripotent stem cells from pig somatic cells. *Proceedings of the*
1172 *National Academy of Sciences of the United States of America* **106:** 10993-10998.
1173 Gallego Romero, I., Ruvinsky, I., and Gilad, Y. 2012. Comparative studies of gene expression and
1174 the evolution of gene regulation. *Nature Reviews Genetics* **13:** 505-516.
1175 Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L.,
1176 Ge, Y., Gentry, J. et al. 2004. Bioconductor: open software development for computational
1177 biology and bioinformatics. *Genome Biol* **5:** R80.
1178 Gilad, Y., Rifkin, S.a., Bertone, P., Gerstein, M., and White, K.P. 2005. Multi-species microarrays
1179 reveal the effect of sequence divergence on gene expression profiles. *Genome Research*
1180 **15:** 674-680.
1181 Hernando-Herraez, I., Prado-Martinez, J., Garg, P., Fernandez-Callejo, M., Heyn, H., Hvilsom, C.,
1182 Navarro, A., Esteller, M., Sharp, A.J., and Marques-Bonet, T. 2013. Dynamics of DNA
1183 methylation in recent human and great ape evolution. *PLoS Genetics* **9:** e1003763.
1184 Hong, S.G., Winkler, T., Wu, C., Guo, V., Pittaluga, S., Nicolae, A., Donahue, R.E., Metzger, M.E.,
1185 Price, S.D., Uchida, N. et al. 2014. Path to the clinic: assessment of iPSC-based cell
1186 therapies in vivo in a nonhuman primate model. *Cell reports* **7:** 1298-1309.
1187 Howden, S.E., Wardan, H., Voullaire, L., McLenachan, S., Williamson, R., Ioannou, P., and Vadolas,
1188 J. 2006. Chromatin-binding regions of EBNA1 protein facilitate the enhanced
1189 transfection of Epstein-Barr virus-based vectors. *Hum Gene Ther* **17:** 833-844.

Israel, M.A., Yuan, S.H., Bardy, C., Reyna, S.M., Mu, Y., Herrera, C., Hefferan, M.P., Van Gorp, S., Nazor, K.L., Boscolo, F.S. et al. 2012. Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. *Nature* **482:** 216-220.

Jacob, F. 1977. Evolution and tinkering. *Science* **196:** 1161-1166.

Jaenisch, R. and Young, R. 2008. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132:** 567-582.

King, M.-C. and Wilson, A. 1975. Evolution at Two Levels in Humans and Chimpanzees. *Science* **188:** 107-116.

Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Paabo, S., Villablanca, F.X., and Wilson, A.C. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc Natl Acad Sci U S A* **86:** 6196-6200.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. 2014. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology* **15:** R29.

Li, H. and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25:** 1754-1760.

Li, Q., Lian, S., Dai, Z., Xiang, Q., and Dai, X. 2013. BGDB: a database of bivalent genes. *Database : the journal of biological databases and curation* **2013:** bat057.

Lian, X., Zhang, J., Azarin, S.M., Zhu, K., Hazeltine, L.B., Bao, X., Hsiao, C., Kamp, T.J., and Palecek, S.P. 2013. Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/ β -catenin signaling under fully defined conditions. *Nature protocols* **8:** 162-175.

Liao, Y., Smyth, G.K., and Shi, W. 2013. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30:** 923-930.

Liu, H., Zhu, F., Yong, J., Zhang, P., Hou, P., Li, H., Jiang, W., Cai, J., Liu, M., Cui, K. et al. 2008. Generation of induced pluripotent stem cells from adult rhesus monkey fibroblasts. *Cell Stem Cell* **3:** 587-590.

Liu, J., Koscielska, K.a., Cao, Z., Hulsizer, S., Grace, N., Mitchell, G., Nacey, C., Githinji, J., McGee, J., Garcia-Arocena, D. et al. 2012. Signaling defects in iPSC-derived fragile X premutation neurons. *Human Molecular Genetics* **21:** 3795-3805.

Lund, R.J., Närvä, E., and Lahesmaa, R. 2012. Genetic and epigenetic stability of human pluripotent stem cells. *Nature reviews. Genetics* **13:** 732-744.

Ma, H., Morey, R., O'Neil, R.C., He, Y., Daughtry, B., Schultz, M.D., Hariharan, M., Nery, J.R., Castanon, R., Sabatini, K. et al. 2014. Abnormalities in human pluripotent cells due to reprogramming mechanisms. *Nature* **511:** 177-183.

Marchetto, M.C.N., Muotri, A.R., and Gage, F.H. 2013a. Proposing a Model for Studying Primate Development Using Induced Pluripotent Stem Cells. In *Programmed Cells from Basic Neuroscience to Therapy* eds F.H. Gage and Y. Christen), pp. 31-39. Springer Berlin Heidelberg, Berlin, Heidelberg.

Marchetto, M.C.N., Narvaiza, I., Denli, A.M., Benner, C., Lazzarini, T.a., Nathanson, J.L., Paquola, A.C.M., Desai, K.N., Herai, R.H., Weitzman, M.D. et al. 2013b. Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature* **503:** 525-529.

