Published in final edited form as:
Cell Stem Cell. 2015 January 8; 16(1): 88–101. doi:10.1016/j.stem.2014.11.005.

Single cell transcriptome analysis reveals dynamic changes in lncRNA expression during reprogramming

Daniel H. Kim[1,*], **Georgi K. Marinov**[1], **Shirley Pepke**[3], **Zakary S. Singer**[1,2], **Peng He**[1], **Brian Williams**[1], **Gary P. Schroth**[4], **Michael B. Elowitz**[1,2], and **Barbara J. Wold**[1,*]

[1]Division of Biology and Biological Engineering, Pasadena, CA 91125, USA

[2]Howard Hughes Medical Institute, Pasadena, CA 91125, USA

[3]Center for Advanced Computing Research California Institute of Technology, Pasadena, CA 91125, USA

[4]Illumina, Inc., San Diego, CA 92122, USA

SUMMARY

Cellular reprogramming highlights the epigenetic plasticity of the somatic cell state. Long noncoding RNAs (lncRNAs) have emerging roles in epigenetic regulation, but their potential functions in reprogramming cell fate have been largely unexplored. We used single-cell RNA sequencing to characterize the expression patterns of over 16,000 genes, including 437 lncRNAs, during defined stages of reprogramming to pluripotency. Self-organizing maps (SOMs) were used as an intuitive way to structure and interrogate transcriptome data at the single-cell level. Early molecular events during reprogramming involved the activation of Ras signaling pathways, along with hundreds of lncRNAs. Loss-of-function studies showed that activated lncRNAs can repress lineage-specific genes, while lncRNAs activated in multiple reprogramming cell types can regulate metabolic gene expression. Our findings demonstrate that reprogramming cells activate defined sets of functionally relevant lncRNAs and provide a resource to further investigate how dynamic changes in the transcriptome reprogram cell state.

INTRODUCTION

Normal embryonic development proceeds through a progressive narrowing of cell fate potential and loss of cellular plasticity, coupled with the acquisition of increasingly specialized, mature cell phenotypes (Hemberger et al., 2009). This progression is reflected in the changing transcriptome and is driven by transcription factors and noncoding RNAs,

*Correspondence: dkim@caltech.edu (D.H.K), woldb@caltech.edu (B.J.W.).

Accession numbers

Sequencing data have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus under accession number GSE55291.

which comprise a regulatory network that is highly resistant to perturbation. Various classes of noncoding regulatory RNAs, including lncRNAs (Lee, 2012; Rinn and Chang, 2012) and small RNAs (e.g. Piwi-interacting RNAs) (Law and Jacobsen, 2010; Moazed, 2009), contribute to the establishment of epigenetic chromatin marks that stabilize cell state. Thus, to reprogram the somatic identity of a cell, the combined effects of epigenetic and regulatory circuit stability must be overcome.

Induced reprogramming to the pluripotent state can be initiated by the enforced expression of Oct4, Sox2, Klf4, and Myc (OSKM) (Takahashi and Yamanaka, 2006). These factors act in conjunction with other transcription factors and multiple chromatin-modifying enzymes (Onder et al., 2012) to initiate a cascade of changes that ultimately convert a somatic cell of limited potential to the pluripotent state (Apostolou and Hochedlinger, 2013; Papp and Plath, 2013; Theunissen and Jaenisch, 2014). Apart from induced pluripotent stem (iPS) cell reprogramming, these defined factors and chromatin regulators have also been shown to facilitate malignant transformation and progression, which might be viewed as a form of pathological reprogramming (Goding et al., 2014; Suva et al., 2013). For example, the catalytic subunit of Polycomb repressive complex 2 (PRC2), Ezh2, enhances the reprogramming potency of OSKM (Buganim et al., 2012) and is also overexpressed in multiple malignancies, including metastatic prostate cancer (Varambally et al., 2002) and lymphomas (Laugesen and Helin, 2014). PRC2 physically associates with lncRNAs in embryonic stem (ES) cells (Guttman et al., 2011; Zhao et al., 2010) and other cell types, and lncRNAs such as Xist and HOTAIR guide PRC2 complexes to their genomic targets (Rinn and Chang, 2012). Notably, loss of Xist can lead to the development of hematologic cancer (Yildirim et al., 2013), while HOTAIR overexpression can facilitate breast cancer metastasis (Gupta et al., 2010). However, only a small fraction of the thousands of mostly uncharacterized lncRNAs are known to affect cell state (Flynn and Chang, 2014), and the ways in which they do so are not fully understood. At this early stage of understanding, a clearer and more comprehensive portrait of lncRNA expression could provide missing information on how a cell overrides its starting identity and redefines a new one, whether in the context of OSKM reprogramming or cellular transformation.

Insights into many aspects of reprogramming are sought at the single-cell level (Buganim et al., 2012; Polo et al., 2012; Smith et al., 2010) because each cell reveals a possibly unique expression state, with its particular repertoire of regulatory factors and target gene behavior. To gain a transcriptome-level understanding of how individual cells are reprogrammed, we used single-cell RNA sequencing (RNA-seq) (Ramskold et al., 2012), augmented by single-molecule RNA-FISH (smFISH) (Raj et al., 2008). Further analysis of induced lncRNAs identified distinct groups with possible roles in suppressing somatic cell identity, conferring greater cellular plasticity, or promoting proliferation and self-renewal. Loss-of-function experiments of induced lncRNAs provided evidence for specific repression of genes characteristic of mature cell fates or regulation of genes involved in metabolic functions. We suggest that lncRNAs identified in the context of somatic cell reprogramming may also act in pathological reprogramming, especially in cancers that engage other parts of the OSKM pathway.

RESULTS

Single-cell analysis of the reprogramming transcriptome

We performed single-cell RNA-seq on cells derived from the 'reprogrammable' mouse (Figure 1A) (Carey et al., 2010). Tail-tip fibroblasts (TTFs) from OSKM transgene-inducible mice were cultured in the presence of doxycycline (dox) under ES cell culture conditions for two weeks. Flow cytometry analysis revealed that these transgene-expressing cells (Figure S1A), which we refer to as transitional cells, did not yet express the SSEA1 antigen (Figure S1B). After three weeks of dox exposure, we obtained early-stage iPS cells that expressed SSEA1 (Figure S1B), exhibited ES cell-like morphology (Figure S1C), and had single-cell cloning efficiencies of ~50% when compared to ES cells (Figure S1D). We then cultured these early-stage iPS cells in the absence of dox for 4-6 additional weeks to profile the transcriptomes of late-stage iPS cells, along with pluripotent ES cells. Together, we generated 81 single-cell RNA-seq datasets (14-21 libraries per cell type) at a sequencing depth of 5-10 million mapped reads per cell. This unbiased experimental approach allowed us to detect and characterize the expression patterns of over 16,000 protein-coding and noncoding genes, including lncRNAs.

We analyzed these high-dimensional data using a self-organizing map (SOM) (Figure S2), which provides an intuitive way to visualize and interrogate single-cell heterogeneity based on the behavior of coordinately expressed gene sets. The SOM displays similarity relationships in a two-dimensional heat map in which spatial proximity reflects expression pattern similarity. We mapped 16,113 genes expressed at >10 RPKM (reads per kilobase per million mapped reads) onto a SOM, where each hexagonal unit represents a set of genes whose expression patterns are most similar to each other. These units are then clustered and are located in the same fixed positions across all single-cell components of the SOM. Each single-cell transcriptome can then be visualized as a component of the SOM (Figure 1A, five representative single-cell components shown for each cell type).

