Published in final edited form as:
Cell. 2012 September 14; 150(6): 1274–1286. doi:10.1016/j.cell.2012.04.040.

Circuitry and dynamics of human transcription factor regulatory networks

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

[1]Department of Genome Sciences, University of Washington, Seattle, WA 98195

[2]Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195

[3]Santa Fe Institute, Santa Fe, NM 87501

[4]Department of Medicine, University of Washington, Seattle, WA 98195

SUMMARY

The combinatorial cross-regulation of hundreds of sequence-specific transcription factors 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 transcription factors, and to analyze the dynamics of these connections across 41 diverse cell and tissue types. We find that human transcription factor 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 the first description of the circuitry, dynamics, and organizing principles of the human transcription factor regulatory network.

INTRODUCTION

Sequence-specific transcriptional 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 sub-networks 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., 2002) 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

¥correspondence: jstam@uw.edu, elbo@uw.edu.
*equal contribution

and robustness (Alon, 2006). However, despite their central biological roles, both the structure of core human regulatory networks and their component sub-networks are largely undefined.

One of the main bottlenecks limiting generation of transcription factor regulatory networks for complex biological systems has been that information is traditionally collected from individual experiments targeting one cell-type and one transcription factor at a time (Davidson et al., 2002; Yuh et al., 1994; Kim et al., 2008; ModENCODE 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 transcription factors and analyzing the effect of these perturbations on the expression of transcription factor genes containing putative *cis*-regulatory binding elements for these factors (Davidson et al., 2002; 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 sub-networks of small numbers of TFs, such as those involved in pluripotency (Kim et al., 2008) or larger scale networks combining several dozen TFs (ModENCODE et al., 2010; Gerstein et al., 2010). However, such approaches are limited by three major factors: (i) the availability of suitable affinity reagents; (ii) the difficulty of interrogating the activities of multiple TFs within the same cellular environment; and, perhaps most critically (iii) 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 et al., 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 multi-cellular 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., 2002; Basso et al., 2005; Boyer et al., 2005; Kim et al., 2008; ModENCODE 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 transcriptional regulatory network analysis suitable for 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 transcription factors (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 potentially all DNA-binding factors in a single experiment (Neph et al., 2012).

By performing systematic analysis of TF footprints in the proximal regulatory regions of each transcription factor 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 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 transcription factor 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 transcription factors, and analyze the dynamics of these connections across 41 diverse cell and tissue types.

RESULTS

Comprehensive mapping of transcription factor networks in diverse human cell types

To generate transcription factor regulatory networks in human cells, we analyzed genomic DNaseI footprinting data from 41 diverse cell and tissue types (Neph et al., 2012). Each of these 41 samples was treated with DNaseI, and sites of DNaseI cleavage along the genome were analyzed using 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., 2012)). Collectively, we detected 45,096,726 footprints, representing cell-selective binding to ~8.4 million distinct 6–40bp genomic sequence elements. We inferred the identity of factors occupying DNaseI footprints using well-annotated databases of transcription factor binding motifs (Wingender et al., 1996; Bryne et al., 2008; Newburger et al., 2009) (**Methods**), and confirmed that these identifications matched closely and quantitatively with ENCODE ChIP-seq data for the same cognate factors (Neph et al., 2012).

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 10kb interval centered on the transcriptional start site) of 475 transcription factor genes with well-annotated recognition motifs (Wingender et al., 1996; Bryne et al., 2008; Newburger et al., 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. 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 sub-networks (Figure 1B–C). The muscle-specific factors MyoD, Myogenin (MYOG), MEF2A, and MYF6 form a network vital for specification of skeletal muscle fate and control of myogenic development and differentiation, which was uncovered using a combination of genetic and physical studies, including DNaseI footprinting (Naidu et al., 1995; Yun and Wold, 1996; Ramachandran et al., 2008). Figure 1B juxtaposes the known regulatory interactions between these factors determined 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 (ES) cells (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 ES cells (Kim et al., 2008) (Figure 1C, top). A nearly identical sub-network emerges from analysis of the transcription factor network computed *de novo* from DNaseI footprints in human ES cells (Figure 1C, bottom).

Critically, both the well-annotated muscle and ES sub-networks are best matched by footprint-derived networks computed specifically from skeletal myoblasts and human ES cells, respectively, vs. other cell types (Figure 1D–E). These findings indicate that network relationships between transcription factors derived *de novo* from genomic DNaseI footprinting accurately recapitulate well-described cell type-selective transcriptional regulatory networks generated using multiple experimental approaches.

Transcription factor regulatory networks show marked cell-selectivity

We next analyzed systematically the dynamics of TF regulatory networks across cell types. 475 transcription factors 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 (Supplemental Figure S1A).

To visualize the global landscape of cell-selective vs. shared regulatory interactions, we first computed the broad landscape of network edges that are either specific to a given cell type, or are found in networks of two or more cell types (Figure 2, Supplemental 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 4 or fewer cell types (Supplemental Figure S1A). By contrast, only 5% of edges were common to all cell types (Supplemental Figure S1A). Interestingly, when comparing networks, we found more common edges than common DNaseI footprints (Supplemental Figure S1B,C), 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 ES cell network, and hematopoietic factors within the network of hematopoietic stem cells (Figure 3A). Next, we plotted the

complete set of regulatory interactions amongst all 475 edges between the same six diverse cell types, exposing a high degree of regulatory diversity (Figure 3B, Supplemental Table S1).

Edges unique to a cell type typically form a well-connected sub-network (Supplemental Figure S1D–F, Supplemental Table S2), implying that cell type-specific regulatory differences are not driven merely by the independent actions of a few transcription factors, but rather by organized TF sub-networks. In addition, the density of cell-selective networks varies widely between cell types (e.g., compare ES cells 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 transcription factor 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 which 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 transcription factors, 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, ES cells, and malignant cells with a ‘de-differentiated’ 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 trans-differentiation among related cell-types (Graf and Enver, 2009).

