

# Distinct transcriptional regulatory modules underlie STAT3's cell type-independent and cell type-specific functions

Andrew Paul Hutchins<sup>1</sup>, Diego Diez<sup>1</sup>, Yoshiko Takahashi<sup>1</sup>, Shandar Ahmad<sup>2</sup>, Ralf Jauch<sup>3</sup>, Michel Lucien Tremblay<sup>4,5</sup> and Diego Miranda-Saavedra<sup>1,\*</sup>

<sup>1</sup>World Premier International (WPI) Immunology Frontier Research Center (IFReC), Osaka University, 3-1 Yamadaoka, Suita, 565-0871 Osaka, <sup>2</sup>National Institute of Biomedical Innovation, 7-6-8 Saito-Asagi, Ibaraki, 567-0085 Osaka, Japan, <sup>3</sup>Laboratory for Structural Biochemistry, Genome Institute of Singapore, 60 Biopolis Street, 138672, Singapore, <sup>4</sup>Goodman Cancer Research Center, McGill University, 1160 Pine Avenue, Montreal, H3A 1A3 QC, Canada and <sup>5</sup>Department of Biochemistry, McGill University, 3655 Promenade Sir William Osler, Montreal, H3G 1Y6, Canada

Received October 2, 2012; Revised November 11, 2012; Accepted November 14, 2012

## ABSTRACT

**Transcription factors (TFs) regulate gene expression by binding to short DNA sequence motifs, yet their binding specificities alone cannot explain how certain TFs drive a diversity of biological processes. In order to investigate the factors that control the functions of the pleiotropic TF STAT3, we studied its genome-wide binding patterns in four different cell types: embryonic stem cells, CD4<sup>+</sup> T cells, macrophages and AtT-20 cells. We describe for the first time two distinct modes of STAT3 binding. First, a small cell type-independent mode represented by a set of 35 evolutionarily conserved STAT3-binding sites that collectively regulate STAT3's own functions and cell growth. We show that STAT3 is recruited to sites with E2F1 already pre-bound before STAT3 activation. Second, a series of different transcriptional regulatory modules (TRMs) assemble around STAT3 to drive distinct transcriptional programs in the four cell types. These modules recognize cell type-specific binding sites and are associated with factors particular to each cell type. Our study illustrates the versatility of STAT3 to regulate both universal- and cell type-specific functions by means of distinct TRMs, a mechanism that might be common to other pleiotropic TFs.**

## INTRODUCTION

The precise spatio-temporal regulation of gene expression programs determines an organism's development and the

interaction with its environment. Transcription factors (TFs) control this process by binding to short DNA sequences (typically 6–8 bp), yet their binding specificities cannot explain the various cell type-specific functions of many TFs. Protein binding microarrays have shown that members of TF families such as homeodomains bind to very similar sequences, which therefore cannot account on their own for the enormous diversity of functional roles of homeodomain TFs during animal development (1,2). Potentially, cell type specificity emerges from the interplay of TF DNA sequence specificity, co-factors and epigenetics (3). However, despite vast efforts to understand the mechanisms that determine cell type-specific TF activity, the exact mechanisms continue to remain frustratingly elusive. A number of studies have shown that key TFs associate locally with co-activators to constitute 'transcriptional regulatory modules' (TRMs) that endow the key TF with cell type-specific functions. An important example was provided in embryonic stem cells (ESCs), where TFs assemble around the core heterodimer SOX2-OCT4 and NANOG (4). In hematopoietic progenitor cells, the TRM centers around GATA2, RUNX1 and SCL/TAL1 (5), whereas in developing B cells the TRM clusters around E2A, EBF1 and FOXO1 (6). Finally, in trophoblast stem cells, the TF core around which the TRM assembles includes SMARCA4, EOMES, TCFAP2A, GATA3 and ETS2 (and possibly STAT3 too) (7). Although experimentally characterized TRMs are very informative as to the co-activators that key TFs need to associate with to perform their biological functions, these TRM models have not yet been able to provide an explanation for how pleiotropic TFs bring about functional specificity in distinct cell types. Examples for the pleiotropic functions of TFs are

\*To whom correspondence should be addressed. Tel: +81 6 6879 4269; Fax: +81 6 6879 4272; Email: diego@ifrec.osaka-u.ac.jp

as follows: (i) the ESC factor SOX2 is also active in neural progenitor cells (8), (ii) the essential hematopoietic factor SCL/TAL1 is also robustly expressed in neural progenitor cells, (iii) the B-cell development factor FOXO1 is known to regulate adipocyte differentiation (9) and (iv) the trophoblast stem-cell factor GATA3 is crucial at various stages of CD4<sup>+</sup> T-cell development (10).

Therefore, a fundamental question in transcriptional regulation is how a given TF can perform highly divergent and at the same time crucial functions across distinct cell types (11). To address this problem, we set out to investigate the mechanisms that enable STAT3 to regulate distinctive gene sets leading to diverse biological outcomes in various cell types. STAT3 has been profiled by ChIP-seq in multiple cell types, including ESCs (4), CD4<sup>+</sup> T cells (12,13), macrophages (14) and AtT-20 corticotroph cells (15). Crucially, for the dissection of cell type-specific functions, STAT3 has radically different roles in each one of these cell types: in ESCs, STAT3 maintains pluripotency (16), whereas in CD4<sup>+</sup> T cells STAT3 drives the differentiation toward Th17 cells (13,17) and is also required for Th2 cells (18). In macrophages, STAT3 is essential for the initiation of the anti-inflammatory response mediated by IL-10 (19,20), and in AtT-20 corticotroph cells, STAT3 promotes adrenocorticotrophic hormone production as part of the hypothalamo–pituitary–adrenal axis in response to stress and inflammation (15,21). Clearly, these diverse functions imply that STAT3 is able to target different enhancers to regulate distinct genes depending on the biological context. Other advantages of using STAT3 as a model to investigate TF functional specificity in the four distinct cellular types described earlier are as follows: (i) STAT3 is an essential regulator in these cell types and cannot be replaced by other factors; (ii) STAT3 is activated upon induction by a cytokine and thus constitutes a natural switch that produces easily distinguishable outcomes and (iii) upon activation, STAT3 initiates a measurable response that is either a developmental program or a response to an environmental stimulus.

Here, we analyze genome-wide STAT3 binding data from ChIP-seq libraries profiled in ESCs, CD4<sup>+</sup> T cells, macrophages and AtT-20 cells and show that STAT3 has two modes of binding: (i) a small number of STAT3-binding sites that are common to all four cell types analyzed which regulate a core set of genes that are pre-bound by E2F1 and which encode a self-regulatory loop for STAT3 and (ii) the larger sets of STAT3-binding events that are cell type-specific and therefore responsible for the distinct biological outcomes of STAT3 in the various cell types. Moreover, by integrating data on predicted TF-binding sites, protein–protein interactions and gene expression, we built TRM models that predict the unique associations of STAT3 with distinct sets of co-factors and which explain how STAT3 directs both its cell type-independent and cell type-specific functions. This is the first example of a TF having both cell type-independent and cell type-specific functions mediated by distinct TRMs, a *modus operandi* that might be shared by other pleiotropic TFs.

## MATERIALS AND METHODS

### ChIP-seq read mapping and peak calling

ChIP-seq reads were mapped to the mouse genome (mm9) using Bowtie (v0.12.7) (22) with the setting ‘best’ and with reads mapping to more than one genomic location being excluded (–m). Peak discovery was performed using MACS (v1.4.1) (23) using the parameters bandwidth = 200 and genomesize = mm. The m-fold parameter was adjusted until MACS could find ~1000–2500 high-quality peaks to construct a model. The *P*-value for each ChIP-seq library was determined by increasing the *P*-value from  $1 \times 10^{-10}$  until the number of peaks discovered by chance alone (false positives) was close to 1% or the *P*-value reached  $1 \times 10^{-5}$ . Detailed genome mapping statistics and the full list of settings used by MACS for all of the ChIP-seq data used in this article are described in Supplementary Table S1.

Gene Ontology (GO) on the ChIP-seq lists was done using GREAT (v2.0.1) (24) using the whole genome as a background. Statistical analysis was performed using SciPy and data visualization used matplotlib, pycairo, R and glbase (<https://bitbucket.org/oaxiom/glbase/wiki/Home/>). For evolutionary conservation analysis, all pre-computed phastCons scores were obtained from the UCSC genome browser.

The random backgrounds used for motif enrichment analysis or for comparison were generated by randomly sampling 1 or 10% of the appropriate control ChIP-seq library. Sequence libraries were made unique, allowing only a single read per genomic position prior to sampling random sites. This removed repeats that typically attract a large number of sequence tags and give spurious regions of overlap. For the ‘shared overlap’, ‘any two cell types’ and ‘any three cell types’, the backgrounds were combined in equal proportions relative to the size of the contribution of the cell-type STAT3 binding to the overlap.

### Gene expression data

Expression profiles were obtained from the Gene Expression Omnibus for ESCs (GSE27708) (25), IL-21-treated anti-CD3/CD28-activated-CD4<sup>+</sup> T cells (GSE19198) (12), IL-6-treated anti-CD3/CD28-activated-naive CD4<sup>+</sup> T cells (GSE21671) (13), IL-10-treated peritoneal exudate cells macrophages (GSE31531) (14), IL-6-treated liver cells (GSE21060) (26) and AtT-20 cells (GSE19042) (21). For ESCs, AtT-20 and CD4<sup>+</sup> T cells, raw data were processed using the BrainArray custom CDFs (27), while the RNA-seq macrophage data for PEC macrophage RNA-seq, transcripts were quantified using RSEM (28) and annotated to Ensembl genes (release 66).

### Motif discovery and identification of TRMs

HOMER (v3.6) (29) was used for *de novo* motif discovery with default parameters, with  $\pm 200$  bp around the peak summits of the list of STAT3 binding peaks and a random background sampling of 1% of the appropriate control ChIP-seq library. TRMs were identified by integrating motif enrichment, gene expression and protein–protein interaction information. *De novo* motifs discovered by

HOMER that did not resemble STAT3 or a STAT3 half-site were searched again for enrichment using FIMO (30) to remove motifs with a  $z$ -score  $< 5.0$  in all STAT3 binding lists. *De novo* motifs detected by HOMER were annotated with Tomtom (MEME suite) (matching motifs with a  $q < 0.05$  were kept) (31) using a library of motifs obtained from the JASPAR and UniPROBE databases (32,33). Known motifs were mapped to human genes and pairwise similarities among motifs were computed with Tomtom to determine clusters of TFs with similar binding preferences. The original gene annotations for the motifs were mapped to the human homologs using the Biomart tool from Ensembl. These mapped genes were filtered by expression in the appropriate cell type to reduce redundancy. Based on the distribution of intensities, genes were considered not expressed if the normalized intensity value was  $< 5.2$  for ESCs,  $< 4.1$  in CD4<sup>+</sup> T cells or  $< 3.9$  in AtT-20 cells. For the macrophage RNA-seq data, those transcripts with a transcripts-per-million value  $< 1.0$  were disregarded. The human protein-protein interaction network was obtained from the BioGRID database (34) and nodes were removed for non-expressed genes in each of the data sets. Additionally, the nodes UBC and SUMO2 were also removed as they connect to 60 and 8% of the entire interactome, respectively. Full details on our method for reconstructing TRMs will be published elsewhere.

