Stem Cell Research (2014) 12, 622–629

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/scr

SHORT REPORT

Primate iPS cells as tools for evolutionary analyses

Stephanie Wunderlich[a,g,1], Martin Kircher[b,1,2], Beate Vieth[b,c], Alexandra Haase[a,g], Sylvia Merkert[a,g], Jennifer Beier[a,g], Gudrun Göhring[d,g], Silke Glage[e,g], Axel Schambach[f,g], Eliza C. Curnow[h], Svante Pääbo[b], Ulrich Martin[a,g,*,1], Wolfgang Enard[b,c,**,1]

[a] *Leibniz Research Laboratories for Biotechnology and Artificial Organs (LEBAO), Department of Cardiothoracic, Transplantation and Vascular Surgery, Hannover Medical School, 30625 Hannover, Germany*
[b] *Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany*
[c] *Department Biology II, Ludwig Maximilian University Munich, D-82152 Martinsried, Germany*
[d] *Department Cell and Molecular Pathology, Hannover Medical School, 30625 Hannover, Germany*
[e] *Institute for Laboratory Animal Science, Hannover Medical School, 30625 Hannover, Germany*
[f] *Department of Experimental Hematology, Hannover Medical School, 30625 Hannover, Germany*
[g] *REBIRTH-Cluster of Excellence, Germany*
[h] *Washington National Primate Research Center, Seattle, USA*

Received 18 December 2013; received in revised form 31 January 2014; accepted 1 February 2014
Available online 8 February 2014

Abstract Induced pluripotent stem cells (iPSCs) are regarded as a central tool to understand human biology in health and disease. Similarly, iPSCs from non-human primates should be a central tool to understand human evolution, in particular for assessing the conservation of regulatory networks in iPSC models.

Here, we have generated human, gorilla, bonobo and cynomolgus monkey iPSCs and assess their usefulness in such a framework. We show that these cells are well comparable in their differentiation potential and are generally similar to human, cynomolgus and rhesus monkey embryonic stem cells (ESCs). RNA sequencing reveals that expression differences among clones, individuals and stem cell type are all of very similar magnitude within a species. In contrast, expression differences between closely related primate species are three times larger and most genes show significant expression differences among the analyzed species. However, pseudogenes differ more than twice as much, suggesting that evolution of expression levels in primate stem cells is rapid, but constrained. These patterns in pluripotent stem cells are comparable to those found in other tissues except testis. Hence, primate iPSCs reveal insights into general primate gene expression evolution and should provide a rich source to identify conserved and species-specific gene expression patterns for cellular phenotypes.

* Correspondence to: U. Martin, Leibniz Research Laboratories for Biotechnology and Artificial Organs (LEBAO), Department of Cardiothoracic, Transplantation and Vascular Surgery, Hannover Medical School, 30625 Hannover, Germany.
** Correspondence to: W. Enard, Department Biology II, Ludwig Maximilian University Munich, D-82152 Martinsried, Germany.
E-mail addresses: martin.ulrich@mh-hannover.de (U. Martin), enard@bio.lmu.de (W. Enard).
[1] Equal contribution.
[2] Present address: Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA.

http://dx.doi.org/10.1016/j.scr.2014.02.001

Introduction

Pluripotent stem cells (PSCs) allow experimental access to cell types and differentiation stages that are difficult or impossible to investigate in humans. Similarly, PSCs from non-human primates would allow to study the molecular and cellular basis of human-specific traits, as recently shown for the differential regulation of LINE-1 transposons in human and ape PSCs (Marchetto et al., 2013). In addition to studying differences, human and non-human PSCs could be very useful to analyze conserved features of molecular and cellular phenotypes (Enard, 2012). On the level of DNA, evolutionary conservation is an established and highly useful measure e.g. to infer the functional relevance of disease associated mutations (Cooper and Shendure, 2011). In a similar manner, evolutionary conservation should be helpful to infer the functional relevance at the level of molecular and cellular phenotypes and it has been used e.g. to analyze regulatory networks of stress response in fungi (Roy et al., 2013) or to compare transcriptional patterns of human and murine immune cells (Shay et al., 2013). Primate PSCs, in particular induced pluripotent stem cells (iPSCs) and their derivatives could be a practically feasible way to exploit this kind of information for humans' closest relatives. Although it has been shown that iPSCs can be generated from a range of primates (Marchetto et al., 2013; Wu et al., 2010; Wunderlich et al., 2012; Ben-Nun et al., 2011; Liu et al., 2008), a quantitative assessment on how molecular phenotypes like expression patterns differ among primate PSCs has not been described. As a first step in this direction, we have generated and characterized human, gorilla, bonobo and cynomolgus monkey iPSCs. We demonstrate that they have comparable differentiation potentials that are similar to human, cynomolgus and rhesus monkey embryonic stem cells (ESCs), and that their gene expression patterns evolve fast, but under considerable constraint. Our results show that primate iPSCs can provide a rich source to identify conserved and species-specific gene expression patterns for cellular phenotypes in humans' closest relatives.

