Cell

Resource

Circuitry and Dynamics of Human Transcription Factor Regulatory Networks

Shane Neph,[1,5] Andrew B. Stergachis,[1,5] Alex Reynolds,[1] Richard Sandstrom,[1] Elhanan Borenstein,[1,2,4,*] and John A. Stamatoyannopoulos[1,3,*]

[1]Department of Genome Sciences
[2]Department of Computer Science and Engineering
[3]Department of Medicine
University of Washington, Seattle, WA 98195, USA
[4]Santa Fe Institute, Santa Fe, NM 87501, USA
[5]These authors contributed equally to this work
*Correspondence: elbo@uw.edu (E.B.), jstam@uw.edu (J.A.S.)
http://dx.doi.org/10.1016/j.cell.2012.04.040

SUMMARY

The combinatorial cross-regulation of hundreds of sequence-specific transcription factors (TFs) defines a regulatory network that underlies cellular identity and function. Here we use genome-wide maps of in vivo DNaseI footprints to assemble an extensive core human regulatory network comprising connections among 475 sequence-specific TFs and to analyze the dynamics of these connections across 41 diverse cell and tissue types. We find that human TF networks are highly cell selective and are driven by cohorts of factors that include regulators with previously unrecognized roles in control of cellular identity. Moreover, we identify many widely expressed factors that impact transcriptional regulatory networks in a cell-selective manner. Strikingly, in spite of their inherent diversity, all cell-type regulatory networks independently converge on a common architecture that closely resembles the topology of living neuronal networks. Together, our results provide an extensive description of the circuitry, dynamics, and organizing principles of the human TF regulatory network.

INTRODUCTION

Sequence-specific transcription factors (TFs) are the key effectors of eukaryotic gene control. Human TFs regulate hundreds to thousands of downstream genes (Johnson et al., 2007). Of particular interest are interactions in which a given TF regulates other TFs, or itself. Such mutual cross-regulation among groups of TFs defines regulatory subnetworks that underlie major features of cellular identity and complex functions such as pluripotency (Boyer et al., 2005; Kim et al., 2008), development (Davidson et al., 2002a), and differentiation (Yun and Wold, 1996). On a broader level, cross-regulatory interactions among the entire complement of TFs expressed in a given cell type form a core transcriptional regulatory network, endowing the cell with systems-level properties that facilitate the integration of complex cellular signals, while conferring additional nimbleness and robustness (Alon, 2006). However, despite their central biological roles, both the structure of core human regulatory networks and their component subnetworks are largely undefined.

One of the main bottlenecks limiting generation of TF regulatory networks for complex biological systems has been that information is traditionally collected from individual experiments targeting one cell type and one TF at a time (Davidson et al., 2002a; Yuh et al., 1994; Kim et al., 2008; Roy et al., 2010; Gerstein et al., 2010). For example, the sea urchin endomesoderm regulatory network was constructed by individually perturbing the expression and activity of several dozen TFs and analyzing the effect of these perturbations on the expression of TF genes containing putative *cis*-regulatory binding elements for these factors (Davidson et al., 2002b; Yuh et al., 1994). More recently, genome-wide analysis combining chromatin immunoprecipitation of individual TFs with high-throughput sequencing (ChIP-seq) has been used to derive subnetworks of small numbers of TFs, such as those involved in pluripotency (Kim et al., 2008), or larger-scale networks combining several dozen TFs (Roy et al., 2010; Gerstein et al., 2010). However, such approaches are limited by three major factors: (1) the availability of suitable affinity reagents; (2) the difficulty of interrogating the activities of multiple TFs within the same cellular environment; and, perhaps most critically, (3) the sizable number of TFs and cellular states that need to be studied. De novo network construction methods based on gene expression correlations partly overcome the limitation of studying one TF at a time but lack directness and typically require several hundred independent gene expression perturbation studies to build a network for one cell type (Basso et al., 2005; Carro et al., 2010). Similarly, yeast one-hybrid assays offer a high-throughput approach for identifying *cis*-regulatory element binding partners (Walhout,

2006; Reece-Hoyes et al., 2011). However, such assays lack native cellular context, limiting their direct utility for building cell-type-specific networks. Given these experimental limitations, only a handful of well-described multicellular transcriptional regulatory networks have been defined, and those that do exist are often incomplete despite the numerous experiments and extended time (typically years) needed to construct them (Davidson et al., 2002a; Basso et al., 2005; Boyer et al., 2005; Kim et al., 2008; Roy et al., 2010; Gerstein et al., 2010).

Given that the human genome encodes > 1,000 TFs (Vaquerizas et al., 2009) and that human cellular diversity spans hundreds of different cell types and an even greater number of cellular states, we sought to develop an accurate and scalable approach to analyze transcriptional regulatory networks suitable for the application to any cellular or organismal state. The discovery of DNaseI footprinting over 30 years ago (Galas and Schmitz, 1978) revolutionized the analysis of regulatory sequences in diverse organisms and directly enabled the discovery of the first human sequence-specific TFs (Dynan and Tjian, 1983). In the context of living nuclear chromatin, DNaseI treatment preferentially cleaves the genome within highly accessible active regulatory DNA regions, creating DNaseI-hypersensitive sites (DHSs) (Wu et al., 1979; Kuo et al., 1979; Wu, 1980; Stalder et al., 1980). Within DHSs, DNaseI cleavage is not uniform but is rather punctuated by sequence-specific regulatory factors that occlude bound DNA, leaving "footprints" that demarcate TF occupancy at nucleotide resolution (Hesselberth et al., 2009; Pfeifer and Riggs, 1991). DNaseI footprinting is a well-established method for identifying direct regulatory interactions and provides a powerful generic approach for assaying the occupancy of specific sequence elements with *cis*-regulatory functions (Karin et al., 1984; Kadonaga et al., 1987).

DNaseI footprinting has been applied widely to study regulatory interactions between TFs and to identify cell- and lineage-selective transcriptional regulators (Dynan and Tjian, 1983; Karin et al., 1984; Tsai et al., 1989). In the context of the ENCODE Project, we applied digital genomic footprinting (Hesselberth et al., 2009) to delineate millions of human DNaseI footprints genome-wide in 41 diverse cell types. Combining DNaseI footprints with defined TF recognition sequences accurately and quantitatively recapitulates ChIP-seq data for individual TFs, while simultaneously interrogating the genomic occupancy of potentially all expressed DNA-binding factors in a single experiment (Neph et al., 2012a).

By performing systematic analysis of TF footprints in the proximal regulatory regions of each TF gene, we develop a foundational experimental paradigm for comprehensive, unbiased mapping of the complex network of regulatory interactions between human TFs. In such networks, TFs comprise the network "nodes," and the cross-regulation of one TF by another the interactions or network "edges." Furthermore, iterating this paradigm across diverse cell types provides a powerful system for analysis of TF network dynamics in a complex organism. Here, we use genome-wide maps of in vivo DNaseI footprints to assemble an extensive core human regulatory network comprising connections among 475 sequence-specific TFs and analyze the dynamics of these connections across 41 diverse cell and tissue types.

RESULTS

Comprehensive Mapping of TF Networks in Diverse Human Cell Types

To generate TF regulatory networks in human cells, we analyzed genomic DNaseI footprinting data from 41 diverse cell and tissue types (Neph et al., 2012a). Each of these 41 samples was treated with DNaseI, and sites of DNaseI cleavage along the genome were analyzed with high-throughput sequencing. At an average sampling depth of ~500 million DNaseI cleavages per cell type (of which ~273 million mapped to unique genomic positions), we identified an average of ~1.1 million high-confidence DNaseI footprints per cell type (range 434,000 to 2.3 million at a false discovery rate of 1% [FDR 1%]; Neph et al., 2012a). Collectively, we detected 45,096,726 footprints, representing cell-selective binding to ~8.4 million distinct 6–40 bp genomic sequence elements. We used well-annotated databases of TF-binding motifs to infer the identities of factors occupying DNaseI footprints (Wingender et al., 1996; Bryne et al., 2008; Newburger and Bulyk, 2009) (Experimental Procedures) and confirmed that these identifications matched closely and quantitatively with ENCODE ChIP-seq data for the same cognate factors (Neph et al., 2012a).

To generate a TF regulatory network for each cell type, we analyzed actively bound DNA elements within the proximal regulatory regions (i.e., all DNaseI hypersensitive sites within a 10 kb interval centered on the transcriptional start site [TSS]) of 475 TF genes with well-annotated recognition motifs (Wingender et al., 1996; Bryne et al., 2008; Newburger and Bulyk, 2009) (Figure 1A). Repeating this process for every cell type disclosed a total of 38,393 unique, directed (i.e., TF-to-TF) regulatory interactions (edges) among the 475 analyzed TFs, with an average of 11,193 TF-to-TF edges per cell type (Data S1). Given the functional redundancy of a minority of DNA-binding motifs (Berger et al., 2008), in certain cases multiple factors could be designated as occupying a single DNaseI footprint. However, most commonly, mappings represented associations between single TFs and a specific DNA element. Because DNaseI hypersensitivity at proximal regulatory sequences closely parallels gene expression (The ENCODE Project Consortium, 2012), the annotation process we utilized naturally focuses on the expressed TF complement of each cell type, enabling the construction of a comprehensive transcription regulatory network for a given cell type with a single experiment.