Masui, S., Ohtsuka, S., Yagi, R., Takahashi, K., Ko, M.S.H., and Niwa, H. 2008. Rex1/Zfp42 is dispensable for pluripotency in mouse ES cells. *BMC developmental biology* **8:** 45.

McLean, C.Y., Reno, P.L., Pollen, A.a., Bassan, A.I., Capellini, T.D., Guenther, C., Indjeian, V.B., Lim, X., Menke, D.B., Schaar, B.T. et al. 2011. Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471:** 216-219.

Meisner, L.F. and Johnson, J.A. 2008. Protocols for cytogenetic studies of human embryonic stem cells. *Methods* **45:** 133-141.

Merkle, F.T. and Eggan, K. 2013. Modeling human disease with pluripotent stem cells: from genome association to function. *Cell Stem Cell* **12:** 656-668.

Müller, F.-J., Schuldt, B.M., Williams, R., Mason, D., Altun, G., Papapetrou, E.P., Danner, S., Goldmann, J.E., Herbst, A., Schmidt, N.O. et al. 2011. A bioinformatic assay for pluripotency in human cells. *Nature Methods* **8:** 315-317.

Nagy, K., Sung, H.-K., Zhang, P., Laflamme, S., Vincent, P., Agha-Mohammadi, S., Woltjen, K., Monetti, C., Michael, I.P., Smith, L.C. et al. 2011. Induced pluripotent stem cell lines derived from equine fibroblasts. *Stem Cell Reviews* **7:** 693-702.

Nazor, Kristopher L., Altun, G., Lynch, C.L., Tran, H., Harness, Julie V., Slavin, I., Garitaonandia, I., Müller, F.-J., Wang, Y.-C., Boscolo, Francesca S. et al. 2012. Recurrent Variations in DNA Methylation in Human Pluripotent Stem Cells and Their Differentiated Derivatives. *Cell Stem Cell* **10:** 620-634.

Ng, H.-h. and Surani, M.A. 2011. The transcriptional and signalling networks of pluripotency. *Nature cell biology* **13:** 490-496.

Okamoto, S. and Takahashi, M. 2011. Induction of retinal pigment epithelial cells from monkey iPS cells. *Investigative ophthalmology & visual science* **52:** 8785-8790.

Okita, K., Matsumura, Y., Sato, Y., Okada, A., Morizane, A., Okamoto, S., Hong, H., Nakagawa, M., Tanabe, K., Tezuka, K.-i. et al. 2011. A more efficient method to generate integration-free human iPS cells. *Nature Methods* **8:** 409-412.

Orkin, S.H. and Hochedlinger, K. 2011. Chromatin connections to pluripotency and cellular reprogramming. *Cell* **145:** 835-850.

Paige, S.L., Thomas, S., Stoick-Cooper, C.L., Wang, H., Maves, L., Sandstrom, R., Pabon, L., Reinecke, H., Pratt, G., Keller, G. et al. 2012. A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. *Cell* **151:** 221-232.

Paradis, E., Claude, J., and Strimmer, K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20:** 289-290.

Pollard, K.S., Salama, S.R., Lambert, N., Lambot, M.-A., Coppens, S., Pedersen, J.S., Katzman, S., King, B., Onodera, C., Siepel, A. et al. 2006. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443:** 167-172.

Prabhakar, S., Visel, A., Akiyama, J.a., Shoukry, M., Lewis, K.D., Holt, A., Plajzer-Frick, I., Morrison, H., Fitzpatrick, D.R., Afzal, V. et al. 2008. Human-specific gain of function in a developmental enhancer. *Science* **321:** 1346-1350.

R Core Team. 2013. R: A language and environment for statistcal computing. R Foundation for Scientific Computing, VIenna, Austria.

Rada-Iglesias, A., Bajpai, R., Prescott, S., Brugmann, S.a., Swigut, T., and Wysocka, J. 2012. Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell* **11:** 633-648.

Rana, P., Anson, B., Engle, S., and Will, Y. 2012. Characterization of human-induced pluripotent stem cell-derived cardiomyocytes: bioenergetics and utilization in safety screening. *Toxicological Sciences* **130:** 117-131.

Risso, D., Schwartz, K., Sherlock, G., and Dudoit, S. 2011. GC-content normalization for RNA-Seq data. *BMC bioinformatics* **12:** 480.

Roadmap Epigenomics Consortium Kundaje, A. Meuleman, W. Ernst, J. Bilenky, M. Yen, A. Heravi-Moussavi, A. Kheradpour, P. Zhang, Z. Wang, J. et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518:** 317-330.

Rockman, M.V., Hahn, M.W., Soranzo, N., Zimprich, F., Goldstein, D.B., and Wray, G.a. 2005. Ancient and recent positive selection transformed opioid cis-regulation in humans. *PLoS Biology* **3:** e387.

