

Mechanisms of transcription factor-mediated direct reprogramming of mouse embryonic stem cells to trophoblast stem-like cells

Catherine Rhee^{1,2}, Bum-Kyu Lee^{1,2}, Samuel Beck^{1,2}, Lucy LeBlanc^{1,2}, Haley O. Tucker^{1,2} and Jonghwan Kim^{1,2,3,*}

¹Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX 78712, USA, ²Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX 78712, USA and ³Center for Systems and Synthetic Biology, The University of Texas at Austin, Austin, TX 78712, USA

Received April 14, 2017; Revised July 21, 2017; Editorial Decision July 25, 2017; Accepted August 02, 2017

ABSTRACT

Direct reprogramming can be achieved by forced expression of master transcription factors. Yet how such factors mediate repression of initial cell-type-specific genes while activating target cell-type-specific genes is unclear. Through embryonic stem (ES) to trophoblast stem (TS)-like cell reprogramming by introducing individual TS cell-specific ‘CAG’ factors (Cdx2, Arid3a and Gata3), we interrogate their chromosomal target occupancies, modulation of global transcription and chromatin accessibility at the initial stage of reprogramming. From the studies, we uncover a sequential, two-step mechanism of cellular reprogramming in which repression of pre-existing ES cell-associated gene expression program is followed by activation of TS cell-specific genes by CAG factors. Therefore, we reveal that CAG factors function as both decommission and pioneer factors during ES to TS-like cell fate conversion.

INTRODUCTION

Advances in transcription factor (TF)-mediated direct reprogramming have revealed the plasticity of cell identity and the feasibility of cell fate conversion both *in vitro* and *in vivo* (1–4). To successfully convert cell fate, global gene expression programs of the original cells must be altered to a state favorable to a reprogrammed target cell type. However, little is known about the mechanisms of how and to what extent altered expression of TFs modulates changes in global gene expression and cellular characteristics.

Induced pluripotent stem (iPS) cells can be generated from overexpression (OE) of a handful of TFs (e.g. Oct4, Sox2, Klf4 and Myc) in fibroblasts. Reprogramming to iPS cells can be broadly divided into two phases: a long stochastic phase followed by a shorter deterministic phase (5). Re-

cent reports suggested that a gene activation during reprogramming is modulated by which ectopically expressed TFs acting as ‘pioneer’ factors, which initially bind to closed chromatin of genes specific to the target cell type (6). Once bound, pioneer factors interact with various chromatin modifiers to convert closed chromatin in open, thereby activating target cell-specific genes. Oct4, Sox2 and Klf4 are known to function as pioneer factors early in somatic cell reprogramming process (7) as Ascl1, a TF capable of converting fibroblasts to induced neuronal (iN) cells (8). Although activated target cell-specific genes can indirectly affect the suppression of active genes in the initial cells, precise gene repression mechanisms during cellular reprogramming has not been explicitly addressed and it is still ambiguous whether activation and repression of cell type-specific genes occur simultaneously or sequentially.

Several trophoblast-specific TFs, including Arid3a, Cdx2, Gata3, Elf5, Eomes, Id2, Tead4 and Tfap2c, play essential roles in trophoblast (TE) development or trophoblast stem (TS) cell identity and self-renewal (9–12). It was previously shown that induction of a single TF, such as Tfap2c, Cdx2, Gata3 or Arid3a, (13–16) is sufficient to reprogram embryonic stem (ES) cells to TS-like cells, and the resultant altered morphology, functional properties and global gene expression profiles are highly similar to genuine multipotent TS cells (13–16). In particular, TS-like cells generated by OE of Cdx2 and Arid3a were successfully incorporated into the TE of developing embryos and contributed to placental lineages *ex vivo* (16,17), revealing the feasibility of generating functional TS-like cells from ES cells. Thus, we reasoned that such an approach would allow us to thoroughly interrogate mechanisms of transcriptional and epigenetic regulation by OE of TFs during cell fate conversion.

Here, we employed an ES to TS-like cell reprogramming system via OE of three key TE/TS cell-specific TFs—Cdx2, Arid3a and Gata3 (herein referred to as CAG factors) that

*To whom correspondence should be addressed. Tel: +1 512 232 8046; Fax: +1 512 232 8045; Email: jonghwankim@mail.utexas.edu

are well-known for being instrumental in trophoblast differentiation and placental development (13,14,16,17). We investigated the dynamics of CAG factor binding as well as subsequent effects on chromatin accessibility and global gene expression during the early phase of reprogramming. We found that CAG factors orchestrate reprogramming of ES cells to TS-like cells via a two-step mechanism; Repression of ES cell-specific genes through decommissioning of active enhancers in ES cells followed by activation of TS cell-specific genes through the pioneer factor activity.

MATERIALS AND METHODS

Cell culture

Mouse J1 ES cells were cultured in ES+ media, composed of DMEM (Dulbecco's modified Eagle's medium) supplemented with 18% fetal bovine serum (FBS), 2mM L-glutamine, 100 μ M of non-essential amino acid supplement, nucleoside mix (100 \times stock, Sigma), 100 μ M of β -mercaptoethanol, 1000U/ml of recombinant leukemia inhibitory factor (LIF, Chemicon) and 50U/ml of penicillin/streptomycin. ES cells were plated on 0.1% gelatin coated dishes. Mouse TS cells were maintained in TS+ media, at a ratio of 3:7 of TS medium to mouse embryonic fibroblasts (MEF)-conditioned TS medium, supplemented with 25 ng/ml Fgf4 and 1 μ g/ml heparin. TS medium is RPMI 1640 (Roswell Park Memorial Institute medium, Gibco) supplemented with 20% FBS, 100 μ M β -mercaptoethanol, 2 mM L-glutamine, 1 mM sodium pyruvate, penicillin (50U/ml) and streptomycin (50 mg/ml). MEF-conditioned medium is TS medium conditioned by MEF. MEF were treated with mitomycin, followed by culturing for 3 days. The medium was collected every 3 days for three times. 293T cells were maintained in DMEM supplemented with 10% FBS, 2 mM L-glutamine and 50U/ml of penicillin/streptomycin. All cells were incubated at 37°C, 5% CO₂.

Generation of stable cell lines

Individual CAG genes were cloned into the pEF1 α -FLBIO vector and the vector was transfected into ES cells expressing BirA by electroporation as previously described (18). For \sim 10 days after transfection, cells were selected in ES media supplemented with puromycin (Invitrogen) and geneticin (Gibco). After picking single colonies, cells were expanded for additional days to reach proper cell numbers to perform further analyses. In the meantime, positive OE clones were confirmed by reverse transcriptase-quantitative polymerase chain reaction (RT-qPCR) and western blotting. We defined passage 0 as the time point that we confirmed the stable OE of CAG factors by these two techniques. Primer sequences for PCR are listed in Supplementary Table S1.

Generation of inducible cell lines

Coding sequence of Arid3a amplified by PCR using mouse cDNA clone (Origene, MC205205) as a template, and then cloned into the pLVX-TRE3G-ZsGreen1 vector (Clontech, 631361). Clones were sequence-verified prior to use.