To examine how individual transcriptomes from different experimental time points were related, we performed principal component analysis (PCA), which revealed a clear separation between TTFs/transitional cells and early- and late-stage iPS/ES cells when projected onto the first two principal components (Figure 1B). We next examined genes with the highest correlation to *Oct4* expression and identified many genes expected to be involved in reprogramming and pluripotency, including Nanog, Cdh1, Tdh, Utf1, Sall4, Tet1, Rex1, Trim71, Fbxo15, Dppa2, Tcl1, and *Tcfcp2l1* (Figure 1C) (Orkin and Hochedlinger, 2011). The single-cell RNA-seq data also allowed us to identify genes of potential interest that may also contribute to somatic cell reprogramming (Figure 1C). Many of the genes that correlated most strongly with *Oct4* expression were heterogeneously expressed at lower levels in transitional cells but then became coherently expressed at higher levels in both early- and late-stage iPS cells (Figure 1C). To examine all protein-coding genes that were specifically activated upon OSKM induction, we performed hierarchical clustering on genes that were expressed at >10 RPKM in transitional cells and iPS cells, while being expressed at <10 RPKM in primary TTFs. This group included 1,835 protein-coding genes (Figure 1D).

Activation of Ras signaling pathways during reprogramming

To examine how individual cells responded to OSKM induction, we analyzed single-cell components of the SOM. After two weeks of OSKM expression, transitional cells strongly expressed a cluster of genes (cluster: 0) that was significantly enriched for the gene ontology (GO) term “Ras protein signal transduction” (Bonferroni-corrected P-value: 5.50e-04) (Figure 2). 35 genes were annotated to this GO term, including *Nrg1*, which enhances self-renewal and proliferation of tumor-initiating cells (Lee et al., 2014), and *Git1* (Huang et al., 2014) and *Rap1a* (Bailey et al., 2009), which are involved in cancer cell metastasis. 19 lncRNAs (Table S1) were also coordinately expressed with these genes, suggesting their possible involvement in the regulation of Ras signaling pathways. Additionally, transitional cells exhibited strong upregulation of 16 lncRNAs (Table S1) within a cluster of genes (cluster: 21) significantly enriched for the GO term “regulation of signal transduction” (Bonferroni-corrected P-value: 5.55e-03) (Figure 2). 100 genes were annotated to this term, such as the oncogenes *Myc* and *Hras1*, and the Ras activator *Rasgrp3* (Yang et al., 2010). One cluster in particular (cluster: 24) was significantly enriched for the GO term “intracellular signal transduction” (Bonferroni-corrected P-value: 3.52e-07) (Figure 2) and contained numerous Ras-related genes, including *Rras, Rheb*, and several members of the Ras oncogene family (e.g. *Rab2a*, *Rab11a*) and Ras homolog family (e.g. *Rhog*, *Rhoa*, *Rhoc*, *Rhoq*) of genes (Karnoub and Weinberg, 2008). Taken together, these results show that activation of Ras signaling pathways and associated lncRNAs are early molecular events in this reprogramming paradigm.

Coordinate expression of pluripotency factors and noncoding genes

Early-stage iPS cells strongly expressed numerous pluripotency factors and related genes in one prominent cluster (cluster 7), including Oct4, Sox2, Nanog, Lin28a, Rex1, Esrrb, Lifr, Fbxo15, Sall4, Epcam, Dppa2, Dnmt3l, Tet1, Jarid2, Klf5, Trim71, Nodal, Tcfcp2l1, Tcl1, Cdh1, Tdh, Utf1, and *Eras* (Figure 2). However, the most significantly enriched GO term in this cluster was for “unannotated” (Bonferroni-corrected P-value: 0) and included 285 noncoding genes of unknown function, underscoring the need for a more in-depth characterization of the noncoding transcriptome. Among these genes were 29 lncRNAs (Table S1) and 27 processed transcripts, which are also classified as a type of lncRNA in the Ensembl annotation (Flicek et al., 2014), as well as numerous pseudogenes. A nearby cluster (cluster: 2) in the SOM also contained the *Meg3* and *Rian* lncRNAs, along with 32 additional lncRNAs (Table S1), *Klf4*, and *Kras*. Moreover, two clusters (cluster: 9, cluster: 10) directly adjacent to the pluripotency cluster (cluster: 7) were significantly enriched for the core PRC2 components *Ezh2*, *Suz12*, and *Eed* and 14 lncRNAs (Figure 2 and Table S1). Taken together, these results demonstrate that early-stage iPS cells exhibit a significant upregulation of hundreds of noncoding genes, including lncRNAs, whose reprogramming functions are largely unknown.

LncRNAs activated during reprogramming

To systematically characterize the expression patterns of lncRNAs activated during reprogramming (Ladr), we focused further analysis on the lncRNA transcriptome. We mapped 437 lncRNAs expressed at >10 RPKM onto a new, lncRNA-specific SOM (Figure

3A, five representative single-cell components shown for each cell type) to assess single-cell heterogeneity and to analyze individual lncRNA variation. Of the 437 lncRNAs, we identified 312 activated lncRNAs that were expressed at >10 RPKM in transitional cells and/or iPS and ES cells, while being expressed at <10 RPKM in TTFs (Figure 3B and Table S2), indicating the dynamic nature of the lncRNA landscape during somatic cell reprogramming.

To assess the potential functions of activated lncRNAs, we first examined previously identified catalogs of Polycomb-associated lncRNAs (Guttman et al., 2011; Zhao et al., 2010). This revealed substantial overlap between the PRC2 transcriptome and activated lncRNAs (n=150) (Figure 4A and Table S3), suggesting that many of these Polycomb-bound lncRNAs may repress developmental genes. Although several coherent clusters of PRC2-bound lncRNAs were strongly upregulated in early- and late-stage iPS cells relative to transitional cells, hierarchical clustering analysis suggested that these lncRNAs were heterogeneously expressed in individual cells. To examine whether this apparent heterogeneity could be attributed mainly to technical noise or to biological variation, we calculated the coefficient of variation (CV) for lncRNAs expressed in late-stage iPS cells and determined the relationship between the CV and mean expression level. Numerous lncRNAs (e.g. Ladr83, Ladr91, Ladr55, Ladr1, *Rian*, *Meg3*) exhibited greater variability when compared to the average noise for a given mean expression level, strongly suggesting that biological variation is indeed elevated for these lncRNAs (Figure 4B). Consistent with our single-cell RNA-seq analysis, clonal iPS cell lines exhibit functionally relevant heterogeneity in *Meg3* lncRNA expression, where iPS cell lines that lack *Meg3* expression are unable to generate all-iPS cell mice (Stadtfeld et al., 2012).

We also examined single-cell lncRNA heterogeneity at the transcriptome- and single-molecule levels using the lncRNA SOM and smFISH, respectively. For the lncRNA SOM, we calculated the component (single-cell) variance for each hexagonal cluster of lncRNA genes (Figure 4C). The lncRNA Ladr83 mapped to a cluster with high variance, while Ladr49 mapped to a lower variance cluster (Figure 4C; Figures S3A and S3B). We validated these lncRNA SOM results using smFISH as an orthogonal approach to determine the number of Ladr83 and Ladr49 molecules in hundreds of cells (n=351) (Figure 4D). Ladr83 was more heterogeneously expressed in both early- and late-stage iPS cells relative to ES cells, with a subset of late-stage iPS cells expressing high levels of Ladr83 (61-113 lncRNA molecules/cell) when compared to ES cells (median: 14 lncRNA molecules/cell). However, the distributions of Ladr49 expression in both early- and late-stage iPS cells were indistinguishable from ES cells (Figure 4E). Taken together, these results demonstrate the utility of the SOM in characterizing cell-to-cell variation and suggest that defined sets of lncRNAs are heterogeneously expressed in iPS cells.