#### ChIP-qPCR of STAT3-bound sites in macrophages

PEC macrophages were purified from pooled 6- to 8-week-old male C57Bl6JcL mice (CLEA Japan) and treated with IL-10 (R&D Systems). Formaldehyde cross-linked chromatin was extracted and subjected to ChIP using antibodies to STAT3 (SantaCruz, sc-482), E2F1 (Millipore, 05-379) and GFP (Santa Cruz, sc-8334), as previously described (14). qPCR was performed on an ABI 7900, using SYBR qPCR mix (TOYOBO) according to the manufacturer's instructions. Primers used in this study are described in Supplementary Table S2.

## RESULTS

### A catalogue of STAT3 genomic binding sites in ESCs, CD4<sup>+</sup> T cells, macrophages and AtT-20 cells

To date, the genome-wide binding pattern of STAT3 has been reported by ChIP-seq in ESCs (4), CD4<sup>+</sup> T cells (12), Th17 cells (13), macrophages (14) and AtT-20 corticotroph cells (15). Although each publication provides an alignment to the genome and a set of characterized peaks, these were performed at different times where the sequencing reads were aligned to different mouse genome assemblies (mm8 and mm9), using various software tools (ELAND, AceView and bowtie) and a variety of peak discovery algorithms (MACS, ChIPseq and AceView). In order to compare the various ChIP-seq libraries, these and their corresponding control libraries (raw sequence reads) were uniformly reanalyzed, except for the STAT3 ChIP-seq library prepared in Th17 cells as it lacks a paired control library (13). Nevertheless CD4<sup>+</sup> T

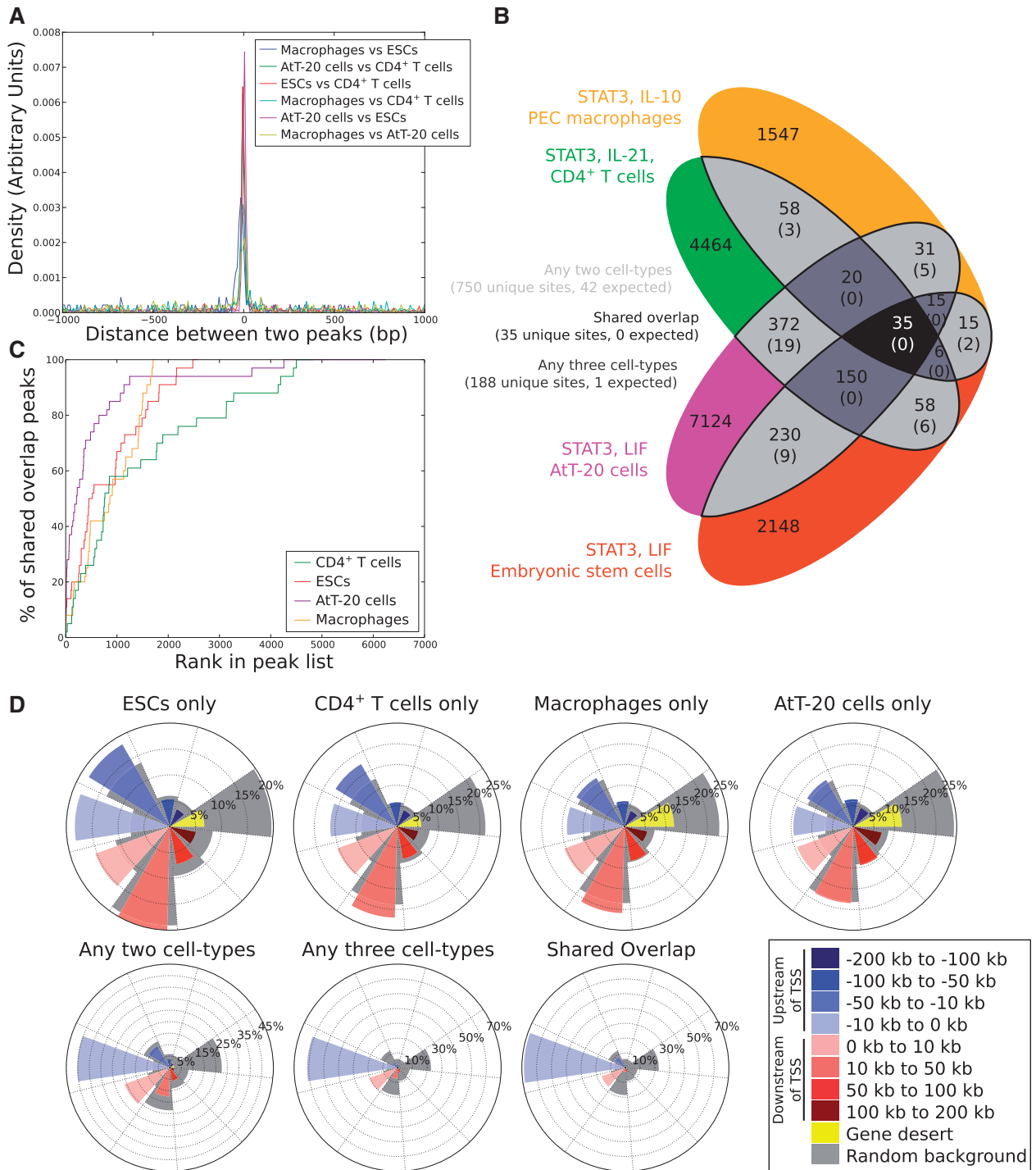
cells contain small numbers of Th17 cells, meaning that the CD4<sup>+</sup> T-cell ChIP-seq library will contain many Th17-specific STAT3-binding sites. For each cell type, ChIP-seq library replicates were merged into a single fastq file and bowtie (22) was used to align the reads to the mm9 version of the mouse genome. The peak discovery tool used was MACS (23), which reports the number of peaks discovered by chance alone by reversing the experimental and control libraries. At the default cutoff ( $P = 1 \times 10^{-5}$ ), MACS has a tendency to inflate the number of peaks reported for larger sequence libraries at the expense of increasing the number of false positives. To correct for this bias, we adjusted the MACS  $P$ -value cutoff to limit the number of peaks discovered by chance alone (false positives) to  $\sim 1\%$  (Supplementary Figure S1 and Supplementary Table S1). Thus, we report 2651 (ESCs), 5152 (CD4<sup>+</sup> T cells), 1724 (macrophages) and 7982 (AtT-20) peaks. Of the original sets of peaks defined in the respective publications, our new peak lists contain 66% of the ESCs peaks, 81% of the CD4<sup>+</sup> T-cell peaks, 87% of the macrophage peaks and 100% of the AtT-20 peaks. Although the majority of the peaks that were missed are lower ranked peaks (Supplementary Figure S2), we also report new additional peaks in CD4<sup>+</sup> T cells (756), ESCs (245), macrophages (372) and AtT-20 cells (4881). Similarly, most of these new peaks are lower ranked peaks by fold enrichment of the experimental libraries over the control libraries (Supplementary Figure S2).

### STAT3 regulates a core set of genes across all four cellular types

STAT3 peaks were defined by extending the summits determined by MACS 200 bp either side of each summit, and overlapping peaks were merged by taking the midpoint between the two summits. Whenever peaks overlapped, they tended to be close (Figure 1A) and only 35 peaks were found to be common to all four libraries (Figure 1B). These 35 peaks overlapping across all four libraries were highly ranked STAT3-binding sites, with one-half of these 35 overlapping peaks being listed in the top 1000 STAT3 peaks in all four libraries (Figure 1C). Moreover, a Monte Carlo simulation of the overlap suggested that the number of overlapping peaks that would be expected among all four libraries by chance alone is zero (expected overlapping sites:  $0.002 \pm 0.045$ ) (Figure 1B). Therefore, the 35 STAT3-binding events shared by all four ChIP-seq libraries are not random, while most STAT3-binding events are highly specific to each cell type (Supplementary Figures S3 and S4; Supplementary Table S3). Conservation of transcriptional regulation can also be gene-centric, where regulatory TF-binding sites do not overlap but still regulate the same nearby gene. This has been reported for SMAD3 (35) and during evolution for CEBPA (36). However, we did not find robust evidence for systematic gene-centric regulation as the number of gene-centric observations was close to the values expected by chance alone (Supplementary Figure S5).

The distribution of STAT3-binding events relative to gene locations is typical of other TF ChIP-seq libraries,





**Figure 1.** STAT3 binds to a small but significant ‘shared overlap’ of binding sites in divergent cell types, but is otherwise strongly cell type specific. (A) Summary of the distances between the overlapping peaks for each pair of STAT3 ChIP-seq libraries. The distance was measured as the number of base pairs between the summits of the two overlapping peaks. (B) Overlap of the binding peaks in three STAT3 libraries culled from mouse ESCs, CD4<sup>+</sup> T cells, macrophages and AtT-20 cells. Peaks were considered to overlap if their peak summits were within 200 bp of one another. The overlap was simulated by generating lists of faux ChIP-seq peaks followed by the assessment of their overlap (this was performed 1000 times to generate the number of overlaps expected by chance-value listed in brackets). (C) The ‘shared overlap’ sites are more likely found in the most highly ranked STAT3-binding sites. STAT3-binding sites were ranked by fold enrichment and then the cumulative overlap of STAT3 peaks appearing in all four cell types was plotted for ESCs, CD4<sup>+</sup> T cells, macrophages and AtT-20 cells. (D) Genome distributions of STAT3-binding sites orientated with respect to the nearest gene to the STAT3-binding site. Coloured bars describe the distance from the STAT3-binding sites to the nearest TSS, as described in the key. The gray regions denote a random background and represent the expected distribution of peaks were the binding sites randomly distributed across the sequenceable genome.