Material and methods

Generation, characterisation, culture and differentiation of iPS cells

iPS cells were generated, cultivated and differentiated using standard procedures described in detail in the Supplemental experimental procedures.

Sequencing and primary data processing

Sequencing libraries were pooled in equimolar ratio and sequenced according to the manufacturer's instructions for single read multiplex experiments with 76 + 7 cycles on the Genome Analyzer IIx platform (v5 sequencing chemistry and v4 cluster generation kit). The standard protocol was followed except that an indexed control φX174 library (index TTGCCGC) was spiked into each lane, yielding approximately 1% control reads in each lane. Sequencing data was analyzed starting from BCL and CIF intensity files generated by the Illumina Genome Analyzer SCS 2.9/RTA 1.9 software. Ibis 1.1.6 was trained from the φX174 control reads of all lanes aligned to their reference sequence and then used to re-call bases and PHRED-like quality scores from intensity files (Kircher et al., 2009). Only reads matching one of the used indexes or one-edit variants (Meyer and Kircher, 2010) were further processed and subjected to an adapter and chimera filter; removing artificial sequences and trimming adapter sequence starting at the read end (Kircher et al., 2011). From this data, reads with more than 5 bases below a quality score of 15 as well as reads shorter than 60 bases after trimming were discarded, resulting in 104 million reads with an average of 4.3 million reads per library (Table S1). For the primate tissue data (Brawand et al., 2011), Ibis called 76-bp raw reads were processed as described above. In addition, reads with low sequence complexity were removed by discarding reads with sequence entropy below 0.85, as these might result from the read out of image artifacts (Kircher et al., 2011) and could not be removed by a barcode filter used for the newly generated data. Data is available at GEO (http://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE50781.

Mapping and expression level estimation

Human Ensemblv66 gene models (Curwen et al., 2004) for GRCh37 were transferred to the coordinates of the chimpanzee genome (CHIMP2.1.4), the gorilla genome (gorGor3.1) and the rhesus macaque genome (MMUL_1) using Ensembl Compara v66 six primate EPO whole genome alignments (Paten et al., 2008, 2009). Since this version of Ensembl did not include a chimpanzee mitochondrial genome, the Ensembl Compara mitochondrial alignments were obtained from the v64 release. Further, the CHIMP2.1.4 genome was supplemented by the previously used mitochondrial genome. Exon start and end coordinates were independently transferred using the EPO alignments and only genes were used where (1) all exons and their respective start and end coordinates were assigned to a single chromosome/contig, (2) exons were on the same strand in each single species and (3) adjacent exons were not more than 1 Mb apart in the three species under consideration. Tophat 1.4.1 (Trapnell et al., 2009) was used to align sequences for each sample first to the inferred transcriptome of each species and the remaining reads to the reference sequence. We also aligned all samples to the mouse annotation (Ensembl v66) and mouse reference genome (musMus37). We removed all reads from the species alignments which aligned with the same or fewer mismatches to the mouse genome, resulting in an average of 3.2 million reads that could be mapped to the species genome per library (Table S1). Afterwards, gene models were quantified using Cufflinks 1.3.0 (Roberts et al., 2011) with activated fragment bias correction and correction for reads mapping to multiple genomic locations. For analysis, we only considered genes as detected with an FPKM value greater or equal to the size of the 95% FPKM-confidence interval (Nagaraj et al., 2011). This resulted in 18,214 genes detected in at least one library and 2877 genes detected in all 23 libraries (see Table S2 for expression estimates of all genes).