Schmidt, D., Wilson, M.D., Spyrou, C., Brown, G.D., Hadfield, J., and Odom, D.T. 2009. ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods* **48:** 240-248.

Scotland, K.B., Chen, S., Sylvester, R., and Gudas, L.J. 2009. Analysis of Rex1 (zfp42) function in embryonic stem cell differentiation. *Developmental dynamics* **238:** 1863-1877.

Smyth, G.K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3:** Article3.

Sommer, C.A., Christodoulou, C., Gianotti-Sommer, A., Shen, S.S., Sailaja, B.S., Hezroni, H., Spira, A., Meshorer, E., Kotton, D.N., and Mostoslavsky, G. 2012. Residual expression of reprogramming factors affects the transcriptional program and epigenetic signatures of induced pluripotent stem cells. *PLoS ONE* **7:** e51711.

Son, M.-Y., Choi, H., Han, Y.-M., and Cho, Y.S. 2013. Unveiling the critical role of REX1 in the regulation of human stem cell pluripotency. *Stem cells (Dayton, Ohio)* **31:** 2374-2387.

Song, Q. and Smith, A.D. 2011. Identifying dispersed epigenomic domains from ChIP-Seq data. *Bioinformatics* **27:** 870-871.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131:** 861-872.

Takahashi, K. and Yamanaka, S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126:** 663-676.

Teschendorff, A.E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., and Beck, S. 2013. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29:** 189-196.

Tohyama, S., Hattori, F., Sano, M., Hishiki, T., Nagahata, Y., Matsuura, T., Hashimoto, H., Suzuki, T., Yamashita, H., Satoh, Y. et al. 2013. Distinct metabolic flow enables large-scale purification of mouse and human pluripotent stem cell-derived cardiomyocytes. *Cell Stem Cell* **12:** 127-137.

Tomioka, I., Maeda, T., Shimada, H., Kawai, K., Okada, Y., Igarashi, H., Oiwa, R., Iwasaki, T., Aoki, M., Kimura, T. et al. 2010. Generating induced pluripotent stem cells from common marmoset (Callithrix jacchus) fetal liver cells using defined factors, including Lin28. *Genes Cells* **15:** 959-969.

Trapnell, C., Pachter, L., and Salzberg, S.L. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25:** 1105-1111.

Wamstad, J.a., Alexander, J.M., Truty, R.M., Shrikumar, A., Li, F., Eilertson, K.E., Ding, H., Wylie, J.N., Pico, A.R., Capra, J.a. et al. 2012. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* **151:** 206-220.

Wang, G., McCain, M.L., Yang, L., He, A., Pasqualini, F.S., Agarwal, A., Yuan, H., Jiang, D., Zhang, D., Zangi, L. et al. 2014. Modeling the mitochondrial cardiomyopathy of Barth syndrome with induced pluripotent stem cell and heart-on-chip technologies. *Nat Med* **20:** 616-623.

Wu, Y., Mishra, A., Qiu, Z., Farnsworth, S., Tardif, S.D., and Hornsby, P.J. 2012. Nonhuman primate induced pluripotent stem cells in regenerative medicine. *Stem cells international* **2012:** 767195.

Wu, Y., Zhang, Y., Mishra, A., Tardif, S.D., and Hornsby, P.J. 2010. Generation of induced pluripotent stem cells from newborn marmoset skin fibroblasts. *Stem Cell Res* **4:** 180-188.

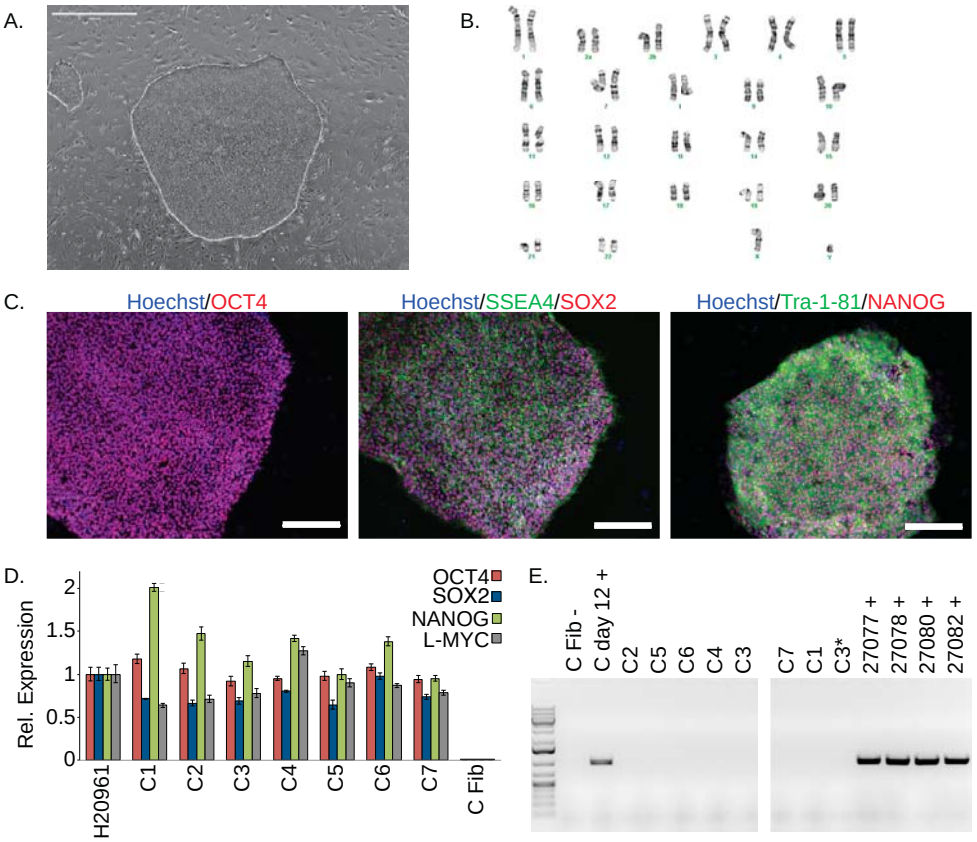
Wunderlich, S., Kircher, M., Vieth, B., Haase, A., Merkert, S., Beier, J., Göhring, G., Glage, S., Schambach, A., Curnow, E.C. et al. 2014. Primate iPS cells as tools for evolutionary analyses. *Stem cell research* **12:** 622-629.

