nature
genetics

Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells

Sabine Loewer[1–4], Moran N Cabili[5,6], Mitchell Guttman[5,7], Yuin-Han Loh[1–4], Kelly Thomas[5,8], In Hyun Park[1–4,12], Manuel Garber[5], Matthew Curran[1,3], Tamer Onder[1–4], Suneet Agarwal[1–3], Philip D Manos[1,3,4], Sumon Datta[1,3,4], Eric S Lander[5–7], Thorsten M Schlaeger[1,3,4], George Q Daley[1–4,8,9] & John L Rinn[5,10,11]

The conversion of lineage-committed cells to induced pluripotent stem cells (iPSCs) by reprogramming is accompanied by a global remodeling of the epigenome[1–5], resulting in altered patterns of gene expression[2,6–9]. Here we characterize the transcriptional reorganization of large intergenic non-coding RNAs (lincRNAs)[10,11] that occurs upon derivation of human iPSCs and identify numerous lincRNAs whose expression is linked to pluripotency. Among these, we defined ten lincRNAs whose expression was elevated in iPSCs compared with embryonic stem cells, suggesting that their activation may promote the emergence of iPSCs. Supporting this, our results indicate that these lincRNAs are direct targets of key pluripotency transcription factors. Using loss-of-function and gain-of-function approaches, we found that one such lincRNA (lincRNA-RoR) modulates reprogramming, thus providing a first demonstration for critical functions of lincRNAs in the derivation of pluripotent stem cells.

Cellular reprogramming demonstrates the remarkable plasticity of cell fates, as illustrated by the isolation of iPSCs from fibroblasts[6–9]. Molecular analysis of epigenetic modifications has revealed a near-complete remodeling of the epigenome during reprogramming[1–4,12], resulting in the conversion of lineage-specific protein-coding gene and microRNA expression profiles similar to those seen in embryonic stem cells (ESCs)[2,6–9]. We and others have recently discovered a new class of lincRNAs that are expressed in a cell-type–specific manner[13] and which can associate with epigenetic regulators[11,14–16] involved in pluripotency and lineage commitment[17,18].

To date, it is not known whether large-scale transcriptional changes induced by reprogramming apply to lincRNAs and whether these changes have any functional relevance. To test this, we compared the transcriptional profiles of human lincRNAs alongside protein-coding genes across fibroblasts, their derivative iPSCs and ESCs. We reprogrammed four primary fibroblast lines[7] and validated the functionality of the resulting iPSC lines (**Supplementary Figs. 1** and **2**). We then performed DNA microarray analysis of the parental fibroblasts, seven of their derivative iPSC lines and two ESC lines. Consistent with previous studies, analysis of the gene expression profiles revealed that all iPSCs were similar to ESCs[19,20] and were distinct from fibroblasts (**Fig. 1a** and **Supplementary Fig. 3**). We detected 3,694 genes upregulated and 3,283 genes downregulated in iPSCs and ESCs compared with fibroblasts (greater than twofold, $P < 0.05$; **Fig. 1b**). Taken together, our fibroblast-derived iPSCs fulfill functional criteria of *bona fide* iPSCs[20] and exhibit a uniform protein-coding gene expression profile similar to ESCs.

To explore the expression of lincRNAs, we designed a microarray probing ~900 lincRNAs in the human genome[11] and analyzed their expression in the above cell lines. The global lincRNA expression profiles of the iPSCs were very similar to those of ESCs and were distinct from those of fibroblasts (**Fig. 1c**). We observed 133 lincRNAs that were induced and 104 lincRNAs that were repressed (greater than twofold, familywise error rate (FWER) < 0.05) across all iPSCs and ESCs compared with fibroblasts (**Fig. 1d,e** and **Supplementary Table 1**). Similar to the case with protein-coding genes, direct reprogramming resulted in concomitant activation or repression of numerous lincRNAs consistent with a reactivation of the ESC state.

To exclude the possibility that reprogramming-induced changes in lincRNA expression reflect the opening and closing of chromatin domains of neighboring protein-coding genes, we analyzed the correlation of expression between each reprogrammed lincRNA and its neighboring genes and found no significant correlation ($P = 0.999$; **Fig. 1f**). This indicates an independent and cell-type–specific regulation of lincRNA expression.