Analysis of expression patterns

For all analyses, we used log2-transformed FPKM values since we are interested in relative differences. We excluded all 15 mitochondrially encoded genes since their larger sequence divergence causes mapping and hence quantification problems. PCA (Fig. S6A), clustering analysis (Fig. 2A) and the distance for the technical replicates were done on the 2877 genes detected in all 23 RNA-Seq libraries. For all other purposes, we calculated FPKM values combining reads from technical replicates (Table S2). Hence, distances for PSCs (Figs. 2B and S6B and S6C) were calculated based on the 3414 genes detected in all 19 samples. For clustering and distance calculations including PSCs and tissues (Figs. 3, S7 and S8), we used the 2016 genes detected in all 63 tissue samples. Average linkage clustering (Figs. 2A and 3A) of Spearman rank correlations and distance calculations were done on mean centered samples and mean centered genes using R 2.13.1 (R Core Team, 2013). When averaging the pairwise distances (Fig. S6B) for different groups (Fig. 2B), we averaged the four pairwise distances of the replicates (N = 4), the average of the distances among clones for each species (N = 4), the average of the distances between individuals among the same PSC (4 pairwise comparisons for human individuals, 1 for macaque, N = 5), the average of the distances between ESCs and iPSCs (6 human, 1 macaque, N = 7), the average among all samples within a species (N = 5 species) and the average of the pairwise distance between each species pair (N = 6 when grouping the two macaque species together). The between and within distances for the tissues were calculated accordingly. However, since only one individual per species is available for testis, expression distance within a species needed to be imputed from the average expression distance in the other tissues. Since testis tends to have less variation in gene expression than other tissues (Khaitovich et al., 2005a), this is conservative.

For expression analysis on the single gene level, we performed a one-way ANOVA with the factor species (combining rhesus and cynomolgus samples as before) on the FPKM values using R (R Core Team, 2013). The same was done for 100 random permutations of species labels and the FDR was estimated by the method of Benjamini and Yekutieli (Benjamini and Yekutieli, 2001) as implemented as part of the FUNC package (Prufer et al., 2007). For the enrichment analysis, genes were ranked according to their Mean Sum of Square for the factor species, following the logic that differences within species are mainly non-genetic and hence need to be subtracted from the total variance. Significant GO categories (biological process only, term tables from 7/29/2012) with more than 20 detected genes were tested for high ranking and low ranking values using the Wilcoxon Rank test as implemented in FUNC (Prufer et al., 2007). To identify the most specific categories we used the Refinement algorithm of FUNC (p-value cut-off of 0.01) and sorted categories based on their rank enrichment (Sum of ranks per gene in the group divided by the sum of ranks of all genes detected and annotated in the biological process ontology) as shown in Table S3. The enrichment was overall significant as based on the global p-value given by FUNC from permutations across genes ($p < 0.001$ for high-ranking and low ranking categories). However the 100 random permutations show a similar profile of p-values resulting in a global p-value of 0.6 and 0.4 for high ranking and low ranking categories, respectively.