De Novo-Derived Networks Accurately Recapitulate Known TF-to-TF Circuitry

To assess the accuracy of cellular TF regulatory networks derived from DNaseI footprints, we analyzed several well-annotated mammalian cell-type-specific transcriptional regulatory subnetworks (Figures 1B and 1C). The muscle-specific factors MyoD, Myogenin (MYOG), MEF2A, and MYF6 form a network that was uncovered using a combination of genetic and physical studies, including DNaseI footprinting, and is vital for specification of skeletal muscle fate and control of myogenic development and differentiation (Naidu et al., 1995; Yun and Wold, 1996; Ramachandran et al., 2008). Figure 1B juxtaposes the known regulatory interactions between these factors determined

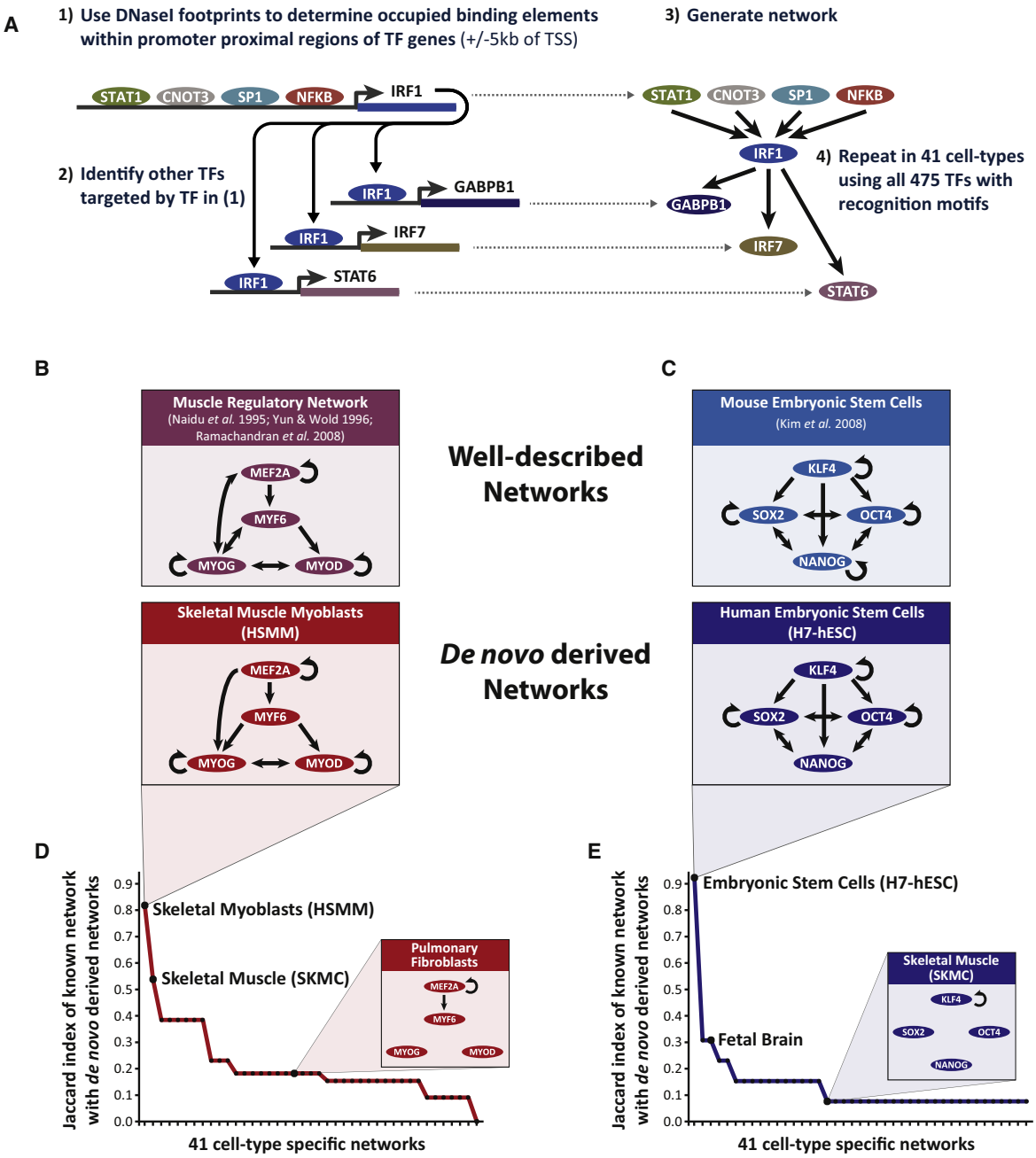

Figure 1. Construction of Comprehensive Transcriptional Regulatory Networks

(A) Schematic for construction of regulatory networks using DNaseI footprints. Transcription factor (TF) genes represent network nodes. Each TF node has regulatory inputs (TF footprints within its proximal regulatory regions) and regulatory outputs (footprints of that TF in the regulatory regions of other TF genes). Inputs and outputs comprise the regulatory network interactions "edges." For example: (1) In Th1 cells, the *IRF1* promoter contains DNaseI footprints matching four regulatory factors (STAT1, CNOT3, SP1, and NFKB). (2) In Th1 cells, IRF1 footprints are found upstream of many other genes (for example, *GABP1*, *IRF7*, *STAT6*). (3) The same process is iterated for every TF gene in that cell type, enabling compilation of a cell-type network comprising nodes (TF genes) and edges (regulatory inputs and outputs of TF genes). (4) Network construction is carried out independently using DNaseI footprinting data from each of 41 cell types, resulting in 41 independently derived cell-type networks.

(B and C) Comparison of well-annotated versus de novo-derived regulatory subnetworks.

(B) Muscle subnetwork. Top, experimentally defined regulatory subnetwork for major factors controlling skeletal muscle differentiation and transcription. Arrows indicate direction(s) of regulatory interactions between factors. Bottom, regulatory subnetwork derived de novo from the DNaseI footprint-anchored network of skeletal myoblasts closely matches the experimentally annotated network.

(C) Pluripotency subnetwork. Top, regulatory subnetwork for major pluripotency factors defined experimentally in mouse ESCs (Kim et al., 2008). Bottom, regulatory subnetwork derived de novo from human ESCs is virtually identical to the annotated network.

in the aforementioned studies (Figure 1B, top) with the nearly identical interactions derived de novo from analysis of the network computed using DNaseI footprints mapped in primary human skeletal myoblasts (HSMM) (Figure 1B, bottom).

OCT4, NANOG, KLF4, and SOX2 together play a defining role in maintaining the pluripotency of embryonic stem cells (ESCs) (Takahashi and Yamanaka, 2006; Takahashi et al., 2007), and a network comprising the mutual regulatory interactions between these factors has been mapped through systematic studies of factor occupancy by ChIP-seq in mouse ESCs (Kim et al., 2008) (Figure 1C, top). A nearly identical subnetwork emerges from analysis of the TF network computed de novo from DNaseI footprints in human ESCs (Figure 1C, bottom).

Critically, both the well-annotated muscle and ES subnetworks are best matched by footprint-derived networks computed specifically from skeletal myoblasts and human ESCs, respectively, versus other cell types (Figures 1D and 1E). These findings indicate that network relationships between TFs derived de novo from genomic DNaseI footprinting accurately recapitulate well-described cell-type-selective transcriptional regulatory networks generated with multiple experimental approaches.

TF Regulatory Networks Show Marked Cell Selectivity

We next analyzed systematically the dynamics of TF regulatory networks across cell types. Four hundred and seventy-five TFs theoretically have the potential for 225,625 combinations of TF-to-TF regulatory interactions (or network edges). However, only a fraction of these potential edges are observed in each cell type (~5%), and most are unique to specific cell types (Figure S1A available online).

To visualize the global landscape of cell-selective versus shared regulatory interactions, we first computed the broad landscape of network edges that are either specific to a given cell type or found in networks of two or more cell types (Figure 2; Table S1). This revealed that regulatory interactions were in general highly cell selective, though the proportion of cell-selective interactions varied from cell type to cell type. Network edges were most frequently restricted to a single cell type, and collectively the majority of edges were restricted to four or fewer cell types (Figure S1A). By contrast, only 5% of edges were common to all cell types (Figure S1A). Interestingly, when comparing networks, we found more common edges than common DNaseI footprints (Figures S1B and S1C), implying that a given transcriptional regulatory interaction can be generated using distinct DNA-binding elements in different cell types.

