Supplemental Figure Legends

Figure S1.

A. In vitro model system for Th17 cell differentiation from naïve $CD4^+$ T cell precursors and schematic for experimental protocol. Th0 cultures provide control cells that receive TCR activation in the absence of exogenous polarizing cytokines (IL-6 + TGFβ).

B. Western blot analysis of Th17 polarization time series starting from FACS-purified naïve CD4 T cells (time=0).

C. Recovery of cognate consensus motifs from TF-ChIP-Seq.

D. High degree of co-occupancy among Th17 lineage TFs. ChIP-seq binding tracks are displayed for core TFs, CTCF, and p300 at selected Th17 loci in both non-polarized Th0 and Th17 conditions. Visualized using the Integrative Genomics Viewer (IGV; Broad Institute).

E. High-order pCRMs are not correlated with proximity to TSS. Bar chart of proportion of proximal versus distal pCRMs with respect to increasing order of occupancy.

Figure S2. Cooperative occupancy by BATF and IRF4.

A. Genome-wide interdependence of IRF4 and BATF co-occupancy in Th0 cells. Box plots displaying the fold change in ChIP-seq reads for IRF4 in *Batf* wild-type (wt) versus knockout (ko) and for BATF in *Irf4* (wt/ko) for 48h Th0 cultured cells. Differences in ChIP-seq reads are assayed within relevant pCRM regions. Three sub-types of pCRMs were interrogated: BATF or IRF4 alone; BATF and IRF4 alone; and BATF, IRF4, plus additional TFs (+) as indicated by color-coding. Displayed is the data distribution: median (line), 25^{th} to 75^{th} percentile (box) +/- 1.5 Interquartile range (whiskers). To compute fold change in ChIP values, reads localized to a given pCRM were normalized by library size (i.e. reads per million; RPM) prior to calculations.

B. Interdependent binding of IRF4 and BATF at selected loci in Th17 cells.

Figure S3. Genome-wide requirement for Th17 TFs for accessibility and TF occupancy.

A. Spatial correlations between Th17 TFs within pCRMs occupied by all five TFs. Bar chart plots the occurrence with which the summit of a given TFs occupies relative position 1 through 5 when the summits of all five are ordered from 5' to 3'.

B. IRF4 and BATF regulate chromatin accessibility at TCR-induced cis regions that are co-occupied by Th17 TFs. FAIRE signal at Th17 5TF+p300 pCRMs was compared between WT and *Irf4*$^{-/-}$ and *Batf*$^{-/-}$ Th0 and Th17 polarized T cells. pCRMs were divided according to their accessibility status in naïve CD4^{+} T cells: constitutive pCRMs are accessible in naive cells (2,930 regions, left panel), while induced pCRMs are not (1,575 regions, right). Biological replicate samples were averaged, and normalized FAIRE reads were aligned around the median summit position of overlapping TF binding peaks, +/- 2,000bp.

C. Limited requirement for RORγt for p300 occupancy as compared to IRF4, BATF, and STAT3. Differential occupancy of p300 in TF wild-type versus deficient 48h Th17 polarization cultures is displayed as scatter plots of fold change versus significance. Various pCRM subtypes are compared as indicated in the figure. Percentage of pCRMs with differential ChIP are indicated in plot.

D. Limited requirement for RORγt for IRF4 and STAT3 occupancy and for presence of H3K4me2 and H3K4me3 modifications. Scatter plots as in (C).

Figure S4. Relationship between core TF regulation and expression of target genes in the Th17 network.

A. Heat map summarizing genome-wide regulatory inputs for the core TF network displayed in Figure 3A. Orange represents repression and blue represents activation. Rows are target genes.

B. Heat map of activation and repression inputs for highly regulated genes (4 or 5 inputs) by core TFs. Orange represents repression and blue represents activation. View is limited to genes with KC scores >1.5 for a given TF-target regulatory interaction. Also displayed is the fold change (FC) in expression observed in Th17 relative to Th0 cells.

C. Regulatory interactions shared by STAT3, IRF4, BATF, and RORγt (as in Figure 3C). Node color depicts the extent to which genes are differentially expressed in RORγt deficient Th17 cells when compared to wild type Th17 cells. Different modes of RORγt regulation are as indicated.

D. Effect of individual core TFs on target gene transcription. Positive and negative regulation by STAT3, IRF4, and BATF is well correlated with expression changes associated with Th17 differentiation. In contrast, the regulatory effect of RORγt is consistent with a modulatory role. Box plots display the fold change in expression of both TF activation (green) and repression (red) targets for Th17 relative to Th0 culture conditions. Displayed is the data distribution: median (line), 25^{th} to 75^{th} percentile (box) +/- 1.5 Interquartile range (whiskers). Genes co-regulated by either 4 or 5 of the TFs are considered in this analysis.

Figure S5. aucROC performance and comparison of two meta-analysis strategies: rank-based and Fisher's method

A. Recovery of validated biologically relevant Th17 targets based on integration of regulatory models from multiple functionally relevant TFs and of multiple data types. The graph

summarizes area under curve (auc) of receiver operator curve (ROC) plot results indicating the degree to which 74 literature-based Th17-relevant genes are enriched as top network predictions under different data combinations. Individual TFs (scatter plot) versus combined TFs (bar plots) are compared for each combination of data types. As a reference, differential expression in Th17 vs. Th0 (dotted line) and random performance (dashed line; based on 200 simulations) are provided.

B. The KCRI network selectively recovers genes linked to SNPs that are associated with Th17-implicated inflammatory disease in GWAS studies. Gray bars and scatter plots in each column correspond to the recovery of SNP-associated disease-relevant targets as top predictions within the ranked list of the KCRI network scores, using the aucROC analysis. Gene lists of disease-associated SNPs were compiled from the National Human Genome Research GWAS Catalog.

C. Distributions of genome-wide TF -> target gene scores (-log10 p-values or absolute z-scores) for five TFs based on method [data-source] used.

D. Area under curve of Precision Recall for the Rank based approach used in this work and Fisher's method for combining p-values (pseudo z-scores were transformed to p-values before combining assuming a normal distribution). Fisher's method achieves better performance in combining TFs -> target p-values over single datasets (compare blue bars), but does not perform well when incorporating p-values from distinct data sources (compare yellow and red bars) due to inherent differences in p-value distributions derived from different data types and methods.

Figure S6. Gain and loss of function screens identify regulators of Th17 specification.

A. Overexpression screen of network TF candidates as putative Th17 subset regulators. Bar charts show the percent of IL-17A-producing or IFNγ-producing cells relative to the control

empty vector after transduction of retroviruses encoding candidate factors and Th17 polarization for 48h, or Th1 polarization for 5 days, respectively. Results are mean ± s.e.m. for four biological replicates, each conducted in duplicate. Significance at *P<0.05 and **P<0.01 by T-test.

B. Representative flow cytometric analysis for IL-17A and Foxp3 expression in Th17-polarized cultures transduced with the indicated cDNAs in a retroviral vector also encoding the Thy1.1 reporter. Cells were gated for Thy1.1 expression to analyze proportions that were IL-$17A^+$.

C. Knock-down efficiency of target mRNAs in the siRNA screen. Data represent reads per kilobase million (RPKM) expression values for the target TF in siRNA knock-down Th17 cultures relative to a non-targeting control. Analysis is at 24h post Th17 differentiation.

D. Western-blot analysis of RORγt protein levels for *Rorc* knock-down at indicated times post initiation of Th17 polarization.

E. JMJD3 regulates the expression of many RORγt and STAT3 targets in Th17 cells. Network representation is as in Figure 3. Due to space constraints, the display is limited to genes that have differential expression in Th17 relative to Th0 cells (z-score >2.5, <-2.5 based on statistical analysis of microarray for 8 independent experiments).