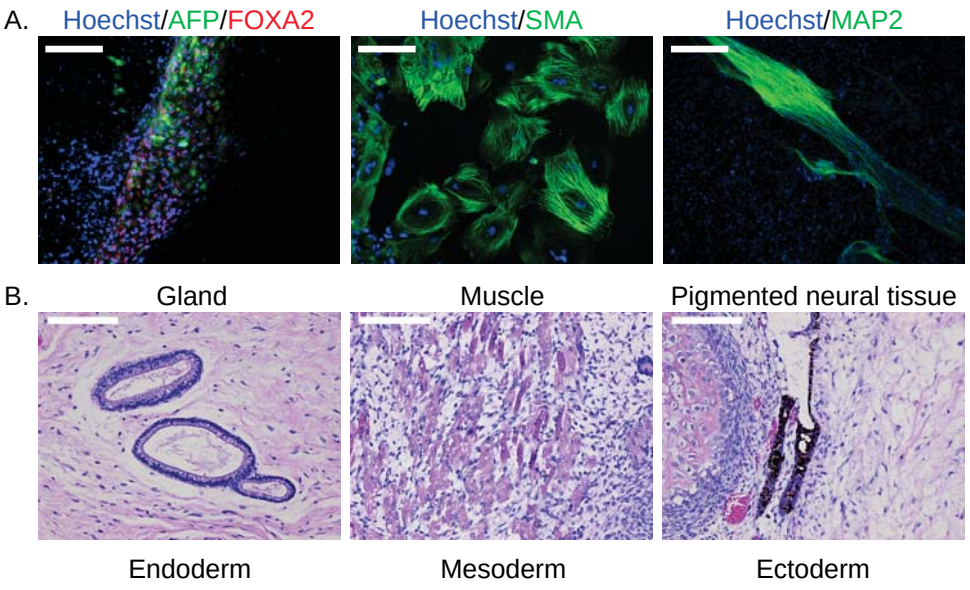
Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D. et al. 2013. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153:** 1134-1148.

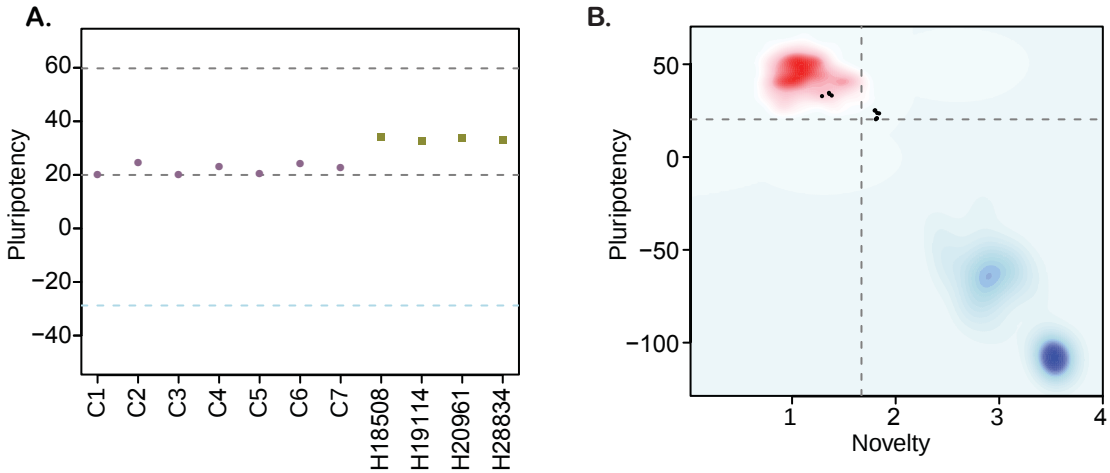
Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9:** R137.