Suppression of lineage-specific genes by activated lncRNAs

To examine the functional roles of Ladr49 and Ladr83, we used a pool of 2-4 small interfering RNAs (siRNAs) per lncRNA to knock down their expression levels in late-stage iPS cells (Figure S4A). While knockdown of Ladr49 or Ladr83 had modest effects on reprogramming efficiencies (Figure S4B), Ladr49 knockdown resulted in the upregulation of

muscle-related genes in iPS cells (Figure 5A). GO analysis revealed significant enrichment for the terms “contractile fiber,” “sarcomere,” and “striated muscle thin filament”. These muscle-related genes were strongly expressed in TTFs and clustered together in the SOM, while they were expressed at low levels in late-stage iPS cells (Figure 5B). These results indicate that Ladr49 is required to repress a subset of the myogenic program during somatic cell reprogramming.

Notably, Ladr83 knockdown resulted in the upregulation of muscle genes that were also upregulated upon Ladr49 loss-of-function (Figure 5C), suggesting that lncRNAs may act cooperatively and/or redundantly during cell fate reprogramming. However, both Ladr83 and Ladr49 knockdowns did not significantly perturb expression levels of the reprogramming/pluripotency genes *Oct4*, *Sox2*, *Klf4*, *Myc*, *Lin28*, and *Nanog* (Figure 5C), consistent with a specific role for these lncRNAs in repressing developmental genes during reprogramming. When we analyzed the differential expression of all Ensembl-annotated lncRNAs in populations of ES cells and TTFs by RNA-seq, Ladr49 and Ladr83 were among the 48 most significantly upregulated lncRNAs in ES cells versus TTFs (Figure 5D). 22 of these lncRNAs (46%), including Ladr49 and Ladr83, were previously shown to physically associate with PRC2 (Guttman et al., 2011; Zhao et al., 2010) (Figure 5E), suggesting that other PRC2-bound lncRNAs are likely to have functional roles in regulating lineage-specific genes.

To examine the cell-type specificity of activated lncRNAs, we also analyzed the transcriptome data from iPS cells derived from CD117+ hematopoietic progenitor cells (HPCs) (Chang et al., 2014). 130 lncRNAs were significantly upregulated during HPC reprogramming into iPS cells (Figure 6A), with 64 lncRNAs overlapping with the activated lncRNAs that we characterized during TTF reprogramming (Figure 6B). Of the remaining 66 lncRNAs specifically activated in HPC-iPS cells (Table S2), we examined one lncRNA in particular, Ladr317, using smFISH in early- and late-stage iPS cells derived from TTFs (Figure 6C). Ladr317 was expressed at both stages, with lower levels in early-stage iPS cells relative to ES cells. However, late-stage iPS cells exhibited comparable levels of Ladr317 when compared to ES cells. Loss-of-function experiments in late-stage iPS cells revealed that Ladr317 is required to repress genes involved in interferon signaling: *Irgm1*, *Usp18*, and *Ifit3* (Figure 6D), consistent with a functional role in repressing hematopoietic lineage-specific genes. When we examined lncRNAs activated during HPC reprogramming in our single-cell data from TTF reprogramming, only a small fraction of these lncRNAs were expressed in TTF-iPS cells (Figure 6E), suggesting that most of these lncRNAs are not involved in silencing fibroblast lineage genes.

Regulation of metabolic genes by activated lncRNAs

To explore the potential functions of lncRNAs activated in multiple reprogramming cell types, we performed hierarchical clustering on lncRNAs upregulated in both TTF-iPS and HPC-iPS cells. Many of these overlapping lncRNAs were strongly upregulated in early- and late-stage TTF-iPS cells (Figure 7A), in contrast to the same analysis performed using lncRNAs specific to HPC reprogramming (Figure 6E). We selected two of these robustly expressed lncRNAs for loss-of-function analysis, based additionally on their sequence

conservation in the human genome (Figure S5A) (Kuhn et al., 2013) and their high correlation with *Oct4* expression (Figure S5B). Knockdown of either Ladr86 or Ladr91 (Figures S3C, S3D, and S4A) repressed a common set of mitochondrial genes. Downregulated genes in both experiments were enriched for the GO terms “membrane-bounded organelle” and “mitochondrion” (Figures 7B and 7C), suggesting that these lncRNAs are involved in regulating metabolic aspects of reprogramming in at least two different lineages, and perhaps very broadly.

DISCUSSION

Normal somatic cells are understood to have restricted developmental and epigenetic plasticity (Hemberger et al., 2009). Multiple mechanisms stabilize the mature cell state and must be overcome during cellular reprogramming. To examine both known and novel regulators of cell state conversion, we measured the transcriptomes of individual cells at defined stages of reprogramming to pluripotency. As shown in earlier studies (Apostolou and Hochedlinger, 2013; Papp and Plath, 2013; Theunissen and Jaenisch, 2014), we observed coordinated expression of known OSKM pathway genes in early- and late-stage iPS cells, while identifying new co-expressed protein-coding and lncRNA genes.

Activation of many lncRNA and Ras-related genes are early molecular events that are prominent in transitional cells. While the significance of this observation is yet to be established, findings from other systems suggest that such mechanisms might be common to multiple reprogramming systems. For example, reprogramming by somatic cell nuclear transfer (SCNT) (Gurdon and Melton, 2008) may also involve lncRNA and Ras-related genes upregulated in our transitional cells, based on mapping oocyte single-cell RNA-seq data onto our SOM (Figures 2 and S2B) (Ramskold et al., 2012). Similarities between SCNT and iPS cell reprogramming are also highlighted by iPS cell activation of germ cell-related genes (Figures S6A, S6B, and S6C) and expression of Piwi-interacting RNA (piRNA)-like small RNAs (Figures S6D, S6E, and S6F), which are also detected in oocytes (Tam et al., 2008) and human iPS cells (Marchetto et al., 2013). Given their known functions in imprinting (Watanabe et al., 2011) and epigenetic regulation (Stuwe et al., 2014), piRNAs may also be involved in regulating imprinted loci during reprogramming. Because SCNT-derived pluripotent stem cells have less epigenetic and imprinting defects than iPS cells, investigating oocyte-enriched noncoding RNAs in OSKM-mediated reprogramming may provide insights for minimizing unwanted epigenetic signatures.

Our loss-of-function experiments argue that at least some OSKM-induced lncRNAs are functionally relevant in reprogramming. We have shown that individual lncRNAs are required to properly silence lineage-specific genes. OSKM-induced variation in lncRNA expression levels, however, might contribute to a form of “epigenetic memory” in iPS cells (Kim et al., 2010; Polo et al., 2010). For example, lower levels of Ladr49, generated by experimental knockdown, gave elevated expression of specific muscle-related genes in iPS cell populations. Failure to activate Ladr49 in iPS cells might introduce a bias favoring muscle derivatives and/or produce an inappropriate mixed phenotype. This scenario is consistent with prior reports that differences in OSKM stoichiometry during reprogramming

can produce varying and functionally relevant effects on the expression of specific imprinted lncRNAs (Carey et al., 2011; Stadtfeld et al., 2010).

OSKM-mediated reprogramming also results in the elevated expression of specific lncRNAs in iPS cells relative to ES cells (Figure 4E) (Loewer et al., 2010), but the functional consequences of these observations are unclear. One potential role for such lncRNAs is suggested by studies of lincRNA-RoR, which is required for iPS cell reprogramming (Loewer et al., 2010). LincRNA-RoR acts as a repressor of the tumor suppressor gene p53 (Zhang et al., 2013), conferring reprogrammed cells with the ability to overcome p53-mediated apoptosis. p53 inactivation, whether through genetic mutation or experimental perturbation, promotes cancer and somatic cell reprogramming (Krizhanovsky and Lowe, 2009). Additionally, leukemia cells activate the expression of many novel, unannotated lncRNAs that are regulated by oncogenic Notch signaling (Trimarchi et al., 2014). We also found that perturbing signaling pathways in iPS cells under “2i” conditions (Mek and Gsk3 inhibition) (Ying et al., 2008) activates a unique set of lncRNAs (Figure S7). Collectively, these results suggest that lncRNAs activated during reprogramming might also participate in oncogenic signaling pathways. Finally, we identified Ladr lncRNAs that are activated in multiple reprogramming cell types. Loss-of-function experiments for Ladr86 and Ladr91 affected genes acting in mitochondrial functions, including electron transport (Figures 7B and 7C). While the downstream consequences are unknown, a prior study linked aberrant mitochondrial membrane potential in pluripotent stem cells with elevated teratoma frequency (Schieke et al., 2008).