[1]Stem Cell Transplantation Program, Division of Pediatric Hematology and Oncology, Manton Center for Orphan Disease Research, Children's Hospital Boston and Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [2]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, USA. [3]Harvard Stem Cell Institute, Cambridge, Massachusetts, USA. [4]Stem Cell Program, Children's Hospital Boston, Boston, Massachusetts, USA. [5]The Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [6]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA. [7]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [8]Division of Hematology, Brigham and Women's Hospital, Boston, Massachusetts, USA. [9]Howard Hughes Medical Institute, Chevy Chase, Maryland, USA. [10]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts, USA. [11]Department of Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA. [12]Present address: Yale Stem Cell Center, Department of Genetics, Yale School of Medicine, New Haven, Connecticut, USA. Correspondence should be addressed to G.Q.D. (george.daley@childrens.harvard.edu) or J.L.R. (jrinn@broadinstitute.org).
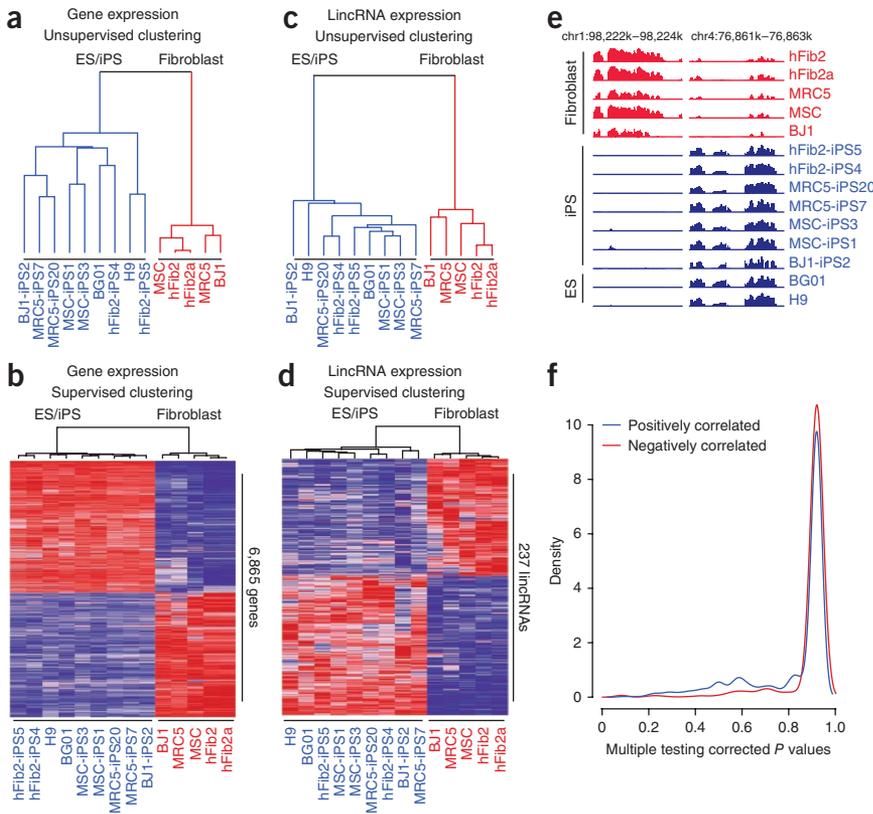
Figure 1 Direct reprogramming of fibroblasts converts both protein-coding genes and lincRNA expression to a pluripotent cell-specific profile. (**a,c**) Unsupervised hierarchical clustering of protein-coding gene expression (**a**) and lincRNA expression (**c**) segregates fibroblasts (red) from ESCs and fibroblast-derived iPSCs (blue). (**b,d**) Supervised hierarchical clustering analysis identified 6,865 protein-coding genes (**b**) and 237 lincRNAs (**d**) that are differentially expressed between ESCs and iPSCs, and fibroblasts (genes, greater than twofold, $P < 0.05$; lincRNAs, greater than twofold, FWER < 0.05). Expression values are represented in shades of red and blue relative to being above (red) or below (blue) the median expression value across all samples (log scale 2, from -3 to $+3$). hFib2 fibroblasts are represented as two replicates (hFib2 and hFib2a). (**e**) Examples of reprogrammed lincRNAs; left, lincRNA expressed in all fibroblasts is repressed in all pluripotent cells; right, a pluripotent cell-specific lincRNA that becomes activated during reprogramming. Expression values for each tiled probe (*x* axis) are displayed as normalized hybridization intensity (*y* axis). (**f**) Correlation analysis of lincRNAs and neighboring genes. Density plot of multiple testing–corrected P values (*x* axis) for lincRNAs that are positively (blue) or negatively (red) correlated with their protein-coding gene neighbors.

We sought to identify lincRNAs with potentially important functions in ESCs and iPSCs. Among the many pluripotency-associated lincRNAs (**Fig. 1d** and **Supplementary Fig. 4**), we searched for those that were expressed in both ESCs and iPSCs but which showed elevated levels in iPSCs relative to ESCs, reasoning that their higher expression there may have conferred a selective advantage

on emerging iPSCs. We identified 28 lincRNAs that showed greater expression in fibroblast iPSCs relative to ESCs (greater than twofold, FWER < 0.05; **Fig. 2a**), and we refer to these as 'iPSC-enriched' lincRNAs hereafter.

We hypothesized that if iPSC-enriched lincRNAs are important for reprogramming, they should be elevated in iPSCs independent
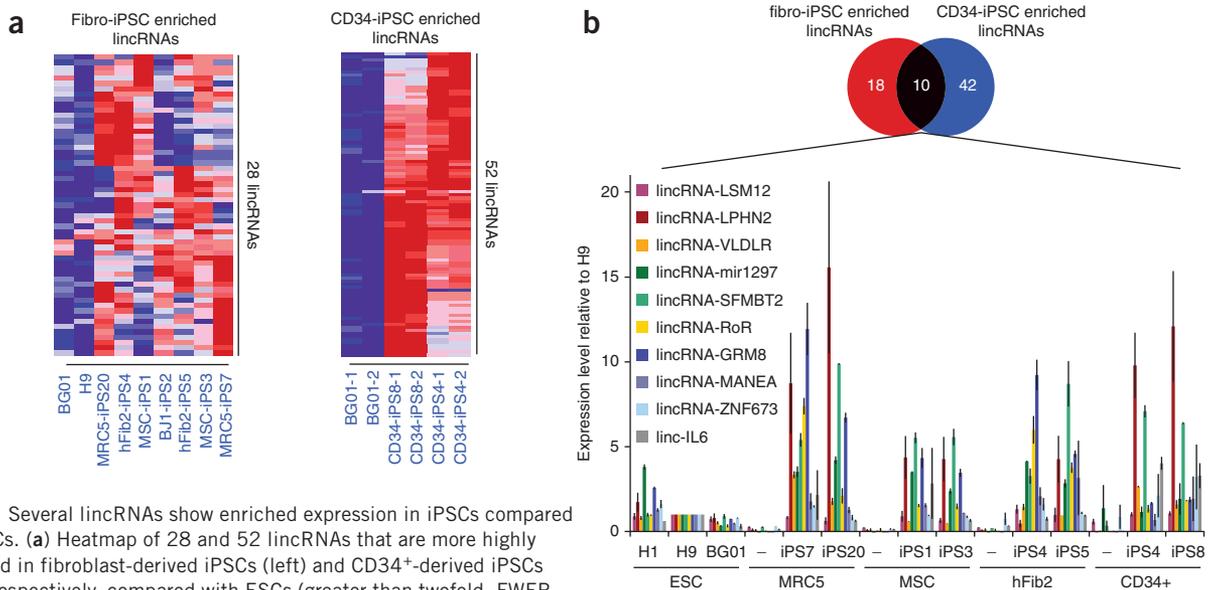