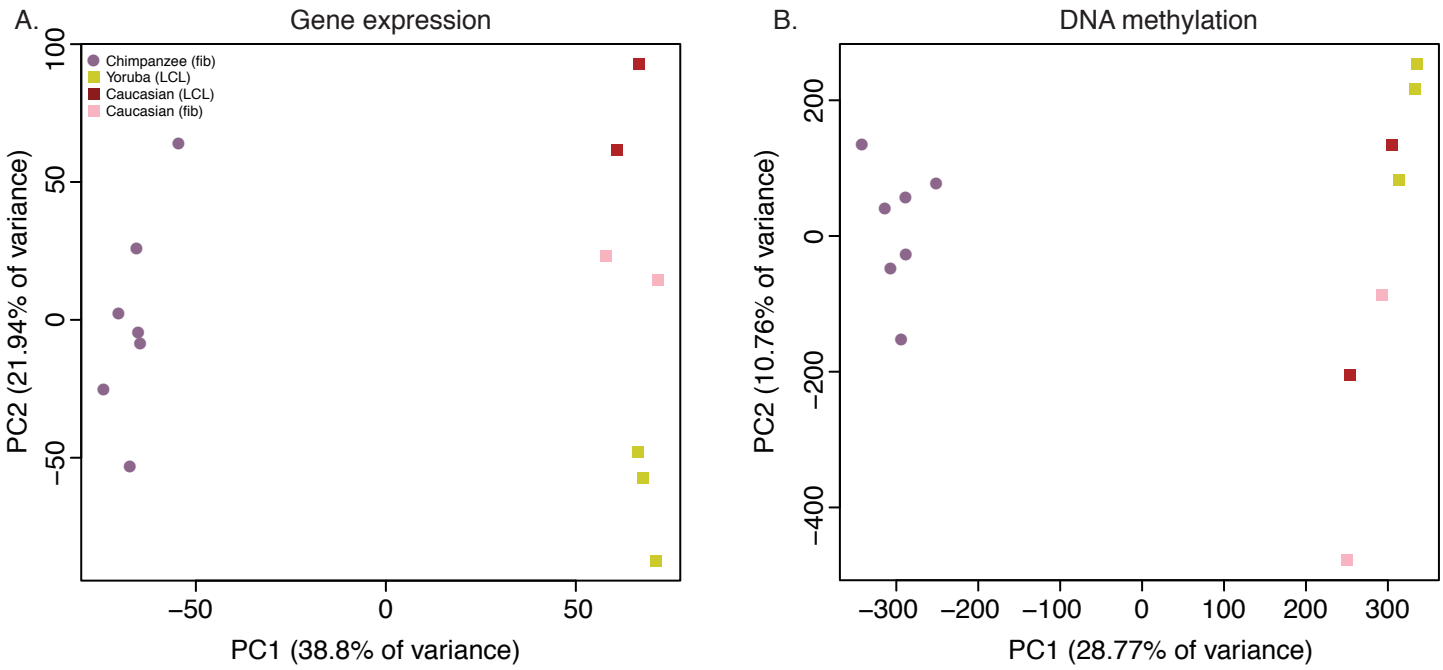
Zhou, X., Cain, C.E., Myrthil, M., Lewellen, N., Michelini, K., Davenport, E.R., Stephens, M., Pritchard, J.K., and Gilad, Y. 2014. Epigenetic modifications are associated with inter-species gene expression variation in primates. *Genome Biology* **15:** 547.