To identify the individual transcription factors 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, agreeing 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, while the cardiac developmental regulator GATA4 (Molkentin et al., 1997) 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 (Supplemental 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 footprinting) to elucidate multiple intricate transcriptional regulatory relationships.

Network analysis reveals cell-type specific behaviors for widely expressed TFs

Many transcription factors 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 level, and a cadre of major transcriptional regulators has been defined, including TAL1/SCL, PU.1, ELF1, HES1, MYB, GATA2 and GATA1 (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 sub-networks comprising the aforementioned seven regulators in five hematopoietic and one non-hematopoietic cell type (Figure 5A). For each cell type sub-network, 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 7-member hematopoietic regulatory sub-network 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 sub-networks 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 sub-network specifically within erythroid cells (Figure 5A), consistent with its defining role in erythropoiesis. In both B-cells and T-cells, the sub-network takes on a directional character, with PU.1 in a superior position. By contrast, the network is largely absent in non-hematopoietic 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 feed-forward loops (FFLs). Feed-forward loops represent one of the most important network motifs in biological and regulatory systems and comprise a three node structure in which information is propagated 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 feed-forward loops containing RAR-α at each of the three different positions (top vs. middle vs. bottom; Figure 5B, top). In most cell-types, RAR-α chiefly participates in feed-forward loops 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 feed-forward loops, 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 which results in a RAR-α/PML fusion protein that misregulates RAR-x003B1; target sites (Grignani et al., 1993; Grignani et al., 1998). Our results suggest that in APL cells, RAR-α is additionally altering the basic organization of the regulatory network. Critically, we identified the prominent role of RAR-α in APL using DNaseI footprint-driven network analysis without any prior knowledge of its role in 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 transcription factor 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 feed-forward loop 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 found at higher than expected frequency within diverse biological networks (Milo et al., 2002; Milo et al., 2004). We therefore sought to analyze the topology of the human transcription factor regulatory network, and to compare it with those of well-annotated multi-cellular 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 3-node networks motifs found in perhaps the best annotated multi-cellular 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, Supplemental 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., 2002), the *Drosophila* developmental transcriptional network (Serov et al., 1998), and the mammalian signal transduction network (Milo et al., 2004) (Supplemental 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 re-computed this network solely from scanned transcription factor 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 if the observed similarity to the neuronal network was a collective property of human transcription factor 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 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, we asked whether a common core of regulatory interactions might be driving the conserved network architecture, by comparing feed-forward loops from biologically similar cell types with one another. This comparison revealed marked diversity among different cellular TF networks (Figure 6D,E), considerably exceeding that observed among individual edges (Supplemental Figure S3C,D). Indeed, only ~0.1% of all observed FFLs across 41 cell types (784 / 558,841) were common to all cell types (Figure 6F and Supplemental Figure S3E). Moreover, only a minority of the TFs represented within a given cellular network contribute to the enriched network motifs (Supplemental 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

Transcription factor 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 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 transcription factor regulatory networks ideal for governing complex processes such as pluripotency (Boyer et al., 2005; Kim et al., 2008), development (Davidson et al., 2002) 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 represent the first genome-scale human transcriptional regulatory networks, and are among the largest described in any organism. 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 multi-cellular 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 feed-forward loops (FFLs) to being a strong driver of FFLs. This finding provides novel 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 transcription factors 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 10kb region centered on the transcriptional start site (TSS), 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 (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., 2002; 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, 2012).

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) highlighting the complexity that exists within neighboring tissue from the same organ.

In summary, our results provide the first description of the circuitry, dynamics, and organizing principles of the human transcription factor 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 mapped motif-binding protein information found in TRANSFAC to 538 coding genes, using GeneCards (Rebhan et al., 1997) and UniProt Knowledgebase (Magrane et al., 2011). 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 (Supplemental Figure S3B and data not shown).

We symmetrically padded the transcriptional start sites (TSSs) of the remaining genes by 5,000 nt 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 1e-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., 2012). Each network contained 475 vertices, one per gene. A directed edge was drawn from a gene-vertex to another when a motif instance, potentially bound by the first gene's protein product, was found within a DNaseI footprint contained within 5,000 nt of the second gene's TSS, indicating regulatory potential. Supplemental 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-vertex (sum of in and out edges) in a cell type and calculated the proportion of edges for that TF relative to all edges (normalized network degree (NND)). We computed the pairwise euclidean distances between cell types using the NND vectors and grouped the cell types using Ward clustering (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 (3-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 multi-cellular information processing networks described in Milo et al., 2004.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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.

REFERENCES

Alon, U. An Introduction to Systems Biology: Design Principles of Biological Circuits. 1 st ed. Chapman and Hall/CRC; 2006.

Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X, et al. Diversity and Complexity in DNA Recognition by Transcription Factors. Science. 2009; 324:1720–1723. [PubMed: 19443739]

Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. Nature Reviews Genetics. 2004; 5:101–113.

Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. Nature Genetics. 2005; 37:382–390. [PubMed: 15778709]

Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Peña-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, et al. Variation in Homeodomain DNA Binding Revealed by High-Resolution Analysis of Sequence Preferences. Cell. 2008; 133:1266–1276. [PubMed: 18585359]

Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nature Biotechnology. 2010; 28:1045–1048.