Figure S7. Fosl2 restricts the plasticity of Th17 subset cells.

A. Dysregulated cytokine production in the absence of Fosl2. Fosl2 wild-type and deficient naïve $CD4^+$ T cells were polarized under Th1, Th2, and Th17 conditions for 6 days. Flow cytometric analysis was then performed for IL-17A, IL-4, and IFNγ.

B. Fosl2 restricts IFNγ production among IL-17A-producing cells. Fosl2 wild-type and deficient naïve $CD4^+$ T cells were polarized under Th17 conditions (20ng/mL IL-6 and 0.3ng/mL TGFβ + blocking antibodies for IFNγ and IL-4) for three days. Thereafter, the media was replaced with cytokines for either (a) Th17-; (b) Th1- (10ng/mL IL-12); or (c) Th2- (2ng/mL IL-4) promoting conditions for an additional three days prior to analysis.

C. De novo motif analysis of high confidence binding regions for BATF and Fosl2 ChIP-seq experiments of Th17 cells showing that the AP-1 consensus motif is recovered in both instances.

D. Regulation of core Th17 TFs by Fosl2. Edges represent integration of data from ChIP-seq and KO RNA-seq differential expression; line weight is relative to network score; FDR <5%. Nodes are colored to indicate the differential expression in Th17 relative to Th0 (blue=upregulated, orange=downregulated in Th17 cells).

Table S1. Literature curated validation list for genes with critical influence for Th17 development or function.

The list of known Th17-relevant genes used for computational validations is provided, including the Pubmed ID (PMID) for the supporting literature.

Table S2. Enrichment scores for identification of novel TFs and regulators.

Candidate genes used in various biological screens (gain- and loss-of function) are highlighted in green, purple, and blue depending on the criteria used for their selection. Positive controls for TF recovery (STAT3, BATF, IRF4, Maf, and RORC) are highlighted in yellow.

Table S3. List of experimental libraries for ChIP-seq and RNA-seq

Table S4. TF Summed scores for KC and KCRI networks.

Supplemental Experimental Procedures

Mice

Mice were bred and maintained in the animal facility of the Skirball Institute (Langone Medical Center, NYU) in specific pathogen-free conditions. C57Bl/6, and *Hif1a*$^{fl/fl}$ (Ryan et al., 2000) mice were obtained from Jackson laboratories. *Rorc(t)* knock-out mice harboring a GFP reporter cDNA at the translation initiation site have been described (Eberl et al., 2004). Mutant strains were kindly provided by the following researchers: *Stat3*$^{fl/fl}$ (Lee et al., 2002), D. Levy (NYU); *Irf4*$^{fl/fl}$,(Klein et al., 2006) R. Dalla-Favera (Columbia University); *Maf* $^{fl/fl}$ (Wende et al., 2012), C. Birchmeier (MDC, Germany); *Batf*$^{fl/fl}$ (Schraml et al., 2009), K.M. Murphy (Washington University); and *Fosl2*$^{fl/fl}$ (Karreth et al., 2004), E. Wagner (CNIO, Spain). *Irf4*$^{fl/fl}$ mice were mated with EIIa-Cre transgenic mice to obtain fully IRF4 null animals. All animal procedures in accordance with protocols approved by the Institutional Animal Care and Usage Committee of New York University.

Cell culture

Naïve CD4$^+$ T cells were purified by cell sorting from spleen and lymph nodes as previously described (Ivanov et al., 2006) using the Aria II (BD). Briefly, red blood cells were cleared from organ cell suspensions using ACK lysis buffer (Lonza). The resulting leukocytes were depleted of B220$^+$ and CD8$^+$ cells by magnetic-activated cell sorting (MACS, Miltenyi) according to the product protocol. The negative fraction was cell surface stained using antibodies specific for CD4, CD25, CD44, and CD62L, and CD4$^+$CD25$^-$CD62L$^+$CD44$^{lo/-}$ naïve CD4$^+$ T cells were isolated by cell sorting using the Aria II to greater than 98% purity based on post-sort analysis. Naïve CD4+ T cells were cultured in 48-, 24-, or 12- well plates coated with an anti-hamster IgG

secondary antibody (MP Biomedicals), in complete IMDM media (containing 10% FCS) containing soluble anti-CD3ε (0.25μg/mL) and anti-CD28 (1μg/mL) for TCR stimulation. Cultures were supplemented as follows, or as indicated in figures: with anti-IL4 (2μg/mL) and anti-IFNγ (2μg/mL) for Th0 conditions and additionally with IL-6 (20ng/mL; eBioscience) and TGFβ (0.3ng/mL; PeproTech) for Th17 conditions; or 5ng/mL TGFβ for iTreg conditions. For Th1 differentiation, IL-4 and anti-IFNγ (2μg/mL) were added; for Th2 differentiation, IL-12 and anti-IL-4 (2μg/mL) were added; cytokine concentrations as labeled in figures. Unless otherwise indicated antibodies were purchased from eBioscience.

Antibodies, surface and intracellular staining

For analysis of cytokine production, cells were incubated for 4-5 h with phorbol 12-myristate 13-acetate (50ng/mL; Sigma), ionomycin (500ng/mL; Sigma), and GolgiStop (BD) at 37°C in a tissue culture incubator. Surface cell staining was carried out with fluorescence-labeled antibodies in PBS containing 0.5% BSA and 2mM EDTA at 4°C for 20 min. For live cell analysis or sorting, cells were washed once in staining buffer and resuspended in 200ng/mL of DAPI in staining buffer to exclude dead cells. For intracellular staining, cells were first stained with the fixable Aqua dead cell exclusion kit (Invitrogen), washed twice with PBS, and resuspended in Fixation-Permeabilization solution (Cytofix/Cytoperm kit; BD Biosciences or eBioscience) and intracellular staining was carried out according to the manufacturer's protocol. All fluorescence-labeled antibodies were purchased from eBioscience. An LSR II (BD Biosciences) was used for flow cytometric acquisition, followed by analysis with FlowJo software (Tree Star). All analysis plots are gated to exclude dead cells.

Chromatin Immunoprecipitation (ChIP).

TF ChIP-Seq was performed in biological duplicate as described (Johnson et al., 2007) with the following modifications. For each ChIP, 20-80 million cells were cross-linked with paraformaldehyde; chromatin was isolated and fragmented with a Vibra-Cell VCX130PB (Sonics & Materials). Following immunoprecipitation, the protein-DNA crosslinks were reversed and DNA was purified. DNA from control samples was prepared similarly but without immunoprecipitation. Histone ChIP of native chromatin was performed as previously described (Kirigin et al., 2012). Sequencing libraries were made from the resulting DNA fragments for both ChIP and controls as described (Reddy et al., 2012). The ChIP-seq libraries were sequenced with single-end 36 bp reads on an Illumina GAIIx or single-end 50 bp reads on an Illumina HiSeq 2000.

Commercial antibodies used for ChIP for each protein were as follows: IRF4 (IRF-4 M-17; Santa Cruz Biotech, sc-6059), BATF (BATF; Santa Cruz Biotech, sc-100974), STAT3 (Stat3 C-20; Santa Cruz Biotech, sc-482), p300 (p300 C-20; Santa Cruz Biotech, sc-585), Maf (Bethyl Laboratories, A300-613A), FOSL2 (Fra-2 Q-20; Santa Cruz Biotech, sc-604), HIF1α (Novus, NB100-105), ETV6 (TEL; Santa Cruz Biotech, sc-8546), JMJD6 (abcam, ab64575), NRF2 (H-300, Santa Cruz Biotech, sc-13032), H3K4me2 (Millipore, 07-030), and H3K4me3 (Millipore, 05-745R). The anti-RORγ rabbit polyclonal antibody was raised against amino acids 79-301 (Covance) and affinity purified antibody was isolated from serum using the same immunogen. The JMJD3 affinity purified antibody was kindly provided by G. Natoli (IFOM-IEO, Italy). The specificity of each transcription factor antibody was validated by immunoblot or conventional ChIP assay comparing wild-type to factor-deficient (or knock-down) Th17 subset polarized cells.