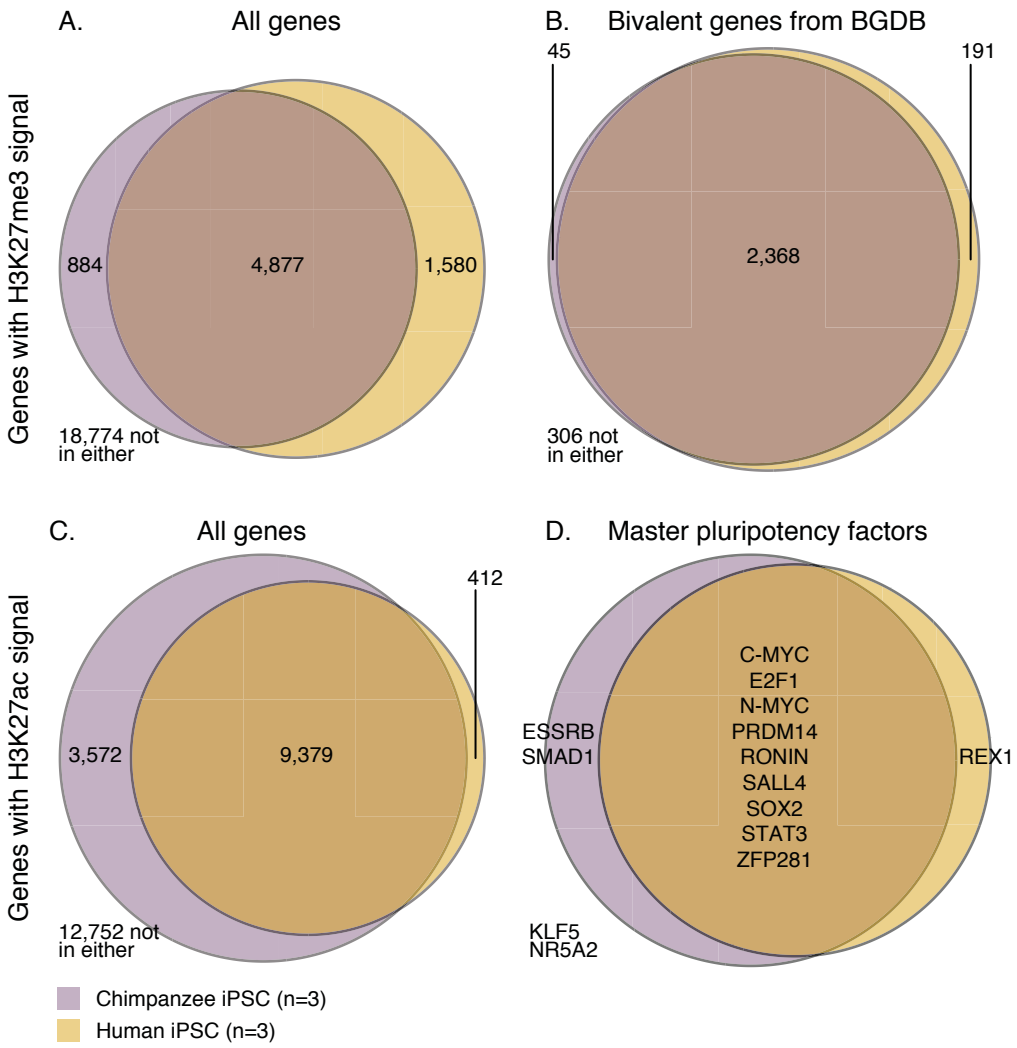
Ziller, M.J., Müller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C.B., Bernstein, B.E., Lengauer, T. et al. 2011. Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genetics* **7:** e1002389.