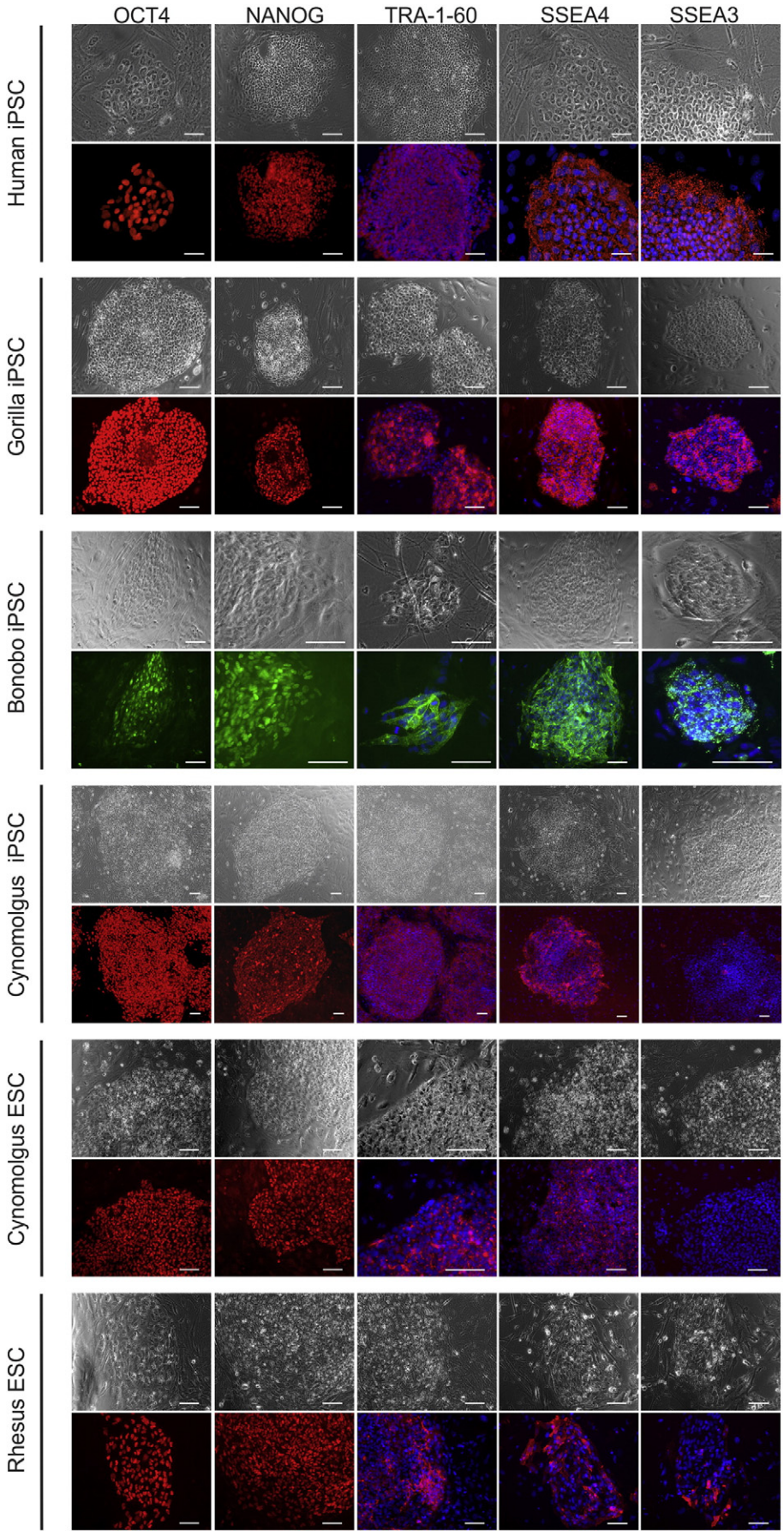
Results and discussion

Primate iPSCs proliferate and differentiate similar to human iPSCs and primate ESCs

We generated iPSCs from endothelial cells or fibroblasts of a gorilla (*Gorilla gorilla*), a bonobo (*Pan paniscus*) and a cynomolgus monkey (*Macaca fascicularis*). The culture characteristics and morphology of those iPSCs were very similar to human iPSCs and to human and macaque ESCs (Fig. 1) and generally resembled human pluripotent stem cells (PSCs) much more than murine PSCs. Immunocytological staining of iPSCs revealed expression of pluripotency markers including OCT4, NANOG, Tra-1-60, SSEA-3 and SSEA-4. As recently described (Suemori et al., 2001), SSEA-3 expression could not be observed on cynomolgus PSCs, and sporadically only on rhesus monkey cells (Fig. 1). Endogenous expression levels of *OCT4*, *SOX2*, *NANOG* and *LIN28* in iPSCs were similar to ESC lines (data not shown). This suggests that the cells have been fully reprogrammed into iPSCs with normal karyotypes (Figs. S1A–S5A).

To assess the differentiation capacity of the iPSCs, we characterized their embryoid body formation *in vitro*. Transcripts of markers for all three germ layers could be detected at levels comparable to ESCs (data not shown), congruent with the immunocytological results (Figs. S1B–S5B). Transplantation into immunodeficient SCID-beige mice resulted in the formation of typical teratomas containing derivatives of all three germ layers (Figs. S1C–S4C). In summary, the generated iPSCs of bonobo, gorilla and cynomolgus monkey have characteristics that are well comparable to human and macaque pluripotent stem cells with respect to their morphology, culture characteristics, their expression of pluripotent markers and their differentiation capacities.

Figure 1 Pluripotent stem cells from humans, great apes and macaques show comparable expression of different ES cell markers including endogenous pluripotency factors. Immunostaining for the ES cell markers OCT4, NANOG, Tra-1-60, SSEA-4 and SSEA-3 (red or green) are shown for human iPS cells (hCBiPSCs) clone 1 passage 7 (Haase et al., 2009), gorilla iPS cells (goiPSCs) clone 1 passage 12, bonobo iPS cells (boiPSCs) clone 2d passage 13, cynomolgus iPS cells (cyiPSCs) clone 2 passage 5 (Wunderlich et al., 2012), cynomolgus embryonic stem cells (cyESCs) MF12 passage 69 and rhesus embryonic stem cells (RESCs) 366.4 passage 53. Nuclei are stained with DAPI (blue). Note that the strong expression of transgenes and consequently red reporter fluorescence in the bonobo iPS cells make the green immunostaining necessary. Scale bars represent 100 μm. Images of human iPSCs are adapted from (Haase et al., 2009) and images from cynomolgus iPSCs are adapted from (Wunderlich et al., 2012).