In addition, ChIP-Seq was performed in knock-out cells for the core TFs to provide an additional negative control for each ChIP-seq.

FAIRE-seq

FAIRE was performed as previously described (Simon et al., 2012). FAIRE reads were mapped using Bowtie (-k 1 –best) (Langmead et al., 2009) on the Galaxy platform (Goecks et al., 2010). For visualization of FAIRE signal around pCRMs, normalized alignment files were prepared using HOMER (Heinz et al., 2010), and heatmaps were made using SEQMINER (Ye et al., 2011).

Co-immunoprecipitations

Naïve CD4 positive T cells were sorted and cultured under Th17 polarizing conditions for 48h prior to assay. Whole cells lysates were prepared with high salt buffer (10mM Tris, 420mM NaCl, 0.5% NP40, 1mM EDTA), sonicated, spun to remove insoluble particles, and diluted to a final concentration of 150mM NaCl for co-immunoprecipitation. Endogenous IRF4 was immunoprecipitated using anti-IRF4 antibody in the presence or absence of 50ug/ml ethidium bromide. Co-IP pulled-downs were resolved by SDS electrophoresis and anti-BATF (Santa Cruz Biotech) and STAT3 (Cell Signaling) antibodies were used for western blot detection.

Luciferase Assay

pCRM activity was assessed using luciferase reporter assays by cloning the ChIP-defined genomic region (average of approx. 750bp) upstream of a minimal promoter driving a luciferase gene (pGL4.23[luc2/minP]; Promega). Importantly, pCRMs were selected in a non-biased

manner based on ranked average binding ChIP p-values for occupying TFs; genomic coordinates are as follows:

TF order	pCRM genomic coordinates
1	chr3:78762692-78763270
1	chr1:43196384-43196904
1	chr18:75739077-75739648
1	chr5:138093690-138094176
1	chr10:94880166-94880778
2	chr17:55841782-55842285
2	chr11:109482382-109482946
2	chr11:44503354-44503971
2	chr3:103188727-103189193
3	chr1:155559667-155560272
3	chr1:184030446-184030992
3	chr4:59821894-59822386
3	chr12:33858479-33858979
4	chr15:9457605-9458144
4	chr11:44449580-44450262
4	chr12:101985209-101985820
4	chr14:52642471-52643107
5	chr1:20730409-20731129
5	chr1:20730409-20731129
5	chr5:53980000-53980448
5	chr1:146086088-146086770
5	chr5:53881767-53882558
5	chr9:107215041-107215628
5	chr12:74988875-74989601
5	chr13:16733224-16734291

pCRM activity was assessed using luciferase reporter assays by cloning the ChIP-defined genomic region upstream of a minimal promoter driving a luciferase gene (pGL4.23[luc2/minP]; Promega). Naïve $CD4^+$ T cells were sorted and cultured under Th2, or Th17 polarizing conditions for 48h prior to being harvested for electroporation. Briefly, 5 million cells were pre-incubated with 10μg of pCRM-pGL4minP or empty pGL4minP construct and 2μg of renilla luciferase plasmid in 500μl of RPMI on ice. Cells were electroporated using a BioRad Electroporator at 300V and 750μF. After 10 min of recovery on ice, cells were placed into pre-warmed polarizing culture medium under TCR and cytokine stimulation conditions (Th2 or

Th17). 24h post electroporation, cells were collected and luciferase assays were performed using the Dual Luciferase Reagents (Promega). Firefly luciferase values were normalized to renilla luciferase values for each sample and expressed as fold change over empty pGL4-minP. pGL4-minP harboring regions from the *Il17a* locus: Il17a-5 (a known enhancer) (Wang et al., 2012) and Il17a-19 (a non-TF occupied conserved region 19kb upstream of TSS) served as positive and negative controls, respectively.

Preparation of RNA-Seq libraries

mRNA was prepared from total RNA by poly-A selection and cDNA synthesis was carried out as described (Mortazavi et al., 2008). The resulting dsDNA was prepared for sequencing by ligation of Illumina sequencing adapters, selection of 225 bp fragments from a 2% agarose SizeSelect E-Gel (Invitrogen), and amplification with 15 cycles of PCR using Illumina paired-end primers. Alternatively, some libraries were made using the Nextera tagmentation protocol described (Gertz et al., 2012). The RNA-seq libraries were sequenced with single-end 36 bp reads on an Illumina GAIIx or single-end 50 bp reads on an Illumina HiSeq 2000. Biological duplicates were carried out for each experiment. Sequence reads were mapped to the mus musculus genome (version mm9) with Bowtie (version 0.12.7) and with the following settings: -k 1 --best. The --phred33-quals or --phred64-quals parameter was set as needed depending on the format of the input fastq file. Anywhere between 8.5M and 78.7M reads aligned per library. Read counts for annotated genomic features were computed using the htseq-count script from the HTSeq (version 0.5.3p3) software suite with parameters: --stranded=no --mode=union.

siRNA knock-downs

For knockdown of network genes in T cell polarization cultures, naïve C57/Bl6 CD4 T cells were sort purified and cultured for 16-18h in RPMI/Th0 conditions. 2 million stimulated cells were transfected with 300pmol of control siRNAs for Ccr6, Rorc, and a non-Targeting pool (pool #2; SMARTpool siRNA; Dharmacon), in addition to SMARTpool siRNAs for network target genes (Dharmacon). Transfections were performed using the Amaxa Mouse T cell Nucleofector Kit with the X-001 program (Amaxa) according the manufacture's protocol. After a 4h recovery at 37°C, cells were stimulated in Th17 conditions in RPMI media. RNA was prepared from cells collected at 24h post polarization to assess knockdown efficiency and for RNA-Seq. Flow cytometric analysis for IL-17A and Foxp3 24h post polarization; viability was assessed by Aqua exclusion (Invitrogen) and cell counts by Accucount particles.

Collection of GWAS and SNP data for network validation:

Disease-associated SNPs compiled from the National Human Genome Research GWAS Catalog (available at www.genome.gov/gwastudies; accessed Feb 29, 2012). For each condition, gene lists were produced by selecting catalog-annotated human genes within 100 kb of associated SNPs. In cases where a SNP falls between two loci, the closest gene was chosen for association. Gene lists were used with no regard to human-mouse synteny.

Retroviral gene transfer

Retroviral constructs were generated by subcloning of the cDNA of interest into MSCV-Thy1.1 5' of the internal ribosomal entry site, permitting the bicistronic expression of candidate TFs and cell surface Thy1.1. Retrovirus was generated by transfection of retroviral constructs into the

PlatE producer cell line (Morita et al., 2000); viral supernatants were used at 48h post transfection. FACS sorted naïve CD4 T cells were stimulated under Th0 conditions for 20-24h prior to retroviral transduction. For gene transfer, cells were spin transduced for 2 hr at 2500rpm with viral supernatants in the presence of 6.7 ug/mL of polybrene (hexadimethrine bromide, Sigma), and media was replaced with T cell polarization media for differentiation to Th17 and control Th1. Cells were harvested after 48h (Th17) and 5 days (Th1) for flow cyotmetric analysis of cytokine production.