Figure 2 Several lincRNAs show enriched expression in iPSCs compared with ESCs. (**a**) Heatmap of 28 and 52 lincRNAs that are more highly expressed in fibroblast-derived iPSCs (left) and CD34+-derived iPSCs (right), respectively, compared with ESCs (greater than twofold, FWER < 0.05). Expression values are represented in shades of red and blue relative to being above (red) or below (blue) the median expression value across all samples (log scale 2, from -3 to $+3$). (**b**) Above, the venn diagram shows ten lincRNAs that are commonly enriched in fibroblast- and CD34+-derived iPSCs. Below, qRT-PCR validation of the ten commonly enriched lincRNAs (named according to their 3′ protein-coding gene neighbor) across three human ESC lines (H1, H9, BG01), fibroblasts (MRC5, MSC, hFib2) and CD34+ cells, and their derivative iPSC lines. Expression values are represented relative to the RNA levels in H9 ESCs.
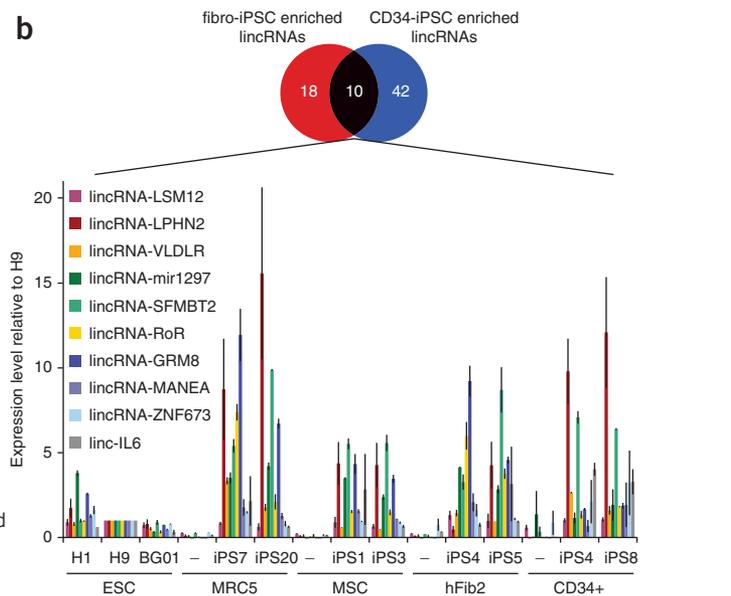
Figure 3 Transcriptional regulation of iPSC-enriched lincRNAs. (**a**) iPSC-enriched lincRNA loci are bound by pluripotency transcription factors. Above, lincRNA loci demarcated by domains enriched in histone H3K4me3-indicating RNA polymerase II promoters and H3K36me3-indicating regions of transcriptional elongation[10,29] in human ESCs (green and blue, respectively). Below, ChIP in hFib2-iPS5 cells followed by quantitative PCR analysis detects binding of OCT4, SOX2 and NANOG within lincRNA-SFMBT2, lincRNA-VLDLR and lincRNA-RoR regions close to lincRNA promoter regions (peaks of H3K4me). ChIP enrichment values are displayed normalized to a control region (chromosome 12, positions 7,839,777–7,839,966; hg18); anti-GFP ChIP was used as a negative control. Positions of ChIP-PCR fragments are indicated by black lines. (**b**) Changes in iPSC-enriched lincRNA levels upon siRNA-mediated knockdown of OCT4 in iPSC. Above, qRT-PCR of *OCT4*, *NANOG* and *LMNA* transcript levels upon depletion of OCT4. Below, qRT-PCR of iPSC-enriched lincRNA levels upon depletion of OCT4. Transcript levels are displayed relative to non-targeting control siRNAs (ctrl siRNA) (*n* = 3; error bars, ± s.e.m). (**c**) iPSC-enriched lincRNA expression during embryoid body differentiation. Above, qRT-PCR analysis monitoring transcript levels of pluripotency markers (*OCT4* and *NANOG*) and the differentiation marker *LMNA* over a 10-day differentiation time-course. Below, qRT-PCR analysis of iPSC-enriched lincRNAs. RNA levels are depicted relative to undifferentiated cells on day 0 (*n* = 3; error bars, ± s.e.m).
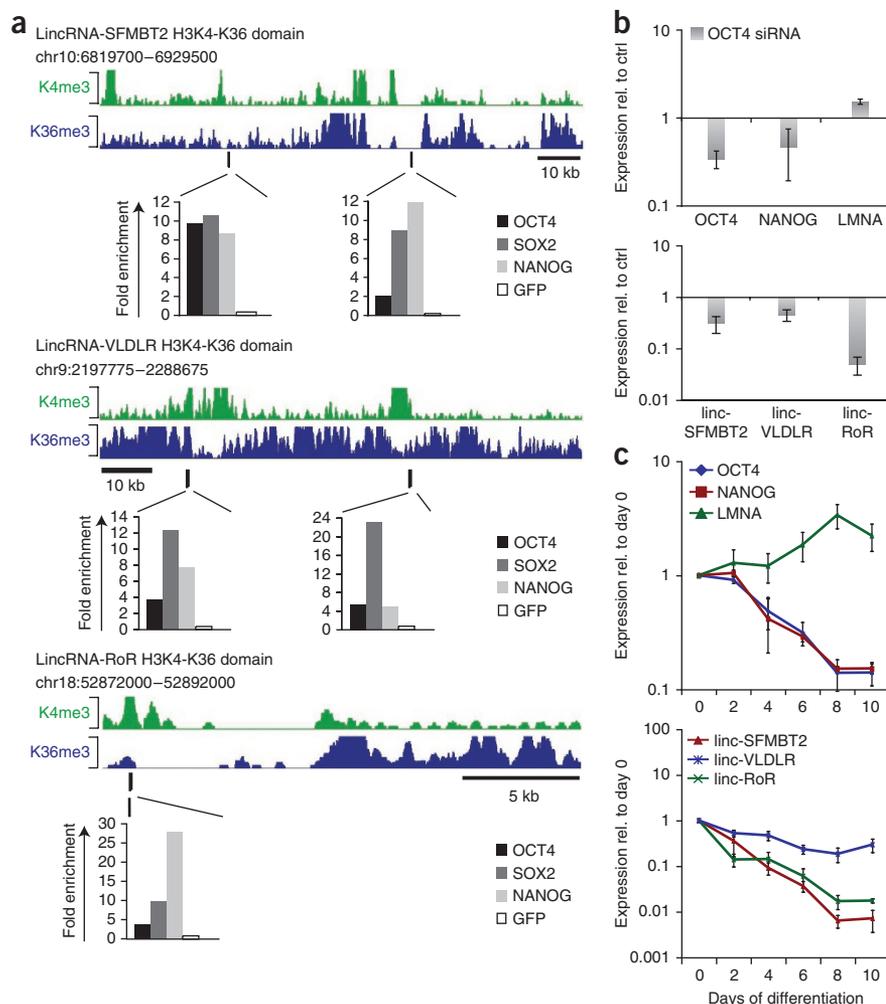
of the cell of origin. To test this, we profiled lincRNA expression in CD34[+] hematopoietic stem and progenitor cells, two CD34[+] iPSC lines[21], and ESCs using the same approach as above. Like fibroblast iPSCs, CD34[+] iPSCs had similar global lincRNA expression profiles as ESCs which were distinct from those of CD34[+] cells (**Supplementary Fig. 4**). Ten of the twenty-eight lincRNAs elevated in fibroblast iPSCs were also elevated in CD34[+] iPSCs (**Fig. 2b** and **Supplementary Fig. 5**). This overlap was statistically significant (*P* < 0.0001). We independently validated the levels of eight out of ten common iPSC-enriched lincRNAs by quantitative RT-PCR (qRT-PCR) (**Fig. 2b**) and detected considerable variation in expression. Positive selection for minimal RNA levels and the absence of counter selection against higher expression during reprogramming may be the cause for this variability. Collectively, these results show that numerous lincRNAs are tightly associated with the pluripotent state, including a subset of lincRNAs that are consistently enriched in iPSCs independent of the cell of origin.

