Published in final edited form as:

Stem Cells. 2015 February ; 33(2): 367–377. doi:10.1002/stem.1870.

Interrogation of a Context-Specific Transcription Factor Network Identifies Novel Regulators of Pluripotency

Ritu Kushwaha^{1,2}, Nirmala Jagadish^{1,2}, Manjunath Kustagi^{2,6}, Mark J. Tomishima³, Geetu Mendiratta¹, Mukesh Bansal⁶, Hyunjae R. Kim⁶, Pavel Sumazin⁶, Mariano J. Alvarez⁶, Celine Lefebvre⁶, Patricia Villagrasa-Gonzalez⁶, Agnes Viale⁴, James E. Korkola¹, Jane Houldsworth¹, Darren R. Feldman⁵, George J. Bosl⁵, Andrea Califano^{6,7,8,*}, and R. S. K. Chaganti^{1,5,8,*}

¹Cell Biology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021

³Stem Cell Core Facility, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021

⁴Genomics Core Facility, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021

⁵Department of Medicine, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021

⁶Center for Computational Biology and Bioinformatics

⁷Department of Systems Biology, Columbia University, 1130 St. Nicholas Avenue, New York, NY 10032

Abstract

The predominant view of pluripotency regulation proposes a stable ground state with coordinated expression of key transcription factors (TFs) that prohibit differentiation. Another perspective suggests a more complexly regulated state involving competition between multiple lineage-specifying TFs that define pluripotency. These contrasting views were developed from extensive analyses of TFs in pluripotent cells *in vitro*. An experimentally-validated, genome-wide repertoire of the regulatory interactions that control pluripotency within the *in vivo* cellular contexts is yet to be developed. To address this limitation, we assembled a TF interactome of adult human male germ cell tumors (GCTs) using the Algorithm for the Accurate Reconstruction of Cellular Pathways (ARACNe) to analyze gene expression profiles of 141 tumors comprising pluripotent and differentiated subsets. The network (GCT^{Net}) comprised 1305 TFs, and its Ingenuity Pathway analysis identified pluripotency and embryonal development as the top functional pathways. We experimentally validated GCT^{Net} by functional (silencing) and biochemical (ChIP-seq) analysis of

*Correspondence: chagantr@mskcc.org (R.S.K.C). califano@c2b2.columbia.edu (A.C).

^{2,8} Equal Contribution

Author Contributions: RSKC: conception and design, financial support, data interpretation and analysis, manuscript writing, final approval of the manuscript; AC: conception and design, financial support, data interpretation and analysis, manuscript writing; RK, NJ, MK: conception and design, data interpretation and analysis, manuscript writing; manuscript writing: MJK, GM, MB, HRK, PS, MJA, CL, PV-G, AV, JEK, JH, DEF: data interpretation and analysis; GJB: data interpretation and analysis, financial support.

the core pluripotency regulatory TFs *POU5F1*, *NANOG*, and *SOX2* in relation to their targets predicted by ARACNe. To define the extent of the *in vivo* pluripotency network in this system, we ranked all TFs in the GCT^{Net} according to sharing of ARACNe-predicted targets with those of *POU5F1* and *NANOG* using an Odds-Ratio analysis method. To validate this network, we silenced the top 10 TFs in the network in H9 ES cells. Silencing of each led to downregulation of pluripotency and induction of lineage; 7 of the 10 TFs were identified as pluripotency regulators for the first time.

INTRODUCTION

Recent studies have provided deep insights into the core pluripotency regulatory circuitry of the embryonic stem (ES) cell state involving interaction between transcription factors (TFs), noncoding RNAs and Polycomb group (PcG) proteins (Marson et al., 2008; Jaenisch and Young, 2008; Schuettengruber and Cavalli, 2009; Young, 2011). These studies have led to the concept that pluripotency is a stable ground state maintained by the coordinated expression of TFs, such as the canonical triad *POU5F1*, *NANOG*, and *SOX2*, which prohibits differentiation. In contrast, others have argued for an unstable nature of the pluripotency state, with competition between lineage-specifying TFs acting to maintain a delicate regulatory balance necessary for its maintenance (Loh and Lim, 2011). Consistent with such a view, Shu et al (2013) showed that lineage specifiers that are pluripotency rivals and are also not commonly expressed in ES cells can reprogram mouse somatic cells to pluripotency. These contrasting views suggest that in-depth elucidation of the mechanisms underlying pluripotency require a far more detailed view of the regulatory programs that control it, *in vitro* as well as *in vivo*. Although we and others have shown that evaluating the overlap of regulatory programs controlled by TF-pairs is critical in assessing their epistatic, synergistic, or complementary role in regulating specific pathophysiologic processes (Carro et al., 2010; Ravassi et al., 2010; Della Gatta et al., 2012), an experimentally-validated, genome-wide repertoire of transcriptional interactions (*i.e.*, *interactome*) representing pluripotency and differentiation processes in stem cells is not yet available to allow these kinds of analyses to be performed. Thus, over all, the complete regulatory control landscape of the pluripotent cell remains largely elusive.

Recently, we have successfully assembled several cell-context-specific transcriptional networks for mammalian cells using the Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNe) (Basso et al., 2005, 2010; Palomero, et al., 2006; Lim et al., 2009; Margolin et al., 2009; Zhao et al., 2009; Carro et al., 2010; Lefebvre et al., 2010; Cadeiras et al., 2011). These networks withstood rigorous experimental validation and their analysis helped elucidate previously unsuspected regulatory mechanisms (Zhao et al., 2009; Carro et al., 2010; Lefebvre et al., 2010; Della Gatta et al., 2012). To apply these methods to the reconstruction and interrogation of an interactome representing the *in vivo* regulatory landscape of pluripotent cells, we chose the adult human male germ cell tumor (GCT) as the model system. GCTs originate from lineage-restricted germ cells and a subset, embryonal carcinoma (EC), re-acquires pluripotency comparable to that of the blastocyst. EC cells represent a biological counterpart of ES cells; including presenting very similar gene expression profiles (GEPs) (Sperger et al., 2003). GCTs comprise two main histologic

subsets, seminoma (SEM) and non-seminoma (NS) (Ulbright, 1993; Chaganti and Houldsworth, 2000). SEM presents as germ cell-like, non-differentiating tumors and NS comprises EC and its developmental derivatives teratoma, which displays somatic lineage differentiation, and choriocarcinoma and yolk sac tumor, which display features of extra-embryonal differentiation. EC development *in vivo* involves downregulation of pluripotency and upregulation of lineage differentiation.