EAE Induction

For induction of EAE, mice were immunized subcutaneously on day 0 with 200 μg/mouse MOG 35-55 peptide (UCLA peptide synthesis facility), emulsified in CFA (CFA supplemented with 2 mg/ml Mycobacterium tuberculosis), and injected intravenously on days 0 and 2 with 200 ng/mouse of pertussis toxin (Sigma Aldrich). The following scoring system used was 0—no disease, 1—limp tail, 2—weak/partially paralyzed hind legs, 3—completely paralyzed hind legs, 4—complete hind and partial front leg paralysis, 5—complete paralysis/death. Mice with disease levels 4 and 5 were considered moribund and were euthanized.

Isolation of Mononuclear Cells from Spinal Cords

Before spinal cord (SC) dissection, mice were perfused with 30 ml of cold Ca^{2+}/Mg^{2+}-free PBS. The spinal columns were dissected, cut open, and intact SCs separated carefully from the vertebrae. The SCs were cut into several small pieces and placed in 2 ml digestion solution containing 10 mg/ml Collagenase D (Roche) in PBS with 5% FCS. Digestion was performed for 30 min at 37°C. Digestion was terminated by the addition of EDTA to a final concentration of

12.5 mM and an additional 5 minute incubation. The resulting digested tissue was passed through a 70 um cell screen. The cells were washed once in PBS, placed in 10 ml of 38% Percoll solution, and pelleted for 30 min at 2000 rpm with no brake. Cells pellets were washed once in PBS, re-suspended in FACS buffer or T cell medium and stimulated for assessment of cytokine production and Foxp3 expression as described above.

Computational methods:

Primary data processing of ChIP-seq and RNA-seq experiments

Sequence reads were mapped to the mus musculus genome (version mm9) with Bowtie (version 0.12.7) (Langmead et al., 2009) and with the following settings: -k 1 --best. The --phred33-quals or --phred64-quals parameter was set as needed depending on the format of the input fastq file. Anywhere between 8.5M and 78.7M reads aligned per library. ChIP-seq datasets were further processed to call peaks with the MACS software (version 1.4.0 20110619) using the settings: -p 1e-10 -m 15,30 -s 36 -g mm --bw=200 (Zhang et al., 2008). All were processed against an appropriate control. RNA-seq datasets were also processed through Tophat (version 1.2.0) with settings: -a 10 -g 20 --no-novel-juncs -G refseqGeneAnnot.gtf (Trapnell et al., 2009). Tophat results were then pipelined to Cufflinks (version 0.9.3) with the settings: -M 20101217_rRNA_tRNA_mask.gtf -G refseqGeneAnnot.gtf (Trapnell et al., 2010). Absolute read counts for annotated genomic features were computed using the htseq-count script from the HTSeq (version 0.5.3p3) software suite with parameters: --stranded=no --mode=union.

Network inference via integration of ChIP-seq, RNA-seq and microarray data.

Overview of integrative network inference: Here we describe how we scored $TF \rightarrow target\ gene$ regulatory interactions based on the four main complementary data types that include the majority of the data collected in this study. We integrate: 1) ChIP-seq for TFs, 2) RNA-seq following knock-out of TFs, 3) RNA-seq of Th17 differentiation (time series) and steady state data for other CD4+ subsets, and, 4) Immgen data, a publically available microarray compendium spanning the hematopoietic differentiation tree (Heng and Painter, 2008). We combine these data types into a multi-support directed regulatory network that accurately

predicts the regulatory events responsible for specifying the Th17 lineage. Recent work has clearly demonstrated the utility of combining data types as diverse as TF binding, motif conservation, and chromatin modifications for prediction of regulatory interactions (Ivanov et al., 2006; Marbach et al., 2012a; Ouyang et al., 2009; Park et al., 2005; Zhou et al., 2009; Zhou et al., 2007; Zhou et al., 2010). Note that we used an older release of the Immgen dataset (dated to March 2011) as the Immgen rules require users not to publish results based on data within six months of its release.

Structure: In the next four sections we describe how we calculated $TF \rightarrow target\ gene$ confidence scores for each individual data source using a method of our own construction. For each data source we store the confidence scores in an $M \times N$ matrix, $S(D_i)$, where M is the number of genes, N is the number of TFs, and $D_i \in [KO, ChIP, RNAseq, Immgen]$ is the data-type in question. Then we map confidence scores (p-values for ChIP and knock-out, or pseudo z-scores for RNA-seq and Immgen) for each matrix, $S(D_i)$, to rank-based quantile scores that we store in an $M \times N$ matrix, $Q(D_i)$. The previous step is required for data integration. Thereafter we show how we integrated scores over multiple data sources by summing (element by element) $\sum_{D_i \in D^*} Q(D_i)$, $D^* \subseteq [KO, ChIP, RNAseq, Immgen]$, for any data combination we tested in this work.

ChIP-seq, defining $\boldsymbol{TF \rightarrow target\ gene}$ association scores:

Scoring of a given TF's association with a target gene in a given ChIP-seq experiment is usually defined (at least in part) based on TF-binding site proximity to a target gene's transcription start site (TSS), and is often given a binary value indicating if a regulatory interaction exists or not (Boyer et al., 2005; Chen et al., 2008; Marson et al., 2008). However, these commonly used

formulations discount the majority of TF binding data that is not proximal to TSS and may be important for regulation, and do not provide a continuous score which is needed for ranking possible target genes by confidence. Recent works have started to address these limitations by considering wider regions around the TSS of genes, and assign a continuous confidence score for putative TF-target gene interactions (Ouyang et al., 2009; Rutz et al., 2011). Here, we use a TF-gene association score that includes regulatory regions that are far from the promoter, as we know that active regulatory elements are often in introns or distal regulatory regions. We thus examined the full gene (TSS to end of last exon) plus 10kb on each side and used a scoring scheme that integrated all peaks found for a given TF in that region, normalizing for the number of peaks of similar strength expected by chance. It is important to note that we primarily use this p-value score as part of a larger integrative framework (integrated with KO and time series expression data) to form our final network; we can thus initially trade sensitivity for accuracy and recover accuracy via our subsequent integration with other data types.

Let g be a possible target gene of TF x and L_g be the genomic region surrounding gene g (as defined above, gene +/- 10kb). Let $|L_g|$ be the genomic span of L_g in bps. Also, let S_x be the set of peaks identified by MACS for TF x across the genome, and S_x^{g} be a subset of S_x that denotes those peaks that locate in L_g. Assuming a naïve null hypothesis that peaks are distributed randomly across the genome, the probability of observing $|S_x^g|$ peaks in L_g follows the Poisson distribution with an expected number of occurrences $\lambda = |L_g| \times \rho(x)$, where $\rho(x)$ is the expected number of peaks per bp for TF x. A simple way to estimate $\rho(x)$ is to divide the total number of peaks by the genome size, $\rho(x) = \frac{|S_x|}{G}$. Then the probability of observing $|S_x^g|$ peaks in L_g is, $Poisson(n \geq |S_x^g|, \lambda = |L_g| \times \rho(x))$. However, this simple formulation does not differentiate

between a gene region that has n strong peaks and the same gene region with □ weak peaks. We thus calculated $\rho(x)$ in a manner that would incorporate the binding significance of peaks found in L_g, as follows:

$$\rho(x) = \frac{\left|a \in S_x : -\log_{10} pval(a) \geq mean(-\log_{10} pval(S_x^g))\right|}{G}$$

where G is the mappable size of the genome, and the numerator specifies the total number of peaks (genome-wide) with significance equal or greater to the average significance of peaks found in the region of L_g. We then defined the TF-gene association score s to be -log_{10} of the p-value of observing S_x^g in L, given $\rho(x)$, which can be calculated as:

$$s(x \rightarrow g \mid ChIPseq\ x) = -\log_{10}(\ Poisson(n \geq \left|S_x^g\right|, \lambda = \left|L_g\right| \times \rho(x))).$$

Although we chose the above ChIP scoring scheme (which considers the entire gene body +/-10kb) to achieve greater sensitivity, our subsequent computational and experimental validation of our network models revealed that this gene-wide ChIP-seq scoring scheme was overall more successful than a more traditional TSS proximal scoring scheme (considering a region of +/-5kb around TSS) at identifying known Th17 target genes as top scoring hits, as measured by both the area-under-curve of precision recall (accuracy) and Receiver operator curves (sensitivity), respectively (data not shown).