To explore the regulatory interaction dynamics of limited sets of related factors, we plotted the regulatory network edges connecting four hematopoietic regulators and four pluripotency regulators in six diverse cell types (Figure 3A). This analysis clearly highlighted the role of cell-type-specific factors within their cognate cell types: regulatory interactions between pluripotency factors within the ESC network and hematopoietic factors within the network of hematopoietic stem cells (Figure 3A). Next, we plotted the complete set of regulatory interactions among all 475 edges between the same six diverse cell types, exposing a high degree of regulatory diversity (Figure 3B; Table S1).

Edges unique to a cell type typically form a well-connected subnetwork (Figures S1D–S1F; Table S2), implying that cell-type-specific regulatory differences are not driven merely by the independent actions of a few TFs but rather by organized TF subnetworks. In addition, the density of cell-selective networks varies widely between cell types (e.g., compare ESCs to skeletal myoblasts in Figure 3B). These observations underscore the importance of using cell-type-specific regulatory networks when addressing specific biological questions.

Functionally Related Cell Types Share Similar Core Transcriptional Regulatory Networks

We next sought to determine the degree of relatedness between different TF networks. To obtain a quantitative global summary of the factors contributing to each cell-type-specific network, we computed for each cell type the normalized network degree (NND)—a vector that encapsulates the relative number of interactions observed in that cell type for each of the 475 TFs (Alon, 2006). To capture the degree to which different cell-type networks utilize similar TFs, we clustered all cell-type networks based on their NND vector (Figure 4A). The resulting network clusters—obtained from an unbiased analysis—strikingly parallel both anatomical and functional cell-type groupings into epithelial and stromal cells; hematopoietic cells; endothelia; and primitive cells including fetal cells and tissues, ESCs, and malignant cells with a ''dedifferentiated'' phenotype (Figure 4A; compare the manually curated groupings in Figure 2). This result suggests that transcriptional regulatory networks from functionally similar cell types are governed by similar factors. Furthermore, this result suggests a framework for understanding how minor perturbations in network composition might enable transdifferentiation among related cell types (Graf and Enver, 2009).

To identify the individual TFs driving the clustering of related cell-type networks, we computed the relative NND (i.e., the normalized number of connections) of each TF across the 41 cell types. This approach uncovered numerous specific factors with highly cell-selective interaction patterns, including known regulators of cellular identity important to functionally related cell types (Figure 4B). For instance, PAX5 is most highly connected in B cell regulatory networks, concordant with its function as a major regulator of B-lineage commitment (Nutt et al., 1999). Similarly, the neuronal developmental regulator POU3F4 (Shimazaki et al., 1999) plays a prominent role specifically in hippocampal astrocyte and fetal brain regulatory networks, whereas the cardiac developmental regulator GATA4 (Molkentin et al., 1997)

(D and E) De novo-derived subnetworks in (B) and (C) match the annotated networks in a cell-specific fashion. Vertical axes: Jaccard index, a measure of network similarity, comparing the annotated subnetwork with regulatory interactions between the four factors derived de novo from each of 41 cell types independently (horizontal axes). For the annotated muscle subnetwork, the highest similarity is seen in skeletal myoblasts, followed by differentiated skeletal muscle. By contrast, subnetworks computed from fibroblasts are largely devoid of relevant interactions. For the annotated pluripotency subnetwork, the highest similarity is seen in human ESCs (H7-ESC).

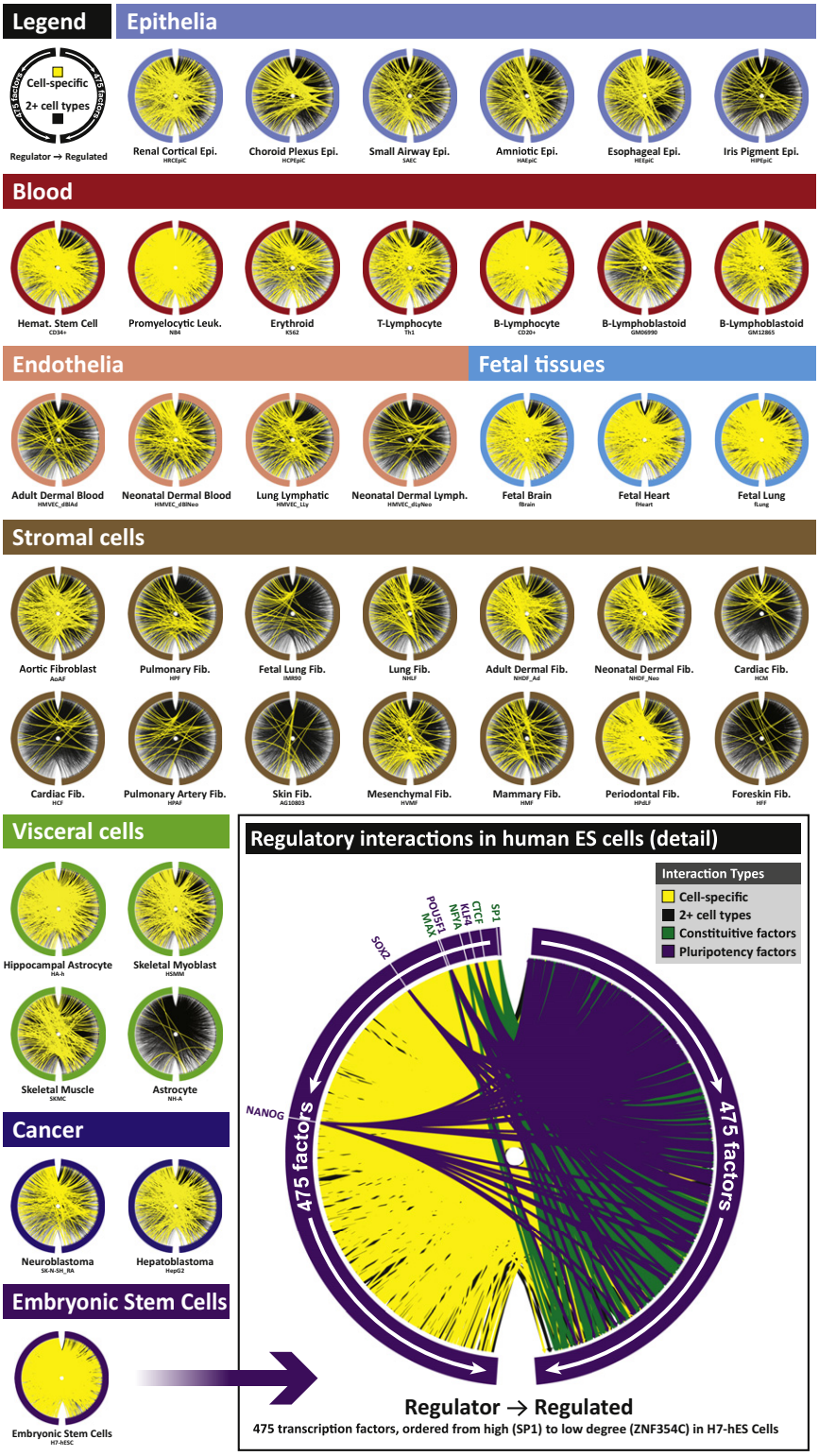

Figure 2. Cell-Specific versus Shared Regulatory Interactions in TF Networks of 41 Diverse Cell Types

Shown for each of 41 cell types are schematics of cell-type-specific (yellow) versus -nonspecific (black) regulatory interactions between 475 TFs. Each half of each circular plot is divided into 475 points (not visible at this scale), one for each TF. Lines connecting the left and right half-circles represent regulatory interactions between each factor and any other factors with which it interacts in the given cell type. Yellow lines represent TF-to-TF connections that are specific to the indicated cell type. Black lines represent TF-to-TF connections that are seen in two or more cell types. The order of TFs along each half-circular axis is shown in Table S1 and represents a sorted list (descending order) of their degree (i.e., number of connections to other TFs) in the ESC network, from highest degree on top (SP1) to lowest degree on bottom (ZNF354C). Cell types are grouped based on their developmental and functional properties. Insert on bottom right shows a detailed view of the human ESC network and highlights the interactions of four pluripotent (KLF4, NANOG, POU5F1, SOX2) and four constitutive factors (SP1, CTCF, NFYA, MAX) with purple and green edges, respectively.

shows the highest relative network degree in cardiac and great vessel tissue (fetal heart, cardiomyocytes, cardiac fibroblasts, and pulmonary artery fibroblasts).

In addition to these known developmental regulators, the network analysis implicated many regulators with previously unrecognized roles in specification of cell identity. For instance, HOXD9 is highly connected specifically in endothelial regulatory networks, and the early developmental regulator GATA5 (MacNeill et al., 2000) appears to play a predominant role in the fetal lung network (Figure 4B), providing functional insight into the role of GATA5 as a lung tissue biomarker (Xing et al., 2010). In addition to factors with strong cell-selective connectivity, we found a number of TFs with prominent roles in all 41 cell-type networks, including several known ubiquitous transcriptional and genomic regulators such as SP1, NFYA, CTCF, and MAX (Figure S2).