We used ARACNe to analyze a large GEP dataset derived from fresh-frozen GCT biopsies representing all differentiation lineages and normal testis (Korkola et al., 2005, 2006, 2009). This analysis yielded the first *de novo*, genome-wide transcriptional interactome (GCT^{Net}), representing the context-specific regulatory programs that determine the *in vivo* behavior of these tumors, including control of pluripotency, differentiation, and tumorigenesis. We evaluated the quality of the interactome with specific focus on established pathways controlling pluripotency and developed an experimentally validated map of the transcriptional programs controlled by *POU5F1*, *NANOG*, and *SOX2*. Time -series GEP analysis followed by silencing of these factors, as well as genome-wide chromatin immunoprecipitation (ChIP) assays, were used to validate their ARACNe-inferred targets and provide further insight into the specific, time dependent, functions regulated by them during loss of pluripotency and differentiation *in vitro* in NT2/D1 EC cells. Analysis of all genetic programs co-regulated by TF pairs *in vivo* in the interactome revealed the first genome-wide map of cooperative regulation in a pluripotent cell and its developmental derivatives and provided critical insights to identify additional candidate pluripotency regulators. Specifically, by ranking all TFs in the network based on the statistical significance of the programs they co-regulate with *POU5F1* and *NANOG*, we identified several novel candidate pluripotency regulators. Functional validation of the 10 most significant regulators identified in this approach by lentivirus-mediated short hairpin RNA (shRNA) silencing in H9 ES cells showed that each of them was necessary to maintain the pluripotency state. These 10 TFs included 3 that were previously identified, and 7 that were previously not identified, as pluripotency regulators. These results thus demonstrate that the GCT^{Net} interactome represents bona fide regulatory interactions in this tumor system; they further elucidate the novelty and complexity of the pluripotency control system.

MATERIALS AND METHODS

Additional details about the experimental procedures can be found in the Supplemental Information.

Gene expression profiles (GEPs) of GCT

The GEP data from 141GCT samples previously reported by us was obtained using Affymetrix U133A+B microarrays (data available in GEO database, accession No GSE 3218, Korkola et al., 2006, 2009).

Cell culture

NT2/D1 cells were maintained and propagated as described previously (Houldsworth et al., 2002). The H9 human ES cells were obtained from the MSKCC Stem Cell Core Facility.

ChIP-seq and ChIP PCR

ChIP assay of NT2/D1 and H9 cells was performed according to described methods (Schmidt et al., 2009).

Lentivirus-mediated shRNA silencing of *POU5F1*, *NANOG* and *SOX2*

NT2/D1 ES cells stably transduced by lentiviral vectors were collected on days 2, 3, 6, and 7 from *POU5F1*, *NANOG* and *SOX2* silenced cells for RNA isolation and protein estimation. RNA was converted into double-stranded cDNA and cRNA and checked on an Agilent bioanalyzer, fragmented, and hybridized to Affymetrix HG-U133Plus microarrays.

ARACNe analysis

The raw Affymetrix CEL files obtained after the imaging protocol were further processed to obtain probe-set level consensus using the Affymetrix MAS5 protocol (affy package from Bioconductor; <http://www.bioconductor.org>, (Gautier, L et al., 2004). This was followed by a log2 transformation and then quantile normalization as implemented in geWorkbench (<http://www.geworkbench.org>). The probe-set level expression values were then profiled (Margolin et al., 2006) to obtain a MI vs p-value curve. This curve described MI values to be used as a threshold corresponding to a user specified p-value for ascribing significance to 'edges' in the resulting ARACNe network. A p-value cutoff threshold of 0.05 (Bonferroni corrected for multiple testing) was used when running ARACNe (threshold corresponded to a Mutual Information value of 0.489). A Data Processing Inequality (DPI) value of 0 was used in the analysis. ARACNe was run in bootstrapping mode with a 100 separate runs of sample with replacement modifications to the original normalized expression dataset. The consensus network from the 100 separate runs was obtained considering all edges occurring in at least 95 of the 100 separate bootstrapped networks, and represents the GCT^{Net} interactome.

RESULTS

Assembly of a core transcriptional network for GCTs

The ARACNe algorithm was used to analyze 135 GEPs derived from GCT biopsies, representing pluripotent and differentiation-associated histologic subsets of this malignancy, and from 6 normal testes as controls, to produce a comprehensive map of transcriptional interactions. Out of 3,099 probe sets annotated as TFs in the Affymetrix gene expression platform (HG-U133A+B) by Gene Ontology (GO) (Liu et al., 2003), ARACNe predicted targets for 2,612 probe sets, corresponding to 1,305 distinct TFs that virtually recapitulate all TFs expressed in this cellular context. A mutual information threshold (MI) 0.489 was selected, corresponding to a $p = 0.05$, the Bonferroni-corrected p-value, for the inference of any false positive interactions in this dataset. ARACNe inferred ~260,000 unique regulatory interactions between the 1,305 TFs and 17,025 target genes, which together comprise the complete GCT^{Net}. To determine whether the GCT^{Net} provided an adequate regulatory context to study pluripotency and lineage differentiation, we first selected the 130 TFs (top 10%) with the largest regulon size (i.e., the number of their ARACNe-inferred targets). Analysis of GO biological categories revealed that these regulators were highly enriched in

those controlling pluripotency, cell fate, development and differentiation (Table S1A). Similarly, the most enriched pathway in Ingenuity analysis of the top 130 TFs was “Role of Oct4 in Mammalian Stem Cell Pluripotency” (Table S1B). Taken together, these data confirm that known pluripotency and differentiation-related regulators are especially well represented in the GCT^{Net} suggesting that this interactome provides a valuable environment to study these processes.

To perform a systematic, albeit non-comprehensive quality control analysis for the inferred interactome, we undertook a detailed experimental validation of the ARACNe-inferred targets of the 3 most relevant TFs in the pluripotency pathway, *POU5F1*, *NANOG*, and *SOX2*. For this, we first analyzed the regulatory programs represented by their targets in the GCT^{Net} and then proceeded with their experimental validation by ChIP-seq and profiling assays, following their shRNA-mediated silencing using NT2/D1 EC and H9 ES cells, which were selected as the closest in vitro counterparts of the *in vivo* GCT cells for studying programs controlling pluripotency and differentiation.