<u>RNA-seq (wild type vs. knock out), defining TF → target gene association scores from TF KO RNA-seq data:</u>

To determine TF → target gene associations scores based on knock-out data, we performed RNA-seq knock-out experiments for key TFs (same TFs as in the ChIP-seq experiments) under

Th17 stimulating conditions. Let x denote a knocked-out TF and g denote a putative target gene. For each gene g under x wild-type vs. knock-out conditions we computed a fold change, and a corresponding p-value using DEseq (Anders and Huber, 2010), a program to calculate the significance of differential expression from RNA-seq. We then used -$\log_{10}$ of the p-value reported by DEseq as a confidence score for the association between TF x and gene g:

$$s(x \rightarrow g \mid Knockout\ x) =$$

$$-\log_{10}\left(DEseq_{pvalue}\left(\frac{reads\left(g(x^{wt})\right)}{reads\left(g(x^{ko})\right)}\right) \times sign\left(\log_2\left(\frac{reads\left(g(x^{wt})\right)}{reads\left(g(x^{ko})\right)}\right)\right)\right),$$

where $reads\left(g(x^{wt})\right)$ and $reads\left(g(x^{ko})\right)$ denote the number of reads sequenced corresponding to the mRNA of gene g in wild-type and knockout cells, respectively (adjusted for differences in library size). Note that we multiplied each regulatory interaction confidence score by the sign of the fold change to indicate activating from repressing interactions (positive and negative scores, respectively). Also note that individual pairs of knock-out vs. wild-type experiments of at least two biological repeats were run separately through DEseq to determine p-values from each pair, which were then combined using Fisher's method for combining p-values (Fisher, 1925). We used this meta-analysis procedure since we found that inherent systematic biases between biological replicates (such that the knock-out of one experiment is more correlated with its corresponding wild-type control than with the knock-out of the other replicate) can significantly degrade DEseq performance (data not shown).

Using the Inferelator to derive networks from our RNA-seq data compendium (time-series, knock-outs, and other CD4+ lineages):

We collected various RNA-seq experiments including time series for Th17 (or Th0 as a control) specification *in vitro*, knockouts as described above, and additional RNA-seq for alternative CD4+ lineages. This resulted in a Th17 and T-cell focused compendium of 155 RNA-seq experiments. We used our RNA-seq data and the 2011 version of the ImmGen dataset as input to the Inferelator to learn additional regulatory relationships (thus expanding the coverage of our network). The Inferelator can also provide further support for regulatory relationships that were learned from the knock out and ChIP data (complementary estimates of the strength, timing, and directionality of these interactions). We have previously shown that the Inferelator is an effective (top performing when compared to many alternative methods) general method for leveraging diverse data types, such as time-series and knockouts, to learn global transcriptional regulatory networks (Bonneau et al., 2007; Bonneau et al., 2006; Gilchrist et al., 2006; Madar et al., 2009; Madar et al., 2010; Marbach et al., 2012b). For a detailed description of the current method we refer the reader to (Madar et al., 2010).

The current version of the Inferelator is composed of two core methods that we have shown to be mutually reinforcing: time-lagged Context Likelihood of Relatedness (tlCLR) (Madar et al., 2010), an extension of the CLR method (Faith et al., 2007) that explicitly uses time series data alongside steady-state data for computing time-lagged mutual information, and the Inferelator 1 (Bonneau et al., 2006). which learns regulatory dynamics as well as network topology by explicitly using time-series data to parameterize a linear ordinary differential expression model using the elastic net (Zou and Hastie, 2005), an $l1$- and $l2$-norm constrained model selection method. The Inferelator takes as input a genome-wide data set of transcriptome

data (typically microarrays or RNA-seq), which can contain time-series data as well as steady-state perturbation data (e.g. knockouts), and outputs a ranked list of regulatory interactions based on confidence scores. We denote the Inferelator-generated scores for TF x regulating gene g as:

$$s(x \rightarrow g | RNAseq) = Inferelator(x \rightarrow g | RNAseq) \times sign(cor(x, g)).$$

Note that we multiplied each regulatory interaction confidence score by the sign of the correlation coefficient between the TF and the putative target gene to differentiate putative activating from repressing interactions (positive and negative scores, respectively).

Using the Inferelator to derive Th17 relevant networks from the Immgen public data (multiple immune lineages):

The version of Immgen data we use dates to March 2011 and has expression data for over 167 distinct immune cells or conditions (Heng and Painter, 2008), not-including the Th17 cell population that we examined herein. As with the RNA-seq transcriptome data, we used the Inferelator to score regulatory interactions:

$$s(x \rightarrow g | Immgen) = Inferelator(x \rightarrow g | Immgen) \times sign(cor(x, g)).$$

As specified above we multiplied confidence scores from the Inferelator by the correlation sign to indicate activating from repressing interactions (positive and negative scores, respectively).

Combining $\boldsymbol{TF \rightarrow target\ gene}$ scores from multiple data sources:

Regulatory network inference based on any single data source alone has strong limitations that are the result of 1) the many layers of regulation comprising biological regulatory networks, 2) systematic errors associated with individual data sources, and 3) methodological constraints. In our case, ChIP-seq for a single TF will inform us of direct regulatory interactions, but these interactions may or may not be functional. Knock out data, on the other hand, will identify regulatory interactions that are functional but may or may not be direct. Correlative and time series analyses based on large compendia of transcriptome data suffer high false positive rates, due primarily to identifiability problems. The latter, although having more false positives, can still provide a boost to regulatory interactions found by ChIP-seq and KO, and more importantly provide information about regulatory information for which ChIP or KO data is not available. In our integrative regulatory network inference scheme regulatory interactions with support from multiple data sources are typically higher in accuracy than even the most confident predicted regulatory edges derived from single data-types; this is the basis for the integrative score we describe below.

When combining regulatory network scores derived from disparate data types one faces several challenges. Two strategies for combining different metrics (where each metric is a separate score of a TF->target pair) are: 1) parametric approaches such as converting each metric to a similar numerical space or metric, such as p-values or Z-scores and then performing the appropriate meta-analysis to combine metrics, and 2) converting each metric to ranks and then averaging ranks across data/support types. Recently, ranked based methods proved effective in learning regulatory networks from complementary data-sources (Madar et al., 2010; Marbach et al., 2012a; Marbach et al., 2012b; Prill et al., 2010). Rank-based methods for combining disparate

measures are robust to cases where p-value (or other significance values) range over many orders of magnitude and differ in range dramatically between data sources. Here we used a relative rank (i.e. quantile) method for combining network metrics derived from four distinct data sources into a final network. Let $D_i \in [KO, ChIP, RNAseq, Immgen]$ be one of the data sources we integrate over, $S(D_i)$ be an $M \times N$ matrix with rows representing genes and columns TFs, and let each entry $s_{g,x}(D_i)$ hold the confidence score for TF x regulating gene g based on data type D_i, i.e.:

$s_{g,x}(D_i) = S(x \rightarrow g | D_i)$.