Taken together, our results suggest a model in which somatic cells activate lineage-specific lncRNAs to repress developmental genes, while also upregulating a common set of lncRNAs that facilitate metabolic reprogramming in a lineage-independent manner (Figure 7D). Our single-cell transcriptome analysis of induced pluripotency provides a resource to further examine the molecular mechanisms by which lncRNAs reprogram somatic cell state. In addition to lncRNAs, other classes of noncoding genes, such as pseudogenes, are also activated during reprogramming and have emerging roles in cancer development (Cooke et al., 2014; Han et al., 2014). Further studies of the noncoding transcriptome in various reprogramming contexts are likely to advance our basic understanding of cell state plasticity and cellular transformation.

EXPERIMENTAL PROCEDURES

iPS cell reprogramming

Tail-tip fibroblast (TTF) cultures were established from 3-8 day old reprogrammable mice homozygous for both the tet-inducible OSKM polycistronic cassette and the ROSA26-M2rtTA allele (Carey et al., 2010). Maintenance of animals and tail tip excision were performed according to a mouse protocol approved by the Caltech Institutional Animal Care and Use Committee (IACUC). TTFs (+ doxycycline), iPS cells, and ES cells were cultured in ES medium (DMEM, 15% FBS, sodium bicarbonate, HEPES, nonessential amino acids, penicillin-streptomycin, L-glutamine, b-mercaptoethanol, 1000 U/ml LIF) and grown on 6-well plates coated with 0.1% gelatin and irradiated MEF feeder cells (GlobalStem). For “2i” conditions, iPS cells were grown in ESGRO-2i medium (Millipore). For lncRNA loss-of-

function, iPS cells were transfected with siRNAs (IDT) using Lipofectamine RNAiMAX (Life). For SSEA-1 detection, StainAlive SSEA-1 DyLight 488 antibody (Stemgent) was used to detect SSEA-1 positive cells at specified time-points during reprogramming, which were isolated using flow cytometry on an iCyt Mission Technology Reflection Cell Sorter inside a Baker Bioguard III biosafety cabinet.

Single-cell and bulk sample cDNA synthesis and amplification

cDNA synthesis was performed using the Smart-Seq protocol as previously described (Ramskold et al., 2012). Briefly, the SMARTer Ultra Low RNA kit for Illumina sequencing (Clontech) was used to generate and amplify cDNA from single cells isolated using a micromanipulator or from bulk samples. Intact single cells were deposited directly into hypotonic lysis buffer. Poly(A)+ RNA was reverse transcribed through oligo dT priming to generate full-length cDNA, which was then amplified using 20-22 cycles. cDNA length distribution was assessed using High Sensitivity DNA kits on a Bioanalyzer (Agilent), and only samples showing a broad length distribution peak centered at 2kb were subsequently used for library generation.

Single-cell and bulk sample RNA-seq library generation and sequencing

Single-cell and bulk sample RNA-seq libraries were constructed using the Nextera DNA Sample Prep kit (Illumina). Briefly, cDNA was ‘tagmentated’ at 55° C with Nextera transposase, and tagmented DNA was purified using Agencourt AMPure XP beads (Beckman Coulter). Purified DNA was amplified using 5 cycles of Nextera PCR, and library quality was assessed using High Sensitivity DNA kits on a Bioanalyzer (Agilent). Libraries were sequenced on the Illumina HiSeq2000. Single-end reads of 50bp or 100bp length were obtained.

Read mapping and expression quantification

All reads were trimmed down to 50bp (if necessary) and mapped to the mouse genome (version mm9) with TopHat (Trapnell et al., 2009) (version 1.2.1) while supplying splice junctions annotated in the ENSEMBL63 set of transcript models. RPKMs for the ENSEMBL63 annotation were obtained using Cufflinks (Trapnell et al., 2010) (version 1.0.3) with otherwise default settings. For downstream analysis, the biotype classification of genes and transcripts in the ENSEMBL annotation was used to identify noncoding genes. Hierarchical clustering was carried out using Cluster 3.0 (de Hoon et al., 2004) and visualized using Java Treeview (Saldanha, 2004). For differential expression analysis, we aligned reads against the refSeq mouse transcriptome using Bowtie version 0.12.7 (Langmead et al., 2009). Expression levels were then estimated using eXpress (Roberts and Pachter, 2013) (version 1.3.0), with gene-level effective counts and RPKM values derived from the sum of the corresponding values for all isoforms of a gene. The effective count values were then used as input to DESeq (Anders and Huber, 2010) to assess differential expression. For ChIP-seq analysis, sequencing data were downloaded from accession numbers GSM307140, GSM623989, GSM307137, GSM307138, EMTAB-1600, GSM307155, and GSM623991. Reads were extracted using the fastq-dump program in the SRA ToolKit and mapped to the mm9 version of the mouse genome using Bowtie 0.12.7 with the following settings: “-v 2 -k 2 -m 1 -t --best -- strata”, i.e. retaining only unique

reads and allowing for up to 2 mismatches in a read. Enriched regions were called using ERANGE 3.2 (Johnson et al., 2007) with the following settings: “-minimum 2 -ratio 3 -shift learn -revbackground”.

Self-organizing maps

The 5000 genes with the greatest variance among the libraries were used for training a self-organizing map. Prior to SOM training, the data vectors were normalized on a gene-by-gene basis by subtracting each vector mean and dividing by its standard deviation. The SOM was constructed using the R package ‘kohonen.’ The total number of map units was set to the heuristic value 5*sqrt(N), where N is the number of data vectors. The map grid was initialized with the first two principal components of the data multiplied by a sinusoidal function to yield smooth toroidal boundary conditions. Training lasted 200 epochs (presentations of the data) during which the radius within which units were adapted toward the winning unit decreased linearly from h/8 to 2 units, where h is the map height (always chosen as the direction of largest length). Further analysis, including clustering and visualization, was performed with custom python code. Clusters were seeded by the local minima of the u-matrix, with a value for each unit defined as the average of the vector difference between that unit's prototype and its six neighbors on the hexagonal grid. All other unit prototypes were then assigned to clusters according to the minimum vector distance to a seed unit. The lists of clustered genes were submitted to the Princeton GO TermFinder (Boyle et al., 2004) server (http://go.princeton.edu) in order to determine enriched terms.

Single-molecule fluorescence in situ hybridization

smFISH was performed as previously described (Raj et al., 2008). Up to 48 DNA probes per target mRNA or lncRNA were synthesized and conjugated to Alexa fluorophore 488, 555, 594, or 647 (Life Technologies) and then purified by HPLC. Cells were trypsinized, fixed in 4% Formaldehyde, and permeabilized in 70% ethanol overnight. Cells were then hybridized with probe overnight at 30° C, in 20% Formamide, 2X SSC, 0.1g/ml Dextran Sulfate, 1mg/ml E.coli tRNA, 2mM Vanadyl ribonucleoside complex, 0.1% Tween 20 in nuclease free water. Samples were washed twice in 20% Formamide, 2X SSC, and Tween 20 at 30° C, and then twice in 2X SSC+0.1% Tween at RT. 1μl of hybridized cells was placed between #1 coverslips and flattened. Automated grid-based acquisition was performed on a Nikon Ti-E with Perfect Focus System, Semrock FISH filtersets, Lambda LS Xenona Arc Lamp, 60x 1.4NA oil objective, and Coolsnap HQ2 camera. Semi-automated dot detection and segmentation was performed using custom-built MATLAB software with a Laplacian-of-Gaussian Kernel, using Otsu's method to determine “dotness” threshold across all cells in the dataset.