Lentiviral production and infection were performed as described below. ES cells were plated at \sim 1 \times 10⁶ cells per one well of 6-well plate with virus-containing supernatant (co-infection of pLVX-TRE3G-ZsGreen1 and pLVX-EF1 α -Tet3G). The cells were placed under neomycin and geneticin selection for 2 days followed by replacement of medium with fresh ES cell medium plus 1 μ g/ml doxycycline (dox). ES cell medium containing dox was changed each day for \sim 6 days. Ectopic expression levels were measured by RT-qPCR. Primer sequences for PCR are listed in Supplementary Table S1.

Lentiviral production and infection

Lentiviral production was performed in 293T cells. 293T cells were plated at \sim 6 \times 10⁶ cells per 100 mm dish and incubated overnight. Cells were transfected with 6 μ g of pLVX-TRE3G-ZsGreen1 or pLVX-EF1 α -Tet3G vectors (for the inducible cell line, Clontech) with 4 μ g of pCMV- Δ 8.9 and 2 μ g of VSVG helper plasmids using Fugene (Promega), according to the manufacturer's instructions. After 15 h, 293T medium was replaced with ES medium. The supernatants containing viral particles were harvested after 48 h of transfection, filtered through 0.45 μ m pore-size cellulose acetate filters and supplemented with polybrene (Millipore). Infections were performed with cells plated at a density of \sim 1 \times 10⁶ cells for one well of 6-well plate with virus-containing supernatant supplemented with polybrene (Millipore).

Quantitative gene expression analysis

For RT-qPCR, total RNA was isolated using the RNeasy Plus Mini Kit (Qiagen). A total of 500 ng RNA was then used to synthesize cDNA with qScript cDNA supermix (Quanta). RT-qPCRs were performed using PerfeCTa SYBR Green FastMix (Quanta) with 2 μ l of 20 \times diluted cDNA. Using Primer 3 (19), RT-qPCR primers were designed to amplify an \sim 100 bp region containing the junction between two exons. For chromatin immunoprecipitation-qPCR (ChIP-qPCR), the primers were designed to amplify \sim 100 bp regions centered on the putative binding sites. The Ct values obtained from qPCR were normalized against *Gapdh* for gene expression and against proximal promoter of *Gfilb* to determine relative gene expression and ChIP enrichment, respectively. Primer sequences for qPCR are listed in Supplementary Table S1.

Western blotting

Proteins were lysed from cultured cells using RIPA Lysis and Extraction Buffer (G-Biosciences) with protease inhibitor cocktail (Roche), followed by incubation at 100°C for 20 min. Cell lysates were separated by electrophoresis on 4–20% gradient polyacrylamide gels and transferred to PVDF membranes. The blots were blocked with TBS-T (20 mM Tris-HCl, pH 7.6, 136 mM NaCl and 0.1% Tween-20) containing 5% BSA (bovine serum albumin) for 1 h and then incubated with primary antibody solution at 4°C overnight. After washing with TBS-T, the membranes were incubated with HRP-conjugated secondary antibodies for 1 h at room temperature (RT). Then the membrane

was washed with TBS-T, followed by developing with ECL reagents (GE Healthcare). Antibodies used for western blotting were anti-Cdx2 (1:1000, ab88129, Abcam), anti-Arid3a (1:5000; (20)), anti-Gata3 (1:1000, SC-9009, Santa Cruz), anti- β -actin (1:10000, ab20272, Abcam) and anti-Hdac1 (1:1000, ab7028, Abcam).

Co-immunoprecipitation

One-step affinity purification with streptavidin-agarose beads (Invitrogen) using $\sim 2 \times 10^7$ cells from ES cell lines expressing BirA only or BirA plus biotin-tagged proteins was performed as previously described (21). The final pulled-down proteins were eluted in Laemmli buffer prior to western blotting.

RNA-sequencing analysis

Libraries were prepared with an RNA library preparation kit (E7490, NEB) using 1 μ g of RNA obtained by the RNeasy Mini Kit (Qiagen). RNA-seq libraries were sequenced with a 1 \times 75-bp strand-specific protocol on a NextSeq 500 (Illumina). Data were analyzed using a high-throughput next-generation sequencing analysis pipeline: FASTQ files were aligned to the mouse genome (mm9, NCBI Build 37) using TopHat2 (22). Gene expression profile for the individual samples was calculated with Cufflinks (23) as RPKM values.

ChIP-sequencing

ChIP assays were performed with ES cell lines expressing BirA only (reference) or both BirA and biotin-tagged proteins (samples) as previously described (24) using streptavidin magnetic particles (Roche). Additional ChIP assays were carried out in ES, OEAr3a.ES+, OEAr3a.TS+ and TS cells using Arid3a (20), Hdac1 (ab7028, Abcam) and H3K27ac (ab4729–100, Abcam) antibodies. ChIP-seq libraries were generated using ChIP-seq library prep kits (NEB) according to the manufacturer's instructions. ChIP-seq libraries were sequenced using Illumina NextSeq 500 at the Genomic Sequencing and Analysis Facility (GSAF), The University of Texas at Austin.

ATAC-sequencing

ATAC assays were performed as previously described (25). We used $\sim 50,000$ cells to start the experiment. The cells were incubated with transposition reaction mix for 30 min followed by PCR reaction for 18 cycles. ATAC samples in the ~ 250 bp range were isolated using E-gel size select chromatography. ATAC-seq was performed using Illumina NextSeq 500.

Identification of ChIP- and ATAC-sequencing peaks

FASTQ files were aligned to the mouse genome (mm9, NCBI Build 37) using Bowtie 2.2.5. For the identification of peaks, model-based analysis with the ChIP-seq (MACS 2.1.1) peak caller (26) was used with a default setting (macs2 callpeak -t Sample.sam -c Control.sam -f SAM -g mm -n)

with keeping five duplicates. The ChIP-seq signals resulting from BirA only expressing ES cells and Mock samples were subtracted to remove background noise for ChIP by incubation with streptavidin magnetic particles and native antibodies, respectively. We readily identified several thousands of significant individual CAG occupied loci upon filtering high-quality peaks based on *P*-values. Genes occupied by TFs within a 10 Kb window (8 Kb upstream to 2 Kb downstream of transcriptional start site (TSS)) were considered as target genes and used for subsequent analyses. In order to summarize the peak signal enrichment over control experiments, *z*-score bedGraph files were produced, and following background subtraction, the bedGraph files were constructed using MACS2 version 2.1.1 with 'bdgcmp -m logLR' command. Peaks were visualized using the Integrative Genome Viewer from the Broad Institute.

Motif analysis

Both *de novo* and known motifs in peak regions were identified using Homer (27) with the fragment size for motif finding set as 100 bp.

Heat maps

All heat maps were generated using JavaTreeview (28).

Gene ontology (GO) analysis

DAVID 6.7 (29) was used for analysis of differentially expressed genes obtained from the expression profile data. TF targets and Genomic Regions Enrichment of Annotation Tool (30) were used for ChIP-seq peak analysis.

Correlation maps

Common binding sites of the indicated TFs were identified by peak calling with an overlap analysis. The significance overlaps of binding sites of two TFs were calculated for with a paired-end Pearson correlation coefficient.

Principle component analysis (PCA)

The R prcomp was used to perform principle component analysis (PCA) using the RNA-seq data (for Figure 1E).

Dendrogram

R hcluster was used to generate dendrograms of RNA- and ATAC-seq data.