OCT4
NANOG
TRA-1-60
SSEA4
SSEA3
Human iPSC
Gorilla iPSC
Bonobo iPSC
Cynomolgus iPSC
Cynomolgus ESC
Rhesus ESC

Gene expression in primate PSCs evolves rapidly, but under constraint

To characterize and quantify molecular differences between the cells, we measured gene expression by mRNA sequencing in five gorilla, two bonobo and three macaque iPS clones as well as in three iPS clones from two human individuals, three human embryonic stem (ES) cell lines and three macaque ES cell lines (Tables S1 and S2). A visual inspection of the data by principal component analysis (Fig. S6A) and hierarchical clustering (Fig. 2A) of the expression profiles shows that samples of the same species group together. To make a more quantitative assessment of the differences, we calculated the average squared Euclidian distance per gene between each pair of samples (Fig. S6B), which is a suitable measure of gene expression divergence (Khaitovich et al., 2004; Pereira et al., 2009). Congruent with the pattern observed in the principal component and clustering analyses, we find that distances within a species are very similar between clones of the same individual and between clones of different individuals as well as between iPSCs and ESCs (Fig. 2B). It is noteworthy, that the distances between clones derived from one individual are as big as the distances between different individuals, suggesting that the majority of gene expression differences within a species are not genetic. Further, it is relevant that distances among ESCs and iPSCs are as big as distances between ESCs and iPSCs. This supports the view that systematic expression differences between iPSCs and ESCs are not as important as the variation among individual PSC clones (Yamanaka, 2012; Bock et al., 2011) and suggests that this is a general phenomenon across primates (Fig. S6C).

Compared to expression distances within species, the distances between the closely related species are more than 3-fold larger (Fig. 2B). Hence, the genetic differences between species do have a strong impact on PSC gene expression despite the similar cellular characteristics described above. However, it is not clear to what extent these expression levels are constrained, i.e. how much bigger the distance would be, if gene expression levels could vary freely. Pseudogenes have been suggested as a measure for this neutral rate of gene expression evolution and were found to differ among primates to a similar extent as other genes (Khaitovich et al., 2005b). However, microarrays were used in these analyses for which the background hybridization due to the expression of the parental genes is difficult to estimate. This is not the case for RNA-Seq data and hence we compared expression distances of the 40 pseudogenes that were expressed in all our 19 PSC samples to the remaining 3374 non-pseudogenes (>99% protein-coding). While the distances within species were very similar in the two groups, pseudogenes had a 2.6-fold larger distance between species (Fig. 2C). The distance between species includes the distance within species and this within distance is largely non-genetic as shown above. Hence, to estimate the expression distance due to genetic differences between species (expression divergence) it is necessary to subtract the distance within species from the distance between species. In the six comparisons among the analyzed species, non-pseudogenes have on average 27% (range 21% to 40%) of the divergence of pseudogenes, suggesting that more than half of the amount of possible, i.e. neutral, expression divergence is subject to negative selection in primate PSCs. In summary, we find that gene expression patterns between PSCs from primates evolve rapidly, but under considerable amounts of constraint.

Rate of gene expression evolution is comparable in PSCs and differentiated tissues

In order to relate the expression patterns of stem cells to those observed in differentiated tissues, we used published RNA-Seq data (Brawand et al., 2011) from brain (cerebellum and cortex), kidney, heart, liver and testis from humans, bonobos, gorillas and macaques. These 44 samples were generated and analyzed in the same way as the stem cell data (Table S2). Based on the 2023 genes detected in all 63 samples, expression patterns cluster generally according to their tissue and then according to their species (Fig. 3A), as found previously (Brawand et al., 2011). To quantify differences, we used again the average squared Euclidian distance per gene between each pair of samples (Fig. S7). Distances within species were generally smallest in PSCs, congruent with a smaller environmental component in cell culture compared to postmortem tissue samples (Fig. S8). When subtracting this distance within species from the distance between species, the resulting expression divergence of PSCs was very similar compared to the other tissues (Fig. 3B), except for testis which has a higher divergence in agreement with previous results (Brawand et al., 2011; Khaitovich et al., 2005a). Hence, gene expression diverges in PSCs as quickly as in most differentiated tissues. This is interesting since one could argue that PSCs might be much less affected by selection for different functions than differentiated tissues given that they need to proliferate and remain pluripotent in all the analyzed species. Therefore, their rate of expression divergence could be regarded as a lower bound for how much expression patterns diverge in primates without changes in cell function. Since this rate is not significantly different from most other tissues, the finding confirms the view that the majority of expression differences between primates are selectively neutral with testis as a notable, possible exception (Brawand et al., 2011; Khaitovich et al., 2004, 2005b).