Biddie SC, John S, Sabo PJ, Thurman RE, Johnson TA, Schiltz RL, Miranda TB, Sung M-H, Trump S, Lightman SL, et al. Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. Molecular Cell. 2011; 43:145–155. [PubMed: 21726817]

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell. 2005; 122:947–956. [PubMed: 16153702]

Bryne JC, Valen E, Tang M-HE, Marstrand T, Winther O, da Piedade I, Krogh A, Lenhard B, Sandelin A. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Research. 2008; 36:D102–D106. [PubMed: 18006571]

Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, Snyder EY, Sulman EP, Anne SL, Doetsch F, Colman H, et al. The transcriptional network for mesenchymal transformation of brain tumours. Nature. 2010; 463:318–325. [PubMed: 20032975]

Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh C-H, Minokawa T, Amore G, Hinman V, Arenas-Mena C, et al. A genomic regulatory network for development. Science. 2002; 295:1669–1678. [PubMed: 11872831]

Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh C-H, Minokawa T, Amore G, Hinman V, Arenas-Mena C, et al. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. Developmental Biology. 2002; 246:162–190. [PubMed: 12027441]

Dynan WS, Tjian R. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell. 1983; 35:79–87. [PubMed: 6313230]

Galas DJ, Schmitz A. DNAase footprinting a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Research. 1978; 5:3157–3170. [PubMed: 212715]

Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, et al. Integrative Analysis of the Caenorhabditis elegans Genome by the modENCODE Project. Science. 2010; 330:1775–1787. [PubMed: 21177976]

Graf T, Enver T. Forcing cells to change lineages. Nature. 2009; 462:587–594. [PubMed: 19956253]

Grignani F, Ferrucci PF, Testa U, Talamo G, Fagioli M, Alcalay M, Mencarelli A, Peschle C, Nicoletti I. The acute promyelocytic leukemia-specific PML-RAR alpha fusion protein inhibits differentiation and promotes survival of myeloid precursor cells. Cell. 1993; 74:423–431. [PubMed: 8394219]

Grignani F, De Matteis S, Nervi C, Tomassoni L, Gelmetti V, Cioce M, Fanelli M, Ruthardt M, Ferrara FF, Zamir I, et al. Fusion proteins of the retinoic acid receptor-alpha recruit histone deacetylase in promyelocytic leukaemia. Nature. 1998; 391:815–818. [PubMed: 9486655]

Groudine M, Weintraub H. Propagation of globin DNAase, I-hypersensitive sites in absence of factors required for induction: a possible mechanism for determination. Cell. 1982; 30:131–139. [PubMed: 6290075]

Hesselberth JR, Chen X, Zhang Z, Sabo PJ, Sandstrom R, Reynolds AP, Thurman RE, Neph S, Kuehn MS, Noble WS, et al. Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nature Methods. 2009; 6:283–289. [PubMed: 19305407]

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

Jolma A, Kivioja T, Toivonen J, Cheng L, Wei G, Enge M, Taipale M, Vaquerizas JM, Yan J, Sillanpää MJ, et al. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Research. 2010; 20:861–873. [PubMed: 20378718]

Kadonaga JT, Carner KR, Masiarz FR, Tjian R. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell. 1987; 51:1079–1090. [PubMed: 3319186]

Karin M, Haslinger A, Holtgreve H, Richards RI, Krauter P, Westphal HM, Beato M. Characterization of DNA sequences through which cadmium and glucocorticoid hormones induce human metallothionein-IIA gene. Nature. 1984; 308:513–519. [PubMed: 6323998]

Kim J, Chu J, Shen X, Wang J, Orkin SH. An extended transcriptional network for pluripotency of embryonic stem cells. Cell. 2008; 132:1049–1061. [PubMed: 18358816]

Kuo MT, Mandel JL, Chambon P. DNA methylation: correlation with DNase I sensitivity of chicken ovalbumin and conalbumin chromatin. Nucleic Acids Research. 1979; 7:2105–2113. [PubMed: 523315]

MacNeill C, French R, Evans T, Wessels A, Burch JB. Modular regulation of cGATA-5 gene expression in the developing heart and gut. Developmental Biology. 2000; 217:62–76. [PubMed: 10625536]

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. Science. 2002; 298:824–827. [PubMed: 12399590]

Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, Ayzenshtat I, Sheffer M, Alon U. Superfamilies of evolved and designed networks. Science. 2004; 303:1538–1542. [PubMed: 15001784]

Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, Eaton ML, Landolin JM, Bristow CA, Ma L, et al. modENCODE Consortium. Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science. 2010; 330:1787–1797. [PubMed: 21177974]

Molkentin JD, Lin Q, Duncan SA, Olson EN. Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. Genes & Development. 1997; 11:1061–1072. [PubMed: 9136933]

Naidu PS, Ludolph DC, To RQ, Hinterberger TJ, Konieczny SF. Myogenin and MEF2 function synergistically to activate the MRF4 promoter during myogenesis. Molecular and Cellular Biology. 1995; 15:2707–2718. [PubMed: 7739551]

Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, Sandstrom R, Johnson AK, Humbert R, et al. An expansive human regulatory lexicon encoded in transcription factor footprints. (Submitted).

Newburger DE, Bulyk ML. UniPROBE: an online database of protein binding microarray data on protein–DNA interactions. Nucleic Acids Research. 2009; 37:D77–D82. [PubMed: 18842628]

Nutt SL, Heavey B, Rolink AG, Busslinger M. Commitment to the, B-lymphoid lineage depends on the transcription factor Pax5. Nature. 1999; 401:556–562. [PubMed: 10524622]

Orkin SH. Transcription Factors and Hematopoietic Development. Journal of. Biological Chemistry. 1995; 270:4955–4958. [PubMed: 7890597]

Pfeifer GP, Riggs AD. Chromatin differences between active and inactive X chromosomes revealed by genomic footprinting of permeabilized cells using DNase I and ligation-mediated PCR. Genes & Development. 1991; 5:1102–1113. [PubMed: 2044957]

Ramachandran B, Yu G, Li S, Zhu B, Gulick T. Myocyte enhancer factor 2A is transcriptionally autoregulated. The Journal of Biological Chemistry. 2008; 283:10318–10329. [PubMed: 18073218]

Reece-Hoyes JS, Diallo A, Lajoie B, Kent A, Shrestha S, Kadreppa S, Pesyna C, Dekker J, Myers CL, Walhout AJM. Enhanced yeast one-hybrid assays for high-throughput gene-centered regulatory network mapping. Nature Methods. 2011; 8:1059–1064. [PubMed: 22037705]

Rigaud G, Roux J, Pictet R, Grange T. In vivo footprinting of rat TAT gene: dynamic interplay between the glucocorticoid receptor and a liver-specific factor. Cell. 1991; 67:977–986. [PubMed: 1683601]

Serov VN, Spirov AV, Samsonova MG. Graphical interface to the genetic network database GeNet. Bioinformatics (Oxford, England). 1998; 14:546–547.