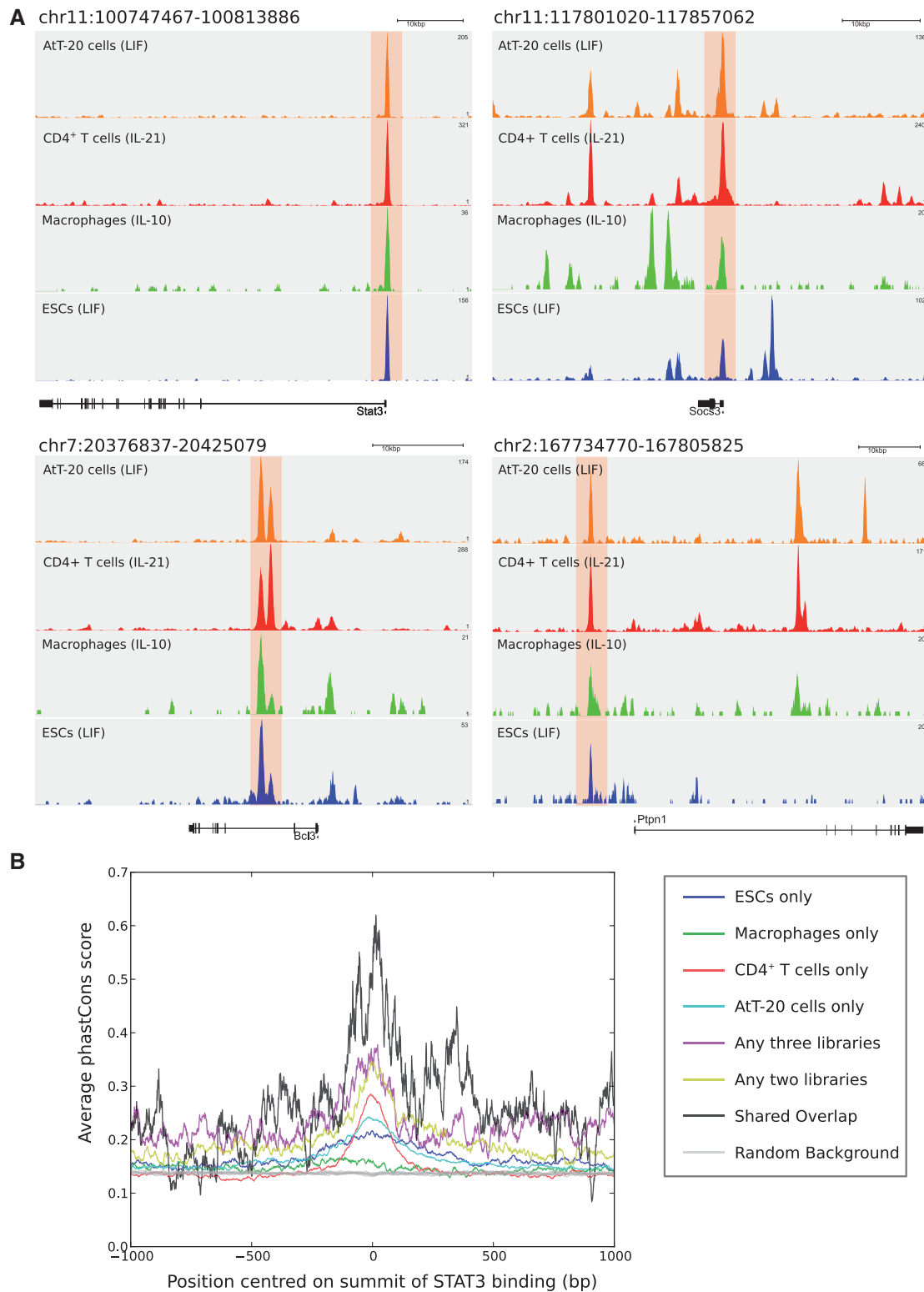
with one-third of STAT3 peaks located within 10 kb of the nearest TSS (Figure 1D and Supplementary Table S4). However, the ‘shared overlap’ of 35 STAT3-binding events across all four cellular types has a very strong bias toward the TSS, with 30 of 35 (~86%) STAT3 peaks being located within 10 kb of the TSS. As proximity to the TSS has been linked with the control of gene expression (4,37), this suggests that the 35 STAT3 peaks likely regulate the expression of the genes they lie closest to. Many of these genes are essential for STAT3 function, including Stat3 itself, Socs3, Bcl3 and Ptpn1 (Figure 2A). Not surprisingly the shared 35 STAT3-binding events show a much greater degree of evolutionary conservation than any set of cell type-specific STAT3 peaks (Figure 2B). Most of the 35 STAT3-binding events shared across all four cellular types are located within 10 kb of a TSS, although the assignment between TF-binding events and gene regulation is still an open question as no generally applicable model has yet been described. The association by proximity of TF-binding events with the genes nearby shows that typically only 10–20% of such genes are differentially expressed (10,14,38). To investigate the effects of the 35 STAT3-binding events on the expression of the closest genes, we analyzed expression data for the four cellular types analyzed here plus two additional conditions where STAT3 is also activated by a cytokine: naive CD4<sup>+</sup> T cells stimulated by IL-6 (13) and liver cells stimulated by IL-6 (for which the STAT3 genome-wide binding profile is not known) (26). Gene expression values of the cytokine-stimulated cells were combined and ranked by fold-change relative to controls, while the expression data from ESCs were inverted (as the withdrawal of LIF causes ESCs to differentiate). Remarkably, about one-half of the genes are up-regulated upon cytokine stimulation in all six conditions (Figure 3B) and strikingly only a few genes are down-regulated. This indicates that a core unit of STAT3 regulation occurs in all these biological contexts and cellular types, including the well-characterized targets of STAT3, Socs3 and Bcl3, which are up-regulated in all six conditions.

The 35 STAT3-binding events common to all four cell types regulate a diverse set of genes involved in the maintenance of specific cellular activities. GO analysis using GREAT (24) reported an over-representation of signal transduction pathways related to STAT3 (Figure 3A). The promoters of Stat3 and Stat1 are both bound and induced by the recruitment of STAT3 in all four cell types (Figure 3B). Moreover, it is remarkable to find that three genes involved in three different pathways that negatively regulate JAK-STAT signaling are induced by STAT3, suggesting that STAT3 is acting to regulate its own activity. These include PTP1B (Ptpn1), a protein tyrosine phosphatase reported to dephosphorylate JAK2 and TYK2 and also phospho-STAT3 (39). Additionally, SOCS3 is part of the ubiquitin pathway that negatively regulates JAK-STAT signaling by degrading both JAKs and STATs (40). A third mechanism of negative feedback is the suppressive function of the RNA-binding protein Zfp36 (Tristetraprolin) which negatively regulates IL-10 signaling in macrophages (41). Zfp36 could thus be playing an identical role in all four

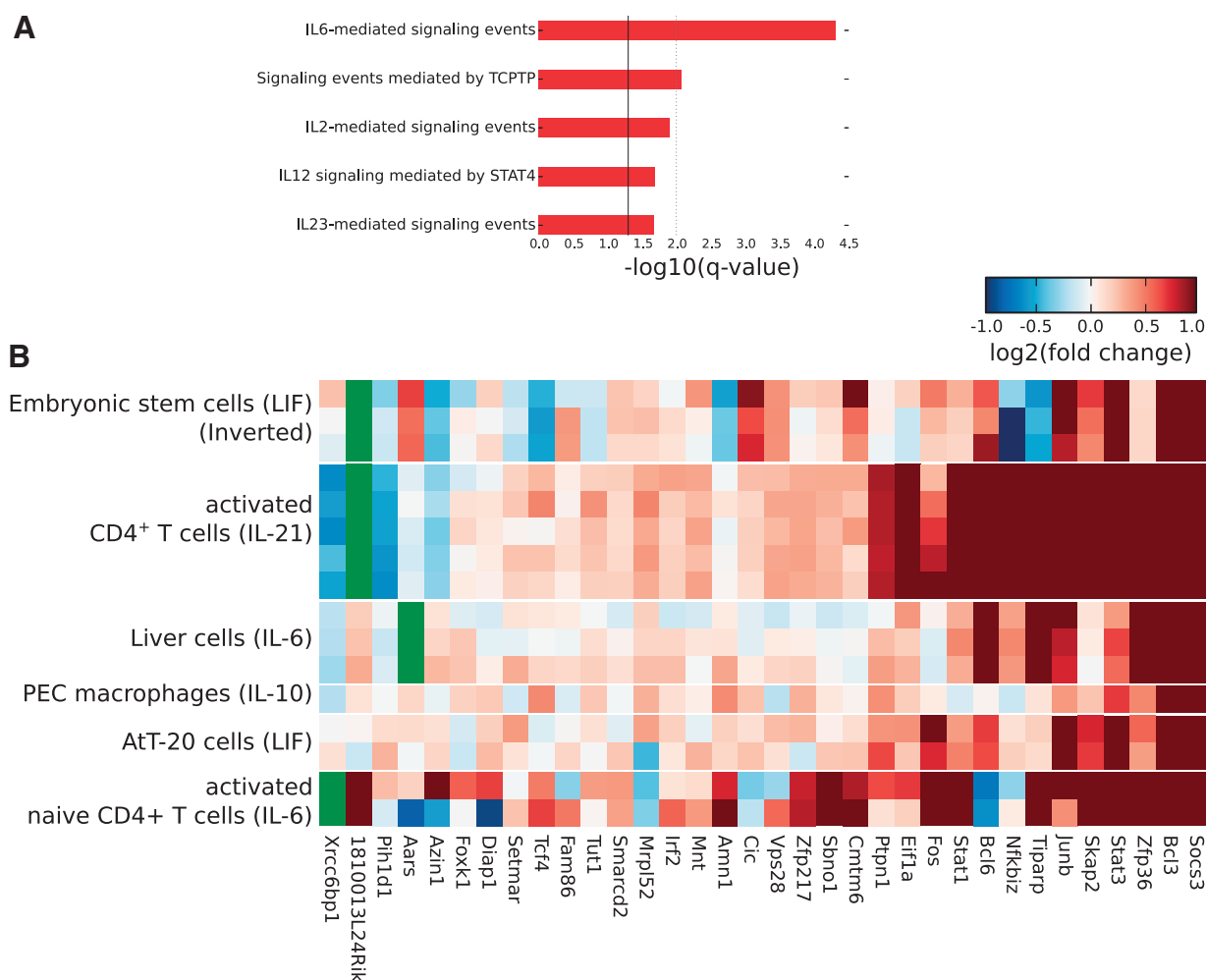
cell types described here. The remaining genes include over a dozen TFs, such as Stat3, Stat1, Bcl3, Bcl6, Tcf4, Cic, Sbn1 and the AP-1 family members Fos and Junb (GO:0003676: ‘nucleic acid binding’, GREAT FDR  $q = 3.6 \times 10^{-2}$ ), which deserve further characterization. In addition to these, our STAT3 target set comprises several genes encoding proteins involved in DNA replication and repair, protein translation, cytoskeletal reorganization and protein trafficking and notably six genes that encode proteins involved in metabolism (summarized in Supplementary Table S5). Our findings on the STAT3 transcriptional program common to all four cellular types indicate that STAT3 establishes its own regulatory network by the following: (i) perpetuating its own transcription, (ii) being a master regulator of other TFs working downstream of it, (iii) stimulating the transcription of cytoplasmic enzymes that control STAT3’s activity and (iv) ensuring an efficient and robust cellular division program and the maintenance of a stable cell type. The STAT3 transcriptional program involves many levels of cellular control from basic DNA replication and chromatin remodeling, to the cell metabolic pathways producing key metabolites needed for increasing the transcriptional and translational processes essential in cell division and maintenance.

#### **A distinct TRM defines cell type-independent STAT3 binding across all four cellular types**

Since STAT3 binds to 35 identical sites across four distinct cellular types, it is reasonable to assume that it does so by assembling around a TRM that is common to all cellular types. To reconstruct the putative TRM that directs the expression of the genes regulated by the 35 STAT3-binding events, we integrated over-represented TF-binding sites co-occurring with STAT3-binding sites, protein–protein interaction data and expression data (as detailed in the ‘Materials and Methods’ section). The resulting TRM (Figure 4A and Supplementary Figure S6) is unique to the 35 STAT3-binding events shared by all cellular types and contains a number of co-factors and other proteins that are known to bind to STAT3 experimentally and whose corresponding genes are expressed in all four cellular types. The co-TFs that appear to work together with STAT3 in this TRM include many ‘general’ TFs known to operate in a variety of biological contexts. The cell type-independent TRM contains the TFs MYC, E2F1 and KLF4, all of which have been profiled by ChIP-seq in ESCs (4). These ChIP-seq libraries were re-analyzed as before (full details included in Supplementary Table S1) and we determined the co-occupancy of the 35 STAT3-binding events by MYC, E2F1 and KLF4 in ESCs. We found that 34 of 35 (97%) of the STAT3 peaks are co-occupied (within 800 bp of each other) by n-MYC, E2F1 or KLF4 and often by several of them (Figure 4B and Supplementary Figure S7), but only 19 of 35 (54%) are co-occupied by one of the ESC-specific factors ESRRB, SOX2, OCT4 or NANOG (Figure 4B). It must be noted that co-occupancy is not a factor of the number of peaks of the ChIP-seq libraries as ESRRB has the largest number of peaks (56/136) as



**Figure 2.** The shared overlap of STAT3-binding sites controls the expression of a core set of genes important for STAT3 function. **(A)** Genomic views of STAT3 ChIP-seq sequence tag densities for four STAT3-binding sites common to ESCs, CD4<sup>+</sup> T cells, macrophages and AtT-20 cells. STAT3 binding is shown around the TSS of Stat3 and Socs3, within the second intron of Bcl3 and ~6 kb 5' of the TSS of Ptpn1. **(B)** Average evolutionary conservation for the different categories of STAT3-binding events. STAT3 peak summits were extended by 1 kb either side and the Euarchontoglires evolutionary conservation scores were annotated as determined by phastCons in the UCSC Genome Browser (<http://genome.ucsc.edu/>).