Identification and distribution of differentially expressed genes reveal conserved biological processes

Next, we investigated expression differences in the PSCs on the level of genes and functional categories of genes using all 6486 genes detected in more than 50% of all samples and at least once per species. A large fraction of genes differ in their expression levels between species as e.g. 5845 genes are affected by the factor species at a false discovery rate (FDR) of 5% (Storey's q-value, Storey and Tibshirani, 2003). Also a more conservative estimate that assesses the FDR based on permutations of species labels (Benjamini and Yekutieli, 2001; Prufer et al., 2007), results in 3406 genes. Hence, it is less relevant whether a gene does or does not differ significantly among species, but rather how much it differs. Therefore, we ranked genes according to their mean sum of squares (MSq) for the factor species, i.e. genes that differ a lot between the species rank high and genes with

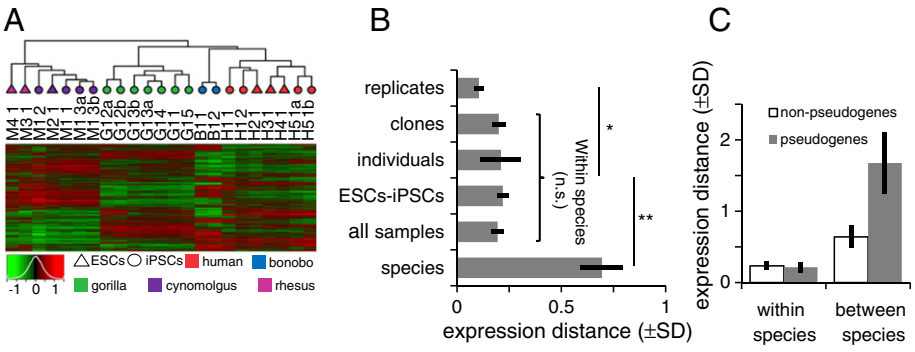


Figure 2 RNA-Seq of primate pluripotent stem cells. (A) Average linkage clustering of rank correlations of the 2877 genes detected in all 23 samples. Technical replicates are indicated by lower case letters. Expression levels (log2-scale) that are higher and lower than the gene average are shown in red and green, respectively. (B) Expression distances per gene are averaged from pairwise distances (Fig. S6B) for six different groups of comparisons (N = 4–7, see Materials and methods for details). Note that the distance among individuals is calculated within the same stem cell type and that for the comparisons between species, the two closely related macaque species were grouped together. Pairwise t-tests are all below 0.01 (**) for comparisons to the species distance, all below 0.05 (*) for comparisons to the replicate distance and all above 0.05 (n.s.) for comparisons within the species. (C) Expression distance as in (B) plotted separately for the 3374 non-pseudogenes and the 40 pseudogenes that are expressed in all samples.

very similar levels between species rank low. An enrichment analysis for variable (i.e. high-ranking) and conserved (i.e. low-ranking) groups of genes, revealed that by far the most significant (Wilcoxon rank test, $p < 10{-}19$) and strongest enrichment (3.3-fold higher average MSq), is seen for the group of 98 expressed pseudogenes. More than 80% of all non-pseudogenes show MSqs below the pseudogene average. This confirms the analysis of expression divergence of pseudogenes above and shows that also on the level of single genes, the vast majority of gene expression levels in PSCs evolve under constraint. Biological processes enriched for constrained genes (i.e. small MSq) include *regulation of chromatin* and *regulation of transcription*, indicating that these processes might evolve under particular constraint in PSCs (see Table S3 for results of all 1703 analyzed categories). Biological processes enriched for genes with large differences between species include e.g. *negative regulation of epithelial cell proliferation*. The 33 detected genes annotated in the group *stem cell maintenance* are not very enriched for large or small expression differences, indicating that this process has not been particularly affected during primate gene expression evolution. In general, there is no strong enrichment for functional categories since permutations of species labels result in similar numbers of conserved or variable categories (p = 0.4 and 0.6, respectively), indicating that factors influencing expression divergence are distributed across many annotated biological processes. However, analyses on their differentiated derivatives will