Small RNA sequencing and analysis

Oxidation and beta-elimination of small RNAs were performed as previously described (Ameres et al., 2010). The Illumina-compatible NEBNext Small RNA Sample Prep Set 1 (New England Biolabs) was used to prepare small RNA libraries for sequencing on the Illumina platform. Sequencing adapters were removed from reads by finding the 3'-most complete match to the adapter sequence and trimming the read after that position. The

resulting were first mapped to the collection of ribosomal repeats (annotated using the RepeatMasker file downloaded from the UCSC genome browser), snoRNAs and snRNAs in the mouse genome (version mm9) using Bowtie version 0.12.7 in order to remove common contaminant reads. The unmapped reads from this filtering step were then aligned against the mm9 genome to determine the number of mappable reads. Both bowtie mapping steps were carried out with the following settings: '-v 0 -a -t --best --strata', i.e. no mismatches and an unlimited number of locations to which a read could map to. Enrichment of repeat classes in sequencing was estimated by calculating RPM (Reads Per Million mapped reads) scores for each individual repeat annotated in the UCSC RepeatMasker file, then summing over all the instances of each repeat class in order to derive a total repeat class RPM score.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank Ellen Rothenberg, Paul Sternberg, and Chuck Murry for helpful discussions, Jorge Mata and Scott Wang for tail tip excision, Josh Verceles, Diana Perez, and Rochelle Diamond for cell sorting, Igor Antoshechkin for sequencing, Henry Amrhein and Diane Trout for data curation, Shujun Luo for single-cell protocol development, and Maria Moon and Hank Huang for illustration and graphic design. D.H.K. was supported by a Fellowship Award from the Damon Runyon Cancer Research Foundation and the Beckman Fellows Program at the California Institute of Technology. M.B.E. is an Investigator of the Howard Hughes Medical Institute. This work was supported by the Bren Chair, Caltech Beckman Institute, and NIH NHGRI funding to B.J.W.

REFERENCES

Ameres SL, Horwich MD, Hung JH, Xu J, Ghildiyal M, Weng Z, Zamore PD. Target RNA-directed trimming and tailing of small silencing RNAs. Science. 2010; 328:1534–1539. [PubMed: 20558712]

Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11:R106. [PubMed: 20979621]

Apostolou E, Hochedlinger K. Chromatin dynamics during cellular reprogramming. Nature. 2013; 502:462–471. [PubMed: 24153299]

Bailey CL, Kelly P, Casey PJ. Activation of Rap1 promotes prostate cancer metastasis. Cancer research. 2009; 69:4962–4968. [PubMed: 19470770]

Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, Sherlock G. GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. Bioinformatics. 2004; 20:3710–3715. [PubMed: 15297299]

Buganim Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, Ganz K, Klemm SL, van Oudenaarden A, Jaenisch R. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell. 2012; 150:1209–1222. [PubMed: 22980981]

Carey BW, Markoulaki S, Beard C, Hanna J, Jaenisch R. Single-gene transgenic mouse strains for reprogramming adult somatic cells. Nat Methods. 2010; 7:56–59. [PubMed: 20010831]

Carey BW, Markoulaki S, Hanna JH, Faddah DA, Buganim Y, Kim J, Ganz K, Steine EJ, Cassady JP, Creyghton MP, et al. Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. Cell Stem Cell. 2011; 9:588–598. [PubMed: 22136932]

Chang G, Gao S, Hou X, Xu Z, Liu Y, Kang L, Tao Y, Liu W, Huang B, Kou X, et al. High-throughput sequencing reveals the disruption of methylation of imprinted gene in induced pluripotent stem cells. Cell Res. 2014; 24:293–306. [PubMed: 24381111]

Cooke SL, Shlien A, Marshall J, Pipinikas CP, Martincorena I, Tubio JM, Li Y, Menzies A, Mudie L, Ramakrishna M, et al. Processed pseudogenes acquired somatically during cancer development. Nature communications. 2014; 5:3644.

de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. Bioinformatics. 2004; 20:1453–1454. [PubMed: 14871861]

Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, et al. Ensembl 2014. Nucleic Acids Res. 2014; 42:D749–755.

Flynn RA, Chang HY. Long noncoding RNAs in cell-fate programming and reprogramming. Cell Stem Cell. 2014; 14:752–761. [PubMed: 24905165]

Goding CR, Pei D, Lu X. Cancer: pathological nuclear reprogramming? Nat Rev Cancer. 2014; 14:568–573. [PubMed: 25030952]

Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010; 464:1071–1076. [PubMed: 20393566]

Gurdon JB, Melton DA. Nuclear reprogramming in cells. Science. 2008; 322:1811–1815. [PubMed: 19095934]

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature. 2011; 477:295–300. [PubMed: 21874018]

Han L, Yuan Y, Zheng S, Yang Y, Li J, Edgerton ME, Diao L, Xu Y, Verhaak RG, Liang H. The Pan-Cancer analysis of pseudogene expression reveals biologically and clinically relevant tumour subtypes. Nature communications. 2014; 5:3963.

Hemberger M, Dean W, Reik W. Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. Nat Rev Mol Cell Biol. 2009; 10:526–537. [PubMed: 19603040]

Huang WC, Chan SH, Jang TH, Chang JW, Ko YC, Yen TC, Chiang SL, Chiang WF, Shieh TY, Liao CT, et al. miRNA-491-5p and GIT1 serve as modulators and biomarkers for oral squamous cell carcinoma invasion and metastasis. Cancer research. 2014; 74:751–764. [PubMed: 24335959]

Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. Science. 2007; 316:1497–1502. [PubMed: 17540862]

Karnoub AE, Weinberg RA. Ras oncogenes: split personalities. Nat Rev Mol Cell Biol. 2008; 9:517–531. [PubMed: 18568040]

Kim K, Doi A, Wen B, Ng K, Zhao R, Cahan P, Kim J, Aryee MJ, Ji H, Ehrlich LI, et al. Epigenetic memory in induced pluripotent stem cells. Nature. 2010; 467:285–290. [PubMed: 20644535]

Krizhanovsky V, Lowe SW. Stem cells: The promises and perils of p53. Nature. 2009; 460:1085–1086. [PubMed: 19713919]

Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. Briefings in bioinformatics. 2013; 14:144–161. [PubMed: 22908213]

Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25. [PubMed: 19261174]

Laugesen A, Helin K. Chromatin repressive complexes in stem cells, development, and cancer. Cell Stem Cell. 2014; 14:735–751. [PubMed: 24905164]

Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 2010; 11:204–220. [PubMed: 20142834]

Lee CY, Lin Y, Bratman SV, Feng W, Kuo AH, Scheeren FA, Engreitz JM, Varma S, West RB, Diehn M. Neuregulin autocrine signaling promotes self-renewal of breast tumor-initiating cells by triggering HER2/HER3 activation. Cancer research. 2014; 74:341–352. [PubMed: 24177178]

Lee JT. Epigenetic regulation by long noncoding RNAs. Science. 2012; 338:1435–1439. [PubMed: 23239728]

Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, Park IH, Garber M, Curran M, Onder T, Agarwal S, et al. Large intergenic non-coding RNA RoR modulates reprogramming of human induced pluripotent stem cells. Nat Genet. 2010; 42:1113–1117. [PubMed: 21057500]

Ma H, Morey R, O'Neil RC, He Y, Daughtry B, Schultz MD, Hariharan M, Nery JR, Castanon R, Sabatini K, et al. Abnormalities in human pluripotent cells due to reprogramming mechanisms. Nature. 2014; 511:177–183. [PubMed: 25008523]