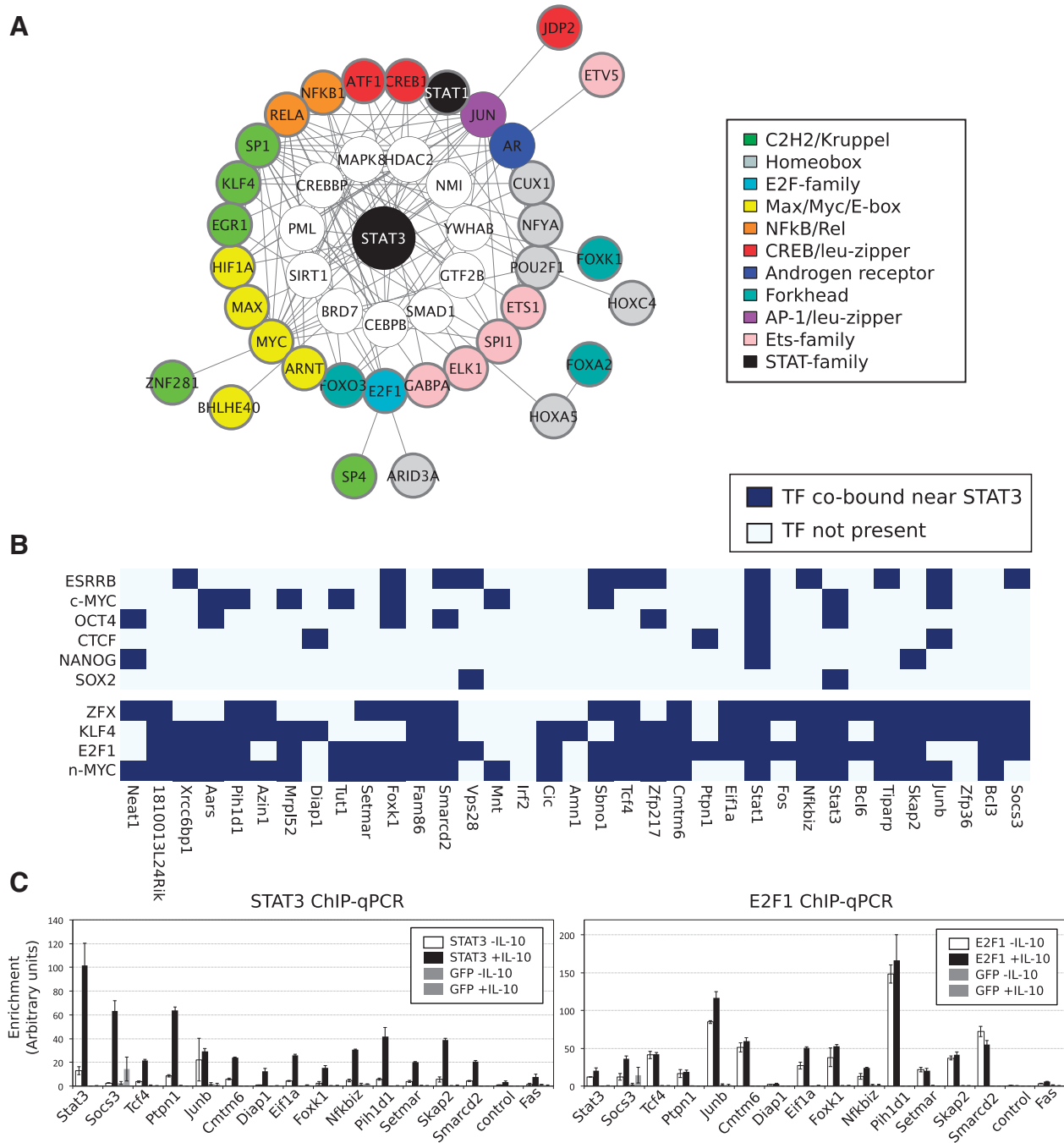
**Figure 3.** The shared overlap STAT3-binding sites co-regulate a set of key genes important for STAT3 function in multiple cell types. **(A)** Significantly over-represented terms from the ‘Pathway Commons’ category for the genes associated with the shared overlap. Shown here are the top five terms only. A significant  $q$ -value of 0.05 is represented by a solid black line and a  $q$ -value of 0.01 by a dotted gray line. **(B)** The expression of the closest genes within 200 kb of the shared overlap was measured in a series of gene expression microarray data sets which show the activation of STAT3 or the loss of STAT3 activity (LIF withdrawal from ESCs). Expression data are reversed in ESCs for clarity, but is otherwise down-regulated upon removal of LIF whereas all other treatments show up-regulation in response to cytokine stimulation.  $P$ -values are from a Wilcoxon test between the treated and untreated conditions: CD4<sup>+</sup> T cells treated with IL-21 ( $P = 5.59 \times 10^{-5}$ ) (GSE19198), naïve CD4<sup>+</sup> T cells treated with IL-6 ( $P = 2.45 \times 10^{-3}$ ) (GSE21671), ESCs upon withdrawal of LIF from the medium ( $P = 0.027$ ) (GSE27708), peritoneal macrophages stimulated with IL-10 ( $P = 2.14 \times 10^{-4}$ ) (GSE31529), AtT-20 cells treated with LIF ( $P = 1.00 \times 10^{-4}$ ) (GSE19042) and liver cells stimulated with IL-6 ( $P = 0.032$ ) (GSE21060). Genes marked in green do not have a corresponding probe on the microarray.

opposed to E2F1 (11448 peaks) (Supplementary Table S1). Finally, for a limited number of STAT3-binding sites in the shared overlap, we also probed their occupancy in IL-10 treated macrophages by ChIP-qPCR (Figure 4C). As expected, STAT3 is specifically recruited to all of the 14 sites we probed by ChIP-qPCR and remarkably E2F1 is not only bound at these same sites but is actually pre-bound at these sites close to STAT3 binding (with the exception of Diap1, Figure 4C). STAT3 is therefore specifically recruited to genomic loci that already have E2F1 bound. This may explain why several of these genes show very rapid induction of expression upon stimulation with IL-10 (14,42) since they are poised for expression by the presence of E2F1. In summary, STAT3 appears to use a TRM, binding close to E2F1 and possibly MYC, to regulate a core set of genes that tune

the JAK-STAT pathway and to promote cell proliferation while counteracting differentiation.

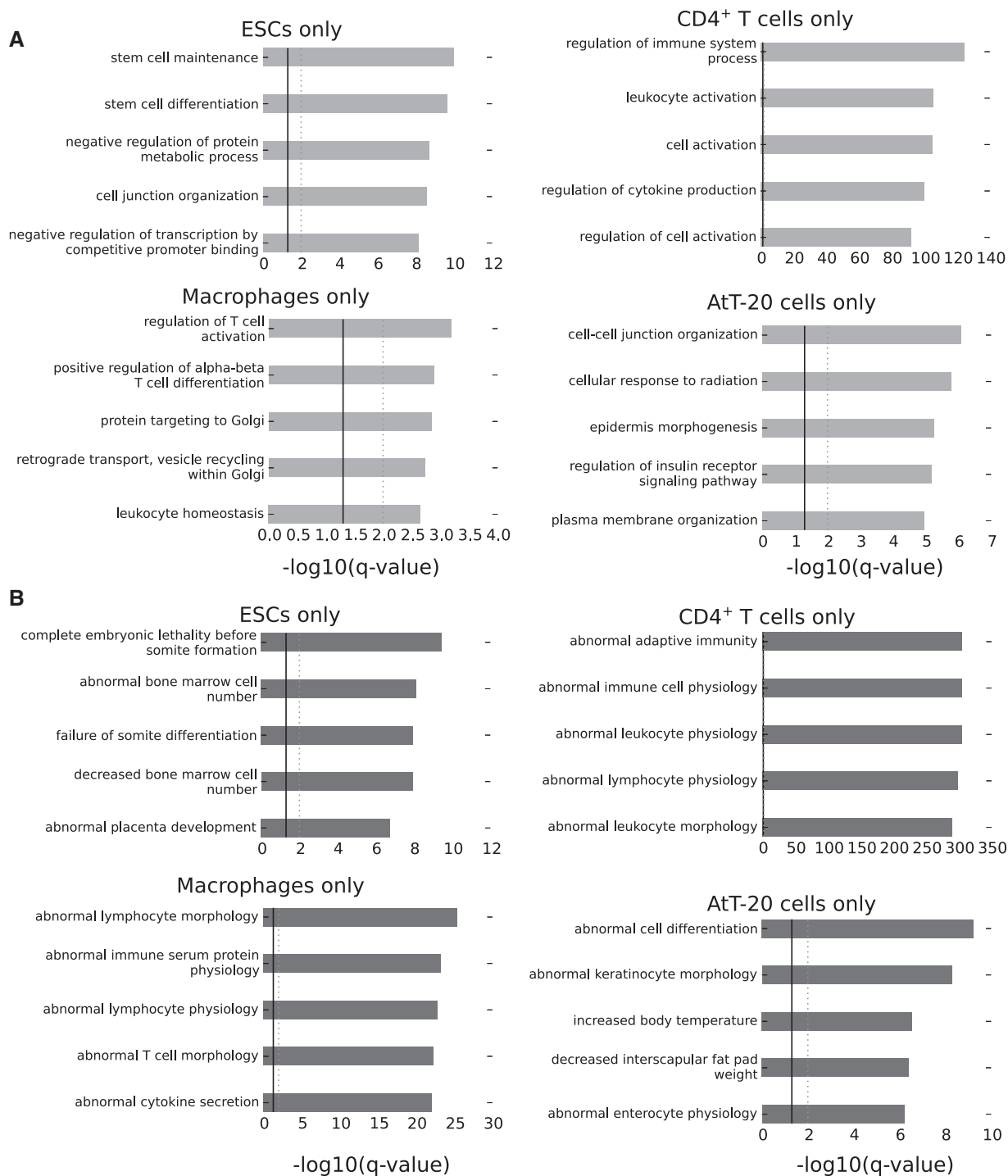
#### Cell type-specific binding events determine the various functions of STAT3 in ESCs, CD4<sup>+</sup> T cells, macrophages and AtT-20 cells

Since the 35 STAT3-binding events shared across all four cell types encode a self-regulatory program for STAT3 that is cell type-independent, the cell type-specific STAT3-binding events should be related to the specific functions of STAT3 in each of the four distinct cell types. GO analysis of the cell type-specific STAT3-binding events using the ‘Biological Process’ category shows that the most over-represented terms in ESCs include distinctive ESC functions such as ‘stem-cell maintenance’ and ‘stem-cell differentiation’ (Figure 5A).