Public datasets used

Published ChIP-seq datasets obtained from the Gene Expression Omnibus (GEO) database under the accession numbers GSE31039 (H3K27ac, ES cells), GSE42207 (H3K27ac, TS cells) and GSE11724 (Oct4, ES cells) were used for the analysis shown in Figures 3 and 4.

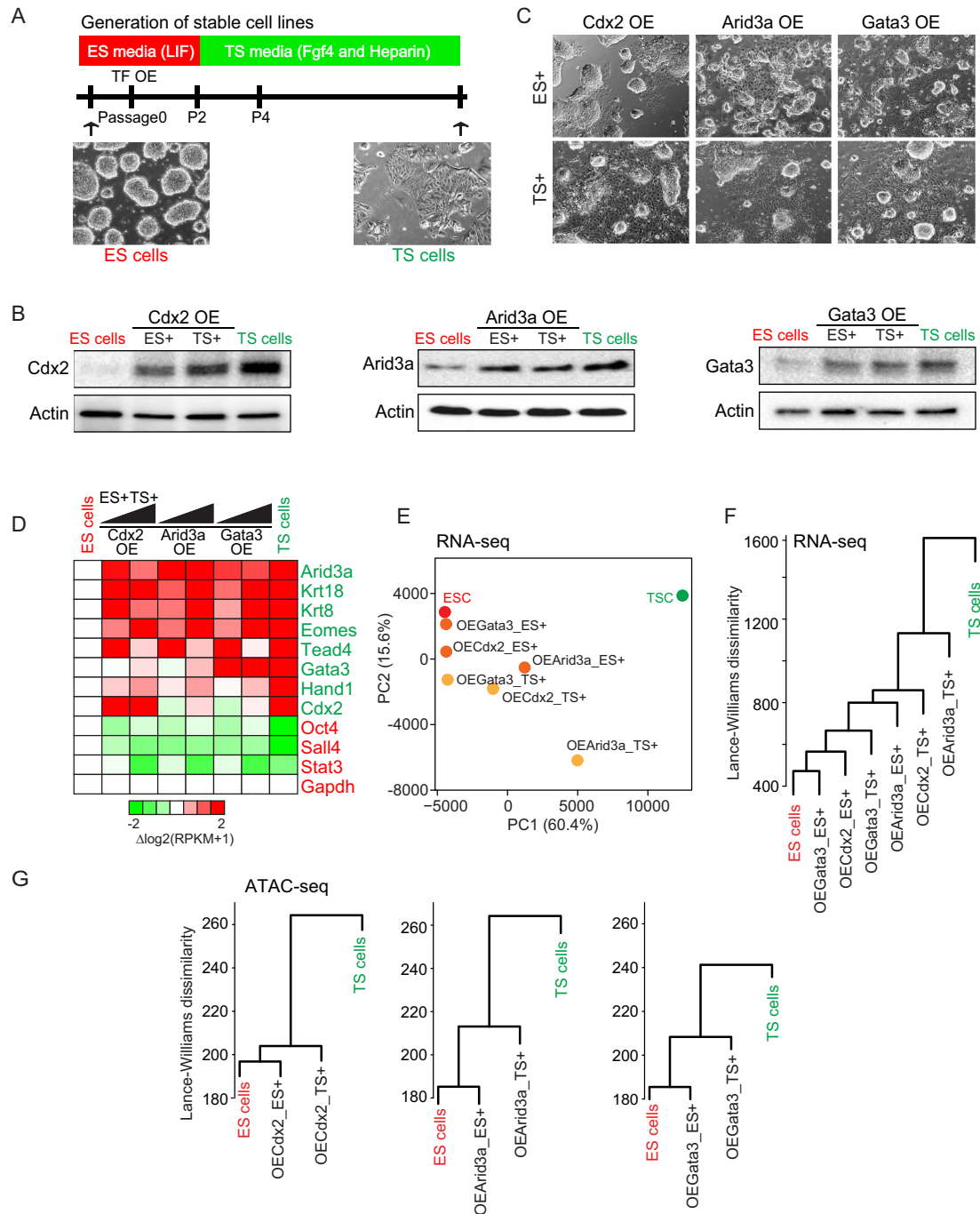


Figure 1. Overexpression of CAG factors in ES cells promotes changes in global gene expression and chromatin landscape toward TS cells. See also Supplementary Figure S1. (A) Schematic diagram of the experimental design employed to generate TS-like cells from ES cells. Constitutive expression was used to produce stable cell lines. At passage 2 (day 4 after the validation of OE clones), ES media was replaced with TS media. Cell morphology of ES and TS cells are shown at the bottom. (B) Western blotting showing protein levels of the CAG factors in ES cells and TS cells at different stages (ES+ and TS+) of the reprogramming upon OE of CAG factors. β -Actin was used as a loading control. (C) Bright field images depicting changes in cell morphologies associated with conversion of ES to TS-like cells at different stages of the reprogramming. (D) A heatmap showing relative gene expression levels of ES- and TS cell-specific genes in CAG OE cells relative to ES cells ($\Delta\log_2(\text{RPKM}+1)$). ES- and TS cell-specific genes are labeled in red and green to the right side of the heatmap, respectively. (E) PCA of time-course RNA-seq data showing gradual transition of the transcriptome from ES toward TS-like cells. (F and G) Dendrograms showing a similarity of expression profile (F) and chromatin landscapes (G) among ES, TS and CAG-mediated reprogrammed cells.

RESULTS

Ectopic expression of individual CAG factors in ES cells promotes changes in global gene expression and chromatin landscape toward TS cells

OE of either of the aforementioned TE/TS cell-specific CAG factors previously was shown to convert mouse ES cells to TS-like cells (13,14,16). To determine how these CAG factors initiate reprogramming, we designed our experimental set-up as shown in Figure 1A. Individual CAG factors were expressed under the control of constitutively active pEF1 α promoters. After establishing OE clones, the cells were maintained for 4 days (two passages) in ES cell culture media (ES+) and the media was replaced with media for TS cell culture (TS+) to optimize conditions for reprogramming (detailed procedures in ‘Materials and Methods’ section). During the ES to TS-like cell fate conversion, we monitored changes in cellular morphology, global gene expression, chromatin accessibility and CAG factor occupancy at passage 2 (day 4; ES+) and passage 4 (day 8; TS+).

We first confirmed OE of individual CAG factors in ES cells by western blotting (Figure 1B and C). Consistent with previous reports (13,14,16), OE of CAG factors in ES cells promoted dramatic morphological changes. Cells transitioned from spheroid undifferentiated colony morphology to a flattened, epithelial-like morphology, indicative of an early stage of cellular conversion from ES to TS-like cells; this occurred even before replacing ES+ media with TS+ media (Figure 1C, upper panels). Under TS+ media conditions, cells adopted even more prominent TS-like cell morphology (Figure 1C, lower panels). Parallel to these morphological changes, TE/TS cell-specific marker genes also were upregulated over time (Supplementary Figure S1), indicating that our experimental system was valid for investigating the early stage cell fate conversion of ES to TS-like cells.