ARACNe-inferred targets of *POU5F1*, *NANOG*, and *SOX2* direct distinct regulatory pathways in the GCT^{Net}

Among the 17,025 regulatory targets represented in the GCT^{Net}, 835 (5%) were found to be either individual or joint targets of *POU5F1*, *NANOG*, and *SOX2*; individually, the regulons of these 3 TFs included 338, 376, and 307 genes, respectively. Interestingly, while *POU5F1* and *NANOG* shared 127 common targets (max overlap $\omega \approx 38\%$, $p = 0$, i.e., -value below machine precision by Chi-square test), *POU5F1* and *SOX2* as well as *NANOG* and *SOX2* shared only 40 ($\omega \approx 13\%$, $p < 5.1\text{E-}56$) and 41 ($\omega \approx 13\%$, $p < 2.7\text{E-}62$) common targets, respectively, suggesting a smaller albeit still highly significant regulatory program overlap; finally, the intersection of all 3 regulons included 21 genes (**Figure 1A, Table S2A and S2B**). Taken together, these results suggest that, in addition to a small core shared program regulated by all 3 TFs, *SOX2* controls a partially orthogonal transcriptional program, while *NANOG* and *POU5F1* display virtually overlapping programs in this interactome. Considering that both experimental and computational methods to dissect molecular interactions have relatively high false negative rates (FNRs), the observed overlap between *POU5F1* and *NANOG* programs is extremely high, suggesting that they control almost identical programs. Assuming an FNR of 30% for ARACNe, the true overlap would be $\Omega = 78\%$, while at 37% FNR, it would be $\Omega = 100\%$.

Interestingly, while GO categories enriched in targets individually controlled by either *POU5F1* or *NANOG* included regulation of cell and nucleic acid metabolism (**Figure 1B and 1C, Tables S2C and S2D**), common targets of both TFs were highly enriched in categories representing control of stem cell and/or developmental related functions. In sharp contrast, *SOX2* targets were mainly involved in developmental processes (**Figure 1D, Table S2E**). Finally, the 21 target genes that were inferred as targets of all 3 TFs (*RHOBTB3*, *RIF1*, *NONO*, *DLG3*, *TTC3*, *MRS2L*, *TBC1D23*, *MED12*, *JARID2*, *PBX3*, *USP28*, *A4GALT*, *FGF4*, *ATP12A*, *CSH1*, *HLA-DPB2*, *ZNF589*, *PDPN*, *DPPA4*, *DND1*, *TERF1*) preside over important functions associated with regulation of stem cell maintenance and development, differentiation and morphogenesis, compartment specification, cell cycle and mitosis, and

regulation of transcription, suggesting a core stem cell regulatory program controlled by all 3 TFs in this biological system.

ARACNe-inferred targets of *POU5F1*, *NANOG*, and *SOX2* are validated by ChIP-seq and ChIP-PCR

Analysis of DNA-binding sites predicted by ChIP-Seq assays in both NT2/D1 EC and H9 ES cells for each of the 3 TFs revealed their significant enrichment in proximal promoter regions (10kb, 5kb, and 2kb) of ARACNe-inferred targets, despite the differences between the *in vivo* context from which these were inferred and the *in vitro* context in which they were validated (**Figure S1 A-H**). Finally, we performed individual ChIP-PCR assays for the top 10 ARACNe-inferred targets of each TF, in both NT2/D1 EC and in H9 ES cells in a very small region [-1,000bp,+500bp] surrounding the transcription start site (TSS) (**Figure S1G and S1H and Table S2F-K**). This analysis confirmed specific binding of *POU5F1* and *SOX2* to the promoter of each of their top 10 targets, in both cell lines. *NANOG* was found to bind specifically to the promoter of 9 out of 10 of its ARACNe-inferred targets in both cell lines and to the promoter of *DND1*, in H9 ES but not NT2/D1 EC cells. These data suggest that actual binding specificity of these TFs to their ARACNe-inferred targets is significantly higher than that revealed by ChIP-related analysis, due to the previously reported false negative rates of this genome-wide technique (Margolin et al., 2009).

Functional pathways mediated by *POU5F1*, *NANOG*, and *SOX2* are identified by target analysis following their shRNA mediated silencing

To functionally characterize the ARACNe-inferred targets of the 3 TFs, we achieved their stable silencing by lentivirus-mediated shRNA in NT2/D1 EC cells, followed by GEP. Cells were harvested on day 6, following puromycin selection of clones that integrated the construct. Downregulation of the silenced TFs was assessed by Western blot (data not shown). Enrichment of ARACNe-inferred targets in the differentially expressed genes following silencing of each TF was highly significant by gene set enrichment analysis (GSEA) (Subramanian et al., 2005) ($P_{POU5F1} = 3.8E-18$, $P_{NANOG} = 5.4E-19$, $P_{SOX2} = 1.72E-12$, respectively) (**Figures 2 A-C and Table S3A**). Based on GO enrichment analysis, categories of downregulated *POU5F1* and *NANOG* targets were primarily enriched in cellular organization and DNA metabolism categories, whereas upregulated targets were enriched in cellular metabolism categories (**Figures 2 D-G and Table S3A**). Conversely, downregulated *SOX2* targets were enriched in regulation of histone methylation and modification and stem cell differentiation and chromosome organization categories, whereas upregulated targets were enriched in control of nervous system development, neurological system process, regulation of axon extension, and chondrocyte differentiation categories, in addition to other developmental processes (**Figures 2H and I, Figure S2 and Table S3A**).