Shimazaki T, Arsenijevic Y, Ryan AK, Rosenfeld MG, Weiss S. A role for the POU-III transcription factor Brn-4 in the regulation of striatal neuron precursor differentiation. The EMBO Journal. 1999; 18:444–456. [PubMed: 9889200]

Slattery M, Riley T, Liu P, Abe N, Gomez-Alcala P, Dror I, Zhou T, Rohs R, Honig B, Bussemaker HJ, et al. Cofactor Binding Evokes Latent Differences in DNA Binding Specificity between Hox Proteins. Cell. 2011; 147:1270–1282. [PubMed: 22153072]

Spemann H. Über die Determination der ersten Organanlagen des Amphibienembryo I–VI. Archiv für Entwicklungsmechanik der Organismen. 1918; 43:448–555.

Stalder J, Larsen A, Engel JD, Dolan M, Groudine M, Weintraub H. Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. Cell. 1980; 20:451–460. [PubMed: 7388947]

Sucov HM, Lou J, Gruber PJ, Kubalak SW, Dyson E, Gumeringer CL, Lee RY, Moles SA, Chien KR, Giguere V, et al. The molecular genetics of retinoic acid receptors: cardiovascular and limb development. Biochemical Society Symposium. 1996; 62:143–156. [PubMed: 8971347]

Swiers G, Patient R, Loose M. Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. Developmental Biology. 2006; 294:525–540. [PubMed: 16626682]

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007; 131:861–872. [PubMed: 18035408]

Takahashi K, Yamanaka S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell. 2006; 126:663–676. [PubMed: 16904174]

The ENCODE Project Consortium. A User's Guide to the Encyclopedia of DNA Elements (ENCODE). PLoS Biology. 2011; 9 e1001046+

The ENCODE Project Consortium. Initial Analysis of the Encyclopedia of DNA Elements in the Human Genome. (Submitted).

Tsai SF, Martin DI, Zon LI, D'Andrea AD, Wong GG, Orkin SH. Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. Nature. 1989; 339:446–451. [PubMed: 2725678]

Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. Nature Reviews Genetics. 2009; 10:252–263.

Waddington, CH. Introduction to Modern Genetics. George Allen & Unwin; 1939.

Waddington, CH. Organisers and Genes. Cambridge University Press; 1940.

Waddington, CH. The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology. George Allen & Unwin; 1957.

Walhout AJM. Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. Genome Research. 2006; 16:1445–1454. [PubMed: 17053092]

White JG, Southgate E, Thomson JN, Brenner S. The Structure of the Nervous System of the Nematode Caenorhabditis elegans. Philosophical Transactions of the Royal Society B: Biological Sciences. 1986; 314:1–340.

Wingender E, Dietze P, Karas H, Knüppel R. TRANSFAC: A Database on Transcription Factors and Their DNA Binding Sites. Nucleic Acids Research. 1996; 24:238–241. [PubMed: 8594589]

Wu C, Bingham PM, Livak KJ, Holmgren R, Elgin SC. The chromatin structure of specific genes: I. Evidence for higher order domains of defined DNA sequence. Cell. 1979; 16:797–806. [PubMed: 455449]

Wu C. The 5’ ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. Nature. 1980; 286:854–860. [PubMed: 6774262]

Xing Y, Li C, Li A, Sridurongrit S, Tiozzo C, Bellusci S, Borok Z, Kaartinen V, Minoo P. Signaling via Alk5 controls the ontogeny of lung Clara cells. Development. 2010; 137:825–833. [PubMed: 20147383]

Yuh CH, Ransick A, Martinez P, Britten RJ, Davidson EH. Complexity and organization of DNA-protein interactions in the 5’-regulatory region of an endoderm-specific marker gene in the sea urchin embryo. Mechanisms of Development. 1994; 47:165–186. [PubMed: 7811639]

Yun K, Wold B. Skeletal muscle determination and differentiation: story of a core regulatory network and its context. Current Opinion in Cell Biology. 1996; 8:877–889. [PubMed: 8939680]

HIGHLIGHTS

- Extensive transcription factor regulatory networks for 41 human cell and tissue types
- Transcription factor networks are highly cell-selective and highlight novel regulators of cellular identity
- Network analysis identifies cell-selective functions for commonly-expressed regulators
- The circuit architecture of human transcription factor networks mirrors living neuronal networks