To determine to what degree CAG factors affect global gene expression and chromatin configuration during the early stage of reprogramming (ES+ and TS+), we profiled global gene expression by RNA-seq and chromatin accessibility by ATAC-seq. Indeed, even at this early stage of reprogramming, global mRNA expression had begun to transition from an ES cell state to a TS-like cell state. As shown in Figure 1D, we observed reduced levels of key ES cell-specific markers (including Oct4, Sall4, and Stat3) and induced levels of TE/TS cell-specific markers (including Eomes, Krt8, and Tead4) in CAG factor OE cells as compared to untreated ES cells. Principal component analysis (PCA) of expression profiles also showed a significant shift in the global expression patterns of ES cells toward TS cells (Figure 1E). Notable among the CAG factors, Arid3a induced an expression pattern most similar to that of bona fide TS cells (Figure 1F). Consistent with the global gene expression data, OE of individual CAG factors induced alterations in ES cell chromatin accessibility similar to that of TS cells (Figure 1G). Although the degrees of global changes in chromatin accessibility were minor compared to changes in expression levels, still, these changes were associated with crucial markers of successful ES to TS-like cell fate conversion, such as Oct4, Nanog, Cdx2, Gcm1, Eomes, Hand1 and Tead4. Col-

lectively, these results indicate that individual expression of CAG factors initiates ES to TS-like cellular conversion by not only modulating global gene expression, but also by reshaping the chromatin landscape.

CAG factors directly regulate both ES and TS cell-specific genes during ES to TS-like cell reprogramming

To examine how OE of CAG factors mediates global transcriptional changes, we determined which classes of genes were directly regulated by CAG factors via ChIP-seq. We found several thousands of statistically significant individual CAG factor binding sites (detailed in ‘Materials and Methods’ section). As previous studies suggesting that cellular identity is largely determined by distal regulatory elements, especially enhancers, that are targets of cell type-specific TFs (31), we found that >60% of CAG target sites are distal regions to their respective TSSs, primarily distributed within intergenic (~40%) and upstream (~20%) regions during the early stage (day 4, passage 2) of reprogramming (Supplementary Figure S2A).

We next performed gene ontology (GO) analysis of the genes associated with the targets of each CAG factor. As shown in Figure 2A, CAG factor targets were largely associated with both ES cell-specific terms (e.g. stem cell maintenance and blastocyst development) and TS cell-specific terms (e.g. embryonic placenta development and trophectodermal cell differentiation). The results indicate that, individual CAG factors act as distal enhancer binding proteins and directly regulate both ES and TS cell-specific genes during the early cell fate conversion. Combined analysis of ChIP-seq and RNA-seq data revealed that both activated and repressed genes upon OE of each TF are largely direct targets of each factor (Figure 2B and Supplementary Figure S2B), indicating that each CAG factor displays dual functions during the reprogramming.

To further characterize the relationship between differentially expressed genes upon OE of each factor and its occupancy, we analyzed the ChIP-seq data by applying moving window averages (window size: 250; bin size: 1; Figure 2C and Supplementary Figure S2C). We found that peak intensities, which indicate occupancy strength of the factor to DNA, are generally correlated with greater changes in gene expression (Figure 2C and Supplementary Figure S2C). Then we examined the target occupancy of CAG factors on the genes sorted by their relative expression (TS cells over ES cells) to determine which classes of genes are directly regulated by CAG factors (Figure 2D and Supplementary Figure S2D). The results revealed that the genes highly expressed in ES cells are, in fact, strong direct targets of CAG factors, whereas TS cell-specific TFs are less strongly occupied by the CAG factors (Figure 2D and E; Supplementary Figure S2D). Overall, our analysis revealed significant repression of ES cell-specific genes along with a slight induction of TS cell-specific genes at the early stage of programming, which is mediated by direct binding of CAG factors (Figure 2F and G; Supplementary Figure S2D and E).