Marchetto MC, Narvaiza I, Denli AM, Benner C, Lazzarini TA, Nathanson JL, Paquola AC, Desai KN, Herai RH, Weitzman MD, et al. Differential L1 regulation in pluripotent stem cells of humans and apes. Nature. 2013; 503:525–529. [PubMed: 24153179]

Moazed D. Small RNAs in transcriptional gene silencing and genome defence. Nature. 2009; 457:413–420. [PubMed: 19158787]

Nichols J, Smith A. Naive and primed pluripotent states. Cell Stem Cell. 2009; 4:487–492. [PubMed: 19497275]

Onder TT, Kara N, Cherry A, Sinha AU, Zhu N, Bernt KM, Cahan P, Marcarci BO, Unternaehrer J, Gupta PB, et al. Chromatin-modifying enzymes as modulators of reprogramming. Nature. 2012; 483:598–602. [PubMed: 22388813]

Orkin SH, Hochedlinger K. Chromatin connections to pluripotency and cellular reprogramming. Cell. 2011; 145:835–850. [PubMed: 21663790]

Papp B, Plath K. Epigenetics of reprogramming to induced pluripotency. Cell. 2013; 152:1324–1343. [PubMed: 23498940]

Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, Borkent M, Apostolou E, Alaei S, Cloutier J, et al. A Molecular Roadmap of Reprogramming Somatic Cells into iPS Cells. Cell. 2012; 151:1617–1632. [PubMed: 23260147]

Polo JM, Liu S, Figueroa ME, Kulalert W, Eminli S, Tan KY, Apostolou E, Stadtfeld M, Li Y, Shioda T, et al. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat Biotechnol. 2010; 28:848–855. [PubMed: 20644536]

Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods. 2008; 5:877–879. [PubMed: 18806792]

Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, Daniels GA, Khrebtukova I, Loring JF, Laurent LC, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol. 2012; 30:777–782. [PubMed: 22820318]

Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu Rev Biochem. 2012; 81:145–166. [PubMed: 22663078]

Roberts A, Pachter L. Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Methods. 2013; 10:71–73. [PubMed: 23160280]

Saldanha AJ. Java Treeview--extensible visualization of microarray data. Bioinformatics. 2004; 20:3246–3248. [PubMed: 15180930]

Schieke SM, Ma M, Cao L, McCoy JP Jr. Liu C, Hensel NF, Barrett AJ, Boehm M, Finkel T. Mitochondrial metabolism modulates differentiation and teratoma formation capacity in mouse embryonic stem cells. J Biol Chem. 2008; 283:28506–28512. [PubMed: 18713735]

Smith ZD, Nachman I, Regev A, Meissner A. Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. Nat Biotechnol. 2010; 28:521–526. [PubMed: 20436460]

Stadtfeld M, Apostolou E, Akutsu H, Fukuda A, Follett P, Natesan S, Kono T, Shioda T, Hochedlinger K. Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature. 2010; 465:175–181. [PubMed: 20418860]

Stadtfeld M, Apostolou E, Ferrari F, Choi J, Walsh RM, Chen T, Ooi SS, Kim SY, Bestor TH, Shioda T, et al. Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells. Nat Genet. 2012; 44:398–405. S391–392. [PubMed: 22387999]

Stuwe E, Toth KF, Aravin AA. Small but sturdy: small RNAs in cellular memory and epigenetics. Genes Dev. 2014; 28:423–431. [PubMed: 24589774]

Suva ML, Riggi N, Bernstein BE. Epigenetic reprogramming in cancer. Science. 2013; 339:1567–1570. [PubMed: 23539597]

Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006; 126:663–676. [PubMed: 16904174]

Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, et al. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature. 2008; 453:534–538. [PubMed: 18404147]

Theunissen TW, Jaenisch R. Molecular control of induced pluripotency. Cell Stem Cell. 2014; 14:720–734. [PubMed: 24905163]

Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009; 25:1105–1111. [PubMed: 19289445]

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010; 28:511–515. [PubMed: 20436464]

Trimarchi T, Bilal E, Ntziachristos P, Fabbri G, Dalla-Favera R, Tsirigos A, Aifantis I. Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. Cell. 2014; 158:593–606. [PubMed: 25083870]

Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, Ghosh D, Pienta KJ, Sewalt RG, Otte AP, et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature. 2002; 419:624–629. [PubMed: 12374981]

Watanabe T, Tomizawa S, Mitsuya K, Totoki Y, Yamamoto Y, Kuramochi-Miyagawa S, Iida N, Hoki Y, Murphy PJ, Toyoda A, et al. Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus. Science. 2011; 332:848–852. [PubMed: 21566194]

Yang D, Kedei N, Li L, Tao J, Velasquez JF, Michalowski AM, Toth BI, Marincsak R, Varga A, Biro T, et al. RasGRP3 contributes to formation and maintenance of the prostate cancer phenotype. Cancer research. 2010; 70:7905–7917. [PubMed: 20876802]

Yildirim E, Kirby JE, Brown DE, Mercier FE, Sadreyev RI, Scadden DT, Lee JT. Xist RNA is a potent suppressor of hematologic cancer in mice. Cell. 2013; 152:727–742. [PubMed: 23415223]

Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, Smith A. The ground state of embryonic stem cell self-renewal. Nature. 2008; 453:519–523. [PubMed: 18497825]

Zhang A, Zhou N, Huang J, Liu Q, Fukuda K, Ma D, Lu Z, Bai C, Watabe K, Mo YY. The human long non-coding RNA-RoR is a p53 repressor in response to DNA damage. Cell Res. 2013; 23:340–350. [PubMed: 23208419]

Zhang J, Nuebel E, Daley GQ, Koehler CM, Teitell MA. Metabolic regulation in pluripotent stem cells during reprogramming and self-renewal. Cell Stem Cell. 2012; 11:589–595. [PubMed: 23122286]

Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M, Lee JT. Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell. 2010; 40:939–953. [PubMed: 21172659]

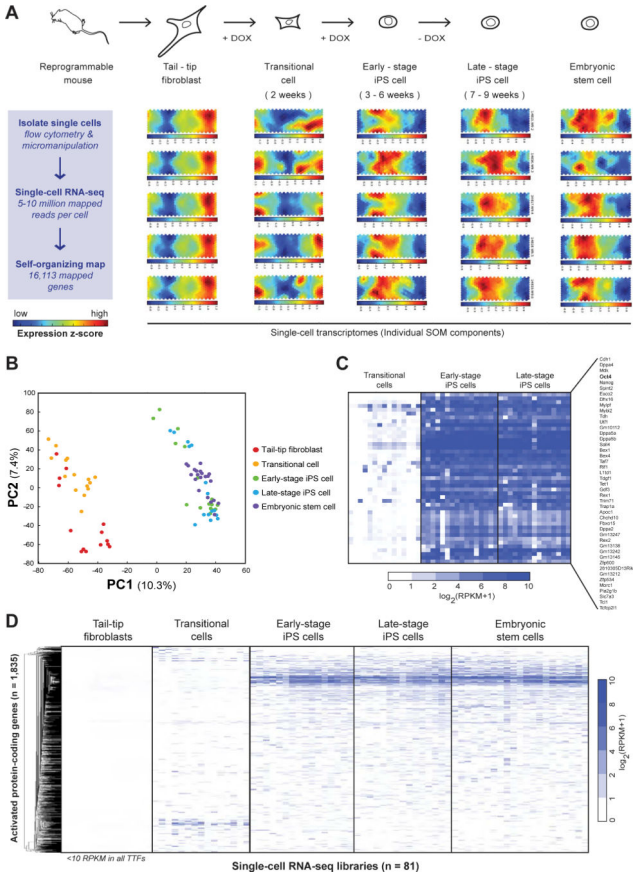

Figure 1. Single-cell transcriptome analysis of cellular reprogramming

(A) Schematic illustration of reprogramming experiments and single-cell RNA-seq analysis using the self-organizing map (SOM). Five representative SOM components are shown for each cell type. (B) Principal component (PC) analysis of 81 single-cell transcriptomes using all genes expressed at >1 RPKM. (C) Heatmap of protein-coding genes that correlate most highly (Pearson correlation) with *Oct4* expression during the reprogramming time course. (D) Hierarchical clustering of activated protein-coding genes during the reprogramming time course using genes expressed at >10 RPKM in non-TTFs and <10 RPKM in all TTFs. See also Figure S1.