If iPSC-enriched lincRNAs are important for iPSC derivation, we suspected that a link with the pluripotency network may exist. To test this, we first intersected previously published OCT4 binding regions in ESCs[22] with iPSC-enriched lincRNA loci (demarcated by domains of histone H3K4 and H3K36 methylation[10,11], named according to their neighboring 3′ gene) and identified three overlapping loci: lincRNA-SFMBT2, lincRNA-VLDLR and lincRNA-RoR (formerly called lincRNA-ST8SIA3). We performed independent ChIP-qPCR to validate the binding of OCT4 and probed for SOX2 and NANOG occupancy at these sites. All three transcription factors

occupied these regions, coinciding with or being in close proximity to lincRNA promoters (peaks of H3K4me (ref. 10); **Fig. 3a** and **Supplementary Fig. 6**).

To determine whether expression of iPSC-enriched lincRNAs is dependent on pluripotency transcription factors, we depleted OCT4 in iPSCs and ESCs using short interfering RNAs (siRNAs) and monitored the levels of iPSC-enriched lincRNAs. We verified OCT4 knockdown and induction of the differentiation marker *LMNA* (**Fig. 3b** and **Supplementary Fig. 7**). Levels of all three iPSC-enriched lincRNAs dropped within 72 h (**Fig. 3b** and **Supplementary Fig. 7c**). To further verify that downregulation of iPSC-enriched lincRNAs is caused by perturbation of the pluripotency network, we induced embryoid body formation as a distinct pathway of differentiation. Again, levels of all three iPSC-enriched lincRNAs dropped within two days (**Fig. 3c** and **Supplementary Fig. 7d**). The expression of these lincRNAs thus appears to be controlled by pluripotency transcription factors in ESCs and iPSCs.

We then turned to investigate the functional roles of iPSC-enriched lincRNAs in the reprogramming process. To this end, we generated short hairpin RNA (shRNA)-expressing lentiviruses targeting lincRNA-RoR and lincRNA-SFMBT2, which showed the strongest response to embryoid body differentiation and OCT4 knockdown and validated each knockdown relative to a nontargeting control shRNA (**Fig. 4a** and **Supplementary Fig. 8a**). To test the effect of lincRNA depletion on reprogramming, we infected dH1f fibroblasts[7] with

Figure 4 LincRNA-RoR expression modulates reprogramming. (**a**) qRT-PCR verifies lincRNA-RoR knockdown with Linc-sh1 and Linc-sh2 in hFib2-iPS5 cells relative to a non-targeting shRNA control ($n = 2$, error bar, ± s.e.m.). (**b**) Quantification of Tra-1-60+ iPSC colonies upon knockdown of lincRNA-RoR relative to the control (day 21; $n = 4$; error bar, ± s.e.m). (**c**) Quantification of cell numbers on days 6 and 7 of reprogramming in lincRNA-RoR shRNA samples relative to the control ($n = 4$; error bar, ± s.e.m). (**d**) Images showing quarters of Tra-1-60 stained reprogramming plates upon infection of a non-targeting control and two lincRNA-RoR targeting shRNAs. Arrowheads mark Tra-1-60+ iPSC colonies. (**e**) Structure of the lincRNA-RoR locus. Green and blue, demarcation of the H3K4me-H3K36me domain in ESCs. Red, structure of lincRNA-RoR RNA. The asterisk marks the position of OCT4-SOX2-NANOG binding (**Fig. 3a**). Right, RNA hybridization of lincRNA-RoR detects a 2.6-kb transcript in hFib2-iSP5 but not in dH1f (for full-length blot, see **Supplementary Fig. 12**). (**f**) qRT-PCR verifies lincRNA-RoR overexpression from a retroviral vector (pBabe-lincRNA-RoR) compared with pBabe-puro and pBabe-puro-GFP vectors in dH1f relative to the levels in H9 ESCs and hFib2-iPS5 ($n = 2$; error bars, ± s.e.m.). (**g**) Quantification of Tra-1-60+ iPSC colonies upon overexpression of lincRNA-RoR compared to pBabe and pBabe-GFP controls ($n = 5$; error bar, ± s.e.m.). (**h**) Quantification of cell numbers on days 6 and 7 in lincRNA-RoR–overexpressing cells and controls. Cell numbers are relative to the pBabe control (day 28 ± 2 days; $n = 5$; error bar, ± s.e.m.). (**i**) Image of quarter-plates of Tra-1-60 stained colonies (arrowheads) in pBabe, pBabe-GFP and pBabe-lincRNA-RoR infected samples. Statistical analysis was performed using a Student's t-test.