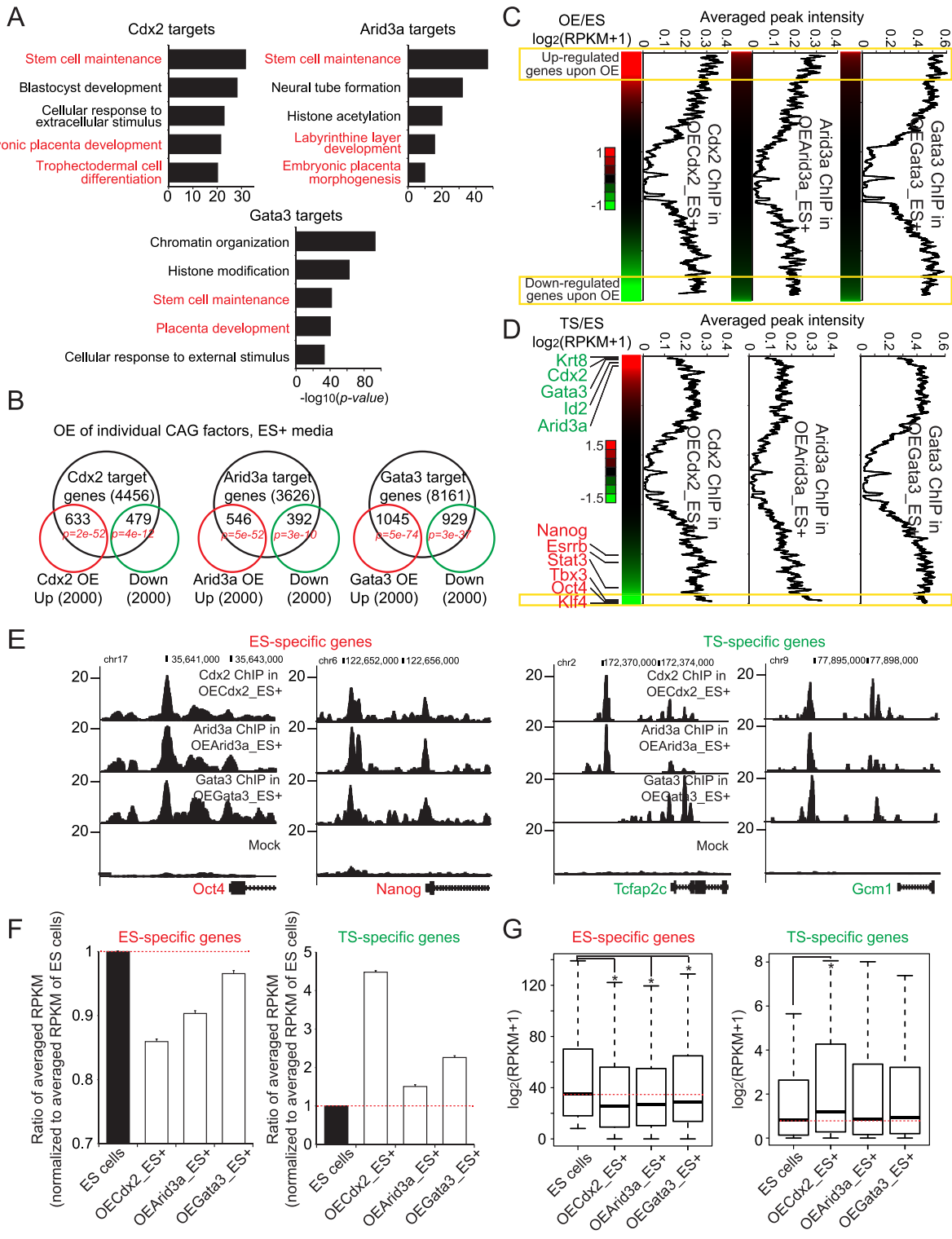


Figure 2. CAG factors exert dual functions during reprogramming. See also Supplementary Figure S2. (A) Bar graphs showing enriched GO terms of individual CAG factor targets. ES- and TS cell-specific functions are highlighted in red. (B) Venn diagrams depicting overlaps among genes bound by each CAG factor and genes that are either up- or downregulated in ES cells upon OE of individual CAG factors. *P*-values were calculated using hypergeometric tests. (C and D) Heatmaps ranking relative gene expression of CAG OE cells relative to ES cells (C) and TS cells relative to ES cells (D). Averaged occupancy scores calculated by moving average (window size, 250; bin size, 1) for individual CAG factors are plotted in black lines. Yellow boxes indicate 2000 genes from panel B (C) and 150 genes from panel F (D). (E) Signal track images depicting occupancy of each CAG factor at both ES- and TS cell-specific genes upon OE in ES cells. (F) Bar graphs showing average expression levels of the top 150 ES cell-specific genes and the top 150 TS cell-specific genes upon OE of each CAG factor in ES cells. Error bars indicate standard deviation between 150 genes. (G) Boxplots showing the distribution of expression levels of ES- and TS cell-specific genes upon OE of individual CAG factors in ES cells. The red dotted lines indicate median expression of ES- or TS cell-specific genes in ES cells. All data presented in the figure were obtained under ES+ culture condition (4 days under ES+ condition).

Each CAG factor binds both open and closed chromatin during the early stage of reprogramming

Previous studies introduced the ‘pioneer factor’ concept as factors capable of binding directly to heterochromatin to activate silenced target cell-specific genes during reprogramming (7,32). In order to investigate whether CAG factors function as pioneer factors and to what degree CAG factors bind to open and closed chromatin, we examined chromatin openness of individual CAG factor binding sites during the reprogramming process. Interestingly, we found that individual CAG factors prefer to bind open chromatin while they also occupy closed chromatin (Figure 3A–C); i.e. >70% of Cdx2 and Gata3 binding events occurred within open regions (Figure 3A–C) although ~53% of Arid3a targets are open regions. This suggested that CAG factors have additional functions distinctly from pioneer factors. Unexpectedly, all CAG factors show fewer target peaks at the later stage of reprogramming (TS+) compared to the earlier stage (ES+), indicating that CAG factors might be crucial during the initiation of reprogramming but play a more limited role thereafter or this may be due to the more relaxed chromatin structure of ES cells (Figure 3A–C). Regardless of the chromatin status (open versus closed), CAG factors preferentially bind to distal regulatory elements (Supplementary Figure S3A and B).

To further explore this observation, we examined whether CAG factors share common binding sites. Comparison of individual CAG factor binding sites revealed that CAG factors share a significant number of target sites regardless of their chromatin accessibility (Supplementary Figure S3C). CAG common peaks within open chromatin are predominantly located at distal elements as expected (91%). Interestingly, common closed chromatin peaks are exclusively located within distal elements (100%)—a strong indication of pioneer factor-like activity (Supplementary Figure S3D). We further found that common open chromatin loci are associated primarily with ES cell-specific genes, as evidenced by GO term enrichment (e.g. stem cell maintenance and cell fate specification) as well as DNA-binding motifs of ES cell core TFs (Supplementary Figure S3E and F). This suggested that CAG factors promote reprogramming by occupying open regulatory regions of the ES cell-specific genes prior to repressing them. On the other hand, common closed chromatin-associated target genes are involved in regulation of chromosome organization and other chromatin regulatory-related terms (Supplementary Figure S3E). Thus, CAG factors seem to regulate the chromatin landscape by activating genes implicated in chromatin remodeling or modification.

To further address this interpretation, we performed motif analyses of individual CAG binding sites associated with the above observed open and closed chromatin regions. As shown in Supplementary Figure 3D–F, top-ranked motifs associated with open CAG binding sites are ES cell-specific motifs (such as Oct4 and Sox2), whereas closed CAG target sites are enriched in TS cell-specific motifs (including Cdx2, Gata3 and Tead4). Unlike their closed common CAG targets that are mostly enriched with chromosome organization, their unique targets might have high portion of TS cell-specific genes based on their enriched TS cell-specific

motifs. These results imply that while CAG factors can bind to close chromatin by recognizing their own motif, they can also occupy open target sites without their own motifs probably via protein–protein interactions.

Arid3a promotes cell fate conversion by repression of ES cell-specific genes followed by activation of TS cell-specific genes

Among the CAG factors, Arid3a OE produced TS-like cells most similar to *bona fide* TS cells at the transcriptional and epigenetic level compared to OE of Cdx2 or Gata3 (Figure 1). Therefore, we investigated Arid3a OE-mediated reprogramming in greater depth. We performed a *de novo* motif search for Arid3a binding sites over the course of reprogramming using the HOMER motif search tool (27). Consistent with our previous study (16), the motif constituting the predominant binding sites for Arid3a in ES cells is highly similar to the motif employed by Oct4 (Supplementary Figure S4A) and binding motif preferences of Arid3a gradually changed over time during reprogramming (Supplementary Figure S4A). We also found that the most significant *de novo* motif of Arid3a identified in TS cells is similar to the previously known motif for Tead family proteins, which are known to be essential for TE specification *in vivo*. This prompted additional motif search (27) for previously established TFs spanning the binding sites of Arid3a. The analysis revealed that motifs for ES cell-specific pluripotency factors are enriched during the earlier time point of reprogramming (OEArid3a_ES+) (Figure 3G). Conversely, TS cell-specific motifs were over-represented at the later reprogramming time point (OEArid3a_TS+) (Figure 3G). These data clearly demonstrate that Arid3a dynamically switches its target site specificity during the course of reprogramming, perhaps by switching its protein interaction partners or by recognizing degenerate motifs.

To further investigate this gradual shift in target specificity and to elucidate its effect on associated target gene activity, we assessed Arid3a occupancy over the course of reprogramming with the levels of H3K27ac—a histone mark enriched among active enhancers (33) by clustering analysis. As shown in Figure 3H, we observed seven distinct patterns of Arid3a occupancy during the course of cell fate conversion. Peaks belonging to class I and IV are mainly associated with gene related to stem cell development (Supplementary Figure S4B), and strong Arid3a occupancy signals were predominantly observed in these classes at the early reprogramming time point (ES+). However, at the later time point (TS+), Arid3a occupancy shown in class I gradually disappears as its occupancy shifts to class II and III genes (Supplementary Figure S4A). In contrast to classes I and IV, genes belonging to class II and III are associated with TS cell-specific functions, such as placenta and labyrinthine development (Supplementary Figure S4B) and Arid3a occupancy gradually increases at the later stage. These data reveal dynamic changes in global Arid3a occupancy; i.e. Arid3a initially occupies ES cell-specific genes, presumably to repress them, followed by occupying TS cell-specific genes for their activation.