Together, the above results demonstrate the ability of transcriptional networks derived from genomic DNaseI footprinting to pinpoint known cell-selective and ubiquitous regulators of cellular state and to implicate analogous yet unanticipated roles for many other factors. It is notable that the aforementioned results were derived independently of gene expression data, highlighting the ability of a single experimental paradigm (genomic DNaseI

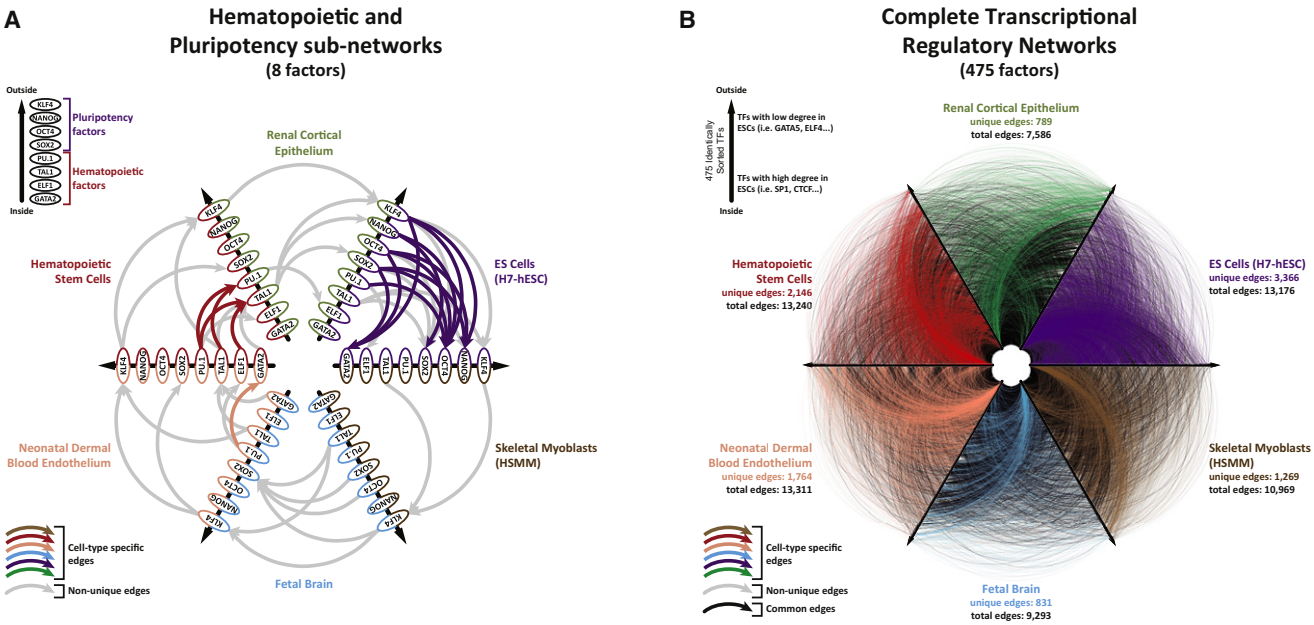

Figure 3. Transcriptional Regulatory Networks Show Marked Cell-Type Specificity

(A) Cross-regulatory interactions between four pluripotency factors and four hematopoietic factors in regulatory networks of six diverse cell types. All eight factors are arranged in the same order along each axis. Regulatory interactions (i.e., from regulat*or* to regulate*d*) are shown by arrows in clockwise orientation. Cell-type-specific edges are colored as indicated, whereas regulatory interactions present in two or more cell-type networks are shown in gray.

(B) Cross-regulatory interactions between all 475 TFs in regulatory networks of six diverse cell types. The 475 TFs are arranged in the same order along each axis, regulatory interactions directed clockwise. Edges unique to a given cell-type network are colored as indicated in the legend, whereas regulatory interactions present in two or more networks are colored gray. Interactions present in all six cell-type networks are colored black.

See also Figure S1 and Table S2.

footprinting) to elucidate multiple intricate transcriptional regulatory relationships.

Network Analysis Reveals Cell-Type-Specific Behaviors for Widely Expressed TFs

Many TFs are expressed to varying degrees in a number of different cell types (Vaquerizas et al., 2009). A major question is whether the function of widely expressed factors remains essentially the same in different cells, or whether such factors are capable of exhibiting important cell-selective actions. To explore this question, we sought to characterize the regulatory diversity between different cell types within the same lineage. Hematopoietic lineage cells have been extensively characterized at both the phenotypic and the molecular levels, and a cadre of major transcriptional regulators, including TAL1/SCL, PU.1, ELF1, HES1, MYB, GATA2, and GATA1, has been defined (Orkin, 1995; Swiers et al., 2006). Many of these factors are expressed to varying degrees across multiple hematopoietic lineages and their constituent cell types.

We analyzed de novo-derived subnetworks comprising the aforementioned seven regulators in five hematopoietic and one nonhematopoietic cell type (Figure 5A). For each cell-type subnetwork, we also mapped the normalized outdegree (i.e., the number of outgoing connections) for each factor (Figure 5A). This analysis revealed both subtle and stark differences in the organization of the seven-member hematopoietic regulatory subnetwork that reflected the biological origin of each cell type. For example, the early hematopoietic fate-decision factor PU.1 appears to play the largest role in the subnetworks generated from hematopoietic stem cells ($CD34^+$) and promyelocytic leukemia (NB4) cells (Figure 5A). The erythroid-specific regulator GATA1 appears as a strong driver of the core TAL1/PU.1/HES1/MYB subnetwork specifically within erythroid cells (Figure 5A), consistent with its defining role in erythropoiesis. In both B cells and T cells, the subnetwork takes on a directional character, with PU.1 in a superior position. By contrast, the subnetwork is largely absent in nonhematopoietic cells (muscle, HSMM) (Figure 5A, bottom right). These findings demonstrate that analysis of the network relationships of major lineage regulators provides a powerful tool for uncovering subtle differences in transcriptional regulation that drive cellular identity between functionally similar cell types.

We next extended this analysis to determine whether we could identify commonly expressed factors that manifest cell-type-specific behaviors. For example, the retinoic acid receptor-alpha (RAR-α) is a constitutively expressed factor involved in numerous developmental and physiological processes (Sucov et al., 1996). Rather than simply measuring the degree of connectivity of RAR-α to other factors across different cell types, we sought to quantify the behavior of RAR-α within each cellular regulatory network by determining its position within feedforward loops (FFLs). FFLs represent one of the most important network motifs in biological and regulatory systems and comprise a three-node structure in which information is propagated

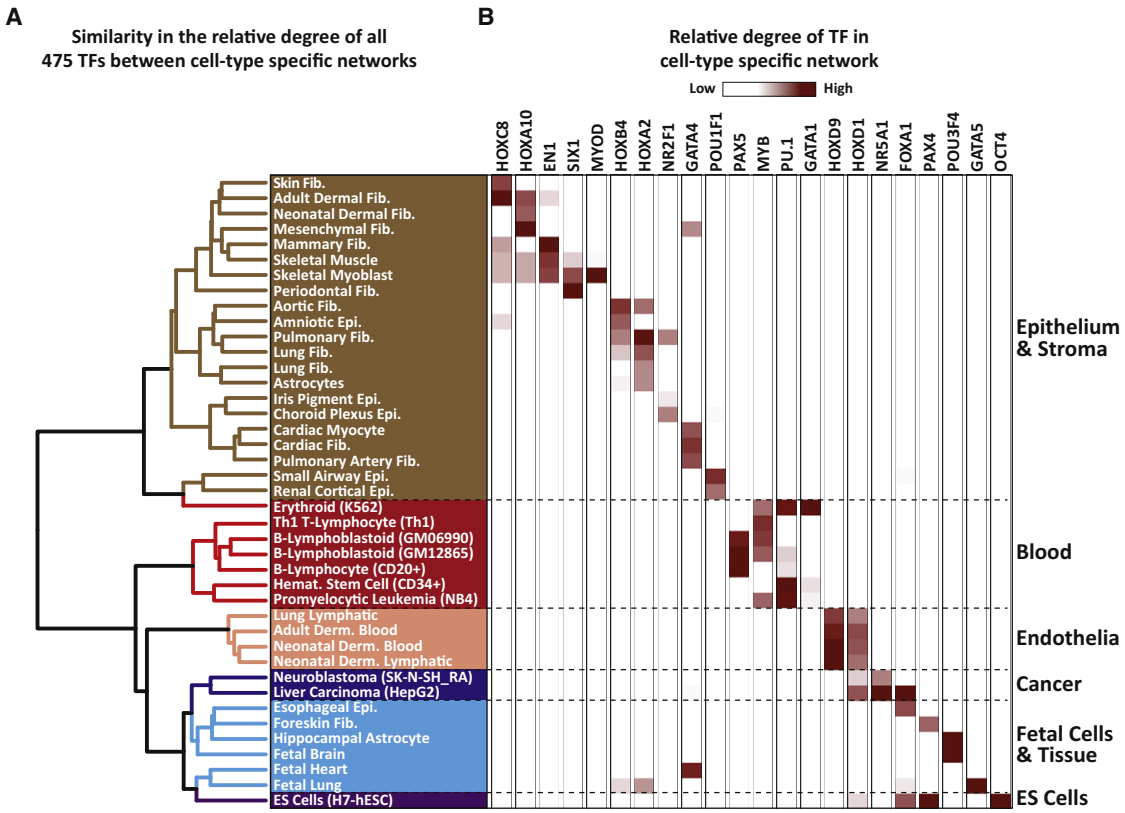

Figure 4. Functionally Related Cell Types Share Similar Core Transcriptional Regulatory Networks

(A) Clustering of cell-type networks by normalized network degree (NND). For each of 475 TFs within a given cell-type network, the relative number of edges was compared between all 41 cell types using a Euclidean distance metric and Ward clustering. Cell types are colored based on their physiological and/or functional properties.