**Figure 4.** The shared overlap of STAT3 binding in all four cell types forms a cell type-independent regulatory network with MYC and E2F1. (A) HOMER was used to generate *de novo* motifs from the list of 35 STAT3-binding sites common to all four cell types. Motifs resembling STAT3 or a STAT3 half-site were removed and over-represented motifs were collected and annotated to genes. Interaction networks were constructed by interrogating the PPI network for proteins interacting with those representing the enriched motifs. TFs were clustered together by motif similarity and coloured by the cluster they belong to: white-colored nodes do not have a representative motif in the databases or do not bind to DNA directly; proteins with a bold circle have a motif enriched in that cell type, while proteins with no bolded circle have no discovered motif but were linked to STAT3 through the PPI network. Proteins in the network were filtered by gene expression and here we present the union of the network in all four cell types (the separate networks are presented in Supplementary Figure S6) (B) ChIP-seq data from ESCs (GSE11431) were re-analyzed and binding sites overlapping within 400 bp were collected. The heatmap shows the 35 STAT3-binding sites together with the other TFs bound in the vicinity of STAT3. (C) We designed primers for 14 STAT3-binding sites shared between all four cell types and performed ChIP-qPCR. Macrophages were treated for 4 h with IL-10 (black bars) or left untreated (white bars) and chromatin was harvested. ChIP was performed using antibodies against STAT3 (left panel), E2F1 (right panel) or GFP (as a control in both panels). Each group of bars represents a STAT3-binding site and is labeled with the name of the nearest gene.





**Figure 5.** STAT3 biological specificity is found within the cell type-specific lists of genomic binding sites. Gene Ontology term enrichment analysis was done using GREAT with default parameters. Over-represented terms displayed here are from the ‘Biological Process’ category (A) and the Mouse genome informatics phenotype::genotype category (B). A significant  $q$ -value of 0.05 is represented by a solid black line and a  $q$ -value of 0.01 by a dotted gray line.

Likewise, in CD4<sup>+</sup> T cells and macrophages terms pertaining to relevant processes were recovered. For AtT-20 cells, as a pituitary epithelial cell line, there are several terms related to epithelial cell function, including ‘cell–cell junction organization’, ‘cellular response to radiation’, ‘plasma membrane organization’ and particularly ‘regulation of insulin receptor signaling pathway’, indicating the

role that the pituitary plays in responding to insulin. The over-represented terms from the Mouse Genome Database genotypes::phenotypes (43) are also indicative of specific STAT3 functions in the four cellular types, including ‘abnormal cytokine secretion’ (macrophages), ‘abnormal adaptive immunity’ (CD4<sup>+</sup> T cells), ‘complete embryonic lethality’ (ESCs) and ‘increased body

temperature' (AtT-20 cells, since one of the functions of the pituitary is to regulate body temperature) (Figure 5B). Collectively, the GO analyses of the cell type-specific STAT3-binding events suggest that the functional specificity of STAT3 in each cell type is contained within the cell type-specific lists of STAT3-binding events and not within the shared overlap of peaks common to all four cellular types, which as we have previously shown has diverse cell type-independent functions.

### Non-canonical DNA binding is prevalent across distinct cellular types but is not sufficient to explain STAT3's divergent functions

The DNA binding preferences for paralogous classes of TF tend to be rather uniform, making it difficult to attribute biological specificity to DNA base changes in TF-binding motifs. A clear example is provided by homeodomain TFs, all of which have strikingly similar DNA binding preferences despite encompassing all of mammalian development (2). The case of STAT3 is particular because although STAT3 has a well-defined canonical motif (TTCnnnGAA), it has nevertheless been reported to use a unique variant motif (TTAnnnGAA) to regulate the *Prdm1* gene in CD4<sup>+</sup> T cells (12).

The general motif recovered *de novo* in each set of cell type-specific STAT3-binding events is the prototypical STAT3 motif (Figure 6A), which is shared by other STAT factors too (44). Since alternative motifs appear to be important for STAT3 function, we investigated the presence of all possible STAT3 motif variants in the set of STAT3 peaks in each cellular type. This was done by counting STAT3 motifs where single base pairs were individually mutated and *z*-scores were derived by comparing against randomly selected sets of background sites drawn from the respective ChIP-seq control libraries. As expected, the STAT3 canonical binding site was found in all four cellular types and the shared overlap (Figure 6B). The variant TTAnnnGAA previously characterized in CD4<sup>+</sup> T cells (12) was over-represented not only in CD4<sup>+</sup> T cells but also in ESCs and AtT-20 cells (but not in macrophages). The variant STAT3 motif TGCnnnGAA is over-represented in all four cellular types, suggesting that this is a common alternative mode of DNA binding by STAT3. Finally, the motif CTCnnnGAA appears to be a variant used by STAT3 exclusively in AtT-20 cells (Figure 6B).

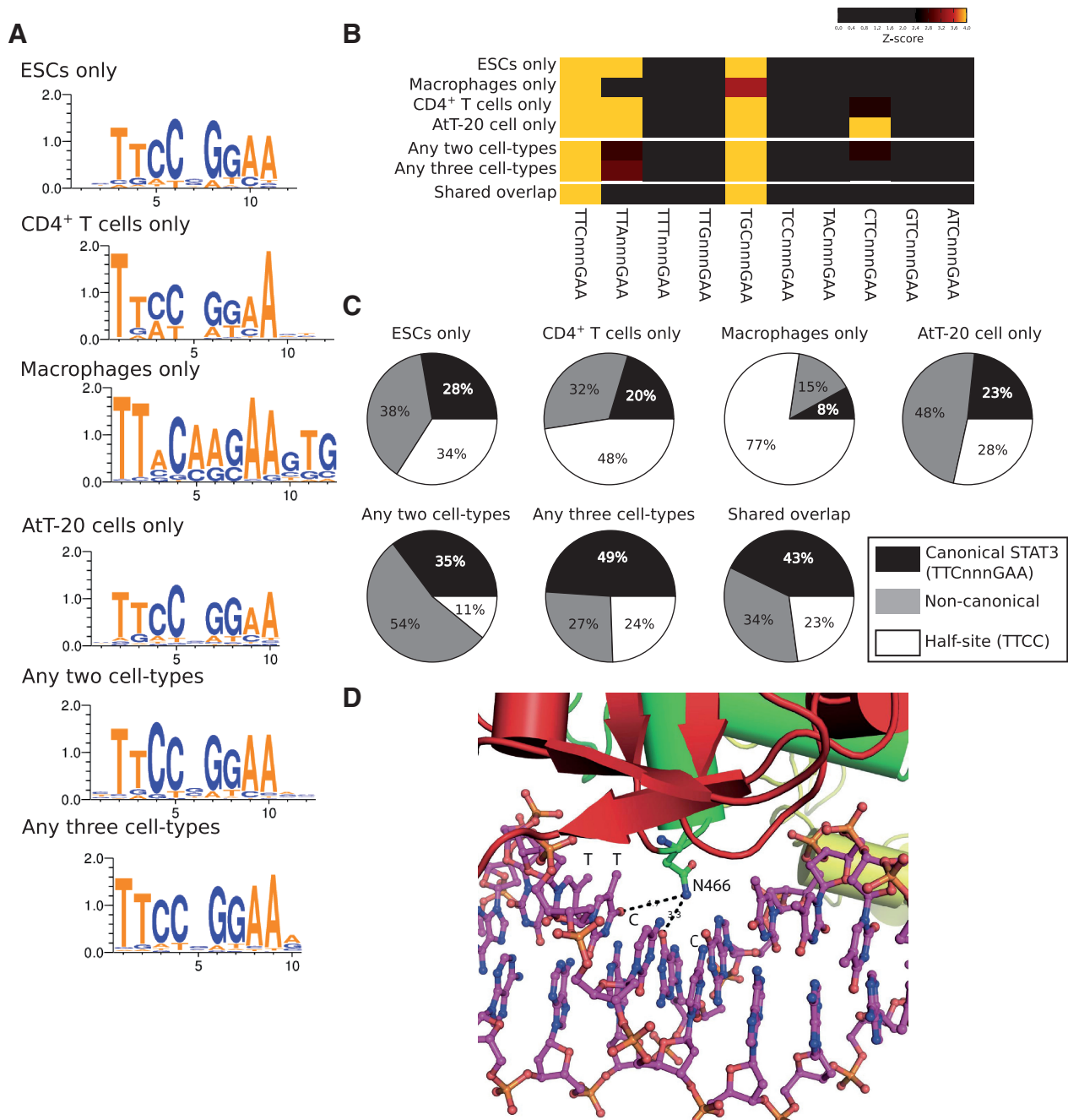
Moreover, we quantified the usage of the various canonical and variant motifs for each set of STAT3 peaks (Figure 6C). All STAT3 peaks contain putative half-sites (TTCC), often several of them per peak. At most 28% of STAT3 peaks harbor a perfect canonical site (in ESCs) while non-canonical sites are especially prevalent in AtT-20 cells (48% of sites). The set of 35 STAT3-binding events shared by all four cellular types turned out to be especially conservative in the use of canonical motifs, with 43% containing a canonical STAT3 motif. The binding of STAT3 to variant motifs could be linked to specific functions: for instance, the binding of STAT3 to canonical TTCnnnGAA might facilitate the recruitment of generic co-activators, such as p300, whereas the

binding to the variant motif TTAnnnGAA (as is the case in CD4<sup>+</sup> T cells) might lead to the recruitment of specific regulatory complexes. The differential recruitment of co-activators could be allosterically induced by different DNA ligands, as demonstrated for the glucocorticoid receptor (45), a model that provides a conceptual framework for understanding the functional plasticity of pleiotropic TFs such as STAT3. Here, the binding of STAT3 to variant DNA motifs induces conformational changes on the binding proteins, which in turn affect the recruitment of co-activators. Indeed, in the case of STAT3, specific base pairs that vary between the canonical and the variant motifs TTAnnnGAA and TGCnnnGAA (i.e. T[TC] versus T[TA] or T[GC]) are directly contacted by Asn466 of the connector domain of STAT3, in which case Asn466 must structurally rearrange to accommodate the altered chemical environment introduced by these variant motifs (Figure 6D) (46). These structural rearrangements could translate into global structural changes to the connector domain as well as the neighboring and C-terminal SH2 domain that mediates interactions with other proteins by recognizing phosphotyrosine residues. A further effect of specific variant motifs and specific co-factor interactions might be the stabilization of the TRM and the tethering of STAT3 to DNA. Indeed, Husby *et al.* (47) recently proposed that the two halves of the STAT3 homodimer do not make identical base pair contacts on a TTAGnGGAA variant motif. We found no clear link between the presence of variant motifs and specific biological functions, although there were several prominent examples including the aforementioned *Prdm1*, *I117a* and *I117ra* (both of which have nearby STAT3-binding sites with the variant motif TTAnnnGAA) and *Cd28*, which is associated with the STAT3 variant motif TGCTggGAA. The presence of alternative STAT3-binding motifs highlights the usage of variant motifs in regulating specific genes, although no clear pattern that explains how variant motifs encode the cell type-specific functions of STAT3 could be discerned.

### Distinct TRMs determine STAT3's cell type-specific functions

In the absence of a clear contribution by variant DNA motifs to explain STAT3's cell type-specific functions, an attractive model providing biological specificity is one where various factors cluster together in a cooperative manner to provide biological specificity (48). This 'piling-up' of TFs and co-factors has been demonstrated by ChIP-seq experiments in ESCs, B cells and blood stem-cell precursors (4,5,29). One limitation of this approach, however, is that the identities of biologically relevant TFs need to be known in advance before performing any ChIP-seq experiment that will eventually allow one to reconstruct a transcriptional regulatory network (49).