Since the H3K27ac signature is associated with active enhancers, we inspected its enrichment near Arid3a binding sites with associated chromatin accessibility. We observed

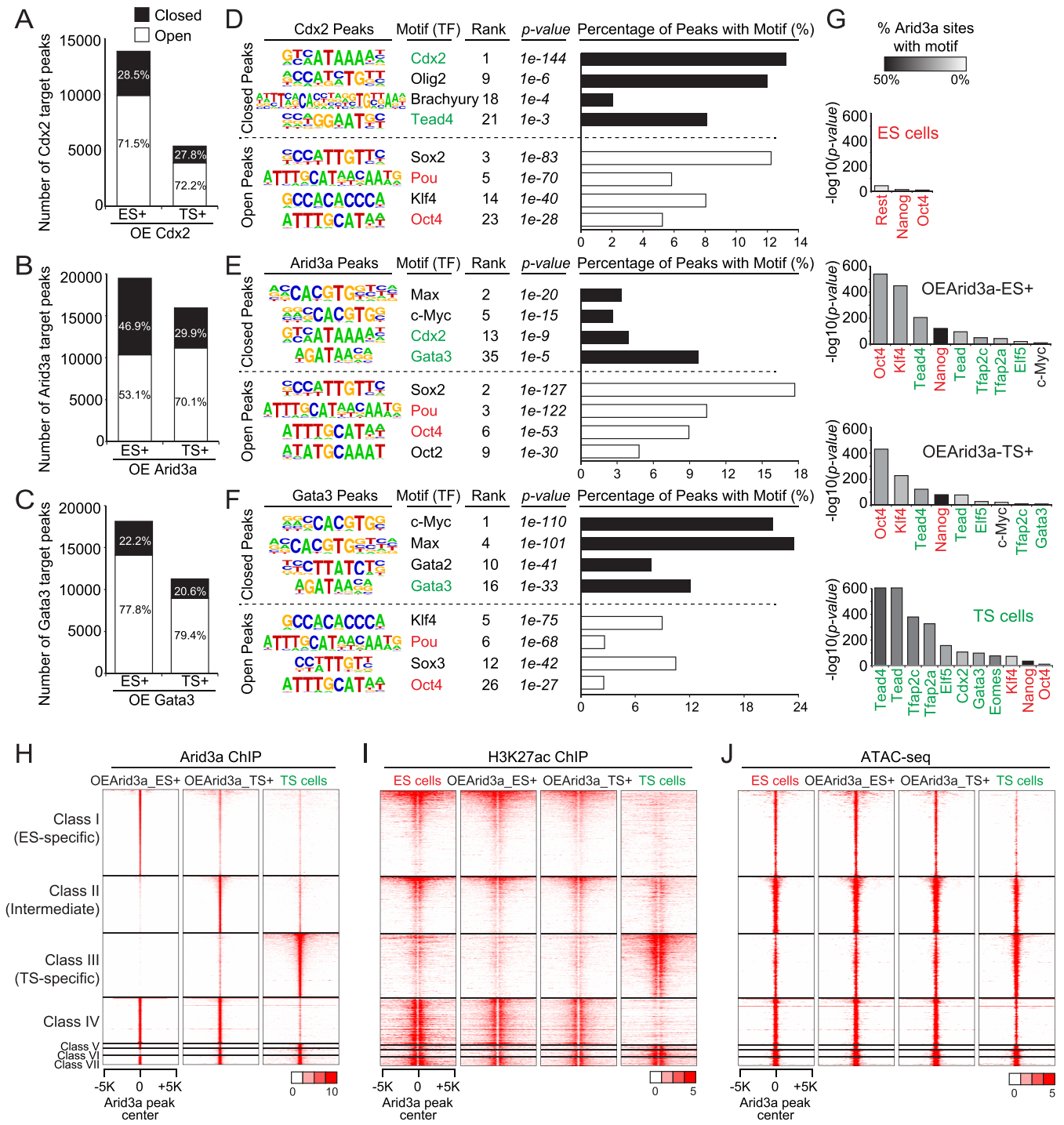


Figure 3. Enrichment of CAG binding motifs depends on the chromatin landscape. See also Supplementary Figure S3. (A–C) Bar graphs showing the number of Cdx2 (A), Arid3a (B) and Gata3 (C) target genes whose regulatory regions are associated with open or closed chromatin at the different stages of the reprogramming. (D–F) Upon OE of individual CAG factors in ES cells, significantly enriched motifs of Cdx2 (A), Arid3a (B) and Gata3 (C) binding sites that reside within open or closed chromatin are shown. Bar graphs (right panel) provide the percentages of individual CAG binding site with the motifs. (G) Bar graphs of enriched motifs within Arid3a binding sites in ES, TS and two different stages of reprogrammed ES cells. (H and I) Heatmaps showing multiple clustered Arid3a binding sites (H) and H3K27ac signatures (I) that are dynamically changed during the reprogramming. (J) Chromatin landscape changes in concert with changes in Arid3a occupancy plotted as a line graph showing average ATAC-seq peak scores within regions of Arid3a binding site classes shown in (H).

a strong positive correlation between the H3K27ac signature and Arid3a occupancy at each stage of reprogramming (Figure 3H and I). As reprogramming proceeded, the H3K27ac marks gradually disappeared from class I peaks (Figure 3I), suggesting that the initial binding of Arid3a to class I peaks ultimately removes H3K27ac signals to promote the suppression of ES cell-specific genes. Conversely, genes associated in class III gradually gain H3K27ac marks along with Arid3a occupancy during the course of reprogramming, indicating that Arid3a binding positively affects the activation of these TS cell-specific genes (Figure 3I). Taken together, our results indicate that Arid3a first represses ES cell-specific genes by reducing levels of H3K27ac, then it activates TS cell-specific genes by increasing H3K27ac. Since Arid3a itself does not possess acetyltransferase or deacetylase activity, our results imply that Arid3a may recruit proteins with these enzymatic activities. The feasibility of such a mechanism was provided by our demonstration that Arid3a recruits Hdac1 and Hdac2 to catalyze target repression (16).

Consistent with changes in Arid3a occupancy during reprogramming, we found that chromatin accessibility is also altered (Figure 3J and Supplementary Figure S4C). By analyzing ATAC-seq data, we observed the decreased accessibility along with loss of Arid3a occupancy during reprogramming in classes I (ES cell-specific) and IV. Conversely, class III (TS cell-specific) genes gained chromatin accessibility as Arid3a occupancy increased, consistent with classical pioneer factor activity (Supplementary Figure S4C). Thus, during the course of reprogramming, Arid3a switches its distal regulatory targets from ES cell-specific to TS cell-specific genes, resulting in a sequential shift in the transcriptional, epigenetic and chromosomal landscapes from an ES to a more TS-like cell profile.

CAG factors deactivate pre-existing ES cell-specific enhancers

Based on our findings, we reasoned that CAG factors might regulate the activity of ES cell-specific enhancers. To investigate this possibility, we analyzed the overlap of CAG factor targets with ES cell-specific enhancers, as defined by the histone enhancer mark, H3K27ac as well as ES cell-specific ‘super-enhancers’, a cluster of enhancers densely occupied by the ES cell-specific TFs, P300 and Mediator (34). We found that >70% of CAG factor targets overlap with active enhancers defined by H3K27ac in ES cells, and that ~50% of ES cell-specific super-enhancers are also occupied by CAG factors upon their OE in ES cells (Figure 4A and B). In addition, binding sites for the ES cell-specific TF Oct4 also displayed significant overlap with CAG binding sites (Figure 4C and D).