Note that for $Di = [KO, ChIP]$ most TFs were not measured and will thus have no regulatory information, i.e. columns in S that correspond to these 'missing' TFs will only have zero values. We then convert all non-zero confidence scores into quantile scores that range from zero (lowest confidence) to 1 (highest confidence), in a procedure that we describe below.

Let $Q(D_i)$ be an $M \times N$ matrix, with each entry, $q_{g,x}(D_i)$, equal to 1 minus the rank, in descending order, of the absolute confidence score $|s_{g,x}(D_i)|$, divided by the total number of non-zero scores in $S(D_i)$:

$$q_{g,x}(D_i) = \left(1 - \frac{rank\left(\left|s_{g,x}(D_i)\right|\right)}{|a \in S(D_i) : a \neq 0|}\right) \times sign(S_{g,x}(D_i))$$

Note that under this formulation $q_{g,x}(D_i) \in [-1,1]$, where negative scores indicate repression, positive scores activation, the absolute values indicate the confidence level. All zero confidence regulatory interactions from before remain zero after mapping to quantiles.

We can now proceed to combine results over combinations of data types. Let $D^* \subseteq [KO, ChIP, RNAseq, Immgen]$ indicate the subset of data sources. We defined the data-combined $M \times N$ score matrix $C(D^*)$, with each entry representing the combined scores over D^*, as:

$$c_{g,x}(D^*) = \sum_{D_i \in D^*} q_{g,x}(D_i)$$

In this manner we calculated the ranked regulatory interaction lists for each TF over every possible data combination (used in Figure 4B).

Although this rank-based approach is simple, one complication does exist: $q_{g,x}(D_i = ChIP)$ is always greater or equal to zero but can equally indicate activation or repression (as ChIP support for a regulatory interaction is in line with both a repression and activation). Thus, when calculating $c_{g,x}(D^*)$ for a data combination that included ChIP, we determined the sign of the ChIP score to be in line with the sign of $c_{g,x}(D^*)$ with all other data types except ChIP (e.g. for $c_{g,x}(D^* = [ChIP, KO, Rnaseq, Immgen])$, we first determined the sign of $c_{g,x}(D^* = [KO, Rnaseq, Immgen])$, and then added the ChIP score with the same sign. Note that under this integrative formulation a $TF \rightarrow target\ gene$ interaction that receives contradicting repressive or activation inputs from $KO, RNAseq, or\ Immgen$ data, also receives a lower confidence score (i.e. the null hypothesis used is that a coherent regulation does not exist, rather than, a regulation does not exist). This consideration of regulation sign significantly boosted performance for combinations that involved the more general transcriptome data of *RNAseq* and *Immgen* (data not shown).

Combining $\boldsymbol{TF \rightarrow target\ gene}$ scores over multiple TFs:

We compute a simple score that identifies genes regulated by many of the core Th17 TFs (BATF, IRF4, STAT3 ,c-MAF, and RORC), as these genes are more likely to be Th17 relevant. We used this simple score to prioritized genes for further study (Figure 3B). Given our integrated $TF\ x \rightarrow gene\ g$ score $c_{g,x}(D^*)$, and a combination of TFs, X =(BATF, IRF4, STAT3, c-MAF, RORC), we calculated multiple TF scores as:

$$c_{g,X}(D^*) = \sum_{x \in X} c_{g,x}(D^*)$$

These TF-sum scores correspond to the bars shown in figures 4B (for any data combination, D^*) and 4D (for $D^* = [ChIP, KO, RNAseq, Immgen]$). The majority of the Th17 relevant genes identified in our creation of the Th17 target benchmark consisted almost entirely of genes that are up regulated in Th17 cells. Therefore we use only positive network scores (activating) when calculating precision-recall with this benchmark. When repression scores are included, absolute performance is slightly decreased but the relative ranking of methods combinations is unaffected by inclusion of repressive network edges (showing that combining TFs and all four data types helps recover Th17 genes).

Comparison of a Rank-based meta-analysis with Fisher's Method for combining p-values:

The rank based approach described above is a non-parametric statistical method. We chose it as the distribution and type of scores derived from the four data sources and methods combinations vary by several orders of magnitude in scale (Fig S4C), and because the null hypotheses for each data type are different (e.g. a TF-gene binding does not exist for ChIP, and a TF-gene expression

dependency does not exist for KO RNA-seq), hampering the use of methods that assume p-values are generated by a similar distribution resulting from comparison to the same null hypothesis. To assess if our rank-based strategy was indeed more suitable than a parametric alternative we compared its performance to Fisher's method for combining p-values (Fisher, 1925). Pseudo z-scores from inferelator were converted to p-values assuming a normal distribution to allow the Fisher's method to be applied over all data sets. Let x denote a TF and g denote a putative target gene. Let $D_i \in [KO, ChIP, RNAseq, Immgen]$ be one of the data sources we integrate over, $P(D_i)$ be an $M \times N$ matrix with rows representing genes and columns TFs, and let each entry $p_{g,x}(D_i)$ hold the p-value for TF x regulating gene g based on data type D_i, i.e.: $p_{g,x}(D_i) = pvalue(x \rightarrow g|D_i)$.

We can now proceed to combine p-values over data sources using Fisher's method as follows. Let $D^* \subseteq [KO, ChIP, RNAseq, Immgen]$ indicate a subset of data sources to combine over. We defined the data-combined $M \times N$ test statistic matrix $T(D^*)$, with each entry representing the χ^2 test statistic as: $T_{g,x}(D^*) = -2 \sum_{D_i \in D^*} \ln\left(p_{g,x}(D_i)\right)$. We then used these test statistic scores to calculate the data integrated p-values assuming a χ^2 distribution with $k = 2|D^*|$ degrees of freedom, $T(D^*) \sim \chi^2(k)$.

We similarly combined $TF \rightarrow target\ gene$ p-values over TFs for a given data subset. Let X =(BATF, IRF4, STAT3, c-MAF, RORC), then the test statistic matrix is $T(D^*) = -2 \sum_{x \in X^*} \ln\left(p_{g,x}(D^*)\right)$, and the combined p-values can be calculated assuming a χ^2 distribution with $k = 2|X|$ degrees of freedom, $T(D^*) \sim \chi^2(k)$. Results of comparing Fisher's method to the ranked based method show that Fisher's approach can be better for combining p-values within a single data source (blue bars in Figure S4C correspond to combinations of single data-types where the performance of Fishers method is better or comparable to our rank based method).,

Our rank based method significantly out-performed Fisher's method when combining scores from different data sources where distributions of p-values vary (e.g. compare performance for the top performing full combination of all data , KCRI in Figure S4C).

Assigning peaks from multiple ChIP-seq experiments into putative Cis Regulatory Modules (pCRMs)

We clustered peaks of multiple TFs that co-localized over small genomic regions into putative Cis Regulatory Modules (pCRMs) (Chen et al., 2008). TFs peaks were joined into a single pCRM if the distance between their peak summits was less than 100 bp. Additional TF peaks were added to a growing pCRM if their summit lay within 100bp of any peak within that pCRM. Simplified pseudo-code for this method for grouping ChIP-seq peaks into pCRMs is presented below.

n = number of TFs to be clustered into pCRMs (number of ChIP-seq experiments)
d = user defined parameter, max distance of a summit to closest neighbor summit in pCRM (set to 100bp in this work).
S = a list. Each element S_i (i=1:N) is a vector of summits belonging to TF_i
M = the output list. Each element will correspond to a single pCRM

As input we have a list S of n vectors, one vector of summits per TF.