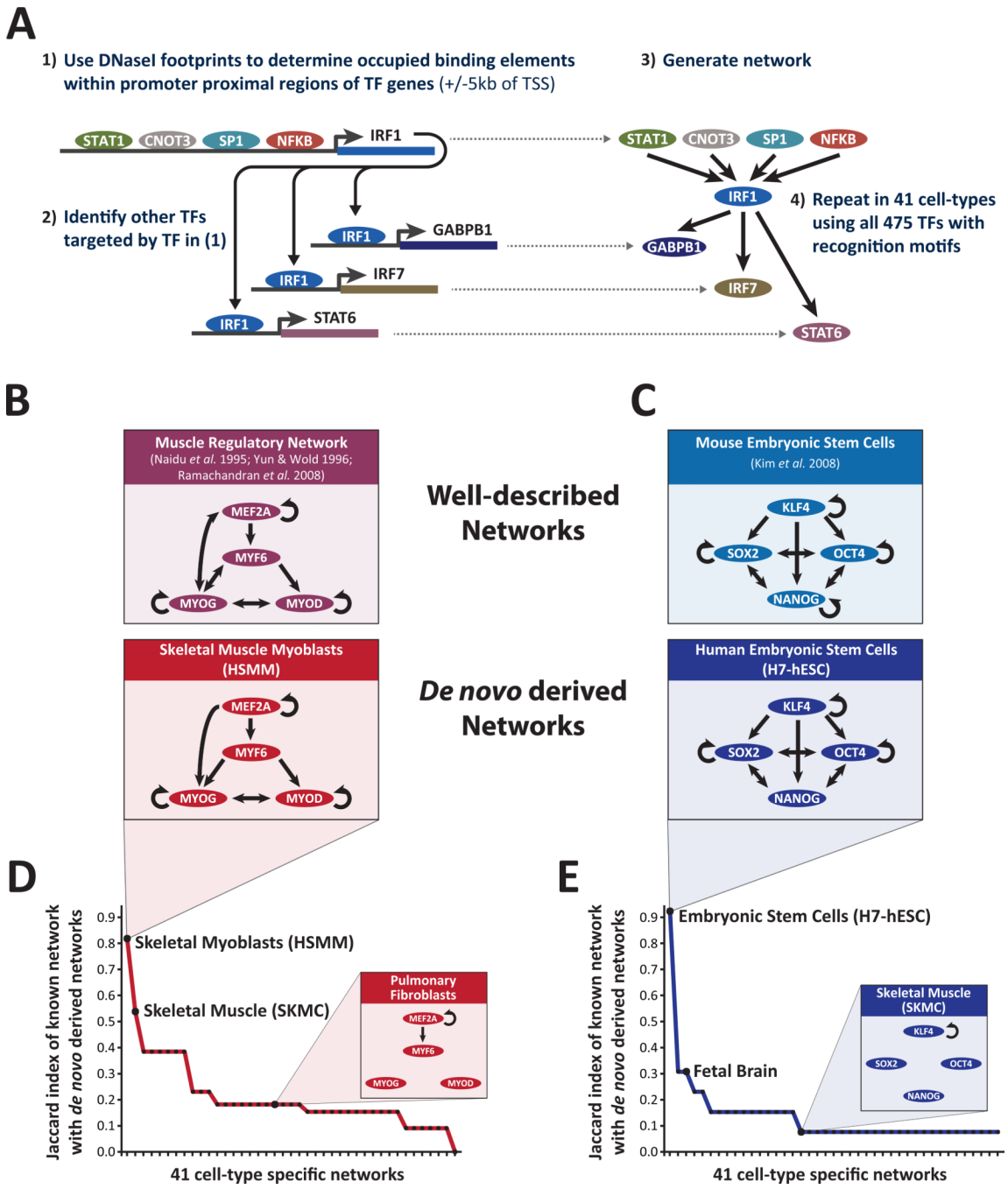

Figure 1. Construction of comprehensive transcriptional regulatory networks

(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 vs. de novo-derived regulatory sub-networks.*
(B) *Muscle sub-network. Top*, experimentally-defined regulatory sub-network for major factors controlling skeletal muscle differentiation and transcription. Arrows indicate direction(s) of regulatory interactions between factors. *Bottom*, regulatory sub-network derived *de novo* from the DNaseI footprint-anchored network of skeletal myoblasts closely matches the experimentally annotated network.
(C) *Pluripotency sub-network. Top*, regulatory sub-network for major pluripotency factors defined experimentally in mouse ES cells (Kim et al. 2008). *Bottom*, regulatory sub-network derived *de novo* from human ES cells is virtually identical to the annotated network.
(D,E) *De novo*-derived sub-networks in (B) and (C) match the annotated networks in a cellspecific fashion. *Vertical axes*: Jaccard index, a measure of network similarity, comparing the annotated sub-network 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, sub-networks computed from fibroblasts are largely devoid of relevant interactions. For the annotated pluripotency sub-network, the highest similarity is seen in human ES cells (H7-ESC).