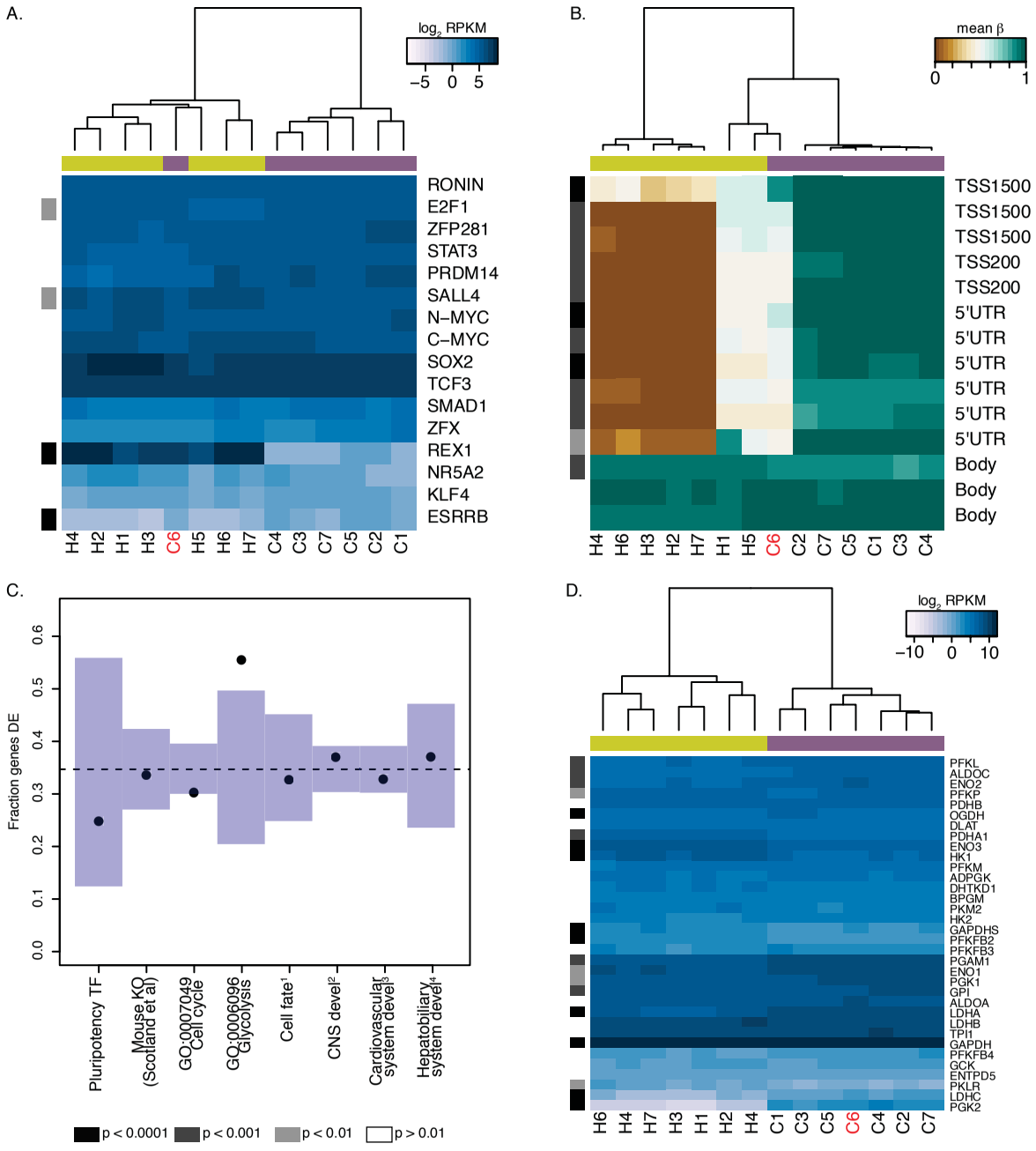

A.
B.
C.
Hoechst/OCT4
Hoechst/SSEA4/SOX2
Hoechst/Tra-1-81/NANOG
D.
Rel. Expression
2
1.5
1
0.5
0
OCT4
SOX2
NANOG
L-MYC
H20961
C1
C2
C3
C4
C5
C6
C7
C Fib
E.
C Fib -
C day 12 +
C2
C5
C6
C4
C3
C7
C1
C3*
27077 +
27078 +
27080 +
27082 +