both the shRNA-expressing and the reprogramming viruses[7,9] and scored emerging iPSC colonies based on Tra-1-60 marker expression (at day 21)[20]. Interference with lincRNA-SFMBT2 did not affect iPSC colony formation (**Supplementary Fig. 8b,c**), suggesting that lincRNA-SFMBT2 is not essential, or alternatively, that its moderate reduction was insufficient to perturb reprogramming. In contrast, knockdown of lincRNA-RoR resulted in a significant twofold to eightfold decrease of iPSC colonies relative to the control, whereas progenitor cells were unaffected ($P < 0.01$; **Fig. 4b,c,d** and **Supplementary Table 2**). The resulting iPSC colonies fulfilled the additional criteria of fully reprogrammed cells (**Supplementary Fig. 9**). These results demonstrate a functional requirement of lincRNA-RoR expression for iPSC derivation.

Several studies have established critical roles of cell proliferation and a bypass of senescence during the early stages of reprogramming[23–27]. We therefore examined if knockdown of lincRNA-RoR compromised cell growth of fibroblasts or cells during this window, and we failed to detect significant differences in cells infected with the lincRNA-RoR–targeting virus compared with the control (**Fig. 4c** and **Supplementary Fig. 10**). In addition, the kinetics of reprogramming upon knockdown of lincRNA-RoR was similar to the control (**Supplementary Fig. 11**). Collectively, these findings point to a specific inhibition of the reprogramming process rather than a delay of iPSC formation upon loss of lincRNA-RoR.

Intrigued by this phenotype, we used 5′ and 3′ rapid amplification of complementary DNA (cDNA) ends to clone the full-length transcript of lincRNA-RoR (**Fig. 4e**), which recovered a 2.6-kb long RNA comprised of four exons (**Fig. 4e**, shown in red). We did not detect any clones that were spliced to protein-coding genes or intact open reading frames, and we confirmed the presence of a single transcript of expected length by RNA blotting (**Fig. 4e** and **Supplementary Fig. 12**).

We next used a complementary gain-of-function approach to test whether elevated lincRNA-RoR expression might enhance reprogramming. We infected dH1fs with empty pBabe-puro retrovirus, GFP-expressing virus or lincRNA-RoR–expressing virus, we selected transgenic cells, and we documented 25-fold to 70-fold overexpression of lincRNA-RoR relative to the levels in H9 ESCs (**Fig. 4f**). We induced
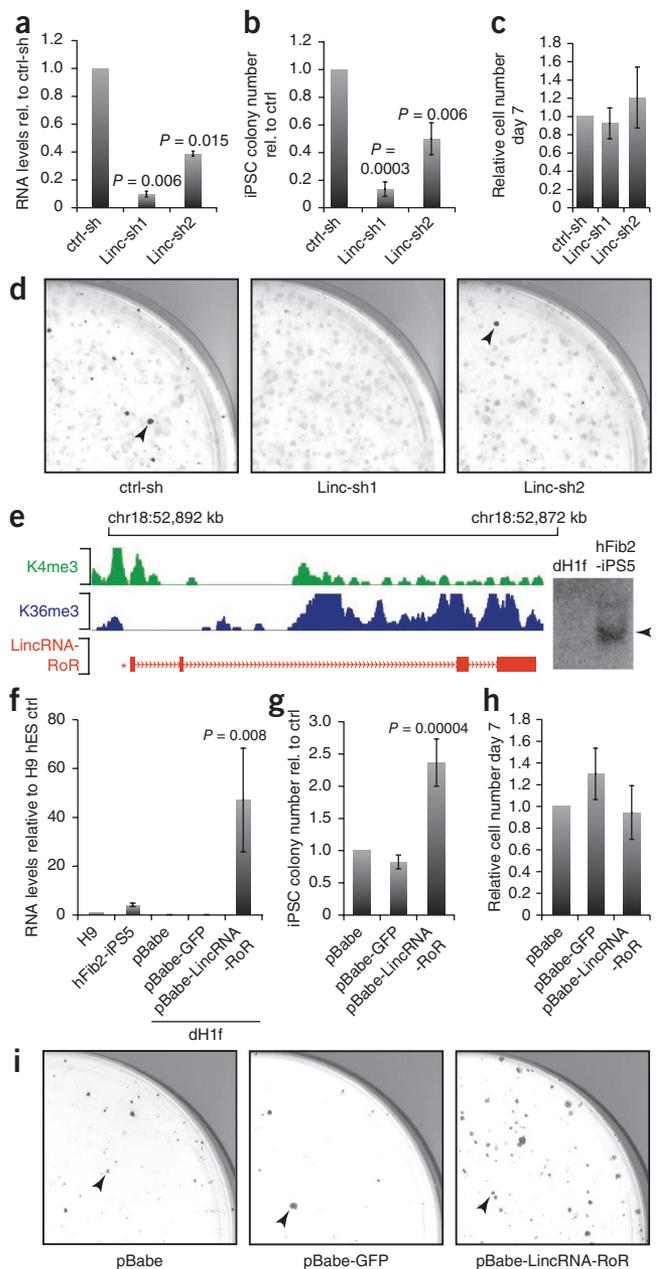
reprogramming in these stable cell lines and consistently observed a more than twofold increase in iPSC colony formation (at day 28 ± 2 days) ($P < 0.001$; **Fig. 4g**). This was not associated with significant changes in cell growth of fibroblasts or cells at the early stages of reprogramming (**Fig. 4h** and **Supplementary Fig. 10**). Thus, overexpression of lincRNA-RoR positively affects the establishment of iPSCs during reprogramming (**Fig. 4g,i**) in addition to having possible functions in iPSC maintenance. Supporting these latter functions, transient knockdown of lincRNA-RoR in ESCs and established iPSCs resulted in a growth deficiency linked with elevated apoptosis (**Supplementary Fig. 13**).