1. Combine and sort ALL summits from S into one ordered (5' to 3') vector S^{ord}.

Note that S^{ord} contains summits of multiple TFs ordered by their bp positions. We also store the name of the TF for each peak and the p-value of the peak in identically ordered vectors.

2. Coalesce all peaks into pCRMs

```
For i in 1:( (length(S^ord)-1) ) {
        # if next summit is less than d bp away
        If( (S^ord [i+1]- S^ord [i]) < d ) {
                Add peak i+1 to current pCRM list M
        } else {
                Initialize a new pCRM for peak i+1 in M
        }
}
```

De novo Motif detection:

TF binding DNA motifs were identified by the online version of MEME-ChIP de novo motif analysis under default parameters (Machanick and Bailey, 2011). For each TF we chose the best 500 peaks (highest -log10 p-value), focusing each motif search on the DNA sequence that spanning the 100bp centered at peak summit. For IRF4 and BATF motif analysis, peaks belonging to 4 sub-types of pCRMs were considered: BATF alone; IRF4 alone; BATF and IRF4 alone; and BATF, IRF4, plus additional ChIPed TFs (one or more of: STAT3, MAF, or RORC).

Differential ChIP: comparing ChIP-seq for TF-x in TF-y deficient mice

In order to test the extent of influence TF y has on the genomic binding distribution and strength of TF x, and to assess if a p air of TFs (x, y) acts cooperatively, we compared ChIP-seq for $TF\ x$ in wild-type to ChIP-seq of x in y deficient mice (knockout). Let M be the set of pCRMs determined based on ChIP-seq experiments for Batf, Irf4, cMaf, Stat3, and Rorc, and define the genomic start and end bp positions of each pCRM $m \in M$ as the extremum $5'$ and $3'$ bp positions of the individual peaks found in m. To control for indirect effects y may have on the binding profile of x (i.e. y deficient mice may have an altered expression level for x) we subdivided M into three subsets: M_x - the set of singleton pCRMs containing peaks only for x (here we do not expect y to have a direct influence on x), $M_{x,y}$ - the set of pCRMs containing peaks for both x and y but no other TF (here we aim to test if y has a direct effect on x), and $M_{x,y,+}$ - the set of pCRMs containing peaks for both x and y and at least one additional ChIPed TF (here we aim to test how much of the effect of y on x is dependent on other factors). We can now calculate the number of reads per million (RPM) found in m for each ChIP-seq experiment: x in y wild-type background, $rpm(x|y^{wt}, m)$, and x in y deficient mice, $rpm(x|y^{ko}, m)$. For

each pCRM we then determine the fold change in binding as:

$FC_m(x|y) = \log_2\left(\frac{rpm(x|y^{wt},m)}{rpm(x|y^{ko},m)}\right)$.

We computed significance scores based on a Poisson distribution with a dynamic background model, similar to the background model scheme employed by MACS (Zhang et al., 2008). This score was used in the volcano plots shown in Fig 2. This score accounts for the fact that some areas of the genome are generally more accessible and thus may collect more mapped reads irrespective of the ChIPed TF. Thus, we calculated lambda (the parameter in the Poisson distribution controlling the expected distribution of reads counts for a given region) based on the genome-wide number of reads or the local read count. We considered two cases depending on the sign of $FC_m(x|y)$. If $FC_m(x|y) \geq 0$, i.e. in the pCRM m there was a stronger binding signal for TF x under TF y wild-type conditions, then we define:

$\lambda_m^{dymamic} = \max(\lambda_m^{local}(x|y^{ko}), \lambda_{BG}^{global}(x|y^{wt}))$,

where $\lambda_m^{local}(x|y^{ko})$ is the number of RPMs found in the DNA region of m for x in y knock-out background , and $\lambda_{BG}^{global}(x|y^{wt})$ is the genome-wide expected RPM given x in y wild-type background.

We then define the significance of $FC_m(x|y)$ to be:

$p_m(x|y) = Poisson(n \geq rpm(x|y^{wt}, m); \lambda_m^{dymamic})$.

Conversely, if $FC_m(x|y) < 0$, then $\lambda_m^{dymamic} = \max(\lambda_m^{local}(x|y^{wt}), \lambda_{BG}^{global}(x|y^{ko}))$, and $p(x|y) = Poisson(n \geq rpm(x|y^{ko}, m); \lambda_m^{dymamic})$.

Identification of additional Th17 core TFs:

We developed a simple procedure to use our Inferelator networks to identify additional regulators that act similarly to core TFs (BATF, IRF4, STAT3, c-MAF, and RORC). Recently, a similarly motivated method to identify master regulator TFs has been shown to be successful (Carro et al., 2010; Lefebvre et al., 2010). To this end we defined a ranked reference list of Th17 relevant genes (targets of known Th17 TFs in our networks), and queried additional TFs target repertoire for significant overlap with this list. We generated this starting Th17-relevant reference gene list from the KO and ChIP-seq network surrounding BATF, IRF4, STAT3, c-MAF, and RORC. We then determined for each query TF (within the set of several hundred TFs in the Inferelator network model), a ranked target gene list from the Inferelator generated network scores, $s(x \rightarrow g | Immgen)$. We have previously shown that $TF \rightarrow target\ gene$ predictions made by the Inferelator are highly accurate for top ranking predictions, and thus construct this score so that it emphasizes top ranked regulatory interaction for these core TFs. We restrict this analysis (for each TF) the top 100 to 300 TF->target pairs ranked by Inferelator score. Then, for each TF x we can calculate the enrichment of these n top ranked target genes in the reference list of Th17 relevant targets. We used three metrics to determine the recovery performance significance: 1) area under curve of Precision Recall curves, 2) area under curve of Receiver Operator curve, and 3) Gene Set Enrichment analysis (Subramanian et al., 2005). All three methods return a value between 0 to 1 that determines the level of agreement between the ranked reference list and the TF top n Inferelator targets as move from top ranked predictions to the n'th prediction (0 no enrichment), 1 (full agreement; all n genes recovered first by the ranked reference list). To determine p-values for each metric we run 20,000 simulations with a random set of n genes. The geometric mean of the three distinct p-values was used as a

final score to rank TFs for further study. We chose n=200 as this value recovered the five positive control core TFs: BATF, IRF4, STAT3, c-MAF, and RORC, as top enriched TFs. This score was used to guide our iterative experimental design and was used to identify Hif1a and Fosl2, as well as several of the additional TFs, as high priority candidates for second and third rounds of additional ChIP-seq, KD and KO experiments.

Supplemental References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome biology *11*, R106.

Bonneau, R., Facciotti, M.T., Reiss, D.J., Schmid, A.K., Pan, M., Kaur, A., Thorsson, V., Shannon, P., Johnson, M.H., Bare, J.C.*, et al.* (2007). A predictive model for transcriptional control of physiology in a free living cell. Cell *131*, 1354-1365.

Bonneau, R., Reiss, D.J., Shannon, P., Facciotti, M., Hood, L., Baliga, N.S., and Thorsson, V. (2006). The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. Genome biology *7*, R36.

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

Carro, M.S., Lim, W.K., Alvarez, M.J., Bollo, R.J., Zhao, X., Snyder, E.Y., Sulman, E.P., Anne, S.L., Doetsch, F., Colman, H.*, et al.* (2010). The transcriptional network for mesenchymal transformation of brain tumours. Nature *463*, 318-325.

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.

Eberl, G., Marmon, S., Sunshine, M.J., Rennert, P.D., Choi, Y., and Littman, D.R. (2004). An essential function for the nuclear receptor RORgamma(t) in the generation of fetal lymphoid tissue inducer cells. Nat Immunol *5*, 64-73.

Faith, J.J., Hayete, B., Thaden, J.T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J.J., and Gardner, T.S. (2007). Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS biology *5*, e8.