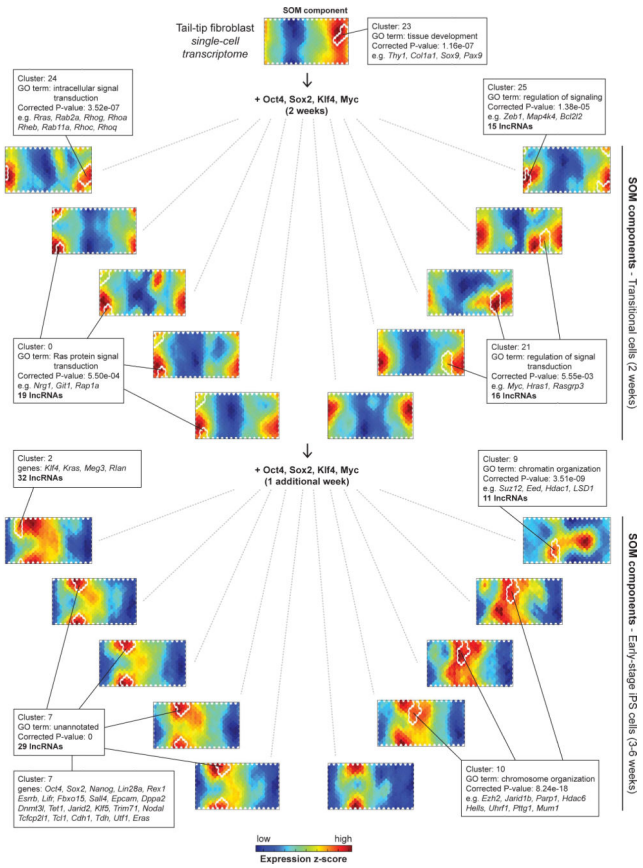

Figure 2. Analysis of transcriptome dynamics using the self-organizing map

Single-cell transcriptomes depicted as individual components of the self-organizing map (SOM). Boxes represent individual cells at defined stages of reprogramming, with clusters outlined in white. Significantly enriched gene ontology (GO) terms for indicated clusters are shown (Bonferroni-corrected P-values). See also Figure S2 and Table S1.

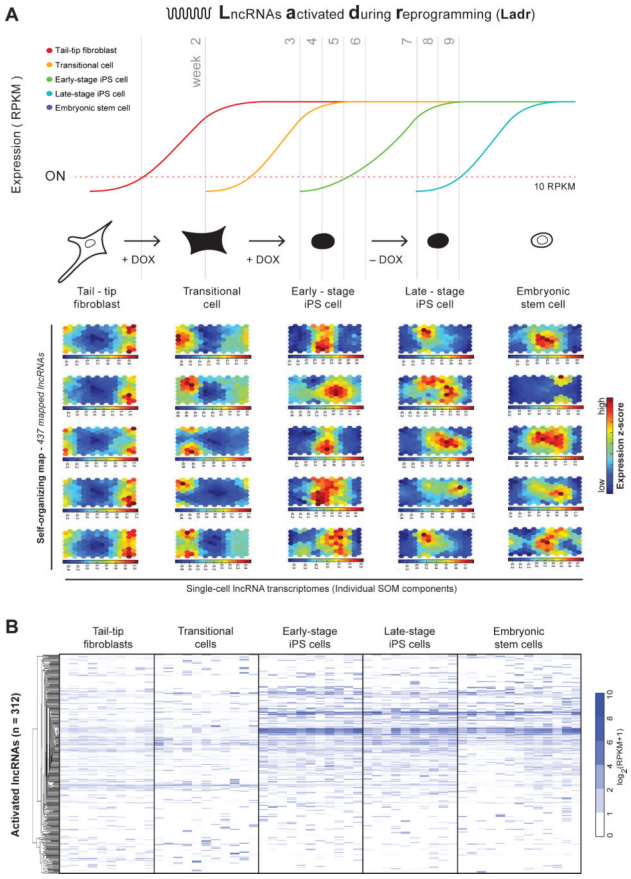

Figure 3. LncRNAs activated during reprogramming

(A) Schematic illustration of lncRNAs activated during defined stages of reprogramming (>10 RPKM) and lncRNA transcriptome analysis using the self-organizing map (SOM). Five representative SOM components are shown for each cell type. (D) Hierarchical clustering of activated lncRNA genes during the reprogramming time course using genes expressed at >10 RPKM in non-TTFs and <10 RPKM in all TTFs. See also Table S2.

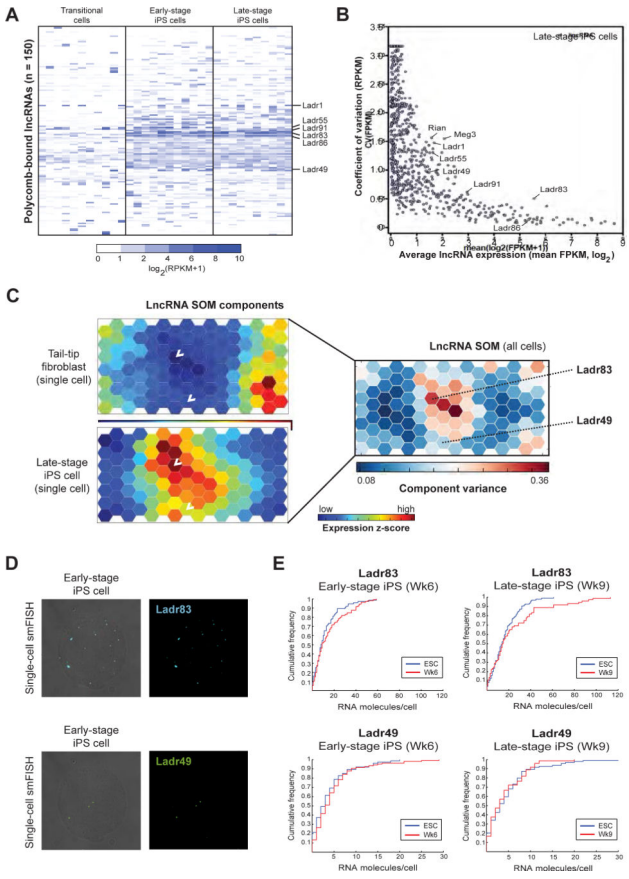

Figure 4. Activated lncRNA expression variability in iPS cells

(A) Heatmap of activated lncRNAs previously shown to physically associate with Polycomb repressive complex 2 in pluripotent stem cells. (B) Plot showing the relationship between average lncRNA expression and coefficient of variation in single late-stage iPS cells . (C) Individual cell components of the lncRNA SOM (left) and component variance for each hexagonal unit of the lncRNA SOM (right). (D) Representative phase contrast (merged with fluorescence) and fluorescence images of a single early-stage iPS cell using smFISH. Scale bar, 10 um. (E) Cumulative distribution function plots of lncRNA molecules per cell, as determined by smFISH. See also Figure S3 and Table S3.