In addition to analysis of these genes following stable silencing, we also analyzed the 7-day time-course following silencing. Cells were harvested on days 2, 3, 6, and 7 following transfection to produce GEPs. Controls included non-transfected cells and cells transfected with non-target RNA. At each time point, the corresponding TF protein levels were also assessed by Western blot (data not shown). We first examined the overall pattern of expression change of target genes during the time-course for all 3 TFs. Most genes showed

progressive up- or downregulation starting on day 2. A smaller proportion of genes however showed a more complex oscillatory pattern, showing up- or downregulation at day 2 but reversing this trend at day 7. These patterns are illustrated in an animation (**Figures S3**). Intriguingly, despite the fact that silencing of each TF induced loss of pluripotency and differentiation, the specific time-dependent programs as well as their enrichment in specific functional categories were TF- specific. Targets of *POU5F1* and *NANOG* that were downregulated by day 2 were predominantly enriched in regulators of pluripotency, cell cycle, chromatin modification, and metabolism, whereas upregulated targets were enriched in regulation of metabolism and signal transduction. For *SOX2*, instead, development related and signal transduction genes were highly enriched in the up- as well as in the downregulated targets.

An analysis of variance (ANOVA) (Anscombe, 1948) of the GEPs at the 4 time points versus controls for each TF yielded a set of genes, whose expression was most variable at these time points (**Figure 3A**). Pair-wise overlap of affected genes following silencing was very significant (p-value below machine precision by Chi-square test). A set of experimentally consistent targets (core program) controlled by each TF was then obtained by selecting their ARACNe-inferred targets whose expression was affected by ANOVA analysis (**Figures 3C-E**). Enrichment of TF-specific core programs in GO categories was also computed and compared confirming the previous results, (**Tables S3B-D**). The overlap of enriched GO categories for *NANOG* and *POU5F1* core programs (p-value below machine precision by Chi-square test) was much more extensive than for the overlap of *SOX2* and *NANOG* core programs (p = 4.66E-239) and for *SOX2* and *POU5F1* core programs (p = 6.05E-134, **Figure 3B**; **Table S3E**). These results suggest that experimental validation of ARACNe-inferred targets by TF silencing substantially confirms their functional organization.

Finally, we analyzed the time series GEPs by Principal Component Analysis (PCA) (Pearson, 1901). On average, PCA identified 3 statistically significant components capturing the gene expression trend across the 4 discrete time-points, representing the 7 days following the shRNA-mediated TF silencing, with 2 components capturing most of the data variance. **Figures 3F and G** project the cell-state trajectory over the most significant components (PC1 and PC2) across the four time-points following silencing of the individual TFs. As shown, the time-dependent trajectory following *SOX2* silencing is highly distinct from those following silencing of *NANOG* and *POU5F1*, which instead are strikingly similar. Thus, PCA analysis further confirmed that *NANOG* and *POU5F1* have largely overlapping programs, while *SOX2* controls a relatively independent program across the time-course, which induces loss of pluripotency while resulting in a different cell endpoint state, as shown in **Figure 3H**.

Taken together, these assays support the accuracy of the ARACNe-inferred transcriptional programs and suggest that the GCT^{Net} interactome constitutes a reasonably accurate and comprehensive map of the transcriptional interactions that regulate GCT behavior, including control of their pluripotent state.

GCT^{Net} interrogation identifies novel pluripotency regulators

We proceeded to test whether GCT^{Net} interrogation could help identify novel pluripotency regulators based on the statistical significance of the overlap of their transcriptional targets with those of *NANOG*, *POU5F1*, and *SOX2*. To assess whether such an analysis could be performed for any TF of interest without requiring additional experimental data, we designed it to depend only on the availability of the GCT^{Net}, without requiring any experimental signature following shRNA mediated silencing of a TF of interest.

Specifically, for any pair of TFs represented in the GCT^{Net} we computed their Transcriptional Overlap Odds Ratio, *TOOR*, *i.e.*, the ratio between the number of overlapping targets of two TFs divided by the number of overlapping targets expected by chance, as well as the associated p-value. From more than 800,000 possible combinations of the 1,305 TFs in the interactome, we considered only TF-pairs with at least 20 overlapping targets and selected the 10% most significant ones (*TOOR* = 8.1), resulting in 12,120 significant pairs, covering 910 of the 1,305 TFs. These pairs were displayed in a proximity map, where the distance between two TFs is inversely proportional to the statistical significance of their *TOOR* (*i.e.*, the larger the overlap the closer the TFs). Interestingly, most of the previously reported ES cell state regulators formed distinct clusters in this map (**Figure 4**). A conclusion of this analysis is that many TFs in the GCT^{Net} have significantly overlapping programs with established ES cell regulators, including *POU5F1*, *NANOG*, and *SOX2*, suggesting that if these are indeed involved in pluripotency control, then regulation of the ES cell state may indeed be much more complex than currently appreciated.

To test this hypothesis, we proceeded to identify and experimentally validate the 10 TFs with the most significant regulatory program overlap with *POU5F1* and *NANOG*. We selected this specific pair given their established relevance in pluripotency control and the fact that their regulatory programs are virtually overlapping. However, the same analysis could have been easily performed for any desired TF or TF cluster in the GCT^{Net}.

To compute the statistical significance of the pooled odd ratios for each TF with two distinct TFs (*i.e.*, *NANOG* and *POU5F1*) we used the Mantel–Haenszel statistics (Mantel and Haenszel, 1959). (**Table S4**). Based on this analysis, the ten TFs with the most significant pooled *TOOR* included: *PRDM14*, *UTF*, *ZNF296*, *ZNF618*, *ZNF208*, *ZNF589*, *ETV4*, *TEAD4*, *VENTX*, and *TFCP2L1*.

Analysis of *POU5F1/NANOG* program overlap identifies novel pluripotency regulators

To investigate their role as pluripotency regulators, we performed lentivirus-mediated shRNA silencing of these 10 TFs in H9 ES cells and evaluated lineage initiation by monitoring RNA and protein levels of a panel of relevant marker genes as a read out for loss of pluripotency. For each TF, a minimum of four distinct shRNA hairpins were first evaluated for their silencing efficiency. The 2 hairpins inducing the most efficient TF silencing were then used in the loss of pluripotency assay, which was performed in triplicate. Cells were harvested on day 5 following puromycin selections of clones that integrated the construct and TF silencing was confirmed by qRT-PCR. Cell cycle analysis of cells silenced for the 4 zinc finger TFs showed decrease in S-phase cells (Figure S4). RNA

from the silenced cells was assayed for expression of multiple lineage markers by qRT-PCR, including those for neuronal (*NES*, *NEFM*, *NEUROD*, *PAX6*, *PROM1*, and *GFAP*), ectodermal (*OLIG2*, *KTR7*, *KTR19*, and *CDH3*), mesodermal (*T*, *TWIST1*, *MSX1*, and *FOXF2*), endodermal (*AFP*, *GATA4*, *DAB2*, *NODAL*, *SOX7*, *SOX17*, and *GATA6*), and trophoectodermal (*HAND1*, *EOMES*, and *CDX*) lineages. As shown in **Figure S5**, each lineage was expressed in the silenced cells of each TF as indicated by expression of one or more of these markers, thus phenocopying the pattern observed in cells silenced for *POU5F1* and *NANOG* (controls). Such multilineage marker expression of ES cells silenced for *POU5F1* and *NANOG* was previously shown in several studies (Niwa et al., 2000; Mitsui et al., 2003; Chazaud et al., 2006; Ivanova et al., 2006; Wang et al., 2012).