(B) Relative degree of master regulatory TFs in cell-type networks. Shown is a heatmap representing the relative normalized degree of the indicated TFs between each of the 41 cell types. For a given TF and cell type, high relative degree indicates high connectivity with other TFs in that cell type. Note that the relative degree of known regulators of cell fate such as MYOD, OCT4, or MYB is highest in their cognate cell type or lineage. Similar patterns were found for other TFs without previously recognized roles in specification of cell identify.

See also Figure S2.

forward from the top node through the middle to the bottom node, with direct top node-to-bottom node reinforcement (Milo et al., 2002; Alon, 2006). For each cell type, we quantified the number of FFLs containing RAR-α at each of the three different positions (top versus middle versus bottom; Figure 5B, top). In most cell types, RAR-α chiefly participates in FFLs at "passenger" positions 2 and 3 (Figure 5B). However, within blood and endothelial cells, RAR-α switches from being a passenger to being a driver (top position) of FFLs. Strikingly, in acute promyelocytic leukemia (APL) cells, RAR-α acts as a uniquely potent driver of FFLs, occurring exclusively in the driver position—a feature unique among all cell types (Figure 5B). APL is characterized by an oncogenic t(15;17) chromosomal translocation that results in a RAR-α/PML fusion protein that misregulates RAR-α target sites (Grignani et al., 1993, 1998). Our results suggest that in APL cells, RAR-α is additionally altering the basic organization of the regulatory network. Critically, using DNaseI footprint-driven network analysis, we identified the prominent role of RAR-α in APL cells without any prior knowledge of the role of RAR-α in the oncogenic transformation of APL cells. This suggests that network analysis is capable of deriving vital pathogenic information about specific factors in abnormal cell types, given a sufficient analyzed spectrum of normal cellular networks. On a more general level, the aforementioned results show clearly that marked cell-selective functional specificities of commonly expressed proteins can be exposed by analyzing factors within the context of their peers.

The Common "Neural" Architecture of Human TF Regulatory Networks

Complex networks from diverse organisms are built from a set of simple building blocks termed network motifs (Milo et al., 2002). Network motifs represent simple regulatory circuits, such as the FFL described above. The topology of a given network can be reflected quantitatively in the normalized frequencies (normalized z-score) of different network motifs. Specific well-described motifs including FFL, "clique," "semi-clique," "regulated mutual," and "regulating mutual" are recurrently

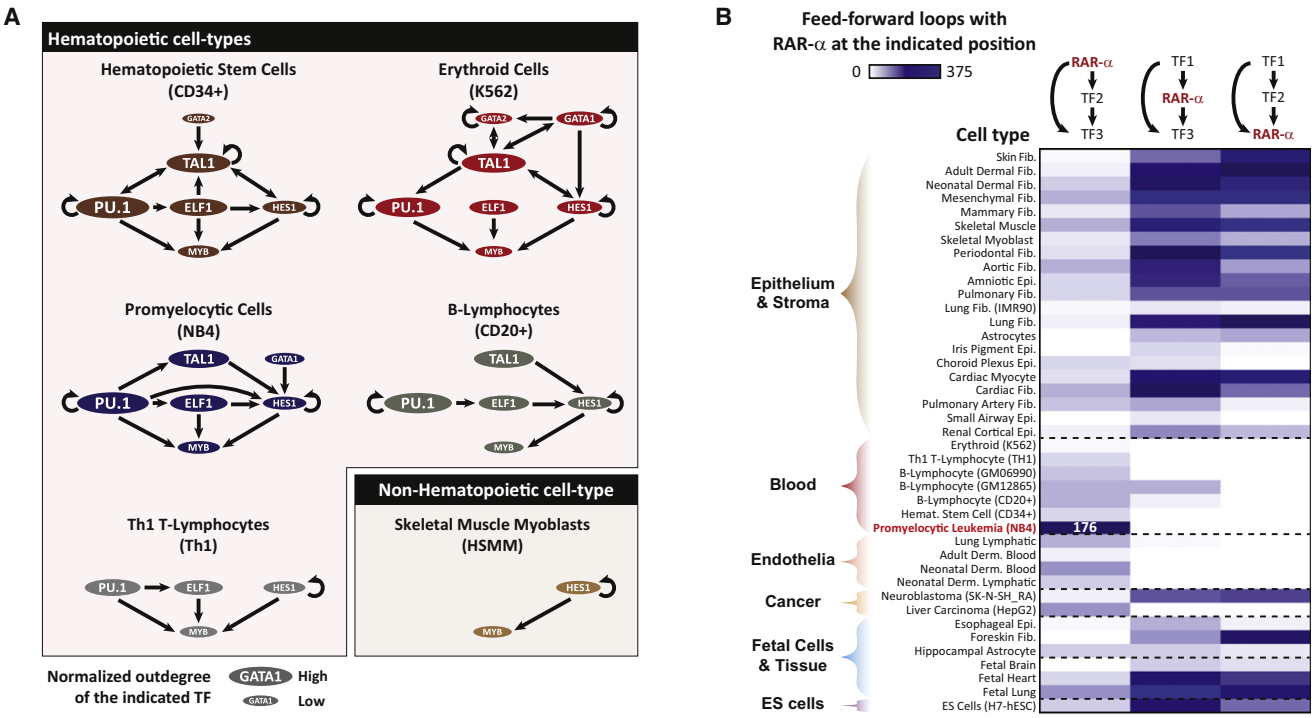

Figure 5. Cell-Selective Behaviors of Widely Expressed TFs

(A) Shown are regulatory subnetworks comprising edges (arrows) between seven major hematopoietic regulators in five hematopoietic and one non-hematopoietic cell types. For each TF, the size of the corresponding colored oval is proportional to the normalized out-degree (i.e., out-going regulatory interactions) of that factor within the complete network of each cell type. The early hematopoietic fate decision factor PU.1 appears to play the largest role in hematopoietic stem cells ($CD34^+$) and in promyelocytic leukemia (NB4) cells. The erythroid-specific regulator GATA1 appears as a strong driver of the core TAL1/PU.1/HES1/MYB network specifically within erythroid cells. In both B cells and T cells, the subnetwork takes on a directional character, with PU.1 in a superior position. By contrast, the network is largely absent in nonhematopoietic cells (muscle, HSMM, bottom right).

(B) Heatmap showing the frequency with which RAR-α is positioned as a driver (top) or passenger (middle or bottom) within FFLs mapped in 41 cell-type regulatory networks. Note that in most cell types, RAR-α participates in FFLs at "passenger" positions 2 and 3. However, within blood and endothelial cells, RAR-α switches from being a passenger of FFLs to being a driver (top position) of FFLs. In acute promyelocytic leukemia cells (NB4), RAR-α acts exclusively as a potent driver of FFLs. Cell types are arranged according to the clustered ordering in Figure 4.

found at higher than expected frequencies within diverse biological networks (Milo et al., 2002, 2004). We therefore sought to analyze the topology of the human TF regulatory network and to compare it with those of well-annotated multicellular biological networks.

We first computed the relative frequency and relative enrichment or depletion of each of the 13 possible three-node network motifs within each cell-type regulatory network. Next, we compared the results for each cell-type network with the relative enrichment of three-node network motifs found in perhaps the best annotated multicellular biological network, the *C. elegans* neuronal connectivity network (White et al., 1986). This comparison revealed striking similarity between the topologies of human TF networks and the *C. elegans* neuronal network (Figure 6A; Table S3). Remarkably, in spite of their cell selectivity, the topologies of each TF network were nearly identical. Notably, the human TF regulatory network topology also closely resembles that of other well-described networks, including the sea-urchin endomesoderm specification network (Davidson et al., 2002a), the *Drosophila* developmental transcriptional network (Serov et al., 1998), and the mammalian signal transduction network (Milo et al., 2004) (Figure S3A), consistent with universal principles for multicellular biological information processing systems (Milo et al., 2004).