To gain insight into which cellular pathways are affected by lincRNA-RoR knockdown, we performed microarray gene expression analysis. Consistent with its apoptotic phenotype, knockdown of lincRNA-RoR led to upregulation of genes involved in the p53 response, the response to oxidative stress and DNA damage-inducing agents, as well as cell death pathways (**Supplementary Table 3**). Notably, simultaneous

knockdown of p53 partially rescued the apoptotic phenotype caused by ablation of lincRNA-RoR (**Supplementary Fig. 14**). Taken together, these results suggest that lincRNA-RoR plays a role in promoting survival in iPSCs and ESCs, likely by preventing the activation of cellular stress pathways including the p53 response.

Our transcriptional profiling approach has revealed numerous lincRNAs that are part of the transcriptional repertoire of human ESCs and are induced during reprogramming of different cell types. We have identified several iPSC-enriched lincRNAs that appear to be directly regulated by the pluripotency network. Notably, we found no direct syntenic correlates of the ten iPSC-enriched lincRNAs expressed in mouse ESCs (with the exception of lincRNA-VLDLR). Similar to what has been described for protein-coding genes[28], the transcriptional networks of lincRNAs in ESCs may have become rewired, conferring species-specific regulation.

The modulation of reprogramming by lincRNA-RoR provides the first functional example of a lincRNA in establishing iPSCs, and we therefore name it lincRNA-RoR for 'regulator of reprogramming'. Future studies will be required to decipher the molecular mechanism by which lincRNA-RoR acts and to gain a global understanding of lincRNA function in the establishment and maintenance of pluripotency. One possibility is that pluripotency-associated lincRNAs interface with chromatin-modifying complexes to assist in the regulation of the distinct epigenetic architecture in pluripotent cells. Supporting this, previous studies have demonstrated critical roles for chromatin-modifying complexes in the establishment and maintenance of pluripotency, and numerous lincRNAs can interact with these complexes to impart target specificity[11,15,16]. Here we demonstrate the modulation of reprogramming by a large non-coding RNA, supporting the notion that lincRNAs represent an additional layer of complexity in the networks controlling cellular identity.

URLs. ImageJ, http://rsbweb.nih.gov/ij/.

METHODS
Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession Numbers. All primary data are deposited in the Gene Expression Omnibus under accession number GSE24182.

Note: Supplementary information is available on the Nature Genetics website.

1. Ball, M.P. *et al.* Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* **27**, 361–368 (2009).
2. Chin, M.H. *et al.* Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* **5**, 111–123 (2009).
3. Doi, A. *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* **41**, 1350–1353 (2009).
4. Maherali, N. *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55–70 (2007).
5. Mikkelsen, T.S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
6. Lowry, W.E. *et al.* Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc. Natl. Acad. Sci. USA* **105**, 2883–2888 (2008).
7. Park, I.H. *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**, 141–146 (2008).
8. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
9. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
10. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
11. Khalil, A.M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. USA* **106**, 11667–11672 (2009).
12. Mikkelsen, T.S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).
13. Ponting, C.P., Oliver, P.L. & Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **136**, 629–641 (2009).
14. Nagano, T. *et al.* The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717–1720 (2008).
15. Rinn, J.L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
16. Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J. & Lee, J.T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).
17. Boyer, L.A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349–353 (2006).
18. Lee, T.I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006).
19. Assou, S. *et al.* A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. *Stem Cells* **25**, 961–973 (2007).
20. Chan, E.M. *et al.* Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. *Nat. Biotechnol.* **27**, 1033–1037 (2009).
21. Loh, Y.H. *et al.* Generation of induced pluripotent stem cells from human blood. *Blood* **113**, 5476–5479 (2009).
22. Marson, A. *et al.* Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521–533 (2008).
23. Hong, H. *et al.* Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature* **460**, 1132–1135 (2009).
24. Kawamura, T. *et al.* Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature* **460**, 1140–1144 (2009).
25. Li, H. *et al.* The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature* **460**, 1136–1139 (2009).
26. Marion, R.M. *et al.* A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature* **460**, 1149–1153 (2009).
27. Utikal, J. *et al.* Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. *Nature* **460**, 1145–1148 (2009).
28. Kunarso, G. *et al.* Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* **42**, 631–634 (2010).
29. Dinger, M.E. *et al.* Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* **18**, 1433–1445 (2008).

ONLINE METHODS

All primer, siRNA, RNA probe, cDNA and cloning sequences are listed in **Supplementary Table 4**.

Microarray analysis. Total RNA was isolated using RNA Stat-60 (Tel-Test) and was DNase treated (DNAfree, Ambion). For protein-coding gene expression analysis, total RNA was hybridized to AffymetrixU133Plus2.0 chips and processed as described[10]. For lincRNA expression analysis, total RNA was amplified using MessageAmp II (Ambion), labeled and hybridized to lincRNA arrays as described[11].

Statistical analysis. Affymetrix gene expression arrays were normalized as described[11]. Differentially expressed genes were identified using a Student's t-test (two-tailed, two-sample equal variance). LincRNA microarray probe intensities were quantile normalized and log transformed, and significantly enriched lincRNA regions were identified as described[11]. Differentially expressed lincRNAs (**Supplementary Table 1**) were identified using a Student's t-test, and significance was estimated using 1,000 permutations of class labels to control for a familywise error rate (FWER < 0.05).

Unsupervised and supervised hierarchical clustering of gene and lincRNA expression profiles were performed using GenePattern[30].

To compute the correlation between lincRNAs and neighboring protein-coding genes, we computed a Pearson correlation coefficient for each lincRNA to both its left and right neighboring gene across the full datasets. We then permuted gene locations and computed the same correlation coefficient for each lincRNA against randomized gene neighbors. We performed 1,000 permutations and assessed the statistical significance of this interaction by comparing the observed scores to the randomly permuted scores.

Statistical significance of overlapping iPSC-enriched lincRNAs in fibroblast iPSCs and CD34+iPSCs. Random simulations were used to calculate the probability of obtaining an overlap as large as we identified while at the same time controlling for the set size of both differentially expressed and upregulated lincRNAs in each iPSC type.

Statistical analysis of iPSC colony yield. Each sample was normalized to the total number of iPSC colonies within one experiment to weigh out the variations in colony numbers across experiments. The resulting values representing fractions of colony numbers within each experiment were then used for statistical analysis using a Student's t-test (two-tailed, two-sample equal variance).