We then further confirmed the RNA expression results by immunofluorescence analysis of H9 ES cells following silencing of each TF, using a representative antibody for each lineage, including PAX6 (neuronal), HAND1 (trophoectodermal), NODAL (endodermal), KRT7 (ectodermal), and MSX1 (mesodermal). As shown in **Figure 5**, silencing of *ZNF208* resulted in lineage marker expression patterns very similar to those following *POU5F1* silencing. Silencing of other TFs, including *TEAD4*, *UTF1*, *ZNF296*, *TFCP2L1*, *ZNF589* and *PRDM14*, also displayed similar lineage expression patterns, whereas silencing of *ETV4* and *VENTX* induced a pattern dominated by expression of endodermal lineage markers (**FigureS5 and S6**). We further quantitated the protein expression of selected lineage markers by Western blotting (**Figure S7**). We also noted that lineage markers were expressed in clonal clusters, within silenced cell colonies.

Overall, these results clearly establish that each of the 10 computationally inferred TFs play an individual role in pluripotency maintenance. It is noteworthy that 4 of these TFs are zinc finger genes of poorly annotated function (*ZNF296*, *ZNF618*, *ZNF208*, *ZNF589*) and 3 of them are developmental regulators (*ETV4*, *TEAD4*, and *VENTX*) with no previous indication of a role in pluripotency regulation. The remaining 3, *UTF1*, *PRDM14*, and *TFC2PL1*, have previously been established to play roles in pluripotency regulation in the *POU5F1*-*NANOG*-*SOX2* pathway (Okuda et al., 1998; Loh et al., 2006; Nishimoto et al., 2013; Chen et al., 2008, van den Berg et al., 2010). These results thus strongly support the hypothesis that pluripotency regulation is more complex than recognized and involves a broad spectrum of partners of the canonical regulators, and possibly non-partners as well, as was indicated in the studies of Shu et al. 2013.

DISCUSSION

In this study, we used our previously published GEP data from adult human male GCTs to reverse-engineer and experimentally validate a genome-wide transcriptional regulatory network in a cellular context that is relevant to the study of pluripotency, differentiation, and tumorigenesis *in vivo*. This interactome provides the first *de novo* genome-wide resource representing the *in vivo* regulatory landscape of pluripotent cells for the study of processes relevant to pluripotency maintenance and differentiation. As shown in prior related efforts in cancer, including glioma (Zhao et al., 2009; Carro et al., 2010; Sumazin et al., 2011), T cell acute lymphoblastic leukemia (T-ALL) (Palomero et al., 2006; Della Gatta et al., 2012), and B cell lymphoma (Mani et al., 2008; Lefebvre et al., 2010), we expect that the availability of

comprehensive and accurate models of regulation will be transformational in the elucidation of a variety of *in vivo* pluripotent cell related mechanisms.

Analysis of the GCT^{Net} revealed multiple pathways of pluripotency, differentiation, and tumorigenesis simultaneously operating in this complex biological system. Notably, “Role of *OCT4* in Mammalian Embryonic Stem Cell Pluripotency” emerged as the most significantly enriched functional pathway in the GCT^{Net}, confirming that this interactome is uniquely suited for the study of pluripotency and lineage differentiation *in vivo*, thereby establishing a strong biological rationale for the validation of the canonical TF triad *POU5F1*, *NANOG*, and *SOX2* in this system. The ARACNe-inferred interactions of the 3 TFs in the network, validated by rigorous functional assays *in vitro* by silencing experiments offer, for the first time, a glimpse into their core regulatory programs, with insights into specific time-dependent functions that they regulate during loss of pluripotency and lineage specification.

Our experiments validated the accuracy of ARACNe inferences. Indeed, single gene ChIP-PCR assay results showed that all 3 canonical pluripotency TFs bind to virtually all of their top ARACNe-inferred targets. This result highlights an advantage of the regulatory models developed by genome-wide ARACNe-based target identification compared to those assembled by highthroughput ChIP approaches for a handful of genes. The latter identify, for a given TF, an enormous number of binding targets, often in the thousands. Yet, among these, only a handful is functionally regulated by the TF in a specific context. The ARACNe-inferred targets, on the other hand, represent expressed functional targets compared to binding targets identified by ChIP-Chip, as is previously highlighted in the study of *BCL6* regulation in human B cells (Basso et al., 2010).

To avoid the issue that TF binding does not imply regulation, we performed functional validation of the ARACNe-inferred targets by 7-day time course GEP following shRNA-mediated silencing of the 3 TFs in NT2/D1 cells. High enrichment of ARACNe-inferred targets in differentially expressed genes was observed following shRNA-mediated silencing of each TF. These assays confirmed the very high enrichment of ARACNe-inferred targets in differentially regulated genes following silencing of the corresponding TF, and revealed enrichment of several functional categories among the inferred targets. This included downregulation of cell cycle and stem cell differentiation and upregulation of metabolic gene categories. Developmental gene categories showed a mixed effect with both up- and downregulation, as expected, since these are related to multiple developmental pathways, only some of which are represented among the differentiated cells. Silencing of *SOX2* showed a rather distinct pattern of expression, compared to *NANOG* and *POU5F1*, including upregulation of genes predominantly related to developmental process categories. PCA also showed that *NANOG* and *POU5F1* programs are substantially overlapping, while *SOX2* controlled an orthogonal program across the time-series and induced a distinct phenotypic endpoint at terminal differentiation by day 14.