To test the sensitivity of the above findings to the manner in which the human transcriptional regulatory networks were determined, we recomputed this network solely from scanned TF-binding sites within the promoter-proximal regions of each TF gene, without considering whether the motifs were localized within DNaseI footprints. Using this approach, the remarkable similarity of the footprint-derived TF networks to the neuronal network was almost completely lost (Figure 6B). This result affirms the criticality of in vivo footprints for biologically meaningful network inference.

Next, we sought to determine whether the observed similarity to the neuronal network was a collective property of human TF networks. To test this, we computed a transcriptional regulatory network from the combined regulatory interactions of all 41 cell types and determined the enrichment of network motifs within this network. The resulting network topology diverges considerably from that of the neuronal network (Figure 6C), far more so than was observed for any individual cell type (Figure 6A). This

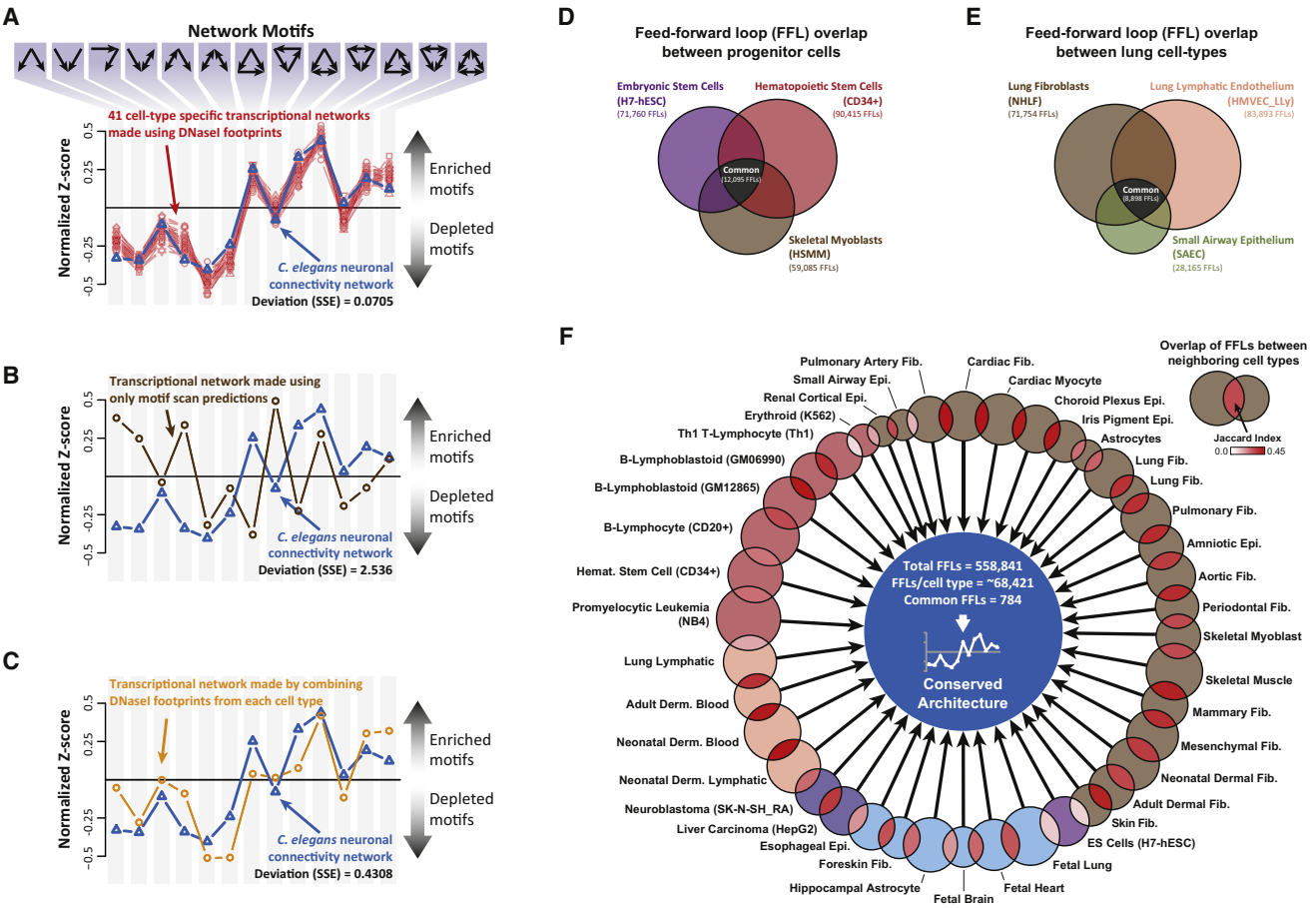

Figure 6. Conserved Architecture of Human TF Regulatory Networks

(A) Shown is the relative enrichment or depletion of the 13 possible three-node architectural network motifs within the regulatory networks of each cell type (red lines), compared with the relative enrichment of the same motifs in the *C. elegans* neuronal connectivity network. Note that the network architecture of each individual cell type closely mirrors that of the living neuronal network (average SSE of only 0.0705).

(B) Enrichment of each triad network motif for a TF network computed using only motif scan predictions within ± 5 kb of TF promoters (brown line). The resulting network bears little resemblance to the *C. elegans* network (blue line) (SSE of 2.536).

(C) The relative enrichment of different triad network motifs is shown for a TF regulatory network generated by pooling DNaseI footprints from all 41 tested cell types into a single archetype (orange line). The resulting topology diverges considerably from that of the neuronal network, far more so than was observed for any individual cell type (SSE of 0.4308).

(D and E) Network architectures are highly cell specific. (D) Overlap of FFLs identified in three different progenitor cell types—ESCs (H7-hESC), hematopoietic stem cells ($CD34^+$), and HSMM. Note that most FFLs are restricted to an individual cell type.

(E) Overlap of FFLs identified in three pulmonary cell types—lung fibroblasts (NHLF), small airway epithelium cells (SAECs), and pulmonary lymphatic endothelium cells (HMVEC_LLy). Highly distinct architectures are present even among cell types from the same organ structure.

(F) Overlap of FFLs from networks of neighboring cell types, following the ordering and coloration shown in Figure 4A. The size of each circle is proportional to the number of FFLs contained within the network of the corresponding cell type. The color of the intersection region between adjacent cell types indicates the Jaccard index between FFLs from those two cell types (see legend in upper right). The average number of FFLs in each network, the total number of FFLs across all networks, and the number of common FFLs across all networks are indicated in the center of the graph.

See also Figure S3 and Table S3.

result suggests that the regulatory interactions within each cell-type network are independently balanced to achieve a specific architecture, and that pooling multiple cellular networks together degrades this balance.

Finally, to assess whether a common core of regulatory interactions might be driving the conserved network architecture, we compared FFLs between biologically similar cell types. This comparison revealed marked diversity among different cellular TF networks (Figures 6D and 6E), going beyond that observed among individual edges (Figures S3C and S3D). Indeed, only ~0.1% of all observed FFLs across 41 cell types (784/558,841) were common to all cell types (Figures 6F and S3E). Moreover, only a minority of the TFs represented within a given cellular network contribute to the enriched network motifs (Figure S3F). These findings indicate that the conserved "neuronal" network architecture (Figure 6A) of the human TF regulatory network is

specified independently in each cell type using a distinct set of balanced regulatory interactions.

DISCUSSION

TF regulatory networks are foundational to biological systems. Collectively, our results highlight the power of regulatory networks derived from genomic DNaseI footprint maps to provide accurate large-scale depictions of regulatory interactions in human cells, and they suggest that such interactions are governed by a core set of organizing principles shared with other multicellular information processing systems.

In a classic treatise, Waddington proposed that the epigenetic landscape of a cell is ''buttressed'' by complex interactions among multiple regulatory genes (Waddington, 1939, elaborated in Waddington, 1957). These genes—now recognized as sequence-specific transcriptional regulators—form an extended ''cognitive'' network that enables the simultaneous integration of multiple internal and external cues and conveys this information to specific effector genes along the genome. Consequently, transcriptional regulatory networks influence both the current chromatin landscape of a cell as well as its epigenetic state, imparting a type of ''memory'' that may impact subsequent cellular fate decisions (Waddington, 1957; Groudine and Weintraub, 1982). Such characteristics render TF regulatory networks ideal for governing complex processes such as pluripotency (Boyer et al., 2005; Kim et al., 2008), development (Davidson et al., 2002a), and differentiation (Yun and Wold, 1996). However, despite their central role in human pathology and physiology, human transcriptional regulatory networks are presently poorly understood.

The networks we describe here for 41 diverse cell types provide an extensive description of the circuitry, dynamics, and organizing principles of the human TF regulatory network. The derivation of regulatory networks from genomic DNaseI footprint maps provides a general, scalable solution for mapping and analyzing cell-selective transcriptional regulatory networks in complex multicellular organisms. By comparison, generation of networks of this size across 41 cell types using traditional approaches such as perturbation or ChIP-seq would have required nearly 20,000 individual experiments. By contrast, the approach we describe can readily scale beyond the 475 factors analyzed in the current study and is constrained only by the availability of accurate TF recognition sequences.