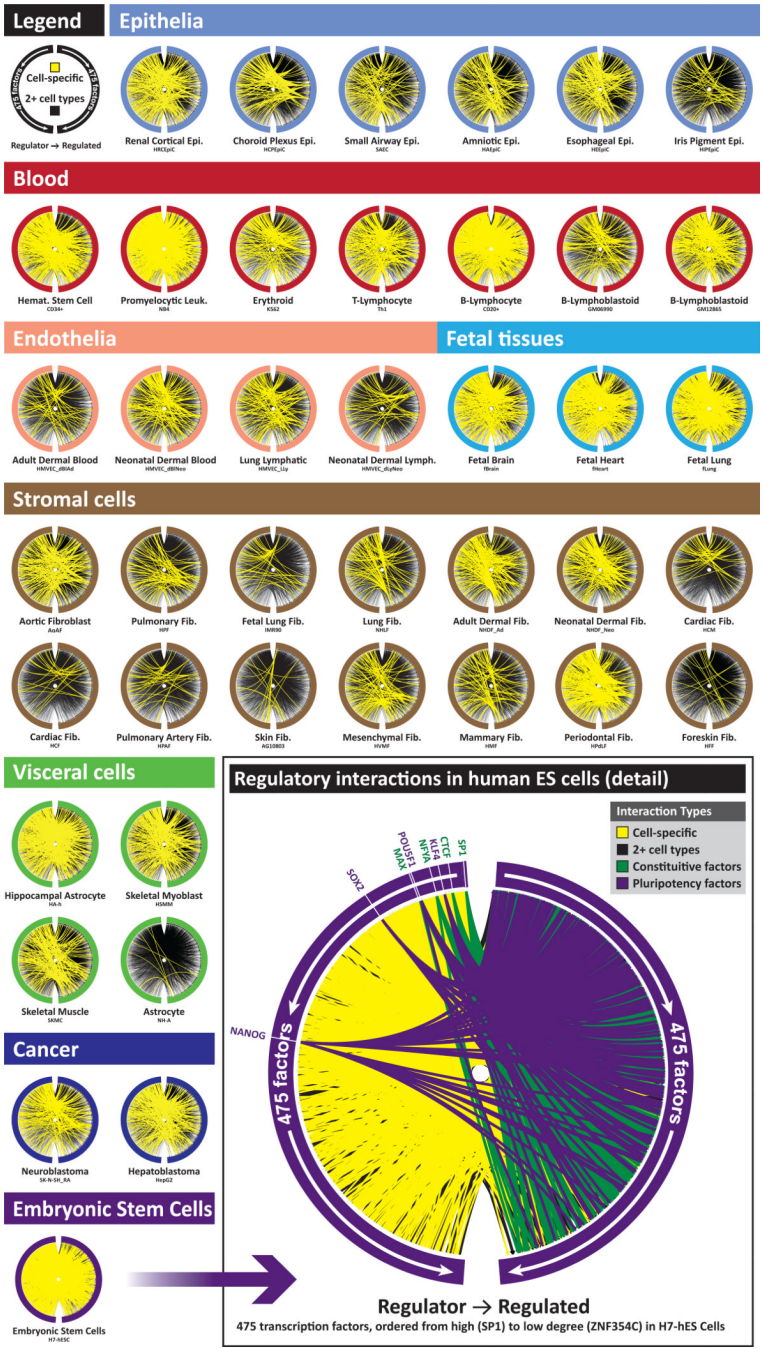

Figure 2. Cell-specific vs. shared regulatory interactions in TF networks of 41 diverse cell types

Shown for each of 41 cell types are schematics of cell type-specific vs. non-specific (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 transcription factor. 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 Supplementary Table 1, and represents a sorted list (descending order) of their degree (i.e., number of connections to other TFs) in the ES cell 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 ES cell 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.

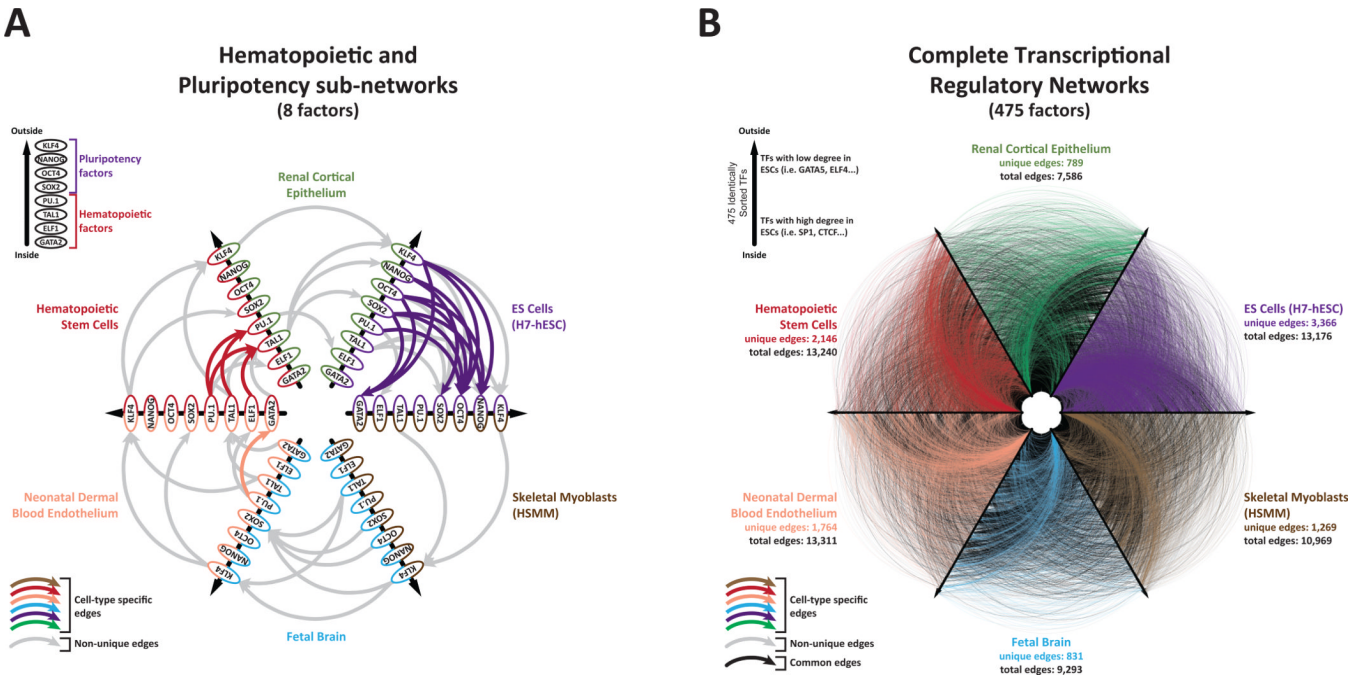

Figure 3. Transcriptional regulatory networks show marked cell-type specificity

(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 regulat*ed*) are shown by arrows in clock-wise orientation. Cell type-specific edges are colored as indicated, whereas regulatory interactions present in two or more cell type networks are shown in grey. **(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 grey. Interactions present in all six cell type networks are colored black. (See also Supplementary Figure S1 and Supplementary Table S2).