Analysis of overlap of ARACNe-inferred targets between TFs in the interactome, implying target sharing, led to several important insights and discoveries. First, it showed that 910 of the 1305 TFs in the interactome (69%) presented highly significant regulatory program

overlap, thus emphasizing the importance of considering transcriptional programs jointly regulated by multiple TFs as well as the context-specific complexity of combinatorial TF regulation in the dissection of biological process of interest. Indeed, focusing on regulatory program overlap between the *NANOG/POU5F1* pair and other TFs, we identified a large number of TFs as high probability candidate pluripotency regulation co-factors. This suggests a much broader pluripotency regulatory network than currently recognized involving both previously identified co-regulators as well as a large number of novel regulators. As a proof of principle, as well as to show that results stemming from the analysis of the proposed pluripotent tumor system regulatory network could be effectively replicated in non-cancer human pluripotent stem cells, we successfully investigated the role of the top 10 of these TFs in pluripotency regulation in H9 ES cells.

Indeed, lentivirus mediated shRNA silencing of each of these candidate regulators elicited loss of pluripotency markers and gain of multilineage differentiation markers, recapitulating the behavior of ES cells in which *POU5F1* and *NANOG* were silenced, as controls. The top 2 of these TFs, *PRDM14* and *UTF1*, have previously been shown to be important ES cell state regulators that interact with the core regulators (Okuda et al., 1998; Zhao et al., 2008; Chia et al., 2010; Ma et al., 2011; Jia et al., 2012; Chan et al., 2012; Grabole et al., 2013; Yamaji et al., 2013). *TFPC2L1* has been shown to have significant target sharing with *POU5F1* and *NANOG* and play a role in pluripotency regulation (Chen et al., 2008; van den Berg et al., 2010; Ye et al., 2013). However, the 4 human zinc finger TFs (*ZNF296*, *ZNF618*, *ZNF208*, and *ZNF589*) have not been previously functionally characterized, although the murine *Zfp296* has recently been shown to induce reprogramming of murine fibroblasts earlier and more efficiently in combination with *Oct4*, *Sox2*, *Klf4* and *C-myc* suggesting a role in pluripotency regulation for this gene (Fischedick et al., 2012). Interestingly, the KRAB- containing ZNF proteins have recently been shown to activate the endogenous *POU5F1* promoter thereby suggesting a role for these genes in regulating pluripotency (Juarez-Moreno et al., 2013; Ji et al., 2014). The remaining 3 TFs (*ETV4*, *TEAD4*, and *VENTX*) are functionally divergent and have not been previously reported as ES cell state regulators. *ETV4* is an oncogene that is rearranged and/or overexpressed in prostate and other neoplasms (Oh et al., 2012), *TEAD4* specifies trophoderm in preimplantation mouse embryos (Nishioka et al., 2008), and *VENTX* is a homeobox TF that promotes expansion of human hematopoietic stem cells (Gao et al., 2012). In addition, human as well as *Xenopus* orthologs of *VENTX* have been shown to be targets of *POU5F1* and *NANOG* suggesting that this gene may play a role in pluripotency regulation (Jung et al., 2010; Scerbo et al., 2012). Thus, 7 of the top 10 predicted pluripotency regulators, based on the overlap of their regulatory programs with those of the 3 core regulators, are shown to be novel, bona fide pluripotency regulators. The specific mechanism(s) by which these novel TFs play a role in lineage induction and whether they can by themselves or in combination with the core regulators reprogram somatic cells to pluripotency are issues that need to be addressed. Our results presented here nevertheless vividly emphasize the complexity of the pluripotency network in the *in vivo* context.

CONCLUSION

This study represents the development of the first experimentally validated genome-wide map of transcriptional interactions in a pluripotent tumor system *in vivo*. We show here the feasibility of systematically unraveling the complexity of context-specific gene interaction in a biological system that can be used to study processes of interest. The extensive validation of the functional targets of the classical triad of core pluripotency regulatory TFs and the discovery of an extensive network of novel pluripotency regulators that interact with them suggest that a systems biology approach as used here has merit and can help further elucidate *in vivo* regulation of the stem cell state and lineage differentiation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This study was supported by the NIH grants CA008748, CA121852, CA109755, GM091240, NYSTEM grant CO29153, and the Byrne Fund. The MSKCC Genomics Core Facility performed the microarray and ChIP-seq. We thank the MSKCC Molecular Cytology Core Facility for help with IF analysis and imaging. The authors declare no conflicts of interest.

REFERENCES

- Anscombe FJ. The Validity of Comparative Experiments. *Journal of the Royal Statistical Society Series A. (General)*. 1948; 111:181–211.
- Basso K, Margolin AA, Stolovitzky G, et al. Reverse engineering of regulatory networks in human B cells. *Nat Genet*. 2005; 37:382–390. [PubMed: 15778709]
- Basso K, Saito M, Sumazin P, et al. Integrated biochemical and computational approach identifies BCL6 direct target genes controlling multiple pathways in normal germinal center B cells. *Blood*. 2010; 115:975–984. [PubMed: 19965633]
- Cadeiras M, von Bayern M, Sinha A, et al. Drawing networks of rejection - a systems biological approach to the identification of candidate genes in heart transplantation. *J Cell Mol Med*. 2011; 15:949–956. [PubMed: 20497491]
- Carro MS, Lim WK, Alvarez MJ, et al. The transcriptional network for mesenchymal transformation of brain tumors. *Nature*. 2010; 463:318–325. [PubMed: 20032975]
- Chaganti RS, Houldsworth J. Genetics and biology of adult human male germ cell tumors. *Cancer Res*. 2000; 60:1475–1482. [PubMed: 10749107]
- Chan YS, Göke J, Lu X, et al. A PRC2-dependent repressive role of PRDM14 in human embryonic stem cells and induced pluripotent stem cell reprogramming. *Stem Cells*. 2012; 31:682–692. [PubMed: 23280602]
- Chazaud C, Yamanaka Y, Pawson T, et al. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev Cell*. 2006; 10:615–624. [PubMed: 16678776]
- Chen X, Xu H, Yuan P, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*. 2008; 33:1106–1117. [PubMed: 18555785]
- Chia NY, Chan YS, Feng B, et al. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature*. 2010; 468:316–320. [PubMed: 20953172]
- Della Gatta G, Palomero T, Perez-Garcia A, et al. Reverse engineering of TLX oncogenic transcriptional networks identifies RUNX1 as tumor suppressor in T-ALL. *Nat Med*. 2012; 18:436–440. [PubMed: 22366949]