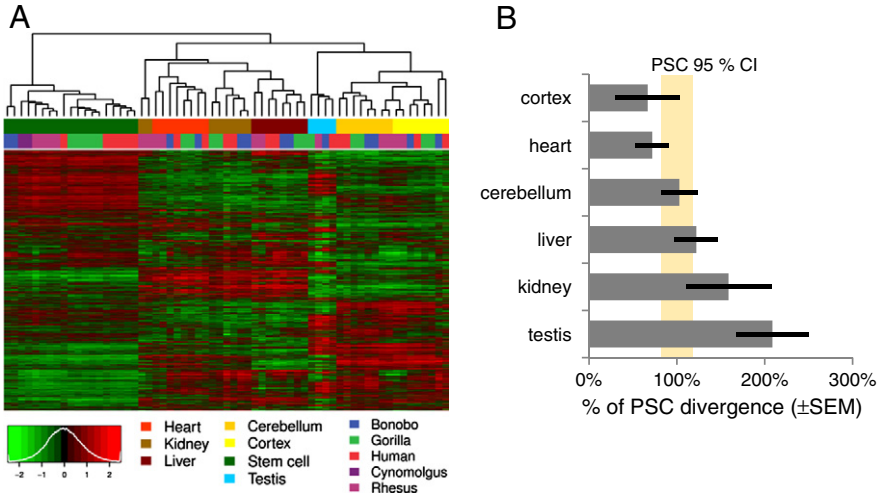


Figure 3 Expression evolution of primate PSCs and adult tissues. (A) Average linkage clustering of rank correlations of the 2016 genes detected in all 63 tissue samples. (B) Expression divergence in tissues relative to PSCs. For each species comparison, the divergence of the tissue (i.e. the distance within species subtracted from total distance between species that are plotted in Fig. S8) was scaled to the PSC divergence. Plotted are averages (±SEM) of the six pairwise species comparisons. The 95% confidence interval (CI) for the PSC divergence is shown in orange.

be needed to find out if this is generally the case and whether this is also true for the conservation and change of co-regulated gene networks.

Conclusions

In summary, we show that iPSCs are a practical and promising tool to analyse molecular and cellular phenotypes among primates. Despite very similar morphology, culture characteristics and general differentiation potential, expression differences between PSCs of the analyzed primates are more than 3-fold larger than differences between individual PSC clones. But since pseudogenes differ twice as much, expression levels are also constrained in their evolution. These patterns are of similar magnitude as in various somatic tissue types. Thus, primate iPSCs are likely to be informative for identifying conserved and derived expression patterns. Maybe the most informative will be the use of such cells to analyze the evolution of gene expression and regulatory networks during cell differentiation (e.g. Roy et al., 2013; Gifford et al., 2013). Hence, we think that primate iPSCs will become an important tool for understanding primate evolution on a cellular and molecular level.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.scr.2014.02.001.

Acknowledgments

We thank Ines Bliesener and Victor Wiebe (MPI-EVA) for their excellent technical support. This work was funded by the Max Planck Society (M.K., W.E. and S.P.) and by the German Research Foundation (Cluster of Excellence REBIRTH, EXC 62/3).

References

Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 29, 1165–1188.

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., Loring, J.F., 2011. Induced pluripotent stem cells from highly endangered species. Nat. Methods 8, 829–831.

Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H., Gnirke, A., Eggan, K., Meissner, A., 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., Csardi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M., Albert, F.W., Zeller, U., Khaitovich, P., Grutzner, F., Bergmann, S., Nielsen, R., Paabo, S., Kaessmann, H., 2011. The evolution of gene expression levels in mammalian organs. Nature 478, 343–348.

Cooper, G.M., Shendure, J., 2011. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. Nat. Rev. Genet. 12, 628–640.

Curwen, V., Eyras, E., Andrews, T.D., Clarke, L., Mongin, E., Searle, S.M., Clamp, M., 2004. The Ensembl automatic gene annotation system. Genome Res. 14, 942–950.

Enard, W., 2012. Functional primate genomics—leveraging the medical potential. J. Mol. Med. 90, 471–480.

Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R., Zhang, X., Coyne, M., Fostel, J.L., Holmes, L., Meldrim, J., Guttman, M., Epstein, C., Park, H., Kohlbacher, O., Rinn, J., Gnirke, A., Lander, E.S., Bernstein, B.E., Meissner, A., 2013. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell 153, 1149–1163.

Haase, A., Olmer, R., Schwanke, K., Wunderlich, S., Merkert, S., Hess, C., Zweigerdt, R., Gruh, I., Meyer, J., Wagner, S., Maier, L.S., Han, D.W., Glage, S., Miller, K., Fischer, P., Scholer, H.R., Martin, U., 2009. Generation of induced pluripotent stem cells from human cord blood. Cell Stem Cell 5, 434–441.