Our analysis of transcriptional regulatory interactions in a network context has uncovered several novel features of human transcriptional regulation, some quite striking.

First, we observed that human transcriptional regulatory networks are markedly cell type specific, with only ~5% of all regulatory interactions common across the 41 tested cell types. This finding highlights the regulatory diversity within humans and underscores the importance of analyzing cell-selective regulatory networks when addressing specific biological questions.

Second, by detecting factors that predominantly contribute to the transcriptional regulatory networks of only one or a few cell types, we identified both known and novel regulators of cellular identity (Figure 4B). Differences between cell types thus encode a surprisingly rich landscape of information concerning differentiation and developmental processes, and this landscape can be systematically mined for regulatory insights.

Third, we found that commonly expressed TFs within a given cell lineage play distinct roles in the governance of regulatory networks of different cells within that lineage. Our analysis discovered that in acute promyelocytic leukemia cells, the widely expressed RAR-α shifts from being a passenger of FFLs to being a strong driver of FFLs. This finding provides insights into the broader—and more fundamental—regulatory alterations that accompany the RAR-α/PML fusion protein unique to acute promyelocytic leukemia. On a general level, our results show that commonly expressed proteins may display highly cell-selective actions, and that such activities may be brought to light by analyzing TFs in the context of their peers.

Finally, in marked contrast to the high regulatory diversity between cell types, we found that all cell-type regulatory networks converge on a common network architecture that closely mirrors the topology of the *C. elegans* neuronal connectivity network and those of other multicellular information processing systems (Milo et al., 2004), highlighting a fundamental similarity in the structure and organizing principles of these biological systems. Strikingly, this common architecture is independently fashioned in each cell type and results from the delicate balance of distinct regulatory interactions.

Despite the experimental and computational advantages and successes of our approach, a number of additional steps could be used to refine and improve our regulatory interaction networks. First, as noted above, our approach is limited by the availability of recognition sequences for specific TFs. The pending availability of both more and higher-quality recognition sequences through approaches such as protein-binding microarrays (Berger et al., 2008; Badis et al., 2009) and SELEX-seq (Jolma et al., 2010; Slattery et al., 2011) promises to expand considerably the horizons of human transcriptional network analysis. Such refined data may enable differentiation of factors that currently appear to bind similar recognition sequences. Second, the model that we described undervalues the role of distal regulatory elements, which can exert major influences on gene expression. Because enhancers can act over long distances, association of a given distal regulatory element with a specific TF gene is at present difficult. We therefore focused on footprints in DHSs within a 10 kb region that is centered on the TSS—a region in which most regulatory interactions are expected to be directed to the local TSS. Although large numbers of distal regulatory DNA regions marked by DNaseI-hypersensitive sites are now available through the ENCODE (The ENCODE Project Consortium, 2012) and Roadmap Epigenomics (Bernstein et al., 2010) projects, the assignment of distal regulatory elements to their cognate gene(s) has proven to be a formidable challenge. Third, the approach we utilized does not take into account indirect regulatory interactions (e.g., tethering) that may affect the expression of a given TF gene (Davidson et al., 2002b; Rigaud et al., 1991; Biddie et al., 2011). Systematic cross-comparisons between DNaseI footprint and TF ChIP-seq data drawn from the same cell type should enable recognition of such indirect interactions and derivation of rules (e.g., tethering partners) that may enable larger-scale modeling of such interactions (Neph et al., 2012a).

In order to interpret human regulatory networks at the organismal level, it will be necessary to analyze cell-selective regulatory networks within the context of surrounding tissues (Barabási and Oltvai, 2004). As initially described by Spemann over 90 years ago, the identity of a given cell can be largely dictated by its surrounding tissue (Spemann, 1918). Consequently, during both normal development and physiological function, the regulatory landscape of one cell type may become intricately dependent upon that of its neighbors (reviewed in Waddington, 1940). In this context, it is notable that we observed large diversity between the regulatory landscapes of distinct lung cell types (Figure 6E), which highlights the complexity that exists within neighboring tissue from the same organ.

In summary, our results provide a description of the circuitry, dynamics, and organizing principles of the human TF regulatory network. Systematically applied, the approach we have described has the potential to expand greatly our horizons on the mechanism, architecture, and epistemology of human gene regulation.

EXPERIMENTAL PROCEDURES

Regulatory Network Construction

We used GeneCards (Rebhan et al., 1997) and UniProt Knowledgebase (Magrane and Consortium, 2011) to map motif-binding protein information found in TRANSFAC to 538 coding genes. Some genes were indistinguishable when viewed from a potential motif-binding event perspective, as their respective gene products were annotated as binders to the same set of motif templates by TRANSFAC. In such cases, we chose a single gene, randomly, as a representative and removed others, which reduced the number of genes from 538 to 475. Networks built by removing the first redundant motif, alphabetically, or by including all redundant motifs showed very similar properties to the one described in this paper (Figure S3B and data not shown).

We symmetrically padded the TSSs of the remaining genes by 5 kb and scanned for predicted TRANSFAC motif-binding sites using FIMO (Bailey et al., 2009), version 4.6.1, with a maximum p value threshold of 1×10^{-5} and defaults for other parameters. For each cell type, we filtered putative motif-binding sites to those that overlapped footprints as previously described (Neph et al., 2012a). Each network contained 475 nodes, one per gene. A directed edge was drawn from a gene node to another when a motif instance, potentially bound by the first gene's protein product, was found within a DNaseI footprint contained within 5 kb of the second gene's TSS, indicating regulatory potential. Table S3 shows the number of such edges in every cell-type-specific network.

Network Clustering

We counted the total number edges for every TF gene node (sum of in and out edges) in a cell type and calculated the proportion of edges for that TF relative to all edges (NND). We used the NND vectors to compute the pairwise euclidean distances between cell types, and we used Ward clustering to group the cell types (Ward, 1963). We observed similar cluster patterns when comparing normalized in-degree, normalized out-degree, or un-normalized total degree (results not shown).

Triad Significance Profiles

We removed self-edges from every network and used the mfinder software tool for network motif analysis (Milo et al., 2004). A z-score was calculated over each of 13 network motifs of size 3 (three-node network motifs), using 250 randomized networks of the same size to estimate a null. We vectorized z-scores from every cell type and normalized each to unit length to create triad significance profiles (TSP) as described in Milo et al. (2004). We computed the average TSP over all cell-type-specific regulatory networks and compared to the TSP of the highly curated multicellular information processing networks described in Milo et al. (2004).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, three figures, three tables, and one data file and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2012.04.040.

ACKNOWLEDGMENTS

We thank Dr. Sam John for critical reading of the manuscript and our colleagues for many helpful and insightful observations and discussions. This work was supported by NIH grant HG004592 to J.A.S. All data are available through the ENCODE data repository at the UCSC genome browser (http://genome.ucsc.edu). E.B. is an Alfred P. Sloan Research Fellow. A.B.S. was supported by an NIDDK F30 fellowship (FDK095678A).

Received: January 17, 2012
Revised: March 19, 2012
Accepted: April 23, 2012
Published online: September 5, 2012

WEB RESOURCES

A TF network browser is available for visualization and comparison of cell-type TF networks and subnetworks: http://www.regulatorynetworks.org/.
The browser can be used to display the wiring of any selected group of TFs and to examine their dynamics across different cell types.

REFERENCES

Alon, U. (2006). An Introduction to Systems Biology: Design Principles of Biological Circuits, First Edition (Boca Raton, FL: Chapman and Hall/CRC).

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., et al. (2009). Diversity and complexity in DNA recognition by transcription factors. Science *324*, 1720–1723.

Barabási, A.-L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nat. Rev. Genet. *5*, 101–113.

Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. Nat. Genet. *37*, 382–390.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Peña-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., et al. (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell *133*, 1266–1276.

Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., et al. (2010). The NIH Roadmap Epigenomics Mapping Consortium. Nat. Biotechnol. *28*, 1045–1048.

Biddie, S.C., John, S., Sabo, P.J., Thurman, R.E., Johnson, T.A., Schiltz, R.L., Miranda, T.B., Sung, M.-H., Trump, S., Lightman, S.L., et al. (2011). Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. Mol. Cell *43*, 145–155.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947–956.

Bryne, J.C., Valen, E., Tang, M.-H.E., Marstrand, T., Winther, O., da Piedade, I., Krogh, A., Lenhard, B., and Sandelin, A. (2008). JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Res. *36*(Database issue), D102–D106.