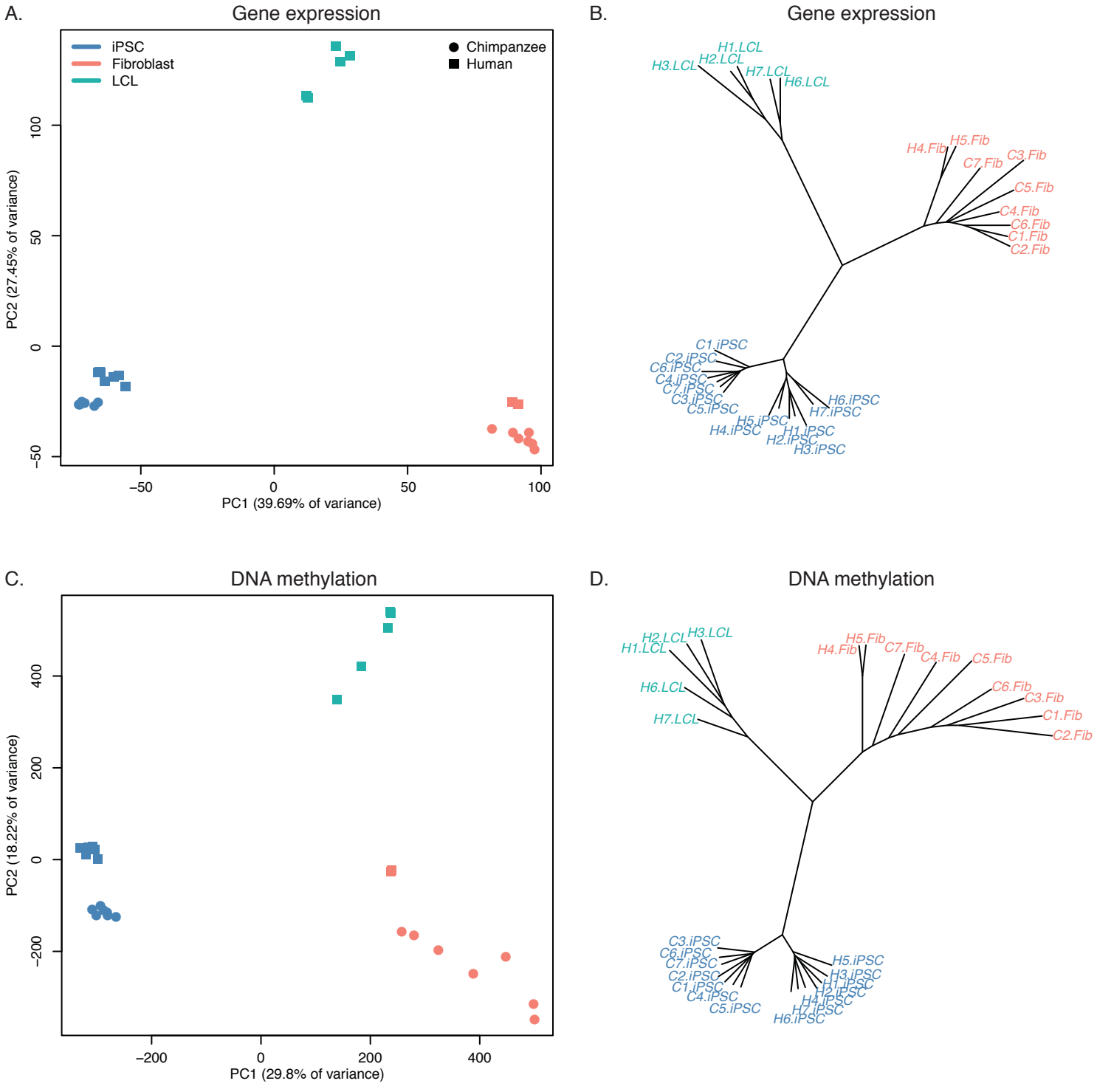

A.
Hoechst/AFP/FOXA2
Hoechst/SMA
Hoechst/MAP2
B.
Gland
Muscle
Pigmented neural tissue
Endoderm
Mesoderm
Ectoderm


A.
Pluripotency
60
40
20
0
−20
−40
C1
C2
C3
C4
C5
C6
C7
H18508
H19114
H20961
H28834
B.
Pluripotency
50
0
−50
−100
1
2
3
4
Novelty


A.
Gene expression
Chimpanzee (fib)
Yoruba (LCL)
Caucasian (LCL)
Caucasian (fib)
PC2 (21.94% of variance)
PC1 (38.8% of variance)
B.
DNA methylation
PC2 (10.76% of variance)
PC1 (28.77% of variance)


A.
All genes
Genes with H3K27me3 signal
884
4,877
1,580
18,774 not
in either
B.
Bivalent genes from BGDB
45
191
2,368
306 not
in either
C.
All genes
Genes with H3K27ac signal
412
3,572
9,379
12,752 not
in either
D.
Master pluripotency factors
ESSRB
SMAD1
C-MYC
E2F1
N-MYC
PRDM14
RONIN
SALL4
SOX2
STAT3
ZFP281
REX1
KLF5
NR5A2
Chimpanzee iPSC (n=3)
Human iPSC (n=3)


A.
log2 RPKM
−5 0 5
RONIN
E2F1
ZFP281
STAT3
PRDM14
SALL4
N−MYC
C−MYC
SOX2
TCF3
SMAD1
ZFX
REX1
NR5A2
KLF4
ESRRB
H4 H2 H1 H3 C6 H5 H6 H7 C4 C3 C7 C5 C2 C1
B.
mean β
0 1
TSS1500
TSS1500
TSS1500
TSS200
TSS200
5'UTR
5'UTR
5'UTR
5'UTR
5'UTR
5'UTR
Body
Body
Body
H4 H6 H3 H2 H7 H1 H5 C6 C2 C7 C5 C1 C3 C4
C.
Fraction genes DE
0.0 0.1 0.2 0.3 0.4 0.5 0.6
Pluripotency TF
Mouse KO (Scotland et al)
GO:0007049 Cell cycle
GO:0006096 Glycolysis
Cell fate1
CNS devel2
Cardiovascular system devel3
Hepatobiliary system devel4
p < 0.0001
p < 0.001
p < 0.01
p > 0.01
D.
log2 RPKM
−10 0 10
PFKL
ALDOC
ENO2
PFKP
PDHB
OGDH
DLAT
PDHA1
ENO3
HK1
PFKM
ADPGK
DHTKD1
BPGM
PKM2
HK2
GAPDHS
PFKFB2
PFKFB3
PGAM1
ENO1
PGK1
GPI
ALDOA
LDHA
LDHB
TPI1
GAPDH
PFKFB4
GCK
ENTPD5
PKLR
LDHC
PGK2
H6 H4 H7 H3 H1 H2 H4 C1 C3 C5 C6 C4 C2 C7


A.
Gene expression
iPSC
Fibroblast
LCL
Chimpanzee
Human
PC2 (27.45% of variance)
PC1 (39.69% of variance)
B.
Gene expression
H1.LCL
H2.LCL
H3.LCL
H7.LCL
H6.LCL
H4.Fib
H5.Fib
C7.Fib
C3.Fib
C5.Fib
C4.Fib
C6.Fib
C1.Fib
C2.Fib
C1.iPSC
C2.iPSC
C6.iPSC
C4.iPSC
C7.iPSC
C3.iPSC
C5.iPSC
H4.iPSC
H5.iPSC
H2.iPSC
H1.iPSC
H3.iPSC
H6.iPSC
H7.iPSC
C.
DNA methylation
PC2 (18.22% of variance)
PC1 (29.8% of variance)
D.
DNA methylation
H2.LCL
H3.LCL
H1.LCL
H6.LCL
H7.LCL
H5.Fib
H4.Fib
C7.Fib
C4.Fib
C5.Fib
C6.Fib
C3.Fib
C1.Fib
C2.Fib
C3.iPSC
C6.iPSC
C7.iPSC
C2.iPSC
C1.iPSC
C4.iPSC
C5.iPSC
H5.iPSC
H3.iPSC
H1.iPSC
H2.iPSC
H4.iPSC
H7.iPSC
H6.iPSC