Fisher, R.A. (1925). Statistical Methods for Research Workers (Edinburgh: Oliver and Boyd).

Gertz, J., Varley, K.E., Davis, N.S., Baas, B.J., Goryshin, I.Y., Vaidyanathan, R., Kuersten, S., and Myers, R.M. (2012). Transposase mediated construction of RNA-seq libraries. Genome Res *22*, 134-141.

Gilchrist, M., Thorsson, V., Li, B., Rust, A.G., Korb, M., Roach, J.C., Kennedy, K., Hai, T., Bolouri, H., and Aderem, A. (2006). Systems biology approaches identify ATF3 as a negative regulator of Toll-like receptor 4. Nature *441*, 173-178.

Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol *11*, R86.
Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.
Heng, T.S., and Painter, M.W. (2008). The Immunological Genome Project: networks of gene expression in immune cells. Nature immunology *9*, 1091-1094.
Ivanov, II, McKenzie, B.S., Zhou, L., Tadokoro, C.E., Lepelley, A., Lafaille, J.J., Cua, D.J., and Littman, D.R. (2006). The orphan nuclear receptor RORgammat directs the differentiation program of proinflammatory IL-17+ T helper cells. Cell *126*, 1121-1133.
Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497-1502.
Karreth, F., Hoebertz, A., Scheuch, H., Eferl, R., and Wagner, E.F. (2004). The AP1 transcription factor Fra2 is required for efficient cartilage development. Development *131*, 5717-5725.
Kirigin, F.F., Lindstedt, K., Sellars, M., Ciofani, M., Low, S.L., Jones, L., Bell, F., Pauli, F., Bonneau, R., Myers, R.M.*, et al.* (2012). Dynamic microRNA gene transcription and processing during T cell development. J Immunol *188*, 3257-3267.
Klein, U., Casola, S., Cattoretti, G., Shen, Q., Lia, M., Mo, T., Ludwig, T., Rajewsky, K., and Dalla-Favera, R. (2006). Transcription factor IRF4 controls plasma cell differentiation and class-switch recombination. Nat Immunol *7*, 773-782.
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.
Lee, C.K., Raz, R., Gimeno, R., Gertner, R., Wistinghausen, B., Takeshita, K., DePinho, R.A., and Levy, D.E. (2002). STAT3 is a negative regulator of granulopoiesis but is not required for G-CSF-dependent differentiation. Immunity *17*, 63-72.
Lefebvre, C., Rajbhandari, P., Alvarez, M.J., Bandaru, P., Lim, W.K., Sato, M., Wang, K., Sumazin, P., Kustagi, M., Bisikirska, B.C.*, et al.* (2010). A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. Mol Syst Biol *6*, 377.
Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696-1697.
Madar, A., Greenfield, A., Ostrer, H., Vanden-Eijnden, E., and Bonneau, R. (2009). The Inferelator 2.0: a scalable framework for reconstruction of dynamic regulatory network models. Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Conference *2009*, 5448-5451.
Madar, A., Greenfield, A., Vanden-Eijnden, E., and Bonneau, R. (2010). DREAM3: network inference using dynamic context likelihood of relatedness and the inferelator. PloS one *5*, e9803.
Marbach, D., Costello, J.C., Kuffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R., Aderhold, A., Allison, K.R., Bonneau, R.*, et al.* (2012a). Wisdom of crowds for robust gene network inference. Nature methods.

Marbach, D., Roy, S., Ay, F., Meyer, P.E., Candeias, R., Kahveci, T., Bristow, C.A., and Kellis, M. (2012b). Predictive regulatory models in Drosophila melanogaster by integrative inference of transcriptional networks. Genome research.
Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J.*, et al.* (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521-533.
Morita, S., Kojima, T., and Kitamura, T. (2000). Plat-E: an efficient and stable system for transient packaging of retroviruses. Gene Ther *7*, 1063-1066.
Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods *5*, 621-628.
Ouyang, Z., Zhou, Q., and Wong, W.H. (2009). ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America *106*, 21521-21526.
Park, H., Li, Z., Yang, X.O., Chang, S.H., Nurieva, R., Wang, Y.H., Wang, Y., Hood, L., Zhu, Z., Tian, Q.*, et al.* (2005). A distinct lineage of CD4 T cells regulates tissue inflammation by producing interleukin 17. Nat Immunol *6*, 1133-1141.
Prill, R.J., Marbach, D., Saez-Rodriguez, J., Sorger, P.K., Alexopoulos, L.G., Xue, X., Clarke, N.D., Altan-Bonnet, G., and Stolovitzky, G. (2010). Towards a rigorous assessment of systems biology models: the DREAM3 challenges. PloS one *5*, e9202.
Reddy, T.E., Gertz, J., Pauli, F., Kucera, K.S., Varley, K.E., Newberry, K.M., Marinov, G.K., Mortazavi, A., Williams, B.A., Song, L.*, et al.* (2012). Effects of sequence variation on differential allelic transcription factor occupancy and gene expression. Genome Res *22*, 860-869.
Rutz, S., Noubade, R., Eidenschenk, C., Ota, N., Zeng, W., Zheng, Y., Hackney, J., Ding, J., Singh, H., and Ouyang, W. (2011). Transcription factor c-Maf mediates the TGF-beta-dependent suppression of IL-22 production in T(H)17 cells. Nat Immunol *12*, 1238-1245.
Ryan, H.E., Poloni, M., McNulty, W., Elson, D., Gassmann, M., Arbeit, J.M., and Johnson, R.S. (2000). Hypoxia-inducible factor-1alpha is a positive factor in solid tumor growth. Cancer Res *60*, 4010-4015.
Schraml, B.U., Hildner, K., Ise, W., Lee, W.L., Smith, W.A., Solomon, B., Sahota, G., Sim, J., Mukasa, R., Cemerski, S.*, et al.* (2009). The AP-1 transcription factor Batf controls T(H)17 differentiation. Nature *460*, 405-409.
Simon, J.M., Giresi, P.G., Davis, I.J., and Lieb, J.D. (2012). Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. Nat Protoc *7*, 256-267.
Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.
Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.
Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.

Wang, X., Zhang, Y., Yang, X.O., Nurieva, R.I., Chang, S.H., Ojeda, S.S., Kang, H.S., Schluns, K.S., Gui, J., Jetten, A.M.*, et al.* (2012). Transcription of Il17 and Il17f is controlled by conserved noncoding sequence 2. Immunity *36*, 23-31.
Wende, H., Lechner, S.G., Cheret, C., Bourane, S., Kolanczyk, M.E., Pattyn, A., Reuter, K., Munier, F.L., Carroll, P., Lewin, G.R.*, et al.* (2012). The transcription factor c-Maf controls touch receptor development and function. Science *335*, 1373-1376.
Ye, T., Krebs, A.R., Choukrallah, M.A., Keime, C., Plewniak, F., Davidson, I., and Tora, L. (2011). seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res *39*, e35.
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 biology *9*, R137.
Zhou, L., Chong, M.M., and Littman, D.R. (2009). Plasticity of CD4+ T cell lineage differentiation. Immunity *30*, 646-655.
Zhou, L., Ivanov, II, Spolski, R., Min, R., Shenderov, K., Egawa, T., Levy, D.E., Leonard, W.J., and Littman, D.R. (2007). IL-6 programs T(H)-17 cell differentiation by promoting sequential engagement of the IL-21 and IL-23 pathways. Nat Immunol *8*, 967-974.
Zhou, X., Sumazin, P., Rajbhandari, P., and Califano, A. (2010). A systems biology approach to transcription factor binding site prediction. PloS one *5*, e9878.
Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. J Roy Stat Soc B *67*, 301-320.