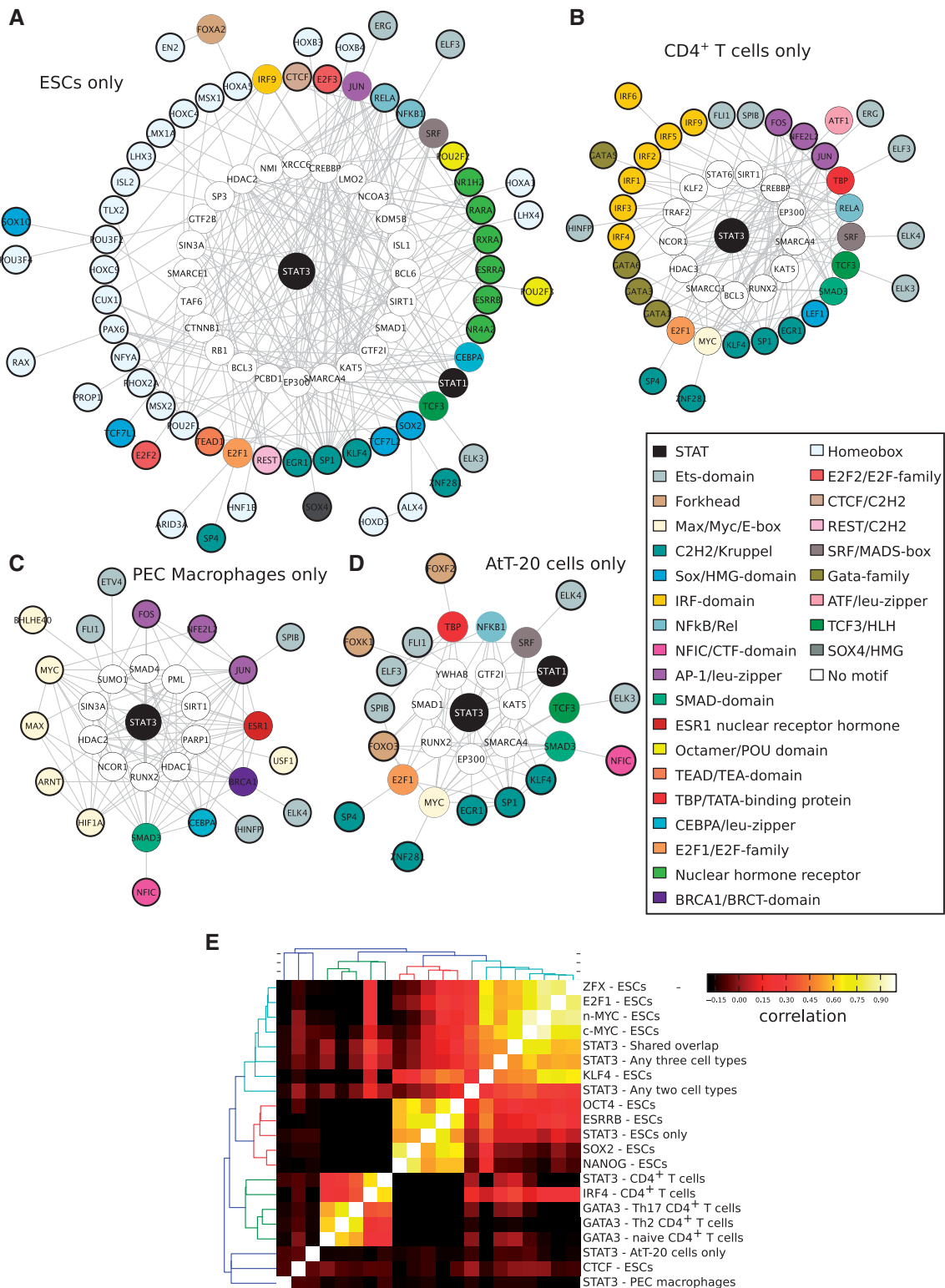
By employing the same data integration and analysis techniques that we used to reconstruct the cell type-independent TRM regulating the shared overlap of 35 STAT3-binding events (Figure 4A), we generated models for the cell type-specific binding events of STAT3 in ESCs, CD4<sup>+</sup> T cells, macrophages and AtT-20 cells



**Figure 6.** STAT3 uses alternative (non-canonical) modes of binding to DNA. (A) *de novo* generated motifs from HOMER are very similar in the six lists of STAT3 binding. Motifs were generated from the entire lists of STAT3-binding sites for each category, except for CD4<sup>+</sup> T cells where the top 1000 sites were used (as for the entire list we could only identify a STAT3 half-site). (B) z-score heatmap to show over representation of variant STAT3 motifs in the STAT3-binding sites in the various cell types. Variant motifs are presented here as all single base pair mutations of one-half of the STAT3 heterodimeric motif. (C) Pie charts showing the frequency of the DNA words TTCnnnGAA (canonical STAT3) or a non-canonical STAT3 binding: TTAnnnGGA and TGCnnnGGA for ESCs, CD4<sup>+</sup> T cells, any two cell types and any three cell types; TGCnnnGAA for macrophages and the shared overlap; and TTAnnnGGA, TGCnnnGGA and CTCnnnGAA for AtT-20 cells. (D) Cartoon representation of the Asn466 amino acid of STAT3 making contact with the DNA base pairs [PDB entry 1bg1 (46)].

(Figure 7A–D). We managed to recover the well-characterized ESC transcriptional regulatory network (4) comprising OCT4 (represented here by the motif POU2F2), SOX2, ESRRB, KLF4 and SMAD1 and a number of homeodomain proteins that may correspond to NANOG (Figure 7A). In addition, we identified

TEAD1, a protein recently shown to be essential to maintain ESC pluripotency (50) and also identified in human ESCs as a potential key TF (51), and REST, whose role is to repress specific genes in ESCs and so block differentiation (52). This network is reminiscent of the ESC networks determined by mass spectrometry



**Figure 7.** Discrete transcriptional regulatory modules (TRM) determine STAT3 cell type-specific functions. TRMs for each cell type were generated by collecting de novo motifs generated by HOMER and excluding motifs resembling STAT3 or a STAT3 half-site. Motifs were then scanned again and compared to a random background generated from the control background libraries to generate a *z*-score. Motifs with a *z*-score < 5.0 were removed and finally motifs were annotated to genes. These genes were filtered by expression data derived from the appropriate cell type and interaction networks were constructed by interrogating the PPI space. TFs were clustered together by motif similarity and coloured by the cluster they belong to. Proteins with a bold circle have an motif enriched in that particular cell type, while proteins without a bold circle have no enriched motif but were linked to STAT3 through the PPI network. White-coloured nodes do not have a representative motif in the motif databases or do not bind to DNA directly. The networks are divided into the cell type-specific lists for ESCs (A), CD4<sup>+</sup> T cells (B), macrophages (C) and AtT-20 cells (D). (E) Pairwise correlation heatmap of the ChIP-seq peaks from ESCs, CD4<sup>+</sup> T cells and the STAT3-binding sites described here. The lists of peaks were assessed for overlap in an all-against-all comparison and Pearson correlation scores were used to measure the frequencies of overlapping peaks when compared to all other overlaps.



(53,54), although it is only a partial match possibly because STAT3 was not used as a bait in the generation of those networks.

The transcriptional regulatory network of CD4<sup>+</sup> T cells is much less well characterized, but we managed to identify the only known STAT3-binding partner in the CD4<sup>+</sup> T cells, IRF4 (12) (Figure 6B). Our model of the TRM of CD4<sup>+</sup> T cells is completed by members of the TF families AP-1, several ETS-related TFs that are known to play important roles in T-cell biology (55) and a diversity of other TFs. ATF1 is recruited to the *Ifng* promoter in CD4<sup>+</sup> T cells and with CREB represses *Ifng* expression in naive CD4<sup>+</sup> T cells (56). Tantalizingly, STAT3 is also bound at the *Ifng* promoter and may co-operate with ATF1 to regulate *Ifng* expression. Finally, we also identify multiple GATA factors associated with STAT3 (Figure 7B), of which GATA3 is known to have multiple roles in T-cell development (10).

In macrophages, several classes of TF were identified (Figure 7C), the most important of which are members of the AP-1 family (FOS, NFE2L2 and JUN), where STAT3 might in turn be regulating the expression of FOS and NFE2L2 as it binds very close to their TSS and CEBPA, which is known to reprogram pre-B cells into macrophages (57).

The STAT3 transcriptional regulatory network of AtT-20 cells is less well characterized than those of ESCs and CD4<sup>+</sup> T cells, and our TRM model suggests many potential partners for STAT3 in AtT-20 cells (Figure 7D). We did not recover any motif matching the glucocorticoid receptor-binding site, which although it is linked to the function of STAT3 in AtT-20 cells, it co-localizes with STAT3 in only 21% of STAT3-binding sites, of which ~50% require the presence of glucocorticoid to co-recruit GR and STAT3 (15).

Finally, a pairwise correlation among the TF-specific binding events derived from CHIP-seq libraries done in ESCs, and all of the STAT3-binding events derived likewise (across all four cellular types) show the distinct TRMs characteristic of ESCs and T cells (Figure 7E). These TRMs contain both the previously reported TRM specific to ESCs (OCT4, SOX2, ESRRB and NANOG) and the 'MYC-regulated' block (MYC, E2F1). Additionally, a new TRM consisting of STAT3, IRF4 and a more loosely associated GATA3 emerged in various T-cell subsets. STAT3 binding in macrophages and AtT-20 cells have no other partner TFs from the published CHIP-seq libraries and hence form isolated clusters. These results demonstrate that STAT3 forms distinct cell type-specific TRMs to execute a diversity of gene expression programs.

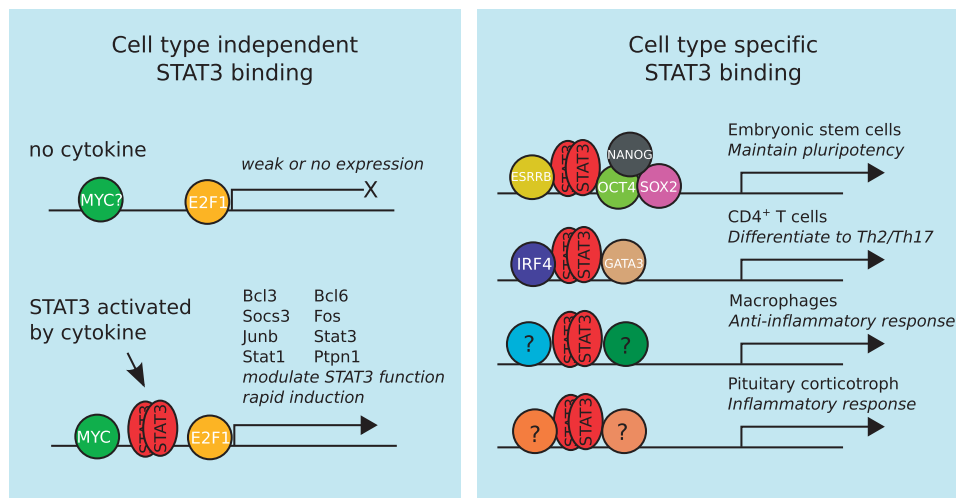
## DISCUSSION

In this study, we have investigated the factors that determine the functions of STAT3 in various cell types. STAT3 binding is predominantly cell type-specific, with just a small but significant number (35) of binding sites shared among ESCs, CD4<sup>+</sup> T cells, macrophages and AtT-20 cells. This shared overlap appears to be a mode of

auto-regulation for STAT3 function and targets many genes that are all directly up-regulated by STAT3 binding. Additionally, STAT3 also co-binds with E2F1, which is present at the same STAT3-binding sites in at least ESCs and macrophages, and we demonstrate that at least in macrophages E2F1 is even pre-bound to the future STAT3-binding sites. This may explain why so many of these STAT3-regulated genes are rapidly induced in <1h upon stimulation with a cytokine. The recruitment of STAT3 to DNA via variant motifs does not appear to be a major factor regulating STAT3's functions across diverse cell types. Nevertheless, a more plausible explanation is provided by the assembly of distinct TRMs around STAT3 in the four cell types. We produced models of TRMs by integrating TF-binding site data, protein-protein interaction and expression data that provide an explanation for STAT3's specific biological functions across distinct cell types. Using just the STAT3-binding events in ESCs, we succeeded in recovering the ESCs-regulatory network and also propose STAT3-based TRMs for CD4<sup>+</sup> T cells, macrophages and AtT-20 cells.