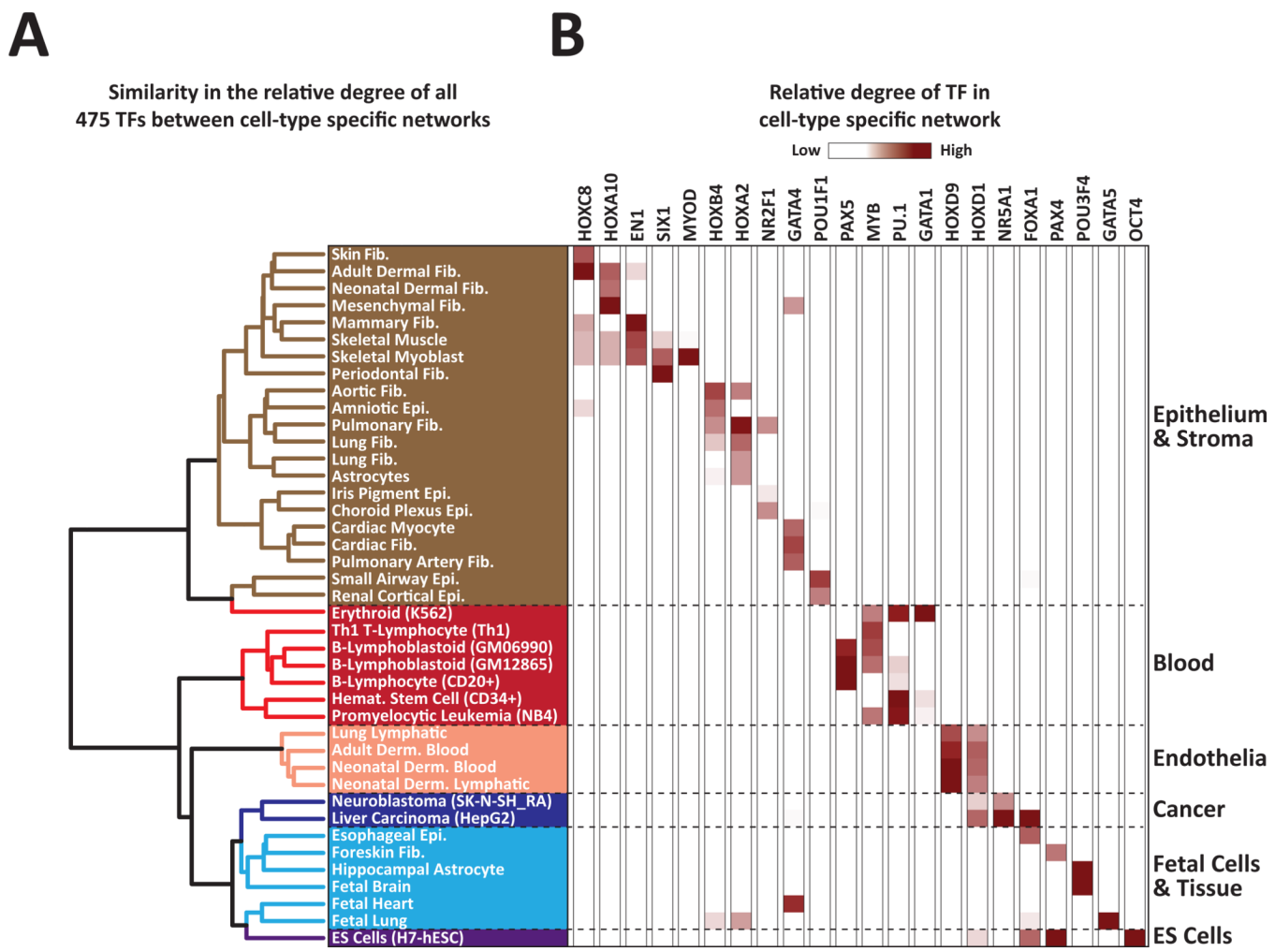

Figure 4. Functionally related cell types share similar core transcriptional regulatory networks

(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 identiy. (See also Supplemental Figure S2).

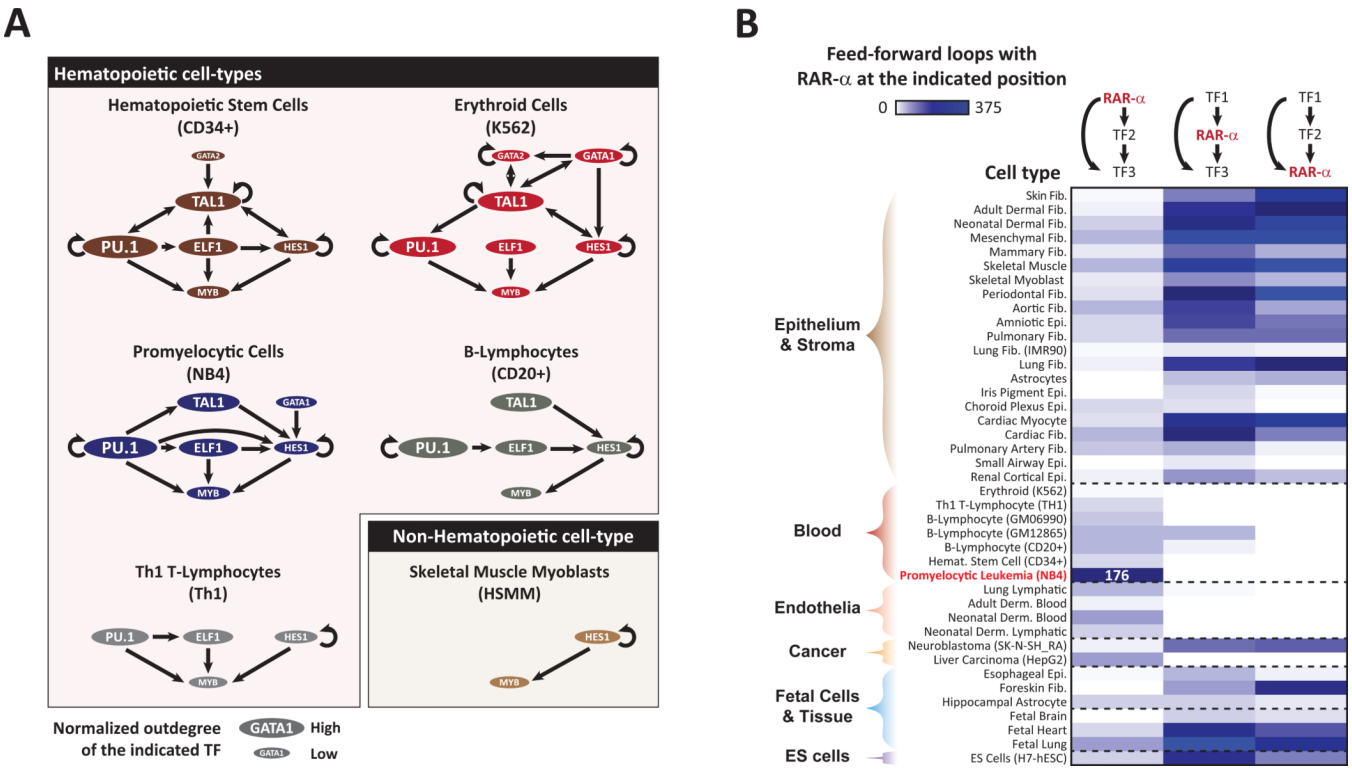

Figure 5. Cell-selective behaviors of widely expressed TFs

(A) Shown are regulatory sub-networks 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 erythroidspecific 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 sub-network takes on a directional character, with PU.1 in a superior position. By contrast, the network is largely absent in non-hematopoietic cells (muscle, HSMM, bottom right).