Khaitovich, P., Weiss, G., Lachmann, M., Hellmann, I., Enard, W., Muetzel, B., Wirkner, U., Ansorge, W., Paabo, S., 2004. A neutral model of transcriptome evolution. PLoS Biol. 2, E132.

Khaitovich, P., Hellmann, I., Enard, W., Nowick, K., Leinweber, M., Franz, H., Weiss, G., Lachmann, M., Paabo, S., 2005a. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science 309, 1850–1854.

Khaitovich, P., Paabo, S., Weiss, G., 2005b. Toward a neutral evolutionary model of gene expression. Genetics 170, 929–939.

Kircher, M., Stenzel, U., Kelso, J., 2009. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. Genome Biol. 10, R83.

Kircher, M., Heyn, P., Kelso, J., 2011. Addressing challenges in the production and analysis of Illumina sequencing data. BMC Genomics 12, 382.

Liu, H., Zhu, F., Yong, J., Zhang, P., Hou, P., Li, H., Jiang, W., Cai, J., Liu, M., Cui, K., Qu, X., Xiang, T., Lu, D., Chi, X., Gao, G., Ji, W., Ding, M., Deng, H., 2008. Generation of induced pluripotent stem cells from adult rhesus monkey fibroblasts. Cell Stem Cell 3, 587–590.

Marchetto, M.C., Narvaiza, I., Denli, A.M., Benner, C., Lazzarini, T.A., Nathanson, J.L., Paquola, A.C., Desai, K.N., Herai, R.H., Weitzman, M.D., Yeo, G.W., Muotri, A.R., Gage, F.H., 2013. Differential L1 regulation in pluripotent stem cells of humans and apes. Nature 503, 525.

Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. 2010 (pdb prot5448).

Nagaraj, N., Wisniewski, J.R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Paabo, S., Mann, M., 2011. Deep proteome and transcriptome mapping of a human cancer cell line. Mol. Syst. Biol. 7, 548.

Paten, B., Herrero, J., Beal, K., Fitzgerald, S., Birney, E., 2008. Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. Genome Res. 18, 1814–1828.

Paten, B., Herrero, J., Beal, K., Birney, E., 2009. Sequence progressive alignment, a framework for practical large-scale probabilistic consistency alignment. Bioinformatics 25, 295–301.

Pereira, V., Waxman, D., Eyre-Walker, A., 2009. A problem with the correlation coefficient as a measure of gene expression divergence. Genetics 183, 1597–1600.

Prufer, K., Muetzel, B., Do, H.H., Weiss, G., Khaitovich, P., Rahm, E., Paabo, S., Lachmann, M., Enard, W., 2007. FUNC: a package for detecting significant associations between gene sets and ontological annotations. BMC Bioinforma. 8, 41.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L., Pachter, L., 2011. Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. 12, R22.

Roy, S., Wapinski, I., Pfiffner, J., French, C., Socha, A., Konieczka, J., Habib, N., Kellis, M., Thompson, D., Regev, A., 2013. Arboretum: reconstruction and analysis of the evolutionary history of condition-specific transcriptional modules. Genome Res. 23, 1039–1050.

Shay, T., Jojic, V., Zuk, O., Rothamel, K., Puyraimond-Zemmour, D., Feng, T., Wakamatsu, E., Benoist, C., Koller, D., Regev, A.,

2013. Conservation and divergence in the transcriptional programs of the human and mouse immune systems. Proc. Natl. Acad. Sci. U. S. A. 110, 2946–2951.

Storey, J.D., Tibshirani, R., 2003. Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U. S. A. 100, 9440–9445.

Suemori, H., Tada, T., Torii, R., Hosoi, Y., Kobayashi, K., Imahie, H., Kondo, Y., Iritani, A., Nakatsuji, N., 2001. Establishment of embryonic stem cell lines from cynomolgus monkey blastocysts produced by IVF or ICSI. Dev. Dyn. 222, 273–279.

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

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

Wunderlich, S., Haase, A., Merkert, S., Beier, J., Schwanke, K., Schambach, A., Glage, S., Gohring, G., Curnow, E.C., Martin, U., 2012. Induction of pluripotent stem cells from a cynomolgus monkey using a polycistronic simian immunodeficiency virus-based vector, differentiation toward functional cardiomyocytes, and generation of stably expressing reporter lines. Cell Reprogram 14, 471–484.

Yamanaka, S., 2012. Induced pluripotent stem cells: past, present, and future. Cell Stem Cell 10, 678–684.