Carro, M.S., Lim, W.K., Alvarez, M.J., Bollo, R.J., Zhao, X., Snyder, E.Y., Sulman, E.P., Anne, S.L., Doetsch, F., Colman, H., et al. (2010). The transcriptional

network for mesenchymal transformation of brain tumours. Nature *463*, 318–325.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., et al. (2002a). A genomic regulatory network for development. Science *295*, 1669–1678.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., et al. (2002b). A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. Dev. Biol. *246*, 162–190.

Dynan, W.S., and Tjian, R. (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell *35*, 79–87.

Galas, D.J., and Schmitz, A. (1978). DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res. *5*, 3157–3170.

Gerstein, M.B., Lu, Z.J., Van Nostrand, E.L., Cheng, C., Arshinoff, B.I., Liu, T., Yip, K.Y., Robilotto, R., Rechtsteiner, A., Ikegami, K., et al; modENCODE Consortium. (2010). Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science *330*, 1775–1787.

Graf, T., and Enver, T. (2009). Forcing cells to change lineages. Nature *462*, 587–594.

Grignani, F., Ferrucci, P.F., Testa, U., Talamo, G., Fagioli, M., Alcalay, M., Mencarelli, A., Grignani, F., Peschle, C., Nicoletti, I., et al. (1993). The acute promyelocytic leukemia-specific PML-RAR alpha fusion protein inhibits differentiation and promotes survival of myeloid precursor cells. Cell *74*, 423–431.

Grignani, F., De Matteis, S., Nervi, C., Tomassoni, L., Gelmetti, V., Cioce, M., Fanelli, M., Ruthardt, M., Ferrara, F.F., Zamir, I., et al. (1998). Fusion proteins of the retinoic acid receptor-alpha recruit histone deacetylase in promyelocytic leukaemia. Nature *391*, 815–818.

Groudine, M., and Weintraub, H. (1982). Propagation of globin DNAase I-hypersensitive sites in absence of factors required for induction: a possible mechanism for determination. Cell *30*, 131–139.

Hesselberth, J.R., Chen, X., Zhang, Z., Sabo, P.J., Sandstrom, R., Reynolds, A.P., Thurman, R.E., Neph, S., Kuehn, M.S., Noble, W.S., et al. (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nat. Methods *6*, 283–289.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497–1502.

Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G., Enge, M., Taipale, M., Vaquerizas, J.M., Yan, J., Sillanpää, M.J., et al. (2010). Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res. *20*, 861–873.

Kadonaga, J.T., Carner, K.R., Masiarz, F.R., and Tjian, R. (1987). Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell *51*, 1079–1090.

Karin, M., Haslinger, A., Holtgreve, H., Richards, R.I., Krauter, P., Westphal, H.M., and Beato, M. (1984). Characterization of DNA sequences through which cadmium and glucocorticoid hormones induce human metallothionein-IIA gene. Nature *308*, 513–519.

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.

Kuo, M.T., Mandel, J.L., and Chambon, P. (1979). DNA methylation: correlation with DNase I sensitivity of chicken ovalbumin and conalbumin chromatin. Nucleic Acids Res. *7*, 2105–2113.

MacNeill, C., French, R., Evans, T., Wessels, A., and Burch, J.B. (2000). Modular regulation of cGATA-5 gene expression in the developing heart and gut. Dev. Biol. *217*, 62–76.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. Science *298*, 824–827.

Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M., and Alon, U. (2004). Superfamilies of evolved and designed networks. Science *303*, 1538–1542.

Magrane, M., and Consortium, U. (2011). UniProt Knowledgebase: a hub of integrated protein data. Database: The Journal of Biological Databases and Curation *2011*, bar009.

Molkentin, J.D., Lin, Q., Duncan, S.A., and Olson, E.N. (1997). Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. Genes Dev. *11*, 1061–1072.

Naidu, P.S., Ludolph, D.C., To, R.Q., Hinterberger, T.J., and Konieczny, S.F. (1995). Myogenin and MEF2 function synergistically to activate the MRF4 promoter during myogenesis. Mol. Cell. Biol. *15*, 2707–2718.

Neph, S.J., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., Sandstrom, R., Johnson, A.K., Humbert, R., et al. (2012a). An expansive human regulatory lexicon encoded in transcription factor footprints. Nature http://dx.doi.org/10.1038/nature11212.

Newburger, D.E., and Bulyk, M.L. (2009). UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. Nucleic Acids Res. *37*(Database issue), D77–D82.

Nutt, S.L., Heavey, B., Rolink, A.G., and Busslinger, M. (1999). Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. Nature *401*, 556–562.

Orkin, S.H. (1995). Transcription factors and hematopoietic development. J. Biol. Chem. *270*, 4955–4958.

Pfeifer, G.P., and Riggs, A.D. (1991). Chromatin differences between active and inactive X chromosomes revealed by genomic footprinting of permeabilized cells using DNase I and ligation-mediated PCR. Genes Dev. *5*, 1102–1113.

Ramachandran, B., Yu, G., Li, S., Zhu, B., and Gulick, T. (2008). Myocyte enhancer factor 2A is transcriptionally autoregulated. J. Biol. Chem. *283*, 10318–10329.

Reece-Hoyes, J.S., Diallo, A., Lajoie, B., Kent, A., Shrestha, S., Kadreppa, S., Pesyna, C., Dekker, J., Myers, C.L., and Walhout, A.J.M. (2011). Enhanced yeast one-hybrid assays for high-throughput gene-centered regulatory network mapping. Nat. Methods *8*, 1059–1064.

Rigaud, G., Roux, J., Pictet, R., and Grange, T. (1991). In vivo footprinting of rat TAT gene: dynamic interplay between the glucocorticoid receptor and a liver-specific factor. Cell *67*, 977–986.

Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., Lin, M.F., et al; modENCODE Consortium. (2010). Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science *330*, 1787–1797.

Serov, V.N., Spirov, A.V., and Samsonova, M.G. (1998). Graphical interface to the genetic network database GeNet. Bioinformatics *14*, 546–547.

Shimazaki, T., Arsenijevic, Y., Ryan, A.K., Rosenfeld, M.G., and Weiss, S. (1999). A role for the POU-III transcription factor Brn-4 in the regulation of striatal neuron precursor differentiation. EMBO J. *18*, 444–456.

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H.J., and Mann, R.S. (2011). Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. Cell *147*, 1270–1282.

Spemann, H. (1918). Über die Determination der ersten Organanlagen des Amphibienembryo I–VI. Arch. Entwicklungsmech. Org. *43*, 448–555.

Stalder, J., Larsen, A., Engel, J.D., Dolan, M., Groudine, M., and Weintraub, H. (1980). Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. Cell *20*, 451–460.

Sucov, H.M., Lou, J., Gruber, P.J., Kubalak, S.W., Dyson, E., Gumeringer, C.L., Lee, R.Y., Moles, S.A., Chien, K.R., Giguere, V., and Evans, R.M. (1996). The molecular genetics of retinoic acid receptors: cardiovascular and limb development. Biochem. Soc. Symp. *62*, 143–156.

Swiers, G., Patient, R., and Loose, M. (2006). Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. Dev. Biol. *294*, 525–540.

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.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

The ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature http://dx.doi.org/10.1038/nature11247.

Tsai, S.F., Martin, D.I., Zon, L.I., D'Andrea, A.D., Wong, G.G., and Orkin, S.H. (1989). Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. Nature *339*, 446–451.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. *10*, 252–263.

Waddington, C.H. (1939). Introduction to Modern Genetics (New York: The Macmillan Company).

Waddington, C.H. (1940). Organisers and Genes (Cambridge, UK: Cambridge University Press).

Waddington, C.H. (1957). The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology (Victoria, Australia: George Allen & Unwin).

Walhout, A.J.M. (2006). Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. Genome Res. *16*, 1445–1454.

White, J.G., Southgate, E., Thomson, J.N., and Brenner, S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans. Philos. Trans. R. Soc. Lond. B Biol. Sci. *314*, 1–340.

Wingender, E., Dietze, P., Karas, H., and Knüppel, R. (1996). TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res. *24*, 238–241.

Wu, C. (1980). The 5′ ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. Nature *286*, 854–860.

Wu, C., Bingham, P.M., Livak, K.J., Holmgren, R., and Elgin, S.C. (1979). The chromatin structure of specific genes: I. Evidence for higher order domains of defined DNA sequence. Cell *16*, 797–806.

Xing, Y., Li, C., Li, A., Sridurongrit, S., Tiozzo, C., Bellusci, S., Borok, Z., Kaartinen, V., and Minoo, P. (2010). Signaling via Alk5 controls the ontogeny of lung Clara cells. Development *137*, 825–833.

Yuh, C.H., Ransick, A., Martinez, P., Britten, R.J., and Davidson, E.H. (1994). Complexity and organization of DNA-protein interactions in the 5′-regulatory region of an endoderm-specific marker gene in the sea urchin embryo. Mech. Dev. *47*, 165–186.

Yun, K., and Wold, B. (1996). Skeletal muscle determination and differentiation: story of a core regulatory network and its context. Curr. Opin. Cell Biol. *8*, 877–889.