Based on these observations, we addressed two potential mechanisms of CAG-mediated repression of ES cell-specific genes. First, CAG factors might compete with ES cell-specific TFs, such as Oct4, to bind to ES cell-specific enhancers. To test this, we determined Oct4 occupancy upon OE of CAG factors in ES cells. Indeed, Oct4 occupancy was dramatically decreased, suggesting that there may be some level of binding competition among Oct4 and CAG factors at ES cell-specific enhancers (Figure 4E). However, we pre-

viously observed slightly decreased levels of Oct4 expression upon OE of Arid3a in ES cells (16); thus, we cannot rule out the possibility that decreased Oct4 occupancy is a result of reduced levels of Oct4. Alternatively, CAG factors may deactivate ES cell-specific enhancers by modifying histone signatures. To test this, we overexpressed Arid3a in ES cells using a dox—inducible system. Cells were treated with dox for 4 days, followed by dox withdrawal for four additional days and induction of Arid3a was monitored by the expression of ZsGreen1 (Figure 4F). We determined by quantitative PCR whether enhancer-associated H3K27ac signatures are modulated by CAG factors. As shown in Figure 4G, induction of Arid3a significantly increased the levels of H3K27ac at the enhancers of TS cell-specific genes. However, only modest changes in H3K27ac marks within ES cell-specific enhancers were observed. Upon dox removal, the levels of H3K27ac were decreased, indicating Arid3a-dependent regulation of H3K27ac on TS cell-specific genes.

Since Hdacs are responsible for deacetylation of H3K27ac (35), we additionally performed Hdac1 ChIP in these Arid3a inducible cells to monitor the changes in Hdac1 enrichment at the regulatory regions of ES cell-specific and TS cell-specific genes. We observed strong enrichment of Hdac1 at the enhancers of ES cell-specific genes without significant enrichment at the regulatory elements of TS cell-specific genes in early stage of the reprogramming (Figure 4H). Conversely, the shutdown of Arid3a induction by dox withdrawal reduced enrichment of Hdac1 at ES cell-specific enhancers (Figure 4H). Together, our results provide evidence that OE of Arid3a in ES cells initiates ES to TS-like cell reprogramming through changes in histone signatures on the enhancers, and this process is reversible at least at the early stage of reprogramming.

DISCUSSION

Overall, our discoveries led us to a stepwise model (Figure 4I) in which, initially, CAG factors predominantly occupy open chromatin of ES cell-specific genes to deactivate them. Then, they switch their binding site preference onto the regulatory elements of TS cell-specific genes for activation. Both deactivation and activation appear to be associated, if not mediated, by histone modifications, as levels of H3K27ac within enhancers change significantly over the course of reprogramming. This mechanism requires dynamic and sequential occupancy changes within the binding of CAG factors, as they must adjust their target specificity from ES cell-specific to TS cell-specific in a timely, ordered manner. These events, which underlie the reprogramming of ES cells to TS-like cells, uniquely expand the established mechanisms of pioneer factor-mediated transactivation. However, it is unclear if this is yet limited to reprogramming from ES cells to TS-like cells. Our stepwise and integrative dissection of the initial steps of reprogramming may help improve both the efficiency and fidelity of the reprogramming process as well as facilitate stem cell therapies in patients whom rely on changing the fate of their own cells for treatment.