Protein-coding genes deregulated upon lincRNA-RoR knockdown. Affymetrix gene expression arrays were normalized as described[11]. PaGE analysis was used to identify genes that are differentially expressed in lincRNA-RoR siRNA knockdown iPSC while comparing with control samples[31] (false discovery rate (FDR) < 0.2, default parameters). Gene set enrichment analysis was performed using 100 permutations of gene sets and a t-test as test statistics (FDR < 0.2).

qRT-PCR. cDNA was synthesized with SuperScript II (Invitrogen) and qPCR was performed using the Brilliant SYBR Green QPCR mix. Relative expression values were calculated ($\Delta\Delta$CT method) using GAPDH or β-ACTIN as a normalizer.

Immunostaining. Cells were fixed with 4% p-formaldehyde and stained with biotin-anti-Tra-1-60 (eBioscience, #13-8863-82) and streptavidin horseradish peroxidase (HRP) (Biolegend, #405210) diluted in PBS (3%), FCS (0.3%) Triton X-100. Staining was developed with the Vector labs DAB kit (#SK-4100), and iPSC colonies quantified with ImageJ software.

Cell culture and siRNA transfection. iPSCs and hESCs were cultured and embryoid body differentiation in suspension was performed as described[7]. For transfections or infections, iPSCs were dissociated with Accutase and plated in mTeSR media (STEMCELL Technologies) with 10 μM Y-27632 (Calbiochem) at 35K–50K cells per 24 wells or 100K per 6 wells pre-coated with matrigel (BD Bioscience-#345277). One hundred nanomolar OCT4- (ref. 32), GAPDH- or non-targeting siRNAs were transfected using DharmaFECT 1 (Dharmacon-#T-2001-01). See **Supplementary Table 4c** for siRNA sequences.

ChIP assays. ChIP was performed as described[33]. Chromatin extracts were immunoprecipitated using anti-OCT4 (Santa Cruz, sc-8628), anti-SOX2 (R&D Systems, AF-2018), anti-NANOG (R&D Systems, AF-1997) or anti-GFP (Santa Cruz, sc-9996) control. Fold enrichments were calculated by determining the ratio of immunoprecipitated DNA to input and normalizing to the levels observed at a control region. See **Supplementary Table 4a** for primer sequences.

Reprogramming assays. Reprogramming infections were performed as described[7]. For knockdown studies, dH1f were infected with a reprogramming virus alongside lentivirus expressing non-targeting control (SHC002V) or shRNAs targeting lincRNAs. shRNAs were designed using the iRNAi software, cloned into pLKO.1-puro (Addgene) and verified by sequencing. See **Supplementary Table 4c** for shRNA sequences. After infection, cells were grown for 6 days in αMEM 10% FCS, dissociated, counted and plated onto mouse embryonic fibroblasts. Twenty-four hours later, cells were cultured in hESC media; colonies were scored between days 21 and 28. For overexpression studies, lincRNA-RoR cDNA was cloned with EcoRI into pBabe-puro (Addgene-#1764) and verified by sequencing. dH1f were infected twice with pBabe-puro, pBabe-puro-GFP and pBabe-lincRNA-RoR retrovirus, and 1 μg/ml puromycin was added 48 h later. Puromycin was removed for 48 h before reprogramming infections.

Cloning of lincRNA-RoR. 5′ and 3′ Rapid Amplification of cDNA Ends (RACE) were performed using the First Choice RLM-RACE kit (Ambion). See **Supplementary Table 4a** for primer sequences. PCR products were cloned into pCR4-TOPO using the Zero Blunt PCR TOPO cloning kit (Invitrogen). Resulting pCR4-lincRNA-RoR clones were sequenced. LincRNA-RoR cDNA was isolated from pCR4-lincRNA-RoR with EcoRI digest and cloned into pBabe-puro. Resulting pBabe-lincRNA-RoR clones were sequenced.

RNA blot. Total RNA was isolated from dH1f and hFib2-iPS5 cells and genomic DNA was removed (see above). Twenty micrograms of RNA per sample were resolved on a 1% denaturing agarose gel. Following blotting using the Turbo Blotting System (Whatman), the membrane was ultraviolet crosslinked and pre-hybridized for 1 h at 42 °C in ULTRAhyb Ultrasensitive Hybridization Buffer (Ambion, #AM8669). A radiolabeled probe directed against lincRNA-RoR was hybridized overnight. The membrane was washed three times for 5 min in ×2 SSC/0.1% SDS at 42 °C, then three times for 15 min in ×0.1 SSC/0.1% SDS. To generate the radiolabeled probe, a 231-bp fragment of lincRNA-RoR was PCR amplified from pCR4-lincRNA-RoR. A fifty nanogram PCR fragment was used to produce the probe using Ready-To-Go DNA Labeling Beads (GE Healthcare-#27-9240-01).

Cells used for microarray expression analysis. *Fibroblasts.* The fibroblasts used were MRC5 (fetal lung fibroblasts) passage (P) 8; MSC (mesenchymal stem cells derived from bone marrow) P9; hFib2 P13 (adult forearm fibroblasts); and BJ1 P7 (neonatal foreskin fibroblasts)[34].

CD34+cells. The CD34+ cells used were mobilized peripheral blood CD34+ cells (AllCells, mBP014F). Cells were thawed and cultured as described[21] for two days before RNA isolation.

hESCs. The hESCs used were H1 (NIH code WA01) P29; H9 (WA09) P49; and BG01 P35.

iPSCs. The iPSCs used were MRC5-iPS7 P13; MRC5-iPS20 P14; MSC-iPS1 P13; MSC-iPS3 P10; hFib2-iPS4 P27; hFib2-iPS5 P21; BJ1-iPS2 P16 (ref. 34); CD34-iPS4 P9; and CD34-iPS8 P9 (ref. 21).