(B) Heatmap showing the frequency with which the retinoic acid receptor-alpha (RAR-α) is positioned as a driver (top) or passenger (middle or bottom) within feed-forward loops (FFLs) mapped in 41 cell type regulatory networks. Note that in most cell-types, RAR-α participates in feed-forward loops 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 feed-forward loops. Cell types are arranged according to the clustered ordering in Figure 4.

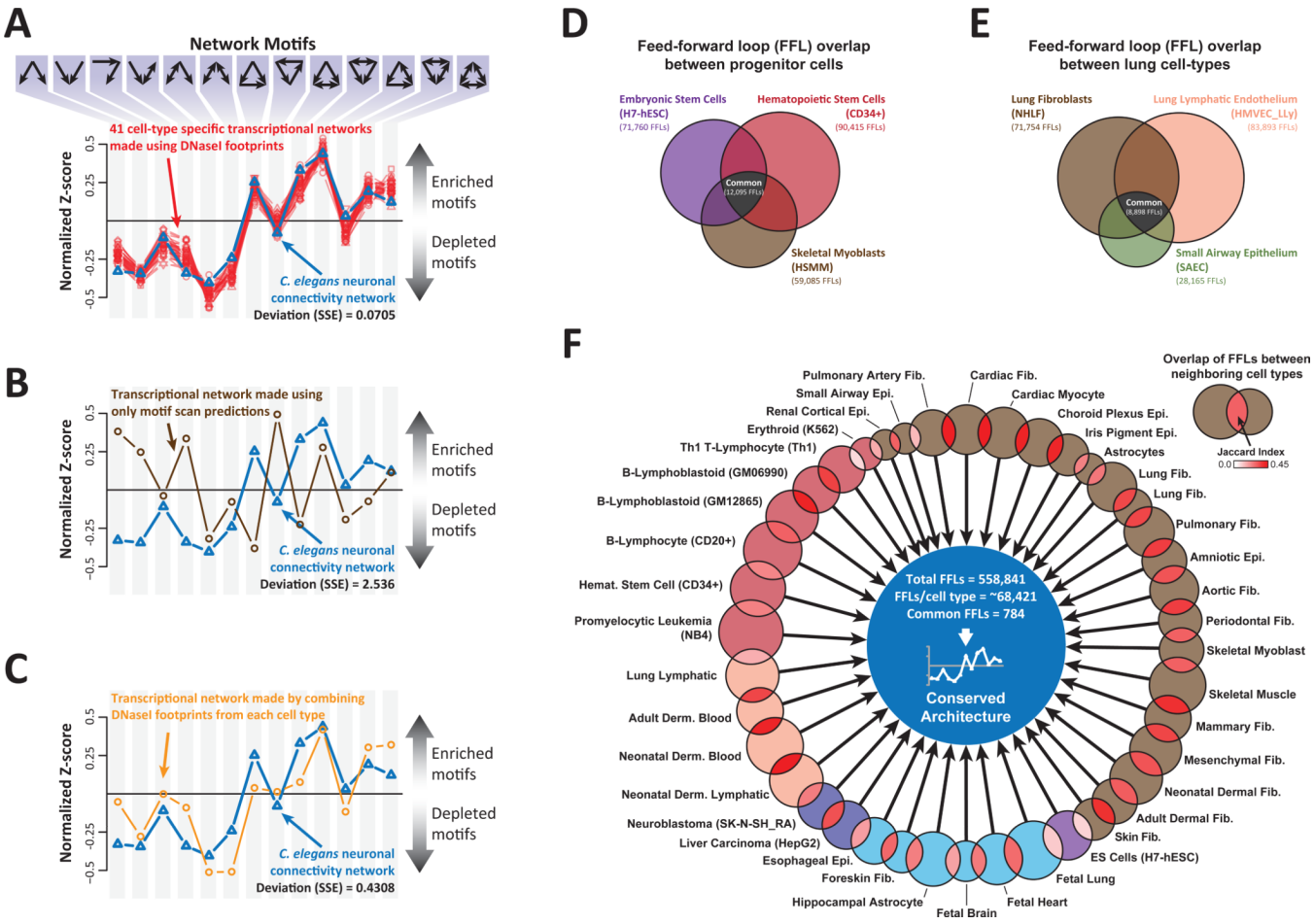

Figure 6. Conserved architecture of human transcription factor regulatory networks

(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 summed squared error (SSE) of only 0.0705).

(B) Enrichment of each triad network motifs for a transcription factor network computed using only motif scan predictions within +/− 5kb of TF promoters (brown line). The resulting network bares 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 transcription factor 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** – **E**) *Network architectures are highly cell-specific*

(D) Overlap of feed-forward loops (FFLs) identified in three different progenitor cell types - embryonic stem cells (H7-hESC), hematopoietic stem cells (CD34+) and skeletal muscle myoblasts (HSMM). Note that most FFLs are restricted to an individual cell type.

(E) Overlap of feed-forward loops (FFLs) identified in three pulmonary cell types - lung fibroblasts (NHLF), small airway epithelium (SAEC), and pulmonary lymphatic

endothelium (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 of panel F). 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 is indicated in the center of the graph. (See also Supplemental Figure S3 and Supplemental Table S3).