- Fischedick G, Klein DC, Wu G, et al. Zfp296 is a novel, pluripotent-specific reprogramming factor. *PLoS One*. 2012; 7:e34645. [PubMed: 22485183]
- Gao H, Wu X, Sun, Y, et al. Suppression of homeobox transcription factor VentX promotes expansion of human hematopoietic stem/multipotent progenitor cells. *J Biol Chem*. 2012; 287:29979–29987. [PubMed: 22791709]
- Grabole N, Tischler J, Hackett JA, et al. Prdm14 promotes germline fate and naive pluripotency by repressing FGF signaling and DNA methylation. *EMBO Rep*. 2013; 14:629–637. [PubMed: 23670199]
- Gautier L, Cope L, Bolstad BM, et al. affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. 2004; 3:307–315. [PubMed: 14960456]
- Houldsworth J, Heath SC, Bosl GJ, et al. Expression profiling of lineage differentiation in pluripotential human embryonal carcinoma cells. *Cell Growth Differ*. 2002; 13:257–264. [PubMed: 12114215]
- Ivanova N, Dobrin R, Lu R, et al. Dissecting self-renewal in stem cells with RNA interference. *Nature*. 2006; 442:533–538. [PubMed: 16767105]
- Jaenisch R, Young R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell*. 2008; 132:567–582. [PubMed: 18295576]
- Ji Q, Fischer LA, Brown RC, et al. Engineered zinc-finger transcription factors activate OCT4 (POU5F1), SOX2, KLF4, c-MYC (MYC) and miR302/367. *Nucleic Acids Res*. 2014; 42:6158–6167. [PubMed: 24792165]
- Jia J, Zheng X, Hu G, et al. Regulation of pluripotency and self-renewal of ESCs through epigenetic-threshold modulation and mRNA pruning. *Cell*. 2012; 151:576–589. [PubMed: 23101626]
- Juárez-Moreno K, Erices R, Beltran AS, et al. Breaking through an epigenetic wall: re-activation of Oct4 by KRAB-containing designer zinc finger transcription factors. *Epigenetics*. 2013; 8:164–76. [PubMed: 23314702]
- Jung M, Peterson H, Chavez L, et al. A data integration approach to mapping OCT4 gene regulatory networks operative in embryonic stem cells and embryonal carcinoma cells. *PLoS One*. 2010; 5:e10709. [PubMed: 20505756]
- Korkola JE, Houldsworth J, Dobrzynski D, et al. Gene expression-based classification of nonseminomatous male germ cell tumors. *Oncogene*. 2005; 24:5101–5107. [PubMed: 15870693]
- Korkola JE, Houldsworth J, Chadalavada RS, et al. Down-regulation of stem cell genes, including those in a 200-kb gene cluster at 12p1331, is associated with in vivo differentiation of human male germ cell tumors. *Cancer Res*. 2006; 66:820–827. [PubMed: 16424014]
- Korkola JE, Houldsworth J, Feldman DR, et al. Identification and validation of a gene expression signature that predicts outcome in adult men with germ cell tumors. *J Clin Oncol*. 2009; 27:5240–5247. [PubMed: 19770384]
- Lefebvre C, Rajbhandari P, Alvarez MJ, et al. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol Syst Biol*. 2010; 6:377–386. [PubMed: 20531406]
- Lim WK, Lyashenko E, Califano A. Master regulators used as breast cancer metastasis classifier. *Pac Symp Biocomp*. 2009; 14:504–515.
- Liu G, Loraine AE, Shigeta R, et al. NetAffx: Affymetrix probesets and annotations. *Nucleic Acids Res*. 2003; 31:82–86. [PubMed: 12519953]
- Loh KM, Lim B. A precarious balance: pluripotency factors as lineage specifiers. *Cell Stem Cell*. 2011; 8:363–369. [PubMed: 21474100]
- Loh YH, Wu Q, Chew JL, et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet*. 2006; 4:431–440. [PubMed: 16518401]
- Ma Z, Swigut T, Valouev A, et al. Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. *Nat Struct Mol Biol*. 2011; 18:120–127. [PubMed: 21183938]
- Mani KM, Lefebvre C, Wang K, et al. A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. *MolSyst Biol*. 2008; 4:169–177.
- Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst*. 1959; 22:719–748. [PubMed: 13655060]