Fluorescent immunostaining and live-cell imaging. *Immunostaining.* Cells grown in 96-well plates (Matrix-#4940) coated with hESC-qualified matrigel (BD Biosciences-#345277) were fixed for 20–30 min with 4% p-formaldehyde/PBS (+/+), washed several times with PBS (+/+) and incubated overnight at 4 °C with primary antibody and Hoechst diluted in 3% donkey serum/3% BSA Fraction VII/0.01% Triton X-100/PBS (+/+) (Alexa 647-coupled anti-SSEA4 (1:100), BD Biosciences #560219; Alexa 555-coupled anti-Tra-1-60, BD Biosciences #560121 (1:75); Hoechst, Invitrogen #H3570 (1:20,000)). After

several washes with PBS (+/+), images were acquired using a BD Pathway 435 imager equipped with a ×10 objective. Areas corresponding to 18.6 mm^2 were imaged. Four images were acquired per frame (Hoechst, GFP, Alexa Fluor 555, Alexa Fluor 647). GFP acquisition settings were optimized for detection of high-level proviral GFP expression.

Live-cell imaging. Cells grown on MEFs in 6-well plates were incubated with Alexa 647-coupled anti-TRA-1-60 (BD Biosciences #560122, 1:75) and Alexa 555-coupled anti-SSEA4 (BD Biosciences #560218, 1:100) for 2 h at 37 °C. Where applicable, Hoechst (Invitrogen #H3570, 1:20,000) was added after 1.5 h for the remaining 30 min. Cells were washed three times with PBS before 1 ml of fresh phenol red-free media was added per well, and images were acquired using a BD Pathway 435 imager equipped with a ×10 objective. Four areas corresponding to 53.21 mm^2 were imaged per 6 wells. Two or three images were acquired per frame (Alexa Fluor 555, Alexa Fluor 647, with or without Hoechst). Post-acquisition image processing was performed using ImageJ (flatfield-correction, background subtraction; see URLs) and/or Adobe Photoshop (pseudocoloring, multi-color composites).

Teratoma formation assay. iPSCs grown on matrigel were harvested with dispase (Invitrogen-#17105-041, 1 mg/ml in DMEM/F12). Cell clumps from one 6-well plate were resuspended in 50 µl DMEM/F12, 100 µl collagen I (Invitrogen-#A1064401) and 150 µl hESC-qualified matrigel (BD Biosciences-#354277). Cell clumps were then injected into the hind limb femoral muscles (100 µl suspension per leg) of Rag2 γ/c mice. After 6–8 weeks, teratomas were harvested and fixed in 4% p-formaldehyde overnight. Samples were then embedded in paraffin, and sections were stained with hematoxylin/eosin (Rodent Histopathology Core, Harvard Medical School, Boston, MA, USA).

Protein blot. Cells were lysed on ice in PBS/1% Triton X-100 containing protease inhibitor cocktail (Roche). Forty micrograms of protein was loaded per well of a 10% SDS polyacrylamide gel (BioRad, #345-0011). OCT4 was detected using a monoclonal mouse human OCT4 antibody (Santa Cruz Biotechnology, #sc-5279, 1:1,000), and the GAPDH loading control was detected using a rabbit antibody (Santa Cruz Biotechnology, #sc-255778, 1:1,000). Secondary antibodies used were horseradish peroxidase–coupled rabbit or mouse antisera (GE Healthcare, #NA934V or NA931V, 1:5,000). Proteins were detected using the Amersham ECL detection kit as described by the manufacturer.

Flow cytometric analysis. iPSCs were dissociated to single cells with Accutase, washed in PBS and stained with Alexa488-coupled AnnexinV and propidium iodide according to the manufacturer's instructions (Invitrogen Vybrant Apoptosis Assay Kit #2, #V13241). Samples were analyzed using a FACSCalibur flow cytometer and data was processed using FloJo Software.

Production of viral supernatants. 293T cells were plated at a density of 2.5 × 10^6 cells per 10-cm dish. The next day, cells were transfected with 2.5 µg viral vector, 2 µg Gag-Pol vector (pCMV-dR8.2 dvpr Addgene #8455, pUMVC Addgene #8449, or ps-PAX2 #12260) and 0.2 µg VSV-G plasmid (pCMV-VSV-G Addgene #8454, or pMD2.G Addgene #12259) using 15 µl Fugene 6 (Roche Applied Science #1181509001) in 50 µl DMEM per plate. Supernatant was collected 48 h and 72 h post-transfection and filtered through 45 µm pore size filters. For concentration, viral supernatants were centrifuged at 70,000 *g* at 4 °C for 90 min using a Beckman XL-90 ultracentrifuge. Reprogramming viruses were either retroviral[7] (MOI 2.5) or lentiviral[35] (Addgene #21162, 21164; 100 µl supernatant).

30. Reich, M. *et al.* GenePattern 2.0. *Nat. Genet.* **38**, 500–501 (2006).
31. Grant, G.R., Liu, J. & Stoeckert, C.J. Jr. A practical false discovery rate approach to identifying patterns of differential expression in microarray data. *Bioinformatics* **21**, 2684–2690 (2005).
32. Matin, M.M. *et al.* Specific knockdown of Oct4 and beta2-microglobulin expression by RNA interference in human embryonic stem cells and embryonic carcinoma cells. *Stem Cells* **22**, 659–668 (2004).
33. Loh, Y.H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **38**, 431–440 (2006).
34. Park, I.H., Lerou, P.H., Zhao, R., Huo, H. & Daley, G.Q. Generation of human-induced pluripotent stem cells. *Nat. Protoc.* **3**, 1180–1186 (2008).
35. Yu, J. *et al.* Human induced pluripotent stem cells free of vector and transgene sequences. *Science* **324**, 797–801 (2009).

Erratum: Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells

Sabine Loewer, Moran N Cabili, Mitchell Guttman, Yuin-Han Loh, Kelly Thomas, In Hyun Park, Manuel Garber, Matthew Curran, Tamer Onder, Suneet Agarwal, Philip D Manos, Sumon Datta, Eric S Lander, Thorsten M Schlaeger, George Q Daley & John L Rinn

Nat. Genet. **42**, 1113–1117 (2010); published online 7 November 2010; corrected after print 3 December 2010.

In the version of this article initially published, the corresponding author designation was incomplete. The corresponding authors should be George Q. Daley and John L. Rinn. The error has been corrected in the HTML and PDF versions of the article.