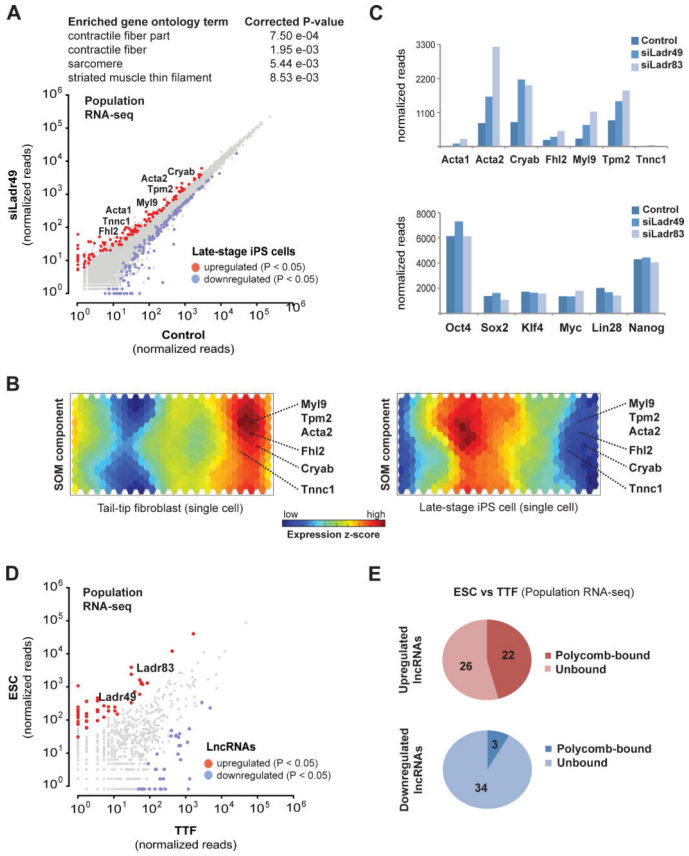

Figure 5. Repression of developmental genes by Ladr49 and Ladr83

(A) Differential expression analysis of significantly upregulated or downregulated genes in late-stage iPS cells deficient for Ladr49, as determined by population level RNA-seq, and gene ontology (GO) analysis for significantly enriched GO terms in upregulated genes. Bonferroni-corrected P-values are shown. (B) Individual cell SOM components of indicated cell types with muscle-related genes labeled. (C) Effective read counts (normalized reads) for muscle-related (top) and reprogramming/pluripotency (bottom) genes in late-stage iPS cells, as determined by population level RNA-seq. (D) Differential expression analysis of significantly upregulated or downregulated lncRNA genes in ES cells versus TTF cells, as determined by population level RNA-seq. (E) Fraction of upregulated or downregulated lncRNAs that are physically associate with Polycomb repressive complex 2 in ES cells. See also Figure S4.

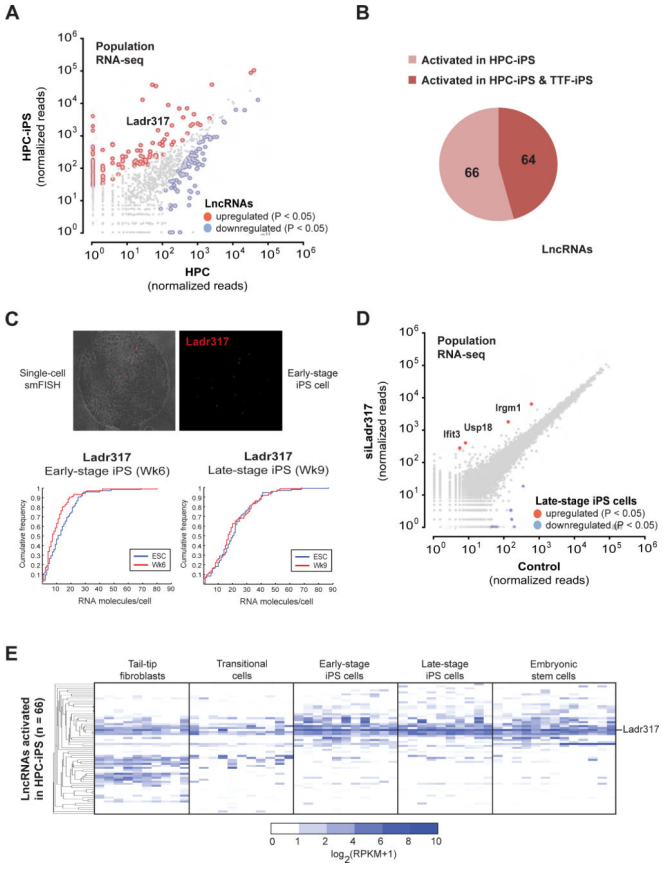

Figure 6. Lineage-specific role for Ladr317 in hematopoietic reprogramming

(A) Differential expression analysis of significantly upregulated or downregulated lncRNA genes in HPC versus HPC-iPS cells, as determined by population level RNA-seq. (B) Fraction of lncRNAs that are upregulated in HPC-iPS or both HPC-iPS and TTF-iPS cells. (C) Representative phase contrast (merged with fluorescence) and fluorescence images of a single early-stage iPS cell using smFISH, and cumulative distribution function plots of lncRNA molecules per cell, as determined by smFISH. Scale bar, 10 um. (D) Differential expression analysis of significantly upregulated or downregulated genes in late-stage iPS cells deficient for Ladr317, as determined by population level RNA-seq. (E) Hierarchical clustering of lncRNAs activated in HPC-iPS cells, as determined by single-cell RNA-seq. RPKM, reads per kilobase per million mapped reads.

Figure 7

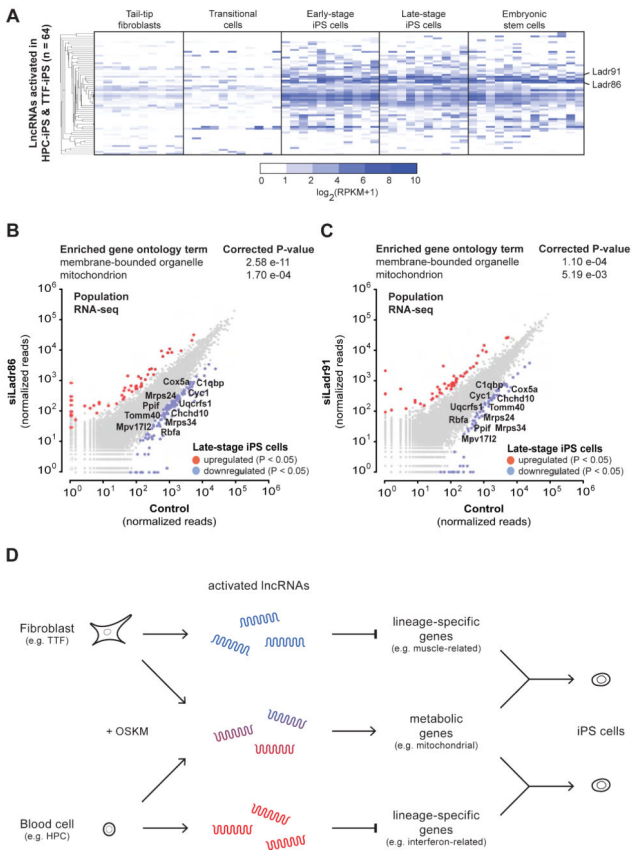

Figure 7. Regulation of metabolic genes by Ladr86 and Ladr91

(A) Hierarchical clustering of lncRNAs activated in both HPC-iPS and TTF-iPS cells, as determined by single-cell RNA-seq. RPKM, reads per kilobase per million mapped reads. Differential expression analysis of significantly upregulated or downregulated genes in late-stage iPS cells deficient for Ladr86 (B) or Ladr91 (C), as determined by population level RNA-seq, and gene ontology (GO) analysis for significantly enriched GO terms in downregulated genes. Bonferroni-corrected P-values are shown. (D) Model showing lineage-specific and lineage-independent functions of activated lncRNAs during fibroblast and blood cell reprogramming. See also Figures S5-S7 and Tables S4-S5.