- Margolin AA, Wang K, Lim WK, et al. Reverse engineering cellular networks. *Nat Protoc.* 2006; 1:662–671. [PubMed: 17406294]
- Margolin AA, Palomero T, Sumazin P, et al. ChIP-on-chip significance analysis reveals large-scale binding and regulation by human transcription factor oncogenes. *Proc Natl Acad Sci USA.* 2009; 106:244–249. [PubMed: 19118200]
- Marson A, Levine SS, Cole MF, et al. Connecting micro RNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell.* 2008; 134:521–533. [PubMed: 18692474]
- Mitsui K, Tokuzawa Y, Itoh H, et al. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell.* 2003; 113:631–642. [PubMed: 12787504]
- Nishimoto M, Katano M, Yamagishi T, et al. In vivo function and evolution of the eutherian-specific pluripotency marker UTF1. *PLoS One.* 2013; 8:e68119. [PubMed: 23874519]
- Nishioka N, Yamamoto S, Kiyonari H, et al. Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mech Dev.* 2008; 125:270–283. [PubMed: 18083014]
- Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet.* 2000; 24:372–376. [PubMed: 10742100]
- Oh S, Shin S, Janknecht R. ETV1, 4 and 5: an oncogenic subfamily of ETS transcription factors. *Biochim Biophys Acta.* 2012; 1826:1–12. [PubMed: 22425584]
- Okuda A, Fukushima A, Nishimoto M, et al. UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells. *EMBO J.* 1998; 17:2019–2032. [PubMed: 9524124]
- Palomero T, Lim WK, Odom DT, et al. NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. *Proc Natl Acad Sci USA.* 2006; 103:18261–18266. [PubMed: 17114293]
- Ravasi T, Suzuki H, Cannistraci CV, et al. An atlas of combinatorial transcriptional regulation in mouse and man. *Cell.* 2010; 140:744–752. [PubMed: 20211142]
- Pearson K. "On Lines and Planes of Closest Fit to Systems of Points in Space" (PDF. *Philosophical Magazine.* 1901; 2:559–572.
- Scerbo P, Girardot F, Vivien C, et al. Ventx factors function as Nanog-like guardians of developmental potential in *Xenopus*. *PLoS One.* 2012; 7:e36855. [PubMed: 22606298]
- Schmidt D, Wilson MD, Spyrou C, et al. ChIP-seq: Using high-throughput sequencing to discover protein–DNA interactions. *Methods.* 2009; 48:240–248. [PubMed: 19275939]
- Schuettengruber B, Cavalli G. Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development.* 2009; 136:3531–3542. [PubMed: 19820181]
- Shu J, Wu C, Wu Y, et al. Induction of pluripotency in mouse somatic cells with lineage specifiers. *Cell.* 2013; 153:963–975. [PubMed: 23706735]
- Sperger JM, Chen X, Draper JS, et al. Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc Natl Acad Sci USA.* 2003; 100:13350–13355. [PubMed: 14595015]
- Subramanian A, Tamayo P, Mootha V, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005; 102:15545–15550. [PubMed: 16199517]
- Sumazin P, Yang X, Chiu HS, et al. An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell.* 2011; 147:370–381. [PubMed: 22000015]
- Ulbright TM. Germ cell neoplasms of the testis. *Am J Surg Pathol.* 1993;11075–11091.
- van den Berg DL, Snoek T, Mullin NP, et al. An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell.* 2010; 6:369–381. [PubMed: 20362541]
- Wang Z, Oron E, Nelson B, et al. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell.* 2012; 10:440–454. [PubMed: 22482508]
- Yamaji M, Ueda J, Hayashi K, et al. PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells. *Cell Stem Cell.* 2013; 12:368–382. [PubMed: 23333148]

- Ye S, Li P, Tong C, et al. Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1. *EMBO J.* 2013; 32:2548–2560. [PubMed: 23942238]
- Young RA. Control of embryonic stem cell state. *Cell.* 2011; 144:940–954. [PubMed: 21414485]
- Zhao X, D'Arca D, Lim WK, et al. The N-Myc-DLL3 cascade is suppressed by the ubiquitin ligase Huwe1 to inhibit proliferation and promote neurogenesis in the developing brain. *Dev Cell.* 2009; 17:210–221. [PubMed: 19686682]
- Zhao Y, Yin X, Qin H, et al. Two supporting factors greatly improve the efficiency of human iPSC generation *Cell Stem Cell.* 2008; 3:475–479.

Figure 1. (A). Unique and interacting targets of *POU5F1*, *NANOG*, and *SOX2* identified by ARACNe (see Table S2A-B). (B), (C) and (D) GO analysis of *POU5F1*, *NANOG* and *SOX2* target genes. Blue bars represent observed percentage of the genes for the given TF in a given GO category. Red bars represent the percentage of genes in the whole genome in a given GO category. Significance (P-value) of enrichment is based on a Fischer Exact Test (also see Table S2 C-E).

Figure 2. Silencing of *POU5F1*, *NANOG* and *SOX2* in NT2/D1 EC cells
(A), (B) and (C) GSEA of *POU5F1*, *NANOG* and *SOX2* target genes on the differential expression of genes in the *POU5F1*, *NANOG* and *SOX2* knockdown cells respectively when compared to controls. The bar code plot indicates the position of target genes. Genes with largest $-\log_{10}$ (p-value) are on the left and those with a value of 0 are on the right). The y-axis is the GSEA enrichment score. The curve represents a running weighted (weight = 1) cumulative enrichment score. The red vertical bars show position of genes in set S1. The small black bars above the red bars show positions of ChIP-PCR validated ARACNe targets. The grey scale bar at the top shows deeper shades of grey for larger local change in enrichment score. The title contains the GSEA Normalized Enrichment Score and the GSEA p-value for the enrichment of set S1 in set S2. **(D)** and **(E)** GO enrichment of downregulated and upregulated targets of *POU5F1*. **(F)** and **(G)**. GO enrichment of downregulated and upregulated targets of *NANOG*. **(H)** and **(I)**. GO enrichment of downregulated and upregulated targets of *SOX2*. The blue bars represent observed percentage of targets genes in a particular GO category. The red bars represent the percentage of genes in the whole genome in a particular GO category (as annotated by GO). The significance (p-value) of this enrichment is based on a Fischer Exact Test (also see Table S3 and Figure S2).

Figure 3. ANOVA and PC analysis of silencing time series of *POU5F1*, *NANOG* and *SOX2* in NT2/D1 EC cells

(A). ANOVA genes (with 0.05 p-value cut-off) in silencing time series *POU5F1*, *NANOG* and *SOX2* in NT2/D1 EC cells. (C), (D) and (E). GSEA of *POU5F1*, *NANOG* and *SOX2* ARACNe target genes on the all the genes in the genome ranked by their ANOVA score in their respective silencing time series in NT2/D1 EC cells. The bar code plot indicates the position of respective ARACNe target genes. (B). Overlapping categories from GO of the intersection of ARACNe targets and ANOVA genes of *POU5F1*, *NANOG* and *SOX2* silencing time series. (F-G). Principal component contribution in silencing time series of *POU5F1*, *NANOG*, and *SOX2*. The ANOVA F-test p-value is also shown. (H). Scatter-plot of the 2 main principal components in the PCA. (see Table S3B-E).

Figure 4.

Layout of all the TFs in the genome based on the overlap of their targets (quantified by an odds-ratio of overlap) . TFs placed closer have a larger odds-ratio of overlap of their targets. The ES cell state regulator TFs are represented by large circles colored in green. (see Table S5)

Figure 5. Immunofluorescence staining showing lineage development in (A) *POU5F1* and (B) *ZNF208* silenced H9 ES cells. Staining with DAPI, specific lineage marker primary antibodies PAX6 (neuronal), HAND1 (Trophoectodermal), NODAL(Endodermal), KRT7(Ectodermal), and MSX1(mesodermal) and the merged staining respectively, is shown. Scale bar for panel a and b in A and B is 100μm and for panel c, d and e in A and B is 50μm. (see Figure S5)