The cell type-specific TRM models of STAT3 predict an important level of epigenetic regulation, as these contain many histone-modifying enzymes, several histone deacetylases (HDAC2, HDAC3) histone acetyltransferases (EP300) and remodeling enzymes (SMARCE1, SMARCA4, SMARCC1) (Figure 7A–D). Additionally, in human ESCs, SIN3A, EP300 and HDAC2 are known to be associated with the ESC regulatory network (51), of which we identified SIN3A and EP300 (Figure 7A). SIN3A has been proposed as a master regulator of STAT activity, particularly in guiding STAT3 to the *Socs3* gene (58). Histone modifications likely play a major role in determining cell type specificity either by blocking other STAT3-binding sites from becoming available or by pre-marking sites that STAT3 can be recruited to. Vallania *et al.* (59) computationally predicted ~1.3 million STAT3-binding sites in the mouse genome. We nevertheless know from CHIP-seq experiments that STAT3 is binding to a fraction of these sites in distinct cell types. Although other members of the STAT family that share a very similar DNA-binding motif with STAT3 (44) might actually be occupying these other sites, in our opinion this possibility does not entirely explain why all those potential binding sites are unavailable to STAT3. Therefore, epigenetic regulation is probably a major factor controlling STAT3 transcriptional programs and has previously been shown for members of the STAT family of TFs (17). It remains to be determined whether chromatin is initially available for STAT3 recruitment, whether STAT3 binds to inaccessible sites but recruits chromatin modifiers to locally open the chromatin or a combination of both.

A number of TFs have been profiled by CHIP-seq in multiple cell types and a spectrum is emerging that goes from the exclusively cell type-invariant TFs to other TFs that are exclusively cell type-specific and a number of TFs that lie in between these two extremes. Among the remarkably cell type-invariant TFs is CTCF (60), whose primary role appears to be in maintaining the architecture



**Figure 8.** A putative model explaining how STAT3 can perform both cell type-independent and cell type-specific functions by assembling around distinct TRMs. STAT3 binding to the genome occurs in two distinct ways: (i) a cell type-independent mode that is primarily concerned with the regulation of STAT3's own activity and (ii) a number of cell type-specific modes that execute distinct transcriptional programmes in various cell types.

of the genome. However, CTCF can also act as a typical cell type-specific TF. REST is another role appears to be in maintaining the architecture of the genome. REST is another cell type-independent TF with important roles in ESCs and neural precursors (52). Among the cell type-specific TFs are SMAD3, which only shares three genomic binding sites across mouse ESCs, myotubes and pro-B cells and only 1% of sites appear in two or more cell types (35). Additionally, SMAD1 and TCF7L2 show only a small overlap in binding sites in two different human cell lines in the erythroid lineage (61) and TCF7L2 similarly to STAT3 shows only a small overlap in common between six cells lines (62). MYC lies somewhere in the center of the specificity spectrum (60), thus reflecting its dual role as a cell type-specific TF and a 'global' regulator of transcription potentially regulating ~15% of human genes (63). Finally, GATA3 binding was explored in multiple CD4<sup>+</sup> T-cell subtypes and shown to be predominantly cell type-specific even within closely related cells of the T-cell lineage (10). Therefore, we could say that both GATA3 and STAT3 are predominantly cell type-specific although they share a limited number of binding sites that are common to all cell types where they have been profiled. These observations place GATA3 and STAT3 between MYC and SMAD1/SMAD3/TCF7L2.

## CONCLUSIONS

Our analyses lead us to propose a dual model for STAT3 transcriptional regulation: (i) a limited, cell type-independent and evolutionarily conserved mode whereby STAT3 binds to DNA close to (and may co-operate with) MYC and E2F1 (and possibly other factors too) to regulate the expression of a set of genes in multiple cell types that regulate STAT3's own signaling activity and (ii) a broader set of cell type-specific binding modes controlled by distinct cell type-specific TRMs that form around STAT3 to enact cell type-specific functions (Figure 8).

This dual model of transcriptional regulation by means of assembling distinct TRMs around a master TF could be a general mechanism to regulate the transcriptional programs of other pleiotropic TFs.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–5, Supplementary Figures 1–7 and Supplementary References [64–78].

## ACKNOWLEDGEMENTS

We would like to thank Mr Shaq Liu for excellent systems support and Ms Mineko Tanimoto for secretarial support.

## FUNDING

Japan Society for the Promotion of Science (JSPS) through the WPI-IFReC Research Program and a Kakenhi grant; Kishimoto Foundation; ETHZ-JST Japanese-Swiss Cooperative Program [to D.M.S.]; Jeanne and Jean-Louis Lévesque Chair in Cancer Research and Canadian Cancer Society Research Institute [#700922 to M.L.T.]. Funding for open access charge: JSPS.

*Conflict of interest statement.* None declared.

## REFERENCES

- 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.
- Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Pena-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.

3. Arvey, A., Agius, P., Noble, W.S. and Leslie, C. (2012) Sequence and chromatin determinants of cell-type-specific transcription factor binding. *Genome Res.*, **22**, 1723–1734.
4. Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J. *et al.* (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, **133**, 1106–1117.
5. Wilson, N.K., Foster, S.D., Wang, X., Knezevic, K., Schutte, J., Kaimakis, P., Chilarska, P.M., Kinston, S., Ouwehand, W.H., Dzierzak, E. *et al.* (2010) Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell*, **7**, 532–544.
6. Lin, Y.C., Jhunjhunwala, S., Benner, C., Heinz, S., Welinder, E., Mansson, R., Sigvardsson, M., Hagman, J., Espinoza, C.A., Dutkowsky, J. *et al.* (2010) A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat. Immunol.*, **11**, 635–643.
7. Kidder, B.L. and Palmer, S. (2010) Examination of transcriptional networks reveals an important role for TCFAP2C, SMARCA4, and EOMES in trophoblast stem cell maintenance. *Genome Res.*, **20**, 458–472.
8. Graham, V., Khudyakov, J., Ellis, P. and Pevny, L. (2003) SOX2 functions to maintain neural progenitor identity. *Neuron*, **39**, 749–765.
9. Nakae, J., Kitamura, T., Kitamura, Y., Biggs, W.H. 3rd, Arden, K.C. and Accili, D. (2003) The forkhead transcription factor Foxo1 regulates adipocyte differentiation. *Dev. Cell*, **4**, 119–129.
10. Wei, G., Abraham, B.J., Yagi, R., Jothi, R., Cui, K., Sharma, S., Narlikar, L., Northrup, D.L., Tang, Q., Paul, W.E. *et al.* (2011) Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. *Immunity*, **35**, 299–311.
11. Neph, S., Stergachis, A.B., Reynolds, A., Sandstrom, R., Borenstein, E. and Stamatoyannopoulos, J.A. (2012) Circuitry and dynamics of human transcription factor regulatory networks. *Cell*, **150**, 1274–1286.
12. Kwon, H., Thierry-Mieg, D., Thierry-Mieg, J., Kim, H.P., Oh, J., Tunyaplin, C., Carotta, S., Donovan, C.E., Goldman, M.L., Taylor, P. *et al.* (2009) Analysis of interleukin-21-induced Prdm1 gene regulation reveals functional cooperation of STAT3 and IRF4 transcription factors. *Immunity*, **31**, 941–952.
13. Durant, L., Watford, W.T., Ramos, H.L., Laurence, A., Vahedi, G., Wei, L., Takahashi, H., Sun, H.W., Kanno, Y., Powrie, F. *et al.* (2010) Diverse targets of the transcription factor STAT3 contribute to T cell pathogenicity and homeostasis. *Immunity*, **32**, 605–615.
14. Hutchins, A.P., Poulain, S. and Miranda-Saavedra, D. (2012) Genome-wide analysis of STAT3 binding in vivo predicts effectors of the anti-inflammatory response in macrophages. *Blood*, **119**, e110–e119.
15. Langlais, D., Couture, C., Balsalobre, A. and Drouin, J. (2012) The Stat3/GR interaction code: predictive value of direct/indirect DNA recruitment for transcription outcome. *Mol. Cell*, **47**, 38–49.
16. Niwa, H., Burdon, T., Chambers, I. and Smith, A. (1998) Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev.*, **12**, 2048–2060.
17. O’Shea, J.J., Lahesmaa, R., Vahedi, G., Laurence, A. and Kanno, Y. (2011) Genomic views of STAT function in CD4<sup>+</sup> T helper cell differentiation. *Nat. Rev. Immunol.*, **11**, 239–250.
18. Stritesky, G.L., Muthukrishnan, R., Sehra, S., Goswami, R., Pham, D., Travers, J., Nguyen, E.T., Levy, D.E. and Kaplan, M.H. (2011) The transcription factor STAT3 is required for T helper 2 cell development. *Immunity*, **34**, 39–49.
19. Murray, P.J. (2006) Understanding and exploiting the endogenous interleukin-10/STAT3-mediated anti-inflammatory response. *Curr. Opin. Pharmacol.*, **6**, 379–386.
20. Takeda, K., Clausen, B.E., Kaisho, T., Tsujimura, T., Terada, N., Forster, I. and Akira, S. (1999) Enhanced Th1 activity and development of chronic enterocolitis in mice devoid of Stat3 in macrophages and neutrophils. *Immunity*, **10**, 39–49.
21. Langlais, D., Couture, C., Balsalobre, A. and Drouin, J. (2008) Regulatory network analyses reveal genome-wide potentiation of LIF signaling by glucocorticoids and define an innate cell defense response. *PLoS Genet.*, **4**, e1000224.
22. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
23. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
24. McLean, C.Y., Bristol, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M. and Bejerano, G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.*, **28**, 495–501.
25. Ho, L., Miller, E.L., Ronan, J.L., Ho, W.Q., Jothi, R. and Crabtree, G.R. (2011) esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nat. Cell Biol.*, **13**, 903–913.
26. Ramadoss, P., Chiappini, F., Bilban, M. and Hollenberg, A.N. (2010) Regulation of hepatic six transmembrane epithelial antigen of prostate 4 (STEAP4) expression by STAT3 and CCAAT/enhancer-binding protein alpha. *J. Biol. Chem.*, **285**, 16453–16466.
27. Dai, M., Wang, P., Boyd, A.D., Kostov, G., Athey, B., Jones, E.G., Bunney, W.E., Myers, R.M., Speed, T.P., Akil, H. *et al.* (2005) Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.*, **33**, e175.
28. Li, B. and Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.
29. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell*, **38**, 576–589.
30. Grant, C.E., Bailey, T.L. and Noble, W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017–1018.
31. Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. and Noble, W.S. (2007) Quantifying similarity between motifs. *Genome Biol.*, **8**, R24.
32. 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**, D77–D82.
33. Portales-Casamar, E., Thongjuea, S., Kwon, A.T., Arenillas, D., Zhao, X., Valen, E., Yusuf, D., Lenhard, B., Wasserman, W.W. and Sandelin, A. (2010) JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **38**, D105–D110.
34. Stark, C., Breitkreutz, B.J., Chatr-Aryamontri, A., Boucher, L., Oughtred, R., Livstone, M.S., Nixon, J., Van Auken, K., Wang, X., Shi, X. *et al.* (2011) The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res.*, **39**, D698–D704.
35. Mullen, A.C., Orlando, D.A., Newman, J.J., Loven, J., Kumar, R.M., Bilodeau, S., Reddy, J., Guenther, M.G., DeKoter, R.P. and Young, R.A. (2011) Master transcription factors determine cell-type-specific responses to TGF-beta signaling. *Cell*, **147**, 565–576.
36. Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S. *et al.* (2010) Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science*, **328**, 1036–1040.
37. Ouyang, Z., Zhou, Q. and Wong, W.H. (2009) ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. *Proc. Natl Acad. Sci. USA*, **106**, 21521–21526.
38. Ramagopalan, S.V., Heger, A., Berlanga, A.J., Maugeri, N.J., Lincoln, M.R., Burrell, A., Handunnetthi, L., Handel, A.E., Disanto, G., Orton, S.M. *et al.* (2010) A ChIP-seq defined genome-wide map of vitamin D receptor binding: associations with disease and evolution. *Genome Res.*, **20**, 1352–1360.
39. Gu, F., Dube, N., Kim, J.W., Cheng, A., Ibarra-Sanchez Mde, J., Tremblay, M.L. and Boisclair, Y.R. (2003) Protein tyrosine phosphatase 1B attenuates growth hormone-mediated JAK2-STAT signaling. *Mol. Cell Biol.*, **23**, 3753–3762.
40. Shuai, K. and Liu, B. (2003) Regulation of JAK-STAT signalling in the immune system. *Nat. Rev. Immunol.*, **3**, 900–911.