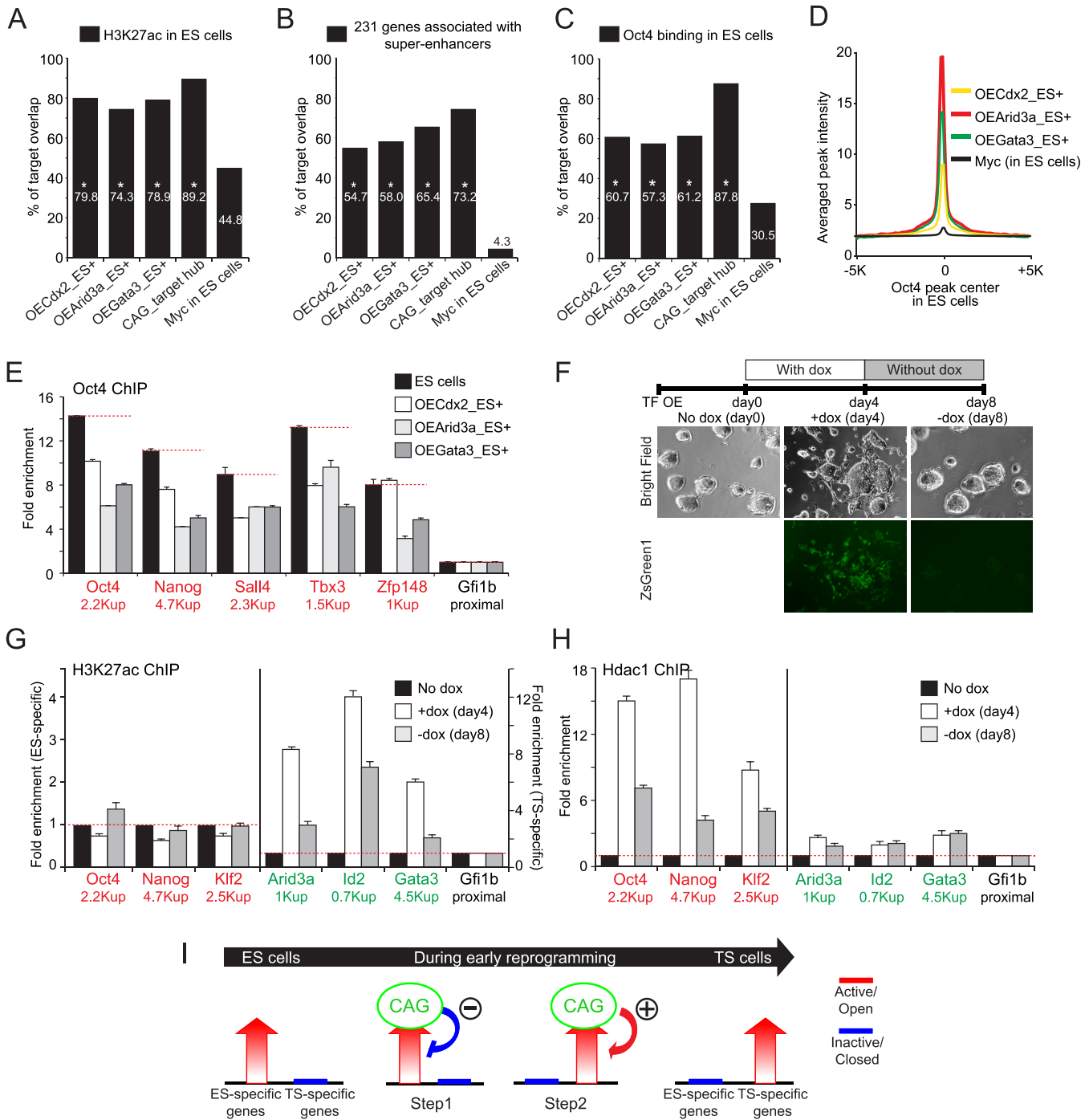


Figure 4. CAG factors decommission ES cell-specific enhancers (A–C) Bar graphs depicting overlaps of individual CAG factor binding sites and target hub upon OE of each factor in ES cells with H3K27ac enriched sites (A), super-enhancers (B) and Oct4 binding sites (C) in ES cells. Myc binding sites in ES cells are used as negative control. Asterisks mark significant overlaps (P -value < 0.0001). (D) Average occupancy profiles of each CAG factor centered on Oct4 binding sites in ES cells following OE of CAG factors in ES cells. (E) Oct4 ChIP-qPCR plotted to show relative Oct4 occupancy in the regulatory regions of ES cell-specific genes upon OE of CAG factors. Error bars depict standard deviations of biological triplicates. (F) Cell morphologies of ES cells and *Arid3a*-inducible ES cells upon either treatment with doxycycline (dox) for 4 days or following withdrawal of dox. (G) ChIP-qPCR showing relative levels of H3K27ac in the regulatory regions of ES cell-specific genes (red) and TS cell-specific genes (green). (H) ChIP-qPCR showing relative levels of Hdac1 occupancy in the regulatory regions of ES cell-specific genes (red) and TS cell-specific genes (green). (I) Two-step mechanism of CAG factor-mediated reprogramming of ES cells to TS-like cells.

ACCESSION NUMBER

Raw and processed RNA-, ChIP- and ATAC-seq data have been deposited at the public server GEO database under accession number GSE90752.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Drs Janet Rossant (U of Toronto) and Guang Hu (NIH/NIEHS) for providing TS cells; the Genome Sequencing and Analysis Facilities (U of Texas) for sample processing for RNA-, ChIP-, and ATAC-seq samples and NGS.

FUNDING

National Institutes of Health (NIH) [R01GM112722]; Burroughs Wellcome Fund (to J.K.); NIH [R01CA31534]; Cancer Prevention and Research Institute of Texas [RP16704, RP100612, RP120348]; Marie Betzner Morrow Centennial Endowment (to H.O.T.). Funding for open access charge: NIH [R01GM112722].

Conflict of interest statement. None declared.

REFERENCES

- Davis, R.L., Weintraub, H. and Lassar, A.B. (1987) Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*, **51**, 987–1000.
- Takahashi, K. and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, **126**, 663–676.
- Kulesa, H., Frampton, J. and Graf, T. (1995) GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboplasts, and erythroblasts. *Genes Dev.*, **9**, 1250–1262.
- Nerlov, C. and Graf, T. (1998) PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. *Genes Dev.*, **12**, 2403–2412.
- Stadtfeld, M., Maherali, N., Breault, D.T. and Hochedlinger, K. (2008) Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell*, **2**, 230–240.
- Zaret, K.S. and Carroll, J.S. (2011) Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.*, **25**, 2227–2241.
- Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M. and Zaret, K.S. (2015) Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell*, **161**, 555–568.
- Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., Chanda, S., Fuentes, D.R., Giresi, P.G., Ng, Y.H., Marro, S., Neff, N.F. *et al.* (2013) Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell*, **155**, 621–635.
- Jen, Y., Manova, K. and Benezra, R. (1997) Each member of the Id gene family exhibits a unique expression pattern in mouse gastrulation and neurogenesis. *Dev. Dyn.*, **208**, 92–106.
- Russ, A.P., Wattler, S., Colledge, W.H., Aparicio, S.A., Carlton, M.B., Pearce, J.J., Barton, S.C., Surani, M.A., Ryan, K., Nehls, M.C. *et al.* (2000) Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature*, **404**, 95–99.
- Ng, R.K., Dean, W., Dawson, C., Lucifero, D., Madeja, Z., Reik, W. and Hemberger, M. (2008) Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nat. Cell Biol.*, **10**, 1280–1290.
- Yagi, R., Kohn, M.J., Karavanova, I., Kaneko, K.J., Vullhorst, D., DePamphilis, M.L. and Buonanno, A. (2007) Transcription factor TEAD4 specifies the trophoblast lineage at the beginning of mammalian development. *Development*, **134**, 3827–3836.
- Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R. and Rossant, J. (2005) Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell*, **123**, 917–929.
- Ralston, A., Cox, B.J., Nishioka, N., Sasaki, H., Chea, E., Rugg-Gunn, P., Guo, G., Robson, P., Draper, J.S. and Rossant, J. (2010) Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development*, **137**, 395–403.
- Kuckenberger, P., Buhl, S., Woynecki, T., van Furden, B., Tolkunova, E., Seiffe, F., Moser, M., Tomilin, A., Winterhager, E. and Schorle, H. (2010) The transcription factor TCFAP2C/AP-2gamma cooperates with CDX2 to maintain trophoblast formation. *Mol. Cell Biol.*, **30**, 3310–3320.
- Rhee, C., Lee, B.K., Beck, S., Anjum, A., Cook, K.R., Popowski, M., Tucker, H.O. and Kim, J. (2014) Arid3a is essential to execution of the first cell fate decision via direct embryonic and extraembryonic transcriptional regulation. *Genes Dev.*, **28**, 2219–2232.
- Rhee, C., Edwards, M., Dang, C., Harris, J., Brown, M., Kim, J. and Tucker, H.O. (2017) ARID3A is required for mammalian placenta development. *Dev Biol.*, **422**, 83–91.
- Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W. and Orkin, S.H. (2006) A protein interaction network for pluripotency of embryonic stem cells. *Nature*, **444**, 364–368.
- Koressaar, T. and Remm, M. (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics*, **23**, 1289–1291.
- Webb, C.F., Bryant, J., Popowski, M., Allred, L., Kim, D., Harris, J., Schmidt, C., Miner, C.A., Rose, K., Cheng, H.L. *et al.* (2011) The ARID family transcription factor bright is required for both hematopoietic stem cell and B lineage development. *Mol. Cell Biol.*, **31**, 1041–1053.
- Kim, J., Woo, A.J., Chu, J., Snow, J.W., Fujiwara, Y., Kim, C.G., Cantor, A.B. and Orkin, S.H. (2010) A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell*, **143**, 313–324.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.*, **14**, R36.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J. and Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, **28**, 511–515.
- Kim, J., Chu, J., Shen, X., Wang, J. and Orkin, S.H. (2008) An extended transcriptional network for pluripotency of embryonic stem cells. *Cell*, **132**, 1049–1061.
- Buenrostro, J.D., Wu, B., Chang, H.Y. and Greenleaf, W.J. (2015) ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.*, **109**, 1–9.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
- Saldanha, A.J. (2004) Java Treeview—extensible visualization of microarray data. *Bioinformatics*, **20**, 3246–3248.
- Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
- McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M. and Bejerano, G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.*, **28**, 495–501.
- Heinz, S., Romanoski, C.E., Benner, C. and Glass, C.K. (2015) The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.*, **16**, 144–154.
- Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J.A., Mahony, S., Pugh, B.F., Lee, D., Kaestner, K.H. and Zaret, K.S. (2016) The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol. Cell*, **62**, 79–91.

33. Creighton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A. *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 21931–21936.
34. Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I. and Young, R.A. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, **153**, 307–319.
35. Urvalek, A.M. and Gudas, L.J. (2014) Retinoic acid and histone deacetylases regulate epigenetic changes in embryonic stem cells. *J. Biol. Chem.*, **289**, 19519–19530.