41. Gaba, A., Grivennikov, S.I., Do, M.V., Stumpo, D.J., Blackshear, P.J. and Karin, M. (2012) Cutting edge: IL-10-mediated tristetraprolin induction is part of a feedback loop that controls macrophage STAT3 activation and cytokine production. *J. Immunol.* **189**, 2089–2093.
42. El Kasmi, K.C., Smith, A.M., Williams, L., Neale, G., Panopoulos, A.D., Watowich, S.S., Hacker, H., Foxwell, B.M. and Murray, P.J. (2007) Cutting edge: a transcriptional repressor and corepressor induced by the STAT3-regulated anti-inflammatory signaling pathway. *J. Immunol.* **179**, 7215–7219.
43. Blake, J.A., Bult, C.J., Eppig, J.T., Kadin, J.A. and Richardson, J.E. (2009) The Mouse Genome Database genotypes::phenotypes. *Nucleic Acids Res.* **37**, D712–D719.
44. Wei, L., Vahedi, G., Sun, H.W., Watford, W.T., Takatori, H., Ramos, H.L., Takahashi, H., Liang, J., Gutierrez-Cruz, G., Zang, C. *et al.* (2010) Discrete roles of STAT4 and STAT6 transcription factors in tuning epigenetic modifications and transcription during T helper cell differentiation. *Immunity* **32**, 840–851.
45. Meijings, S.H., Pufall, M.A., So, A.Y., Bates, D.L., Chen, L. and Yamamoto, K.R. (2009) DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science* **324**, 407–410.
46. Becker, S., Groner, B. and Muller, C.W. (1998) Three-dimensional structure of the Stat3beta homodimer bound to DNA. *Nature* **394**, 145–151.
47. Husby, J., Todd, A.K., Haider, S.M., Zinzalla, G., Thurston, D.E. and Neidle, S. (2012) Molecular dynamics studies of the STAT3 homodimer:DNA Complex: relationships between STAT3 mutations and protein-DNA recognition. *J. Chem. Inform. Model.* **52**, 1179–1192.
48. Ogata, K., Sato, K. and Tahirov, T.H. (2003) Eukaryotic transcriptional regulatory complexes: cooperativity from near and afar. *Curr. Opin. Struct. Biol.* **13**, 40–48.
49. Miranda-Saavedra, D. and Gottgens, B. (2008) Transcriptional regulatory networks in haematopoiesis. *Curr. Opin. Genet. Dev.* **18**, 530–535.
50. Lian, I., Kim, J., Okazawa, H., Zhao, J., Zhao, B., Yu, J., Chinnaiyan, A., Israel, M.A., Goldstein, L.S., Abujarour, R. *et al.* (2010) The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation. *Genes Dev.* **24**, 1106–1118.
51. Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T.W., Greven, M.C., Pierce, B.G., Dong, X., Kundaje, A., Cheng, Y. *et al.* (2012) Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* **22**, 1798–1812.
52. Johnson, R., Teh, C.H., Kunarso, G., Wong, K.Y., Srinivasan, G., Cooper, M.L., Volta, M., Chan, S.S., Lipovich, L., Pollard, S.M. *et al.* (2008) REST regulates distinct transcriptional networks in embryonic and neural stem cells. *PLoS Biol.* **6**, e256.
53. Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M.M. and Choudhary, J. (2010) An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell* **6**, 382–395.
54. Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W. and Orkin, S.H. (2006) A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**, 364–368.
55. Russell, L. and Garrett-Sinha, L.A. (2010) Transcription factor Ets-1 in cytokine and chemokine gene regulation. *Cytokine* **51**, 217–226.
56. Zhang, F., Wang, D.Z., Boothby, M., Penix, L., Flavell, R.A. and Aune, T.M. (1998) Regulation of the activity of IFN-gamma promoter elements during Th cell differentiation. *J. Immunol.* **161**, 6105–6112.
57. Xie, H., Ye, M., Feng, R. and Graf, T. (2004) Stepwise reprogramming of B cells into macrophages. *Cell* **117**, 663–676.
58. Icardi, L., Mori, R., Gesellchen, V., Eyckerman, S., De Cauwer, L., Verhelst, J., Vercauteren, K., Saelens, X., Meuleman, P., Leroux-Roels, G. *et al.* (2012) The Sin3a repressor complex is a master regulator of STAT transcriptional activity. *Proc. Natl Acad. Sci. USA* **109**, 12058–12063.
59. Vallania, F., Schiavone, D., Dewilde, S., Pupo, E., Garbay, S., Calogero, R., Pontoglio, M., Provero, P. and Poli, V. (2009) Genome-wide discovery of functional transcription factor binding sites by comparative genomics: the case of Stat3. *Proc. Natl Acad. Sci. USA* **106**, 5117–5122.
60. Lee, B.K., Bhinge, A.A., Battenhouse, A., McDaniel, R.M., Liu, Z., Song, L., Ni, Y., Birney, E., Lieb, J.D., Furey, T.S. *et al.* (2012) Cell-type specific and combinatorial usage of diverse transcription factors revealed by genome-wide binding studies in multiple human cells. *Genome Res.* **22**, 9–24.
61. Trompouki, E., Bowman, T.V., Lawton, L.N., Fan, Z.P., Wu, D.C., DiBiase, A., Martin, C.S., Cech, J.N., Sessa, A.K., Leblanc, J.L. *et al.* (2011) Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell* **147**, 577–589.
62. Frieze, S., Wang, R., Yao, L., Tak, Y.G., Ye, Z., Gaddis, M., Witt, H., Farnham, P.J. and Jin, V.X. (2012) Cell type-specific binding patterns reveal that TCF7L2 can be tethered to the genome by association with GATA3. *Genome Biol.* **13**, R52.
63. Dang, C.V., O'Donnell, K.A., Zeller, K.I., Nguyen, T., Osthus, R.C. and Li, F. (2006) The c-Myc target gene network. *Semin. Cancer Biol.* **16**, 253–264.
64. Basso, K. and Dalla-Favera, R. (2012) Roles of BCL6 in normal and transformed germinal center B cells. *Immunol. Rev.* **247**, 172–183.
65. Boulon, S., Bertrand, E. and Pradet-Balade, B. (2012) HSP90 and the R2TP co-chaperone complex: building multi-protein machineries essential for cell growth and gene expression. *RNA Biol.* **9**, 148–154.
66. Diani-Moore, S., Ram, P., Li, X., Mondal, P., Youn, D.Y., Sauve, A.A. and Rifkin, A.B. (2010) Identification of the aryl hydrocarbon receptor target gene TipARP as a mediator of suppression of hepatic gluconeogenesis by 2,3,7,8-tetrachlorodibenzo-p-dioxin and of nicotinamide as a corrective agent for this effect. *J. Biol. Chem.* **285**, 38801–38810.
67. Jimenez, G., Shvartsman, S.Y. and Paroush, Z. (2012) The Capicua repressor—a general sensor of RTK signaling in development and disease. *J. Cell Sci.* **125**, 1383–1391.
68. Kannan, Y., Yu, J., Raices, R.M., Seshadri, S., Wei, M., Caligiuri, M.A. and Wewers, M.D. (2011) IkkappaBzeta augments IL-12- and IL-18-mediated IFN-gamma production in human NK cells. *Blood* **117**, 2855–2863.
69. Lessard, L., Stuiblé, M. and Tremblay, M.L. (2010) The two faces of PTPIB in cancer. *Biochim. Biophys. Acta* **1804**, 613–619.
70. Lores, P., Visvikis, O., Luna, R., Lemichez, E. and Gacon, G. (2010) The SWI/SNF protein BAF60b is ubiquitinated through a signalling process involving Rac GTPase and the RING finger protein Unkempt. *FEBS J.* **277**, 1453–1464.
71. Maldonado, V. and Melendez-Zajgla, J. (2011) Role of Bcl-3 in solid tumors. *Mol. Cancer* **10**, 152.
72. Passon, D.M., Lee, M., Rackham, O., Stanley, W.A., Sadowska, A., Filipovska, A., Fox, A.H. and Bond, C.S. (2012) Structure of the heterodimer of human NONO and paraspeckle protein component 1 and analysis of its role in subnuclear body formation. *Proc. Natl Acad. Sci. USA* **109**, 4846–4850.
73. Perez-Leal, O. and Merali, S. (2012) Regulation of polyamine metabolism by translational control. *Amino Acids* **42**, 611–617.
74. Takano, A., Zochi, R., Hibi, M., Terashima, T. and Katsuyama, Y. (2010) Expression of strawberry notch family genes during zebrafish embryogenesis. *Dev. Dyn.* **239**, 1789–1796.
75. Walker, W., Zhou, Z.Q., Ota, S., Wynshaw-Boris, A. and Hurlin, P.J. (2005) Mnt-Max to Myc-Max complex switching regulates cell cycle entry. *J. Cell Biol.* **169**, 405–413.
76. Watanabe, N., Madaule, P., Reid, T., Ishizaki, T., Watanabe, G., Kakizuka, A., Saito, Y., Nakao, K., Jockusch, B.M. and Narumiya, S. (1997) p140mDia, a mammalian homolog of *Drosophila* diaphanous, is a target protein for Rho small GTPase and is a ligand for profilin. *EMBO J.* **16**, 3044–3056.
77. Wyman, S.K., Knouf, E.C., Parkin, R.K., Fritz, B.R., Lin, D.W., Dennis, L.M., Krouse, M.A., Webster, P.J. and Tewari, M. (2011) Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. *Genome Res.* **21**, 1450–1461.
78. Zhou, L., Zhang, Z., Zheng, Y., Zhu, Y., Wei, Z., Xu, H., Tang, Q., Kong, X. and Hu, L. (2011) SKAP2, a novel target of HSF4b, associates with NCK2/F-actin at membrane ruffles and regulates actin reorganization in lens cell. *J. Cell. Mol. Med.* **15**, 783–795.